IntechOpen

Risk Management for the Future

Theory and Cases

Edited by Jan Emblemsvåg

RISK MANAGEMENT FOR THE FUTURE – THEORY AND CASES

Edited by **Jan Emblemsvåg**

Risk Management for the Future - Theory and Cases
http://dx.doi.org/10.5772/1809
Edited by Jan Emblemsvag

Contributors

Luca Franzi, Assed Naked Haddad, Claudia R. V. Morgado, Erick Galante, Rafaell De Oliveira Pinto Caldas, Irina Voronova, Roland Iosif Moraru, Eivind Lars Rake, Ganthan Narayana Samy, Rabiah Ahmad, Zuraini Ismail, Deck, Thierry Verdel, Abd Rahman Ahlan, Yusri Arshad, Lu, Andre, Roberte Manigat, Jochen Seidel, Paul Dostal, Florian Imbery, Roberta Pellegrino, Nicola Costantino, Frank Zwißler, Marco Hermann, Nisanci, Patrick Brockett, Linda Golden, Whitley Wolman, Dragutin Vukovic, Hayette Gatfaoui, Giovanni Improta, Antonio Fratini, Maria Triassi, Gianfranco Di Nino, Anna Castagnoli, Rita Maria Melotti, Marco Adversi, Grazia Innocenti, Panagiotis Marhavilas, Dimitrios E. Koulouriotis

Notice

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in Croatia, 2012 by INTECH d.o.o.
eBook (PDF) Published by IN TECH d.o.o.
Place and year of publication of eBook (PDF): Rijeka, 2019.
IntechOpen is the global imprint of IN TECH d.o.o.
Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

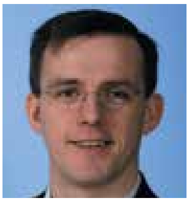
Risk Management for the Future - Theory and Cases
Edited by Jan Emblemsvag

p. cm.

ISBN 978-953-51-0571-8

eBook (PDF) ISBN 978-953-51-5111-1

Meet the editor



Dr Jan Emblemsvåg is Senior Vice-President of Innovation and Process Management at STX OSV AS – one of the major, global offshore shipbuilders in the world. He is responsible for improving the shipbuilding process across eight yards and several product companies. He is a Professor II at Ålesund University College, gives lectures and serves as a management consultant if time permits. He holds his Ph.D. in engineering design, but his research and work is inter-disciplinary ranging from cost management, strategy/risk management, general management to business development. He has published three books – the most notable is one on Activity-Based Life-Cycle Costing published by John Wiley & Sons. He has also published more than dozen journal papers on a variety of subjects.

Contents

Preface

If you think predicting the future is risky, try ignoring it.

The Economist

Risk management is a topic on the agenda of an increasing number of organizations around the world for the last 20 years or so. In fact, due to the large number of corporate scandals, risk management has become central in the boardrooms of large enterprises around the world as some stock exchanges in fact demand risk management in the corporate governance work. Despite this, we have a financial crisis that abundantly illustrated that risks were not properly understood – also in corporations that supposedly were conducting risk management.

While risk management in corporate governance is a relatively new idea, we have been managing risk in engineering for decades. Yet, engineering disasters appears every now and then often indicating (*posteriori*) lack of, or at least insufficient, risk management. There are many other cases in all aspects of human society that could have been mentioned here, as well, but the point is that managing risks is difficult.

This illustrates further important facts about risk – it is pervasive, it is timeless and it is inevitable. The pervasiveness and timelessness of risk means that it is found in all kinds of scholarly disciplines and human endeavors. An important side effect is that it is often slowly emerging, which makes it even harder to address – disasters are rarely due to a single mistake or single source of problems, but due to a complex interplay of factors that by themselves may not have resulted in a disaster.

Furthermore, because it is inevitable, risk has been addressed in a large number of ways. This means that basic terminology is still not unified. Depending on whom you ask, and what background they have, you will get different definitions and approaches towards risk management. The ISO 31000 Risk Management standard has therefore been developed to provide principles and generic guidelines on risk management (without intending to promote uniformity of risk management across organizations). Yet, many find the standard unsatisfactory and therefore find their own ways towards risk management. In this book, we therefore present a flavor of current advances in risk management theory as well as some cases with no attempt to present a unified theory of risk management.

The book is divided into four, broad topics – each covering an entire part of the book. The first topic is Health, Safety and the Environment (HSE) in which we have seven contributions. The opening chapter is written by R.I. Moraru and it concerns the identification of effective practices, processes and structures in occupational health and safety risk management. The authors identifies and argues that there is an urgent need for the formulation and implementation of a new management framework for occupational hazards; one that is appropriate for the new economic and occupational structure of work.

Next, in Chapter 2, A. Haddad, E. Galante, R. Caldas and C. Morgado focus on the development and usage of a risk assessment methodology called Hazard Matrix (HM) and its application in Health, Safety and Environmental Management (HSE). The HM is a prioritization methodology suitable to be used in the analysis phase of a risk management program. The authors argue that the HM in HSE is a very powerful methodology to highlight critical hazards and sectors/areas in a business unit or company under study.

In Chapter 3, P. K. Marhavilas and D.E. Koulouriotis present a new risk assessment framework based on the combination of the deterministic FTA ("fault-tree-analysis") technique and the stochastic TRF ("time at risk failure)" model, and they apply it on an industrial worksite to test its usefulness.

Then, in Chapter 4, G. Improta, A. Fratini and M. Triassi present an example on a possible design and implementation of a Health Technology Assessment (HTA) protocol for the classification of hospitals or health facilities equipment, realized by combining the classic HTA concepts with hierarchic clustering techniques in a multidisciplinary analysis of requirements, cost, impact of logistics, technology associated risks.

Chapter 5 is written by R. Manigat, F. Allix, C. Frochot and J.C. André. They chose to develop a case study on nanomedicine based on nanotechnology, with integrated inputs from each individual of the multidisciplinary team (photo chemist conducting research in basic sciences, risk management specialist, public health medical specialist), in order to develop an interdisciplinary expertise open to large societal needs.

The objective of the 6th chapter, written by A. Castagnoli, M. Adversi, G. Innocenti, G.F. Di Nino and R. M. Melotti, is to update the state of the art on Post-Operative Residual Curarization (PORC) and risk management of patients with persistent neuromuscular blockade. They start by careful reviewing the literature using electronic databases, analyzing original papers, systematic reviews and guidelines and end up by suggesting possible ways to correctly prevent or manage PORC.

The final chapter in the first section – Chapter 7 – is written by E.L. Rake. It describes the assessments on-scene, the arena where the crisis take place, especially assessment carried out by incident commanders and other professional leaders of emergency response units; the police, paramedics and fire brigade. The chapter gives insight in how risk assessment on-scene is performed and how effective risk assessment can be carried out in real time while the crisis unfolds on-scene.

The second part of the book, Part II, concerns Engineering. Here we have five chapters focusing largely on issues pertinent to geology and civil engineering, although there should be good thinking for other engineers as well. The first chapter (Chapter 8) in Part II concerns classic issues like uncertainty and risk. The authors – O. Deck and T. Verdel – focus on clarifying the interactions between risk management and uncertainties within the context of geohazards. Recent trends developed in the field of risk management within the context of mining subsidence hazards, are also discussed.

R. Pellegrino and N. Costantino have written Chapter 9. Here, they develop an approach to analyze real options in real world investment opportunities. It combines two well-known techniques, namely the Monte Carlo simulation for real option pricing and the fuzzy-Delphi method for eliciting probabilistic input parameters, when historical data are missing, from the knowledge of even more than one expert in a consistent, structured and transparent way.

Chapter 10 provides a case from Turkey written by R. Niscanci, V. Yildirim, Y.S. Erbas where the city center of Trabzon was selected as the pilot area for the establishment of a sample fire database based on Geographic Information System (GIS) and as the basis of sample spatial queries in support of fire management. Specifically, an analysis of fire hydrant location was carried out and the related needs were identified.

From fire in Chapter 10, we move to river flooding in Chapter 11. L. Franzi provides in this chapter a concept of Flood Risk Management (FRM) with the aim of replacing the earlier and narrower paradigms of flood defense and flood control. The aim is to show and discuss the state-of-the–art as well as provide a more in-depth description of the FRM relating to the Northern part of Italy. It will be shown, in particular, that the effectiveness of the applied FRM strategies strongly depends on the uncertainties in the flood risk assessment. As a consequence, FRM strategies should be enough flexible to adapt to new circumstances and evidences, taking into account a good balance between planning and civil protection.

Chapter 12 also concerns flood risk management. J. Seidel, P. Dostal and F. Imbery present a case study of the Neckar Catchment in southwest Germany where different methods are used applied to reconstruct and analyze two historical flood events in 1824 and 1882. These results were then used to extend the data series for a gauging station in the Neckar River where modern discharge data exists from 1921 and onwards. In total, the authors illustrate how this information can be used to produce more stable calculation of return times and river discharge characteristics.

Then, in Part III, we change topicality radically and enter the world of Information Management. Here, we have four contributions. The first is made by A.R. Ahlan and Y. Arshad in Chapter 13. Here, they perform a thorough literature review to synthesize the risk factors associated with information technology (IT), or information system (IS), and subsequently categorize or classify them into a few main major themes to guide IT management in managing their risks.

In Chapter 14, P.L. Brockett, L.L. Golden and W. Wolman focus on enterprise cyber risk management and risk mitigation (as opposed to individual consumer cyber risk, which is not addressed in this chapter). They investigate cyber risks including information theft, compromise of consumer information, and the interruption of e-commerce and how these risks affect the economics and security of organizations.

With the development of internet technologies, transfer and storage procedures are becoming more asynchronous, and this introduces new risks in its own right. In Chapter 15, D. Vuković addresses this challenge and investigates what this means in terms of trust in the system and what we can do to the system infrastructure to increase its security and thereby trust. Basically, "could we envision a model for distributed computer system which would foster sociological notions of trust and confidence within the infrastructure?"

In Chapter 16, G.N. Samy, R. Ahmad and Z. Ismail introduce a new method for analyzing information security risk. They adopt a medical approach namely survival analysis and adapting the overall risk management process. Under survival analysis approach, a method which is known as Cox Proportional Hazards (PH) Model can be applied to identify significant information security threats. The overall risk management process is based on ISO 31000:2009.

Our final topic in Part IV is broadly defined as Finance and Economics. Z. Lu and Y. Zhuang start this part of the book with a technical chapter concerning the Capital Asset Pricing Model (CAPM) and how the beta risk is linked to the market condition as measured by the market volatility as modeled in the CAPM. This is a particularly interesting topic in the light of the recent interest in the large and unexpected swings in asset values.

From Chapter 17, Chapter 18 follows quite naturally as H. Gatfaoui assess the impact of the stock market trend on the credit market trend while describing also how the magnitude of stock market moves impacts the magnitude of credit market moves. The importance of this assessment is evident from the recent mortgage subprime crisis and the partly resulting global financial crisis which partly illustrate the weaknesses of prevailing risk management practices where Credit Default Swaps (CDS) or corporate bond spreads become highly sensitive to the stock market trend and/or the corresponding market volatility.

I. Voronova investigates financial risks in the context of non-financial, small and medium-sized enterprises (SME) in Chapter 19. For SMEs the principle of KISS (Keep It Simple, Stupid) are important. The application of these principles in relation to the choice of the methods of financial risks assessment means that mainly simple methods should be used. The author evaluates the development in SMEs in nine East European countries concerning the usage of discriminant and conditional probability methods to assess, predict and manage risks related to liquidity, credit, decreasing financial stability and insolvency/bankruptcy.

Since supply chains are very large systems with a great number of economic transactions, the book is closed off with a chapter that focuses on supply chains. In this final chapter, Chapter 20, F. Zwißler sets out to define basic terms in supply chain risk management before presenting the results of a survey from 2010. From this, he introduces an approach for identifying, assessing, and managing risks in a supply chain, particularly to help SMEs with risk management.

In the *Hitchhiker's Guide to the Galaxy*, Vroomfondel states that "We demand rigidly defined areas of doubt and uncertainty". These rigidly defined areas, constituting science and engineering, have since the Renaissance undoubtedly produced great results in many avenues of human civilization. However, I cannot free my mind from Peter Bernstein's ascertainment that risk management approaches have led us as society to take risks we would otherwise not have embarked upon. It seems that good judgment is always needed and that risk management will always have an element of art.

As editor of the book, I hope you find all these chapters and pages to your satisfaction and a good source of new ideas and fresh thinking to help you in *your* thinking and practice. May we all keep in mind Albert Einstein's cautious words;

Concerns for man and his fate must form the chief interest of all technical endeavors. Never forget this in the midst of your diagrams and equations.

Jan Emblemsvåg
STX OSV AS and Ålesund University College
Norway

Section 1

Health, Safety and the Environment

1

Current Trends and Future Developments in Occupational Health and Safety Risk Management

Roland Iosif Moraru
University of Petroşani
Romania

1. Introduction

Occupational safety and health (OHS) like all facets of business, needs to be properly managed. A company's OSH system helps ensure effective control of OHS risks and continual improvement in OHS performance, prevent work-related illness or injury and to achieve compliance with regulations and standards.

The goals of this chapter are 1) the identification of effective practices, processes and structures in OHS risk management, and 2) using a simple framework to draw together what is known of good and bad practice in this area, particularly in deciding what rules should be explicitly formulated and imposed. We argue that there is an urgent need for the formulation and implementation of a new management framework for occupational hazards; one that is appropriate for the new economic and occupational structure of work. The overall objective is 1) to underpin observations, 2) illustrate typical characteristics of the current situation and 3) indicate directions that could lead to solving these new safety problems. We suggest that this task should initially involve stepping back and revisiting the frame of reference in which the protection against occupational injury is viewed.

In approaching the issue, the chapter, first, attempts to provide a succinct mapping of the environment of occupational risk, through a brief examination of its historical dimensions.

Based on a thorough literature review, the major role of the ISO 31000:2009 standard is emphasized. Given that risk management is an adaptive process and that risk assessment is merely one of its features, the question is what can risk managers do to make their activities more credible and acceptable? A section is devoted to benchmarking organizational practice and risk treatments. This focus also raised the discussion of drawbacks and pitfalls of risk ranking methods. The chapter pays special attention to developing a new understanding of the participatory approach and closes with a comparative analysis which seek further explanation of approaches to occupational health and safety risk management based on two kinds of epistemological assumptions existing in the field, namely constructivism and positivism. This work should assist those practitioners, researchers and other stakeholders within industry who are interested in assessing and managing the existing OHS risks in their organisations, with the intention of identifying the priority areas for focussing improvement effort.

2. Risk management: The need for an evolutionary and multifaceted approach

Regardless of the type and size, any organization faces risks that can affect the achievement of its goals. Therefore, acquiring a coherent system of concepts and rules, generally accepted nationally and internationally, becomes essential for the public and private sector today, regardless of their nature. The approaches to safety seem to put emphasis on management functions, guidelines, industry standards, quality principles, to establish the safety management system, as outlined bellow.

2.1 Premises and brief history

All the activities of an organization involve risks and risk management is the foundation of the decision-making, considering the effects of uncertainty on the objectives. Companies that have applied risk analysis and management for many years also recognize that the change to a "culture of prevention" via "systematic and comprehensive risk management" involves a journey (Hudson, 2003). The model shown in Figure 1 suggests that a move towards an integrated risk management system is multifaceted and evolutionary. As pointed out by (Joy and Griffiths, 2004) the key for success is for companies "to select the method that is designed to suit their needs". They also need to understand the challenges related to risk management of the company. Stakeholders must understand where important decisions are requiring risk to be systematically considered, as well as the current status of their culture or systems, so the next step can be triggered.

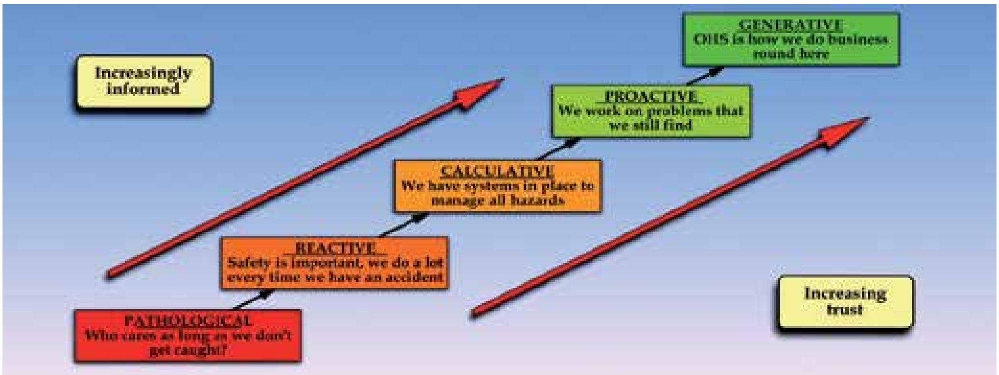

GENERATIVE
OHS is how we do business round here
Increasingly informed
PROACTIVE
We work on problems that we still find
CALCULATIVE
We have systems in place to manage all hazards
REACTIVE
Safety is important, we do a lot every time we have an accident
Increasing trust
PATHOLOGICAL
Who cares as long as we don't get caught?

Fig. 1. Multi-faceted and evolutional journey toward risk management (Adapted from Ayers, 2007)

A discussion of the major challenges related to development of causal models of organizational safety performance and a set of principles to address them have been presented by several authors in separate publications, see for example (Bourrier, 1998; Haines *et al*, 2002; Reason, 1995). The conceptual models proposed for organizational safety performance are naturally heavily influenced by the particular theoretical perspective adopted and the objectives chosen for the model. For example, literature on safety culture (Cox and Flin, 1998) and safety climate, such as (Zohar, 1980; Zohar and Luria, 2004) focuses primarily on the psychological causes of safety, with perception survey as the main measurement method. On the other hand, safety management literature including (Walters *et al*, 2005) primarily considers organizational safety structure and practices using auditing measurement approaches. Yet other disciplines (e.g. Preliminary Risk Analysis) mainly focus on direct causes of accidents such as hardware failures or operational errors, and on a common metric for measuring them (Sage and White, 1980; Reason, 1993).

The best state of health, safety and well-being for the workers and of physical and economic health for the company cannot be reached in once. Effective systems are based on the principle of "Plan – Do – Check - Act" (Deming, 1982). In OSH terms this will require to develop a policy on what is intended to achieve, then a plan of how and when it will be done, including any necessary arrangements. Next is the "doing" phase, when plans are implemented and then check that you have done what you planned to do and that it is effective in controlling risks. Any deficiencies found need to be acted upon and rectified, so that the system performance improves continually. Numerous management practices and processes include elements of risk management and several organizations are resorting to formal management processes for specific circumstances and particular risks, as depicted in Figure 2 (Smith, 2008).

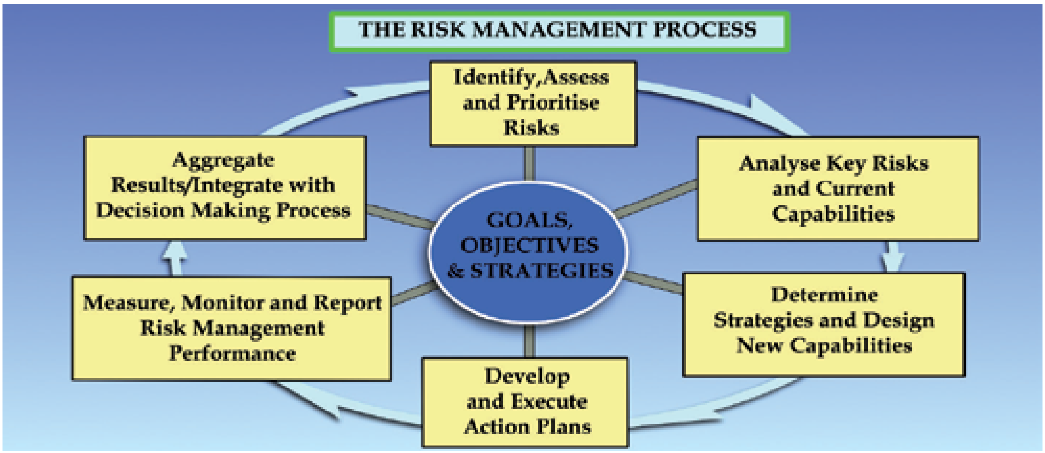

THE RISK MANAGEMENT PROCESS
Identify,Assess and Prioritise Risks
Analyse Key Risks and Current Capabilities
Determine Strategies and Design New Capabilities
Develop and Execute Action Plans
Measure, Monitor and Report Risk Management Performance
Aggregate Results/Integrate with Decision Making Process
GOALS, OBJECTIVES & STRATEGIES

Fig. 2. Risk Management Process – Marsh Perspective (Adapted from Smith, 2008)

Over time, more than 60 Technical Committees and Working Groups of ISO and national standards or regulatory bodies, have addressed the risk management issue one way or another. Numerous multi-sectoral standards have been drafted, e.g. OHSAS 18001, BS 8800, FD X50-252:2006, ISO/IEC 51 guide etc or dedicated to a particular sector. While the AS/NZS 4360:2004 (published in 1995 and amended in 1999) was the most widely used global standard for risk management. Considering the need for unification, Australia proposed to set up a Working Group on Risk Management, aimed at providing practical guidance on the risk management principles for all applications, including small and medium enterprises. Before the first meeting of the Working Group, it was developed a discussion document, based on AS/NZS 4360:2004 (Standards Australia and Standards New Zealand, 2004), using the terminology of ISO/IEC 73. The 20 delegates from 12 national standards associations attended the first meeting held in September 2005 in

Tokyo. The following Working Group meetings were held in Sydney (February 2006), Vienna (September 2006), Ottawa (April 2007), Sanya (December 2007), Singapore (November 2009). Voting began on May 25, 2009 and ended in July 25, 2009; in November 2009, the ISO 31000 standard was issued.

In order to better highlight the evolutionary and multifaceted character of the risk management conceptual models, a brief comparison of AS/NZS 4360 and ISO 31000:2009 standards is performed in the following section.

2.2 AS/NZS 4360 and ISO 31000:2009: A comparison

Basically, we can argue that ISO 31000 is the natural successor to AS/NZS 4360:2004 and although the comparative analysis of the two standards is not the purpose of this work, we consider necessary and useful to highlight the differences. The main elements are shown in Table 1 and the basic terms are defined in Table 2.

Because we are describing a holistic process, the scope of this section is greater than that of some documents which deal with limited scope of the topic. In particular, many texts deal only with the analytical processes of risk assessment, and omit the management and organizational aspects of their implementation. The steps of the risk management process which are often omitted are 1) establishing the context, 2) monitoring and review, and 3) communication and consultation. This trend is particularly valid for the field of OHS risks.

Elements	**AS/NZS 4360:2004**	**ISO 31000:2009**
Application	All organizations, all risks – no exclusion. Australia and New Zealand	All organizations, all risks – no exclusion. All countries
Context for risk management	The organization's objectives	The organization's objectives
Risk Management Process („What you do?")	Core of AS/NZS 4360	Part of ISO 31000
Risk Management Framework („How you do?")	Substantially revised în 2004	Extension of AS/NZS 4360
Risk management principles	Implicitly approached, to some extent	Explicitly and clearly approached
Attributes of enhanced risk management	Not approached	Annex to ISO 31010:2010

Table 1. AS/NZS 4360:2004 and ISO 31000:2009: Differences regarding the main constituents

Term	AS/NZS 4360:2004	ISO 31000:2009
Risk	The chance of something happening that will have an impact on objectives	Effect of uncertainty on objectives
Risk management	The culture, processes and structures that are directed towards realising potential opportunities whilst managing adverse effects	Coordinated activities to direct and control an organization with regard to risk
Risk management framework	Set of elements of an organization's management system concerned with managing risk	Set of components that provide the foundations and organizational arrangements for designing, implementing, monitoring, reviewing and continually improving risk management processes throughout the organization
Risk management policy	Not defined	Overall intentions and direction of an organization related to risk management
Risk management plan	Not defined	Document within the risk management framework, specifying the approach, the management components and resources to be applied to the management of risk
Risk management process	The systematic application of management policies, procedures and practices to the tasks of communicating, establishing the context, identifying, analysing, evaluating, treating, monitoring and reviewing risk	Systematic application of management policies, procedures and practices to the tasks of communicating, consulting, establishing the context, identifying, analyzing, evaluating, treating, monitoring and reviewing risk

Table 2. AS/NZS 4360:2004 şi ISO 31000:2009: Basic definitions

The ISO 31000:2009 Standard is aimed at harmonizing the risk management processes. It is not a substitute for the existing standards, but a a top-level generic document providing a unified and coherent approach of the risk management principles and framework. Its purpose is to contribute to mutual understanding amongst stakeholders rather than provide guidance on risk management practices.

Any application or private sector requirements, brings its particular perceptions and individual criteria and, therefore, one of the key features of the standard is to include "establishing context" as the initial stage of the process, step that allows to "capture"the criteria diversity and complex nature of risks involved in each case. A brief outline of the major requirements of the standard is now needed to highlight its usefulness within the OHS risk management.

2.3 Snapshots from ISO 31000 standard "Risk Management – Principles and guidelines for implementation

The aforementioned standard emphasizes how the organization should understand the specific context in which risk management is implemented throughout the organization, at all levels and in every moment of its existence, to allow:

- Fosterage of proactive, rather than reactive, management;
- Awareness of the need to identify and address risks throughout the organization;
- Improvement of opportunities and threats identification;
- Compliance with relevant legislation and international standards;
- Improvement of corporate governance and stakeholder confidence;
- Solid base of planning and decision making processes;
- Better control systems, learning and organizational resilience, operational efficiency, safety and health, loss prevention and incident management;
- Effective allotment and use of resources.

To be effective risk management must become an integral part of governance, management, reporting processes, policies, philosophy and culture of the organization. As stated in Clause 3 of the standard, the risk management: a) creates value; b) is an integral part of organizational processes and part of decision making; c) explicitly adresses uncertainty; d) is systematic, structured and timely; e) is based on the best available information; f) is tailored; g) takes human and cultural factors into account; h) is transparent and inclusive; h) dynamic, j) iterative and responsive to change; k) facilitates continual improvement and enhancement of the organization (ISO, 2009).

The general framework (see Clause 4 of the standard) supports the organization to effectively manage risks, applying risk management process on different levels, in the specific context, at a given moment. This clause describes the necessary components of a risk management framework and how they inter-relate, as illustrated in Figure 3. The process involves the use of logical and systematic methods for continuous communication and consultation, defining the context for identifying, analyzing, evaluating and treating risks, as well as monitoring and reviewing risks. It includes activities described in the standard through requirements 5.2-5.6, as shown in Figure 4.

In the spirit of this new standard, one may decide to review the foundations of existing processes and practices regarding the OHS risk assessment and management. While it is not specific to a particular industry or sector, the standard can be applied to any organizational entity regardless of the type and nature of the risks. Despite this, the standard is not about promoting uniformity in risk management, because the design and implementation of the framework and management plans should take into account the specific needs of the organization's particular objectives, structure, operations, processes, functions, projects, products, services, goods, and specific practices employed. Some areas, as in OHS, are requiring regulatory criteria that reflect an „aversion" to the predominantly negative consequences of risk. Resorting to the approach proposed in the standard proposed enables the identification and application of such criteria. We argue that the standard supports organizations to comply with legislative requirements and international standards while increasing the performance of the organization. Unfortunately, this argument is not currently possible to qualify as the standard is so new that little reliable empirical research on its usefulness has been conducted

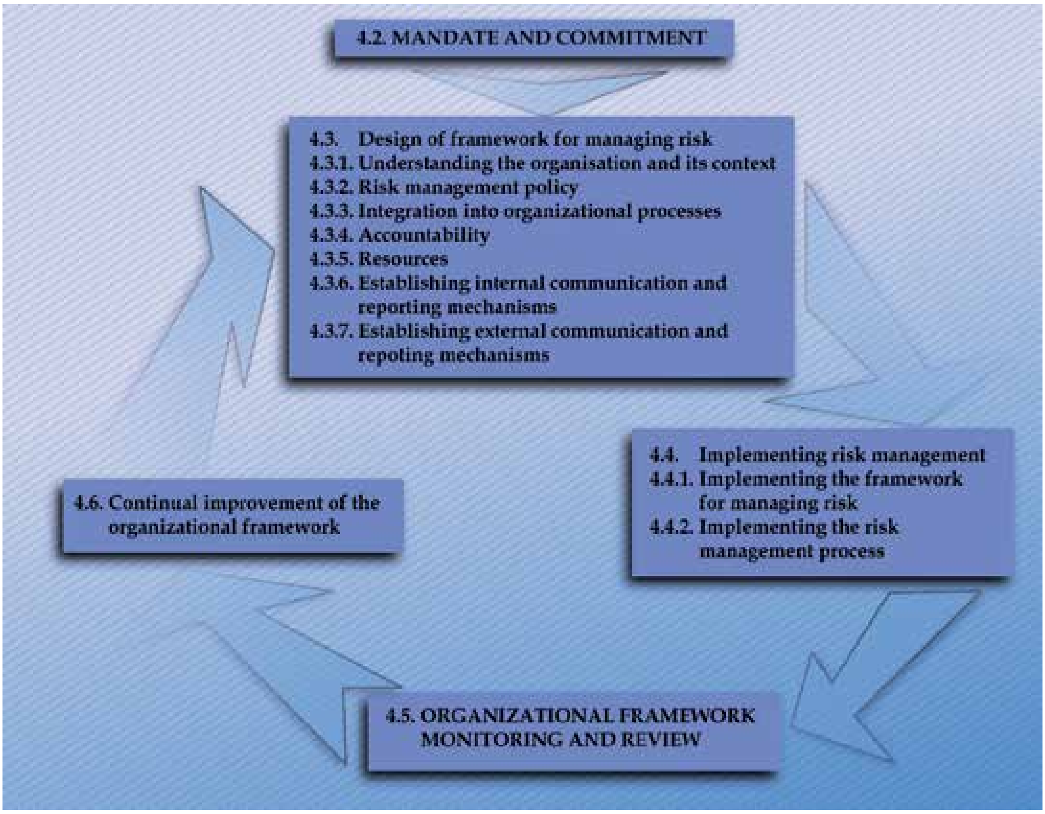

4.2. MANDATE AND COMMITMENT
4.3. Design of framework for managing risk
4.3.1. Understanding the organisation and its context
4.3.2. Risk management policy
4.3.3. Integration into organizational processes
4.3.4. Accountability
4.3.5. Resources
4.3.6. Establishing internal communication and reporting mechanisms
4.3.7. Establishing external communication and repoting mechanisms
4.4. Implementing risk management
4.4.1. Implementing the framework for managing risk
4.4.2. Implementing the risk management process
4.6. Continual improvement of the organizational framework
4.5. ORGANIZATIONAL FRAMEWORK MONITORING AND REVIEW

Fig. 3. Components of the framework for managing risk (Based on ISO 31000:2009)

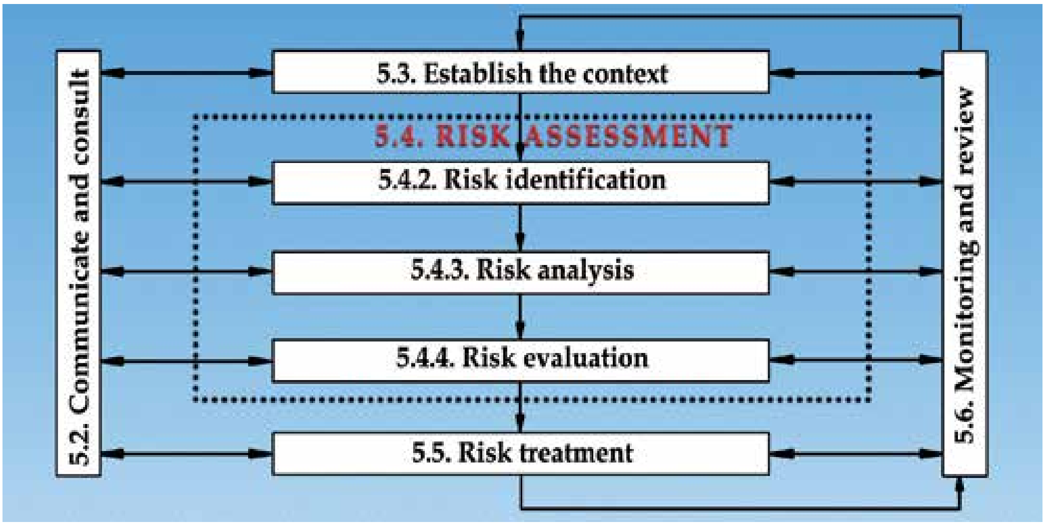

5.2. Communicate and consult
5.3. Establish the context
5.4. RISK ASSESSMENT
5.4.2. Risk identification
5.4.3. Risk analysis
5.4.4. Risk evaluation
5.5. Risk treatment
5.6. Monitoring and review

Fig. 4. The risk management process (Based on ISO 31000:2009)

Assessment of workplace risks is the foundation of a company's OHS risk management. Yet it is surprising how little literature there is about 1) how to conduct risk assessments effectively, 2) how to decide what methods and rules are needed, 3) how to prepare and formulate them and 4) how to promulgate them and ensure they stay appropriate. As such, we are now focusing on the practical application of the risk assessment process, representing a resource for getting up to speed quickly on the different options available and the means to introduce and implement risk management.

3. Current trends and challenges in OHS risk management

Risk assessments are vital support to decision-making process. Risk assessment supports the design review process by providing the underlying analysis on which safety decisions can be made. Risk assessment methods are being deployed in many industries, and the momentum is likely to continue. Although the level of sophistication in risk assessment processes varies the general risk assessment process applies both across and within all industries.

3.1 Occupational risk assessment: Benchmarks for the organization's practice

Modern risk assessment began over three decades ago, with applications in the military and nuclear power (Theys, 1991). In the late 1970s it gradually expanded, and was applied to a vast array of chemical risks. Applications to engineered systems, and in particular infrastructure, are common; examples are given by (Lave and Balvanyos, 1998). Blockley (1992) also devotes a number of chapters to civil engineering topics (e.g. design codes or risk assessment in structural engineering), and several infrastructure-engineering applications (e.g., dam safety, marine structures).

According to ISO 31000:2009 standard, risk depends both on the probability or frequency of an adverse outcome, and also on the severity of that outcome. Risk has similarly been defined generally as "the potential for realization of unwanted, negative consequences of an event" (Moraru and Băbuţ, 2010). More quantitatively, in (Sage and White, 1980) risk is defined as "the probability per unit time of the occurrence of a unit cost burden", and state that it "represents the statistical likelihood of a randomly exposed individual being adversely affected by some hazardous event". Thus, risk has been defined at many different levels of detail. The usage of of the word risk ususlly has negative connotations and risks are regarded as something to be minimized or avoided. The aforemetioned standard recognizes that activities involving risks may lead to impacts that can be positive as well as negative. The processes described herein can be used to exploit opportunities for enhancing organizational outcomes as well as reducing negative consequences.

Risk treatment efforts to achieve acceptable risk must work within the real world constraints of feasibility, practicality and cost. A practical solution to achieving acceptable risk is a good faith application of the hierarchy of controls within the risk assessment process. The number of methods aiming at assessing the risks is definitely greater than the number of methods aiming at preventing them.

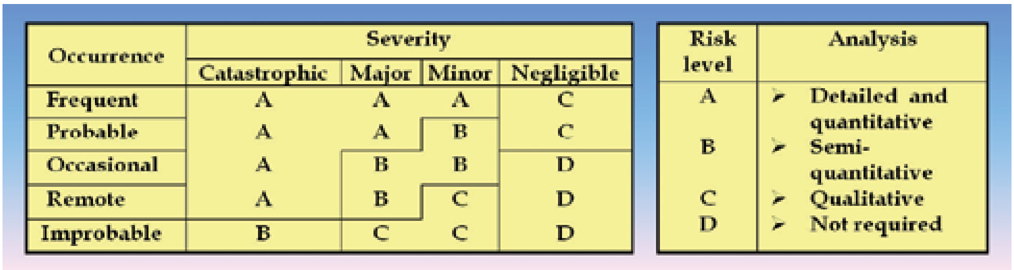
In Romania, since 2006, when the new Occupational Health and Safety Act (Romanian Parliament, 2006) have stated that the risk assessment is compulsory, several approaches were in use but only one method is extended in application. It appears as obvious that a large number of practitioners are resorting to a single method, without considering the great variety of working systems and conditions which are requiring specific approaches and techniques. Methods are used to rank risks and to define priorities for actions - which is desirable - but often this is done by neglecting the analysis of the elements defining these risks and the means of improving the situation. The accident risk management should be seen as the process of providing recommendations on whether to accept or resolve potential consequences of hazards associated with a given activity. It is neither a "science" (in that it provides a precise prediction of future events), nor just "common sense" or "something good managers have always done". It resorts to systematic procedures and specific techniques to analyze safety and occupational health factors, design and construction of equipment, and other situational hazards. As highlighted by Pasman (2009), for this process to be effective, the company culture must be willing to embrace the risk assessment process, and cultural acceptance stems from management leadership. Engineering design needs to change to include the risk assessment process to more effectively move safety into design.

Guidance on how to most effectively introduce the risk assessment process to an organization, and how to conduct them thereafter can be extracted from different sources, but the most valuable information source remains the practical experience gained by effectively performing the risk management. Practical guidance should be provided for Romanian companies get started and make progress in the risk assessment process. Topics addressed include: 1) the time to complete an assessment, 2) forming a team, 3) what to expect, 4) when to stop a risk assessment, 5) what to do in cross industry situations, 6) when to revise an existing risk assessment, 7) making changes to the protocol and 8) results of risk assessment.

When adressing the tool issue, „risk ranking matrix" is the term that describes how risks are ranked in the first instance, employing a method-specific tool. There are many variables, factors and combinations that must be considered in selecting an appropriate tool for further analysis, as that presented in Table 3. The different variables that are used to rank risks are requiring a proper understanding, and the three most common types of risk ranking systems are 1) qualitative, 2) semi-quantitative and 3) quantitative. Given the subjective nature of rating risk, risk scoring systems will likely continue to emerge and proliferate, as users refine and improve their risk assessment process.

This diversification of methods should be considered healthy, due to the variety of working circumstances requiring specific approaches. In time, convergence to one or a few risk scoring systems may occur, as efforts to harmonize and standardize risk assessment methods are made. This process will require some time, particularly in developing countries, as Romania, where the legal compliance is nowadays seen as the main requirement, instead of considering performance as main goal. There is also considerable resistance to creating risk assessment documents from the legal community primarily, due to product liability concerns and economic and financial restraints.

However, good engineering practice, continuous improvement and risk management requirements, all push for documenting processes. Documenting the risk assessment process is required or recommended by every guideline, standard or technical description of risk assessment. There are many risk ranking systems in use, each offering its strengths and weaknesses. This variation reflects the great diversity of opinion on risk assessment. Some of the most significant differences between risk assessment methods used today involve how risk is assessed. As mentioned before, there is a continuum of risk ranking systems from qualitative to quantitative that effectively address a variety of risk assessment applications. Very few benchmarks use quantitative risk ranking systems. However, there is no indication that any particular risk ranking system is better than another for all applications.



Occurrence	Severity			
	Catastrophic	Major	Minor	Negligible
Frequent	A	A	A	C
Probable	A	A	B	C
Occasional	A	B	B	D
Remote	A	B	C	D
Improbable	B	C	C	D

Risk level	Analysis
A	➢ Detailed and quantitative
B	➢ Semi-quantitative
C	➢ Qualitative
D	➢ Not required

Table 3. Matrix helping to identify hazard assessment method classes. Action Guide

Based on the above-mentioned benchmarks, the following principles directed towards practical risk assessment improvement can be stated:

- *Minimize the use of labels***:** the use of labels to describe portions of the risk assessment process should be minimized. The terms used in assessing risk were until the issue of the ISO 31000 Standard very confusing. There existed confusion, or at least no common understanding, as to the meanings of the terms of risk assessment, risk analysis, risk estimation, risk evaluation. Efforts at harmonizing, standardizing or even communicating were severely hampered by the past confusion and different uses of the term „risk assessment” and others. The practitioner trying to conduct a risk assessment does not care about terms or labels. He just wants to know what he need to do to complete an effective risk assessment. Extra terms detract from this objective. Unnecessary terms that add no value should be removed from the risk assessment process. Labels that provide no value only add confusion.
- *Simplify the risk assessment process:* the steps of the risk assessment process should be written using active verbs rather than labels or titles. The steps of the risk assessment process need to be simple and straightforward, and provide the reader very clear direction on what he needs to do.
- *Adopt "risk assessment process"* as overall term: the term „risk assessment process”should be adopted to describe the overall process of identifying hazards, assessing risk and reducing risk. The terms "risk analysis", "risk assessment", "risk estimation" and others have different definitions depending on the industry using them. The two most frequently used terms to describe the overall risk assessment effort are "risk assessment" and "risk management".
- *The risk assessment process goal involves risk treatment*: there is no point in assessing the risks of a system, design, process or product, unless one plans to perform risk reduction. The risk reduction effort is always completed even though not every residual risk requires further risk reduction (the risk may already be acceptable). This implies that risk reduction is a necessary part of, and should be included in, the overall risk management process regardless of the term used to describe that overall process.

- *Adopt the risk assessment process flow chart*, as it is given in clause 5 of the ISO 31000 Standard;
- *Subjective judgment needs to be accepted:* subjectivity is a necessary part of risk assessment. Even in quantitative risk assessments subjective judgment occurs. However, the subjectivity does not diminish the value or credibility of the risk assessment process. Safety is not an absolute state, but a relative one. Engineers, safety practitioners and decision makers need to become comfortable with subjectivity, and recognize that the subjective risk assessments do offer value.
- *Accept uncertainty:* uncertainty enters risk assessment as assumptions, estimates and subjective judgments and lack of precise information. Even in quantitative assessments there often remains substantial uncertainty. Performing a risk assessment does not create the uncertainty. Uncertainty is, and should be accepted as, an integral part of the risk assessment process.

With experience, the risk assessor learns which assessment tool is best for investigating a certain type of activity. The qualities of a good risk analysis are:

- Clear, concise, and a well-defined method that a reviewer or reader can readily understand.
- Orderly and consistent in systematically reviewing the activity or system for risk.
- A closed loop where the assessor reviews each risk control for its impact on the other risks and their controls.
- Objective in that reviewers and users can understand and verify each step of the assessment.

Paradoxically, one of the most important advantages of some risk assessment methods is meanwhile a handicap: a numerically expressed risk assessment. Of course, an approach based on figures allows prioritization, but will not be able to consider certain major aspects, such as ergonomics and psychosocial risks. Certain methods are difficult to apply, or even impossible, for chronic intoxication risk assessment, mental or physical fatigue assessment etc. In the following section, we review the relevant strengths and drawbacks of risk ranking methods, in particular the so called Kinney Method.

3.2 Advantages, limitations and subjectivity in resorting to the Kinney method of risk ranking

As outlined by Honings (2000), there is already a significant time lapse since the introduction of the Kinney method, and also to other similarly structured methods, self - denominated as risk „quantification" methods, are used on large scale in the field of OHS. The goal is nonetheless to analyze, rank and prioritize the risks identified in working place. At the present, more and more experts in the field are developing a criticism in relation to the limits, pitfalls and disadvantages of this category of methods (Koob and Malchaire, 2003), which we further denominate as "Kinney–type". They are increasingly considered as incompleted, non-reliable and detaining a strong subjective character. Other experts are refining their opinions, suggesting the resort to this kind of methods only as complementary or informational tools (Main, 2002; Schwartz, 2002).

The Kinney method was first introduced in USA, being proposed by G.F. Kinney and A.D. Wiruth in a technical document of Naval Weapons Center in California, see (Kinney and Wiruth, 1976). Initially aimed at explosion risk prevention in military industry, the method was rapidly adopted in Europe, with immediate success.

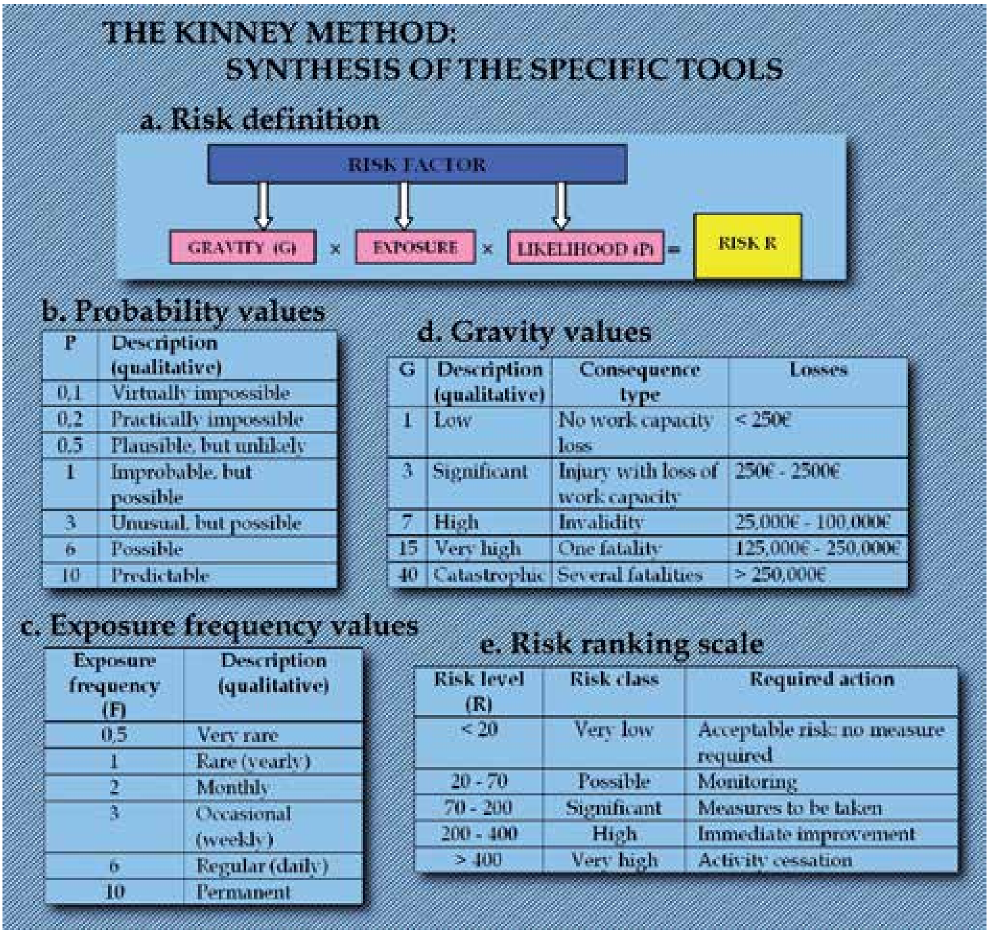
We argue that, indeed, the method can be used in the risk analysis stage, but, considering the specific estimation pattern, it rather evaluates and ranks the risks in order to prioritize them. According to this method, the risk (R) assessment, is achieved considering three parameters: 1) the probability (P) of an accident or damage occurrence, 2) the exposure at risk frequency (F) and 3) the gravity (G) of the induced consequence (see Figure 5.a). The probability of the damage occurring during the exposure to a risk factor describes the accidental, stochastic and uncertain character. Kinney has defined 7 probability classes, to whom he allocated certain numerical values (see Figure 5.b). The exposure frequency expresses the time lapse in which the worker is exposed to the risk factor action; this component is estimated by one of the 6 classes, described in Figure 5.c. The size of damages is expressed by 5 gravity classes, highlighted in Figure 5.d.

Unfortunately, if the evaluation is performed by a single individual, the process will be a flawe. Therefore the need for a multidisciplinary team is obvious. Afterwards, but only after the completion of this first identification phase, will be imagined and developed the risk propagation scenarios. Based on the context setting, the numerical values will be assigned to probability, frequency and gravity; the risk level will be obtained by multiplying these three factors. The value obtained allows then to frame the risks into 5 levels, according to Figure 5.e. If the method is applied by a working team, it is strongly recommended that all the risk values are retained, an average value to be computed, discussed and interpreted within the group session.

The use of the Kinney-type methods can offer seriously differing results, for the same working place or system, if the individuals composing the team are different from one case to another. The practical use of Kinney-type methods indicates that a misuse can lead to a variable risk factor list and diverging scores, according to the competency and expertise of the assessor. On the other hand, even if the figures obtained can be useful in increasing the awareness, the quantification is limited. It gives only the appearance of a mathematical evaluation, without really having the rigour of such an approach.

Table 4 comparatively illustrates the advantages and limitations of Kinney-type methods. However, despite their inherent limitations, Kinney-type methods do possess some advantages, such as accessibility, simplicity in use and fitness for training and teaching workers basic concepts, such as probability, frequency and gravity, in a qualitative manner. Also, these methods can be ideal sensibilization tools for the workers and staff members.

It can be asserted that the main drawback of Kinney-type risk assessment methods is their subjective character. The results, expressed in figures, have a quite low representativity and do not allow the user to compare different working places or enterprises. However, as far as certain aspects are not disregarded or neglected, these methods are sound.



THE KINNEY METHOD:
SYNTHESIS OF THE SPECIFIC TOOLS

a. Risk definition

RISK FACTOR

GRAVITY (G) × EXPOSURE × LIKELIHOOD (P) = RISK R

b. Probability values

P	Description (qualitative)
0,1	Virtually impossible
0,2	Practically impossible
0,5	Plausible, but unlikely
1	Improbable, but possible
3	Unusual, but possible
6	Possible
10	Predictable

c. Exposure frequency values

Exposure frequency (F)	Description (qualitative)
0,5	Very rare
1	Rare (yearly)
2	Monthly
3	Occasional (weekly)
6	Regular (daily)
10	Permanent

d. Gravity values

G	Description (qualitative)	Consequence type	Losses
1	Low	No work capacity loss	< 250€
3	Significant	Injury with loss of work capacity	250€ - 2500€
7	High	Invalidity	25,000€ - 100,000€
15	Very high	One fatality	125,000€ - 250,000€
40	Catastrophic	Several fatalities	> 250,000€

e. Risk ranking scale

Risk level (R)	Risk class	Required action
< 20	Very low	Acceptable risk: no measure required
20 - 70	Possible	Monitoring
70 - 200	Significant	Measures to be taken
200 - 400	High	Immediate improvement
> 400	Very high	Activity cessation

Fig. 5. Tools employed in Kinney risk assessment method

Advantages	Limitations
Numerical	Random data
Simple in use	Cost
Risk ranking	No guarantee for risk identification quality
Allows to assess the prevention-protection measures efficiency	Subjective method (high variability of results)
Risk acceptability evaluation	Unable to council the diverging risk scores
Establish if measures are required	Confusion hazard: P, F and G inaccurately defined
Education, information, reflection	False safety feeling
Employer or financial manager persuasion	Lack of rigour: how the scores differences are interpreted?
-	Applicable just for certain risks (not psychosocial or occupational illnesses, etc)

Table 4. The Kinney method: Advantages and limitations

If traditional risk management approaches have focused mainly on actions to counter hazards, modern methods are promoting actions of prediction, simulation, forecasting risk, reducing the reactive function and enhancing the preventive one. As outlined in previous work (Moraru and Băbuț, 2010 ; Moraru *et al*, 2010), risk management is no longer a narrow, limited approach, targeted exclusively to restrict or control the negative effects of potential events.

3.3 Controls development and decision-making in the risk treatment stage

Risk treatment involves selecting one or more options for risk control and implement these options through an iterative process. Figure 6 shows the sequence of phases which, in its entirety, guides risk treatment stage of the risk management process.

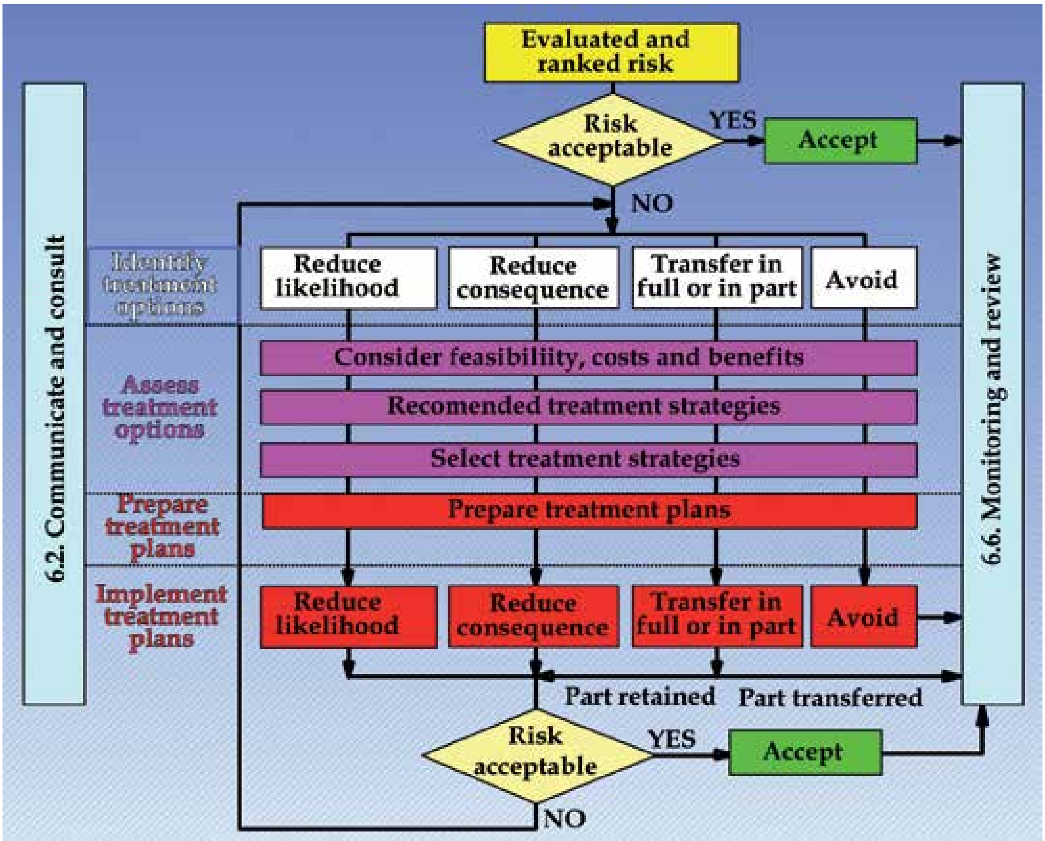

Evaluated and ranked risk
Risk acceptable
YES
Accept
NO
6.2. Communicate and consult
Identify treatment options
Reduce likelihood
Reduce consequence
Transfer in full or in part
Avoid
Assess treatment options
Consider feasibiliity, costs and benefits
Recomended treatment strategies
Select treatment strategies
Prepare treatment plans
Prepare treatment plans
Implement treatment plans
Reduce likelihood
Reduce consequence
Transfer in full or in part
Avoid
Part retained
Part transferred
Risk acceptable
YES
Accept
NO
6.6. Monitoring and review

Fig. 6. The structure of risk treatment stage (Adapted from AFNOR, 2006)

After assessing each hazard, the assessor develops one or more controls that either eliminate the hazard or reduce the risk (likelihood or consequence) of a hazardous incident. A key element of the risk decision is determining if the residual risk is acceptable. The decision maker, based upon the level of risk associated with the mission, must compare and balance the risk against mission objectives. If the the risk level is too high, he or she can direct the development of additional controls or alternate controls, or can modify, change, or reject the course of action.

When developing controls, it is important to try to implement controls based on the mitigation order of precedence (ISO, 2009; Moraru, Băbuț and Cioca, 2010;). The mitigation order of precedence is a prioritized ranking of methods for instituting countermeasures and controls, ranked by effectiveness in reducing the risk associated with an identified hazard. The mitigation order of precedence is discussed below.

- **Design to eliminate hazards:** the most effective method of controlling a hazard is to eliminate it from the system or equipment, by making fundamental changes in the design, process, system, equipment or task. Each situation must be viewed considering not only the hazard being addressed, but also the total situation. An excellent control used in another situation might seem appropriate, yet when viewed holistically in the context of the current task it not only does not work, but also introduces new hazards.
- **Incorporate safety devices:** when the hazards cannot be designed out or eliminated from the process, system or equipment, then safety devices need to be incorporated.
- **Provide warning devices:** warning devices are passive. While they provide notification that a hazardous situation exists, they require the operator to react to a given situation.
- **Develop procedures and provide training:** procedures rely upon the operator executing them. This requires initial training as well as periodic training, to ensure that the operator understands the "why" and the "how" of the procedures. They should be trained in 1) what the hazards are, 2) how to recognize the hazards and 3) what the control procedures are. If they do not understand the consequences, they are less likely to follow procedures. When implementing procedures certain factors need to be considered prior to their development, as those outlined in Table 5.
- **Selection, development and evaluation of controls:** a good understanding of the risk mechanisms facilitates effective development, selection, and prioritization of risk countermeasures and controls. The idea is to brainstorm as many controls and countermeasures as possible.
- **Decision-making:** this involves deciding which countermeasures to use, and in some special situations, requirements may dictate that the hazard and the risk be accepted due to constraints placed on the mission, process, system, or equipment. However, when the hazard is not eliminated or controlled to tolerable limits, the organization's top management needs to decide about the acceptability of the risk based upon mission requirements. Supervision ensures that deviation from standards, complacency, or violations of policies and risk controls are not allowed to threaten safety and health of the workers (U.S. Department of Army, 2010) .

The process for treating risks should be integrated with other planning and management activities at the programme, project or team level. It is important to try to link plans at this level to the corporate strategic plan where possible.

The employee must thus be the main actor - and not only the object - of prevention. This means that participation - and not only consultation - of the employees is indispensable. The successful implementation of OHS standards requires the process to be integrated into the activities of managers and supervisors at all levels of the hierarchy as well as the active participation of employees. The 'key considerations' concerning the strength of a participatory approach in the dynamic context of occupational risk management are described in Section 4, ahead of the review.

Issue	Questions to adress
Targeted working group	What is the structure of the team? What is the at-risk team? How large is the team.
Intervention	Are the reasons for application clearly defined? Are the results repeatable?
Outcome measurement	How do we measure the effectiveness? Have measurable goals and objectives been established?
Implementation process	What are the implementation issues? Are there unresolved issues and questions?
Developing training	What factors to address in the training? How to address those factors? Does the employee need new knowledge to do the procedure? What is that new knowledge?
Leadership	Leadership must supervise their operators to enforce the standard operating procedures. If leadership does not place value and importance on them, operators will not value or implement them.

Table 5. Training procedure development: questions to adress

4. Intervention steps in dynamic OHS risk management: The participatory approach

In the last decade, an increasing attention was given to participatory approach of occupational health and safety risk management (Honings, 2000; Maclagan, 1999; Malchaire, 2007, St. Vincent *et al*, 1998; Walters and Frick, 2000; Walters *et al*, 2005). By “workplace”, we generally understand, in a restrictive way, the place and the conditions in which a worker has to perform a stereotyped task (Moraru & Băbuţ, 2010).

This concept seems now out-of-date and, in the new forms of work organization, the operators work in a group of workplaces, that we will call a “work situation”, where they interfere the ones with the others, as depicted in Figure 7. Moreover, the behavior, satisfaction, quality of work and well-being of any worker do not depend only on the physical factors of his working environment, but also on the work organization, the responsibilities and the collective relations. The collocation “work situation”, firstly introduced by (Malchaire 2007), refers to all the aspects, physical, organizational, psychological or social of the working life (Cioca and Moraru, 2010), that are likely to have an influence on the health, the behavior and the well-being of the employee, see Figure 8.

A quite realistical assumption is that knowledge from what really occurs in the work situation is decreasing from the employee to the expert. On the other hand, qualification in health and safety increases in the opposite direction (Malchaire, 2007). It thus appears logical to consider that the two sets of knowledge - about the work situation and about the principles of the well-being - are complementary. What remains is to organize a cooperation in an interdisciplinary way. Many risk assessments by OHS practitioners or external experts, are undertaken in a given time span, on a specific problem often extracted from its context. This is why they may have very little positive effects, especially if workers are not involved directly in the process.

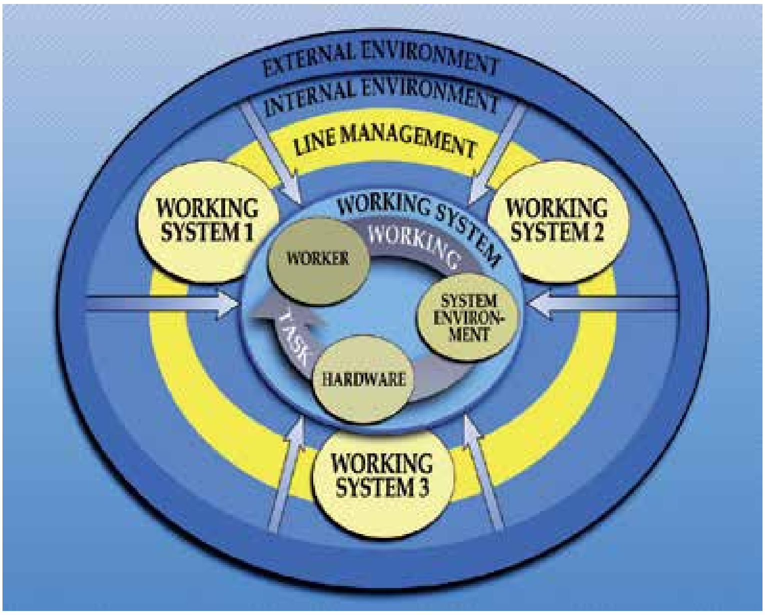
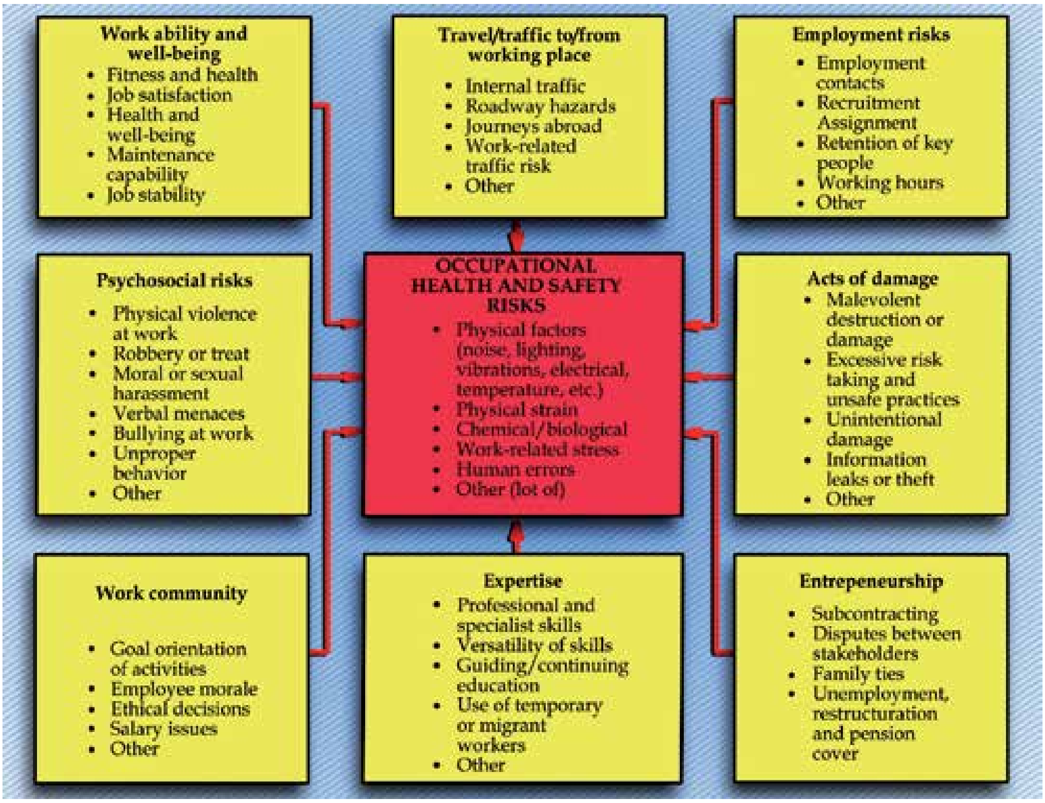


Fig. 7. Graphical representation of a hypothetical working situation



Work ability and well-being
- Fitness and health
- Job satisfaction
- Health and well-being
- Maintenance capability
- Job stability

Travel/traffic to/from working place
- Internal traffic
- Roadway hazards
- Journeys abroad
- Work-related traffic risk
- Other

Employment risks
- Employment contacts
- Recruitment Assignment
- Retention of key people
- Working hours
- Other

Psychosocial risks
- Physical violence at work
- Robbery or treat
- Moral or sexual harassment
- Verbal menaces
- Bullying at work
- Unproper behavior
- Other

OCCUPATIONAL HEALTH AND SAFETY RISKS
- Physical factors (noise, lighting, vibrations, electrical, temperature, etc.)
- Physical strain
- Chemical/biological
- Work-related stress
- Human errors
- Other (lot of)

Acts of damage
- Malevolent destruction or damage
- Excessive risk taking and unsafe practices
- Unintentional damage
- Information leaks or theft
- Other

Work community
- Goal orientation of activities
- Employee morale
- Ethical decisions
- Salary issues
- Other

Expertise
- Professional and specialist skills
- Versatility of skills
- Guiding/continuing education
- Use of temporary or migrant workers
- Other

Entrepeneurship
- Subcontracting
- Disputes between stakeholders
- Family ties
- Unemployment, restructuration and pension cover

Fig. 8. The complex and multifactorial landscape of OHS risks

But workers participation is only effective if the qualification of the employees concerning their work situation, and their integrity are explicitly recognized (Haines *et al*, 2002). The employee "sees" his work situation like a whole and not like a set of distinct and independent facts: he is "being well" or not, he enjoys his job or not. In addition, all aspects of the work situation are inter-related: the noise influences the relations between the people; the technical organization between workstations influences the risks for musculoskeletal disorders; the division of the responsibilities influences the work content, the accidents, etc

The prevention approach consists in seeking the most effective means to reduce the risk, by acting on one or several of its components: elimination of the risk factor, reduction of the exposure, increase of the reliability of the work system. It is thus essential that the analysis of the risk be not simply a recording of its components, but consisting of a careful analysis of the reasons of the exposure, the circumstances of this exposure, the severity of the consequences and the most useful means to reduce them. The final quantitative evaluation of the risk is consequently secondary, the most important thing being to study the components and the details on which it is going to be possible to act. Rather than speaking about risk assessment, it is thus more appropriate to speak about risk management.

Although all the problems are inter-dependent, it is neither realistic nor possible to solve all of them at once. Considering the Swiss-cheese accident model (see Figure 9) and the spirit and structure of the Sobane strategy (see Figure 10), developed by Malchaire (2007) the first stage (called *Screening*) consists, for example, in replacing a defective tool, improving the ventilation system or modifying a hierarchical relation. Even if such a simple measure is essential, it is not sufficient because the primary latent causes were not eliminated and the situation can return to the initial defective state.

The second stage (*Observation*) can consist in reexamining the general work organization, the institutional links between people or in rearranging the operating area. Perhaps a third stage will relate to the workers training (*Analysis*): vocational training to perform the tasks, education to their well-being, leading them to recognize themselves the problems, to manage them directly as they arise, bringing the employees to a degree of self-management of their health, safety and well-being to work. Maybe a next stage (*Expertise*) will relate to the culture of the company, the integration of the concerns of well-being in the overall management of the company.

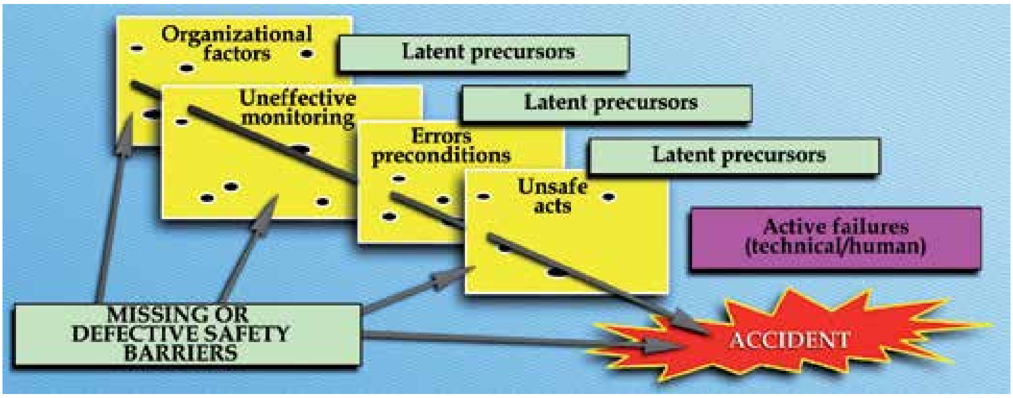


Fig. 9. The Swiss cheese accident model (Adapted from Reason, 1993)

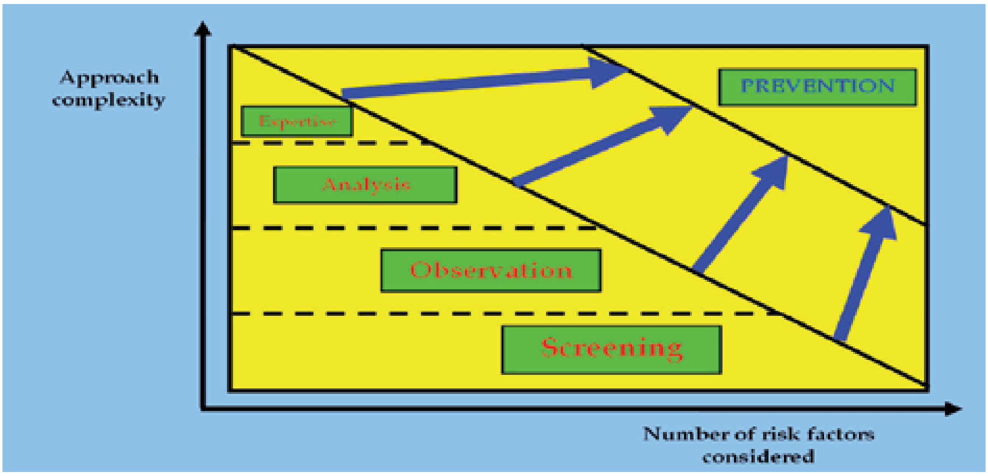


Fig. 10. General layout of the Sobane risk management strategy (Based on Malchaire, 2007)

The knowledge, information and data necessary during the first steps relate primarily to the work situation. Knowledge in ergonomics, medicine and safety is certainly desirable to select the good tool or to ventilate more effectively, but is less essential than the knowledge of the work situation. This first step must therefore be carried out as close as possible of the work situation and its output will be especially a function of intimate knowledge of what occurs in the course of time in this work situation.

Conversely, at a more advanced step of the process, the problems require more qualification in work organization, training or management. The analysis must be finer, more specific and requires tools and competences that only OHS practitioners generally have. According to the step, the necessary competences will thus rather be those of an OHS practitioner or of the workers themselves, these remaining the main actors of the prevention, for whom and by whom prevention is implemented.

In our attempt to highlight the future trends in OHS risk management, we will focus in the next section on two epistemological assumptions existing in the field, presently valid in complex socio – technical systems. Epistemological issues determine how we see the world and is hence of fundamental importance. Our aim is to emphasize that contrary to common beliefs, the positivism and the constructivism are complementary and not competeing.

5. Towards complementarity of positivism and constructivism in OHS risk management

Risk management within complex socio - technical systems must be subordinated to the objectives that form an integrated, coherent system and converging to the overall objectives, so that levels of activity to be mutually supportive (Emmet and Hickling, 1995). Operational safety techniques in use did not appear, indeed, though it seems paradoxical, until during the Second World War, with the emergence of complex systems, which encompass a large number of components. The ’60s have brought new techniques for identifying risks, following the deductive or inductive approaches. In the '70s, great efforts have been made to assess risks to populations near nuclear facilities, being developed many scenarios of accidents related to equipment failures and/or operational errors. In the following decades, methods and techniques of risk assessment and control were widely spread in various industries and now enter new sectors (health, food, water treatment, etc.), including in Romania, following the harmonization of the national legislation with the European Union.

As depicted in Figure 11, technical risks are related to the main features of operational safety. Technical risk control covers all methods, means, analysis, procedures and measures applied throughout the life cycle, to remove or render acceptable the identified risks.

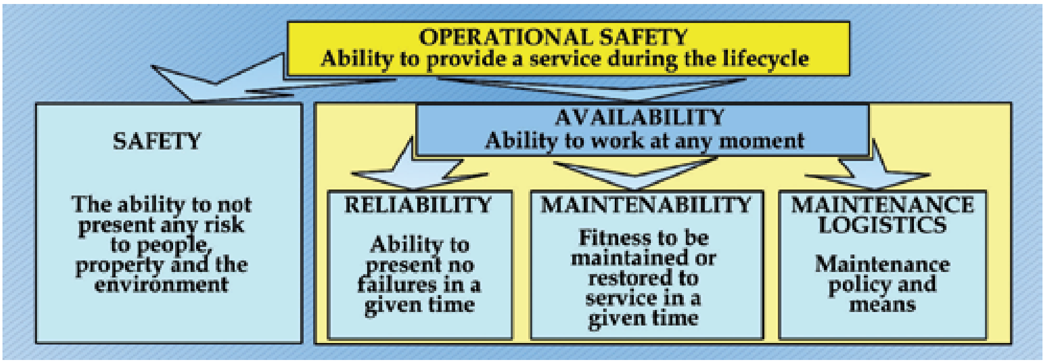

OPERATIONAL SAFETY
Ability to provide a service during the lifecycle
SAFETY
The ability to not present any risk to people, property and the environment
AVAILABILITY
Ability to work at any moment
RELIABILITY
Ability to present no failures in a given time
MAINTENABILITY
Fitness to be maintained or restored to service in a given time
MAINTENANCE LOGISTICS
Maintenance policy and means

Fig. 11. Operational safety components within a complex socio - technical system

As discussed before, the omnipresence of uncertainty about the reliability of data used, adopted measures to limit or minimize the consequences likely to materialize, the degree of subjectivity of the assessments are rending particularly difficult a precise formalization of applied assessment and decision – making procedures. Moreover, we find very often that the existing level of expertise itself is an area of experimentation. Most times, between policy makers and technicians, OHS practitioners and economists, there is a fundamental degree of independence. For various reasons, one can minimize the real risks, for example by overpricing available technological level, leading to a dogmatic extreme reaction (by ignoring his own ignorance) or because of reasons that we might euphemistically call “inconspicuous”. Precaution and prevention is a step fall in anticipation of risk. Precaution can be understood as extending prevention methods applied in the field of uncertainty. Precaution and prevention are thus two sides of prudence that is required to reduce the occurrence of situations likely to cause losses to humans, property and environment.

Until recently, the safety of complex systems were addressed in a manner called "positivist" (Journe, 1997), which is to control risks during the design stage. Developed specifically by engineers and ergonomists, this approach considers technology as 1) a real phenomenon ("*ontological principle*"), 2) having an existence outside the subject that one observes and implements (*principle of objectivity*"), 3) having a fixed operation and successful law of his own ("*wired universe principle*"), 4) which leads to an optimal solution ("*single optimum principle*") (Le Moigne, 1990 b). It is considered that the reliability and safety of a system can be developed by acting on technology, on the working environment and on the proper definition of procedures to follow. The human factor is often identified as "*the weak link*" element, which reduces overall system reliability, as "*a black box capable of unpredictable and irrational behaviour, the origin of the errors, failures and shortcomings*" (Journe, 1997). The operator has only a small margin of action or response to an organizational and technical reality which is external to him, because it was set *a priori* by those who designed the system.

Following the serious incidents that have occurred mainly in the nuclear field (e.g. Three Mile Island in 1979, Chernobyl in 1986 etc), there have been strongly emphasized the limits of such an approach. Due to the ever increasing complexity of systems studied, it becomes obvious the impossibility to predict all possible accident scenarios.

A complementary approach, called "constructivist" was therefore proposed (Le Moigne, 1990a; Theys, 1991, Journe, 1997). This one, considers technique as 1) a construction (*"principle of the built universe"*), 2) including the subject which tries to control it, through the intermediate of his own representation (*"projection and representation principles"*), 3) having a complex operation that can not be broken down into simple, independent elements and 4) which can only lead to more or less satisfactory solutions (*"principle of intelligent action"*). This approach leverages operators role in the complex socio – technical systems reliability and security, seeking to benefit as much of their intelligence and capacity to respond to new situations. Founded on "safety culture" and developing the spirit of initiative within organizations, at the opposite of the quasi-military discipline required in the positivist tradition, this approach gives the central role to the individual and is subject of the "organizational communication" (Zohar, 2010).

In opposition to the positivist conception, the constructivist approach considers that the reliability is, above all, a social and organizational *construction*, which is the final output of the symbolic representations that operators develop together, in action. From this perspective, safety conditions lies, particularly, in organizational and human variables. This organization's ability to remedy failures would be particularly related to a clear definition of each individual's role and a strong personal responsibility. We argue that this approach can be extended to any kind organization, even to small and medium enterprises, to properly manage the OHS risks, following a conceptual model as the one described in Figure 12.

Safety is considered both as 1) a non-event because the result of a safe state is invisible through nature (nothing happens when the situation is under control) and 2) a dynamic state because a seemingly stable system conceals intense internal activity, requiring vigilance and anticipation capacity of the operator. Thus defined, safety is an issue of interaction, development and management of the representations that will give a sense to the situations experienced by operators. It is obvious that, from this perspective, the role of safety culture and climate is ever increasing.

In this constructivist approach, risk control means, above all, the operators ability to anticipate and recover abnormal situations. It requires the skill of the actors to understand the environment in which they work, based on their own experiences, the meaning of various stimuli perceived. It is the product reached through cognitive operations and not the result of a pre-existing sense, independent of themselves. The constructivist approach also put emphasis on the performance of involved actors and the role of everyday communication; it considers communication as part of a process of organizational structure. We would like to outline that this is also consistent with ISO 31000:2009 standard.

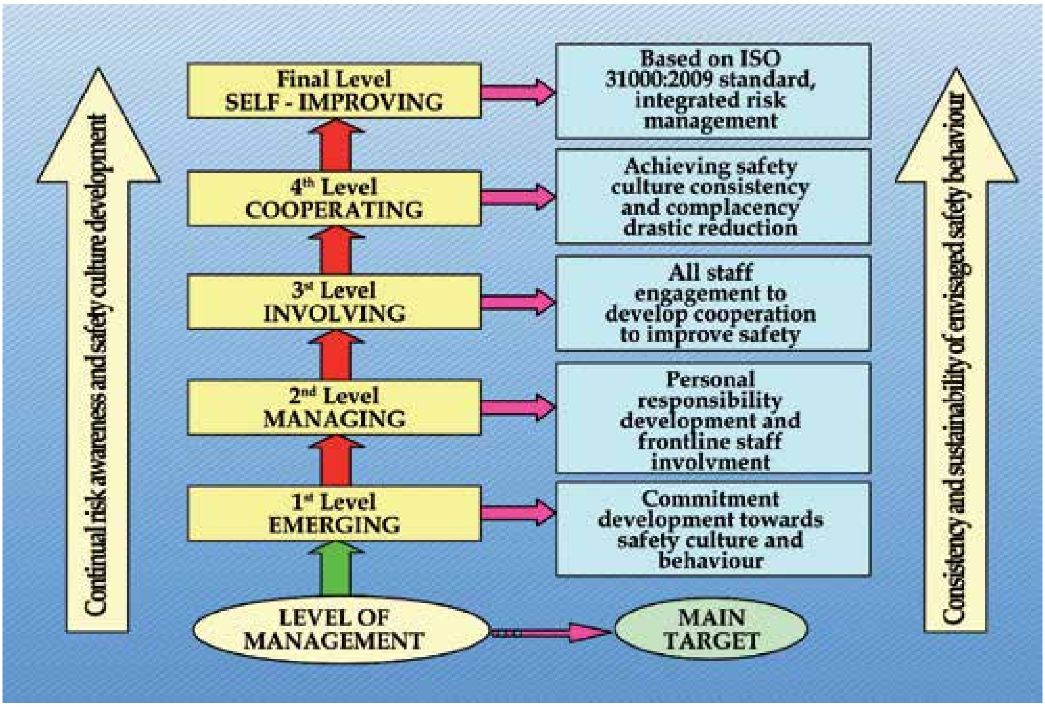

Continual risk awareness and safety culture development
Final Level
SELF - IMPROVING
Based on ISO 31000:2009 standard, integrated risk management
4th Level
COOPERATING
Achieving safety culture consistency and complacency drastic reduction
3rd Level
INVOLVING
All staff engagement to develop cooperation to improve safety
2nd Level
MANAGING
Personal responsibility development and frontline staff involvment
1st Level
EMERGING
Commitment development towards safety culture and behaviour
LEVEL OF MANAGEMENT
MAIN TARGET
Consistency and sustainability of envisaged safety behaviour

Fig. 12. A constructivist conceptual model for OHS risk management

On the other hand, there is always an unexplored ethical dimension over and above the list of causes of occupational accidents. This can result in a cascade of bad decisions being taken at the organizational, human and technical levels, which cause tragic accidents, often with loss of human life. To manage risks properly, it will not only be necessary to develop techniques but also to develop processes, at the personnel level as well as at the organizational level, which will take human nature into account. There is indeed a real difficulty to forecast the operator's representations and reactions, in a context that develop a growing public distrust in the complex high-risk organizations, such as for example nuclear power plants. Positivist and constructivist visions of operational safety, which initially appeared in an opposite, even irreconcilable, may be complementary, we believe, in the minds of some players, including designers of new systems that would like to associate quality in design with operators professionalism, to limit risks to a level still untouched.

6. Closure

This paper reviews the main strands of literature and proposes a framework of steps in good OHS management. It offers a number of suggestions of good practice and illustrates a number of the dilemmas. It also considers that the approach based on ISO 31000:2009 Standard is of paramount importance and may represent a step forward to managing safety, but may not be enough to address the management of risk effectively. There is a need to adopt a systemic approach to safety management. Systemic may be defined as trying to see things as a whole and attempting to see events, including failure, as products of a working of a system and, within that, see fatality/injury/property loss, etc as results of the working of systems. Companies that have applied risk management for many years recognize that the change to a “culture of prevention” via “systematic and comprehensive risk management” involves a journey.

On the other hand, risk assessment and risk management should be described more fully, recognizing that some companies are currently going well beyond the minimum level of practice that is described here. From this perspective also, we consider that the ISO 31000:2009 Standard can be seen as a milestone. Risk management should try to change everyone’s mind set toward proactive, empowered, systems oriented thinking. Risk management should become a continuous process of learning from past experiences.

What is extremely important in seeking to achieve effective risk management is an ongoing consolidation of an organizational culture of risk. Risk management is incompatible with attitudes such as "leave it, that goes like this too”, "someone else looks after, I shouldn’t worry", "we shall live and see", "devil is not quite so black as it seems", "this is a fatality" etc. Risk management is the assumed accountability. This is the difficult problem in the way of implementing effective risk management and not learning the terminology and techniques.

Learning from the past is crucial in making progress. Yet, there are obstacles of various nature such as 1) human factors, 2) technical complexities of how to store and retrieve information, restraints to knowledge management, and last but not least 3) policy and decision making in view of cost–benefit. Over the last decade, wide interest in the role of behaviour has led to the development of numerous safety climate tools and behavioural modification programmes. It must be emphasised that behavioural approaches should not be seen as the panacea for all safety problems. Behaviour modification is not an alternative to sound safety management policies, systems and procedures. However, when these are well established and functioning effectively, behaviour modification can play an important role in achieving further improvements in safety performance.

Also, the move towards continuous management and learning demands participation from employees. Often, however, initiatives are imposed from above. Creativity and initiatives from the shop floor to improve are stifled in such a climate. A different approach is required to encourage further improvement.

This next step involves taking action to ensure that the behaviours of people at all levels within the organisation are consistent with an improving safety culture Once rules become more conceptual and goal-oriented, rather than prescriptive, this cannot work. Nor can it work if the technology is changing or there are many abnormal situations to be coped with. It is hard to have an expert around all the time to make up new rules or modify old ones. The only other alternative is to delegate the rule-making and changing to the work group. Participatory methods are increasingly utilized in improving risk management of work and workplaces. The merits of these methods are widely recognized as a means of promoting initiative of local people and achieving workable solutions. A notable merit is that they contribute to improving various forms of workplaces in their diverse conditions. The influence, though still low, of constructivism allow actors to assign a contextual re-creation and permanent interaction capacity with the working system. Unfortunately, we find that this vision strikes with numerous and powerful resistance, both inside and outside the organization.

The in-depth analysis of the dynamic policy of risk management resulted in proposing a participatory strategy whose purpose is 1) to gradually approach the work situations, 2) to coordinate the cooperation between employees, management and OHS practitioners and 3) to arrive faster and less expensively to effective prevention.

If we put together all of the trends sketched in this chapter, they add up to an improvement in the way OHS is, or should be, managed compared to a decade ago. With an increasingly complex structure of society, trade and industry, communication, political and administrative systems, technical systems and technological change etc, it is getting more and more important to relate to different ethical and political arguments to keep an image of being in control and for being trusted. There is a need for developing a mechanism for a path between the positivistic and rational choice founding of risk assessment and the richness of perceptions and judgements of risk issues. The best answers to risk issues are reached by exposing risk assessments and decisions to intelligent debate, criticism, and amendment by the scientific community and the workers likely to be affected by the risk.

The preliminart conclusion of this work is also an outlook to the future. Although there are many signs of renewed interest in questions regarding OHS risk management , the road to a more full implementation of "premises for risk management" is long and there are many obstacles. One of the main issues to be resolved is how to decide on differing levels of risk depending on the nature of the activity and the importance or benefits of the activity for society.

7. References

Association Francaise de Normalization, AFNOR, (2006). *Management du risque – Approche globale,* ISBN 2-12-169211-8, Saint-Juste-la-Pendue, France

Ayers, G. (2007). Encouraging meaningful and effective consultation about occupational health and safety (OHS) in the construction industry: a recognition of workforce competence. Retrieved June 2011 from: http://web-safety.com/kosha/SY38-03.pdf

Blockley, D. (1992). *Engineering Safety*, McGraw-Hill, London, UK.

Bourrier M. (1998). Elements for designing a self-correcting organisation: examples from nuclear plants. In Hale A.R.& Baram M. *Safety management: the challenge of change*. Pergamon. Oxford.

Cioca, L.I. & Moraru, R.I. (2010). *Psychosocial Occupational Risk Management*, (in Romanian), "Lucian Blaga", Publishing House, ISBN 978-973-739-924-3, Sibiu, Romania

Cox, S. & Flin, R., 1998. Safety culture: philosopher's stone or man of straw? *Work & Stress* 12 (3), 189–201.

Deming, W.E. (1982). *Out of the Crisis*, Center for Advanced Engineering Study, ISBN 0-911379-01-0, Cambridge: MIT Press

Emmett, E. & Hickling, C. (1995). Integrating Management Systems and Risk Management Approaches. *J. of Occupational Health and Safety – Aust. NZ* , 11(6), 617–624

Haines, H., Wilson, J.R.,Vink, P. & Koningsveld, E. (2002). Validating a framework for participatory ergonomics. *Ergonomics* vol 45, nr. 4, pp. 309-327

Honings, C. (2000). Complémentarité des méthodes Kinney et d'analyse participative des risques, *Travail & Bien-être* n°5, décembre 2000, pp. 29-32.

Hudson P.T.W. (2003) Aviation safety culture. *Journal of Aviation Management*, issue 3, 27–48.

ISO (2009). ISO 31000:2009 Risk management -- Principles and guidelines, International Organization for Standardization, Available from: www.iso.org

Journe, B. (1997) Positivisme et constructivisme dans la gestion de surete et de la fiabilite des centrales nucleaires, Constructivisme et Sciences de gestion, Lille, 23 octobre.

Joy, J. & Griffiths, D (2004). National minerals industry safety and health risk assessment guideline, version 3, March, MCA and MISHC, Australia, Retrieved November 2008 at www.planning.nsw.gov.au.

Kinney, G.F. & Wiruth, A.D. (1976) Practical risk analysis for safety management, Naval Weapons Center, California, SUA.

Koob, J.-P. & Malchaire, J. (2003). Fiabilité de la méthode Kinney d'analyse des risques, Essai sur un chantier de pose de conduites enterrées, Université catholique de Louvain, Unité Hygiène et Physiologie du Travail.

Lave, L. B. & Balvanyos, T. (1998). Risk Analysis and Management of Dam Safety, *Risk Analysis*, Volume 18, , pp. 455-462.

Le Moigne, J-L. (1990a). Epistemologies constructivistes et sciences de l'organisation, în Martinet, A.C., *Epistemologies et Sciences De Gestion,* Paris, Economics.

Le Moigne, J-L. (1990b). *La modelisation des systemes complexes*, Paris, Dunod.

Main, B.W. (2002). Risk Assessment is Coming. Are You Ready? *Professional Safety* ,47(7) , pp. 32–37

Maclagan, P. (1999) Corporate social responsibility as a participative process, *Business Ethics: A European Review,* Vol. 8 (1), pp. 43-49.

Malchaire, J. (2007). *Stratégie SOBANE et guide de dépistage DEPARIS,* Service public fédéral Emploi, Travail et Concertation Sociale, UCL, Retrieved September 2010 at www.sobane.be

Moraru, R. & Băbuț, G. (2010). *Participative risk assessment and management: a practical guide,*(in Romanian), ISBN 978-973-677-206-1, Focus Publishing House, Petroşani, Romania

Moraru R.I., Băbuț G.B. & Cioca L.I. (2010). Adressing the human error assessment and management, *Archives of Mining Sciences*, Vol.55 (2010), No.4, pp.873-878.

Pasman, H.J. (2009). Learning from the past and knowledge management: Are we making progress? *Journal of Loss Prevention in the Process Industries* vol.22, pp.672–679

Reason, J. (1993). *L'erreur humaine*, Paris: Presses Universitaire de France

Reason, J. (1995). A system approach to organisational error. *Ergonomics*, vol. 38, nr. 8, pp. 1708-1721.

Romanian Parliament (2006). The 319/2006 Act on Occupational Health and Safety, *Oficial Monitor of Romania*, Part I, nr. 646/26.07.2006

Sage, A. P. & White, E. B., (1980). Methodologies for Risk and Hazard Assessment: A Survey and Status Report, *IEEE Transactions on Systems, Man and Cybernetics*, SMC-10(8), pp. 425-445.

Swartz, G. (2002). Job Hazard Analysis. A Primer on Identifying Hazards. *Professional Safety*, 47(11), pp. 27–33.

Smith, D (2008). Implementing the new BSI code of practice on risk management (BS31100) in the Public Sector, ALARM Scotland Conference, Bournemouth, Retrived July 2011 from : http://www. theirm. org/events/ documents /IRMNoSGroup DougSmith PresentationNov08.pdf

St.Vincent, M., Chicoine, D.&Beaugrand, S. (1998). Validation of a participatory ergonomic approach in industries in the electrical sector. *Int J Ind Ergonomics*, vol. 21, pp. 11-21.

Standards Australia & Standards New Zealand (2004). *Risk management, AS/NZS 4360/2004* Available from *https://infostore.saiglobal.com/store*

Theys, J. (1991). Conquete de la securite, gestion des risques, Paris: L'Harmattan, p.279-298.

U.S. Department of Army (2010). Safety Mishap Risk Management, DA-PAM 385-30, Retrieved August, 2011 from http://www.apd.army.mil/pdffiles/p385_30.pdf

Walters, D, & Frick, K. (2000). Worker participation and the management of occupational health and safety: reinforcing or conflicting strategies? Cht.3, in *Systematic occupational health and safety management; perspectives on an international development* (ed by Frick, K., Jensen, P., L., Quinlan, M., & Wilthagen, T.), Pergamon, Amsterdam.

Walters, D., Nichols, T., Connor, J., Tasiran, A. C. & Cam, S. (2005). *The role and effectiveness of safety representatives in influencing workplace health and safety,* Health and Safety Executive, Retrieved April 2011 from http://hse.gov.uk.research/rrhtm/rr363.htm

Zohar, D. (1980). Safety climate in industrial organizations: theoretical and applied implications. *Journal of Applied Psychology* 65/1980, pp. 96–102.

Zohar, D. & Luria, G., (2004). Climate as a social-cognitive construction of supervisory safety practices: scripts as proxy of behavior patterns. *Journal of Applied Psychology* 89/2004, pp.322–333.

2

Hazard Matrix Application in Health, Safety and Environmental Management Risk Evaluation

Assed Haddad, Erick Galante, Rafaell Caldas and Claudia Morgado
Federal University of Rio de Janeiro
Brazil

1. Introduction

In occupational safety one of the most complicated and harder to achieve goals is to prioritize actions towards risk prevention and mitigation. There are several methodologies to do that. Some of them are expensive; demand some extensive structure for its application and so on. Some others are either weak in results or lack technical or scientific basis. (Haddad *et al*, 2008). The research presented in this chapter emerged in between these methodologies for being not expensive and requiring resources, most of the time, already available in a company's documentation.

More specifically, this chapter focuses on the development and usage of a risk assessment methodology called Hazard Matrix (HM) and its application in Health, Safety and Environmental Management (HSE). The HM is a prioritization methodology suitable to be used in the analysis phase of a risk management program. The application of HM in HSE is a very powerful methodology to highlight critical hazards and sectors/areas in a business unit or company under study.

We will present and explore the HM concept, features, relevance and implementation, under the scope of a risk management process. In order to achieve this, the chapter will follow this sequence, organized in sections:

- HSE aims and structure (Section 2);
- Aspects of risk management necessary for a HSE Program (Section 3);
- The HM concept, structure and applications; (Section 4)
- Two case studies for the clarification of the methodology and to highlight its possibilities of use (Section 5).

The case studies presented come from two different types of industries, in order to demonstrate the comprehensive application of HM and details. Although the methodology lacks a refined or complex mathematical structure, we can refine the presented approach and develop other implementations leading to a high detailed usage. Here we refrain to do that; this is something for future work.

2. HSE aims and structure

Basically, HSE deals with the anticipation and recognition, analysis, evaluation, treatment and communication of hazards and treats in the occupational safety aspect of any company, and by an expanded or more comprehensive perspective, to all surrounding environments. HSE is becoming increasingly important, as well as demanded, in any company's effort towards sustainability and legal adequacy. It is becoming part of its regular operations and is commonly required by regulatory bodies, clients and society.

In order to organize a HSE Program one needs to develop and structure (ISO 14.001:2007, 2007) a set of activities in a close relation with a Risk Management Program (ISO 31.000:2009, 2009). Determination of scope and objectives are necessary for the establishment of such a Risk Management Program. In this phase, each step of this management process must be defined, as well as its expected results. After that, the mapping of all organization's processes and also their interactions must be performed (the ones that interfere in risk management). On this phase of the HSE effort, all activities, operations and the relations between them must be clearly identified, as they represent potential sources of hazard. Moreover, in order to achieve enough knowledge to underlie adequate action of HSE process, data must be collected from the activities and operations, about its **Processes** (inputs and outputs), its **People** who work there, and generally about the **Organization** itself.

After data collection and the processes mapping are done, the management process is able to perform a **Hazard Mapping and Identification**, which consists in identification, localization and classification of the hazards involved in each process activities of the organization. This classification is performed according to two aspects: type of hazard existent, such as physical, chemical, biological etc., and the severity of its impact. The second one is traduced mathematically by weights that can be represented in a boarding selection.

After these entire actions have been performed one can establish the necessary set of information in order to develop the HM (Haddad *et al*, 2008) application in the next stage. With the sequential implementation of the concepts and tools listed one should be able to apply these concepts, use the HM, recognize and prioritize risks, and implement a useful tool in the development of a Risk Management Program addressing HSE issues.

3. Aspects of risk management necessary for a HSE program

System Safety can be described as a sub-discipline of systems engineering that applies scientific, engineering and management principles to ensure adequate safety, the timely identification of hazard risk, and initiation of actions to prevent or control those hazards throughout the life cycle and within the constraints of operational effectiveness, time, and cost (Vincoli, 2006).

One of the key aspects of safety management is the proper management of risk. There are several ways of defining risk, but all of them share a common core: they define risk as combination of its independent variables likelihood and severity. The new ISO 31.000:2009, entitled "Risk Management", presents a series of guidelines and principles for its management. ISO 31.000:2009 defines risk as:

"Risk - effect of uncertainty on objectives

NOTE 1 an effect is a deviation from the expected – positive and/or negative.

NOTE 2 Objectives can have different aspects (such as financial, health and safety, and environmental goals) and can apply at different levels (such as strategic, organization-wide, project, product and process).

NOTE 3 Risk is often characterized by reference to potential events and consequences), or a combination of these.

NOTE 4 Risk is often expressed in terms of a combination of the consequences of an event (including changes in circumstances) and the associated likelihood of occurrence.

NOTE 5 Uncertainty is the state, even partial, of deficiency of information related to, understanding or knowledge of an event, its consequence, or likelihood"

The concept of risk indicates that it relates to activities or tasks to be performed, being intimately related to a human behavior or environment. As an example, a fuel, such as gasoline, isolated from its context, does not present any level of risk by itself. However, when this fuel is considered to be handled at an environment where sources of ignition are present, for example, dealing with gasoline becomes a hazardous activity. The associated risk will vary according to the behavior of the handler and environmental where it is done. As stated by (Ciocoiu and Dobrea 2010) "even apparently insignificant risks have the potential, as they interact with other events and conditions, to cause great damage. "

Due to the very nature of risk, it is present in almost every activity, job or task performed in the modern world. The accelerating pace of business, globalization, the financial crisis, all contribute to the growing number and complexity of risks and to the greater responsibility for managing risks on an enterprise-wide scale (Ciocoiu and Dobrea, 2010). This leads to the need of managing occupational risks ahead in order to assure minimum casualties and optimum performance. However, the risk is also a very complex entity, which directs to difficulties when understanding and managing.

It is commonly accepted that a proper method for assessing and managing risks pass through decomposing the "risk" in its independent variables: *frequency, severity,* and *scenario*. Therefore, the dependent variable "Risk" can be written as:

$$Risk = f\,(frequency,\ severity,\ scenario)$$

Decomposing the risk in independent variables and further analyzing each variable is the core method presented in the MIL STD 882 for preliminary Hazard analysis. This concept has proved to be successful , since it is still used by many standards, including the forth review of MIL STD 882 (United States, 2003), OHSAS 18.001:2007 (OHSAS, 2007) and ISO 31.000:2009 (ISO, 2009)

3.1 The risk management process

Before human's development of *risk management* process and the clear concept of *system safety*, the common technique used to deal with risk was mainly based on a *trial and error* approach. This culture led mankind to fix problems only after its consequences were observed at the already designed running system. In opposition, the idea of minimizing risks until acceptable levels by trying to predict the occurrence of accidents and taking measures in order to avoid them is widely diffused nowadays. This approach characterizes the concept of risk management.

Risk management can be defined as the collection of culture, processes and structures that are directed towards realizing potential opportunities whilst managing adverse effects (Australia, 2004). In this context, it is a complex process and can be understood as a systematic application of management policies, procedures and practices.

The NZS 4360:2004 guideline (AUSTRALIA, 2004) determines that a proper overall risk management process incorporates several other tasks, such as:

- Communication,
- Identification,
- Analysis,
- Evaluation,
- Treatment,
- Monitoring; and
- Reviewing.

The ISO 31.000:2009, in its turn, defines risk management as:

"Risk management:

process of systematic application of management policies, procedures and practices to the activities of communicating, consulting, establishing the context, and identifying, analyzing, evaluating, treating, monitoring and reviewing risk"

Fig. presents an overview of the entire risk management process, based upon the process described in the NZS 4360:2004 guideline (Australia, 2004), using a block diagram representation.

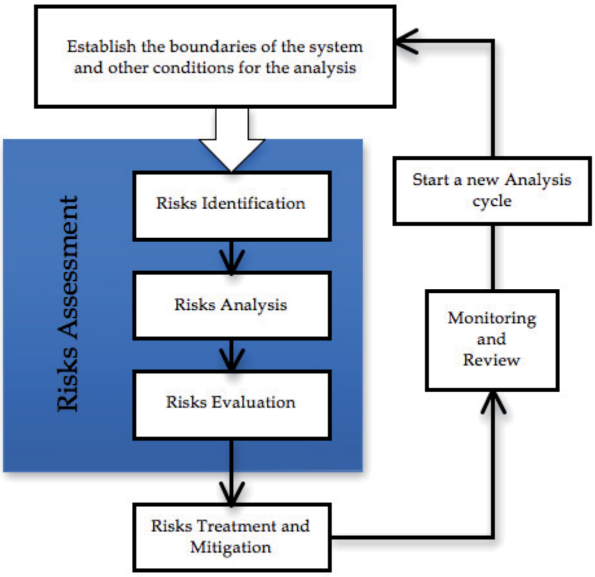

Establish the boundaries of the system and other conditions for the analysis
Risks Assessment
Risks Identification
Risks Analysis
Risks Evaluation
Risks Treatment and Mitigation
Start a new Analysis cycle
Monitoring and Review

Fig. 1. Risk Management overview (based on ISO 31000: 2009)

At this point, it is important to highlight that the success of risk treatment depends directly on how complete its preliminary analysis, which identifies, describes and classifies them. Risk identification of any organization requires the previous knowledge of its activities and processes, of the external market conjuncture and the legal, social, political and cultural environment in which the company is inserted, and for that, it must be clearly defined its strategic objectives (Muniz, 2011).

Within the risk management, which is an interactive process where the risk is studied again after all the mitigations are in place (as indicated by the arrow linking "treat risk" and "monitoring and review" in figure 01), one major phase in the overall process is assess the risk. The Australian and New Zealand Standards AS NZS 4360:2004 defines that this phase comprehends: Identify the risk, analyze the risk and evaluate the risk, and defines its steps as follow:

- Risk analysis: systematic process to understand the nature of the risk and deduce its level;
- Risk identification: the process of determining what, where, when, why and how something could happen:
- Risk evaluation: the process of comparing the level of risk. In many cases the risk evaluation involves establish a priority among several risks.

Furthermore, due to resources limitation, which is unavoidable, addressing all indentified risks becomes a non-realistic approach. As resources can be limited financially, technically, or even related to time or personnel, one will need to implement a system to highlight the most critical hazards and sectors, so that appropriate resources application can be assured. The important role of a prioritization system highlighted above is efficiently accomplished by HM, since its main objective is to establish a priority ranking among risks and sectors.

3.1.1 The importance of prioritization

Companies have one primary goal: profit. This is the reason for their existence and thus almost all decisions taken at business environment unavoidably aim at profits increase (Mello, 2004). In this context, hazard minimization and mainly the prevention and mitigation of its consequences must be established in a systemic and optimized way.

As it is clear that any company works with limited resources, there ought to be a prioritization tool able to identify such as which risks or company's sectors are more critical, on which the plan of risks mitigation can be based and oriented. Therefore, the best utilization of financial, labor or time resources can be assured.

In this context, must be emphasized that there are many aspects in which prioritization can be established. It can be financial aspect, so that more money is spent on the priority issue and it is assigned a larger proportion of the available budget. Alternatively, 'priority' may be temporally defined, as all risk issues cannot be tackled simultaneously. An identified priority issue would thus be address in precedence to another, which may be deferred until a later time. Temporal prioritization is fundamental to the development of long-term strategic risk management plans. (Centre for Environmental & Risk Management, 1997)

Situated right between the Evaluation and the Treatment phases at Risk Management Process, described by Figure 1, risks prioritization becomes a strategic part of this process when it is taken for its real application, since, after risk analysis and evaluation, it is possible that resources limitations naturally lead to a necessary chose of more critical measures to be taken at a first moment, as well as demand a temporal ordination.

Further, studying risk management, one can conclude that risk prioritization is a key aspect of the overall procedure. As risk prioritization and treatment are intrinsically connected to risk analysis and evaluation, the next section briefly presents and discusses methods commonly used to evaluate risks.

3.2 Risk evaluation schemes

In the risk management process one important aspect is the risk analysis, which is done using a chosen methodology, which is sometimes called "risk assessment scheme".

Despite the fact that Hammer's book (reference) is an old reference, it still is quoted and cited in many recent works and manuals of risk assessment, due to its clear ideas and concepts. Although the risk assessment schemes presented by Hammer might be somehow obsolete, the concepts are not.

There are an unlimited number of schemes. Each one has its strengths and weakness. Hammer, in his book, presents several possible schemes, which can be folded into 4 greater categories:

- Analysis in trees
- Analysis in spreadsheets
- Qualitative Analysis
- Quantitative Analysis

It is important to remark that every risk assessment method will fall into two of the four categories. A risk assessment tool will be qualitative or quantitative and formatted as tree or spread sheet. This leads to a classification matrix as shown in Table 1.

	Analysis formatted as tree	Analysis formatted as Spread Sheet
Qualitative Results		
Quantitative Results		

Table 1. Risk Assessment – Classification Table

The schemes of risk analysis under the **tree concept** focus on establishing a chain of events, as well as assess the risk occurrence likelihood. The TNO Red Book **(**Committee For The Prevention of Disasters, 1997**)** presents some fault tree analysis methods. Risk assessment performed using fault trees technique can provide a result as simple as a series of causes and effects (**qualitative approach**) or as complex as the evaluation of risk occurrence probability (**quantitative approach**). It is also important to remark that the analysis complexity varies at the same rate as the analyzed system complexity.

The other category of risk assessment is the spreadsheet. Some of the most used risk assessment schemes fall into the spreadsheet category. This category counts widely used methods like HazOp **(**Committee For The Prevention of Disasters, 1997**)**, FMEA **(**United States. MIL-STD-1629, 2000**),** Hazard Preliminary Analysis (HPA) (United States, MIL-STD 882-D, 2003) and the HAZARD MATRIX (Haddad, 2008).

The majority of the risk schemes available approach the problem using the principle of compartmentalization, studying sub-systems and sub-components, on behalf of simplicity in more complex systems.

This "compartmentalization" is done setting borders in the scenario. These borders can set sub-areas for analysis, sub–system, and sub-component or split the "target" of the analysis using nodes. The main representative tool of this technique that places nods in the flow of matter and energy flows and analyze the risks in each nod is the HazOp. (Committee For The Prevention of Disasters, 1997),

Irrespective of the chosen method to assess risk, a quantitative analysis is to be preferred over any qualitative, mainly because a quantitative approach establishes a hierarchy between risks. The HM is an example of a risk assessment method designed to prioritize risks, hazards and environmental (sectors).

In summary, the risk assessment schemes should be selected matching the strengths and weakness of each scheme with a given scenario and the output requirements:

- **HazOP** - This scheme suite perfectly a scenario where the productive process can be represented (drawn) as a flow chart, due to the easiness of cutting the flow applying nods on it. It is known that a block diagram can be used to represent a "flow chart" and, therefore, allow application of HazOP to a discontinuous process (such as an assembling line), if the level of risk justifies this approach. (Committee For The Prevention of Disasters, 1997),
- **FMEA** - The FMEA is the risk assessment scheme published by the MIL STD 1629 (UNITED STATES, 2000) and used spreadsheets to decompose the risk in its failure modes, causes and consequences. This FMEA suites best when applied to machinery or equipment's.
- **Tree Analysis** - All the assessment methods that fold into this category assess the risk by set a chain of events (causes and consequences). These methods regard to determine the frequency (number) for the likelihood of the risk to occur.
- **Matrix Schemes** - These schemes provide a very flexible approach to the risk and hazards analyses. As a general rule, these tools compensate the lack of deep by decomposing the risk in its independent variables (likelihood and severity) and analyzing each variable separately. These methods are also indicated to assess risk of a particular task in hand. **The hazard matrix proposed in this chapter is one example of this category.**

This section provided an overview about risk management in a general approach. However, risk management procedures may differ depending upon scenario and applications. The next section addresses the risk management and evaluation when applied specifically to environmental risks.

3.3 Environmental risk management and evaluation

When it comes to environmental risk, the scenario is a key part of the entire risk assessment routine. In general, the hazards identification process is the same for both environmental and occupational risk management. The main difference between them will be related to the scope utilized to determine the scenario.

At his point, considering this risk scenario context, it is important to clearly define a limit that, in a simplified way, represents the interface between the industry internal and external environment. This limit is well represented by a key concept: the *industry walls*, which defines a separation between occupational and environmental hazards.

During an occupational risk management process, one is concerned with the effects of the hazard within the boundaries of the enterprise, company, unit or industry. This scope guides all decisions, from which methodology use to what consequences should be neglected.

On the opposite, the scope applied over the scenario when an environmental risk management process is undertaken focus on risks source effects on the external area, which means beyond the *industry walls*. Figure 2 represents the difference between environmental and occupational risk management approaches.

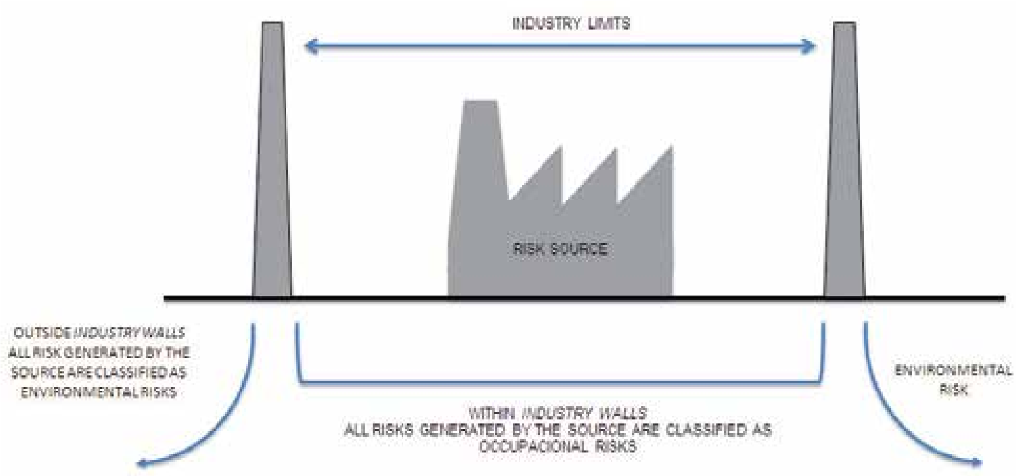

INDUSTRY LIMITS
RISK SOURCE
OUTSIDE *INDUSTRY WALLS* ALL RISK GENERATED BY THE SOURCE ARE CLASSIFIED AS ENVIRONMENTAL RISKS
WITHIN *INDUSTRY WALLS* ALL RISKS GENERATED BY THE SOURCE ARE CLASSIFIED AS OCCUPACIONAL RISKS
ENVIRONMENTAL RISK

Fig. 2. Environmental Risk versus Occupational Risk

4. The Hazard matrix (HM) concept, structure and applications

The HM methodology, based on the work done by (Haddad *el al* 2008), is a valuable tool to allow determination of prioritization among several risks, hazards and sectors within a given system or environment. This complex system can be anything that held more than a single hazard to be prevented or mitigated, from an industry to an office building.

The HM approach is based on the already cited concept of risk as a function of its severity and probability of occurrence. In a simplified view, as seen before, risk can be defined by the product of its two variables. In the HM, the probabilistic factor is represented by the number of workers exposed to the hazard. The severity factor, in turn, is mathematically traduced by a numerical classification of hazards, which will be more specified later.

The analysis starts by dividing the company in sectors, identifying the hazards and sectors its respective sectors of exposure. Due to that, it is likely to use the hazard matrix combined with other risk identification and assessment tools, like FMEA (MIL STD 1629) or HPA (MIL

STD 882). Each sector constitutes a line (from 1 to y) in the hazard matrix, followed by a column that stands for the number of workers on that sector. As for the other columns, they take all the hazards identified in all the sectors, drawing column from 1 to x.

After building the matrix, the hazards identified are preliminarily assessed though out the use of a "Risk Assessment Code" (RAC). This RAC follows a simple criterion, which varies according to the scenario, availability of data or even precision required. The important is that the criteria must be the same for the entire matrix, as well as be able to provide a numerical number.

When it comes to occupational hazards, one valid criterion is presented (as example) in Table 2.

Risk Assessment Code	**Description**
0	This hazard is NOT present in the sector evaluated
1	The exposure of this hazard occurs bellow the action level AND it is occasional
3	The exposure of this hazard occurs bellow the action level AND, continuously
6	The exposure level is between the action level and the Threshold Value Limit (TVL-TWA) or equivalent.
9	The exposure level is above Threshold Value Limit (TVL-TWA) or equivalent.

Note: Action level is half the TVL-TWA value

Table 2. Risk Assessment Codes for Hazardous agents (chemicals, physical or Biological agents).

Therefore, the HM is completed by evaluating the hazards using the RAC pre-selected for a given scenario. Each given position within the matrix corresponds to the hazard in a given sector. That means that the value written in the position (*i,j*) stands for the RAC () which best represents the exposure to the Hazard "j" faced by the workers in sector "i".

In the HM, the Sectors (S) and Number of Workers (W) forms the lines and the hazards (H) the columns that draws the borders of the RAC. Table 3 presents a general HM.

Sector	**Hazards identified**					
Description / Name	**Number of people working**	**H_1**	**H_2**	**H_3**	**…**	**H_x**
S_1	**W_1**	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$	…	$R_{1,X}$
S_2	**W_2**	$R_{2,1}$	$R_{2,2}$	$R_{2,3}$	…	$R_{2,X}$
S_3	**W_3**	$R_{3,1}$	$R_{3,2}$	$R_{3,3}$	…	$R_{3,X}$
…	**…**	…	…	…		…
S_Y	**W_Y**	$R_{Y,1}$	$R_{Y,2}$	$R_{Y,3}$	…	$R_{Y,X}$

Table 3. Hazard Matrix – General Construction model

The next stage of the HM approach is to calculate the **hazard frequency of recurrence**, the **exposure frequency** and the **relevancy percentage**.

The hazard frequency of recurrence measures how intense is the overall exposure to a given risk, while the exposure frequency evaluates which sector represents a more hazardous environmental to work in. Both frequencies take into account the number of works exposed and the intensity of the hazard.

The relevancy percentages are a mathematical composition of both hazard frequency of recurrence and exposure frequency to allow easier understanding and prioritization. Next, we present the hazard frequency calculation in more details.

4.1 Hazard frequency calculation

Taking the first Hazard (H1) in the matrix (Table), its hazard frequency of recurrence is determined by the following calculation:

$$f_{H1} = W_1 * R_{1,1} + W_2 * R_{2,1} + W_3 * R_{3,1} + \ldots + W_y * R_{y,1}$$

Similar calculations are performed to all the other hazards to determine their hazard recurrence frequency:

$$\begin{aligned} f_{H2} &= W_1 * R_{1,2} + W_2 * R_{2,2} + W_3 * R_{3,2} + \ldots + W_y * R_{y,2} \\ f_{H3} &= W_1 * R_{1,3} + W_2 * R_{2,3} + W_3 * R_{3,3} + \ldots + W_y * R_{y,3} \\ f_{Hx} &= W_1 * R_{1,x} + W_2 * R_{2,x} + W_3 * R_{3,x} + \ldots + W_y * R_{y,x} \end{aligned}$$

This allows the establishment of a general rule for determining the **recurrence frequency** of any hazard within a HM as follows:

$$f_{Hj} = \sum_{i=1}^{y} W_i * R_{i,j} \ ,$$

where

$$1 \le j \le x$$

After all the **hazard recurrence frequencies** are determined, a global hazard frequency can be determined as follows:

$$F_H = \sum_{j=1}^{x} f_{Hj}$$

The global hazard frequency is used to calculate the relevancy percentages. The other key figure of a hazard matrix is the exposure frequency, which is further presented in the next section.

4.2 Exposure frequency calculation

Following the determination of all the hazard recurrence frequencies, comes the determination of the **exposure frequencies**. Taking the first sector (S1) in the HM (Table ???) the exposure frequency is determined as follows:

$$f_{s1} = W_1 * R_{1,1} + W_1 * R_{1,2} + \ldots + W_1 * R_{1,x}$$

The other exposure frequencies are determined using similar equations:

$$\begin{aligned} f_{s2} &= W_2 * R_{2,1} + W_2 * R_{2,2} + \ldots + W_2 * R_{2,x} \\ f_{s3} &= W_3 * R_{3,1} + W_3 * R_{3,2} + \ldots + W_3 * R_{3,x} \\ f_{sy} &= W_y * R_{y,1} + W_y * R_{y,2} + \ldots + W_y * R_{y,x} \end{aligned}$$

Analyzing the method to determine each **exposure frequency**, it is possible to determine a general rule for their determination:

$$f_{si} = \sum_{j=1}^{x} W_i * R_{i,j}$$

, where

$$1 \leq i \leq y$$

The global exposure frequency is determined as follows:

$$F_S = \sum_{i=1}^{y} f_{si}$$

Similarly to the global hazard frequency, the global exposure frequency will also be used to calculate the relevancy percentages. This is done next.

4.3 Relevancy percentage calculation

The next stage of the hazard matrix method is determining the **relevancy percentage** , which can be calculated via the following set of equations:

$$\%_{Hj} = \frac{f_{Hj}}{F_H} * 100 = \frac{\sum_{i=1}^{y} W_i * R_{i,j}}{\sum_{i=1}^{x} f_{Hj}} * 100,$$

where

$$1 \leq j \leq x,$$

$$\%_{si} = \frac{f_{si}}{F_s} * 100 = \frac{\sum_{j=1}^{x} W_i * R_{i,j}}{\sum_{i=1}^{y} f_{si}} * 100$$

where

$$1 \leq i \leq y$$

With these calculations performed, we are ready to assemble the entire HM. This is discussed next.

4.4 The complete hazard matrix

Once all the calculations are concluded, the HM from Table 3 can be updated to incorporate these percentages. Table 4 Hazard Matrix – Relevancy Percentages presents this newer and complete matrix.

5. Hazard matrix applications in HSE

Like risk management, hazard matrix application also varies depending upon scenario and scope of analyses. This section addresses these differences and customization required when one applies HM to HSE as well as other applications.

Sector	**Hazards identified**							
Description/ Name	**Number of people working**	**H_1**	**H_2**	**H_3**	**...**	**H_x**		
S_1	**W_1**	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$	…	$R_{1,X}$	**f_{s1}**	**$\%_{s1}$**
S_2	**W_2**	$R_{2,1}$	$R_{2,2}$	$R_{2,3}$	…	$R_{2,X}$	**$f_{S,2}$**	**$\%_{S,2}$**
S_3	**W_3**	$R_{3,1}$	$R_{3,2}$	$R_{3,3}$	…	$R_{3,X}$	**f_{S3}**	**$\%_{S3}$**
...	**...**	…	…	…		…	**...**	**...**
S_Y	**W_Y**	$R_{Y,1}$	$R_{Y,2}$	$R_{Y,3}$	…	$R_{Y,X}$	**f_{SY}**	**$\%_{SY}$**
			f_{H1} **$\%_{H1}$**	**f_{H2}** **$\%_{H2}$**	**f_{H3}** **$\%_{H3}$**	**f_{Hx}** **$\%_{Hx}$**		**100%**

Table 4. Hazard Matrix – Relevancy Percentages

5.1 Environmental risks prioritization

The main difference between applying the HM to prioritize occupational hazards and environmental hazards is the understanding of the scenario. As already discussed in figure 2, one addressing the occupational hazards will analyze risk, frequencies and exposure within the boundaries of an enterprise, Industry or site, while the environmental risk assessment will regard the effects on the exterior of this same enterprise.

Hence, in cases on which the HM is intended to be applied to prioritize and assess environmental risks, some small adaptations are due.

Firstly, the sectors of an enterprise will be replaced to vulnerable areas already identified on the neighborhood (or areas of influence) of a given enterprise, industry or unit. Subsequently, the number of workers will be replaced by an average number of people living / working in that given vulnerable areas. Secondly, the hazards identified will concern to risks that can create effects outside the enterprise, industry or unit boundaries. From these adaptations, the environmental HM is adapted as shown in table 5.

Vulnerable area	Hazards identified							
Description/ Name	**Number of people Living/ working**	**H_1**	**H_2**	**H_3**	**...**	**H_x**		
A_1	**P_1**	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$	…	$R_{1,X}$	**f_{s1}**	**$\%_{s1}$**
A_2	**P_2**	$R_{2,1}$	$R_{2,2}$	$R_{2,3}$	…	$R_{2,X}$	**$f_{S,2}$**	**$\%_{S,2}$**
A_3	**P_3**	$R_{3,1}$	$R_{3,2}$	$R_{3,3}$	…	$R_{3,X}$	**f_{S3}**	**$\%_{S3}$**
...	**...**	…	…	…		…	**...**	**...**
A_Y	**P_Y**	$R_{Y,1}$	$R_{Y,2}$	$R_{Y,3}$	…	$R_{Y,X}$	**f_{SY}**	**$\%_{SY}$**
			f_{H1} **$\%_{H1}$**	**f_{H2}** **$\%_{H2}$**	**f_{H3}** **$\%_{H3}$**	**f_{Hx}** **$\%_{Hx}$**		**100%**

Table 5. Environmental Hazard Matrix

Thirdly, the risk levels need to be re-designed to match environmental events. Due to the complexity of assess effects and predict damages, it is recommended to keep the description of environmental risks as simple as possible. Table 6 presents one suitable example.

Risk Assessment Code	**Description**
0	This hazard is NOT percept in that given vulnerable area
1	This Hazard can damage the environment in a reparable way, without any permanent damage to structures, environmental or people, and without victims.
3	This Hazards can severely damage the environmental and harm people, without loss of lives
9	This hazard can cause immediate death of at least one person within the vulnerable area

Table 6. Environmental Risk Assessment Codes, based upon MIL STD 882 Severity categories.

The mathematical determination of the hazard frequency, exposure frequency, relevancy percentages remains the same as presented in the section 3.

This section presented an application of the Hazard Matrix tool to Environmental Risk Analysis. However, HM is a valuable tool for other scenarios as well. One example of such application of HM is in project management, which is further explored in the next section.

5.2 Projects/temporal applications

One of the most useful and enriching results of a Risk Prioritization tool, which specially includes the HM, is the global and comparative view of the organization's sectors (as well as their interactions) that it provides. Nevertheless, some productive sectors, such as the Construction Industry, have its operation based on projects, where the final product is the main goal of the production activities, which are finished when the product is completed.

Such kind of activities, in terms of risk assessment and prioritization, may demand a specific approach, in order to provide to project's managers its global understanding and the associated risks related to each temporal stage, as well as their interactions along time and its effects on the project.

In this context, in order to meet such demands, a convenient adaptation of HM is suggestively developed. Initially, the **SECTOR** column will be replaced by a **STAGE** column, where will be allocated each stage of the project. Alternatively, it is quite acceptable to create sub-stages, if it seems applicable. The probabilistic factor continues to be well represented by the number of employees who is going to be involved with each stage.

It is important to emphasize that these analysis do not intent to replace the objectives and results of conventional HM, but it only provides, when it is possible and convenient, a different and complementary view of the whole process in a temporal approach. Therefore, the combined use of both types is quite possible.

6. Case studies

In this section, two study cases are presented. Each of the cases of study deals with a different aspect and application of the hazard matrix method.

6.1 Chemical process unit

Urbanski (1964) and (Meyer 1977) present a method to produce lead azide via precipitation using lead nitrate and sodium azide as reactants. This method is largely used in the explosive industry, been lead azide is a chemical compound with explosives properties and largely used as a primer explosive (payload of blasting caps) in mining activities. Due to its (lead azide) explosive properties, "explosion" is expected to be the most relevant hazard. However, the hazard matrix method proves that "perception" alone is not a suitable risk assessment technique.

Since lead azide manufacture involves chemical compounds, the RAC (Risk assessment codes) should take into account the threshold exposure limits as a parameter. Table 7 presents the RAC used in this case of study.

		Risk Assessment Codes
TVL-TWA		9
Action Level		6
		3
		1
		0

Table 7. RAC for a chemical process unit

Taking the work done by (Galante 2008), in which a hazard matrix methodology was applied to this very kind of manufactory, a hazard matrix for a lead azide unit can be written as in table 8.

SECTOR HAZARDS							
Description	Number of Workers	Exposure to Chemicals (Lead salts)	Physical Hazard - Noise	Accident - Explosion	Accident – Electrical Discharge	Fs	%s
Pb(NO3)2 preparation workshop	2	9	1	0	1	22	23%
Na(OH) preparation workshop	2	0	1	0	1	4	4%
Na(N3) preparation workshop	2	0	1	0	1	4	4%
Precipitation reaction workshop	4	6	1	3	1	44	46%
Drying	2	3	1	6	1	22	23%
	Fh	48	12	24	12		
	%H	50%	12.5%	25%	12.5%		100%

Table 8. Hazard Matrix for Lead Azide manufactory

Analyzing the results achieved by calculating the hazard matrix, it is possible to prioritize the sector and the hazards using the relevance percentage:

Sectors:

Precipitation reaction workshop – 46%
Pb(NO3)2 preparation workshop – 23%
Drying – 23%
Na(OH) preparation workshop – 4%
Na(N3) preparation workshop – 4%

Hazards:

Exposure to Chemicals (Lead salts) – 50%
Accident – Explosion – 25%
Physical Hazard – Noise – 12.5%
Accident – Electrical Discharge – 12.5%

From the analysis, it is possible to conclude that the most relevant hazard is related to deal with lead salts. This relates to the toxicity of lead and its effects once within the human body. Since the exposure to lead involves more people, it got a higher relevance percentage, which differs from the "initial guess" that "explosion" would be the most relevant hazard.

As for the sector, the precipitation workshop got the higher relevance, mainly due to the higher number or workers in the sector.

In summary, according to the HM written for this scenario, the top priority when comes to mitigate and control hazards would be address the lead exposure (50% of all the hazards relevancies are due to lead exposure) in general, and in particular its exposure within the precipitation workshop (this area counts for 46% of the exposure relevancies).

6.2 Oil Extraction – Offshore platform

Nowadays modern society and economy is heavily dependent upon oil and its derivatives. From that, it is just logical to assume that oil industry is a major player within the international economy, being present worldwide, from prospection, extraction and processing. At the same time, oils are fuels and by that reason they represent a hazard to be managed. Even more, the large amounts of oil dealt with worldwide exponentially increases the hazard, both occupational (intense labor, heavy machinery, pumps, chemical hazards, fire hazard, and accidents in general.) and environmental (major fires, explosions, major leaks both in land and water).

In a greater scope, there are the oil extraction operations off shore. Extraction platforms stationed off shore are one important representative of this group. These platforms are very complex installations, being capable to drill and build the oil wells, extract, produce and hold both oil and high pressure gases, as well as perform the entire set of required maintenances (Freitas *et al* 2001; Booth and Butler,1992).

Due to the off shore environment in which oil platform operate and the hazards of its operations and products dealt with, as well as economical relevance in modern society, they constitute an extremely important scenario for risk assessment, including hazard matrix ranking system. Muniz (2011) developed his work having an off-shore oil extraction platform as a case of study. The chosen platform is set at the Brazilian coast, near the city of Vitoria.

The work performed by (Muniz 2011), the RAC used were as in table 9

Risk Assessment Code	Description
0	Exposition un-existence
1	Low level of exposition.
3	Medium Level of Exposition
9	High Level of Exposition

Source: Translated from (Muniz 2011)

Table 9. Environmental Risk Assessment Codes.

Using the criteria for RAC as presented in Table 9, (Muniz 2011), it was drawn two hazard matrixes, one analyzing a platform in general (Table 10) and a second one analyzing the drilling system (Table 11), among others.

As a complex system, oil extraction platforms in an off-shore environment have a large number of areas and hazards to be dealt with. The data presented in Tables 10 and 11 are nothing but a small portion of the work developed by (Muniz 2011). Therefore, a HM is a remarkable tool to guide the risk assessment process, as well as, the team responsible for risk control and mitigation.

In the early stages of the risk management program, by the use of HM, several hazards can be ignored, sectors and areas of lower relevance percentages can be neglected and all the effort and resources can be oriented to those hazards and areas that represent the most relevant within this vast and complex system.

SECTOR HAZARDS																			
	Physical	Chemical	Biological	Ergonomic	Accidents														
Description	Number of Workers	Noise	Heat	H2S, CH4	Metal dust and fumes	Organic Vapors	Cleaning Products	Chemical Products	Contaminated Air	Chemical Intoxication	Position and Movement	Lighting	Machinery and Equipment	Fire and Explosion	Falling Objects	Transportation People and Cargo	Electricity	Fs	%s
Administration	6	1	0	0	0	0	1	0	1	0	1	1	0	1	0	0	1	42	14
"Hostelling"	13	1	3	0	0	0	3	0	1	0	3	1	3	3	0	0	1	247	8
External Area	16	3	1	0	9	0	0	0	0	0	1	0	9	3	9	9	1	720	23.5
Drilling System	20	3	3	3	0	9	0	3	0	1	9	1	9	9	3	9	3	1300	42
Engine Room	10	9	9	0	0	9	3	9	1	3	9	1	3	9	1	1	9	760	24.8
	Fh	217	205	60	144	270	75	150	29	50	331	49	393	363	214	334	185		
	%H	7.1	6.7	2	4.7	8.8	2.4	4.9	0.9	1.6	10.8	1.6	12.8	11.8	7	10.9	6		100%

Source: Translated from (Muniz 2011)

Table 10. Hazard Matrix – GENERAL

SECTOR	HAZARDS																		
	Physical	Chemical	Biological	Ergonomic	Accidents														
Description	Number of Workers	Noise	Heat	H2S, CH4	Metal dust and fumes	Organic Vapors	Cleaning Products	Chemical Products	Contaminat ed Air	Chemical Intoxication	Position and Movement	Lighting	Machinery and Equipment	Fire and Explosion	Falling Objects	Trans-portation People and Cargo	Electricity	Fs	%s
Driller's Cabin	4	1	0	0	0	1	0	0	1	0	0	0	0	9	0	0	1	48	4.1
Energy Room	4	1	0	0	0	0	0	0	0	0	0	1	0	3	0	0	3	32	2.7
Probe Floor	10	3	1	3	0	9	0	1	0	1	9	0	3	9	3	9	0	510	43.2
Sieves Area	10	9	9	1	0	9	0	3	0	3	9	3	9	3	1	0	0	590	50
	Fh	128	100	40	0	184	0	40	4	40	180	34	120	168	40	90	12		
	%H	10.8	8.5	3.4	0	15.6	0	3.4	0.3	3.4	15.3	2.9	10.2	14.2	3.4	7.6	1		100

Source: Translated from (Muniz 2011)

Table 11. Hazard Matrix – DRILLING SYSTEM

7. Limitations and aspects of hazard matrix

Since Hazard Matrix is a prioritization tool, as it aims to provide a priority ranking among identified hazards and sectors, they must be previously identified and inserted in the HM. Therefore, HM utilization in combination with a Risk Identification tool, such as Preliminary Hazard Analysis (PHA) is essential. In this sense, failures in the risk identification phase, as well as the exclusion of some specific risk or sector identification will certainly jeopardize the accuracy and reach of HM prioritization results and global analysis.

Under the approach of decomposing risk into its independent variables, as discussed earlier, HM prioritization of hazards are intrinsically related to the factors severity and probability of occurrence. As HM prioritizes the most critical hazards, the ones whose product of severity x probability have a high value would be, at a first moment, a priority for the treatment phase highlighted by HM. However, at a second moment, after risks mitigation under the orientation of HM prioritization is done, hazards whose severity and probability were lower initially will appear as priorities in a second HM utilization. The transition happens because, during the risk mitigation phase, severity and probability factors of the most critical risks are reduced. Consequently, in a second HM utilization, they will not appear in the top of priority ranking provided by the method.

In addition, even though the HM Environmental application provides an efficient alternative analysis, one of the most relevant limitations of the methodology is related to the difficulty to compare efficiently in prioritization terms both Environmental and Occupational risks in the same HM.

Besides, although HM provides a relative comparison between hazards or sectors that ordinates, in percentage, the most critical sectors/hazards to be prioritized, when the difference between percentages are very small, the differentiation of relevance between this sectors/hazards and its effects on the risks mitigation plans must be analyzed carefully, since this numerical analysis is relative. Therefore, in this situation, other factors that concern at these sectors/hazards must be evaluated and combined to the HM's result.

In this sense, the role of HM alternative applications, such as Project and Environmental presented earlier become a useful complementary analysis. By providing an auxiliary view of the global situation, they may compensate the conventional HM restrictions and limitations such as the ones listed above, reducing the probability of errors.

8. Closure and future work

Some future aspects of the application of the HM methodology should be listed as: 1) a deeper study in the application of the weights used inside the HM; 2) improvements done in the data collection of hazards and sectors, in order to enhance the results achieved (having the burden of a higher price of investment) and 3) the development of a better integration between risks pertaining the occupational aspect of the analysis (said to be inside the walls) and those risks which expose a larger community (said to be outside the walls).

As it stands, the HM concept and application in HSE is a powerful yet simple form of decision making in an occupational risk assessment. It is fully integrated in the Risk Management Program of any company. Resources and structure to do this are quite commonly already available.

The strategic role of a risk prioritization stage at HSE is justified by the inherent limitation of resources with which any company works, as well as the difficulty to compare, in a relevance scale, predicted accidents. As shown in the first study case, human's impression of risk's relevance is not a reliable reference to make decisions and prioritize risk mitigation. In this context, the utilization of the HM, combined with some risk identification tool, such as Hazards Preliminary Analysis (HPA), enables such prioritization, ordering in a critical scale both sectors and hazards already identified and classified.

As there are different applications of HM, it is important to notice that their combined utilization with the original HM model is in a complementary approach. Thus, for the HSE of a company, for example, there can be developed both an occupational and an environmental Hazard Matrix. However, comparison between both environmental and occupational risks, in order to orient the prioritization of risk mitigation plans is still a challenge.

One may say that HM is the chosen method to prioritize risks and to determine strategic resources utilization within risk management. However, HM approach and effects transcend the mere aspect of risk prioritization: It must be considered as an efficient, global and multidisciplinary analysis, connected to plenty aspects of risk management, optimization and resources utilization.

9. References

Australia. (2004). Australian and New Zealand Standards, AS NZS 4360:2004: *Risk Management*. ISBN 0 7337 5904 1

Brazil. Lei 6.514/77, regulamented by Portaria 3.214/78 - NR 23

Brazil. Lei 6.938 de 31 de agosto de 1981 - Política Nacional de Meio Ambiente. Brasilia, 1981.

Bubbico, R.; Marchini, M. (2008). Assessment of an explosive LPG release accident: A Case study. *Journal of Hazardous Material,* No 202, (2008), pp 558-565 , ISSN 0304-3894

Ciocoiu C. N., Dobrea R. C. (2010). The Role of Standardization in Improving the Effectiveness of Integrated Risk Management. *Advances in Risk Management*, Giancarlo Nota (Ed.), ISBN 978-953-307-138-1, InTech, Available from: http://www.intechopen.com/books/show/title/advances-in-risk-management

Centre for Environmental & Risk Management (CERM), School of Environmental Sciences, University of East Anglia. (1997). *Risk Ranking*, Health & Safety Executive, Norwich.

Center for Chemical Process Safety of the American Institute of Chemical Engineers. (1989). *Guidelines for Chemical Process Quantitative Risk Analysis*. New York: CCPS.

Clancy, VJ. (1972). Diagnostic Features of Explosion Damage, *6th International Meeting on Forensic Sciences*, Edinburgh, Scotland,.

Committee For the Prevention of Disasters. (1997). Methods for Determining and processing probabilities (RED BOOK). Netherlands. ISBN 90-12-08543-8

Crippa, C., Fiorentini, L., Rossini, V., Stefanelli, R., Tafaro, S., & Marchi, M. (2009). Fire risk management system for safe operation of large atmospheric storage tanks. *Journal of Loss Prevention in the Process Industries,* Web of science..

Date, P. Lade, R.J. Mitra, G. E Moore, P.E. (2009). *Modelling the risk of failure in explosion protection installations.* Center for the Analysis of Risk and Optimisation Modelling Applications (CARISMA), School of Information Systems, Computing and Mathematics, Brunel University, Middlesex UB8 3PH, UK

De Mello, C.H. (2004), *Uma Ferramenta Computacional para uso da Matriz de Relevância na Avaliação de Riscos Computacionais,* Universidade Federal Fluminense (UFF), Brazil.

Fortuna, A. O. (2000). *Técnicas Computacionais para Dinâmica dos Fluidos: Conceitos Básicos e Aplicações,* Editora da Universidade de São Paulo, ISBN 85-314-0526-2, São Paulo

Fogler, H. S.; Elements of Chemical Reaction Engineering. *Prentice-Hall International Series in the Physical and Chemical Engineering Sciences.* IBSN 0-13-53708-8;

Freitas, C.M., Souza, C.A., Machado, J.M et al. (2001). *Acidentes de trabalho em plataformas de petróleo da Bacia de Campos,* Cad. Saúde Pública, Rio de Janeiro, Brazil.

Eds Frutuoso, P. F, Faertes, D. (2001). *Análise de Confiabilidade e Análise Quantitativa de Risco – Curso de Avaliação e Gerenciamento de Riscos* , Instituto Politécnico da UFRJ.

Galante, E. B. F., Haddad, A. N. (2009). *Risk Analysis procedures for explosives manufacturing. Reliability, Risk and Safety. Theory and Applications.* Volume 2, Pgs 1117 – 1120. Crc press, taylor and francis group. ISBN 978-0-203-85975-9 (e-book) / ISBN 978-0-415 – 55509-8

Nota, G., Gregorio, M. P. D. (2010). *A Model for Process Oriented Risk Management, Advances in Risk Management,* Giancarlo Nota (Ed.), ISBN 978-953-307-138-1, InTech, Available from: http://www.intechopen.com/articles/show/title/a-model-for-process-oriented-risk-managenent

Haddad, A.N., Morgado, C. V., & De Souza, D. I. (2008). Health, Safety and Environmental Management Risk Evaluation Strategy: Hazard Matrix Application Case Studies, *Proceedings of the 2008 IEEE IEEM,* ISBN 978-1-4244-2630-0.

International Standard Organization. (2009). Risk Management – Principles and Guidelines. ISO 31000:2009,

International Standard Organization. (2000). Sistemas de gestão da qualidade – Requisitos. NBR ISO 9001: 2000,

International Standard Organization. (2004). Sistemas da gestão ambiental - Requisitos com orientações para uso. ISO 14001:2004, 2004

Kletz T. A. (1988). *What Went Wrong? – Case Histories of Process Plant Disasters,* Gulf Publishing Company, Texas, EUA

Lees, F.P. (1996). Hazard Identification, Assessment, and Control, vols. 1–3, *Journal of Loss Prevention in the Process Industries* Butterworth-Heinemann, Oxford

Hammer, W. (1972) *Handbook of System and Product Safety.* Prentice-Hall.

Hanea, D., Ale, B. (2009). *Risk of human fatality in building fires: A decision toolusing* Bayesian networks. Web of Science

Meyer, R. (1977). Explosives. Gebr. Diesbach. Germany.

Moran, M J.; Shapiro, Howard N. (1996). *Fundamentals of Engineering Thermodynamics* (3rd ed.), J. Wiley & Sons. ISBN 0-471-07681-3, New York; Toronto

Muniz, T. P. (2011). *Gerenciamento de riscos, uma ferramenta básica de segurança: estudo prático em uma unidade marítima de exploração de hidrocarbonetos*. Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

Sa, T. S. (2008). *Análise Quantitativa de Risco Aplicada àIndústria de Gases,* Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Au, S. K., Wang, Z-H., & Lo, S-M. (2007). *Compartment fire risk analysis by advanced Monte Carlo simulation*, Web of Science

Tannehill, J. C., Anderson, D. A., & Pletcher, R. H. (1997). *Computational Fluid Mechanics and Heat Transfer* (2nd Edition), ISBN 1-56032-046-X, Taylor & Francis, USA.

United Kingdom. (2002). A risk Management Standard, *Association Of Insurance And Risk Managers; National Forum For Risk Management In The Public Sector; Institute Of Risk Management,* United Kingdom: AIRMIC, ALARM, IRM,

Krieger, G.R., Montgomery, J.F. (1997). *Accident Prevention Manual for Business and Industry, Administration & Programs volume (11th edition),* Illinois, EUA

United States. Mil-Std-882d: (2001). Standard Practice for system safety.

United States. Mil-Std-1629: (2000). Standard: Failure Mode and Effect Analysis.

Nfpa. (2002). *SFPE Handbook of Fire Protection Engineering*. National Fire Protection Association, - 3rd Edition. Quincy, Massachusetts. ISBN 087765-451-4,

Vincoli, J. W. (2006). *Basic Guide to System Safety* (second edition), Titusville, John Wiley & Sons, Inc. ISBN-13: 978-0-471-72241-0, Titusville, Florida, USA

Urbanski, T. (1964). *Chemistry and Technology of Explosives,Volume 01,* Pergamon Press, Department of Technology, Politechnika Warszawa, London.

The Deterministic and Stochastic Risk Assessment Techniques in the Work Sites: A FTA-TRF Case Study

P.K. Marhavilas and D.E. Koulouriotis
Democritus University of Thrace, Xanthi,
Greece

1. Introduction

Occupational accidents have a major impact upon human integrity and they also create high costs for the social welfare system in a country. Furthermore, risk analysis is an essential process for the safety policy of a company, having as main aim the effacement of any potential of damage. The diversity in risk analysis approaches is such that there are many appropriate techniques for most circumstances and the choice has become more a matter of taste (Reniers *et al.*, 2005b). The risk assessment is an essential and systematic process for assessing the impact, occurrence and the consequences of human activities on systems with hazardous characteristics (Van Duijne *et al.*, 2008) and constitutes a needful tool for the safety policy of a company. We can consider risk as a quantity, which can be measured and expressed by a mathematical relation, under the help of real accidents' data. The risk assessment is generally achieved by a deterministic and/or a stochastic method. The first one is classified into three main categories; 1) the qualitative, 2) the quantitative, and 3) the hybrid techniques (qualitative-quantitative), while the second one includes the classic statistical approach and the accident forecasting modelling (Marhavilas, 2009a, 2009b; Marhavilas and Koulouriotis, 2007, 2008, 2011, 2012; Marhavilas *et al.*, 2011a, 2011b).

On the other side, few comparative studies have been performed on different stochastic and deterministic risk assessment methods. Thus, most researchers primarily focus on longitudinal surveys concerning an individual method (Zheng and Liu, 2009). However, an individual method cannot achieve the best risk-assessment result in the worksites, and future perspectives should focus on the parallel application of a deterministic (DET) and a stochastic (STO) process (Marhavilas and Koulouriotis, 2012).

In fact, the contribution of the development and elaboration of STODET processes, to the health and safety science, could be focused (Marhavilas and Koulouriotis, 2011) on:

a. The *improvement* of the risk assessment techniques
b. The *comparison* of their outcome risk estimation results
c. The *enrichment* of the scientific literature with new tools

In two recent works, we presented the development and the application of two STODET risk assessment methods based on the combination of special stochastic (STO) and deterministic (DET) processes, like the PRAT-TRF technique (Marhavilas and Koulouriotis, 2011), and the PRAT-TSP-SRE technique (Marhavilas and Koulouriotis, 2012).

Taking into account the above reasons, we proceed to the development of a new STO-DET risk assessment framework by combining the deterministic FTA ("fault-tree-analysis") technique and the stochastic TRF ("time at risk failure)" model, and apply it on the worksite of an industrial productive procedure. The objective of this work is there fore twofold;

a. We present a new risk assessment framework based on the combination of the deterministic FTA ("fault-tree-analysis") technique and the stochastic TRF ("time at risk failure)" model
b. We apply this FTA-TRF process on an industrial worksite to test its usefulness

This chapter consists further of three sections: 1) an overview of the main stochastic and deterministic risk analysis and assessment techniques, 2) the development of a new STODET risk assessment framework based on FTA-TRF combination, and 3) a case study for the simultaneous application of FTA and TRF techniques in industry.

2. Risk analysis and assessment techniques

There are various risk analysis and assessment techniques, which are included in the literature (e.g. Baker *et al.*, 1998; Kontogiannis *et al.*, 2000; Reiners *et al.*, 2005a, 2005b; Marhavilas and Koulouriotis, 2007, 2008, 2011; Marhavilas *et al.*, 2011a, 2011b; Doytchev and Szwillus, 2008; Marhavilas, 2009a, 2009b; Colli *et al.*, 2009; Johansson *et al.*; 2009; Lim and Zhang, 2009). A basic classification of the risk analysis and assessment methodologies based on the literature, includes the deterministic (DET) approach and the stochastic (STO) approach (Marhavilas and Koulouriotis, 2011). Furthermore, DET techniques are classified into three main categories: (a) the qualitative, (b) the quantitative, and (c) the hybrid techniques (qualitative-quantitative, semi-quantitative) (Marhavilas *et al.*, 2011a), while STO method includes the Classic Statistical Approach (CSA) and the Accident Forecasting Modelling (AFM) (Marhavilas and Koulouriotis, 2011). The reader could find a thorough presentation of the main deterministic and stochastic risk assessment and analysis techniques in the work of (Marhavilas and Koulouriotis 2011). Briefly stated, these approaches can be classified as follows:

A. Deterministic Techniques:

A.1 Qualitative Techniques:

- Checklists
- What-if analysis
- Safety audits
- Task Analysis
- STEP technique
- Hazard and Operability (HAZOP) study

A.2 Quantitative Techniques:

- The proportional risk assessment technique (PRAT)
- The decision matrix risk assessment (DMRA)

- Quantitative risk measures of societal risk
- The QRA (Quantitative Risk Assessment) tool
- Quantitative assessment of domino scenarios (QADS)
- The CREA (Clinical Risk and Error Analysis) method
- The weighted risk analysis (WRA)

A.3 Hybrid Techniques:

- Human Error Analysis Techniques (HEAT)
- Fault tree analysis (FTA)
- The ETA method (Event Tree Analysis)
- The RBM Method (Risk-based Maintenance)

B. Stochastic Techniques:

B.1 Classic statistic approach:

- Epistemic Models: The PEA (Predictive, Epistemic Approach) method
- Probability distributions of failure and reliability:
    - Exponential distribution
    - Normal distribution
- Event data-models
    - Constant Failure and Repair Rate Model (Rate Model)
    - Mean Time to Failure and Repair Model (MTTF/MTTR Model)
    - Time at Risk Failure (TRF) Model
    - Rate/MTTR Model

B.2 Accident forecasting modelling:

- Time-Series Stochastic Processes/Time-Series Method (TSM)
- Markov chain analysis
- Grey model
- Scenario analysis
- Regression method
- Neural networks
- Bayesian Networks

Fault tree analysis (FTA) is a deductive technique focusing on one particular accident event and providing a method for determining causes of that event. Fault trees are constructed from events and gates. Basic events can be used to represent technical failures that lead to accidents while intermediate events can represent operator errors that may exacerbate technical failures. The gates of the fault trees can be used to represent several ways in which machine and human failures combine to give rise to the accident. For instance, an AND-gate implies that both initial events need to occur in order to give rise to the intermediate event. Conversely, an OR-gate means that either of two initial events can give rise to the intermediate event. In the context of accident analysis, an OR-gate implies lack of evidence; as more evidence becomes available we can become more certain which of the two initial events were true (Vesely *et al.*, 1981 ; Kontogiannis *et al.*, 2000; Harms-Ringdahl, 2001; Reniers *et al.*, 2005a; Yuhua and Datao, 2005; Hong *et al.*, 2009).

On the other side, we present here basic elements referring to the study of the stochastic behavior of single-component Occupational Health and Safety System (OHSS) concerning the worksite of a company and being subjected to failures (breakdowns) by observing them over a period of time. Let us simplify things by assuming that the system is put to work at the instant t = 0 for the first time and that it presents a single mode of failure. The component, starting a lifetime period at the instant t = 0, is functioning for a certain period of time X_1 (random) at the end of which it breaks down. It remains in this state for a period of time Y_1 (random) during its replacement (or repair) and, at the end of this time, the component is again put to work and so on. In this case, the system is said to be repairable. In the contrary case, when the component breaks down and continues to remain in this state, the system is said to be non-repairable (Limnios, 2007; Haimes, 2009; Marhavilas and Koulouriotis, 2011). Let X be a random variable (r.v.) representing the lifetime of the system with F its cumulative distribution function (c.d.f.): $F(\mathrm{t})=P(X\leq \mathrm{t})$. If F is absolutely continuous, the random variable X has a probability density function (p.d.f.) f and can be written as:

$$f(t) = \frac{d}{dt}F(t) = \operatorname*{Lim}_{\Delta t \to 0} \frac{P(t < X \leq t + \Delta t)}{\Delta t} \tag{1}$$

Reliability: The complementary function of F, noted as $\overline{F}$, is the reliability (or probability of success) of the system, noted as $R(t)$. That is to say:

$$R(t) = \overline{F} = 1 - F(t) = P(X > t) \tag{2}$$

Where:

$$R(t) = \int_t^{\infty} f(u)du \ , \ R(0) = 1 \ , \ R(+\infty) = 0 \tag{3}$$

The exponential distribution is the most frequently used in relation to the reliability of systems. A system whose stochastic behavior is modeled by an exponential distribution is a system without memory, that is to say, for t>0, x>0, we have P(X>t+x | X>t)=P(X>x). For the exponential distribution we have for x≥0:

$$f(t) = \lambda e^{-\lambda t}, F(t) = 1 - e^{-\lambda t}, R(t) = e^{-\lambda t}, \lambda(t) = \lambda \text{ (the failure rate)} \tag{4}$$

Although, this distribution gives good modeling for the lifetime of electronic components, its use in other fields, such as in risk analysis for the modeling of OHSS in the worksites is justified.

Moreover, for a quantitative analysis to be performed, event failure and repair data-models could be specified for the events in the study of the stochastic behavior of single-component occupational health and safety systems (OHSS) being subjected to failures over a period of time. Some of the usual event data-models (Limnios, 2007; Isograph, 2008) are:

Mean Time to Failure and Repair Model (MTTF/MTTR Model): This model is the same as the constant failure and repair rate model described above, except that the parameters entered by the user are the mean time to failure (MTTF) (or mean time between failures (MTBF)) and the mean time to repair (MTTR). These parameters are related to the failure and repair rates by the following expressions:

$$\lambda = \frac{1}{MTTF}, \ \mu = \frac{1}{MTTR} \text{ (component repair rate)} \tag{5}$$

Time at Risk Failure (TRF) Model: This model allows users to specify a 'time at risk' that differs from the system lifetime. The model is useful for representing component failures that only contribute to system failure during certain phases of the lifetime of the system or duration of a mission. The unavailability of events (or the probability of failure) associated with this model are calculated using the expression

$$Q(t) = 1 - e^{-\lambda T} \tag{6}$$

where λ = failure rate, T = time at risk

It is worth noting that most researchers primarily focus on surveys concerning an individual method (Zheng and Liu, 2009). However, an individual method cannot achieve the best risk-assessment result in the worksites, and future perspectives should focus on the parallel application of a deterministic and a stochastic process (Marhavilas and Koulouriotis, 2012). So, we proceed to the development of a new STODET risk assessment framework by combining the deterministic FTA ("fault-tree-analysis") technique and the stochastic TRF ("time at risk failure)" model, and apply it on the worksite of an industrial productive procedure.

3. A risk assessment framework based on FTA-TRF combination

Below, we present a new risk assessment framework based on a stochastic-deterministic (STODET) quantified risk evaluation according to function of Figure 1. In addition, Figure 2 shows the flowchart of this risk assessment framework, as a part of the risk management process, using safety aspects–guidelines of ISO/IEC (1999, 2009), (Høj and Kröger 2002), (BS 8800 2004), (van Duijine *et al.* 2008), (Suddle 2009), (Marhavilas *et al.* 2011b) and (Marhavilas and Koulouriotis 2011). This framework consists of three distinct phases: (a) the risk analysis, (b) the quantified risk evaluation and c) the risk assessment and safety-related decision making. The first phase includes the hazard sources' identification and the risk consideration/calculation, while the second one includes the stochastic and deterministic processes. The module #B emphasizes the application of a STODET quantified risk-evaluation, which is implemented by the simultaneous application and the jointly evaluation of the TRF ("Time at Risk Failure") stochastic model and the deterministic process of FTA ("Fault Tree Analysis").

3.1 Risk analysis

Risk analysis or safety analysis is an approach to identify the factors that may lead to accidents, and constitutes a systematic use of available information to identify hazards ((ISO/IEC, 1999; Marhavilas *et al.*, 2011b). In general, ''danger'' should be defined *as an attribute of substances or processes, which may potentially cause harm*. Furthermore, "risk" has been defined as *the chance that someone or something that is valuated will be adversely affected by the hazard*, and also as *a measure under uncertainty for the severity of a hazard* (Høj and Kröger, 2002) while "hazard" is *any unsafe condition or potential source of an undesirable event with potential for harm or damage* (Reniers *at al.* 2005a).

3.1.1 Identification of hazard sources

Danger can be separated in two major categories: "Direct" and "indirect". Direct danger includes the apparent accidents (fractures, scratches, tool injuries etc) and indirect, the danger which is not apparent and devious and comes from the exposure in sources of hazard, such as electromagnetic radiation, noise, weather conditions and raising weights, that cause hard-hearing, cancer, dizziness, respiratory problems and cardiac problems. The identification of hazard sources is usually comprised of specifying one or more scenarios of risks. A risk scenario describes an interaction between a person and a system or product that possesses hazardous characteristics. It describes the activity of the person(s) involved, the hazard(s), the external factors of the situation and the potential injury. Injury (real accidents') data are the primary source of evidence to establish risk scenarios that describe critical pathways to injury. Furthermore, expert opinions are a significant source for creating risk scenarios. Experts rely on their technical knowledge about the system (with its intrinsic hazards) and the productive process, but they also need to apply their knowledge in order to identify relevant and plausible scenarios, for more information see (BS8800:1996, 1996; ILO-OSH, 2001; BS8800:2004, 2004; BS18004:2008, 2008; BS OHSAS18001:2007, 2007; OHSAS 18002:2008, 2008; Marhavilas *et al.* 2011b).

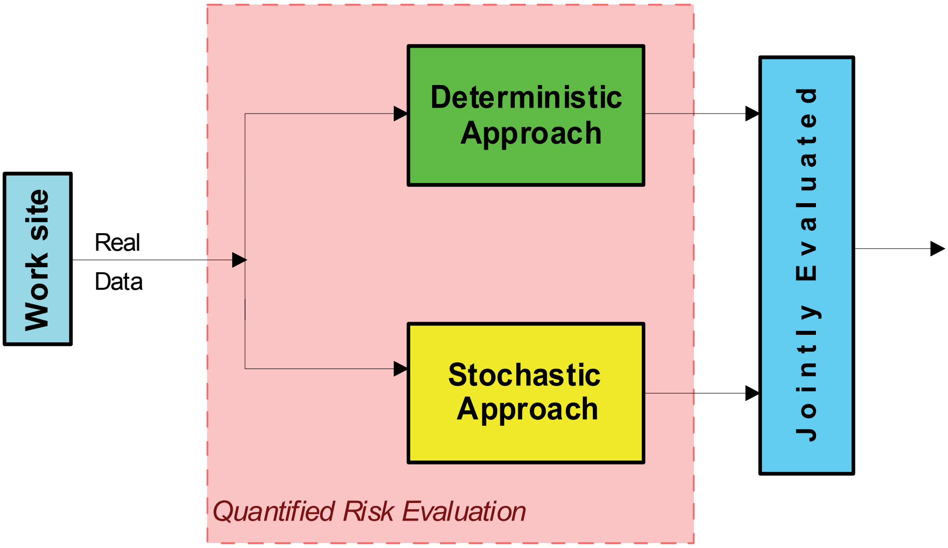


Fig. 1. The combination of a stochastic and a deterministic (STODET) approach in the quantified risk evaluation

The method used to analyze occupational risk follows the algorithm in Figure 2, and in that respect the following must be taken into account: a) gathering of information on the system (by using questionnaires, interviews and checklists) provides the basis for analysis and must be carried out systematically, b) the entire system and its activities should be included in the analysis, which must be designed systematically so as not to overlook important elements, c) the risks to which these hazards give rise must be assessed in a consistent manner, and d) a systematic approach is required even when safety proposals are to be generated and evaluated (Harms-Ringdahl, 2001; Marhavilas *et al.* 2011b; Marhavilas and Koulouriotis, 2012).

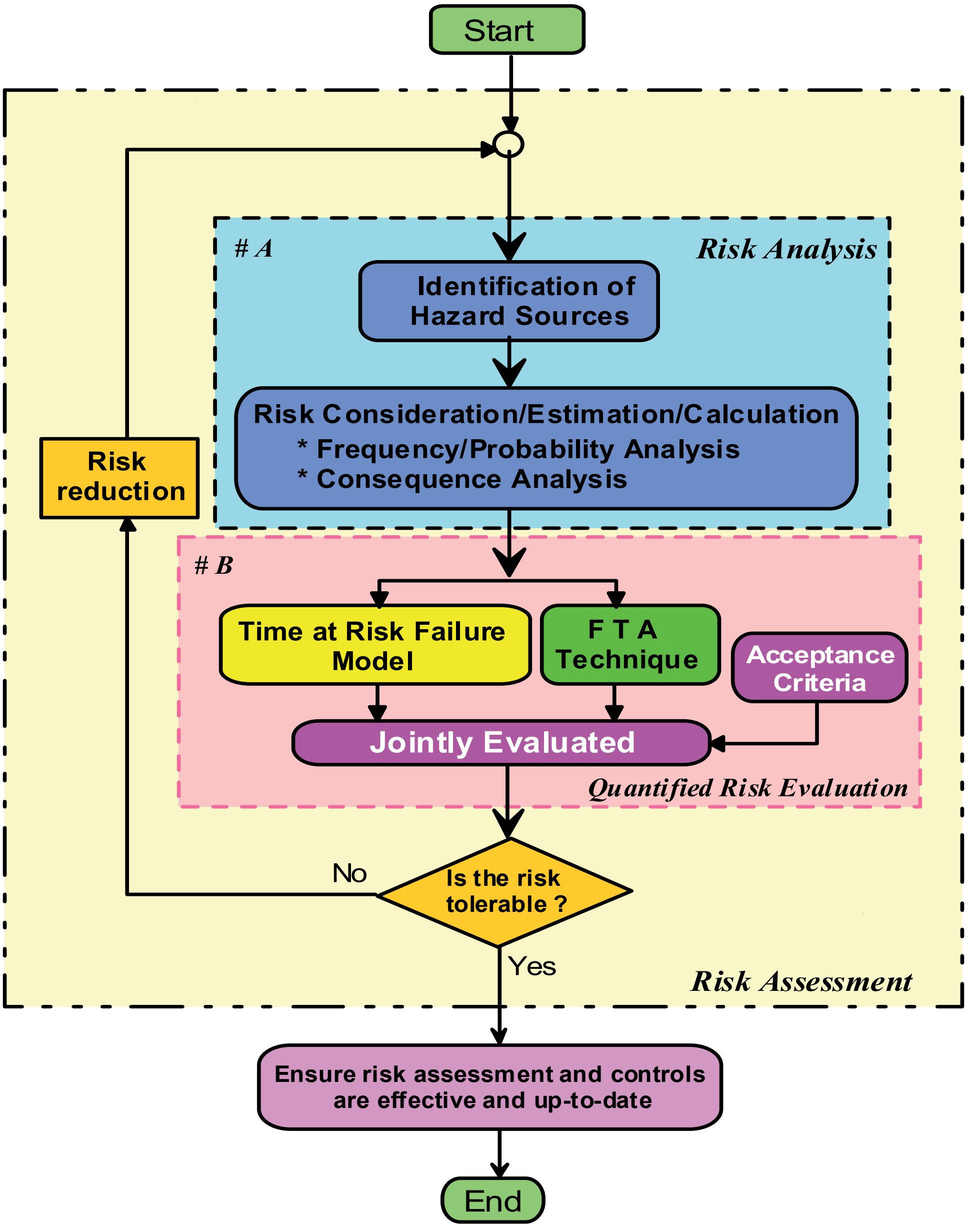

Start
# A
Risk Analysis
Identification of Hazard Sources
Risk Consideration/Estimation/Calculation
* Frequency/Probability Analysis
* Consequence Analysis
Risk reduction
# B
Time at Risk Failure Model
F T A Technique
Acceptance Criteria
Jointly Evaluated
Quantified Risk Evaluation
No
Is the risk tolerable ?
Yes
Risk Assessment
Ensure risk assessment and controls are effective and up-to-date
End

Fig. 2. The flowchart of an alternative risk assessment framework by including a stochastic and a deterministic (STODET) approach, as a part of the risk management process, based on safety aspects–guidelines of (ISO/IEC 1999, 2009), (Høj and Kröger 2002), (BS 8800 2004), (van Duijine *et al.* 2008), (Suddle 2009), (Marhavilas *et al.* 2011b) and (Marhavilas and Koulouriotis 2011).

3.1.2 Risk consideration

The risk consideration is achieved by the following steps (Marhavilas *et al.* 2011b; Marhavilas and Koulouriotis, 2011):

- **Estimation of the likelihood of hazard sources occurrence (P):** The occurrence of injury/damage (or the likelihood of hazard-sources occurrence) may depend on several factors related to the actual interaction of the employee with a hazard source and also to the energy transferred during this interaction. This likelihood depends on the (hidden) potential energy that may become active during unsafe behaviour, the energy absorbing capacity, resilience and other qualities of the human body (Marhavilas *et al.*, 2011b).
- **Estimation of the consequences' severity (S):** The risk assessment techniques require the estimation of the injury's seriousness gradation (i.e. the consequences' severity). Of course, severity is a subjective issue, because some events, such as cuts, possibly have non-serious effects, while others, such as injuring due to slips, may become more significant. To solve this problem, we can gradate the severity of injury (or damage) by specifying the **level** of employee's **inability** in association with the **duration** that the employee is absent from his work according to the obligations of Law 3850/2010 of the Greek State (HR, 2010; Marhavilas *et al.* 2011b).
- **Estimation of the frequency-level of exposure to hazard sources (f)**: The probability that a dangerous scenario may occur, depends on the frequency of exposure to the hazard sources. It is worth to note that we can estimate the gradation of the frequency-level by using information about workers' activities which may give an indication about the frequency of a risky activity (Marhavilas *et al.* 2011b). Furthermore, the gradation of the frequency-level can be illustrated by the *Frequency (or Exposure) Factor* in association with the frequency of appearance of a potential hazard source (or an undesirable event), and according to the results of the work of Marhavilas and Koulouriotis (2008, their Table 3).

3.2 Quantified risk evaluation

Quantified risk evaluation techniques enable risk assessors to scale their appreciation of the severity of the short and long term consequences of accidents and the factors that influence the occurrence of an accident scenario. The methods of quantified risk evaluation need to be as precise as possible to differentiate the risk level of various activities (Marhavilas *et al.*, 2011b). Below, we explain (in association with module #B of Figure 2) the usage and implementation of the STO-DET quantified risk-evaluation process, by the combination and the jointly evaluation of the *TRF* stochastic model and the *FTA* deterministic process.

3.2.1 A stochastic model: "Time at risk failure model"

According to this model the probability of failure is expressed using the relation $Q(t) = 1 - e^{-\lambda T}$ (λ is the failure rate, T is the time of exposure). It is worth to note that there is a magnitude (called as "mean time"), that plays a very important role in connection with the reliability and the probability of failure of the occupational health and safety systems (OHSS). One significant "mean time" is the "*mean time between failures*" (MTBF), which is expressed by the relation: $\lambda = 1 / MTBF$

3.2.2 A deterministic model: “FTA” model

“FTA” is constructed from events and gates. Basic events can be used to represent technical failures that lead to accidents while intermediate events can represent operator errors that may exacerbate technical failures. The gates of the fault trees can be used to represent several ways in which machine and human failures combine to give rise to the accident.

3.3 The decision making

In the risk management, it is fundamental to distinguish between the risk assessment process and the decision-making process (ISO/IEC Guide-73, 2009; Marhavilas *et al.*, 2011b). In particular, the risk assessment is a part of the risk management process, ending up with the decision making (Salvi and Gaston, 2004). In addition, the risk assessment is a tool used to measure the risk, characterized by the likelihood and severity of specific events, and can further be a basis for decision-making (Høj and Kröger, 2002). Risk-based decision-making processes are naturally based on the risk assessment criteria, but could integrate also other criteria that can be cultural, economical, ethical etc (Salvi and Gaston, 2004).

We will now present a case study in order to illustrate our approach, i.e. the simultaneous application of FTA and TRF techniques, in industry.

4. Case study: Application of FTA-TRF on an Industrial worksite

In the following passages, we proceed to the application of FTA-TRF process on the worksite of a tobacco-industry’s chemical-laboratory (TICL).

4.1 Deterministic approach: Application of FTA

Figure 3 shows the FTA construction concerning one of the more important hazard-sources that exist in a tobacco-industry chemical-laboratory i.e. the “*EMPLOYEES AMBUSTION/BURN*”. This hazard-source has been determined by the application of Figure’s 2 risk-analysis phase on the worksite of the chemical-laboratory. In particular, we use in the FTA chart two types of graphic symbols (Limnios, 2007): (i) the OR/AND logic gates/operators, and (ii) the fundamental events (circle for an elementary basic event and rectangle for a top or intermediate event). We have designated the set of basic events of the fault tree by the word “*EVENT*”, that is, *EVENT*= {*EVENT-1*, ..., *EVENT-n*}. The numbers assigned to the basic events on the fault tree correspond to the indices of the events. The OR operator describes the failure of a series system, while the AND one the failure of a parallel system. Similar FTA charts can be constructed for all hazard-sources exist in the industry’s worksite.

4.2 Stochastic approach: Application of the “time at risk failure” model

In this section, we apply the “*TRF*” model by using the methodological background of Sections 2 and 3.2.1. As an example, on a single-component OHSS of TICL, like the “*EMPLOYEES AMBUSTION/BURN*”, the occurrence frequency of E1-event is estimated to be f=2400 yr^{-1} and the estimated likelihood of accidents is P=10^{-4}, which means that the estimated number of accidents (per year) is N=0.24, implying that MTBF=36500 hr and λ=2.73973E-05 hr^{-1}. By using as exposure time (T) the duration of 8760 working hours (w.hrs) i.e. one full-time working year (w.yr), we find that the reliability of TICL OHSS due to E1 is R≅79% and the probability of failure is Q≅21%. Furthermore, Table 1 illustrates the calculated results of the TRF application on the TICL OHSS, concerning all basic events E1-E12 of the FTA construction of Figure 3.

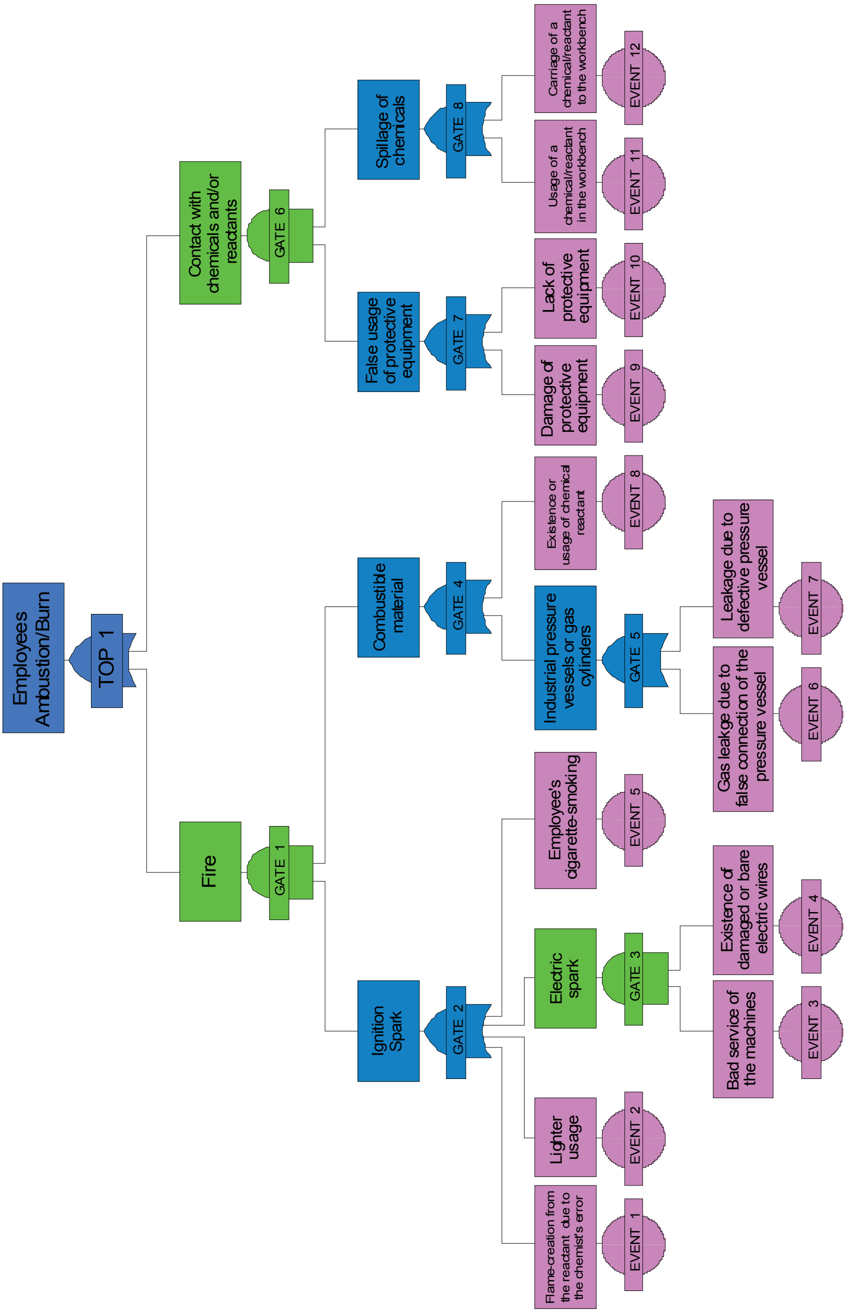

Employees Ambustion/Burn
TOP 1
Contact with chemicals and/or reactants
GATE 6
Fire
GATE 1
Spillage of chemicals
GATE 8
False usage of protective equipment
GATE 7
Combustible material
GATE 4
Ignition Spark
GATE 2
Carriage of a chemical/reactant to the workbench
EVENT 12
Usage of a chemical/reactant in the workbench
EVENT 11
Lack of protective equipment
EVENT 10
Damage of protective equipment
EVENT 9
Existence or usage of chemical reactant
EVENT 8
Industrial pressure vessels or gas cylinders
GATE 5
Leakage due to defective pressure vessel
EVENT 7
Gas leakge due to false connection of the pressure vessel
EVENT 6
Employee's cigarette-smoking
EVENT 5
Electric spark
GATE 3
Existence of damaged or bare electric wires
EVENT 4
Bad service of the machines
EVENT 3
Lighter usage
EVENT 2
Flame-creation from the reactant due to the chemist's error
EVENT 1

Fig. 3. The FTA construction concerning one of the more important hazard-sources exist in a tobacco-industry chemical-laboratory, the "*EMPLOYEE'S AMBUSTION/BURN*"

4.3 Joint evaluation of FTA-TRF combination

We proceed to the joint evaluation of FTA-TRF combination. More specifically, the probabilistic assessment of the FTA consists of calculating the probability of the top event TOP-1 (Figure 3) starting from the probabilities Q_i (i=1, …, 12) of the basic events (E1-E12) which are illustrated in Table 1. This can be done directly because the FTA construction of Figure 3 does not possess any repeated event (according to the rules of Limnios' (2007) work), and it is carried out with a simple approach, which consists of climbing back up the FTA by starting from its primary operators up to the top event and using the Boolean algebra (algebra of events) (Haimes, 2009).

Thus, we define the following set of equations:

Q (TOP1) = Q(G1)+Q(G6)-Q(G1)*Q(G6)

Q(G1)=Q(G2)* Q(G4)

- Q(G2)=Q(E1)+Q(E2)+Q(G3)+Q(E5)- Q(E1)*Q(E2)*Q(G3)*Q(E5)
  - Q(G3)=Q(E3)*Q(E4)
- Q(G4)=Q(G5)+Q(E8)- Q(G5)*Q(E8)
  - Q(G5)=Q(E6)+Q(E7)- Q(E6)*Q(E7)

Q(G6)=Q(G7)*Q(G8)

- Q(G7)=Q(E9)+Q(E10)- Q(E9)*Q(E10)
- Q(G8)=Q(E11)+Q(E12)-Q(E11)*Q(E12)

By using the numbers of Table 1 we take the following results:

Q(G7)=0.21+0.11-0.21*0.11=0.2969
Q(G8)=0.21+0.11-0.21*0.11=0.2969
Q(G6)=0.2969*0.2969=0.08815
Q(G5)=0.0+0.12- 0.0*0.12=0.12
Q(G4)=0.12+0.21-0.12*0.21=0.3048
Q(G3)=0.0*0.0=0.0
Q(G2)=0.21+0.11+0.0+0.38-0.21*0.11*0.0*0.38=0.7
Q(G1)=0.7*0.3048=0.213

So the probability of failure Q of the single-component TICL's OHSS due to the "*EMPLOYEES AMBUSTION/BURN*" hazard source is

Q(TOP1)=0.213+0.08815-0.213*0.08815=**0.282374** or **28.2%**

This means that this is a ***medium-risky hazard source*** because 10%<Q<50% , according to the work of (Marhavilas and Koulouriotis 2011).

The same process for the calculation of Q can be applied in all hazard-sources determined by the risk-analysis on the TICL's OHSS, which could classify them into three categories like in the work of (Marhavilas and Koulouriotis 2011) as follows:

- *High-risky sources* (Q≥50%)
- *Medium-risky sources* (10%<Q<50%)
- *Low-risky sources* (Q≤10%)

EVENT	Occurrence Frequency *(f)* *[yr^{-1}]*	Likelihood *(P)*	Est. number of acc. per year *(N=P*f)* *[acid./yr]*	MTBF *[hr]*	λ=1/MTBF *[hr^{-1}]*	T *[w.hr]*	λ*T	Reliability *R=e$^{-λt}$*	Prob. of failure *Q=1-R*
E1	2,400	10^{-4}	0.24	36,500	2.74E-05	8,760	0.240	0.79	0.21
E2	1,200	10^{-4}	0.12	73,000	1.37E-05	8,760	0.120	0.89	0.11
E3	48	10^{-4}	0.0048	1,825,000	5.48E-07	8,760	0.005	1.00	0.00
E4	48	10^{-4}	0.0048	1,825,000	5.48E-07	8,760	0.005	1.00	0.00
E5	4,800	10^{-4}	0.48	18,250	5.49E-05	8,760	0.480	0.62	0.38
E6	6	10^{-4}	0.0006	14,600,000	6.85E-08	8,760	0.001	1.00	0.00
E7	-	-	-	67,927 (*)	1.47E-05	8,760	0.129	0.88	0.12
E8	2,400	10^{-4}	0.24	36,500	2.74E-05	8,760	0.240	0.79	0.21
E9	240	10^{-4}	0.024	365,000	2.74E-05	8,760	0.240	0.79	0.21
E10	1,200	10^{-4}	0.12	73,000	1.37E-05	8,760	0.120	0.89	0.11
E11	2,400	10^{-4}	0.24	36,500	2.74E-05	8,760	0.240	0.79	0.21
E12	1,200	10^{-4}	0.12	73,000	1.37E-05	8,760	0.120	0.89	0.11

(*) From the technical specifications

Table 1. Depiction of the results of TRF application on the OHSS of a tobacco-industry's chemical-laboratory, concerning all basic events E1-E12 of the FTA construction of Figure 3

5. Discussion

We can consider the risk as a quantity, which can be estimated and expressed by a mathematical relation, under the help of real accidents' data. The risk assessment is generally achieved by a deterministic and/or a stochastic method. The diversity in risk analysis procedures is such that there are many appropriate techniques for any circumstance and the choice has become more a matter of taste. However, an individual method cannot achieve the best risk-assessment result in the worksites and future perspectives should focus on the parallel application of a deterministic technique with a stochastic one.

The objective of this work is twofold a) present of a new risk assessment framework based on the combination of the deterministic FTA ("fault-tree-analysis") technique and the stochastic TRF ("time at risk failure)" model, and b) apply the FTA-TRF process on an industrial worksite.

In particular, the new alternative risk assessment framework we develop is achieved in Figure 1 by the combination of a stochastic and a deterministic process (STODET). This process consists of three distinct phases: (a) the risk analysis, (b) the quantified risk evaluation and c) the risk assessment and safety-related decision making. The first phase includes the hazard sources' identification and the risk calculation, while the second one the stochastic and deterministic processes (Figure 2 illustrates its flowchart as a part of the risk management process). To continue, the STODET quantified risk-evaluation consists of the combined evaluation of the *TRF* ("Time at Risk Failure") stochastic model and the *FTA* ("Fault Tree Analysis") deterministic technique (module #B of Figure 3 emphasizes it).

Furthermore, in order to present a case study, we proceeded to the application of FTA-TRF on the worksite of a tobacco-industry's chemical-laboratory (which is situated in Thrace, Greece) by using real data of undesirable events and accidents. So, the probability of failure Q of the single-component TICL's OHSS due to the "*EMPLOYEES AMBUSTION/BURN*" hazard source was calculated to be Q=28.2%, which means it is a *medium-risky hazard source* because 10%<Q<50% (Marhavilas and Koulouriotis, 2011). The same process for the calculation of Q can be applied in all hazard-sources determined by the risk-analysis on the TICL's OHSS, which could classify them into three categories: (i) high-risky sources (Q≥50%), (ii) medium-risky sources (10%<Q<50%), and (iii) low-risky sources (Q≤10%).

6. Future work and closure

In a future work, we are planning: (i) the development of another risk assessment framework including more stochastic and deterministic techniques, and (ii) the application on other industrial OHSS. This means that we have the ability to combine more different stochastic techniques like Markov chains, the grey model, neural networks, the scenario analysis, the regression method, Bayesian networks etc (Zheng and Liu, 2009; Marhavilas and Koulouriotis, 2011), with more deterministic techniques like DMRA (for more information: Marhavilas *et al.* 2011a, 2011b). In the work of (Marhavilas *et al.* 2011a; see their Table 8), there is a comparison of the various DET methodologies focusing on their advantages/disadvantages, and highlighting areas of future improvements, while in the work of (Zheng and Liu 2009; see their table 8), a comparison of different STO models, a fact which could help the reader to select the best STO-DET combination.

Apart from the exponential distribution, other usual probability distributions dealing with the reliability of health and safety systems which could be applied and tested are the following (Limnios, 2007; Marhavilas and Koulouriotis 2011):

- **Normal distribution**: It is used for modeling the duration and the lifetime of the systems and expressed by the relations

$$f(t)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\ ,\ R(t)=\int_t^{\infty}f(y)dy\ ,\ \lambda(t)=\frac{f(t)}{R(t)} \tag{7}$$

Where μ is the average and σ is the standard deviation.

- **Log-Normal distribution**: It is expressed by the relations

$$f(t)=\frac{1}{\sqrt{2\pi}\sigma t}e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}}\ \text{(t≥0)},\ R(t)=\int_t^{\infty}f(y)dy\ ,\ \lambda(t)=\frac{f(t)}{R(t)} \tag{8}$$

- **Weibull distribution**: Due to the vast variations of form that it can take up according to the values of its parameters, the Weibull distribution is used in many domains of reliability, particularly in those concerned with the reliability of mechanical components. It is expressed by the relations

$$f(t) = \frac{\beta}{\eta^{\beta}}(t-\gamma)^{\beta-1} \cdot e^{-\frac{(t-\gamma)^{\beta}}{\eta}},\ R(t) = e^{-\frac{(t-\gamma)^{\beta}}{\eta}},\ \lambda(t) = \frac{\beta(t-\gamma)^{\beta-1}}{\eta^{\beta}} \quad (9)$$

Where β is the parameter of form, η the parameter of scale and γ the parameter of localization. For β= 1 and γ= 0, we will obtain the exponential distribution.

As a general observation in the end, we believe that the usage of the new STODET alternative risk assessment scheme, presented here, would help industries achieve better occupational risk protection.

7. References

Baker, S., Ponniah, D., Smith, S. (1998). Techniques for the analysis of risks in major projects. Journal of the Operational Research Society, 49, 6, 567-572.

BS 18004:2008 (2008). Guide to achieving effective occupational health and safety performance. ISBN:978 0 580 529108.

BS 8800:1996 (1996). Guide to occupational health and safety management systems. ISBN:0 580 25859 9.

BS 8800:2004 (2004). Guide to occupational health and safety management systems. ISBN:0 580 43987 9.

BS OHSAS 18001:2007 (2007). Occupational health and safety management systems. Requirements. ISBN:978 0 580 59404 5, p.34.

Colli, A., Serbanescu, D., Ale, B.J.M. (2009). Indicators to compare risk expressions, grouping, and relative ranking of risk for energy systems: Application with some accidental events from fossil fuels. Safety Science, 47, 5, 2009, 591-607.

Doytchev, D.E., Szwillus, G. (2008). Combining task analysis and fault tree analysis for accident and incident analysis: A case study from Bulgaria. Accident Analysis and Prevention, doi:10.1016/j.aap.2008.07.014.

Haimes, Y.Y. (2009). Risk modeling, assessment, and management. A John Wiley & Sons Inc. publication, 3rd edition ISBN 978-0-470-28237-3.

Harms-Ringdahl, L. (2001). Safety Analysis, Principles and Practice in Occupational Safety. 2nd edition, ISBN: 9780415236553, p.302, CRC Press.

Hellenic Republic (HR) (2010). Law 3850/2010: Code of Health and Safety in the worksites, National Printing-House, Issue 1, Part 84/2.6.2010, pp. 1721-1750.

Høj, N.P., Kröger, W. (2002). Risk analyses of transportation on road and railway from a European Perspective. Safety Science, 40, 1-4, 337-357.

Hong, E-S., Lee, I-M., Shin, H-S., Nam, S-W., Kong, J-S. (2009). Quantitative risk evaluation based on event tree analysis technique: Application to the design of shield TBM. Tunnelling and Underground Space Technology, 24, 3, 269-277.

ILO-OSH (2001). Guidelines on occupational safety and health management systems. ISBN: 9221116344.

ISO/IEC Guide 51 (1999). Safety Aspects – Guidelines for Their Inclusion in Standards. ISO/IEC (2nd ed.), Geneva.

ISO/IEC 31000. (2009). Risk management-Principles and guideline, ISBN 0 7337 9289 8.

ISO/IEC Guide 73:2009 (2009). Risk management-Vocabulary.

Isograph (2008). Fault Tree+ for windows: Fault Tree Analysis-Event Tree Analysis-Markov Analysis. Isograph Limited FaultTree+V11.2 document, Version 11.2, p.1-325.

Johansson, Ö., Wanvik, P.O., Elvik, R. (2009). A new method for assessing the risk of accident associated with darkness. Accident Analysis and Prevention, doi:10.1016/j.aap.2009.04.003.

Kontogiannis, T., Leopoulos, V., Marmaras, N. (2000). A comparison of accident analysis techniques for safety-critical man-machine systems. International Journal of Industrial Ergonomics, 25, 327-347.

Lim, H.J., Zhang, X. (2009). Semi-parametric additive risk models: Application to injury duration study. Accident Analysis and Prevention 41, 211–216.

Limnios, N. (2007). Fault Trees. ISTE Ltd, UK, ISBN 13: 978-1-905209-30-9.

Marhavilas P.K. (2009a). Health and Safety in the Work–Handling of the Occupational Danger. Tziolas Edition, ISBN 978-960-418-171-1, pages 289.

Marhavilas P.K. (2009b). Risk Estimation in the Greek Constructions' Worksites by using a Quantitative Assessment Technique and Statistical Information of Occupational Accidents. Journal of Engineering Science and Technology Research, Vol. 2, Issue 1, p.p. 51-55, ISSN 1791-2377.

Marhavilas, P.K., Koulouriotis, D.E. (2007). Risk Estimation in the Constructions' Worksites by using a Quantitative Assessment Technique and Statistical Information of Accidents. Technika Chronika Sci. J.TCG, Scientific Journal of Technical Chamber of Greece, Vol I, Issue 1-2, p. 47-60, ISSN 1106-4935.

Marhavilas P.K., Koulouriotis, D.E. (2008). A risk estimation methodological framework using quantitative assessment techniques and real accidents' data: application in an aluminum extrusion industry. Journal of Loss Prevention in the Process Industries, Elsevier, doi:10.1016/j.jlp.2008.04.009, vol. 21, issue 6, p.p. 596-603.

Marhavilas P.K. and D.E. Koulouriotis (2011). Developing a new alternative risk assessment framework in the work sites by including a stochastic and a deterministic process: a case study for the Greek Public Electric Power Provider. Article in press, Safety Science, Elsevier, doi:10.1016/j.ssci.2011.10.0006.

Marhavilas P.K. and D.E. Koulouriotis (2012). A combined usage of stochastic and quantitative risk assessment methods in the worksites: Application on an electric power provider. Reliability Engineering and System Safety, Elsevier, 97 (2012), pp.36-46, doi: 10.1016/j.ress.2011.09.006.

Marhavilas P.K., D.E. Koulouriotis and V. Gemeni (2011a). Risk Analysis and Assessment Methodologies in the Work Sites: On a Review, Classification and Comparative Study of the Scientific Literature of the Period 2000-2009. Journal of Loss Prevention in the Process Industries, Elsevier, DOI: 10.1016/j.jlp.2011.03.004, vol 24, issue 5, pp. 477-523.

Marhavilas, P.K., D.E. Koulouriotis and C. Mitrakas (2011b). On the development of a new hybrid risk assessment process using occupational accidents' data: Application on the Greek Public Electric Power Provider. Journal of Loss Prevention in the Process Industries, Elsevier, DOI 10.1016/j.jlp.2011.05.010, vol 24, issue 5, pp. 671-687.

OHSAS 18002:2008 (2008). Occupational health and safety management systems. Guidelines for the implementation of OHSAS 18001:2007. ISBN: 978 0 580 61674 7, p.88.

Reniers, G.L.L., Dullaert, W., Ale, B.J.M., Soudan, K. (2005a). Developing an external domino prevention framework: Hazwim. Journal of Loss Prevention in the Process Industries, 18, 127-138.

Reniers, G.L.L., Dullaert, W., Ale, B.J.M., Soudan, K. (2005b). The use of current risk analysis tools evaluated towards preventing external domino accidents. Journal of Loss Prevention in the Process Industries, 18, 119-126.

Salvi, Ol. and Gaston, D. (2004). Risk assessment and risk decision-making process related to hazardous installation in France. Journal of Risk Research, 7: 6, 599-608, DOI: 10.1080/1366987042000192192.

Suddle, Sh. (2009). The weighted risk analysis. Safety Science, 47, 668-679.

van Duijne, Fr. H., van Aken, D. and Schouten, E.G. (2008). Considerations in developing complete and quantified methods for risk assessment. Safety Science, 46(2), 245-254.

Vesely, W.E., Goldberg, F.F., Roberts, N.H., Haasl, D.F. (1981). Fault Tree Handbook. US Nuclear Regulatory Commission, Washington, DC.

Yuhua, D., and Datao, Y. (2005). Estimation of failure probability of oil and gas transmission pipelines by fuzzy fault tree analysis. Journal of Loss Prevention in the Process Industries, 18, 83–88.

Zheng X., Liu M. (2009). An overview of accident forecasting methodologies. Journal of Loss Prevention in the Process Industries, 22, 4, 484-491.

4

Health Technology Assessment: An Essential Approach to Guide Clinical Governance Choices on Risk Management

Giovanni Improta[1], Antonio Fratini[2] and Maria Triassi[1]

[1]*Dept. of Preventive Medical Sciences, University of Naples "Federico II", Naples,*
[2]*Dept. of Biomedical, Electronic and Telecommunication Engineering*
University of Naples "Federico II", Naples,
Italy

1. Introduction

Risk management in this chapter is defined as the process of identifying, through the study of all possible sources of errors and problems, the required preventive and corrective actions to reduce risk and whose consequences that compromises the capacity of an organization to reach its own objectives (Del Vecchio and Cosmi, 2003). It has been extensively used in economics, engineering and recently has also been adopted in the fields of public and private health (Gorrod, 2004; Alexander and Sheedy, 2005).

Safety, however, is becoming an imposed target for any health care system so that recent provisions stimulate the application of risk management methodologies (i.e. health risk management) also in clinical environments (Sanfilippo, 2001; Carroll, 2009).

Health Risk Management (HRM) aims to improve the quality of health care, ensure safety and security for patients and sanitary operators encompassing the comprehension of risks associated to the introduction or use of a technology in a clinical environment.

New technologies are not simple objects/products, but like a social practice built within actions and relationships, they are strictly connected to the business setup and act as a basic part of the organization design especially in health care industry.

The knowledge of the risks associated with different technologies (risk assessment) is of extreme importance in the definition of programs and initiatives, at various levels of health care governance (public health authorities, regions government, Ministry of Health), to reduce the incidence of errors and failures.

Healthcare distinguishes itself from other industries in that patient's safety represents a quality dimension of greatest importance.

However, healthcare systems are affected by risks of different nature: risks associated to the personnel professionalism or to the environment appropriateness, risks related to specific equipment use (e.g. magnetic resonance or X-ray), risks related to therapeutic or diagnostic pathways but also risks related to the social, ethical or economic impact to the use of technologies and methodologies in healthcare facilities. Each of these aspects have their own specific aspects of investigation making HRM a complex process.

Healthcare is moving towards increased assistance needs with limited resources, both in economics terms, in personnel or space terms, leading to the usage of specific analyses for the acquisition, evaluation and assessment of medical technologies.

The systematic evaluation of properties, effects or other impacts of a medical (or health) technology with a broad multidisciplinary approach is named Health Technology Assessment (HTA).

The main purpose of HTA is to assist policymaking for technology in health care to achieve the most advantageous resource allocation, evaluating the efficacy and the efficiency of the diagnostic and therapeutic pathways as well as related risks and organizational models.

HTA consists in identifying an analytical methodology that allows the optimization of the product adoption/evaluation process, through a careful study of the effective needs of the users, of the available alternative technologies and the relative operational implications on the setup. This type of evaluation requires an interdisciplinary approach of "policy analysis", studying the aspects of safety, cost, benefits, effectiveness, and include critical evaluations of the actual measures and improving the quality of life.[1] HTA methodology implies to recognize the actual healthcare needs and evaluate how technologies may answer to those needs while considering the overall implication of their use including the associated risks. It may address the direct and intended consequences of technologies as well as their indirect and unintended consequences.

HTA practices have become widespread and are increasingly present in health systems, so that more and more healthcare facilities monitor the global impact of their medical technologies.

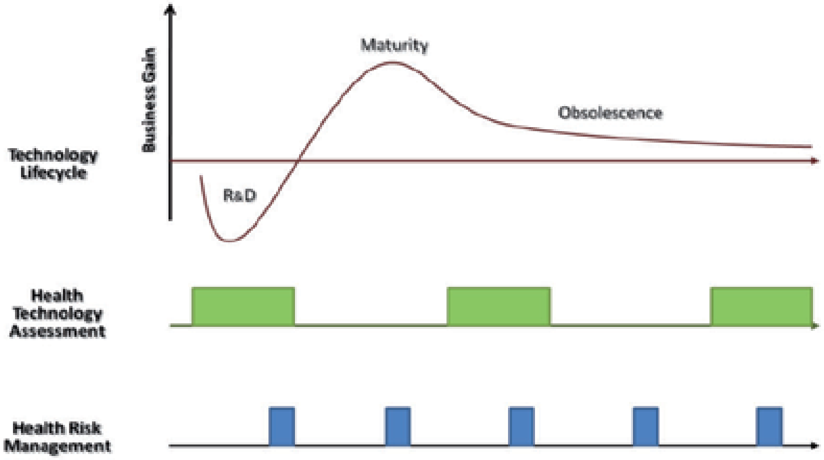


Fig. 1. A link among Technology Lifecycle, Health Technology Assessment and Health Risk Management (derived from Derrico *et al*, 1999)

[1] Canadian Coordinating Office for Heath Technology Assessment 1995

HTA process may be carried out at different technologies lifecycle phases (see Figure 1) to obtain a "dynamic" overview of their potential, usage and intended or unintended consequences.

Since HRM may take advantage from a complete technology impact overview that includes a comprehensive identification of its associated risks, it becomes important to discover HTA methodologies and their application for the evaluation of medical technologies.

The authors aimed to discuss HTA systematic approach and its advantages in assuring correct risk estimation and global patient safety, which is one of the objectives each health organization is aimed to.

This chapter presents an example on a possible design and implementation of a HTA protocol for the classification of hospitals or health facilities equipment, realized by combining the classic HTA concepts with hierarchic clustering techniques in a multidisciplinary analysis of requirements, cost, impact of logistics, technology associated risks.

The rest of the chapter is organized as follows:

- The next section (Section 2) presents a brief review on health technology assessment origins and some fundamental concepts;
- Section 3 – Actual Research – describes the methods followed for the combination of HTA approaches with the hierarchic clustering technique;
- Section 4 (Case study) presents the development of the specific classification protocol and its preliminary application to medical technologies;
- Section 5 (Discussion) reports practical observations and remarks;
- Section 6 (Future work) describes some future and desirable model developments;

2. Origins of health technology assessment

During past decades, health care systems of industrialized countries have focussed on the problem of assuring health services to all citizens while reducing the allocation of economic resources (Fleurette and Banta, 2000; France, 2000; Granados *et al*, 2000; Perlett and Busse, 2000; Jorgensen and Hvenegaard, 2000).

To achieve both the subsistence of the essential health services and the reduction of sanitary costs, almost every state engaged in policies aimed at rationalizing the use of resources by acting on the efficiency of organizations in strengthening service delivery as well as introducing elements of competition between producers or prioritization of health care services to ensure to citizens through public funding. (Sackett, 1980; Banta, 1993, 2000; Battista and Hodge, 1989)

The need to evaluate the effectiveness of different diagnostic and therapeutic protocols and technologies compared to the suffering population and, at the same time, the need to a complete knowledge of the service delivery costs originated a multi-disciplinary research area called "Health Technology Assessment" (Blades, 1986; Birch and Donaldson, 1987; Battista and Hodge, 1999).

Early studies of technology evaluation were performed in the mid-1960s: the Committee on Science and Astronautics of the House of Representatives of the United States Congress, showed the need for new methods of analysis that can clarify the *economic and social impact due to the introduction and development of new technologies*.

In 1972, United States Congress creates the Office of Technology Assessment (OTA) with a pubic law (92-484) with the task of developing and deploying technology assessment, authoritative analysis of complex scientific and technical issues, and demonstrates its usefulness to politicians.

Technical information needed by policymakers is frequently not available, or not in the right form. A policymaker cannot judge the merits or consequences of a technological program within a strictly technical context. He has to consider social, economic, and legal implications of any course of action[2].

In Europe the importance of the assessment was accepted about a decade later than in the U.S., when the World Health Organization (WHO), within the program "Health for Hall"; in a first phase, the response by governments has focussed on introducing policies aimed at controlling the spread of technologies in logic of cost containment.

This phase did not produce significant results in terms of technology assessment, it however allowed the introduction of methods of economic evaluations, and in particular the concept of cost-effectiveness in health care (Weisbrod, 1961; Cochrane, 1972; Bush *et al*, 1973; Sorkin, 1975; Drummond, 1994).

The change in technology and resource allocation assessment was a great enhancement for the development of another key asset: the Evidence-Based Medicine, whose objective is to find a relationship between the empirical evidence and current clinical practice in order to improve the quality and effectiveness of care at the level of individual patients (Sackett *et al.*, 1996).

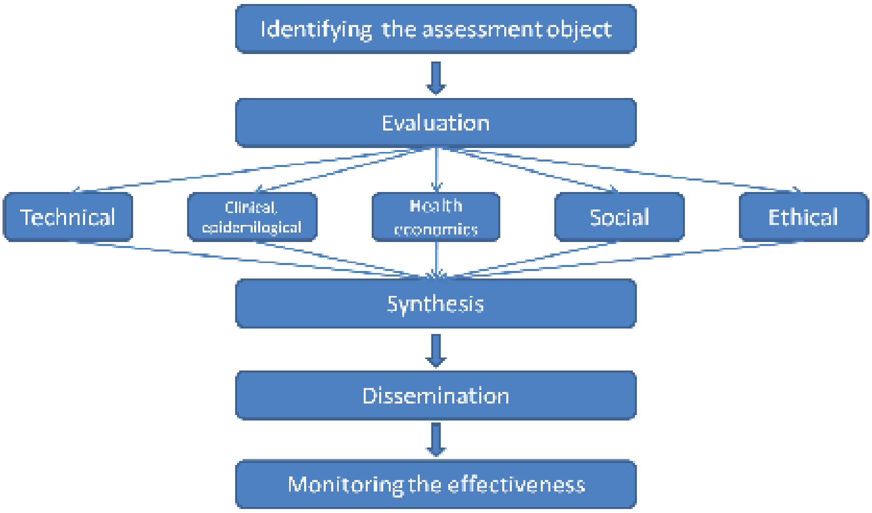


Fig. 2. HTA process steps (derived from Panelius *et al.* 1988, FinOHTA 1997)

[2] Emilio Daddario, U.S. Congress, House of Representatives 1967

During the last decades then, national and international societies contributes to consolidate and to refine HTA approaches and methodologies.

HTA has been recognized as a process that involves a number of different actions rather than a single research. These may include identification of technologies that require an assessment, collection and data analysis, the synthesis of existing information and the distribution of results and subsequent advice and recommendations.

The fundamental steps of an HTA process can be summarized, as well as in a technology assessment, in some main steps that include:

- the *identification* of the assessment object/topic in order to clarify:
  - the problem addressed by the technology;
  - real clinical needs (needs assessment);
  - requirements or constraints the technology under investigation has to fit.
- the *evaluation* of the technology that for HTAs has to include:
  - the collection of key data in terms of general impact: technical, clinical, social, ethical as well as economical; it involves the comparison of different technologies according to criteria of quality, evaluating the clinical efficacy (benefits), safety, clinical outcomes, costs of the entire life cycle of technology;
  - the analysis of all the collected data and the technology rating;

This process, and its multidisciplinary evaluations, characterize (or it should do) all HTA processes.

- the *synthesis* phase includes:
  - the consolidation and synthesis of all the analysis in order to give a synthetic overview of the assessment results;
  - the production of recommendation on the applicability and adoption of the technology;

Finally, the *dissemination* of information phase comes from the evaluation and implementation of the decision maker.

However, as in every dynamic process, *monitoring the effectiveness* of the assessment conclusion helps in refining methodologies and in assuring the correctness of the decision adopted.

Crucial in the entire HTA process is the identification of professionals selected for the study.

HTA analyses uses different methods, depending on the purpose of the assessment; some of them used economic approach: cost-analyses (Drummond *et al.*, 1987), some a managerial approach SWOT (Strengths, Weaknesses, Opportunities and Threats) analyses, logical framework (Armstrong, 1982; Kahveci *et al.*, 2006) and some other a more pluralistic approach: Multiple Attribute Decision Making (MADM) analyses, hierarchical clustering (Agrawal *et al.*, 1991). Each of these methods have their own particular or partial judgment aspects. In the past years scientific community stimulated the debate on other possible approaches and tool in HTA process and hierarchic approach as used in decision support theory have been recently proposed. One of the most interesting and used hierarchic model is the Analytic Hierarchy Process (AHP) that began with (Saaty 1980) and is often referred as his method. In general, hierarchic processes are interesting methodologies since their ability to grade choices in meeting conflicting objectives and its wide variety of decision problems applications.

However, HTA is often adopted in process of technology acquisition and hierarchic approach has been proposed in some cases (Sloane *et al.*, 2003); in our opinion however this approach can also be conveniently used as HTA in case of evaluation on the state of technologies in an hospital or in general any health facilities.

3. Actual research

Every health facility (HF) manages a great number of medical technologies; however, in their "hospital" life cycle, technologies will be replaced by recent ones, especially those with lower efficiency and possible risks related to their age and use.

In general, HF managers demand regular surveys regarding status of the technologies in the same structures; hence, HTA procedure should also be applied to assess the state of capital equipment in order to facilitate further management decisions on technologies investments or to consider equipment immediate disposal, planned disposal or relocation (assigning it to other departments).

The authors intended to present an example of assessment protocol, realized by combining hierarchical clustering techniques with HTA multi-criteria approach, aimed to help clinical engineers and clinicians to concur in manager decisions on the capital equipment of a hospital or a health facility, in particular, the solution proposed was intended to allow an easy equipment ranking in terms of disposal/relocation planning.

The goal of the assessment was to highlight the *state* of single device/technology, classifying it in three main categories:

- **Out of service**: it is not more possible to use the equipment. The state is related to a severe damage, uneconomic repairing action or to a non-safe condition, mainly following a corrective technical maintenance;
- **Planned out of service:** It refers to a condition in which clinicians can safely continue to use the equipment despite having to plan to replace it; the state is related mainly to a reduced utilization or non-utilization, non-compliance to newer technical standards, reduced functional efficiency, uneconomic maintenance or technical obsolescence. In this case the possible alternatives are:
    - *Relocation*: the equipment could respond more effectively to the need of other departments within HF; it complies to technical standard and it is still properly functioning.
    - *Disposal:* the equipment does not respond effectively to the needs of the HF, it could be however conveniently employed by other institutions or health structures; the alternatives in this case are:
        - *Trading:* new technologies vendor discounts or equipment selling.
        - *Donation:* the equipment can be used in other emerging countries, in emergency environments.
- **Full compliance** with the evaluation criteria: The equipment is used properly and efficiently.

In order to achieve the desired classification, HTA and clustering technique methodologies have been combined; the advantages obtained with the combination proposed and the combination methods are completely discussed and clarified in the next paragraphs.

3.1 Adopted methodology

The protocol has been structured following a hierarchical assessment approach, similar to AHP (Saaty, 1980, Saaty, 1982, 1990), based on the definition of the goal to achieve, the criteria and evaluation parameters and their relative and global incidence in the overall decision.

A hierarchical breakdown of the problem in N different criteria (or cluster), which groups properties and attributes of alternatives, helps in a better evaluation of the problem itself.

For each cluster are then recognized properties or attributes (or cluster elements) in a variable number. It is worth mentioning that, in health environments, these properties and attributes are not always directly and objectively measurable (i.e. revenues *versus* degree of patient technology acceptance) and, in order to obtain a comprehensive and concise assessment reducing *subjective* bias, these are aggregated together in clusters.

The non-objective measurable parameters/attributes can also be quantified and then made comparable by using expert opinions expressed in linguistic variables and converted into numerical values (usually using the ordinal scale used in AHP and a pairwise comparison procedure with the aim of producing a square matrix, whose element a_{ij} indicates the relative importance of the element with respect to criteria A_J).

Synthetic assessment of the degree of importance of the single Aj with respect to the others (weights) are calculated by normalizing the global importance of individual factors, i.e. the sum of each element of a row, with respect to the sum; it keeps unchanged the relationships between the factors and makes the sum of all weight obtained equal to 1, which is mathematically convenient in weighted sums.

Assuming gather experts evaluations so that $a_{ik} = a_{ij}*a_{jk}$ (i. e. assuming to know n-1 matrix elements and obtaining the remaining matrix elements from the properties of consistency and reciprocity) is not necessary to evaluate the AHP technique Consistency Index (CI) as it has hypothesised a perfect consistency in judgments (CI = 0).

Weights obtained are aggregated together with the hierarchical Saaty's composition principle, which allows a priority listing of alternatives to the goal.

In our case, the final equipment classification is obtained by *scoring* the equipment based on the evaluated importance of the criteria and their properties which helps in correctly combining the specific characteristics/condition of the equipment under investigation.

Based on this principle, the overall score of the generic alternative A with respect to the goal may be expressed as:

$$C = \frac{\sum_{i=1}^{k} P_i \cdot V_i}{\sum_{i=1}^{k} P_i} \tag{1}$$

where:

k is the cluster numbers
P_i is the weight of cluster
V_i is the total score of equipment with respect to i-cluster
C is the total score of equipment

$$V_j = \frac{\sum_{j=1}^{n} p_j \cdot v_j}{\sum_{j=1}^{n} p_j} \tag{2}$$

where:

n is the element numbers
p_j is the weight of element j with respect to cluster
v_j is the score of element j
V_j is the total score of equipment for that criteria

Finally, the process ends with a classification of the equipment based on its specific *score;* in particular, since we hypothesized four different alternatives, classification is achieved choosing three different thresholds and comparing the obtained equipment score with those values. In case of partial evaluation (that is the evaluation based only on some cluster) the sum will include only the aspect under investigation.

Every assessment strategy however, although theoretically correct, has to be tested in real environment to discover its acceptability and practical applicability; thus, the described methodology has then been examined in an HTA case study of an hospital.

4. Case study

This case study describes the application and refinement of the adopted methodology in a hospital HTA focussed on technology evaluation of hospital equipment: our case consider a preliminary application on an ultrasound device.

The hospital has around 900 beds and serves a population of more than a million people, it is a national centre of excellence and has agreement with different emerging countries for the exchange of personnel and technology.

Among its structures, the hospital includes a biotechnology research centre born in 2000 with a consistent acquisition of specialized personnel already present in a research division founded in 1986 in the same hospital.

The centre is specialized in advanced research on malignant hyperthermia, liver cells mutations, clinical trials on animals, it is also a biomedical data repository.

The centre has its own imaging department that include different technologies, one of which was evaluated with the developed model; this case represented an ideal opportunity to test the acceptance of method by centre employees and the model results.

As in each HTA process (see Figure 2), after a first identification of the assessment object (ultrasound device), a panel of experts (*HTA team*) has to be defined to proceed in concretizing the evaluation strategy and to reveal the relative weight of the different evaluation fields as described above. Still, the process of team choice is of extreme importance for the entire assessment and particular attention have been paid to the definition of the team members.

4.1 HTA team

Different professional were involved in the HTA protocol development; in our opinion HTA team may benefit from including professionals of structures that are frequently involved in the process of acquisition or evaluation and use of technologies. The team was then composed of five professionals: the hospital general manager, the sanitary manager, a clinical engineer, the superintendent and the risk protection and prevention manager.

The choice of including individual used to do assessment processes facilitates the identification and scoring of criteria and their constituents.

Additionally, since periodic controls of hospital technology are mandatory, clinical engineering services hold information related to every equipment existing in the hospital and the collected data is often the result of monitoring procedure and checklist filling, by scoring[3] each voice of these checklist it will immediately score the related cluster attribute.

Once the team has been formed and the information sources

identified, the application of the described methodology for the development of the assessment protocol took place as reported below.

4.2 Protocol development

The multidimensional analysis, typical of a health technology assessment process as mentioned above, determined the identification of five clusters related to key areas of the assessment: S-Safety, L-Legal, O-Organizational, E-Economic, T-technological their constituents.

The protocol was intended for the classification of a wide variety of equipment hence the definition of elements and thresholds was derived from previous experience of different professionals, as discussed in the next paragraphs. The figure below reports some of the selected variables.

[3] The process of assignment of a value to a monitoring checklist field

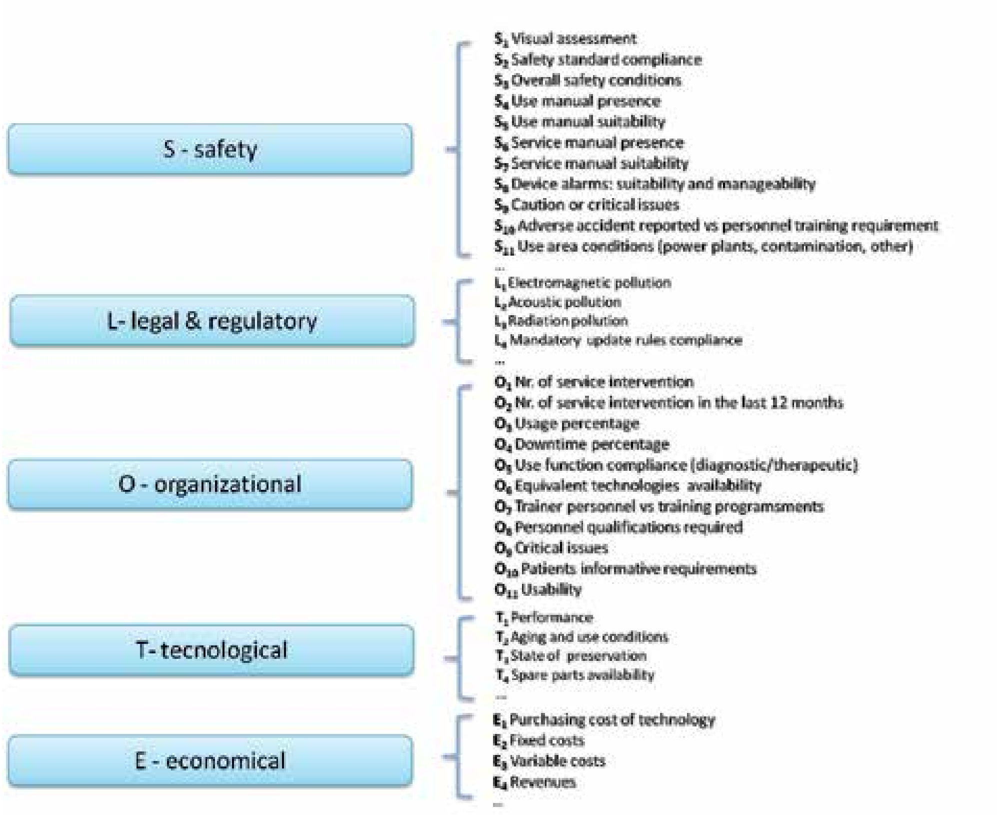

S - safety
S_1 Visual assessment
S_2 Safety standard compliance
S_3 Overall safety conditions
S_4 Use manual presence
S_5 Use manual suitability
S_6 Service manual presence
S_7 Service manual suitability
S_8 Device alarms: suitability and manageability
S_9 Caution or critical issues
S_{10} Adverse accident reported vs personnel training requirement
S_{11} Use area conditions (power plants, contamination, other)
...
L- legal & regulatory
L_1 Electromagnetic pollution
L_2 Acoustic pollution
L_3 Radiation pollution
L_4 Mandatory update rules compliance
...
O - organizational
O_1 Nr. of service intervention
O_2 Nr. of service intervention in the last 12 months
O_3 Usage percentage
O_4 Downtime percentage
O_5 Use function compliance (diagnostic/therapeutic)
O_6 Equivalent technologies availability
O_7 Trainer personnel vs training programsments
O_8 Personnel qualifications required
O_9 Critical issues
O_{10} Patients informative requirements
O_{11} Usability
T- tecnological
T_1 Performance
T_2 Aging and use conditions
T_3 State of preservation
T_4 Spare parts availability
...
E - economical
E_1 Purchasing cost of technology
E_2 Fixed costs
E_3 Variable costs
E_4 Revenues
...

Fig. 3. Evaluation clusters elements chosen for protocol implementation

HTA protocol developed is based on a hierarchical classification of evaluation criteria (S, L, O, E, T) and their constituents (s_i, l_i, o_i, e_i, t_i); in particular, each criteria is compared against the other through a pair wise comparison and weighted in order to obtain a numerical equivalent of its relevance with respect to the totality of the clusters.

However, for this case study, the protocol has been modelled with different pathways since safety (S) and the legal-ethical-social (L) aspects, may in practice individually determine an *Out of order* condition.

A primary classification on legal (L) and safety (S) aspects is realized, and if equipment get an acceptable score the classification goes further with the organizational, economic and technological scoring. A first threshold was then identified for that first classification, meaning the minimum adequacy of the equipment with respect to the compliance to regulatory standards and relative related risks.

In case of non-sufficient evaluation the process stops with an *Out of order* classification outcome, saving time not proceeding through the overall assessment.

However, technology classification is strictly dependant on the weights and ratings assigned to cluster and cluster elements; this is always related to the purpose of the assessment and the HTA team choices that for the specific case study have been further explained.

4.3 Weights and ratings

Cluster weights assignment has been obtained using pair-wise comparison; preferences were expressed with a scale from 1 to 6 as reported in table.

Preference	*Score*
Parity	*1*
Miminal	*2*
Little	*3*
Medium	*4*
Maximal	*5*
Absolute	*6*

Table 1. Preference scored, higher scores mean higher preference

Pair-wise comparison highlighted the substantial equivalence between Legal and Safety criteria (see table 2).

	Legal and social aspects (T)	**Weight %**
Safety (S)	S L 1	50 %

Table 2. Pair wise comparison of safety and legal criteria

A successive pair-wise comparison was performed for organizational (O), economic (E) and technological (T) approach resulting in the percentages reported below.

	Technical Aspects (T)	**Economical Aspects (E)**	**Sum of Scores**	**Weight %**
Organizational Aspects (O)	O 2	O 2	O = 4	PO= 66.67%
	Technical Aspects (T)	TE 1	T = 1	PT= 16.665%
			E = 1	PE= 16.665%

Table 3. Weighting of Technical, Economical and Organizational criteria

On the other hand, cluster elements were compared using a ratings approach instead of pair wise comparison since their large number and their relative incidence on all the identified clusters.

The rating approach allowed us to define individual weight of elements for each criterion; the table depicts an example of weighting procedure for an element with respect to all the entire criteria set.

S_2: Safety standard compliance	**Sum of Scores**	**Weight %**				
	L	O	T	E		
S	S 5	S 5	S 4	S 5	S = 19	$P_{S\mid S2}$ = 77.94 %
	L	L 3	L 3	L 3	L = 9	$P_{L\mid S2}$ = 13.23 %
		O	OT 1	OE 1	O = 3	$P_{O\mid S2}$ = 2.94 %
			T	T 2	T = 3	$P_{E\mid S2}$ = 1.47 %
					E = 1	$P_{T\mid S2}$ = 4.42 %

Table 4. Individual ranking of S_2 parameter

Finally, as will be detailed below, clusters elements rating was achieved by utilizing the data collected by clinical engineering services; each cluster elements (equipment characteristics/conditions) was therefore scored by means of continuous monitoring and reference checklist analysis.

A value v_j for each element of the cluster on a scale from 1 to 10, was assigned where 10 represents the best condition.

Once values and weights are assigned the technology assessment can be finalized: the process proceeds by valuing each technology aspects thorough the pathways described in the next paragraph, the final score is computed and compared with determined thresholds in order to *classify* technology.

4.4 Classification

As already stated, the protocol includes different pathways; this solution was adopted because legal and safety criteria individually assess the compliance of the device to national and internationals rules and the risks associated to the use of equipment itself.

This first protocol step allows us to classify equipment that has to be immediately declared *Out of Service*; scores for each of the step of the protocol were computed as weighted averages based on the obtained cluster estimations and weights previously assigned to each cluster.

A partial evaluation is computed based on the formula reported below to obtain the assessment:

$$C_{SL} = \frac{P_S * V_S + P_L * V_L}{P_S + P_L} \tag{3}$$

A first threshold of sufficiency (S_0=6) was identified and only equipment that passes this score proceed to the rest of classification, otherwise it is classified out of service (see Figure 4). Equipment that proceeds with the rest of assessment process is ranked based on the Economical, Organizational and Technological criteria as shown below;

$$C_{OTE} = \frac{P_O * V_O + P_T * V_T + P_E * V_E}{P_O + P_T + P_E} \tag{4}$$

Based on the obtained score the equipment is classified.

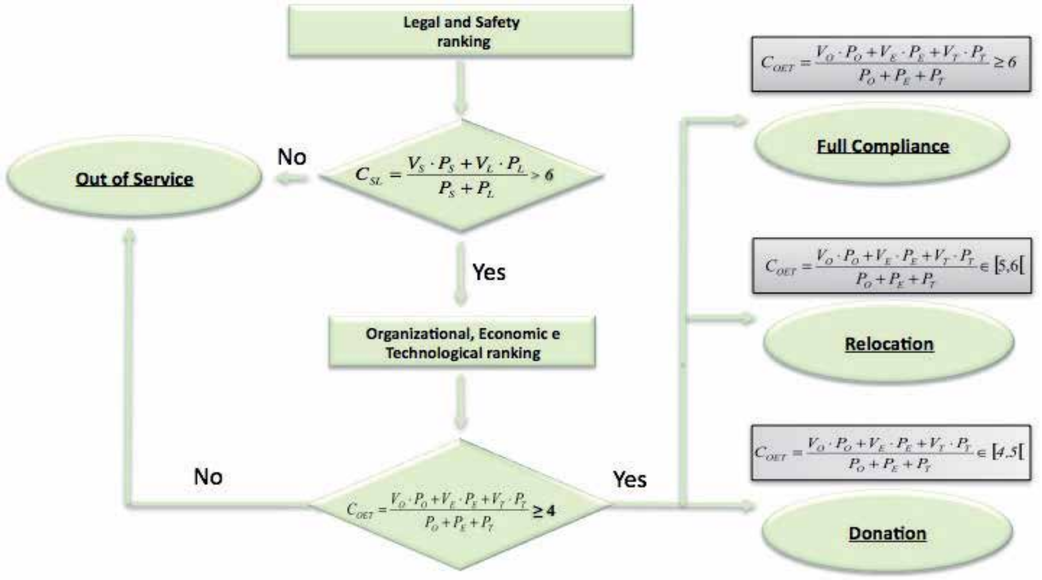

Legal and Safety ranking
No
Out of Service
$C_{SL} = \frac{V_S \cdot P_S + V_L \cdot P_L}{P_S + P_L} > 6$
Yes
Organizational, Economic e Technological ranking
No
$C_{OET} = \frac{V_O \cdot P_O + V_E \cdot P_E + V_T \cdot P_T}{P_O + P_E + P_T} \geq 4$
Yes
$C_{OET} = \frac{V_O \cdot P_O + V_E \cdot P_E + V_T \cdot P_T}{P_O + P_E + P_T} \geq 6$
Full Compliance
$C_{OET} = \frac{V_O \cdot P_O + V_E \cdot P_E + V_T \cdot P_T}{P_O + P_E + P_T} \in [5,6[$
Relocation
$C_{OET} = \frac{V_O \cdot P_O + V_E \cdot P_E + V_T \cdot P_T}{P_O + P_E + P_T} \in [4,5[$
Donation

Fig. 4. Evaluation protocol pathways and classification

For this last classification, three thresholds were established S1 (= 4), S2 (= 5) and S3 (= 6), which led to the identification of 4 classifications:

- *Class 1* with C < S1 that identifies equipment that still be declared "*out of service*";
- *Class 2* with S1 ≤ C < S2 that identifies equipment that does not respond effectively to the needs of that facility (e.g. non complete standards compliance), but could be usefully employed elsewhere (e.g. other country), and then *donated* to other institutions or health care facilities;
- *Class 3* with S2 ≤ C < S3, which identifies the equipment that could be better employed inside the hospital (i.e. based on the percentage/frequency of its usage) and then *relocated* ;
- the *Class 4* with C ≥ S3 identifies the equipment that, at the time of the assessment, fully comply with the adopted constraints and do not need intervention.

However, threshold were chosen on the HTA team practical experience, they have to be confirmed by feedback results obtained by applying protocol in preliminary studies as for our case study as presented in the next section.

4.5 Preliminary application of protocol

The ultrasound is placed at the Biotechnology centre of the hospital. It was purchased in 1998 at a cost of € 75,000, this price refers to the unit acquisition cost and includes two linear probes and a convex probe; moreover, the vendor signed of a full risk contract type on this equipment.[4]

The ultrasound equipment has been classified through the hierarchy protocol proposed, monitoring checklist has been examined and the assessment process started.

As explained in the previous section the first classification was made based on Safety and Legal criteria.

As example, we reported a complete score of ultrasound equipment for the Safety criteria; in case of non applicable evaluation o some of the listed variables the parameter itself is non-considered in the total criteria scoring process. The condition of non-applicability of parameter evaluation has been identified by using hash (#) symbol (see table below).

	Evaluated Parameter	Score
S1	Visual assessment	8
S2	Safety standard compliance	10
S3	Overall safety conditions	10
S4	User manual availability	10
S5	User manual suitability	10
S6	Service manual availability	1
S7	Service manual suitability	#
S8	Device alarms: suitability and manageability	#
S9	Equipment caution or critical issues	7
S10	Update training for the personnel than adverse events involving the equipment	1
S11	State of use compared to the rooms and facilities	10

Table 5. Safety criteria elements evaluation

Safety an Legal properties analysis achieved the first step of the examination protocol; equation 3 was computed with previously evaluated weights (see Table 2) and scores obtained by monitoring checklist recognition.

$$C_{SL} = \frac{P_S \cdot V_S + P_L \cdot V_L}{P_S + P_L} = \frac{7.415 \cdot 50 + 9.238 \cdot 50}{100} = 8.327$$

The value obtained is above S_0 threshold therefore, the assessment process can continue to the evaluation of Organizational, Technological and Economic aspects; in Table 6 are reported the scores of ultrasound with respect to each criteria:

[4]The contract covers preventive and corrective maintenance program for equipment full technical support including workforce and spare parts.

	Criteria	Score
S	Safety	7.415
L	Legal	9.238
O	Organizational	5.204
E	Economic	6.075
T	Technological	5.935

Table 6. Equipment scoring results

As done for the first step and using the previously computed weights of the assessment criteria (see Table 3), ultrasound score was:

$$C_{SL} = \frac{P_o \cdot V_o + P_E \cdot V_E + P_T \cdot V_T}{P_O + P_E + P_T} = \frac{5.204 \cdot 66.67 + 6.075 \cdot 16.665 + 5.935 \cdot 16.665}{100} = 5.471$$

On the base of the considered threshold, the proposed model classified the equipment as planned out of service in particular *Relocation.*

The equipment complies with safety and legal standards but it usage percentage and its efficacy for that centre is not sufficient to maintain in use the device; however, since the overall state of the ultrasound achieves a good usability, the model correctly suggested a relocation of the equipment to another hospital department.

5. Discussion

Technology assessment cannot replace the clinical governance decision makers as these topics are often related to variables dependent on their sensitivity; however, HTA certainly improve management processes through a more effective use of information and knowledge available.

HTA leads to a wider risk analysis and a better health needs assessment, it makes possible an extensive knowledge of the technology characteristics, its effects on individuals health, its economic and/or organizational impact and may allow:

- improved selection processes: for the selection of technologies to adopt through an explicit comparison between the "*needs*" (health needs, resources available);
- efficient management of procurement processes, since a better understanding of the overall characteristics of the technology can enhance negotiation skills in dealing with suppliers;
- the preparation of all the organizational, professional and financial resources necessary for effective and efficient use of technology in order to increase the level of performance provided.

Generally, HTAs are mainly related to technology or equipment purchase; results are presented in the form of reports or indicators to help decision makers in their conclusions.

However, in our knowledge few of them have been dedicated to a classification of the overall state of hospital equipments especially with relocation/donation purposes.

The proposed methodology, based on the requirements and constraints often suggested by decision makers themselves, provides an indicator (a numerical value) through which the equipment may be classified; using the algorithm all the information associated with the assessment are synthesized to allow managers in easily getting an overall picture of capital equipment state and usage implications in the hospital facility.

Modern hospitals own a huge amount of equipments, in some case these technologies an underused or not used properly, in some other case equipments are used still if they do not properly comply with safety standard or simply can not be used because not more compliant with emerging ones. Still, it is not common to find HFs relocating technologies, after a HTA, in order to achieve better performance in a different hospital area.

In all of this cases, by appropriately indicate threshold, the proposed model may evaluate if the technology under investigation meets the requirements/provisions of use, and how much convenient may result a specific intervention (disposal/relocation/replacement).

A considerable importance in the application of this protocol has to be assigned to the design of the evaluation checklist and the p_x weight attribution in pair wise comparisons, appropriate design choices allow for reduction of operator-dependent errors and the correct evaluation of technologies.

In our opinion, the choice of including professionals from the clinical engineering service of the hospital may became a common practice; it gives the chance to score cluster elements in a easy and continuous way. However, clinical engineering services are often outsourced, this leads to a difficult integration with hospital information systems while it also makes difficult equipment monitoring data retrieval.

Monitoring checklists have to be discussed and refinements may be suggested to allow a limited choice to the evaluator, rather than proposing a free answer, reducing the bias of the individual operator with respect to the overall assessment.

The classification system proposed in this paper was built to be easy to use and to allow an immediate interpretation to non-experts. As well as hierarchic methods, the model may led to a broader evaluations (i.e. with a larger number of clusters) or more focused assessments; in general, the use of an automated assessment protocol may be profitably integrated in information systems as a tool for a better management of technologies.

The model has been tested for the evaluation of hospital equipment and gave encouraging feedbacks; it can provide regular analysis on the state of technological equipment of an HF.

A crucial and delicate point was the HTA team building; the evaluations reflect the experience, knowledge and sensibility of individuals involved in the study.

In our opinion, as well as all HTA methodologies the model presented should improve healthcare operators and individual consciousness with respect to the whole healthcare system.

However, the application of any HTA methodology is dependant on the acceptability of users (patients, personnel, managers) and on the integration that the methodology may achieve with HF information systems.

Nevertheless, in this our first experience, the received feedbacks were encouraging and the information level of the structure was fine enough to start thinking a HTA methodology integration.

6. Future work and closure

Efficiency of medical equipment and devices, from which nowadays the totality of medical services depends, has direct and indirect influences on the quality of the offered service in association with diagnostic or therapeutic accuracy, access time to medical services and it influences the safety of personnel that makes use of those facilities.

Evaluation of health technologies must be a continuous activity to be conducted before their introduction but also, and with more consideration, during their entire lifecycle. Organizations with high reliability, where healthcare is included, aim to reach high safety standards by focusing the activity not on avoiding isolated failures, but on the ability of rendering the entire system robust and practical against relative human and operational risks.

This work suggested some new ideas and future development of the proposed model; in particular, more model application are desirable to enhance the protocol efficacy.

In conclusion, in our opinion two main improvements will become widespread in clinical governance:

- evolved monitoring checklist: checklist designed in collaboration with clinical engineers and hospital professionals can contribute to obtain a feasible evaluation of characteristics/condition of capital equipment;
- hospital information systems integration: assessment models may be integrated in HFs' information system to allow continuous
- support decision tool.

In a period in which HF information is increasing, it is reasonable to imagine an evolved information/computer support system able to store data provided by monitoring procedure and to proceed rapidly to technology classification on predetermined rules and thresholds.

Intelligent decision support systems will be able to highlight to HF managers the state and the needs of capital equipment of the entire facility, in a more complete approach to hospital risk management.

7. References

Agrawal, VP, Kohli, V., Gupta, S. *"Computer aided robot selection: A multiple attribute decision making approach"*. 1991, International Journal of Production Research, Vol. 29, pp. 1629-1644

Alexander C. and Sheedy E. "*The Professional Risk Managers Handbook: A Comprehensive Guide to Current Theory and Best Practices"*. 2005, PRMIA Publications.

Armstrong J.S. *"The Value of Formal Planning for Strategic Decisions*". 1982, Strategic Management Journal, issue 3 pp:197–211.

Banta D, Oortwijn W. "*Health Technology Assessment and Health Care in the European Union*". International Journal of Technology Assessment in health Care 2000; 16(2):626-635

Banta H.D. and Luce B. R..: *"Health Care Technology and its Assessment: an international perspective"*. Oxford University Press, 1993, pp 197-203.

Bates D.W. and Gawande A.A.: "*Improving safety with information technology*" New Engl J Med; 2003; 348(25): pp. 2526-34

Battista R. "*Innovation and diffusion of health-related technologies. A conceptual framework"* 1989 Int J Technol Asses Health Care ; 5(2):227-48

Battista R. Hodge M.J " *The Evolving Paradigm of Health Tecnology Assessment : Reflections for the Millennium*", CMAJ ,1999, 160 (10) pp 1464-1467

Birch S, Donaldson C. *"Applications of cost benefit analysis to health care: departures from welfare economic theory."* J Health Economics 1987; 6: 211-25.

Blades CA, Culyer AJ,Wiseman J,Walker A. *"The international bibliography of health economics."* London, Wheatsheaf Books, 1986.

Bush JW, Chen MM, Patrick DL. *Health status index in cost-effectiveness analysis of PKU programme."* In: Berg RL (ed). *"Health status indexes."* Chicago, Hospital Research and Educational Trust, 1973; pp.172-208

Carroll R. *"Risk Management Handbook for Health Care Organizations"* 2009, John Wiley and Sons

Cochrane AL *"Effectiveness and efficacy. Random reflections on health services"* Nuffield provincial hospital trust, 1972 (reprinted in 1989 in association with BMJ), London. Trad it: *"Efficienza ed efficacia. Riflessioni sui servizi sanitari"* Il Pensiero Scientifico Editore, Roma, 1999

Del Vecchio M., Cosmi L. *"Il Risk Management nelle aziende sanitarie"* Mc-Graw Hill, 2003

Drummond M.F., *"Methods for Economic Appraisal of Health technology"*, in "Economic Appraisal of Health Technology in the European Community", ed. Drummond 1987 pp15-48. Oxford: Oxford University Press

Drummond MF. *"Evaluation of health technology: economic issues for health policy and policy issues for economic appraisal."* Social Science and Medicine 1994 ;38(12):1593-1600.

Fleurette F, Banta D. *"Health Technology Assessment in France"*. International Journal of Technology Assessment in health Care 2000; 16(2):400-411.

France G. *"Health Technology Assessment in Italy"*. International Journal of Technology Assessment in health Care 2000; 16(2):459-474.

Goodman CS. TA101. *"Introduction to Health Technology Assessment"*. The Lewin Group, 1998 pp. 106

Gorrod M. *"Risk Management Systems : Technology Trends" 2004,* Basingstoke: Palgrave Macmillan

Granados A, Sampietro Colom L et al. *"Health Technology Assessment in Spain"*. International Journal of Technology Assessment in health Care 2000; 16(2):532-559.

Jorgensen T, Hvenegaard A et al. *"Health Technology Assessment in Denmark"*. International Journal of Technology Assessment in health Care 2000; 16(2):347-381.

Kahveci R. *"Analysis of strengts, weakness, threats and opportunities in development of an HTA program in Turkey"*. 2006, Handb Health Technol Assess Int. 2006; 3: 188.

Kristensen FB, Horder M et al. *"Health Technology Assessment handbook"*. Danish Institute for HTA, 2001.

Perleth M, Busse R. *"Health Technology Assessment in Germany"*. International Journal of Technology Assessment in health Care 2000; 16(2):412- 428.

Saaty T.L., *"The Analytic Hierarchy Process"*, McGraw Hill, New York, 1980.

Saaty T.L., *"How to make a decision: The Analytic Hierarchy Process"*, European Journal of Operational Research, 48(1), 9-26, 1990

Sackett DL. *"Evaluation of health services."* In Last JM (ed). Health and preventive medicine. New York: Appleton-Century Crofts, 1980; 1800-23.

Sackett DL, Rosenberg WMC, Gray JA, Haynes B, Richardson S. *"Evidence based medicine: what it is and what it isn't"* 1996, British Med J, 372(13) pp. 71-

Sanfilippo J.S. *"The risk management handbook for healthcare professionals"*, 2001, Lavoisier

Sloane E.B., Liberatore M.J., Nydick R.L., Luo W., Chung Q.B. *"Using the analytic hierarchy process as a clinical engineering tool to facilitate an iterative, multidisciplinary, microeconomic health technology assessment"*, 2003, Computers & Operations Research, (30) pp. 1447–1465

Sorkin AL. *"Health economics: an introduction. Lexington",* Lexington Books, 1975.

Weisbrod BA. *"Economics of public health"*. Philadelphia, University of Pennsylvania Press, 1961.

Preventing Societal Health Risks Emerging in the Development Of Nanomedicine – What Should Prevail?

Roberte Manigat[1], Florent Allix[2],
Céline Frochot[2] and Jean Claude André[2,3]
[1]Ministère du Travail, de l'Emploi et de la Santé, Paris,
[2]Centre National de Recherche Scientifique LRGP-UPR 3349, Nancy,
[3]Centre National de Recherche Scientifique INSIS, Paris,
France

1. Introduction

The traditional distinction between preventive and curative medicine has probably served, among others, programmatic and educational purposes, but does not look so robust (and valid) when used to back-up choices for innovative technologies or justify managerial economic options. Obviously, the line between both some secondary or tertiary preventive actions and many curative ones is often void, and above all a simple question of optical perspective.

Indeed, in our opinion, prevention and precaution are the underlying humanistic driving forces that have, or should have, always presided to the management of health risks. And this should remain untouched. Is this merely an ideological choice? In any case, the implementation of these concepts is also definitely not detrimental to long term economical considerations, another real asset, nor a limiting factor to progress and innovation.

Nanotechnology refers to the design, characterization, production and application of structures, devices and systems that have novel physical, chemical and biological properties, by controlling shape and size at the nanometer scale[1] (Ebbesen and Jensen, 2006). While this multidisciplinary scientific field is undergoing explosive development, it is also subject to many controversial debates among scientists, but not only[2]. The design

[1] A nanometer (nm) is equal to one billionth of a meter (10^{-9} meter); formerly called millimicron (1/1000 of a micron), the nano scale was denoted by the symbol mμ. As an illustration, the diameter of a red blood cell is roughly 1 000 nanometers (10^{-6} meter).

[2] A 3 days Conference on Risks associated with nanoparticles and nanomaterials was organised by the French National Institute on Research and Security (INRS) in April 2011, to discuss occupational health research issues. A public debate on Nanotechnologies was held in France from October 15th 2009 to February 24th 2010, organized by the National Commission on Public Debate (CNDP) the National official and independent body (full details are available from: http://www.debatpublicnano.org/index.html).

and assembly of sub microscopic devices called nanoparticles, which are 1 to 100 nanometer in diameter (Kateb *et al.*, 2011), are also the object of concern as to the safety of their use, and not only in the general public (Poland *et al.*, 2008; Takagi *et al.*, 2008; Genaidy *et al.*, 2009; Murphy *et al.*, 2011). Potential toxic effects of certain nanoproducts have legitimately conducted either the decision makers or the civil society to mobilise high level expert's investigation not only in France (AFSSET, 2006, 2008; HCSP, 2009) but in most of the OECD member countries (Kaluza *et al.*, 2008; DHHS-CDC-NIOSH, 2009; Ostiguy *et al.*, 2010). In either case, among the multiple fields of application of nanotechnology, we believe that the development of nanomedicine, the application of nanotechnology for the diagnosis and treatment of human disease, should be oriented through the lenses of the above mentioned driving forces, and subsequently handled with special care. If any, this field of research, whether fundamental or applied, calls for the implementation of principles qualified as socially responsible (André, 2008). Charters, codes of conduct, guidelines and other such documents, are slowly adopted by institutions[3] that promote Socially Responsible Research, introduced with various status (contractual or not) in their rules and procedures.

Today, not a single country is exempt from the consequences of the early 21st century financial, economic and social crises which impact the capacities to pursue and meet the goals and objectives set globally, regionally or nationally for the health sector, both in terms of access to the health care system and improvement of the well being of the world population[4]. Funding issues, not only due to a trend towards aging of the world population, are among the challenges faced and choices have to be made among a variety of options. In our opinion, concentrating on the problems for which a sound alternative does not exist while privileging the most effective, efficient and sustainable measures probably represent the best strategic option. Decision makers, industrials, research institutions, non governmental organizations, many stakeholders claim that nanomedicine offers an excellent, innovative, and almost ideal solution to handle a lot of the medical problems faced by the world population, and at costs that can be contained (ETPN, 2006). Thus research in this discipline is evolving at a high speed.

Given our diverse respective skills and competencies, we chose to develop a case study on this specific field of application of nanotechnology, with special integrated inputs from each individual of the multidisciplinary team (photo chemist conducting research in basic sciences, risk management specialist, public health medical specialist), in order to develop an interdisciplinary expertise open to large societal needs.

[3] The European Charter for Researchers, edited in 2005 by the European Commission, has been adopted by many French research institutions, the National Centre for Scientific Research (CNRS) and the National Institute for Research in Agronomy (INRA) among others. Many other countries have also elaborated such document, Australia (Australian Code for Responsible Conduct of Research) and the USA (NIH Policy on instruction in the responsible conduct of research), among others.

[4] The 8 Millennium Development Goals (MDG), along with their quantified targets to be reached by 2015 for the vast majority, were adopted at the United Nations Millennium Summit in September 2000. Three of them, MDG 4 (Reduce child mortality), MDG 5 (Improve mental health), and MDG 6 (Combat HIV/AIDS, malaria and other major diseases) address specifically health issues. According to the last report, released in July 2011, they are unlikely to be met, particularly in the most vulnerable populations.

We will first give an overview of nanomedicine in order to better situate research on photodynamic therapy among the various objects covered by this field of application of nanotechnology. This whole descriptive phase is done in the second section. After introducing the nanomedicine taxonomy, we will present the principal historical and technical elements, starting from the broad perspective of drug delivery systems and/or materials, in the sub-section on biopharmaceutical, and then moving to a more detailed one, in the sub-section on photodynamic therapy. We will present photodynamic therapy more extensively, since the development of nanoparticles for use in this cancer treatment modality, the object of the research conducted by some of the authors[5], is at the foundation of our case study.

In the third section, centred on our case study, the design will be presented. The method and literature review will be introduced first. Then, in addition to a detailed presentation of the general framework, the research conducted by some of the authors, the content and implementation of a pilot survey that we conducted in June/July 2011 will be described. Our principal aim in doing that pilot survey was to dispose of first hand and shared information regarding the state of knowledge of our study object in the general educated public.

The results of our pilot survey will be analyzed in section four.

In the discussion, the following section, based on our professional experiences and using the main findings extracted from our pilot survey as illustrative arguments, we will highlight three management organizational modalities, the precautionary principle, knowledge management, and translational research, that we believe useful to implement in order to improve the risk management process in general, and the quality of our work, in particular.

We will close with some lessons learned from these management modalities, a few broad recommendations, and an opening on ethical and regulatory issues.

2. An overview of nanomedicine

The concept of using nanotechnology in medical research and clinical practice holds great values and opens real perspectives for effective innovations in medical sciences (Sandhiya *et al.*, 2009). Considering the broad spectrum of the series of technologies that can be used individually or in combination to make products and applications and to better understand science, efforts in classification have been initiated (Table 1). One way of characterizing nanotechnology is by utilizing the following four segments: tools, materials, devices & intelligent materials and machines. Those segments support the nanomedicine taxonomy developed by the Canadian Institutes of Health Research and the Canadian NanoBusiness Alliance (Gordon and Sagman, 2003). In a partial nanomedicine technologies taxonomy proposed in the beginning of this century, some of the most interesting and diverse current research projects within several of the 96 existing sub-categories were described (Freitas, 2005). The applications, which pertain to both medical and surgical diagnosis and therapeutic techniques, are as diverse as the development of nanoparticles for diagnostic and screening purposes, artificial receptors, DNA sequencing using nanopores, manufacture of unique drug

[5] This refers to those of us from the Reactions and Chemical Engineering Laboratory of the National Center for Scientific Research.

delivery systems, gene therapy, enablement of tissue engineering, and, single-virus detection (Emerich, 2005). Indeed, nanostructures, whether still under development or already in use, have the potential to play a critical role in the future of medicine, as they can be carriers for drugs, genes, and imaging agents as well as targeting units (Kateb *et al.*, 2011).

The need for the development of a new framework, including regulations, procedures and mechanisms, adapted specifically to govern research in nanotechnology, thus nanomedicine, is as salient in Europe as in the United States. Heightened uncertainty regarding risks, fast-evolving science yielding complex and increasingly active materials, likelihood of research on vulnerable participants including cancer patients, and potential risks to others beyond the research participant are identified as relevant confluent factors in support of the need for an exceptional oversight beyond the existing Common Rules[6] supervised by either local or federal institutions such as the US Department of Health and Human Services, the Federal Drug Administration (FDA), notably its Centres in charge of overseeing drug, biologics, and device approval, the National Institutes of Health or the National Cancer Institute (Wolf and Jones, 2011). Concern regarding potential risks does not involve only the volunteers and candidates recruited to participate to clinical trials, but also workers employed in the manufacturing industries, as well as, during the fundamental research stage, scientists conducting experiments in public or private laboratories.

At the European level also, the existing corpus of regulations needs to be progressively adapted to risk management specifically applicable to research in nanomedicine, even if enforcement will not necessarily be mandatory in each individual Member State. In France, Committees for the Protection of Citizen[7], structures similar to those existing in the US, are in place and the supporting laws under review.

Among the leading areas of research in nanomedicine summarized in the taxonomy (Table 1), biopharmaceutics offers an array of applications, with numerous drugs either already commercialised or still subject to ongoing research programmes, whether in the public or private sectors. Within this research area, the development of new materials or systems to optimize the deliver of the therapeutic agents is a steady option. The development of nanoparticles for photodynamic therapy falls under this category. Thus, in the following two sub-sections, we will highlight historical and technical elements, first on biopharmaceutical, then on photodynamic therapy.

2.1 Biopharmaceutics

Liposomes, spherical vesicles that can be produced from natural nontoxic phospholipids and cholesterol, are made of a lipid bilayer surrounding a water core hosting the drug. The first studies to report the efficiency of liposomes as nanoparticles focused on the improvement of pharmacokinetics and bio-distribution of doxorubicin. Approved by the FDA on November 17th 1995, Doxorubicin HCI Liposome injection, an anti-neoplastic drug administrable intravenously, is probably the first nanoscale drug delivery materials that

[6] Are identified as such the basic procedures existing since 1981 that rely mainly on Institutional Review Boards (IRB) and Data Safety Monitoring Boards or Committees.

[7] Comités de Protection des Personnes (CPP).

Nanotechnology segments & Nanomedicine Taxonomy (Source: Gordon and Sagman, 2003)[8]					
Tools	Materials			Devices	Intelligent Materials & Machines
Microscopy Techniques & Instruments	Raw Nanomaterials	Nanostructured Materials	Nanotubes & Fullerenes	Nano & Micro devices	(no application expected soon)
Biopharmaceutics	**Biopharmaceutics**				**Understanding Basic Life Processes**
Drug Discovery	**Drug Delivery (systems and materials)**				
Genetic testing* Ultra-sensitive Labelling & Detection Technologies High Throughput Arrays & Multiple Analyses ("Lab on a chip" Technologies)	Drug Encapsulation (Neurology, Ophthalmology)	Functional Drug Carriers (HIV/AIDS, Cancer)			Nanoscience in Life Sciences
Surgical Aids	**Implantable Materials**			**Implantable Devices**	**Implantab Materials**
Operating tools	**Tissue repair & replacement**			**Assessment & Treatment**	**Struct. Implant Mat.**
Smart Instruments (Surgical Tools & Optically Guided Surgery)	Implant coatings (e.g. for bone)			Implantable sensors (Blood level substance monitoring)	Smart Materials (Human enhancement)
Surgical Robots [Joystick handles, control console & camera system] (Gall Bladder, Prostate, Colorectal, Esophageal & Gastric bypass)	Tissue Regeneration Scaffolds (corneal, muscle, bladder & bone cells)			Implantable medical devices with sensors to automatically deliver treatment (MEMS)	
Diagnostic Tools	**Structural Implant Materials**			**Sensory Aids**	
Imaging	Bone repair (Substitute & Cement)			Retina Implants	
Nanoparticle Labels	Bioresorbable Materials (Sutures & Orthopaedic Fixation Devices)			(Degenerative diseases)	
Imaging Devices				Cochlear Implants (Hearing lost)	

* Measuring gene sequences (mutations) & expression levels (abundance)

Table 1. A summary of Nanomedicine Taxonomy arranged according to Nanotechnology supporting segments

[8] Original table adapted from the Nanomedicine Taxonomy. Potential inaccuracy may have occurred in the proposed synthesis, and some items of the Nanomedicine Taxonomy may belong to more than the segments under which they appear.

reached the market (Table 2). Soon after during the same year, an antifungal, Amphotericin B Lipid Complex Injection, was approved. A significant number of drug products have been commercialized since then, including on platforms other than liposomal such as nanocrystal platforms, for various ill states (hypercholesterolemia, hypertriglyceridemia, anorexia etc.), and, in 2005, indicated in metastatic breast cancer, the first albumin nanoparticle was approved for human use (Wang and Thanou, 2010; Kateb *et al.*, 2011).

Platform	Generic name	Indication	Route of Administration	Approval date
Liposomal platforms	Liposomal doxorubicin	Ovarian and breast cancer	intra venous (i.v.)	17 November 95
	Daunorubicin citrate liposome injection	Anti neoplastic	i.v.	8 April 96
	Pegylated liposomal doxorubicin hydrochloride	Ovarian and breast neoplasms Kaposi's Sarcoma Multiple myeloma	i.v.	21 June 96*
	Verteporfin for injection	PDT for age-related macular degeneration		12 April 00
	Non-pegylated liposomal doxorubicin formulation	Breast neoplasms	i.v.	13 July 00*
Other Platforms				
	Albumine taxol conjugate	Metastatic breast cancer	i.v.	7 January 05
	Dextran Ferumoxide injectable solution Superparamagnetic Iron Oxides Particles (SPIO)	Magnetic Resonance Imaging (MRI) contrast agent for the liver	i.v.	1 February 96
	Dextran coated iron oxide Ultra small Superparamagnetic Iron Oxides Particles (USPIO)	MRI contrast agent	i.v.	3 March 05

Table 2. Some of the main nanoscale drug product approved for cancer treatment and diagnosis by the Federal Drug Administration in the USA (Zolnik and Sadrieh, 2009 as cited in Kateb *et al.*, 2011) and by the European Medicines Agency* for use in the European Union

Some key properties that make nanoscale liposome an excellent device for targeted drug delivery have been highlighted recently: they are biocompatible, biodegradable, not immunogenic, familiar to and mastered by the scientific community, have known pharmacokinetics, bio-distribution and metabolism. Therefore, from a toxicological stand point, they have advantages over other nanoproducts for medical pharmaceutical applications as well as consumer products in cosmetics and clean technology. Additionally, if introduced in excess, the liposomes may be destroyed by the macrophages, as seen in the natural course of their life cycle (Barenholz, 2010). Convincing, and that may also address the question of eliminating the product, if need be, to deal with an irreversible risk later identified. Thus, the research can be pursued until challenged by new discoveries.

This leads us to introduce reversibility, another important factor that we consider worth exploring, besides the concerns that we previously mentioned as leading our interdisciplinary expertise: prevention and precaution as the underlying humanistic driving forces presiding to the management of health risks at short, medium and long terms, and, strategic research development options which promote the most equitable, effective, efficient and sustainable fields. These factors will be touched upon throughout the following sections, and in particular in the discussion, with extensive developments on three complementary approaches to consider in risk management applied to our case study: the precautionary principle, knowledge translation and translational research.

2.2 Photodynamic therapy

Photodynamic therapy (PDT) is another important emerging research field for the development of nanoscale therapeutics (Bechet *et al.*, 2008; Couleaud *et al.*, 2010b; Vanderesse *et al.*, 2011) that can optimize drug delivery of materials or systems. Photodynamic therapy involves the use of light, photosensitizers and oxygen. The photosensitizers, after excitation with light of an appropriate wavelength, can transfer their energy from their triplet excited state to neighboring oxygen molecules (Ortel *et al.*, 2009). Reactive Oxygen Species (ROS) and singlet oxygen ($^1O_2$), which is commonly accepted to be the main cytotoxic species, are formed and lead to the destruction of cancer cells by both apoptosis and necrosis. PDT efficiency depends on the photosensitizer's ability to produce ROS and $^1O_2$, availability (Verhille *et al.*, 2010), light dose and photosensitizer concentration in the treated area.

Used in dermatology to treat basal cell carcinomas as early as the beginning of last century, the application of this treatment modality stayed somehow confidential for quite a while, but PDT is now established as a clinical treatment modality for various diseases including cancer (Dougherty *et al.*, 1998; Triesscheijn *et al.*, 2006). PDT is currently being developed as a treatment for cancer of the esophagus, bronchi, and bladder, as well as for other non oncological applications, such as age-related macular degeneration (Agostinis *et al.*, 2011), and is also used as a successful non-invasive therapeutic modality for treating cutaneous neoplasm. In France, research on PDT is developed within different laboratories from the National Centre for Scientific Research, with four main hubs located respectively in the Parisian Region (Evry, Orsay), Lille, Nancy and Toulouse. A partnership has been officialised for 4 years in 2007, with the establishment of a research network within the National Centre for Scientific Research's Institute of Chemistry. Recognized by the acronym GDR[9], the objective of the research network is to develop medicines that can be activated by light along with their therapeutic applications.

The first photosensitizer to have received approval for use in PDT was porfimer sodium, authorized for the treatment of superficial bladder cancer in Canada in 1993, and later for early lung and advanced esophageal cancers in Netherlands and Japan (Triesscheijn *et al.*, 2006). Over the past decade, various types of photosensitizers have been developed in order to improve their properties (light absorption and diseased tissues selectivity) compared to the original first-generation ones (Table 3).

[9] Groupement de Recherche (GDR) 3049 PHOTOMED Médicaments Photoactivables – Photochimiothérapies. More information available from the web site at: http://www.gdr-photomed.cict.fr/spip.php?rubrique1.

Photosensitizer*	Cancer types	Chemical structure	Excitation wavelength (nm)	Approved
		First generation		
Porfimer sodium (Photofrin)(HPD)	Bile duct, Bladder, Brain, Lung, Oesophagus, Ovaries	Porphyrin	630	Yes
		Second generation		
ALA	Bladder, Brain, Oesophagus, Skin	Porphyrin precursor (protoporphyrin IX)	635	Yes
ALA esters	Bladder, Skin	Porphyrin precursor (protoporphyrin IX)	635	Yes
Temoporfin (Foscan)(*m*THPC)	Bile duct, Brain, Head and Neck, Lung, Skin	Chlorin	652	Yes
HPPH	Head and neck, Lung, Oesophagus	Chlorin	669	
SnEt2 (Purlytin)	Breast, Skin	Chlorin	660	
Mono-(L)-aspartylchlorin-e6 Talaporfin (LS11, MACE, NPe6)	Brain, Colon, Liver	Chlorin	660	
Ce6-PVP (Fotolon), Ce6 derivatives (radachlorin, photodithazine)	Brain, Nasopharyngeal, Sarcoma	Chlorin	660	
Silicon phtalocyanine	Cutaneous T-cell lymphoma	Phtalocyanine	675	
Padoporfin (Tookad)	Prostate	Bacteriochlorin	762	
Motexafin lutétium (Lutex)	Breast	Texaphyrin	732	
		Third generation		
Verteporfin (liposomal formulation)	Ophtalmic, Pancreas, Skin	Chlorin	690	Yes
		Fourth generation (theranostic)		
		Multifunctional nanoparticle		

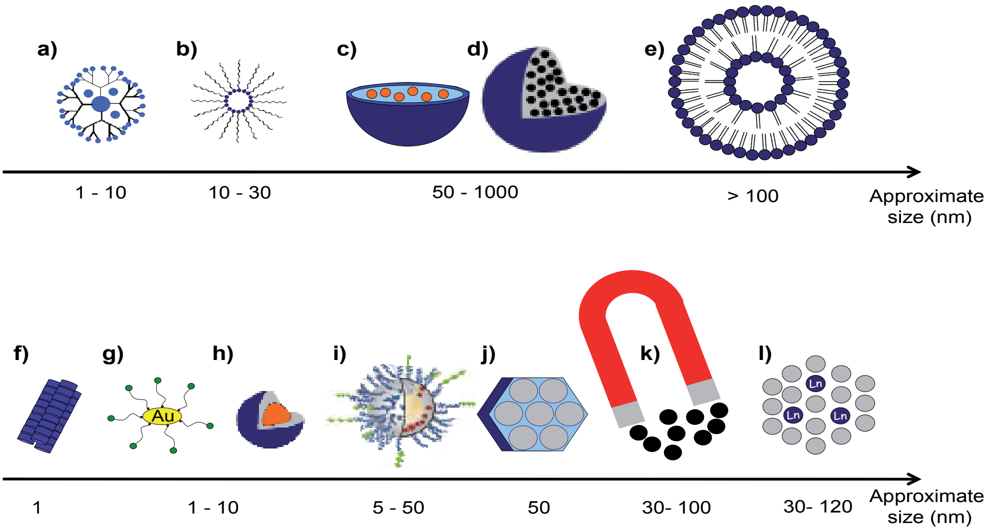
*Abbreviations: HPD [Hematoporphyrin Derivative]; ALA [5-Aminolaevulinic Acid]; *m*THPC [*meta*-tetra (hydroxyphenyl)chlorin] ; HPPH [2-(1-hexyloxyethyl)-2-devinyl-pyropheophorbide]; SnEt2 [tin ethyl etiopurpurin]; Ce6 [chlorin e6]; PVP [polyvinylpyrrolidone]

Table 3. Successive Generations of Photosensitizers used for the diagnostic or treatment of cancer with their main characteristics (Plaetzer *et al.*, 2009; Agostinis *et al.*, 2011)

A second-generation of photosensitizers, such as chlorins and phthalocyanines, present better light absorption in the red part of the visible light spectrum, which commonly represent the therapeutic window because of the slightest absorption of endogenous tissues compounds in this area (mainly hemoglobin and water). Thousands of patients have already been treated with PDT using the first and second generation of photosensitizers for a variety of advanced neoplasms, and the treatment induced a great improvement in their quality of life and lengthened their survival. Despite photophysical enhancements, the ability of these photosensitizers to discriminate between healthy cells and tumors remains low, leading to side effects such as photosensitivity in the presence of sunlight due to skin accumulation.

A third generation of photosensitizers has been developed, based on a strategy of selective delivery of the drug into diseased tissues. They consist of second-generation photosensitizers combined with targeting units such as oligo-nucleotides, peptides or monoclonal antibodies (Chen *et al.*, 2006), or photosensitizers encapsulated into nanoplatforms. Nanoparticles represent emerging photosensitizer carriers really promising for use in PDT. In bio-nanotechnology, their development can overcome most of the shortcomings of classic photosensitizers. Potential advantages of nanoparticles are that a high "payload" can be delivered: they can transport hydrophobic drug in blood, their surface can be modified with functional group, they have a large volume of distribution and are generally taken up efficiently by cells, and, controlled release of drug is possible. It is commonly accepted that strategies used to deliver the photosensitizer specifically to diseased tissues using the target tissue receptors or antigens are termed "active", whereas other formulations that enable parenteral administration and passive targeting are termed "passive".

Many nanoplatforms are being studied in the field of PDT (Figure 1), some of which are biodegradable (Figure 1a-e) and others not (Figure 1 f-l). The first nanodrug that reached the marked for use in PDT, approved in the USA by the FDA in April 2000, was a liposomal formulation, Verteporfin for injection, indicated for the treatment of predominantly classic subfoveal choroidal neovascularisation due to age-related macular degeneration, pathologic myopia or presumed ocular histoplasmosis (Table 2 and Figure 1e).



Biodegradable: a) dendrimers, b) polymeric micelles, c) polymeric capsules, d) polymeric nanospheres, e) liposomes,
Non biodegradable: f) carbon nanotubes, g) gold nanoparticles, h) quantum dots, i) silica nanoparticles, j) zeolite, k) magnetic nanoparticles, l) up-converting nanoparticles

Fig. 1. Main nanoplatforms used for cancer diagnosis and treatment in the field of PDT

Among the many photosensitizing agents that are being elaborated, very few have made it to robust clinical trials. Although third generation photosensitizers have been widely described for selective targeting, very few have been evaluated for clinical applications as their *in vivo* selectivity was not high enough compared to the one demonstrated *in vitro*.

However, the third generation sensitizers can potentially be grafted with targeting moieties and may encapsulate imaging agents and then lead to the fourth generation of photosensitizers. Theranostic nanoparticles can be considered as which can diagnose, deliver targeted therapy and monitor the response to therapy.

Some key properties explain why nanoscale liposomes are an excellent material for targeted drug delivery (Barenholz, 2010) and the successful outcome of research using this approach, with the commercialization in November 1995, almost 17 years ago, of the first nanodrug developed on a liposomal platform, indicated for the treatment of cancer in gynecology. Many other nanodrugs have reached the market since, developed not only on liposomal platforms nor for treating cancers, such as verteporfin for injection, a product indicated for photodynamic therapy (PDT) in ophthalmology (Table 2).

What factors, different in nature but similar by analogy to the key pharmacological factors identified and pushed forward for nanoscale liposomes, are likely to make research on nano PDT a sustainable and successful approach in the development of targeted drug delivery material or systems for the treatment of cancer?

In order to try to indentify such factors and attempt to provide an answer to this question, we designed a case study to explore specific factors known to be crucial to risk management in complex settings involving multidisciplinary team work, complicated decision channels, fast-evolving highly technical new science and heightened uncertainty regarding potential risks but considerable potential human and financial benefits. We also designed a qualitative survey, to be conducted by questionnaire, addressed electronically to a selected population ranging from researchers specialized or not in the field of PDT to the educated social body, to investigate their perception of the research conducted in the field of Nanotechnology and Nanomedicine in general, and PDT in particular. The design of the case study is described in the next section, which starts with a presentation of the methodology used in constructing our interdisciplinary expertise, and the results of the pilot survey are analysed in section four.

Considering the growing interest of the research community in nanomedicine, on one hand, and the complexity of nano PDT, a highly specialized field of research that involves many sectors and multiple scientific disciplines, on the other hand, it is difficult, if not impossible, for each individual researcher involved to apprehend completely the risk management process, in its full scope. Thus, in this chapter, we would like to investigate this further by identifying interventions that could be applied here to facilitate and improve the intake of risk management in the researchers' preoccupation. In the course of the implementation of our interdisciplinary expertise, we have narrowed down the scope of our investigations to address exclusively, but in depth, three structured management interventions and modalities that have demonstrated their usefulness, which will be presented in the discussion (section five).

3. Design of the case study

We start the description of the case study by introducing the method used and the literature that it is based on, before presenting the results of the pilot survey, in the following section (section 4).

3.1 Method and literature review

This interdisciplinary expertise was constructed using many iterative either bi or tri-directional exchanges which can be summarized in three broad steps, mixing both bottom-up and top-down approaches, moving from a large to more narrow perspective, inward and/or outward centre bound:

1. Identification of the scientific object of the case-study and definition of its content.
2. Mutual understanding of the each author's potential contribution for integration: constrains, competencies and operational skills.
3. Implementation by iterative construction.

Starting from our respective knowledge base we first agreed on a common study object: the research on nano PDT conducted at the Reactions and Chemical Engineering Laboratory of the National Centre for Scientific Research[10] located in Nancy. Although the experience was rejuvenating and stimulating, it entailed many challenged, some of which explain the limits of our work. The fact that the research on nano PDT involved other actors and ongoing scientific publications in a highly competitive field was a first obvious constrain that lead us to focus only on specific issues related to risk management. We also narrowed down the spectrum of our case study in order to obtain inputs from each individual of the multidisciplinary team: a photo-chemist conducting research in basic sciences and her post doctorate scholar, a risk management specialist and a public health medical specialist. We also expanded that spectrum, beyond a focus centred on individuals involved in that medical research domain, whether volunteers, patients or practitioners, in order to develop an interdisciplinary expertise open to large societal needs. Finally, among the many types of risks faced by research institutions, we concentrated mainly on internally driven hazard risk (i.e. health and safety issues), but also embraced, to a certain extend, both strategic and operational ones.

Beyond agreement on a common study object, a common understanding of each others' discipline was achieved in a progressive manner. Although the research on nano PDT conducted by some of us is at the foundation of our case study, our work also includes a qualitative survey. Our principal aim in doing so was to dispose of first hand and shared information regarding the state of knowledge of our study object in the general educated public.

We made an extensive use of the Information and Communication Technology (ICT), various software (Microsoft Work and Excel for instance) in particular to generate the various tables and figures, and of the Internet. We exchanged material (texts, tables, figures, articles) *via* e-mail all through the course of our work's development and used Google as a

[10] Laboratoire Réactions et Génie des Procédés (LRGP) du Centre National de Recherche Scientifique (CNRS) - France.

tool to search for articles, retrieve full texts of certain publications and consulted various institutions' website for information relevant to our case study. We held only one whole team physical meeting, but had many 2 or 3 team members work sessions, either in Paris or Nancy, or by telephone.

The literature review does not present any particularity, in the sense that we used well known tools and methods, besides the fact that it was implemented in 2 specific and distinct steps. A specific bibliography is generated on Scopus and Medline, using various key words relevant to the research developed on nano PDT by some of the authors (PDT, nanoparticles, targeting, chemical grafting of fluorescent dyes, spectroscopy, interaction between light and living cells, etc). This is done on a regular base. Additional information is also added in an ongoing manner, during various continuing education operations such as scholarships, participation to national or international conferences and seminars or while tutoring post-doctoral students. Similarly, a steady expertise is withheld by some of the team members in risk management and public health policy applied to nanotechnology and other medical fields. Beside data banks such as those already mentioned (Scopus and Medline), Google Google scholar and Wikipedia are tools used on a regular base to keep abreast with knowledge development in the field of competencies. Update or skill development are also achieved using different continuing education *media*.

The originality of our interdisciplinary expertise lies more on the procedure that we implemented. Since no relevant unique bibliography could be screened and shared by all, and considering the solid knowledge base detained individually, we provided each other with selected articles or information. Once the common scope and general direction were determined, we conducted additional literature research (Scopus, Medline, Google Scholar) for articles and institutional publications to complement our development on health and safety issues in a risk management perspective. The precautionary principle and knowledge translation were the first two concepts that appeared relevant to investigate in our search for an answer to our study question in the particular context of research on nano PDT. We expanded later and more briefly on translational research, as this new management paradigm is distinct from the others and holds interesting features which can also facilitate the risk management process and emphasize the sustainability of the research field.

Today, working across sectors, fields and institutions is more and more frequent and slowly becoming the norm. However creating a common understanding of terms and concepts is difficult, because of the so many different linguistic styles. The limits of a purely disciplinary approach, which we consider improper to conduct relevant successful research in such complex fields, plead for interdisciplinarity in order to achieve a sound and robust risk management process.

In the following two sub-sections, we will describe and analyze successively the two elements that constitute our case study: the research on nano PDT which is at its foundation and the qualitative survey, a pilot survey which constitutes a common baseline of information.

3.2 General framework of the case study

The strategy implemented at the Reactions and Chemical Engineering Laboratory (LRGP) of the National Centre for Scientific Research (CNRS), in close collaboration with the

Laboratory of Molecular Physical Chemistry (LCPM) and the Research Centre for Automatic Control – Centre Alexis Vautrin (CRAN-CAV), an anti cancer centre, both also located in Nancy, and the Physical Chemistry of Luminescent Materials Laboratory (LPCML), located in Lyon, consists in designing, characterizing and *in vivo/in vitro* testing new third and fourth generation photosensitizers to improve their selectivity for tumor vasculature and cells. So far, most of the endeavours in the development of tumor targeting-photosensitizers have focused on the targeting of markers over-expressed by tumor cells themselves, such as lectins (Distasio *et al.*, 2005) or folic acid receptors (Gravier *et al.*, 2008). Another promising strategy consisting in destroying endothelial cells lining angiogenic blood vessels may offer some advantages over the usual approach aimed at direct killing of tumor cells is also being developed. This vascular targeted PDT, designed by the acronym VTP for Vascular Targeted Photodynamic Therapy, could be proposed for a large number of vascularized tumors. Compared to conventional cancer cell-targeting approaches, targeting tumor vasculature is easier to access, more efficient in cancer cell killing, and has a lower likelihood of drug resistance development. This vascular effect can be potentiated either by modulating PDT scheduling or by designing photosensitizers that would localize primarily in the vascular compartment. For several years, the focus has been in developing photosensitizers coupled to vascular-targeting agents to target the neovasculature already formed (Tirand *et al.*, 2006; Thomas *et al.*, 2009). In order to improve PDT performance and more precisely the selectivity of the treatment, the potential of using nanoparticles is being investigated, particularly since 2008 (Bechet *et al.*, 2008; Couleaud *et al.*, 2010b), not only by the LRGP and its partners, but throughout Europe and in most of the OECD member countries.

Thus, starting from no publication prior to 1994, the number of publications on the development of nanoparticles for use in PDT has shown a sharp growth rate between 2002 and 2003 and is still increasing rapidly since 2006, reflecting the growing interest of the scientific community for the development of nano objects in the field of PDT (Figure 2).

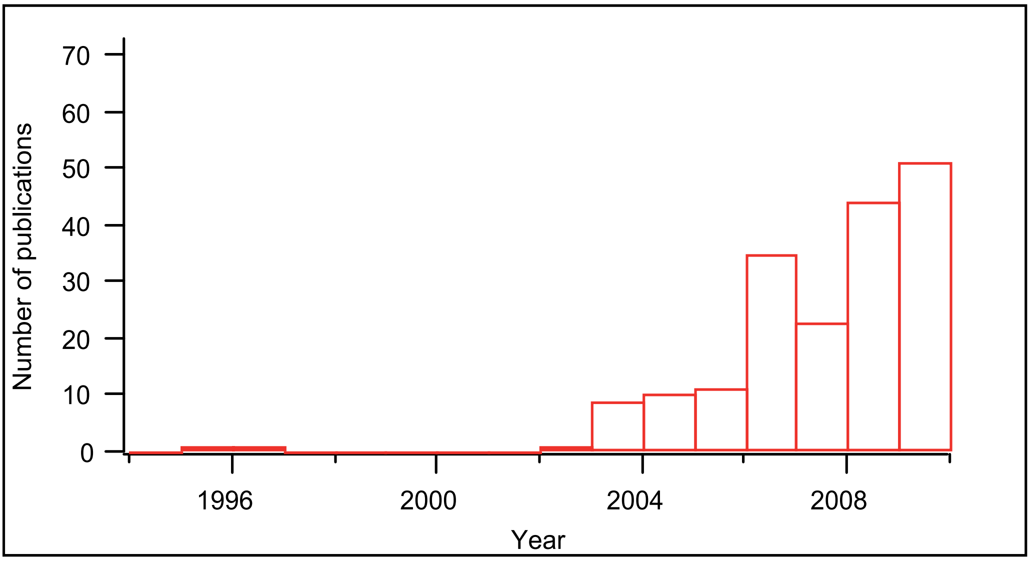


Fig. 2. Evolution of the number of publications on Photodynamic Therapy using nanoparticles from 1994 to 2010 (analysis from Scopus)

The French nano PDT hub located in Nancy develops, in collaboration with its partner in Lyon, multifunctional silica nanoparticles (Couleaud *et al.*, 2011). These non-biodegradable nanoparticles seem to be promised to a long lasting career, as they satisfy all the requirements for an ideal targeted PDT. For instance, Hybrid gadolinium oxide nanoparticles have been suggested to be useful for Magnetic Resonance Imaging (MRI), for both diagnosis and treatment (Faure *et al.*, 2008). Their small size (less than 50 nanometer in diameter), their silica shell that allows the covalent coupling of a photosensitizer, their polyethylene glycol grafting, make them an ideal nanoplatform for both imaging and Vascular Targeted Photodynamic therapy (VTP). Inhalation of crystalline silica is known to cause an inflammatory reaction and chronic exposure leads to lung fibrosis and can progress into the disease, silicosis. The mechanism by which particles are bound and internalized and the reason why particles are toxic is unclear. Amorphous silica has been considered to be a less toxic form, but this view is controversial. In order to inhibit tissue accumulation of the nanoparticles, one way is to promote the urinary excretion of the nanoparticles. This can be accomplished by increasing the solubility of the nanoparticles. In the multifunctional silica nanoparticles developed by the Nancy Hub Research Team & its partner in Lyon, polyethylene glycol grafting allows the covalent coupling of the peptide as well as a good solubility of the nanoparticles. By MRI *in vivo* in mice and *ex vivo* of the organs 24 hours post-intravenous injection, the elimination on the nanoparticles by renal excretion was demonstrated. More recently, the functionalization by peptides of such nanoparticle's photophysical properties as well as cytotoxicity and in vitro photodynamic efficiency, were also reported, describing for the first time the molecular affinity of the functionalized silica nanoparticle to neuropilin-1 molecular target (Couleaud *et al.*, 2011).

Capitalizing on our respective past professional experiences and with the motivation to broaden our views for future developments, we have chosen to investigate some organizational issues that we consider crucial to take in account in order to forestall potential occupational, environmental and more broadly societal health risks emerging in the development of therapeutics using nanomaterials in PDT.

One of the main focuses leading our study is "research and sustainability", notably waste disposal, and, in this particular instance, the elimination of the nanoproduct, both from the body and in the environment. Thus, beyond the individual approach, the core of this work also concentrates on workers' safety, and to a certain extent, exposure of close contacts and community effects. Indeed, since laboratory practices, waste, as well as human subject excretion of nanomaterials can raise environmental concern, analysis of environmental effects should also be examined (Wolf and Jones, 2011).

3.3 The pilot survey

The decision to conduct a qualitative survey rose from the need to dispose of first hand and shared information regarding the level of knowledge of our specific study object, nanomedicine and more precisely PDT using nanoparticles, in the educated public. It was designed to be conducted by questionnaire, addressed electronically to a selected population ranging from researchers specialized or not in the field of PDT to the educated social body, to investigate their perception of the research conducted in the field of nanomedicine in general, PDT in particular, and the associated potential health hazards along with the means of protection.

The questionnaire was elaborated with the intention to investigate different knowledge regarding medical research. A series of questions were organized in three sets that covered the following items:

- five short questions relating to general information on nanomedicine and the pre-requisite specific to the initiation of a clinical trial;
- a broad and detailed table to fill, with the choice given to answer either yes, no or don't know, to the same questions regarding the different phases in the development of a new product in medical research; the following points were investigated: interdisciplinary team work in practice, knowledge of *versus* compliance to the existing regulation on health & safety issues including their implementation in a cross sectoral and environment focused approach;
- two additional questions pertaining to the special field of clinical research, regarding the transition from experiments on animals to trials in human and the implementation of the principle of the necessary informed consent of the participants.

The target population was selected to cover an audience educated enough to be aware of the specific research field represented by nanomedicine. We initially aimed at two distinct groups: the general social body with approximately a master's level of education, investigated *via* Non Governmental Organisations (NGOs), and a population of researchers, either practising or still in training.

Four institutions were selected[11]:

1. the "Académie Lorraine des Sciences", a professional body with a total of approximately 300 members from multiple scientific backgrounds;
2. the "INstitut des Sciences de l'Ingénierie et des Systémes (INSIS)", one of the CNRS's major actor of research acting in engineering sciences and covering roughly 10 000 persons working in more than 150 laboratories all over the country;
3. "Poursuivre";
4. the "Association de Veille et d'Information Civique sur les Enjeux des Nanosciences et des Nanotechnologies (AVICENN)".

We were provided with feedback from only the latter.

The questionnaire was sent *via* e-mail to a representative of each institution, previously contacted, informed and sensitized of the strict deadline. Distribution was agreed to rest upon an intermediate within each of the selected institution, using internal mailing lists and the filled questionnaires were to be returned directly to the organizing team. A three weeks for response delay, going from June 10th to 30th 2011, was initially proposed to the participants, but the very low response rate in the course of implementation led to telephone contact with our correspondents and an additional five days bonus. The following biases were then identified:

[11] The "Académie Lorraine des Sciences" (Academy for Science of the Lorraine Region) is a local regional institution, and the two NGO's, "Poursuivre" and AVICENN (Surveillance and civic information on issues related to nanoscience and nanotechnology) have a national scope and audience.

- the institutions had transferred the questionnaire on a hierarchical mode, and, in most cases, the responsible had delegated the answer to the individual in their team they judged the most competent on the central theme of our survey;
- similarly, and in particular in the case of the NGOs, our correspondents informed us that their activities being extremely far from the subject of PDT, they did not feel legitimate to answer the questionnaire.

Although the total number of individual who actually received the questionnaire to fill is unclear, the response rate, with 21 questionnaires returned filled, among which 14 from the Nancy hub, was obviously certainly very low. This observation is particularly striking for the intended investigation of the general social body, as only one representative of AVICENN, identified as the focal point by the institution, provided us with one duly completed and extensively commented questionnaire that reflected the NGO's standpoint.

4. Results of the pilot survey

In fact, the questionnaire covers a wide sample of disciplinary and applied scientific fields: research, Research and Development (R&D), experiments on animals and clinical research in human, doctor-patient relationship, ethical and regulatory issues relating to social responsibility, etc. It was conceived to collect first hand and shared information on the competencies of different actors or social groups on our specific field of study: photodynamic therapy using nanoparticles in cancer therapy. The previously described limitations and biases, notably in the recruitment of the target population (screening and casting), make our pilot survey totally unsuited to be representative of the expression of an educated population's opinion, but still useful for our initial and immediate purpose. Moreover, this could also be an interesting preliminary and very instructive phase to determine feasibility of a similar survey on a larger scale.

In summary, our survey, implemented from June 10th to July 5th 2011, provided note worthy elements of information that served our need to better situate the use of nanoparticles in PDT as a field of relevant clinical research.

1. Respondents are rather competent on the central theme, master quite well the clinical medical issues, and in particular the classical known difficulties specific to moving from research to the application of the results in current practice (i.e. knowledge translation and translational research).
2. Association between creativity and interdisciplinary approach is subject to a real debate, partly due to lack in fruitful interactions between the many scientific fields and insufficient funding; however, when existing, the interdisciplinary mode of action is centred on the specific application, without an analysis of the related other aspects (i.e. health and safety issues, risk prevention, regulatory and ethical considerations).
3. Precautionary approach indispensable to conducting clinical research seems completely unknown, in particular the health & safety issues specific to each of the phases in a product's development starting from basic research is not the object of much attention from the various categories of responding professionals (researchers, technicians, maintenance, transport and cleaning personals). The discourse refers only to the existence of regulation without explicit mention of risk management in this context of

uncertainties; the precautionary principle for instance is not highlighted and priority seems to be given "only" to the scientific production attached to research. This comment does not apply to the answer provided by AVICENN, since the organisation is grounded on a social mission to inform on the risks related to nanosciences and nanotechnology, and feeds its legitimacy on the surveillance of their applications, thus on short and long term effects in human and the environment.

4. As a whole, risks identified are centred on those incurred by patients who are concerned with the intake of nanodrugs in clinical research, and in particular bioaccumulation, while hazards which may occur in hospital personnel who attend their needs are not mentioned. The same observation applies when it comes to the elimination of those substances which may cause damage to the environment, and the waste generated is not the object of expressed concern. Hence, life cycle analysis is an element that should be induced in the various professional's preoccupations.
5. Doctor-patient relationship in their interactions is well described, as well the role of the doctor in relation to his patients.
6. Interesting comments are made regarding commercialisation of new products after approval by the relevant authorities, but they are situated outside of the scope of our pilot project. They probably translate a context effect with the present debate occurring in France.

Concerning specifically health and safety preventive measures at the workplace and the question investigating knowledge of *versus* compliance to them, besides the fact that in most instance (70%) respondents either declared a lack of knowledge or did not pronounce themselves outside of their own area of practice, some interesting remarks made can be highlighted and commented:

1. The absence of a real culture of prevention in university settings, probably related to the absence of control by the relevant authority, along with a pressure to deliver immediately, which contrasts with the industrial world where risk prevention is rigorously handled, apart probably within subcontracting companies.
2. The need to upgrade the health and safety equipment and installations in the universities to render them appropriate to safe working conditions, leading responsible management to use some of their project funding toward this end, an honourable decision but nevertheless a form of misappropriation of funds.
3. The question of the information of the personnel, and more broadly of the citizens *vis à vis* the traditional dialectic of *a priori* criticism of the precautionary principle is raised. Viewed as a bridle to innovation and the freedom of entrepreneurship by some, the application of the precautionary principle can be an interesting economic and social development tool to avoid health crisis, such as the asbestosis one which negative effects are well known today. Moreover, it can sustain the public's demand of an As Low As Reasonably Achievable (ALARA) type of leadership (Manigat *et al.*, 2010).
4. The notion of independence of the experts in general, and of the scientists in particular, when conducting safety studies, imperious in order for the results to be delivered without pressure to major or minor risks, pleading for funding from independent institutions.

However, no conclusion in term of weight of the above mentioned remarks and comments can be reasonably made, as a good amount of respondents (30%) reported difficulty in understanding the questions that they qualified either unclear or expressed in a too complicated format, and skipped the question. Among those, around 50% expressed their reasons, which can be summarized by the following two explanations: the questionnaire was too long and the time announced to fill it was inadequate, and, the fields investigated were too broad and too distant from their own work environment.

5. Discussion

Whether political, administrative, legal, industrial, scientific or technical, public or private, any kind of organisation claims legitimately to aim at meeting their immediate, intermediate or long-term, more distant objectives. The key element for success is obviously a realistic planning were objectives are set taking in account the required material and human resources. Whether internally or externally driven, four types of generic risks can threaten the achievement of the goals and objective of an organisation: financial, operational, strategic and hazards (IRM, 2002).

Risk management, which can be defined as the set of individual or institutional response to the analysis and assessment of risks, including decisions to reduce or control risks (Ebbesen and Jensen, 2006), covers distinct realities according to the area involved (enterprise, industry, project, technology etc.). However, diverse initiatives have successfully conducted to a comprehensive approach and an internationally recognized level of standardization[12]. Risk management protects and adds value to the organisation and its stakeholders through supporting the organisation's objectives by different means, and notably: i) providing a framework that enables future activity to take place in a consistent and controlled manner, ii) improving decision making, planning and prioritisation, including in the allocation of material and human resources, iii) reducing volatility in the non essential areas, iv) protecting and enhancing assets and the organization's image, v) developing and supporting the knowledge base, and, optimizing operational efficiency (IRM, 2002).

Risk treatment, inclusive of the risk management process (Figure 3), comprises risk control/mitigation as its major element, but extends further to risk avoidance, risk transfer, risk financing etc. However, the standard states specifically that *compliance* with laws and regulations is not an option.

In the following three sub-sections, we will discuss concepts that we believe useful to take in account in order to improve the risk management process, in general, and the quality of our work, in particular. The precautionary principle and knowledge translation will be highlighted first and extensively, as they immediately appeared relevant to investigate in our search for an answer to our study question. We will also expand on translational research, as this new management paradigm, distinct from the others, also holds interesting features which can contribute to make research on nano PDT a sustainable and successful approach in the development of targeted drug delivery material or systems for the treatment of cancer.

[12] ISO/IEC Guide 73:2009. Risk management – Vocabulary, and, ISO/DIS 31000 (2009). Risk management – principles and guidelines on implementation, among others.

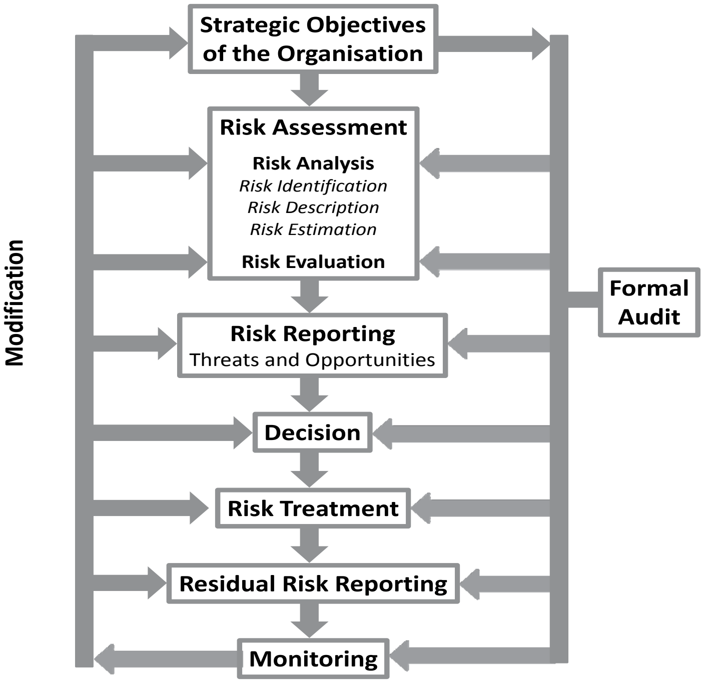


Fig. 3. The Risk Management Process (IRM, 2002)

5.1 The precautionary principle

Like scientists, who include elements of caution in the course of the assessment of scientific data, decision-makers are faced with similar issues, and can rely on the precautionary principle in a structured approach to risk management when shaping, accenting then releasing their decision. Although the two processes do resemble, they differ in the sense that the recourse to the precautionary principle is a particularly relevant strategy to the management of risk when potential dangerous and irreversible effects have been identified, but their scientific estimation holds a certain level of uncertainty. The decisions made using this strategic option can take many forms and does not necessarily have to be a binding legal measure. However, it should be transparent and involve, as early as possible and to a reasonable extent, all interested parties (CEC, 2000). In that respect, the recourse to public consultations, a practice that the Information and Communication Technology (ICT) has considerably facilitated, is spreading rapidly, and in particular within the European Commission.

In France, the principle was introduced in the country's legal regulation by the law related to the reinforcement of the environment's protection dated February 2nd 1995, often referred to as "la Loi Barnier", which expresses the way to explore systems where risks are not mastered. Hence, in its article L200-1, the Rural Code stipulates: The precautionary principle, according to which the absence of certainty, given the state of scientific and technical knowledge, should not delay the adoption of effective and proportionate measures to prevent a risk of serious and irreversible environmental damage at an economically acceptable cost. When applied, it is clearly a means to manage uncertainties and in no way a means of abstention. This new framework developed to implement action is of particular relevance in our contemporary world where innovation is intensively generated, as it leaves room to adapt to change.

At the supra national level, the most recent consensus rely on the February 2000 Communication from the Commission (CEC, 2000), in which all of the major components of the precautionary principle agreed on are specified, and in particular the need for decision makers to be aware of uncertainty attached to the result of the evaluation of the available scientific information and the eminently political responsibility for judging what is an acceptable level of risk for society. Gaps in knowledge and scientific uncertainty highlighted after the evaluation of a potential hazard are among the triggering factors that can lead to the recourse to the precautionary principle.

It is important to point that the general concept of risk management is quite similar to the precautionary principle, although it applies to a context of a more established certainty. The recourse to the precautionary principle, particularly relevant when facing a severe and irreversible potential hazard, can intervene after the risk assessment sequence of the risk management process.

The implementation of the precautionary principle in medical research and public health also calls to revisit the current causal inference methodologies (Kopelman *et al.*, 2004; Weed, 2004) and to refocus on many fundamental ethical principles and values (Jardine *et al.*, 2003). The principle is of extensive use also in the USA and Canada, by both Governments and Non Governmental Organizations, professional bodies, and the civil society, as illustrated by the following examples: the Canadian Environmental Act 1999 that came into force on March 31st 2000; the introduction of the precautionary principle by New York State and San Francisco City & County in their respective regulations in March 2003; the Wingspread Statement on the Precautionary Principle released by a group of scientists, philosophers, lawyers and environmental activists on January 20th 1998 to resume after a three days conference; or the American Public Health Association's resolution on the Precautionary Principle and Children's Health dated November 15th 2000.

Therefore, crucial to the theme of our work, our subsequent development will be centred on management and organizational modalities that have rendered their usefulness: knowledge translation and translational research.

5.2 Knowledge translation

Lack of knowledge, whether poor level of information or apparent inadequate use of knowledge regarding health and safety at the work place, regardless of the product in development's stage and environment (i.e. type of laboratory), was mentioned in a good proportion of the educated populations' responses to our questionnaire. Thus, considering the main characteristics attached to the subject of nanodrugs for PDT in cancer therapy[13], it

[13] One of the most rapidly growing branch of science and fairly new field of research – nanomedicine – but using an old but not frequently applied technology - PDT, numerous actors and stake holders involved, real controversies debated on health risks with sister technologies like carbon nanotubes, incredible economical weight voiced by industrials as to potential benefits challenged by NGO's as to

seems more than relevant, in today's health economic-aware environment, to question the knowledge translation process which provides scientific methods for closing the knowledge to action gap and identify the inherent barriers and facilitators (Straus *et al.*, 2009a).

Knowledge translation is defined as a dynamic and iterative process that includes synthesis, dissemination, exchange and ethically-sound application of knowledge to accelerate the capture of the benefits of research through improved health, more effective services and products, and to strengthen the health care system. This process takes place within a complex system of interactions among researchers and users that may vary in intensity, complexity and level of engagement depending on the nature of the research and the findings as well as the needs of the particular knowledge user (CIHR, 2004).

We adopted the knowledge to action process developed in Canada, as the guiding overarching framework to identify potential areas of possible usefulness for the improvement of such an intervention (Figure 4).

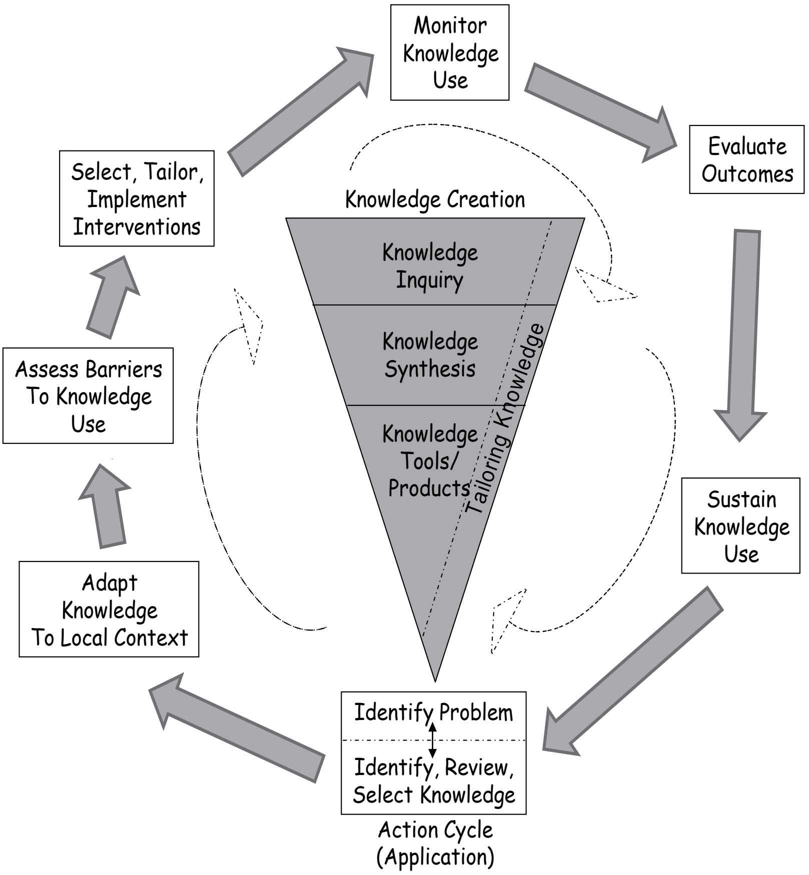

Monitor Knowledge Use
Evaluate Outcomes
Sustain Knowledge Use
Identify Problem
Identify, Review, Select Knowledge
Action Cycle (Application)
Adapt Knowledge To Local Context
Assess Barriers To Knowledge Use
Select, Tailor, Implement Interventions
Knowledge Creation
Knowledge Inquiry
Knowledge Synthesis
Knowledge Tools/ Products
Tailoring Knowledge

Fig. 4. Knowledge to action process (Graham et al., 2006)

long term potential effects detrimental to man, endless amount of available information, huge already existing market applications and developments.

In the course of the knowledge translation process development, whether during the creating or applying phases, some steps are more prone than others to allow identification of knowledge to action gaps, barriers or facilitators, which, when assessed at an early stage, can successfully lead to find suitable solutions. Considering the knowledge issues commonly identified in health and security studies, and also observed in the results of our pilot survey, it is worth stressing that knowledge needs and knowledge use should be the object of special focus, and carefully addressed.

The identification of knowledge to action gaps can occur anytime throughout the course of the knowledge translation development, either when refining distilling and tailoring knowledge to the needs of end-users during the knowledge funnel (i.e. inquiry, synthesis, and generation of user friendly knowledge tools and products), or when applying knowledge during the action cycle while assessing barriers & facilitators to knowledge use or developing mechanisms to sustain ongoing knowledge use (Straus *et al.*, 2009a).

It has been established that gaps between research and practice may result from interacting factors such as limited time and resources of practitioners, insufficient training, lack of feedback and incentives for use of evidence-based practices, inadequate infrastructure and systems organisation to support translation. More recently, the assumption that effectiveness research (intended to measure the impact of an intervention on the target population) naturally and logically follows from successful efficacy research (which purpose is to establish a causal link between and intervention and its outcome), has been challenged in health promotion research, leading to highlighting a disconnection inherent to their very distinct designs and purposes. The moderating factors, that limit robustness across settings, populations, and interventions staff, need to be addresses in efficacy studies (Glasgow *et al.*, 2003).

To sum up, we want to highlight the following remarks:

1. Underuse or overuse of evidence with critical impacts have often been reported, making monitoring of knowledge use a fundamental issue, considering the different types that may be involved; the appropriate method to close or lessen gaps in knowledge use will vary whether dealing with conceptual (to change the levels of knowledge, understanding and attitudes), instrumental (to change behaviour or practice) or strategic (as ammunition in the attainment of power or profit) knowledge use (Straus *et al.*, 2009a; Manigat *et al.*, 2010).
2. Assessing barriers and facilitators to the implementation of shared decision making is crucial, since barriers and facilitators are the salient beliefs of self efficacy, the most important determinant of behavior change after intention, on one hand, and, considering the fact that a given factor can be perceived as both a barrier or a facilitator, on the other hand (Straus *et al.*, 2009a).
3. Interest in sustainability of knowledge use is quite recent (Graham *et al.*, 2006), nevertheless, the consideration of sustainability, which includes discussion of budgetary and human resources, should occur early in the process (Straus *et al.*, 2009a).
4. Lack of knowledge, also identified in our survey concerning health and security issues, is, with variability, one of the two generic sources of uncertainty; it can be either structural or due to unreliability, and in that second instance potentially measurable (Van Asset, 2000, as cited by Hoppe, 2009; Van Asset *et al.*, 2002).

5. Uncertainty in knowledge is now an established triggering factor for the application of the precautionary principle (CEC, 2000).
6. When people are given more freedom to get involved, they more actively engage in finding creative solutions to routine problems and implementing them (Straus *et al.*, 2009a).

Ultimately, and consistent with the vision for knowledge management promoted by the Canadian Institute of Health Research (CIHR), the development of a systematic, integrated approach to accelerate optimal use of the best available research evidence in the interest of health would be particularly useful here. Furthermore, as often reported and also evidenced in the analysis of the information retrieved in our survey, it appears that a real margin of improvement does exists, since gaps in knowledge remain in the course of the implementation in highly interdisciplinary research in nanomedicine. The four comments aforementioned concerning specifically health and security preventive measures at the workplace are valid beyond biomedical research and the field of nanomedicine. Pushed forward with the recent controversies, still ongoing, regarding nanosciences and nanotechnology, they unfortunately reflect the general "traditional" and prevailing situation which critically calls for reconsideration. A deep reflexion could be advantageously developed, centred on the following: a reasoned exploration for the frame of application of the precautionary principle for nanodrugs and nanomedicine; communication on the innovations in the field; the interrelations between the product and technique conceivers and users; and, in order to prevent potential environmental hazards, the life cycle analysis, the development of research on processes to eliminate waste or recycle the material.

5.3 Translational research

Translational Research also aims to bring health information and discoveries to the public to improve well being and health. Although, the term may mean different things to different people, it was initially described as comprised of two areas of translation (Figure 6): the process of applying discoveries generated during research in the laboratory and in preclinical studies to the development of trials and studies in humans and *vice versa* – or from bench to bedside and back (T1 translation), on one hand, and, research aimed at enhancing the adoption of best practice in the community and back – or from research into practice (T2 translation). Two additional translations were later identified to better reflect the complexity of the second component of the translational research paradigm: moving research finding into the daily care of patients (T3), and moving scientific knowledge into the public sector and thereby changing people's everyday life (T4) (Woolf, 2008; Kon, 2008).

Improving the funding balance within translational research, between "bench to bedside (T1)" and "research into practice (T2)", the latter often overshadowed by the primer notably in the United States, was shown to be crucial at a time when experts warn of the fragmented health care system and the huge disparities in access and quality of care. Poverty matters as much as proteomics in understanding disease, even though the scientific discoveries and spectacular new devices are more fascinating to the public and more lucrative to the industry. Additionally, for many diseases, T2 could save more lives than T1 as, for example, greater fidelity in administering aspirin to eligible patients might prevent more strokes than developing more potent antiplatelet agents (Woolf, 2008). Under this translational research paradigm, (i.e. avoiding to neglect historically less glamorous components than bench research which are prone to dramatically improve public health and decrease cost), there will be more funding and consequently more recognition for such work (Kon, 2008).

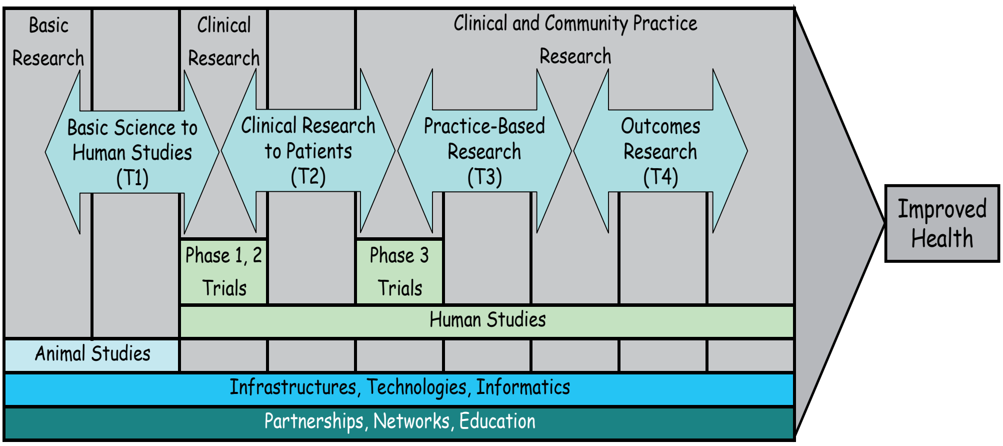

Basic Research
Clinical Research
Clinical and Community Practice Research
Basic Science to Human Studies (T1)
Clinical Research to Patients (T2)
Practice-Based Research (T3)
Outcomes Research (T4)
Phase 1, 2 Trials
Phase 3 Trials
Human Studies
Animal Studies
Infrastructures, Technologies, Informatics
Partnerships, Networks, Education
Improved Health

Fig. 5. The Continuum of Biomedical Research (UTSUMC, 2011)

The research on PDT using nanoparticles developed at the Reactions and Chemical Engineering Laboratory of the National Centre for Scientific Research is also implemented within a network of laboratories and could benefit from implementing the concept.

6. Closure and future work

Building on the study of 14 past cases, drawn as far as in 1896, related to hazards that caused damage to humans and the environment and for which early warnings where available, the European Environmental Agency (EEA, 2001) produced a report that illustrates the relevance of the precautionary principle's implementation. The following twelve late lessons learnt where highlighted from this historical analysis:

1. Acknowledge and respond to ignorance, as well as uncertainty and risk, in technology appraisal and policy-making.
2. Provide adequate long term environmental and health monitoring and research into early warnings.
3. Identify and work to reduce blind spots and gaps in scientific knowledge.
4. Identify and reduce interdisciplinary obstacles to learning.
5. Ensure that real world conditions are adequately accounted for in regulatory appraisal.
6. Systematically scrutinize the claimed justifications and benefits alongside the potential risks.
7. Evaluate a range of alternative options for meeting needs alongside the option under appraisal, and promote more robust, diverse and adaptable technologies so as to minimize the costs of surprises and maximise the benefits of innovation.
8. Ensure use of lay and local knowledge as well as relevant specialist expertise in the appraisal.

9. Take full account of the assumptions and values of different groups.
10. Maintain the regulatory independence of interested parties while retaining an inclusive approach to information and opinion gathering.
11. Identify and reduce institutional obstacles to learning and action.
12. Avoid paralysis by analysis by acting to reduce potential harm when there is reasonable ground for concern.

They all appear relevant to be taken in account in our subject of study with two central themes also quite consistent with the conclusive analysis of our pilot survey: knowledge issues in a scientific field were risks are identified but not yet fully evaluated thus withholding uncertainties, and, obstacles to learning in interdisciplinary settings.

The heightened possible risks widely described for certain nanomaterials in recent publications (Poland *et al.*, 2008; Takagi *et al.*, 2008; Genaidy *et al.*, 2009; Murphy *et al.*, 2011), the multiple sources of uncertainty, either primary due to variability or secondary in relation to limited knowledge, the absence of a robust common and universally accepted terminology, the ongoing debates to agree on appropriate regulation, are so many reasons that makes it legitimate to refer to the precautionary principles in the course of the decision making process when dealing with particles, materials and devices in nanomedicine.

Knowledge management is crucial, so referring to knowledge translation should also carefully be considered in that context of uncertainties and controversial points of view, where knowledge gaps are definitely known to exist. The systematic and full use of the knowledge to action process, funnel and action cycle, may offer a model on which research institutions should consider relying, either for internal evaluation (e.g. when monitoring knowledge use) or during external audits (e.g. for evaluation of outcomes). Although sustaining knowledge use appears as the last of the 7 steps in the action cycle (Figure 5), consideration of sustainability, which include discussion on budgetary and human resources, should occurs early in the process (Straus *et al.*, 2009a).

However, awareness should exist on the fact that a successful implementation of knowledge translation is also a daunting task, considering the challenges to be faced, and in particular the lack of knowledge management and infrastructure, such as the huge volume of research evidence currently produced, access to research evidence, time to read and skills to appraise, understand and apply research evidence. Other challenges identified operate at many other different levels, including the health care system (e.g. financial disincentives), the health care organization (e.g. lack of equipment or personnel), the health care teams and individual professional, and also the patient (Straus *et al.*, 2009b). Although the strategy may not have been extensively studied in the health care research sector, the use of either information specialists (Wilkinson *et al.*, 2009) or knowledge brokering (Oldham and McLean, 1997; Lomas, 2007; Dobbins *et al.*, 2009), as seen in clinical nursing practice (Kent *et al.*, 2009), could be an interesting strategy to implement after further proper assessment of possible added value, in order to enhance knowledge translation.

Since, when people are given more freedom to get involved, they more actively engage in finding creative solutions to routine problems and implementing them (Straus *et al.*, 2009a), management style is also important and the stewardship theory of management, developed as a counter strategy to agency theory in 1997 (Olson, 2008), should be explored. Beyond the afore mentioned need for a relevant level of material and human resources, indeed the five components of the management philosophy of the stewardship theory of management (i.e. performance enhancement, long-term orientation, empowerment, open communication & trusting relationship) could be inspiring to reduce interdisciplinary obstacles to learning as well as institutional obstacles to learning and action.

Although risk management is increasingly recognized as being concerned with both positive and negative aspects of risks, in the safety field, risk management's focus should be on prevention and mitigation of harm, as it is generally recognized that in that given field, consequences are only negative (IRM, 2002). Compared to research in techno-sciences, scientists who initiate research in the biomedical field have obviously integrated ethical issues in their processes (Couleaud *et al.*, 2010a). Among the explanations, we retain the consequences of some major health crisis such as the use of thalidomide and the cancer of the uterus later induced in the offspring or the asbestos induced mesothelioma in factory or construction industry workers. Today, risk analysis seems inclusive of the culture of most medical scientists.

The ethical considerations to take in account when addressing some technologies developed for applications in nanomedicine have been examined, notably for nanosurgery, tissue engineering, diagnostics and targeted drug delivery (Ebbesen and Jensen, 2006). Although ethical considerations to take in account when developing research in the field of nanomedicine are more complex, the knowledge base acquired in the field of bioethics seems sound enough, and the recourse to the four principles of Beauchamp and Childress' theory - doing Good (Beneficence), avoiding harm (Non-Maleficience), respect of autonomy (Autonomy Protection) and justice (Protection of Equity/Justice) - a set of values commonly used today in ethical analyses (Ebbesen and Jensen, 2006; Westerholm, 2010), something to build on. The increasing weight of cancer in the global mortality and morbidity legitimates to search for new drugs and the discovery of new modes of treatment. However, research protocols applied on vulnerable populations have also obvious ethical limitations and also call for a thorough oversight. The need for interdisciplinary research on the ethical, legal and social implication of nanomedicine and enhanced information exchange, particularly on toxicity studies and informed consent procedure with regard to safety, has been underlined by the European Group on Ethics in Science and New Technologies to the European Commission (EGESNT/EC, 2007). The group has also argued that as far as public participation is concerned, transparency, including openness about uncertainties and knowledge gaps, is essential to ensure public trust in nanotechnology. Moreover, the group has stressed that in prospective technology assessment, consideration of social effects should expand to the developing countries. Indeed we agree, as we believe in global adaptation of new technologies to the human species and not in the reverse paradigm (Manigat *et al.*, 2010). A large interdisciplinary scientific debate has to be opened to agree on boundaries and limits, notably to human enhancement, include the decision makers to set strategic orientations and expand to the civil society.

While ethical considerations should have always been consubstantial with the practice of medicine (Hippocrates's Oath, Code of deontology), the need for legislative measures has emerged partly with the growing complexity of medical practice and research in humans, and probably more strongly with the history of past events such as the Tuskegee syphilis experiment in the US (Emanuel and Menikoff, 2011). The urge to adapt the legal framework to the challenges of nanomedicine, maybe even on a regular base, is real, and the debate on the most appropriate reform is an ongoing process in the USA as well as in France and within the European Commission Member States. The call for the development of mandatory nanospecific regulation seems highly legitimate since most existing laws and regulations focus on cleaning up and controlling damage rather than preventing it[14], and is consistent with the claim from the industrials that nanotechnologies are novel technologies. Since the risk management process clearly specifies that compliance with the laws and regulations is not an option, needless to stress that the rules must be carefully conceived in order to be also applicable, and achieving that could be a daunting and lengthy process.

Additionally, monitoring of the insurance industry's policies is worth investigating and can provide useful inputs, as risk treatment can be dealt with using risk financing and insurers do not manage risk coverage haphazardly.

The implementation of this interdisciplinary expertise was a unique experience from which we can highlight the following elements as key messages:

- The limits of a purely disciplinary approach, which we consider improper to conduct relevant successful research in such complex fields, considering the multiplicity of the knowledge gaps and other issues that are at stake and jeopardize a sound risk management process.
- Although the pilot survey was developed for a precise and limited purpose, it could serve as the first step - the feasibility phase – of a qualitative study for a larger research project, if proper funding can be obtained.
- Lessons learnt from the precautionary principle, applied in risk management of many other instances, could be food for though in the development of nano PDT as a relevant and sustainable research field in medicine.
- Translational research, another new management paradigm, should also be explored.
- Other recommendations are expressed (e.g. the use of information specialists or knowledge brokering, the shift to a stewardship style of management), addressed to the decision makers, and it is up to the relevant authority to decide whether or not they want to implement them.
- Moreover, future work can also usefully be conducted, oriented in the two following sectoral directions that represent real openings: ethical and regulatory issues.

"One hopes that our society will be able to muster the collective financial and moral courage to allow such extraordinarily powerful medicine to be deployed for human betterment, with due regard to essential ethical considerations." Robert Freitas

7. Acknowledgements

We want to thank the following individuals and institutions for their support in the implementation of this work: Muriel Barberi-Heyob (Nancy Hub Research Team), Roger

[14] The whole rational, published on the Science & Environmental Health Network, relates to the aforementioned Wingspread Statement on the Precautionary Principle released by a group of scientists, philosophers, lawyers and environmental activists on January 20th 1998 to resume after a three days conference (Wingspread Conference on the Precautionary Principle. Available from: http://www.sehn.org/wing.html).

Salamon (President of the High Council on Public Health), Renée Pomarède (Ministère du travail, de l'emploi et de la santé); important partners in the pilot survey who helped us for the dissemination of the questionnaires: the non governmental organisation AVICENN (Association de Veille et d'Information Civique sur les Enjeux des Nanosciences et des Nanotechnologies) and the National Centre for Scientific Research's institute INSIS (INstitut des Sciences de l'Ingénierie et des Systémes); and Helène Leyland for her invaluable last minute inputs.

8. References

Agence Française de Sécurité Sanitaire de l'Environnement et du travail (AFSSET). (July 2006). Les nanomatériaux. Effets sur la santé de l'homme et sur l'environnement, AFSSET, France. Last consulted on August 30th 2011. Available from: http://www.afsset.fr/upload/bibliotheque/367611898456453755693572842048/nanomateriaux.pdf

Agence Française de Sécurité Sanitaire de l'Environnement et du travail (AFSSET). (July 2008). Les nanomatériaux. Sécurité au travail, AFSSET, France. Last consulted on August 30th 2011. Available from: http://www.afsset.fr/upload/bibliotheque/258113599692706655310496991596/afsset-nanomateriaux-2-avis-rapport-annexes-vdef.pdf

Agostinis, P.; Berg, K.; Cengel, K.A.; Foster, T.H.; Girotti, A.W.; Gollnick, S.O.; Hahn, S.M.; Hamlin, M.R.; Juzeniene, A.; Kessel, D.; Korbelik, M.; Moan, J.; Mroz, P.; Nowis, D.; Piette, J.; Wilson, B.C. & Golab, J. (2011). Photodynamic therapy of cancer : An update, *CancerJournal of Clinicians*, 61, 250-281

André, J.C. (2008). Vers le développement d'une recherche durable... ou vers une (ré)humanisation des sciences des artefacts, *Environnement, Risques et Santé,* 7 , 47-54

Barenholz, YC. (2010). Targeted Nanodrugs: The Drugs of the 21st century? Science, Technology, and Commercialization Venues. Lecture made on February 9th 2010 at the 6th Workshop, The Center for Nanoscience & Nanotechnology, Tel Aviv University, February 9-11, 2010, Maalot. Last consulted on August 30th 2011. Available from Available from: http://www.youtube.com/watch?v=7pvQxL2UnKY

Bechet, D.; Couleaud, P.; Frochot, C.; Viriot, M.L. & Barberi-Heyob, M. (2008). Nanoparticles for photodynamic therapy agent delivery, *Trends in Biotechnology*, 26, 612-621

Canadian Institute of Health Research (CIHR). (2004). Knowledge translation strategy 2004-2009. Innovation in action. Last consulted on August 30th 2011. Available from: http://www.cihr-irsc.gc.ca/e/26574.html

Chen, B.; Pogue, B.W.; Hoopes, P.J. & Hasan, T. (2006). Vascular and cellular targeting for photodynamic therapy. *Critical Reviews in Eukaryotic Gene Expression*;16(4):279-305

Commission of the European Communities (CEC). (2000). Communication from the Commission on the precautionary principle. Brussels, Belgium. Last consulted on August 30th 2011. Available from: http://ec.europa.eu/dgs/health_consumer/library/pub/pub07_en.pdf

Couleaud, P.; Faure, M.; Verhille, M.; Manigat, R. & André, J.C. (2010a). From public to occupational health: towards an inverse push-pull paradigm in nanotechnologies innovation, *G. Ital. Med. Lav. Erg*. 32:4, 400-402

Couleaud, P.; Morosini, V.; Frochot, C.; Richeter, S.; Raehm, L. & Durand, J.O. (2010b). Silica-based nanoparticles for photodynamic therapy applications. *Nanoscale, 2*, (7), 1083-1095

Couleaud, P.; Bechet, D.; Vanderesse, R.; Barberi-Heyob, M.; Faure, A.C.; Roux, S.; Tillement, O.; Porhel, S.; Guillemin, F. & Frochot, C. (2011). Functionalized silica-based nanoparticles for photodynamic therapy, *Nanomedecine*, under press

Department of Health and Human Services/Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health (DHHS-CDC-NIOSH). (2009). Approaches to safe Nanotechnology: Managing the Health and Safety concern Associated with Engineered Nanomaterials. Publication N° 2009-125. Last consulted on August 30th 2011. Available from: http://www.cdc.gov/niosh/docs/2009-125/pdfs/2009-125.pdf

Di Stasio, B.; Frochot, C.; Dumas, D.; Even, P.; Zwier, J.; Müller, A.; Didelon, D.; Guillemin, F.; Viriot, M.L. & Barberi-Heyob, M. (2005). The 2-aminoglucosamide motif improves cellular uptake and photodynamic activity of tetraphenyl porphyrin, *European Journal of Medical Chemistry*, 40, 1111-1122

Dobbins, M.; Robeson, P.; Ciliska, D.; Hana, S.; Cameron, R.; O'Mara, L.; DeCorby, K. & Mercer, S. (2009). A description of a knowledge broker role implemented as part of a randomized controlled trial evaluating three knowledge translation strategies. *Implementation Science,* 4, 23

Dougherty, T. J.; Gomer, C. J.; Henderson, B. W.; Jori, G.; Kessel, D.; Korbelik, M.; Moan, J. & Peng, Q. (1998). Photodynamic therapy, *Journal of the National Cancer Institute*, 90, 12, 889-905

Ebbesen, M. & Jensen, T.G. (2006). Nanomedicine: Techniques, Potentials and Ethical Implications. *Journal of Biomedicine and Biotechnology*, Volume 2006, Article ID51516, Pages 1-11

Emerich, D.F. (2005). Nanomedicine: prospective therapeutic and diagnostic applications, *Expert Opinion in Biological Therapy*, 5(1):1-5

Emanuel, E.J. & Menikoff, J. (2011). Reforming the Regulations Governing Research with Human Subjects. *N Engl J Med*, 2011 Jul 25.[Epub ahead of print. Last consulted on August 30th 2011. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/21787202

European Environment Agency (EEA). (2001). Late lessons from early warnings: the precautionary principle 1896-2000. Environmental issue report n°22. Luxembourg: Office for Official Publications of the European Communities, 2001. ISBN 92-9167-323-4. Last consulted on August 30th 2011. Available from: http://www.eea.europa.eu/publications/environmental_issue_report_2001_22

European Group on Ethics in Science and New Technologies to the European Commission (EGESNT/EC). (2007). Opinion on the ethical aspects of nanomedicine. Last consulted on August 30th 2011. Available from:
http://ec.europa.eu/bepa/european-group-ethics/docs/publications/opinion_21_nano_en.pdf

European Technology Platform on Nanomedicine (ETPM). (2006). Nanomedicine, Nanotechnology for Health. Strategic Research Agenda for Nanomedicine. Last consulted on August 30th 2011. Available from:
ftp://ftp.cordis.europa.eu/pub/nanotechnology/docs/nanomedicine_bat_en.pdf

Faure, A.C.; Hoffmann, C.; Bazzi, R.; Goubard, F.; Pauthe, E.; Marquette, C.; Blum, L.; Perriat, P.; Roux, S. & Tillement, O. (2008). Functionalization of luminescent aminated particles for facile bioconjugation, *ACS nano*, 2(11):2273-82

Freitas, R.A. (2005). What is nanomedicine? *Nanomedicine: Nanotechnology, Biology and Medicine,* 1 (2005) 2-9. Last consulted on August 30th 2011. Available from: http://www.cs.duke.edu/bioComp/referencesFall07/Nanomedicine.pdf

Genaidy, A.; Tolaymat, T.; Sequeira, R.; Rinder, M. & Dionysiou, D. (2009). Health effects of exposure to carbon nanofibers: Systematic review, critical appraisal, meta analysis and research to practice perspectives. *Science of the Total Environment*, 407, 3686-3701

Glasgow, R.; Lichtenstein, E. & Marcus, A. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health,* 93, 1261-1267

Gordon, N. & Sagman, U. (2003). Nanomedicine Taxonomy. Canadian Institutes of Health Research & Canadian NanoBusiness Alliance. Last consulted on August 30th 2011. Available from: http://www.pain.cz/nanomedicina/files/taxanomy.pdf

Graham, I.D.; Logan, J.; Harrison, M.B.; Straus, S.E.; Tetroe, J.; Caswell, W. & Robinson, N. (2006). Lost in Knowledge Translation: time for a Map? *The Journal of Continuing Education in the Health Professions,* 26, 13-24

Gravier, J.; Schneider, R.; Frochot, C.; Bastogne, T.; Schmitt, F.; Didelon, J.; Guillemin, F. & Barberi-Heyob, M. (2008). Improvement of m-THPC-like photosensitizer selectivity with Folate-based Targeted Delivery. Synthesis and in vivo Delivery Studies, *Journal of Medical Chemistry*, 51, 3867-3877

Hoppe, R. (2009). Risk and Uncertainty: Politics and Analysis. *Proceedings of Governing Uncertainty: the contribution of social sciences to the governance of risks in environmental health,* Conference held at Ecole des Mines, Paris, France, July 2009. Last consulted on August 30th 2011. Available from:
http://www.afsset.fr/upload/bibliotheque/935409038664891455468866124930/governing_uncertainty_en.pdf

Haut Conseil de la Santé Publique (HCSP). (2009). Avis relatif à la sécurité des travailleurs lors de l'exposition aux nanotubes de carbone. 7 Janvier 2009. Last consulted on August 30th 2011. Available from :
http://www.hcsp.fr/docspdf/avisrapports/hcspa20090107_ExpNanoCarbone.pdf

Institute of Risk Management (IRM). (2002). A risk management standard. Published by AIRMIC, ALARM, IRM: 2002, London, UK. Last consulted on August 30th 2011. Available from:
http://www.theirm.org/publications/documents/Risk_Management_Standard_030820.pdf

Jardine, C.; Hrudev, S.; Shortreed, J.; Craig, L.; Krewski, D.; Furgal, C. & McColl, S. (2003). Risk management framework for human health and environmental risks. *J Toxicol Environ Health B Crit Rev.* 2003 Nov-Dec;6(6):569-720

Kaluza, S.; Balderhaar, J.K.; Orthen B.; Honnert, B.; Rosell, M. G.; Tanarro, C.; Tejedor, J. & Zugasti, A. (2008). Workplace exposure to nanoparticles. European Risks Observatory Literature Review. European Agency for Safety and Health at Work (EU-OSHA). Last consulted on August 30th 2011. Available from: http://osha.europa.eu/en/publications/literature_reviews/workplace_exposure_to_nanoparticles

Kateb, B.; Chiu, K.; Black, K.L.; Yamamoto, V.; Khalsa, B.; Ljubimova, J.Y.; Ding, H.; Patil, R.; Portilla-Arias, J.A.; Modo, M.; Moore, D.F.; Farahani, K.; Okun, M.S.; Prakash, N.; Neman, J.; Ahdoot, D.; Grundfest, W.; Nikzad, S. & Heiss, J.D. (2011). Nanoplatforms for constructing new approaches to cancer treatment, imaging and drug delivery: what should be the policy? *Neuroimage*, 54 (1), 106-124

Kent, B.; Hutchinson, A.M. & Fineout-Overholt, E. (2009). Getting evidence into practice-understanding knowledge translation to achieve practice change. *Worldviews Evid Based Nurs.* 6(3), 183-185

Kon, A.A. (2008). The Clinical and Translational Science Award (CTSA) Consortium and the Translational Research Model. *American Journal of Bioethics,* 8(3), 58-60

Kopelman, L.M.; Resnick, D & Weed D.L. (2004). What is the role of the precautionary principle in the philosophy od medicine and bioethics? *J Med Philos.,* Jun:29(3):255-8

Lomas, J. (2007). The in-between world of knowledge brokering. *BMJ,* 334:1, 29-32

Manigat, R.; Wallet, F. & André, J.C. (2010). From past to better public health programme planning for possible future global threats: case studies applied to infection control. *Annalli dell' Istituto Superiori di Sanita,* 46, 3, 228-235

Murphy, F.A.; Poland, C.A.; Duffin, R.; Al-Jamal, K.T.; Ali-Boucetta, H.; Nunes, A.; Byrne, F.; Prina-Mello, A.; Volkov, Y.; Li, S.; Mather, S.J.; Bianco, A.; Prato, M.; MacNee, W.; Wallace, W.A.; Kostarelos, K. & Donaldson, K. (2011). Length-Dependent Retention of Carbon Nanotubes in the Pleural Space of Mice Initiates Sustained Inflammation and Progressive Fibrosis on the Parietal Pleura. *The American Journal of Pathology,* Volume 178, Issue 6, Pages 2587-2600

Oldham, G. & McLean, R. (1997). Approaches to Knowledge Brokering. Last consulted on August 30th 2011. Available from:
http://www.iisd.org/pdf/2001/networks_knowledge_brokering.pdf

Olson, K. (2008). The relationship between Stewardship Theory of Management and Employee Engagement: A case study exploration of the leadership philosophy of a professional services firm. *Proceedings of the Midwest Academy of Management 2008 Annual Conference*, St Louis, Missouri, October 2-4, 2008

Ortel, B.; Shea, C.R. & Calzavara-Pinton, P. (2009). Molecular mechanisms of photodynamic therapy. *Frontiers in bioscience : a journal and virtual library*, 14, 4157-4172

Ostiguy, C.; Roberge, B.; Woods, C. & Soucy, B. (2010). Engineered Nanoparticles: Current Knowledge about Occupational Health and Safety Risks and Prevention Measures. Studies and Research Projects. Report R-656. Institut de Recherche Robert-Sauvé en Santé et Sécurité du Travail (IRSST). Last consulted on August 30th 2011. Available from: http://www.irsst.qc.ca/media/documents/PubIRSST/R-656.pdf

Plaetzer, K.; Krammer, B.; Berlanda, J.; Berr, F. & Kiesslich, T. (2009). Photophysics and photochemistry of photodynamic therapy: fundamental aspects. *Lasers Medical Science*, 24(2), 259-268.

Poland, C.A.; Duffin, R.; Kinloch, I.; Maynard, A.; Wallace, W.A.H.; Seaton, A.; Stone, V.; Brown, S.; MacNee, W. & Donaldson, K. (2008). Carbone nanotubes introduced into the abdominal cavity of mice show asbestos-like pathogenicity in a pilot study. *Nature Nanotechnology*, 3, 423-428

Sandhiya, S.; Dkhar, SA. & Surendiran, A. (2009). Emerging trends in nano-medicine : an overview, *Fundamental & Clinical Pharmacology*, 23(3), 263-269

Straus, S.; Tetro, J. & Graham, I. (2009a). Power Point slideshow presentations from *Knowledge translation in health care: moving from evidence to practice,* Wiley-Blackwell, ISBN 978-1-4051-8106-8, UK. Last consulted on August 30th 2011. Available from: http://www.cihr-irsc.gc.ca/e/40618.html

Straus, S.; Tetro, J. & Graham, I. (2009b). Knowledge translation is the use of knowledge in health care decision making. *Journal of Clinical Epidemiology,* 64, 6-10

Takagi, A.; Hirose, A.; Nishimura, T.; Fukumori, N.; Ogata, A.; Ohashi, N.; Kitajima, S. & Kanno, J. (2008). Induction of mesothelioma in p53+/- mouse by intraperitoneal application of multi-wall carbon nanotube*, The Journal of Toxicological Sciences*, vol. 33 : n° 1, 105-116

Tirand, L.; Frochot, C.; Vanderesse, R.; Thomas, N.; Trinquet, E.; Pinel, S.; Viriot, M.L.; Guillemin, F. & Barberi-Heyob, M. (2006). A peptide competing with VEGF165 binding on neuropilin-1 mediates targeting of a chlorin-type photosensitizer and potentiates its photodynamic activity in human endothelial cells, *Journal of Controlled Release*, 111, 153-164

Thomas, N.; Bechet, D.; Becuwe, P.; Tirand, L.; Vanderesse, R.; Frochot, C.; Guillemin, F. & Barberi-Heyob, M. (2009). Peptide conjugated chlorin-type photosensitizer binds neuropilin-1 in vitro and in vivo, *Journal of Photochemistry and Photobiology B: Biology*, 96, 101-108

Triesscheijn, M.; Baas, P.; Schellens, J.H.M. & Stewart, F.A. (2006). Photodynamic therapy in oncology, *The Oncologist*, 11, 1034-1044

UT Southwestern University Medical Centre Web site (UTSUMC). (2011). What is Translational Research*? Continuum of Biomedical Research, Original Figure adapted from the National Center for Research Resources Strategic Plan 2009-2013*, Last consulted on August 30th 2011. Available from: http://www.utsouthwestern.edu/utsw/cda/dept440996/files/489751.html

Van Asselt, M.B.A. (2000), as cited in Hoppe, R. (2009). Risk and Uncertainty: Politics and Analysis. *Proceedings of Governing Uncertainty: the contribution of social sciences to the governance of risks in environmental health,* Conference held at Ecole des Mines, Paris, France, July 2009

Van Asselt, M.B.A.; Huijs, S. & Van't Klooster, S.A. (2002). The intriguing relationship between uncertainties and normativity: the need for pluralistic assessments. *Paper prepared for a workshop Normativitat und Unsicherheit,* University of Stuttgardt, Germany, February 2002. Last consulted on August 30th 2011. Available from: http://www.uni-stuttgart.de/philo/ngm/van.pdf

Vanderesse, R.; Frochot, C.; Barberi-Heyob, M.; Richeter, S.; Raehm, L. & Durad, J.O. (2011). Eds. *Nanoparticles for Photodynamic Therapy Applications*. A. Prokop (ed.), Intracellular Delivery: Fundamentals and Applications, Fundamental Biomedical Technologies 5, DOI 10.1007/978-94-007-1248-5_19, © Springer Science+Business Media B.V.

Verhille, M.; Couleaud, P.; Vanderesse, R.; Brault, D.; Barberi-Heyob, M. & Frochot, C. (2010). Modulation of photosensitization processes for an improved targeted photodynamic therapy. *Current Medicinal Chemistry,* 2010, *17*, (32), 3925-3943

Wang, M. & Thanou, M. (2010). Targeting nanoparticles to cancer, *Pharmacological Research*, 62, 90-99

Weed, D.L. (2004). Methodologic implications of the Precautionary Principle: causal criteria. *Int J Occup Med Environ Health* (2004);17(1):77-81

Westerholm, P. (2010). The ethical challenges of the occupational physician in our time, *G. Ital. Med. Lav. Erg*. 32:4, 403-406

Wilkinson, A.; Papaioannou, D.; Keent, C. & Booth, A. (2009). The role of the information specialist in supporting knowledge transfer: a public health information case study. *Health Information and Libraries Journal,* 26, 118-125

Wolf, S.M. & Jones, C.M. (2011). Designing Oversight for Nanomedicine Research in Human Subjects: Systematic Analysis of Exceptional Oversight for Emerging Technologies, *Journal of Nanoparticle Research,* Forthcoming; Minnesota Legal Studies Research Paper N°. 11-04. Last consulted on August 30th 2011. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1744336

Woolf, S.H. (2008). The Meaning of Translational Research and Why It Matters. *JAMA,* 299(2), 211-213

Post-Operative Residual Curarization (PORC): A Big Issue for Patients' Safety

A. Castagnoli, M. Adversi, G. Innocenti, G.F. Di Nino and R.M. Melotti
Anesthesiology and Intensive Care, S. Orsola-Malpighi Hospital, University of Bologna, Italy

1. Introduction

The post-operative residual curarization (PORC) indicates incomplete recovery of muscle function following the administration of neuromuscular blocking agents. The phenomenon of residual curarization actually exists and has a significant impact which is often underestimated. The complications that can result are severe, in terms of patient's outcomes, costs and legal issues. The neuromuscular blocking agents may give as main side effects hypersensitivity reactions or residual curarization.

Recent surveys have estimated the real impact of PORC to be between 4 and 50%. The absence of residual paralysis can be defined as the patient's ability to breathe and cough normally, and the presence of airway reflexes in order to prevent the inhalation of gastric material. On the other hand, a residual paralysis is associated with hypoxia, weakness leading to respiratory failure and increased perioperative morbidity. It is not uncommon for these features to be more or less obvious in the postoperative period, but symptoms are generally not attributed to residual curarization.

The PORC increases the risk of complications, especially if the patient is suffering from respiratory comorbidity and if after the intervention he's transferred directly to the ward (rather than in a protected environment like a recovery room or ICU) lacking adequate monitoring and oxygen supply.

Subjects such as "postoperative residual curarization" and "inadequate intraoperative curarization" are frequently met by disinterest in anesthetists, since they are problems that are widely underestimated and poorly understood, in our experience. The monitoring of neuromuscular functions is not mandatory in the operating room, and several investigations (Di Marco et al. , 2010; Naguib et al. , 2010) have shown that the equipment necessary to do so is not sufficiently available in the vast majority of hospitals.

However, the Ministry of Health in Italy in 2009 issued a handbook on safety in the operating room with specific objectives, which included preventing damage from anesthesia and guaranteeing vital functions. Among the requirements: "devices to monitor neuromuscular transmission must always be available."

Currently, there is a gap in terms of guidelines and recommendations of scientific societies on this issue, and in daily practice this coincides with a variety of management strategies for neuromuscular blockade. It is hoped that in the near future, guidelines and recommendations will be made in order to enable specific strategies to standardize the management of neuromuscular blocking agents.

The objective of this work is therefore to update the state of the art on PORC and risk management of patients with persistent neuromuscular blockade. We integrate our expertise with significant publications on the subject. We carry out a careful review of the literature using electronic databases, analysing original papers, systematic reviews and guidelines.

We aim to provide evidences, in terms of problem definition and epidemiology, on appropriate neuromuscular monitoring, and pharmacokinetics and dynamics of the most commonly used drugs for neuromuscular blocking and reversal. We also provide a brief description of the "Incident Reporting" system, a useful tool to monitor errors and near misses. We also suggest possible ways to correctly prevent or manage PORC, in order to optimize the use of available resources.

The outline of the following chapter reflects this order: starting from the evidences we found in literature about PORC, first we define the phenomenon and its epidemiology, then we describe pharmacological features of NMBAs, ways of neuromuscular monitoring, and NMBAs reversal drugs. Finally, we propose risk management tools, as the "incident reporting" system, together with a specific algorithm.

2. Literature review

A proper analysis of the problem is essential for the in-hospital management of the patient, as a guarantee of his safety and of the anesthesia best practice. We analyse in detail the incidence of the problem and the perception of those directly involved.

In order to update knowledge of the phenomenon of non-monitored curarization and related adverse events, we performed a literature search aimed at identifying the most relevant and recent evidence. The databases and research strategy are:

- Primary studies and case reports: Medline and Cinhal.
- Systematic reviews: Cochrane Library.
- Best Practice Guidelines: National Guidelines Plan (PNLG): http://www.snlg- iss.it /; Agency for Healthcare Research and Quality (AHRQ): http://www.ahrq.gov; National Guideline Clearinghouse: http : / / www.guideline.gov; Haute Autorité de Sante ``: http://www.has-sante.fr; Expert Consensus Guidelines (EKS): http://www.psychguides. com, American Psychiatric Association, http://www. psych.org, American Academy of Neurology: http:// www.aan.com, National Institute for Health and Clinical Excellence http://www.nice.org.uk; Scottish Intercollegiate Guidelines Network: http://www. sign.ac.uk; Canadian Medical Association: http://www.cma.ca; Canadian Task Force on Preventive Health Care: http:// www.ctfphc.org; New Zealand Guidelines Group: http://www.nzgg . org.nz;
- Search strategy used: "Neuromuscular Blockade" [MeSH] OR "Neuromuscular Blocking Agents" [MeSH] OR "Anesthesia Recovery Period" [MeSH] OR "Neuromuscular Blockade" [TIAB] OR "Neuromuscular Blocking Agents" [TIAB] OR "Residual curarization" [TIAB] OR "PORC" [TIAB]. The selected studies are listed in the bibliography.

We selected the studies basing on the value of the discussed topics and on the structural coherence of the articles. We found a poor perception of the problem but also some potential strategies for the prevention and proper management of PORC: different ways of neuromuscular monitoring, correct knowledge of the pharmacology of different NMBAs and reversals, available systems of incident reporting and appropriate hospital wards where to manage patients.

In this paper we want to integrate the strategies we found in literature in a coherent and concrete path of "Risk Management".

3. Post-operative residual curarization (PORC)

We start out our discussion of PORC by properly defining the term (3.a); afterwards, we go through its epidemiology and related guidelines and legal issues.(3.b and c); then, we end this section describing the pharmacological key features of neuromuscolar blocking agents. (3.d).

3.1 PORC: Post-operative residual curarization – Definition

PORC is an abbreviation for postoperative residual curarization, identified by instrumental signs (TOFRatio <0.9 -1.0) and clinical signs such as:

- evident muscle fatigue or 'fade' due to continuing occupation of presynaptic receptors by molecules of curare
- attenuation of the hypoxic reflex due to the inhibition of functional nicotinic cholinergic receptors of the carotid glomus
- pharyngolaryngeal dysfunction with loss of airway patency and the risk of "aspiration". The muscles involved in swallowing, such as those of the tongue or pharynx, decurarizate with more difficulty and are therefore implicated in the PORC phenomena
- Reduction of the cough reflex, and reduced expansion of the rib cage, with superficial ventilation and often inadequate and decreased clearance of tracheobronchial secretions

Pulmonary postoperative complications (POPC) include acute respiratory failure and aspiration pneumonia. The possibility of an acute respiratory failure, due to the collapse of the hypopharynx and the tongue base, to a deficiency of the diaphragm and intercostal muscle pump and to a compromised hypoxic reflex, suggests to hold the patient in a "recovery room" after the operation, for the purpose of monitoring and prompt treatment. This occurrence, if not readily diagnosed and treated by experienced staff, establishes a condition of severe hypoxia with psychomotor agitation that can result in severe multiple anoxic organ damage and psychiatric post-traumatic syndromes, which are debilitating and difficult to manage.

Pneumonia caused by "aspiration" secondary to inhalation after extubation, due to lack of pharyngolaryngeal muscle function or secondary to repeated microinhalations in the early hours of admission to the ward (due to persistence of partial neuromuscular block) is not clinically detectable in patients not monitored with TOF, while it increases the morbidity and mortality of the patient, frequently requiring prolonged antibiotic therapy and ventilatory support in an intensive environment.

Volatile and intravenous anaesthetics and analgesics effects that persist after surgery (also those used for pain control) contribute to POPC. These drugs help to maintain a modest level of hypnosis with partial abolition of airway protective reflexes. (Fagerlund MJ et al., 2010; Murphy GS et al.,2010; Papazian L et al., 2010)

3.2 Epidemiology and perception of the problem

Several studies (Butterly A. et al. , 2010; Plaud, B et al. , 2010) have documented that neuromuscular block often persists in the post-anaesthetic care unit (PACU), even with the administration of acetylcholinesterase inhibitors. The frequency of this phenomenon, which has been called "residual curarization," "residual neuromuscular block," "postoperative residual curarization," or "residual paralysis," ranges between 4 and 50% depending on the diagnostic criteria, the type of nondepolarizing neuromuscular blocking drug (NMBD) used, the administration of a reversal agent and, to a lesser extent, the use of neuromuscular monitoring.

Murphy and Brull in a 2010 article, report similarly a wide variability of incidence of residual neuromuscular block in literature, with reported frequencies ranging from 2% to 64%. The problem is obviously clinically relevant, because residual paralysis after emergence from anaesthesia is associated with muscle weakness, oxygen desaturation, pulmonary collapse, and acute respiratory failure that could lead to severe permanent brain damage or death.

A recent survey (Naguib et al. 2007) conducted in Europe and the United States shows a general lack of knowledge and consideration of the PORC phenomenon and monitoring of neuromuscular function: the majority of respondents considered the residual curarization to have an incidence lower than 1%; in Europe only one third of the sample, and 10% in the United States, considered it necessary to monitor the neuromuscular block with the train of four (TOF). The decurarization was always completed by only 18% of the sample in Europe and 34% in the United States.

The criteria by which it is more often decided not to decurarizate are the time elapsed between the last dose of neuromuscular blocking and awakening, the total dose administered and subjective clinical parameters indicative of residual weakness. Only 45% of anaesthetists in Europe and 12% in the United States base their decision on the TOF.

Even a recent survey conducted in Italy (Di Marco et al. 2010) shows little understanding of the PORC phenomenon; most anaesthesiologists use clinical trials with subjective interpretation to assess the degree of residual curarization: head raising, protruding of the tongue, hand shaking, opening of the eyes; these are actually examinations which are not very sensitive and specific. When asked about TOF monitoring, only 24% of respondents know that it’s necessary to reach a TOF ratio of at least 0.9 to consider the curarization of a patient to be adequately resolved.

Today, it is necessary to ensure an adequate level of quality and safety in anaesthesia, using objective monitoring of the level of curarization. Among the devices on the market, the TOF ensures maximum reliability, specificity and sensitivity.

Despite extensive documentation of such residual paralysis in the literature, the awarness of its clinical consequences remains surprisingly limited, and the use of NMBDs, neuromuscular monitoring, and reversal agents are dictated more by tradition and local practices than by evidence-based medicine.

In an internet-based survey conducted among European and US anaesthetists, a high percentage of respondents had never observed patients in the postanesthesia care unit with residual neuromuscular weakness after intraoperative administration of a muscle relaxant. Respondents from the US were more likely than their European counterparts to estimate that the incidence of clinically significant postoperative residual neuromuscular weakness was < 1%.

In the light of evidence reported (A.F. Kopman, 2007) it is reasonable to conclude that "few clinicians perceive residual block as an important safety issue", both for the relatively low incidence of this phenomenon in some postsurgical settings, and for the lack of clinical recognition of the phenomenon by the anaesthetist.

Inadequate management of curarization in terms of non TOF monitored onset and offset of curare, is associated as previously mentioned with another typically intraoperative event: inadequate intraoperative neuromuscular blockade, an essential aspect to ensure the success of the surgery involving the curarization of the patient.

Inadequate curarization can result in unexpected and sudden movements of the patient with the risk of laceration of internal organs by the surgeon who's operating, distortion of the limbs which had been fixed, or accidental removal of tools placed by the anaesthesiologist. (Baillard C et al.,2005; Di Marco P et al., 2010; Naguib M et al., 2010; Naguib M et al.,2007)

3.3 Guidelines on PORC and legal issues: State of the art

The issues concerning patient safety are real and pressing. Medicine today faces a growing demand for efficiency, safety and quality.

The increasingly wide circulation of knowledge in the field of health goes hand in hand with an increase in disputes over medical-legal issues. Right or wrong, the media disseminate news about real or presumed cases of medical malpractice almost daily.

Premiums on professional insurance policies are constantly increasing and a few companies have even refused to grant coverage. As complaints and lawsuits continue to pile up, the doctor is no longer seen as the sole undisputed custodian of medical knowledge and place of refuge in case of sickness.

Errors in medicine are possible, but we must try to avoid them. Medical science is not infallible and diseases do not follow mathematical laws, but are subject to the variables in the individual. This line of reasoning, if generally accepted in theory, is easily forgotten when coping with an adverse event that is actually happening, while the search for a guilty party and the quest for indemnity get immediately under way.

Confronted with this situation, in no way can doctors today do without the equipment and procedures necessary to ensure the minimization of risks for the patient.

The need to ensure adequate standards of quality and reproducibility in one's activity has led to a routine use of guidelines, evidence and protocols, which now form the basis of common practice and are the indispensable tools of legal defence against allegations of malpractice.

At the same time there is a growing awareness of the importance of risk management and the need to build default shared paths to prevent and manage the risks arising from medical practice.

In this context, so-called defensive medicine takes hold: in a recent survey conducted by the Order of the Italian Doctors in Rome (2008), more than 80% of doctors considered it a realistic risk to receive a complaint from their patients, whereas only 7% excluded this possibility; 99% of surgeons and 97% of anaesthetists felt intimidated.

The fear of retaliation leads to increased costs due to prescription of drugs, diagnostic tests and specialist advice, in excess of actual needs. Even in the operating room safety is a very significant issue.

In 2008, the Italian Ministry of Labour, Health and Social Policies issued specific recommendations and safety standards, ranging from the correct positioning of the patient, airways management, medication, anaesthesia wake up and documentation.

The anaesthetist is among those who take the greatest risks. The anaesthetist can make mistakes like everybody else, but the consequences can be more serious and expensive.

The most complete data on anaesthesia claims are to be found in the American Society of Anaesthesiologists Closed Claims Project (ASACCP), which has been running effectively for more than 20 years. More recently, Japanese and Danish authors have reported on the scope and costs of anaesthesia-related complications and their medical-legal claims.

In the UK, the National Health Service Litigation Authority (NHSLA) Manages legal claims (both clinical and non-clinical) made against the NHS: a survey done in England on medical-legal disputes recorded from 1995 to 2007 (Cook TM et al. 2009) demonstrates that the majority are related to obstetrics and gynaecology or surgery problems, and 2.5% of the cases is related to anaesthesia. Among these, 11% can be attributed to problems related to drug delivery: needle exchange, incorrect dosage, allergic reactions, adverse events, neuromuscular blockers.

3.4 Neuromuscular blocking agents, key features

3.4.1 Pharmacodynamics and pharmacokinetics

The neuromuscular blocking agents used in anesthesia are also known as muscle relaxants. Through the specific and reversible blockade of the neuromuscular junction they allow to obtain an adequate skeletal muscle relaxation.

The molecules now commercially available belong to two classes according to the mechanism of action: depolarizing agents, among which the only one still sold is succinylcholine and nondepolarizing agent, which include benzylisoquinolinium molecules and steroidal compounds.

3.4.2 Depolarizing muscle relaxants

The depolarizing muscle relaxants (succinylcholine) bind to pre/post synaptic and extra-junctional acetilcholine receptors acting like it, thus depolarizing the terminal plate. Unlike acetylcholine, however, which is metabolized in less than 1 msec, succinylcholine remains bound for a time sufficient to cause desensitization of the receptor and therefore neuromuscular blockade.

The rapid metabolism is by plasma cholinesterase, with a half-life of less than a minute.

The main side effects include bradycardia, anaphylaxis, muscle pain, increased intraocular pressure and intracranial hyperkalaemia.

3.4.3 Non depolarizing muscle relaxants

These competetively bind postsynaptic receptors as antagonists.

There are many molecules on the market, with different characteristics. (Table 1)

The features include the organ-independent elimination of derivatives benzylisoquinolinium, the absence of vagolytic effects, a different release of histamine for different molecules and the substantial lack of accumulation. A distinctive feature of the molecules of this class is the enzymatic degradation that occurs in plasma, through the reaction of Hoffmann, influenced by pH and temperature.

Molecule	ED95 (mcg/kg)	Onset (min)	Time for the recovery of T1 (min)
Atracurium (B)	200	5-6	20-25
Cis-Atracurium (B)	50	5-6	20-25
Vecuronium(S)	80	5-6	20
Rocuronium (S)	300	2-3	20
Mivacurium (B)	80-150	2-3	20
Pancuronium (S)	60	4-5	60

Table 1. Main non depolarizing blocking agents – curares. ED95 = effective dose (ED) that will produce a specified response (neuromuscolar block) in 95 percent of a population. S = aminosteroid; B = benzylisoquinolinium

Atracurium, non-depolarizing benzylisoquinolinium, is metabolized by two non-oxidative mechanisms: esterase hydrolysis (70%) and to a lesser extent by the Hoffmann reaction (30%).

The isomer cis-atracurium, characterized by a greater power of action and a slower onset, is mainly degraded (80%) by the reaction of Hoffmann and causes less histamine release.

The nondepolarizing molecules with aminosteroid structure, metabolized by the liver, include pancuronium, vecuronium and rocuronium. The molecules of this class do not release histamine, depend on liver and kidney for their metabolism and exert a variable degree, specific for each molecule, of muscarinic receptor stimulation.

The clinical profile of the molecules differs significantly in some respects: the vagolyitc effects of pancuronium results in often evident hypertension and tachycardia; vecuronium demonstrates an efficient hepatic metabolism, thus causing fewer problems of accumulation for repeated administration than the others; rocuronium, the newest addition to this family of drugs, is characterized by the lack of production of active metabolites during the metabolic cascade and by a quick onset, which at the dosage of 4ED95, can be compared to succinylcholine.(De Miranda LC et al., 2008)

3.4.4 Use in clinical practice

In choosing the most suitable anesthetic procedure, beyond the specific requirements related to the type of surgery, the characteristics of muscle relaxants anesthesiologists have always sought are the speed of action (onset) and the equally rapid resolution of neuromuscular block (offset).

The clinical use of muscle relaxants, however, is characterized by a considerable variability and a different intensity of response with the same amount of drug used in different patients, this having a multifactorial genesis: determining factors are individual characteristics such as age, race, cardiac output, enzymatic activity, sex, genetic polymorphism, environmental factors, as well as the type of surgery undergone by the patient, the combination of therapies that interfere with the function of the neuromuscular junction (antibiotics, drugs acting on central nervous system) and the pre-existing conditions (electrolyte alterations, obesity, neuromuscular diseases, sepsis). A special mention must be accorded to hypothermia, which alone is able to alter the kinetics, drug distribution and rate of liver degradation of all muscle relaxants.

With regard to obesity, some studies conducted using cis-atracurium and rocuronium have shown a significantly prolonged duration of action of both drugs in patients with severe obesity (BMI> 40) and dose calculated according to the actual body weight, than in obese people with doses calculated on ideal body weight, and in the control group.

Concomitant therapy may modify the action profile of the drug. For example antibiotics like neomycin and streptomycin, and anticonvulsant MAO inhibitors strengthen the action of neuromuscular blockers, while phenytoin and carbamazepine decrease it.

Sympathomimetics reduce the onset of muscle relaxants, but also the duration of action. Halogenated anesthestics enhance the block in proportion to the dose.

Particular attention should be paid to patients suffering from conditions such as multiple sclerosis, myasthenia gravis, Duchenne muscular dystrophy, myotonic dystrophy, spinal cord injury, septic conditions.

All neuromuscular disorders are more or less characterized by the presence of receptors with altered conformation and sensitivity towards neuromuscular blocking agents. When confronted by these anomalies, the anesthesiologist must be particularly careful in preoperative evaluation in order to make the proper choice of anesthetic technique and drugs to be used. A quantitative monitoring of neuromuscular blockade is also mandatory. (Meyhoff CS et al.,2009; Singh H et al., 2009)

4. Neuromuscolar monitoring

During anesthesia we generally induce deep levels of paralysis to facilitate tracheal intubation and surgery.

At the end of surgery, before awakening and extubation, we must obtain a full recovery of neuromuscular and respiratory functions: respiratory depression from residual curarization is a significant cause of mortality and morbidity related to anesthesia.

Why monitor:

1. To improve a patient's safety
2. To analyse onset-time, clinics, duration, recovery index, total duration
3. To better manage medication and NMBAs antagonists
4. To establish the right time to use a reversal
5. To establish the right time to extubate
6. For the continuous infusion of neuromuscular agents, in order to avoid unwanted accumulation
7. To manage particular patients by increasing their safety (myasthenic, Lambert-Eaton syndrome, dystrophic, neuropathy, atypical cholinesterase, burns, etc..)
8. To reduce the incidence of PORC
9. To find the degree of neuromuscolar block and thereby better manage the dosage of the antagonists
10. To collect uniform data on NMBAs and antagonists

(Kopman AF et al., 2010)

4.1 Stimulation modes

There are different ways to stimulate peripheral nerves and thereby muscles (Table 2)

Type	Frequency	Duration	Interval	Repetition	Application
Single twitch	0,1 Hz	0,2 msec	1-10 sec	10-1 sec	Anesthesia induction
Tetanus	50 Hz	5 sec		> 6 min	
TOF	2 Hz	2 sec	10 sec	10 sec	Induction, maintenance, Intubation, Awakening, ICU
PTC	50 Hz	2 sec		> 6 min	Deep block
DBS	50 Hz	40 msec	750 msec	> 6 min	Residual curarization

Table 2. Characteristics of neuromuscular stimuli

Single twitch: single supramaximal stimulus repeated with a frequency between 0.1 and 1 Hz. The intensity of the response is compared with a control obtained before the implementation of the block. It has poor clinical applications because it requires a control value as a starting point reference, before paralyzing the patient, against which to evaluate the muscle response.

Tetanus: Perhaps the most stressful method of stimulating the neuromuscular junction. It is very painful and therefore not suitable for unanaesthetised patients. It has very little place in daily clinical anaesthesia except in the context of post tetanic count. Optimal rate of tetanic stimulation: a frequency of 50Hz is similar to that generated during maximal voluntary effort. 100Hz or 200Hz may be more stressful and thus a more sensitive indicator of smaller degrees of neuromuscular blockade. High frequency stimulation (50Hz or more) results in sustained or tetanic contraction of the muscle during normal neuromuscular transmission, despite decrement in acetylcholine release. During tetanus, progressive depletion of acetylcholine output is balanced by increased synthesis and transfer of transmitter from its mobilization stores. The presence of nondepolarizing muscle relaxants reduces the number of free cholinergic receptors and impairs the mobilization of acetylcholine within the nerve terminal thereby contributing to the fade in the response to tetanic and TOF stimulation. Fade is first noted at 70% receptor occupancy. The tetanic fade ratio at the end of 1 second is comparable to that of T4/T1. It facilitates the neuromuscular response during and after its application, artificially shifting all subsequent neuromuscular events towards normality.

Train of Four: gold standard for the assessment of neuromuscular blockade. The technique records the response of a muscle to an electrical stimulus, quantifying in numerical terms the degree of muscle relaxation.

The transmission of an electrical stimulation through electrodes placed on the skin along the course of peripheral nerves evokes muscle response and acceleration, which can be measured.

The response to stimulation depends on the applied current, the duration, the position of the electrodes, the skin condition (dry/wet), the functionality of the stimulated muscles.

The recovery of neuromuscular function depends on the type of muscle relaxant used, on the duration of administration and on individual variability.

It is performed with the patient asleep because it's particularly painful.

There is a sequence of 4 stimuli at 2 Hz frequence, preceded by the search for a supramaximal stimulus. The 4 stimuli T1-T2-T3-T4 are repeated every 12-15 seconds and the response is measured by acceleromiography: 0 out of 4 responses indicates deep muscle relaxation; 1 out of 4 indicates that the patient is still paralyzed, 2 out of 4 indicate the need for administration of a new dose, 3 out of 4 indicates the possibility of administering the anticholinesterase; 4 out of 4 indicates that we are in an advanced stage of decurarization; then the TOF ratio appears (the percentage ratio between the height of the fourth and the first response) and it is proportional to the degree of recovery. Several studies have shown a good correlation between the TOF-ratio values and clinical signs of decurarizzation: with a TOF-R of 40% the patient is barely able to breathe, with 60% he's able to lift his head and protrude his tongue, with 75% he recovers a valid cough reflex, and only with 90% there is a complete recovery of lung function.

Post-tetanic count: it is used during the deep phase of the block, when there's no response to TOF, to evaluate the time needed before the return of an answer. It depends on the principle of post-tetanic facilitation: we apply a tetanus at 50 Hz for 5 seconds, followed by a pause of 3 seconds and then by 1 Hz stimulation. For each drug, the number of visible post-tetanic contractions is inversely proportional to the time required for the appearance of a response to the TOF.

Double burst stimulation: two short bursts of 50 Hz separated by 750 msec. This is mostly used in visual-tactile evaluations of neuromuscolar blockade.

4.2 Recording modes

Visual and tactile evaluation (for surface muscles): Subjective method and therefore dependent on the individual assessment.

Among the various clinical criteria, there are some proved unreliable (tongue protrusion, eye opening, normal or near normal vital capacity, inspiratory pressure < -25 cm H2O, arm raised toward the opposite shoulder), others not very reliable (head lift test, holding hand for 5 seconds, lifting leg for 5 seconds, maximum inspiratory pressure > - 50 cmH2O), and finally the only one that seems pretty reliable, the tongue depressor test, which seems to correspond to a value of TOF> 0.8 -0.9. (Table 3)

Induction of muscle relaxation
Reduction or disappearance of the respiratory movements
Fasciculations (succinylcholine)
Easy extension of the head
Easy positive pressure mask ventilation
mouth opening
Easy tracheal intubation
Absence of cough during intubation
Maintenance of muscle relaxation
Relaxation of the abdominal muscles
If partial paralysis: maladaptation to mechanical ventilation, diaphragmatic movements, increased peripheral muscle tone, contraction of frontal muscles, eyelid reflex
Recovery of muscle relaxation: Testing of awakening
Ability to lift the eyelids
Ability to protrude the tongue
Ability to cough
Ability to shake hands
Ability to keep an arm raised
Ability to lift head
Vital capacity greater than 15 ml / kg
Inspiratory force of at least -25 cmH20
Tongue-depressor test: ability to squeeze an object between the teeth, resisting the removal

Table 3. Clinical tests used to assess adequacy of reversal for neuromuscolar blockade

Force measurement: a force transducer records the muscular responses to electrical impulses and sends them to a screen. Applicable only to the thumb adductor.

Electromyography: it records the electrical response in the muscle, through the use of properly positioned electrodes.

It can be used at the adductor of the thumb, hypothenar eminence and first dorsal interosseous, innervated by the ulnar nerve.

Accelerometry: it measures the acceleration. From the second law of Newton, we know that the acceleration (a) is proportional to force (F). Provided that the mass (m) remains constant, acceleration becomes a direct measurement of the force of contraction and can be used for the quantification of the neuromuscular block. An acceleration transducer is easily put in place but must move freely for reliable measurement.

Applicable to the adductor of the thumb.

4.3 Muscle choice

In principle, any superficially located peripheral motor nerve may be stimulated, but not all the skeletal muscles show the same sensitivity to neuromuscular blocking agents.

Sites of stimulation most frequently used: ulnar nerve, posterior tibial nerve, common peroneal nerve, facial nerve.

From a practical point of view, no difference in sensitivity exists between the arm (adductor pollicis muscle) and the leg (flexor hallucis brevis muscle).

In some surgical procedures, the patient is positioned in such a way that the arm is not available for monitoring neuromuscular relaxation. The leg is then a good alternative. Another alternative is the stimulation of the facial nerve. Monitoring of the orbicularis oculi muscle can provide also information about the degree of neuromuscular relaxation in those situation where both arm and leg are not available. Not all monitoring equipment is suitable for use with this location.

After the administration we may measure different times as onset, maximum block and recovery. We ideally should measure the function of clinically relevant muscles (abdominal, airways, breathing), but alternatively we can record the activity of muscles with a similar response: for example the eye orbicular is a good indicator of the conditions of intubation, while the thumb adductor is a good indicator of upper airway's muscles.

Adductor of the thumb: accessible during most surgical procedures, innervated by the ulnar nerve. At the induction it has a longer onset than central muscles and recovery is slower than the diaphragm, the rectus abdominis and the adductor of the larynx.

Periorbital muscles: the facial nerve is normally used when the ulnar nerve is not available.

It is located anterior to the ear, where the the electrodes must be put. The transducer must be applied to the orbicularis oculi.

This muscle's function shows a good correlation with the diaphragm. It may overestimate the decurarization, so it's better to perform a calibration to increase its sensitivity.

Muscles of the foot: the posterior tibial nerve can be stimulated behind the inner ankle bone and provokes the flexion of the big toe. The response is comparable to that of the adductor of the thumb.

Since when administering a neuromuscular blocking agent the patient is under general anesthesia, it's not possible to measure actions by asking him to perform volunteer movements. We must therefore stimulate a nerve somehow in order to evoke an automatic and unconscious response; then, in a visual and/or tactile way or, even better, instrumentally, we can detect and then evaluate muscle contractions. The ulnar nerve, whose stimulation causes contraction of the thumb adductor, or the posterior tibial nerve, whose stimulation causes contraction of the flexor allucis brevis, are normally used to monitor curarization and to provide profiles that can be seen to be very similar.

In clinical routine, the ulnar nerve is the most widely used, since it is easily accessible: it runs on the forearm, parallel to the ulnar artery: on stimulation of the ulnar nerve, the adductor muscle contracts. Electrodes are commonly applied to the wrist but alternative locations are possible. The use of a relatively sensitive muscle such as the adductor pollicis has the disadvantage that even after the total elimination of response to TOF or ST, the possibility of movement of the diaphragm, such as coughing or hiccuping, can not be ruled out. Equipment that offers the PTC simulation pattern can of course still monitor the degree of relaxation during intense block. On the positive side, the chance of overdosing the patients during surgery is smaller when the most sensitive muscle is used for monitoring. More important, during recovery, when the adductor pollicis has recovered sufficiently (TOF ratio > 70%), it can safely be assumed that no residual neuromuscular blockade exists in the diaphragm (which recovers faster than the adductor pollicis).

During induction of muscle paralysis, the single stimulus repeated every second (1Hz) is the easiest and most suitable way for observing the exact time of establishment of neuromuscular block, i.e. the onset time. During maintenance of the neuromuscular block it is more convenient to use the TOF, i.e. a train of 4 stimuli (Train of Four). The first one of the 4 stimuli (T1) is sent to the ulnar nerve 13 seconds after the previous 4th TOF stimulation. In addition, the TOF allows the evaluation of a particular phenomenon called "muscle fatigue" or "Fade": when acetylcholine is released in response to depolarization of the motor neuron, it acts on both postsynaptic nicotinic receptors (or post-junctional), and pre-synaptic nicotinic receptors (or pre-junctional). The agonist action of acetylcholine on presynaptic receptors results in a physiological recruitment of presynaptic vesicles such that they can more easily release additional acetylcholine (positive feed-back or presynaptic facilitation, which supports muscle contraction facilitating it). In contrast, in presence of a non-depolarizing curare, these receptors are inefficient to varying degrees: acetylcholine slows down its ready availability and the subsequent release, so that muscle contraction is weakened gradually, also because the muscle fiber is partially blocked by the presence of low or moderate concentrations of curare bound to postsynaptic receptors. The more a curare shows affinity for presynaptic receptors, the more intense the fade effect will be. However, an interval of 10 seconds between one TOF and the next is enough to recover the amount of acetylcholine necessary to produce a new T1, at least of the same height as the previous one. Here is the importance of the TOF, which not only allows one to control the depth of the block (occupancy of postsynaptic receptors) through the first response (T1) but also allows the measurement of the degree of "fade" (occupancy of presynaptic receptors) by TOF ratio (TR), the ratio between the amplitude of the last of the four answers (T4) and the first one (T1). Only when the TR is ≥ 0.9, the recovery from neuromuscular block is sufficient, and the patient can be considered self-sustaining in his respiratory autonomy, at least in muscle tone. Better yet, it's good to aim for full recovery (TR 1.0) before sending the patient to the ward. (Vincenti 2010)

In a 2005 study by Glenn et al. the authors assessed the incidence and severity of residual neuromuscular block at the time of tracheal extubation. Clinicians were instructed to reverse neuromuscular blockade only at a TOF count of 2-3 and to delay tracheal extubation if the TOF ratio was 0.6-0.7, regardless of the clinical evaluation of the patient. In fact, severe residual paresis (TOF 0.7) was noted in 70 Patients (58%) at the time the anesthesia care provider had estimated the block to have been recovered sufficiently enough to exclude residual paralysis. The inability of clinicians in this investigation to detect a residual block in the operating room is not unexpected. A 5-second head lift or hand grip (used by all clinicians in the study) can be maintained in some patients with postoperative TOF ratios as low as 0.25-0.4. In conclusion, the authors determined that significant residual paralysis was present in the majority of patients at the time of tracheal extubation. Despite directing the use of a strict monitoring protocol and reversal of an intermediate-acting muscle relaxant, and the performance of a careful clinical examination for signs of muscle weakness, clinicians were unable to achieve consistently acceptable levels of neuromuscular recovery in the OR. In order for anesthesiologists to be assured that neuromuscular recovery is complete and pharyngeal muscle and respiratory functions have returned to normal, the authors suggest that quantitative neuromuscular monitoring is required. (Murphy et al., 2005; Lee HJ et al.,2009)

5. Reversal of neuromuscular blockade

In order to pharmacologically antagonize the effect of NMBA, we can use specific drugs, which belong to two classes: anticholinesterases and sugammadex. They have completely different pharmacodyanimics and pharmacocynitics.

5.1 Anticholinesterases

The usual practice of anesthesia, as indicated by numerous surveys, contemplates the use of anticholinesterases such as neostigmine, in order to antagonize neuromuscular block. This practice varies in frequency and indications depending on the setting considered, but in recent decades it has become the gold standard treatment for post-operative residual curarization (PORC). Although this practice is called pharmacological antagonism, the actual function of neostigmine is to increase levels of acetylcholine in the synapsis with the aim of displacing curare molecules still there, thus depolarizing the neuromuscular junction. The destiny of curare molecules removed by the junction's receptors is not taken into account. However, in the presence of long-acting curare or of conditions that increase the half-life of administered curare, there is a risk that, once the action of acetylcholine comes to an end, the remnant molecules reoccupy the plaque, thus establishing a neuromuscular late block. In addition, the therapeutic dose of neostigmine is sometimes insufficient to determine an optimal recovery of a neuromuscular function, particularly in the presence of abundant residual curare with a high affinity for pre- or post-synaptic receptors. This condition can lead, in case of awakening and tracheal extubation not monitored by TOF, to residual curarization and real risks of complications. Finally, severe bradycardia and parasympathomimetic effects related to the use of neostigmine may be responsible, if not treated promptly, for an increase in morbidity and mortality of the treated patient.

The recommended dosage to antagonize muscle relaxation due to the curare and curare-like molecules is about 0.5-2 mg, administered by slow intravenous bolus, combined with atropine in order to oppose the muscarinic effects. (Jonsson Fagerlund M et al.,2009)

5.2 Sugammadex

Sugammadex is a new molecule synthesized for the selective reversal of neuromuscular blockade induced by rocuronium or vecuronium, curares belonging to the family of aminosteroids. It is therefore a SBRA: Selective Relaxant Binding Agent.

Chemically it is a gamma-cyclodextrin (suitably modified to improve the encapsulation of aminosteroids) consisting of an inner portion interacting with the lipophilic steroidal structure of rocuronium or vecuronium, and a hydrophilic outer surface that completely dissolves in water. Unlike drugs currently used for the "reversal" of neuromuscular blockade (anticholinesterases such as neostigmine) Sugammadex can antagonize a deep neuromuscular blockade, if administered in appropriate dosage, without having to wait for a partial recovery of neuromuscular function. Sugammadex, which represents a new concept of antagonism, can be called an "injectable receptor. (Vincenti 2010).

Below we list the basic information concerning the recommended doses and safety features of the drug:

- 2 mg/kg in the presence of a spontaneous recovery with the reappearance of T2. To support the use of this dose, 9 dose-finding studies have been conducted: administering 2 mg/kg of sugammadex, the TOF-ratio reaches 0,9 within 2.8 minutes. Doubling the dose, within 2,6.
- 4 mg/kg if the recovery from block induced by rocuronium or vecuronium has reached a value of at least 1-2 as post-tetanic counts (PTC). In literature, the average time to restore a TOF-ratio of 0.9 was 3.2 minutes. Doubling the dose, there is a reduction of 1 minute.
- For immediate reversal of a rocuronium block the dosage is 16 mg/kg. No studies have been conducted with vecuronium.

Respiratory and anesthetic complications, such as the detection of residual blockade or re-emergence of the block, are more frequent with neostigmine, but we need to further confirm these observations. Studies have found hypersensitivity to Sugammadex and a significant prolongation of the QT interval, especially in case of concomitant use of sevoflurane, though without torsades de pointes. The administration must be modified in patients with renal diseases, adjusting the dosage to the actual renal function. The security in dyalazed patients is limited and the use is not recommended in severe renal failure (creatinine clearance < 30 ml/min) (Abrishami et al., 2010; Della Rocca G et al., 2009; Duvaldestin P et al., 2009; Khuenl-Brady KS et al., 2010; Lee C et al., 2009; McDonagh DL et al., 2011; Plaud B et al., 2009; Schaller SJ et al., 2010; Staals LM et al., 2010; White PF et al.,2009)

6. Incident reporting: Perceptions of the problem by operators and development of actions for improvement

The "Risk management in health facilities" program of the Regional Health Agency of Emilia Romagna, Italy, is committed to developing several tools to be used for the identification and analysis of risk in healthcare organizations. Among these, there is one that's particularly useful during risk identification and analysis: the incident reporting system.

The report, done firstly by operators, of significant events (accidents or near misses, which means events that could develop into accidents) is relevant and useful if it is made and inserted into a systematic approach, whose primary aim is to improve patients' and workers' safety in the healthcare facilities.

Incident reporting is the way of collecting reports of adverse events, near misses and accidents in a structured way. It provides a basis for the analysis and preparation of strategies for improvement, aiming to prevent such incidents from happening again in the future. This system, first introduced in the aircraft industry for the voluntary and confidential reporting of events by pilots and air traffic controllers, in order to improve aviation safety, has been imported for some years by the Anglo-Saxon health systems (Australia, Britain, United States) to fit the health organizations, with the aim of improving patients' safety.

The approach is based on the psycho-sociology of organizations and especially on model interpretation of adverse events (accidents with or without damage) or near-events introduced twenty years ago by James Reason and widespread in risk management of organizations that need to ensure high reliability (traffic, space, nuclear, offshore). Based on this approach, already widely known in healthcare organizations, improving the safety of the environment is possible if we admit the possibility that "something can go wrong" and if we use the information derived from the analysis of events to develop corrective actions or improvements.

To achieve this it is necessary, however, to guarantee a friendly and protective context, because what we want to achieve is the development of a widespread culture of risk, based on the voluntary membership of the operators.

The Italian program has been developed referring to foreign operational practices, that have been adapted to the local context. Among these, incident reporting is one of the available instruments, and it is particularly useful during the process of risk management, identification and analysis. It is not the only tool used, but international analyses show that it is particularly useful for long-term factors, to study rare events, or to identify the most frequent causes. Its utility is also supported by its low costs.

In general we can say that the establishment of reporting systems has two main functions:

1. They provide a measure of the reliability of organizations observed (external role)
2. They provide useful information to those who work for the improvement of the organization, particularly in security aspects (internal role)

When the main purpose is to provide reliability, these systems are mandatory by law or specific rules, with specific authorities appointed to investigate and evaluate cases and if required, to impose sanctions. The focus of these systems is on particularly relevant events, usually with serious results, up to death.

The achievable goals can be summarized as follows:

1. to provide users with a minimum level of protection, ensuring that the most serious cases are reported and investigated, and that follow-up actions are taken;
2. to provide healthcare organizations with incentives to improve patient safety, in order to avoid potential sanctions and bad name resulting from negative situations;
3. to enhance the awareness of health care organizations as to developing forms of activity in the field of security.

However, events that have serious consquences are only a small part of all events that occurr daily, since errors and accidents are typically intercepted before they lead to consequences, or the consequences are not critical.

The reporting systems focused on risk management improvement are therefore directed to broaden the traditional field of observation, also considering the events that do not result in damages (near miss) or in minimal ones. The main purpose of such an approach is to better understand the organization, in particular to identify system defects that favor the occurrence of events before they do come about with potentially damaging consequences.

These reporting systems are tipically voluntary, they manage confidential information and don't entail penalties and punishment. The characteristic strengths of this approach can be summarized as follows:

1. to enable the identification of types of events that occur infrequently and therefore can hardly be ascertained by a single structure;
2. to enable the correlation of events so as to identify issues that run throughout the organization;
3. to read a single event in a systemic manner, so that it is no longer perceived as random, but placed against large-scale trends which can be interpreted;
4. to identify unusual or emerging events, which are reported because they are perceived as unusual;
5. to be able to react quickly to situations, because the alert is normally simultaneous with the event.

Even in this case, the people in charge of collecting and analyzing reports may be internal, allowing only consideration about what is happening in a specific organization, or external to the health facilities, which affords a greater quantity of data that may allow one to take advantage of all the strengths of this type of system.

The salient features of an effective system of voluntary reporting are:

1. confidentiality and an absence of punitive behavior, to be pursued as follows:

- creating parallel systems: responsibilities must relate to required systems, that co-exist with so-called "quality" systems, internal to companies, whose sole purpose is to create information aimed at improving the company's safety;
- making the reported events anonymous, through the elimination of all elements of recognition. This process should be done during the analysis of events, since the possibility of reconstructing what happened is crucial for their understanding and for the analysis of related issues.

2. giving priority to reports about situations of near misses rather than those that caused damage;

3. to answer to the information received through the adoption of measures consistent with the findings from reports, letting people know that:

- reports were received;
- reports were elaborated;
- reports were used to design interventions to solve problems that may arise

In summary, the higher values of an incident reporting system are:

- to provide an empirical basis of reference for the design and adoption of corrective actions / improvements;
- to create awareness among the operators on safety issues

The adoption of an incident reporting system provides the operators (those who report) the capability of recognizing the events that should be reported. A moment of initial training and retention of capacity (together with the "restitution of the information", as discussed above) are therefore necessary to create a first basis of attention to safety issues.

The commitment of the direction that establishes and maintains the incident reporting system, is a sign of importance given to the issue.

Finally, the reporting is done by the operator and it thus highlights both his responsibility of identifying events, and his attention to behaviors. It is clear, however, that:

- if the operators ask for activities and initiatives to increase competence and improve risk management skills (regardless of the initial condition), the incident reporting system can help through a work of knowledge, understanding and improvement of security;
- if the operators are not interested in learning more about topics like adverse events, risk, safety, the incident reporting is unable to provide satisfactory results, or - by itself - to create improvements.

Dealing with “negative" (potential or actual incidents) events, in fact, it is not stimulating in a context with lack of tranquillity and will to improve.

To date, the Emilia Romagna risk management system, hasn’t recorded any incident reporting about PORC, this confirming the underestimation and lack of knowledge of the problem.

7. Closure and future work

Both the literature reviewed and our experience suggest the urgent need to formulate appropriate Guidelines and to prepare a Risk Management program for the patient undergoing curarization.

- The evidences discussed above suggest first of all to read up properly on pharmacokinetics and pharmacodynamics of curares that will be used, in order to optimize the dosage and timing of administration according to the patient and the surgical procedure that must be applied.
- It looks mandatory, basing on the evidences, to monitor neuromuscular function using the TOF in order to objectify and document the degree of muscle relaxation achieved (essential also for medico-legal disputes) and to maintain such monitoring up to extubation, in order to reduce the incidence of PORC.

At the end of the intervention it's prudent and advisable to obtain and evaluate both the TOF ratio (TR) and the T1 or single twitch.

- The TOF ratio is useful to assess the extent of the residual "fade", that is suggestive of pre-junctional receptor occupancy; the value of T1 or a "single twitch" should be compared to the baseline value obtained with the patient not yet paralyzed, in order to assess the depth of residual blockade, that is suggestive of the post-junctional receptor occupancy. The monitoring of T1 or the single twitch becomes particularly important in case of PORC due to accidental administration of succinylcholine in patients with plasma cholinesterase deficiency or suffering from Myasthenia Gravis, with unexpected prolongation of neuromuscular blockade. Succinylcholine doesn't provoke "fade", unlike non-depolarizing neuromuscular blocking molecules, and the T1/T4 ratio is therefore normal, while there's a uniform reduction in all twitch (T1: 50%, T2: 50%, T3: 50 %, T4: 50%) that can be detected by monitoring the T1 or a single twitch, that must be compared to the value recorded pre-curarization.
- The preventive use of neostigmine in all patients receiving neuromuscular blockade monitored with TOF does not appear reasonable in light of the possible side effects induced by the drug. It seems rather reasonable to antagonize neuromuscular persistent block (PORC) in a patient extubated with TR> 0.9-1.0. The pharmacological considerations set out above would favor the preferential use of sugammadex in the absence of severe impairment of renal function or pre-existing QT lengthening. However the high cost of supply of the drug make it unavailable in many settings with limited budgets. In the presence of PORC in patients with neuromuscular diseases, sepsis, hypothermia, or morbid obesity, unpredictable and predisposing conditions for a significant prolongation of neuromuscular blockade, it seems reasonable to make a reversal of the block in the first instance directly with Sugammadex, in order to achieve a real antagonism and to prevent the recurrence of a late block. In the absence of these comorbidities, in order to optimize the cost-effectiveness, it seems reasonable to use neostigmine and atropine in the first instance, in the presence of clinical and instrumental partial recovery from the block, and to restrict the use of sugammadex to PORC with a deep block, to blocks not regressed after Prostigmine at therapeutic doses or in patients with unexpected difficult intubation paralyzed with rocuronium or vecuronium.
- The "Incident Reporting" is an usual practice of a hospital that aims to record the operational incidents that occur within it, to analyze it carefully, and to bring back the annual incidence and trends over time discussing it within business meetings with the aim to formulate corrective actions and improvement. Such events would promote a greater awareness of the problem by the professionals involved and would objectify the real incidence of the problem in the structure in which they operate. We suggest to adopt this practice in any hospital setting and to integrate it within a path of "Risk Management" for the prevention of the PORC.
- In case of respiratory complications and/or heart disease attributable to rapidly regressed PORC, it seems reasonable to extend the observation of the patient in a "Recovery Room" or to consider a transfer to intensive environment, where his vital signs can be monitored and he can be assisted, with particolar attention to neuromuscular function, which should be controlled by TOF until complete regression of neuromuscular blockade.

Schematic risk management algorithm:

1. Choose the most suitable curare in relation to the patient and the surgery
2. TOF monitoring? Always recommended!!
3. Intrastigmine or Sugammadex?
4. ICU or "Recovery Room" if PORC
5. Fill in the "Incident Reporting" Card in the event of PORC

Referring to the comments made in this chapter, we consider it essential to structure a proper management path for PORC and POPC. We hope that the awareness of the problem induced by this work and the "Risk Management" tools will help to reduce the epidemiology of PORC and POPC.

We recognize the limitations of our work, in terms of databases explored and search strategy used. We aim to expand in the future to other bibliographic databases and periodically update the location of "Risk Management" on the basis of present knowledge, to ensure the patient undergoing anesthesia and curarization with the most correct and updated medical management, and to allow the nurse's and physicians' team to reach high performance and safety levels. (Brull SJ, et al., 2010)

8. Acknowledgments

We thank the 'Regional Health Agency of Emilia Romagna ASR-RER) for sharing the material relating to Incident Reporting and for helping to implement this tool within the University Hospital S. Orsola-Malpighi, Bologna.

We thank the nursing and anesthesia team of the Prof. Di Nino Operating Room and the Polyvalent Intensive Care Unit of S.Orsola-Malpighi hospital for their interest in the topic and for the support they gave us in the realization of a model of "Risk Management" of the patient under neuromuscular blockade.

9. References

Abrishami A, Ho J, Wong J, Yin L, Chung F. *Cochrane corner: sugammadex, a selective reversal medication for preventing postoperative residual neuromuscular blockade.* Anesth Analg. 2010 Apr 1;110(4):1239. PubMed PMID: 20357160.

Baillard C, Clec'h C, Catineau J, Salhi F, Gehan G, Cupa M, Samama CM. *Postoperative residual neuromuscular block: a survey of management.* Br J Anaesth. 2005 Nov;95(5):622-6. Epub 2005 Sep 23. PubMed PMID: 16183681.

Brull SJ, Murphy GS. *Residual neuromuscular block: lessons unlearned. Part II: methods to reduce the risk of residual weakness.* Anesth Analg. 2010 Jul;111(1):129-40. Epub 2010 May 4. Review. PubMed PMID: 20442261.

A. Butterly , E. A. Bittner , E. George , W. S. Sandberg , M. Eikermann and U. Schmidt . *Postoperative residual curarization from intermediate-acting neuromuscular blocking agents delays recovery room discharge.* British Journal of Anaesthesia 105 (3): 304–9 (2010) Advance Access publication 24 June 2010 . doi:10.1093/bja/aeq157

Claudius C, Garvey LH, Viby-Mogensen J. *The undesirable effects of neuromuscular blocking drugs.* Anesthesia 2009. Mar;64 Suppl 1:10-21. Review. PMID:19222427.

Cook TM, Bland L, Mihai R, Scott S. *Litigation related to anesthesia: an analysis of claims against the NHS in England 1995-2007.* Anesthesia 2009 Jul; 64(7):706-18. Erratum in: Anesthesia. 2009 Sep:64(9):1039-40.

Della Rocca G, Pompei L,. *A novel approach to reversal of neuromuscular blockade.* Minerva Anestesiol. 2009 May;75(5):349-51. Revview. PubMed PMID: 19412157.

De Miranda LC, Barrucand L, Costa J, Ver√ßosa N. *A comparative study between one and two effective doses (ED95) of rocuronium for tracheal intubation*. Rev Bras Anestesiol. 2008 May-Jun;58(3):202-9. English, Portuguese. PubMed PMID: 19378515.

Di Marco P, Della Rocca G, Iannuccelli F, Pompei L, Reale C, Pietropaoli P. *Knowledge of residual curarization: an Italian survey.* Acta Anaesthesiol Scand. 2010 Mar;54(3):307-12. Epub 2009 Oct 15. PubMed PMID: 19839947.

Duvaldestin P, Kuizenga K, Saldien V, Claudius C, Servin F, Klein J, Debaene B, Heeringa M. *A randomized, dose-response study of sugammadex given for the reversal of deep rocuronium- or vecuronium-induced neuromuscular blockade under sevoflurane anesthesia*. Anesth Analg. 2010 Jan 1;110(1):74-82. Epub 2009 Nov 21. PubMed PMID: 19933538.

Fagerlund MJ, Ebberyd A, Schulte G, Mkrtchian S, Eriksson LI. *The human carotid body: expression of oxygen sensing and signaling genes of relevance for anesthesia.* Anesthesiology. 2010 Dec;113(6):1270-9. PubMed PMID: 20980909.

Jonsson Fagerlund M, Dabrowski M, Eriksson LI. *Pharmacological characteristics of the inhibition of nondepolarizing neuromuscular blocking agents at human adult muscle nicotinic acetylcholine receptor.* Anesthesiology. 2009 Jun;110(6):1244-52. PubMed PMID: 19417616.

Khuenl-Brady KS, Wattwil M, Vanacker BF, Lora-Tamayo JI, Rietbergen H, Alvarez-G√≥mez JA. *Sugammadex provides faster reversal of vecuronium-induced neuromuscular blockade compared with neostigmine: a multicenter, randomized, controlled trial*. Anesth Analg. 2010 Jan 1;110(1):64-73. Epub 2009 Aug 27. PubMed PMID: 19713265.

Kopman AF. *Managing neuromuscular block: where are the guidelines?* Anesth Analg. 2010 Jul;111(1):9-10. PubMed PMID: 20576960.

Kopman AF, Lien CA, Naguib M. *Determining the potency of neuromuscular blockers: are traditional methods flawed*? Br J Anaesth. 2010 Jun;104(6):705-10.Epub 2010 Apr 28. PubMed PMID: 20430764.

Lee C, Jahr JS, Candiotti KA, Warriner B, Zornow MH, Naguib M. *Reversal of profound neuromuscular block by sugammadex administered three minutes after rocuronium: a comparison with spontaneous recovery from succinylcholine.* Anesthesiology. 2009 May;110(5):1020-5. PubMed PMID: 19387176.

Lee HJ, Kim KS, Jeong JS, Cheong MA, Shim JC. *Comparison of the adductor pollicis, orbicularis oculi, and corrugator supercilii as indicators of adequacy of muscle relaxation for tracheal intubation.* Br J Anaesth. 2009 Jun;102(6):869-74. Epub 2009 Apr 17. PubMed PMID: 19376787

McDonagh DL, Benedict PE, Kovac AL, Drover DR, Brister NW, Morte JB, Monk TG. *Efficacy, safety, and pharmacokinetics of sugammadex for the reversal of rocuronium-induced neuromuscular blockade in elderly patients.* Anesthesiology 2011 Feb;114(2):318-29. PubMed PMID: 21239968.

Meyhoff CS, Lund J, Jenstrup MT, Claudius C, Sorensen AM, Viby-Mogensen J, Rasmussen LS. *Should dosing of rocuronium in obese patients be based on ideal or corrected body weight*? Anesth Analg. 2009 Sep;109(3):787-92. PubMed PMID: 19690247.

Ministero del Lavoro, della Salute e delle Politiche Sociali. Dipartimento Della Qualità Direzione Generale Della Programmazione Sanitaria, Dei Livelli Di Assistenza E Dei Principi Etici Di Sistema Ufficio Iii, *Manuale per la Sicurezza in sala operatoria: Raccomandazioni e Checklist. Ottobre 2009*

Murphy GS. *Residual neuromuscolar blockade: incidence, assessment, and relevance in the postoperative period*. Minerva Anesthesiol. 2006 Mar;72(3):97-109. Review. PMID:16493386.

Murphy GS, Brull SJ. *Residual neuromuscular block: lessons unlearned. Part I: definitions, incidence, and adverse physiologic effects of residual neuromuscular block.* Anesth Analg. 2010 Jul;111(1):120-8. Epub 2010 May 4. Review. PubMed PMID: 20442260.

Murphy GS, Szokol JW, Marymont JH, Greenberg SB, Avram MJ, Vender JS, Nisman M. *Intraoperative acceleromyographic monitoring reduces teh risk of resiudal neuromuscolar blockade and adverse respiratory events in the postanesthesia care unit.* Anesthesiology 2008. Sep;109(3):389-98. PMID:18719436.

Naguib M, Kopman AF, Lien CA, Hunter JM, Lopez A, Brull SJ. *A survey of current management of neuromuscular block in the United States and Europe*. Anesth Analg. 2010 Jul;111(1):110-9. Epub 2009 Nov 12. PubMed PMID: 19910616.

Naguib M, Kopman AF, Ensor JE. *Neuromuscular monitoring and postoperative residual curarisation: a meta-analysis*. Br J Anaesth. 2007 Mar;98(3):302-16. Review. PubMed PMID: 17307778.

Papazian L, Forel JM, Gacouin A, Penot-Ragon C, Perrin G, Loundou A, Jaber S, Arnal JM, Perez D, Seghboyan JM, Constantin JM, Courant P, Lefrant JY, Gu√©rin C, Prat G, Morange S, Roch A; ACURASYS Study Investigators. *Neuromuscular blockers in early acute respiratory distress syndrome.* N Engl J Med. 2010 Sep 16;363(12):1107-16. PubMed PMID: 20843245.

Plaud B, Debaene, B, Donati F, Marty, J. *Residual Paralysis after Emergence from Anestesia.* Anesthesiology 2010; 112:1013–22

Plaud B, Meretoja O, Hofmockel R, Raft J, Stoddart PA, van Kuijk JH, Hermens Y, Mirakhur RK. *Reversal of rocuronium-induced neuromuscular blockade with sugammadex in pediatric and adult surgical patients*. Anesthesiology. 2009 Feb;110(2):284-94. PubMed PMID: 19194156.

Schaller SJ, Fink H, Ulm K, Blobner M. *Sugammadex and neostigmine dose-finding study for reversal of shallow residual neuromuscular block*. Anesthesiology. 2010 Nov;113(5):1054-60. PubMed PMID: 20885293.

Singh H, Tewari A, Bansal A, Garg S, Gupta M, Adlakha P. *Anaesthesia for a patient with Isaac's syndrome and myasthenia gravis.* Br J Anaesth. 2009 Sep;103(3):460-1. PubMed PMID: 19679588

Sungur Ulke Z, Senturk M. *Mivacurium in patients with myasthenia gravis undergoing video-assisted thoracoscopic thymectomy*. Br J Anaesth. 2009 Aug;103(2):310-1. PubMed PMID: 19596766.

Staals LM, Snoeck MM, Driessen JJ, van Hamersvelt HW, Flockton EA, van den Heuvel MW, Hunter JM. *Reduced clearance of rocuronium and sugammadex in patients with severe to end-stage renal failure: a pharmacokinetic study*. Br J Anaesth. 2010 Jan;104(1):31-9. PubMed PMID: 20007792.

White PF, Tufanogullari B, Sacan O, Pavlin EG, Viegas OJ, Minkowitz HS, Hudson ME. *The effect of residual neuromuscular blockade on the speed of reversal with sugammadex*. Anesth Analg. 2009 Mar;108(3):846-51. PubMed PMID: 19224792.

7

Risk Assessment On-Scene

Eivind L. Rake
Fire Department of South Rogaland/
Stavanger University Hospital, RAKOS
Norway

1. Introduction

The modern community is complex with tight couplings between infrastructures, systems and geographical areas (Perrow, 1999). Accidents and disasters are becoming increasingly global, human made and less observable (nuclear, chemical, and biological etc.)(Beck, 1992). New kinds of crises (Roshental *et al*, 2001) are always upcoming, e.g. Mad Cow disease, viral pandemics, Tsunami and new forms of terrorist attacks. A typical example of the latter is the killing of 69 youths at Utøya, Norway, July 22, 2011. New type of crises, unprecedented, will be seen in the future and challenge the abilities of the societies and communities to cope successfully. Crises rarely correspond with the jurisdictional boundaries of organisation or government (Boin *et al*, 2005) or boundaries of countries. These crises demand considerations other than the preparations common to well – known crises.

Disasters can be described as the ultimate test of plans, preparedness, the emergency management and emergency response capability of a society. The ability to effectively deal with disasters is becoming more relevant because of factors that tend to increase risk and an increased attention and demand from society. The effort to build defences against unconventional threats has not kept pace with the rapid rate of development of new kind of crises. The need for better ways to deal with the potential for catastrophic loss inherent in emergencies and rescue operations has been widely recognised and accepted by government, industry and response units, especially in the aftermath of 9/11, the Tsunami and during the rise of terrorism as witnessed in the bombings in Madrid and London in 2004 and 2005.

To a certain extent we can reduce the numbers of crises that hit, even if we cannot nullify them. Despite the best efforts of society crises will occur and have to be dealt with. The post – event actions, as emergency responses, seek quick and efficient ways to minimize impacts when an accident occurs. The incident, especially in dynamic complex situations, may escalate to a major emergency and even disaster if not handled correctly. The possibility of severe detrimental effect during emergencies is closely tied to the authorities' and response units' opportunity, ability and modus of management, which in turn can act as constraints on subsequent decisions and coping. However, we *can* prepare. If we take the time to make the right preparation now, we may be able to reduce the unwanted consequences a disaster can wreak. Standard procedures and prepared plans act sufficiently in predictable, well – known and routine accidents. The demands of a crisis tend to make specific detailed emergency plans of limited use. A conclusion from the Swedish Tsunami Report (SOUS, 2005) describes the incompleteness of plans;

- A crisis will occur that is new and not predicted and detailed emergency plans don't exist.
- A plan will never give us instructions for every situation that occurs during an unfolding emergency.

New problems must be faced through openness, cooperation and flexibility. The bomb attack in Oslo, July 22, and the following mass execution at Utøya, surprised the response units. They had no sufficient plans but had to deal with the attacks and the following massive need of emergency assistance. Risk management was a vital part of the crises management.

This chapter describe the assessments on-scene, the arena where the crisis take place, especially assessment carried out by incident commanders and other professional leaders of emergency response units; the police, paramedics and fire brigade. The chapter intend to give insight of how risk assessment on-scene is coped with and how effective risk assessment can be carried out in real time while the crisis unfolds on-scene. Initially the command system, the commanders' tasks, and the inherent uncertainty on-scene an accident/crisis is described. It is followed by a description of how decision making on-scene normally is carried out. The challenges of decision making and some basic principles of effective decision making are given. To reduce uncertainty it is important to make satisfying decisions and satisfying risk assessments. A review of risk and risk assessment on-scene ends with a proposal of successful risk management and risk assessment on-scene. By being prepared for unique and sudden scenarios, including the vast number of variables involved, and unprecedented emergencies, we can reduce the uncertainty and thereby the extent of the damage, and increase the probability of a successful crisis management. Proper risk management is the core.

The reminder of the chapter is organised into 5 sections, starting with Section 2 that provides an overview of crises and crisis management. Then, in Section 3 we describe the incident command system on-scene and the need of an incident commander and his responsibility. The following Section 4 discusses uncertainty, a major obstacle to proper decision making, which is presented in Section 5. This section explains some principles of decision making and how effective decision making can be conducted by the incident commander. In Section 6 systematic ways of dealing with risk are pointed out and ideas of successful risk management are presented. Finally, in the last Section 7, we highlight the need of adequate situational awareness and extensive training of risk assessment and do more research to cope successfully with the risks in future incidents and crisis.

2. Crises and crisis management

Boin *et al.* (2005) defines crisis in terms of a discontinuity which usually causes authorities to engage decision making under conditions of uncertainty and time pressure. According to (Rosenthal *et al.* 2001) a crisis can be understood as a period with increasing stress, disturbing society and threatening values and structures in unexpected and unthinkable ways. Crisis management must therefore deal with present risks and avoid risks that can, or will arise. In the experience of the author, crisis management is the continuous process by which all those involved, from an incident commander to groups such as an incident response team and even entire communities, manage hazards in an effort to avoid or ameliorate the impact of disasters resulting from these hazards. The management must also cope with non-routine phenomena and developments during emergencies. We use the word "crisis", but other words are often used to describe unexpected and unintentional occurrences resulting in an immediate threat to human life, or serious damage to property or environment.

Despite the best efforts of society, crises will occur and have to be dealt with. Post-event actions, such as emergency responses, seek quick and efficient ways to minimise impacts when a crisis occurs. An emergency response consists of immediate, time-sensitive actions to be taken during and after the impact to reduce casualties and damage and to respond immediately to the victims to avoid any threatening situations. Quick, appropriate and sufficient relief efforts are typical activity. Response measures include identifying and disseminating the threats and the impact, alerting the responders, searching for and rescuing any trapped victims and providing the necessary care. The response phase includes mobilisation of the necessary emergency services and first responders. The police, the ambulance service and the fire department are typical first responders.

The terror attack in Oslo, Norway, including a bomb explosion and the massacre at Utøya killing 77 persons, pinpoints what crises are about and what we are concerned with: threats (unknown and unforeseen), and undesirable outcomes (injuries, fatalities, depression, political changes etc.). A crisis is an unexpected event that threatens values, such as health, environment or society in general, possibly resulting in undesired outcomes, e.g. causing death. Typically, it is the moment at which a threat is transformed into actual fatalities or other substantial loss. The people affected expect the local authority, and if the local authority fails, the government to avert the threat or at least minimise the damage of the crisis at hand. The authority is normally represented by the emergency services. Crises can be described as the ultimate test of plans, preparedness and the management and emergency response capability of a society. The ability to deal effectively with crises is becoming increasingly relevant because of factors now tending to exacerbate risk and the increased focus on these, with demands for urgent action, especially from the media, the politicians and the population at large. The statement of (Boin *et al.* 2005) "*Crisis management bears directly upon lives of citizens and the well-being of societies*" (p.1), emphasises why crises have to be coped with effectively. Actions by the authorities, the response units, involved persons and organisations need to result in mitigation and success.

For many years, rescue operations have been organised in accordance with strictly hierarchical management structures. There seem to be only minor differences between emergency response units within and between countries with respect to formalised routines. A typical management structure has manuals describing the organisation, leadership and the responsibilities of each of the emergency services at a major incident. LESLP (2008) and (Bigley and Roberts 2001) include manuals and procedures from different management structures. These manuals and procedures intended to gather, coordinate and control the temporary systems of managing personnel and equipment at a wide range of emergencies. The procedures describe the management system and responsibilities and set out the tasks and duties of the commander of the operation, the incident commander.

Crises management is the shorthand phrase for management and coping with non – routine phenomena and development during emergencies. Emergency management on accident scenes is complicated. The consequences may be severe for many people, implying competing and frequently ill-defined goals for the rescue operation. Uncertainties, both in situation assessment and outcome predictions are large, that is; data is missing; information is fragmented and unreliable; the mass of non-relevant data interrupts the focus; and occasionally lack of expertise in vital areas, and lack of resources are problematic. The situations and the risks are constantly changing, with the potential of sudden dramatic occurrences that require an entire rethinking of the rescue operation. There are multiple individuals involved in the emergency organisations and the teams involved are not static; they change from incident to incident. The work domain changes for each emergency situation. Each scenario and the inherent risk are unique and will only in broad general features be known to the combating actors. Crises management and the following operational command and decision making is in general complex, due to the causes above and the high number of feasible courses of action and their implicit representations. Distributed decision making is even more complex, and in a commanding situation, and with multiple actors, even more demanding and exacting.

3. Incident commanding

A crisis calls for management and leadership. On-scene at the emergency, the incident commander is the predetermined manager and leader. He is ultimately responsible for all activities that take place at the incident ground (Bigley and Roberts, 2001). His responders and superiors expect the incident commander to have command and control. His aim is to reduce uncertainty, provide an authoritative account of what is going on, why it is happening, and what has to be done to minimise risks and the following impact. Leadership can be seen as the interaction between the leader and the leadership situation. Fiedler (1996) claims that the most important leadership lesson to be learned so far is that the leadership of groups is a highly complex interaction between an individual and the social and task environment.

The on-scene commanding structure and the incident commander (IC) in particular play an important role in fighting emerging crises. Formal leaders carry a special responsibility for making sure that the tasks of leadership are properly addressed and executed (Boin *et al.*, 2005). In general, the leader affects responders' performances and the outcome. Fiedler (1996) states that how well the leader's particular style, abilities, and background, i.e. experience, contribute to performance is largely contingent on the control and influence the leadership the situation provides for.

A classic example is the Piper Alpha disaster in 1988 (Flin, 2001). A major explosion on an oil and gas production platform resulted in the loss of 165 crewmembers. Lord Cullen concluded, in his public inquiry report (Cullen, 1990), that the number of fatalities was substantially greater than it would have been if the offshore installation manager (the IC on the platform) had taken initiatives to save life. The IC failed, and demonstrated inadequate leadership during the crisis. According to (Fredholm 2006) every response operation needs leadership. The more complex situation and less routine, the more need for coordination, strategic planning, prioritising and decisions to cope with the problems and risks at hand. He calls for distinct and explicit leadership in demanding situations.

The IC's responsibility is to be the commanding officer and have overall management on scene. Overall management includes determining incident objectives and strategy, setting immediate priorities and assigning subsequent priorities, working out an action plan, approving requests for additional resources or for the release of resources, informing agencies and organisations of the incident status and demobilising when appropriate. The IC must establish an appropriate organisation and coordinate the activities for all emergency units. Figure 1 shows an example of such an organisation.

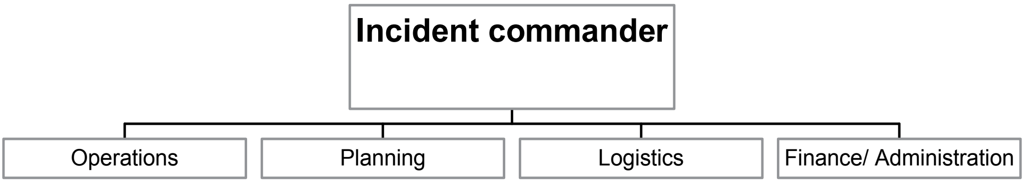


Fig. 1. US Incident Command System (ICS): Basic Functional Structure (FEMA, 2012)

The ICS structure establishes five functional areas: (1) Command, (2) Operations, (3) Planning, (4) Logistics, and (5) Finance/administration for management of all major incidents.

For the purpose of this chapter we will focus only on the Incident commander and the Operations. See (FEMA, 2006) for information about the other 3 areas. Operations (2) manage tactical operations to implement the overall strategic plan. They are responsible for all activities on-scene and run the operational tasks such as life-saving, reducing risks from immediate hazards etc. Operations deals also with the risk and risk assessments on-scene. The IC must ensure the overall safety of the rescuers, the response units, the victims, and any threatened residents or inhabitants. This is a part of the risk management on-scene. The IC normally acts at tactical level and his subordinates, such as the medical officer, at operational level. On the operational level the IC is at the sharp end of the action, located at a command post, and directs the team performing the orders, the decided tasks. The IC's role at tactical level is to implement the plans and achieve the objectives set by the strategic level. The tactical level prioritises plans and coordinates actions on the operational level. The strategic level allocates resources and supports the tactical level just as the tactical level supports the operational level. At major responses and crises the strategic and tactical command is generally located away from the scene.

The major aim of a standardised and hierarchal command structure is to have an effective and predictable command system: a functional system well known to all the responders. Command and control (C2) is the shorthand phrase for the aim of incident management. Leadership is both a position and a process involving collaboration, teamwork, and cooperation. Leadership on-scene an accident can be described, as (Boin *et al.* 2005) do, as a set of strategic tasks that encompass the activities associated with the scenes/stages of management. The leadership function seems pivotal to coping and vital to how the incidents evolve and the risk is managed.

The structure of on-scene command is an "all hazard – all risk" approach to managing crisis response operations as well as non-crisis events. The commander is responsible for coping with all threats on-scene and minimising the risk and consequences. The structure in every western country and all emergency responses are broadly similar. The structure shown in

Figure 1 offers a good example of an effective system. A similar structure is used in Great Britain (bronze, silver and gold) and in the Scandinavian countries.

Flin and Arbuthnot (2002) provide 6 cases described by experienced incident commanders. They represent a variety of disciplines and study commanders from the police, the fire department, the Royal Navy and the Marines, on a passenger airline and from a prison. The commanders reflect (retrospectively) upon extraordinary episodes in which their command skills were tested in a variety of ways and presented as experience – based and the hands-on knowledge of skilled practitioners. These cases are narrative descriptions of past serious incidents. The commanders draw lessons for the future, either rule of thumb or as important lessons. Common conclusions were:

- The importance of being prepared (training, planning, understanding of techniques and staging, learning from earlier occurrences and knowing your team)
- Key personal activities are decision making, communication, information, comprehension of the situation and risk assessment

Flin and Arbuthnot (2002) reflects on issues concerning the training of commanders, identifying some of the key capabilities and skills, such as leadership and team coordination, stress management, situation (risk) management, and decision making. McLennan *et al.* (2003) studied effective incident management and team management during wildfires. They concluded that the incident commander, with his four staff functions, see Figure 1, must develop a common operational picture of the situation by concentrating on the threats and resources.

On call-out, the incident commanders construct their mental maps of the situations from a combination of information from the call centre, knowledge of standard procedures, their expectation of available resources, and personal knowledge of the site (Rake and Njå, 2009). Typical management strategies when reaching the scene involves incremental problem solving within narrow time horizons. The incident commanders pay attention to details rather than considering the overall situation. In general the commanders expect normal situations, i.e. typical accidents they were trained to cope with, and in responses to which their preparations and strategies were standard. A study of incident commanders in real time shows that these incident command strategies are more reactive than proactive, and that the commanders rarely command. Risk management is normally limited to evaluations of the response units' safety (*ibid.*). A risk approach, to be used by crisis managers, incident managers and in situation with important values at stake and different kind of uncertainty, is recommended.

4. Uncertainty

Uncertainty is defined as "*lack of knowledge about the performance of a system (the "world"), and observable quantities in particular*" (Aven, 2003, p. 178). This definition is supported by NATO (2002), describing uncertainty in situations needing command and control. NATO generally defines uncertainty as an "*inability to determine a variable value or system (or nature) or to predict its future evolution*" (p. 249). In the action context, such as on-scene a crisis, (Lipshitz and Strauss 1997) describe uncertainty as a sense of doubt that blocks or delays action. This description goes to the core of on-scene risk management: an action is a result of a decision and a decision is based upon the information and the following assessment.

A crisis, a disaster or an accident, minor or major, can be described as borderless threats, creeping and acute, contending reality claims (uncertainty/surprise), conflicting, they are unplanned, unscheduled, unprecedented and unpleasant to the victims and almost unmanageable events (Rosenthal *et al*. 1989, 2001). Three characteristics must be foremost, both during the preparation and the responses; *1) important values at stake, 2) limited time to deal with the situation, and 3) a great deal of uncertainty involved*. The need for prompt action, from various response units to handle the different tasks in an effective way is obvious. Actors in rescue operation operate within a short time horizon most of the time. They have limited and incomplete information to manage. Thus, it can be suggested that crisis management, preparations and plans, must include the terms: *severe threats*, *uncertainty* and the need for *prompt action*. Uncertainty constitutes a prominent characteristic, an inherent feature, and is the major obstacle to effective decision making, risk assessment and the overall emergency management, especially during the response period (Lewis, 1988). Hansson (1996) presents interesting considerations concerning severe uncertainty related to: the identity of the options not being well determined (*uncertainty of demarcation*), the consequences of at least some options are unknown (*uncertainty of consequences*), it is not clear whether information obtained from others, such as experts and informants, can be relied upon (*uncertainty of reliance*), and the values relevant for the decision are not determined with sufficient precision (*uncertainty of values*). This uncertainty must be dealt with, both during the preparations and the responses, especially when decisions are to be made and executed. Sometimes our uncertainty is regarded as too large and not conducive to making a decision. Decisions are made upon some information. We need more information, the information has to be interpreted, and the decision has to be made in real-time. The decisions must be made in accordance with the demands of the situation. If it is not possible to postpone the decision, the decision maker simply makes his decision. Such decisions have of course a greater uncertainty, in the sense that the background knowledge should have been better. We have to make trade–offs quickly, based on real-time constraints, in order to respond effectively in real-time.

Sometimes our uncertainty is regarded as too large and not conducive to making a proper decision. Klein (1989) lists four sources of uncertainty: (1) *Missing information*, (2) *Unreliable information*, (3) *Ambiguous or conflicting information*, and (4) *Complex information*. Uncertainty can also be described, according to Klein, as a sense of doubt that threatens to block or delay action. Note the distinction between (Klein 1989) and (Lipshitz and Strauss 1997) regarding the description of uncertainty. Klein uses the opening words *threats to block* (it might happen) whereas for Lipshits and Strauss the action verb *block* (it happens). The distinction stresses the importance of the threat, a consequence of the uncertainty.

Orasanu and Connolly (1993) describe uncertainty as incomplete, ambiguous and changing information. At the scene of the crisis the information is fragmented and ambiguous and it is difficult to assemble a clear picture of the dynamic situation. The decision maker can lose valuable information in a critical situation because of overload or deficiency of information. The lack of information, or the overwhelming amount of information, present at the scenes causes problems to the incident commander. This highlights at least two important elements of the information process: (1) *The need to search for meaningful information* and (2) *The processing of information*. According to (Rijpma and Van Duim 2001) crisis management demands rapid information processing to succeed.

Lipshitz and Strauss (1997) identified three forms of uncertainty when studying retrospective reports of decision making under uncertainty: (1) *Inadequate understanding of the situation*, (2) *Lack of information on which to base a decision*, and (3) *Conflict among decision alternatives*. They hypothesised RAWFS heuristics and proposed that decision makers cope with the uncertainty with a heuristic consisting of five strategies: (1) ***R**eduction* (collecting more information), (2) ***A**ssumption based reasoning* (use assumption to close information gaps), (3) ***W**eighing pros and cons* (of at least two alternatives), (4) ***F**orestalling* (generate options), and (5) ***S**uppression* (ignoring uncertainty by suppressing negative information). According to the RAWFS heuristic, the decision process begins with an attempt to understand the situation, recognise or make sense of it. If these tactics succeed, the decision maker initiates a mental process of serial option evaluation.

Lipshitz *et al.* (2007) followed up their hypothesis by studying the fire ground commander at ten incidents. The commander used a helmet-mounted video camera and microphone during the response. The commander reviewed the video and audio records and reported his associated thinking process. In total 150 uncertainties were mapped. They found that the commanders preferred to use reduction tactics, especially information search, and relied on information from other people, such as bystanders, subordinate fire fighters or from other emergency units. When such tactics were impractical, or failed, they switched to assumption-based reasoning. These findings are consistent with RAWFS heuristics. The use of "weighing pros and cons" and "suppression" was not confirmed. Lipshitz *et al.* (2007) explain the absence by the commander's high level of experience and the use of matching. The commander manages to overcome uncertainty and make a satisfactory decision without expending energy on the other two strategies.

These approaches to uncertainty, expressed by Klein (1989), (Lipshitz and Strauss 1997), and (Orasanu and Connolly 1993), have reasonable conformity: information is crucial to success, i.e. to reduce uncertainty to an "acceptable" risk level so that the decision maker can make his decision. There seems to be a general agreement by researchers to manage uncertainty consistently and use uncertainty explicitly as an assessment and decision tool. It is essential to include uncertainty when studying the incident commander in action dealing with multiple risks.

Next, we address how decision making takes place in times of crisis, especially on-scene, because it is vital for effective risk assessment and the overall risk management. It is, after all, no point having identified risks if we do nothing about it. Decisions have to be made and the following section describes how effective decision making could be carried out by the incident commander.

5. Decision making

Decision making during crises, or under disaster conditions, is a complex and multifarious project (Dror, 1988) and is a core part of risk management. Boin *et al.* (2005) see decision making as the critical task of crisis leadership. Wrong decisions, or decisions made too late, may lead to poor management, poor risk assessment, and loss of values. Decision making is described as a cognitive process leading to the selection of a course of action, among alternatives, at least to do or not to do. Every decision making process produces a final choice. It can be an action, immediately executed or intended to be accomplished in the future. It can also be an opinion. Yates (2001) describes a decision as a commitment to an action that is intended to yield satisfying states and makes a distinction between the decision and the following action. Decision making may be described as a tool to help a decision maker, e.g. the commander on-scene a crisis, to reach successfully the goals of the response. The decision process begins when a person, i.e. the decision maker, needs to do something with a “problem” at hand, but he still does not know what. Therefore, decision making is a reasoning process and can be based on explicit or tacit assumptions. Decision making is said to be a psychological construct. This means that we can never "see" a decision, but we can infer from observable behaviour, such as the implementation of the decided action, that a decision has been made. We may then conclude that a psychological event, which we call "decision making", has occurred.

Decision making in the on-scene context can be described as dynamic and may often be a compromise between a good strategy for controlling the decision task, the problem at hand or the event, and a strategy that enables the decision maker to exert some measure of control over the rate at which he/she has to make decisions (Brehmer, 1992). The latter strategy may be useful when important information is missing and the decision ought to be postponed, if possible, to enable the acquisition of more information to improve the decision basis.

Edwards (1962) gives some characteristics of dynamic decision making. Firstly, *a series of decisions is required to reach the goal*, a successful outcome. Secondly, *the decisions are not independent*. One decision leads to a later decision. Thirdly, *the state of the decision problem changes*, both autonomously and as a consequence of the decisions already made. Brehmer (1992) adds a fourth characteristic: *the decision has to be made in real time*, which means that the decision maker is not free to make decisions when he himself wants to. The decision maker is the "owner of the problem" and no one else can make the decision.

To describe effective decision making, a necessity of excellent incident performance, (Cannon-Bowers *et al.* 1996, 1997) identified six attributes of effective decision making that are important to the incident commander:

1. *Flexible*

If possible the decision ought to be an evolving decision, which means that it can be improved by later decisions if it is not sufficient to cope with the situation. Alternative courses of action should as far as possible not be limited

2. *Quick*

On the incident ground, problems often demand rapid responses. The decision must be taken ″now″

3. *Resilient*

The decision should bear challenges. The situation offers resistance and must be overcome

4. *Adaptive*

The decision is not singular or independent of earlier or subsequent decisions or the situation at hand

5. *Risk taking*

On-scene it is impossible to avoid all threats and hazards to the emergency units if the response is to be successful

6. *Accurate*

The input of efforts and units to solve the problem should be sufficient and appropriate. If not, the incident commander may run out of resources before it is necessary. These attributes are logical because an incident commander acts in situations characterised by uncertainty, severe threats and the need for prompt action. The work of Cannon-Bowers *et al*. suffers so far from a lack of empirical background and they define these attributes on the basis of hypotheses.

We can describe the decision process as a longitudinal time process. Time is running, and it is impossible to stop or take time out. In addition, the goal of a successful outcome is not straightforward in a crisis. The decision making is often incremental, in which it is difficult to relate sub-goals to the ultimate one. This is emphasised by (Klein *et al.* 1993), who describe on-scene command situations with ill-defined goals and ill-structured tasks. In order to cope with the event, the on-scene commander has to perceive the real-time situation and its dynamics. The workload on the on-scene commander can be extreme, compounded by *ill-structured problem*, *critical values at stake*, *multiple players involved*, *time constraints* and *competing goals* (Orasanu and Conolly, 1993). The decisions made in the first minutes and hours are crucial to successful mitigation and the overall conclusion of the crisis (Flin, 2001, Kowalski and Vaught, 2001). Weingart and Wyer (2006) describe emergency medicine decision making as critical choices in chaotic environments. In short: on-scene an incident the activities are complex, the stakes are high and the effects on lives potentially significant. There has been a shift towards attempting to understand how decisions are made in the real world. Cannon-Bower *et al.* (1996), describe the development of the Naturalistic Decision Making (NDM) perspective as a paradigm shift. NDM became more influential in explaining management and decision making in command and control situations. Klein (1989) studied experienced fire ground commanders and during his findings and conclusion, introduced NDM. NDM differs clearly from the classic analytical approach, with respect to experience and field settings. NDM is concerned with experienced personnel operating in real life settings rather than studying naive participants, such as students, in a laboratory setting. NDM research studies seek to describe what is already happening in its natural context, as opposed to the traditional approach of prescribing an ideal way of finding the best option or an improvement to the existing strategy on hand.

The essentials of the approach have remained the same since it emerged in the 1980s and consist of three basic principles (Bryant, 2002):

1. Decisions are made by sequential, holistic evaluation of the action against some criterion of acceptability, rather than by comparison of multiple alternatives along multiple dimensions
2. The decision maker relies primarily on recognition-based processes to generate options and compares them to previous personal experiences. On-scene an accident the incident commander identifies a potential course of action by assessing the situation, then recognising past situations that are similar and determining their acceptability to the current situation

3. Mental evaluation is used to evaluate the course. A satisfactory solution is more important than finding an optimal solution (Simon, 1955). If the course is acceptable, the decision will be accomplished. Emergencies and rescue operations demand very rapid responses, and the rescuers accept a solution that merely works without considering whether a better solution exists.

In general, NDM models do not discuss the technological aspects of complex systems and the influence they have on decision making. The decision maker bases much of what is decided on what is perceived and the following cognition. Much of what is perceived is influenced by the design of the technologies, i.e. communication systems, digital maps and GPS, in the system. Technology can aid or impede decision making by representing the environment more or less accurately. Managers may be passive recipients of data presented by the technology or they may be in a position to shape and direct the technology (Shattuk and Miller, 2006).

So far we have discussed how crisis management, and incident commanding, deals with present risk or avoiding risk that might or will arise, during the crisis. We have also looked at uncertainty and its impact on the decision making on-scene. Next, we tie it all together in the risk management process during crisis.

6. Risk and risk assessment

Risk is traditionally understood as the potential negative impact of an activity and some characteristics of value that may arise from some present process or future event. There are many definitions of risk, depending on the specific application and situational context. Generally, risk is related to the expected losses which can be caused by a risky event, and to the probability of this event. The higher the loss and the more likely the event is, the worse the risk. In everyday usage, risk is often used synonymously with the probability or possibility of a loss or threat or suffering harm. Risk is a threat to a successful outcome.

Aven *et al.* (2004) describe the risk as the *combination* of possible consequences of a certain activity and the uncertainty of the consequences – the outcomes. The probability is a subjective measure of uncertainty. The term risk is related to future outcomes and related probabilities. The use of risk analysis and/or risk management offers an interesting approach to dealing with crises and incidents. Comfort (1988) describes analysis of risk as a rational process that results in a powerful goal for action of emergency operations. The dominant concept of risk, applied to safety and emergency management, can be described as *the engineer perspective*. The engineer perspective views risk as an inherent property of the system, in this case the accident scene, and the purpose of a study of the risk is to reveal the true risk. Thus, a sharp distinction is made between the real objective risk and the perceived risk. The focus is on the risk figure, which is an unobservable unknown quantity and thereby difficult to use in decision making on- scene. The engineer perspective is not practically applicable to incident commanders because it is too extensive and time demanding when the situations need rapid decisions. The engineer perspective is described in several textbooks and papers. See for example (Henley and Kumamoto 1981), (Modarres 1993) and (Vose 2000). The fundamental issues of risk are discussed for example in (Apostolakis and Wu 1993), (Hoffman and Kaplan 1999), and (Aven 2003).

An alternative risk perspective focuses on observable quantities, such as (Aven 2003, 2007), and can be described as knowledge and decision oriented analysis. Aven is inspired by *the classic analytic decision perspective*. This perspective searches for methods and models to apply in the case of multiple alternatives, to analyse, compare and choose the "best" solutions or decision. Aven emphasises the uncertainty element and implies that the risks are characterised by the combination of possible consequences with an activity, and the related uncertainties of the future consequences or outcomes. In relation to on-scene activities, such quantities could be the number of victims trapped in an earthquake, volume of gas from a gas leak, diffusion of an ammonia cloud, location of children caught in a fire scenario, materials exposed to fire, occurrence of structure breakage during fire fighting, time and capacity to carry out rescue operations, the number of dead and injured, injury categories and so on. Such quantities are of interest to the incident commander in the real time of the emergency. These quantities are called observables, because we will observe these quantities when the activity has ended. However, they are uncertain during the response activity. The risk and inherent uncertainty is quantitatively expressed by probabilities and the associated predictions of the observables. The uncertainties are assessed and the probabilities assigned. In this sense, the risk is purely epistemic. In other words we are uncertain because we lack sufficient knowledge. Aven (2003) denotes this concept of risk the predictive Bayesian approach to risk. See also (Njå and Rake, 2003).

We need to emphasise that risk analysis is an analytic decision support that describes the risk. Risk analysis results alone cannot be used to make a decision in action. Risk analysis is a decision tool. The frequency of rare failures can be hard to estimate and loss of human life is generally considered unacceptable and these considerations hamper the use of risk analysis on-scene an accident. It is imperative that risk assessment must be a part of the decision process for engendering effective decisions. Even so, a study including real time observations of experienced incident commanders from the police, the ambulance service and the fire department, could not identify any systematic risk management strategies in the observed accident cases, neither by the individual incident commanders nor by rescue teams (Rake and Njå, 2009). Risk as a concept important for on-scene activity was strongly connected with the responders' safety. The responses were mainly reactive, comprising direct action to deal with the visible hazard and the problem at hand. Pro-active strategies aimed at revealing and tackling uncertain events and risks were rarely seen (Rake and Njå, 2009).

How can focus on risk be useful on- scene? By using risk analysis as a decision support in his decision making process the commander can, firstly, *decide what kind of information is needed within specific time frames*, and, secondly, *decide which strategies and measures are to be applied in real-time*. Some systematic ways of dealing with risk are:

1. Risk prevention – strategies to reduce the probability of occurrence of a risk
2. Risk mitigation – strategies to reduce the impact of an occurring risk. For example, fail safe mechanisms in systems are designed for this purpose.
3. Insurance – transfer the risk to a third party.
4. Accept – simply do nothing about the risks. Typically applied to either very small risks or low probability, high impact events for which humans are powerless.

5. Reduce the uncertainties surrounding the situations in order to improve the quality of the decision – this will apply to all four generic strategies listed above.

In addition, it should be stressed the importance of effective communication of the involved risks with other actors. This communication results in attentive responders and managers and this reduces the uncertainties inherent in the situation. Therefore the risk is reduced.

There are few empirical evaluations of the actual influence of risk assessments on the decision making of first responders and incident commanders (Braut *et al*, 2012). This limitation restricts our ability to make broad generalisations. In addition, the complex situation on scene, with multiple actors, critical values at stake, need for prompt action, the dynamic situation, the vast numbers of variables involved and the inherent uncertainty, makes use of the risk concept demanding. However, by reducing uncertainty, and highlighting alternative decision options, the performance on scene will improve and risk reduced. In particular, accident situations regarded as abnormal to the incident manager, or the decision maker on-scene, would benefit from this way of thinking.

Identification of the observable quantities, which are critical values at stake, e.g. number of trapped persons inside a burning building, is a key principle to identify the risks which must be dealt with. This information is the basis for the decision maker's representation of the actions, contingency and outcomes that seem relevant to managing the incident. The decision prospects are assessed in relation to possible outcomes and the assigned probabilities. These assessments form the basis for the final decision, followed by an ongoing feedback process throughout the response. This approach involves and affects the risk and risk assessment.

The rescue operations are normally carried out by conscientious managers, commanders, and rescuers, doing their best to optimize the consequences and mitigate losses. However, the performance of the operations rest on standard procedures and experiences from "normal" responses. Despite the broad spectrum of incident types and conditions, a management has to take charge of the site, assess the situation and implement a plan of action to bring events under control. In familiar action, i.e. well known types of rescue operations, operational decisions are not based on rational situation analysis, only on the information, which in the given context is necessary to distinguish among the perceived alternatives for action. In such actions the experienced manager or decision maker makes his decisions upon recognition of the situation. Is the situation familiar? If so the decision maker selects an action, which he/she "knows" will cope with the urgent situation. In general, a pro – active mental analytic risk approach to the problem at hand will be effective when the situation is unclear and unknown. If the manager or decision maker systematically focuses on *Threats and what's at stake, Decision alternatives, Uncertainty* and *Consequences* it's possible to work pro – active and to cope both with well – known and uncommon situations during the emergencies (Rake, 2004). Even when decision makers have the necessary information and competence their emergency management will not be effective if they are not aware of the need to consider the potential risk involved in the situation and in their decisions. A risk approach is essential.

A core of successful risk assessment and decision making is to map the most likely threats to the future and analyze the subsequent impacts (from the threats) we have to cope with. Concentrate on threats and what's on stake before the incident commander, or decision maker on scene, lists the most relevant decision alternatives to solve the problem at hand or reduce the threats. To analyze the inherent uncertainty and consequences of the alternatives is the final step before the decision is made (Rake, 2004).

At the incident the information is fragmented and ambiguous thus making it difficult, and sometimes impossible, to assemble a complete or clear picture of the hazardous situation and the risks. Further observations, i.e. more information, are not always relevant. They only have value for the managerial problem at hand when the result of an observation could lead the crisis management to make a new or different decision. In a dynamic situation continuous/ persistent observation and information is necessary and a part of the described risk approach. Vital information can also fall victim to the situation and never reach the commander on-scene, or even more often, the emergency management remote from the arena, even if the situation changes relative slowly as in the Asian flow disaster– the tsunami – 2004.12.26 and the following days (Evaluation Norway, 2005).

The expert must recognize the problem/risk/threat even when an explanation is not available. This suggests that tacit knowledge (e.g. knowledge not easily verbalised) may play an important role in effective risk management. Training together with experienced managers and decision makers, followed by evaluation, can be suitable to unmask and transfer such competence.

7. Closure

Even the best management, decision or response may be overwhelmed by the situation over which the decision makers have no control, resulting in an undesirable outcome. However, if the understanding of the situation is impaired, then the ability to predict outcomes of actions is more flawed, and due to this, risks of an accident occurring are increased, independent of the plan or approach to problem solving. The problem is not the faults, errors or omissions made by the decision maker or the management, but to recognise/perceive the faults, and adjust the implementation of the decisions in time to avoid negative consequences, and make new and better decisions.

We suggest that extensive training of risk assessment and problem solving will be valuable and may lead to better risk management on-scene and less undesirable outcomes. We would also emphasize the need for more and focused research. Especially important are the connection between situation assessment and inherent uncertainty, decision making and the risk assessment process.

Research in different settings, as in real time, is also necessary (Rake, 2003). The research should focus on the decisions maker. Normally we investigate incidents when the leaders fails, as in the Piper Alpha disasters (Cullen, 1990). An alternative meaningful approach could be to study successful risk managers.

One interesting research and development project is the BRIDGE project (BRIDGE, 2012) within the EU Seventh Framework Programme. BRIDGE intend to build a system to support interoperability – both technical and social – in large-scale emergency management. The system plans to serve as a bridge between first responder organisations, contributing to an effective and efficient response to natural catastrophes, technological disasters, and large-scale terrorist attacks. Important parts of the project are technical solutions and procedures to avoid risk and cope with the risk on-scene.

8. References

Apostolakis, G. and Wu, J. S. (1993). The interpretation of probability, De Finetti's representation theorem, and their implications to the use of expert opinions in safety assessment. In Barlow, R. E. and Clarotti, C. A. (eds.), *Reliability and Decision Making,* Chapman & Hill, London, pp. 311-322.

Aven, T. (2003). *Foundations of Risk Analysis. A knowledge and Decision-oriented Perspective.*John Wiley & Sons Ltd, Chichester, England

Aven, T. (2007). *A unified framework for risk and vulnerability analysis covering both safety and security*. Reliability Engineering & System Safety 92, pp. 745-754

Aven, T, Boyesen, M., Njå, O., Olsen, K.H. and Sandved, K. (2004). *Societal safety,*Universitetsforlaget, Oslo, Norway. (in Norwegian)

Beck, U. (1992). *Risk society: towards a new modernity.* London: Sage

Bigley, G.A. and Roberts, K.H. (2001). *The Incident Command System: High - Reliability Organizing For Complex and Volatile Task Environments.* Academy of Management Journal, vol 44, pp. 1281-1300

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*, 211-241

BRIDGE (2012). http://www.bridgeproject.eu/en

Braut, G.S., Rake, E.L., Aanestad, R. and Njå, O. (2012). *Risk images as a basis for decisions related to provision of public services.* Risk Management, 14 (1) pp. 60-76

Bryant, D. J.(2002). *Making Naturalistic Decision making "Fast and Frugal".* Proceedings of the 7th International Command and Control Research Program. Defence Research Development Canada, Toronto, Canada. http://www.dodccrp.org/

Boin, A., 't Hart, P., Stern, E. and Sundelius, B. (2005). *The Politics of Crisis Management. Public leadership under Pressure.* Cambridge University Press, Cambridge, UK.

Cannon - Bowers, J.A., Salas, E. and Pruitt, J.S. (1996). *Establishing the boundaries of a paradigm for decision making research.* Human Factors, 38(2) pp. 193-205

Cannon - Bowers, J.A. and Bell, H.H. (1997). Training Decision Makers for Complex Environments: Implications of the Naturalistic Decision Making Perspective. In Zsambok, C.E. and Klein, G. (eds.): *Naturalistic Decision Making.* Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, USA. pp. 99-110

Comfort, L.K. (ed.) (1988). *Managing Disaster. Strategies and Policy Perspectives.* Duke University Press, Durham, USA

Cullen, Lord (1990). *The public inquiry into the Piper Alpha Disasters.* Vol. I and II, HMSO, London, UK

Dror, Y. (1988). Decisionmaking under disaster conditions. In Comfort, L.K. (ed.). *Managing Disaster. Strategies and Policy Perspectives.* Duke University Press, Durham, USA

Edwards, W. (1962). *Dynamic decision theory and probabilistic information processing*. Human Factors, 4, pp. 59-73

Evaluation Norway (2005). *26:12 Rapport fra Evalueringsutvalget for Flodbølgekatastrofen i Sør Asia.* Statens Forvaltningstjeneste, Oslo, Norway

FEMA (2012). Federal Emergency Management Agency.
http://www.fema.gov/pdf/emergency/nims/ NIMS_core.pdf, p. 53

Fiedler, F. (1996). *Research on leadership selection and training: One view of the future.* Administrative Quarterly, Vol. 41, pp. 241-251

Flin, R. and Arbuthnot, K. (eds.) (2002). *Incident command: tales from the hot seat*. Ashgate Publishing Limited, Aldershot

Flin, R. (2001). "Decision Making in Crises: The Piper Alpha Disaster. In Rosenthal, U., Boin, A and Comfort, L.K. (eds.). *Managing crises; Threats, dilemmas opportunities.* Charles C Thomas, Springfield, Ill. USA. pp. 103-119.

Fredholm, L. (2006). Coping of small and major societal crisis. (in Swedish). In Fredholm, L. and Göranson, A-L (eds.). *Management of rescue operations in the complex society*. Rädningsverket, Karlstad, Sweden, pp. 15-30. (in Swedish)

Hansson, S. O., (1996). *Decision Making Under Great Uncertainty*. Philosophy of the Social Sciences, 26 (3), pp. 369-386

Henley, E. J. and Kumamoto, H. (1981). *Reliability Engineering and Risk Assessment.* Prentice-Hall, N.J., USA

Hoffman, F. O. and Kaplan, S. (1999). *Beyond the domain of direct observations: how to specify a probability distribution that represents the state of knowledge about uncertain inputs*. Risk analysis, 19, pp. 131-134

Klein, G. (1989). Recognition – Primed Decisions. *Advances in Man-Machine Systems Research, 5*, 47-92.

Klein, G., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.). (1993). *Decision Making in Action: Models and Methods*. Norwood, N.J.: Ablex Pub

Kowalski, K. M. and Vaught, C. (2001.). Judgement and Decision making under Stress: An Overview for Emergency Managers. In *8th Annual Conference Proceedings* (TIEMS 2001), Oslo, Norway

Lewis, R.G. (1988). Management Issues in Emergency Response. In Comfort, L.K. (ed.). *Managing Disaster. Strategies and Policy Perspectives*. Duke University Press, Durham, USA, pp. 163-179

LESLP (2008). *Major Incident - Procedure Manual*. 7th edition, London Emergency Services Liaison Panel, Metropolitan Police Service, London, UK. See also http://www.leslp.gov.uk/

Lipshitz, R., Omodei, M., McClellan, J and Wearing, A. (2007). What's burning? The RAWFS heuristics on the fire ground. In Hoffman, R. R. (ed) (2007). *Expertise out of context*. Lawrence Erlbaum Associates, New York, USA. pp. 97-111

Lipshitz, R., and Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes, 69*(2), 149-163.

McLennan, J., Omodei, M.M, Holgate, A.M., and Wearing, A.J. (2003). *Human Information Processing aspects of Effective Emergency Incident Management Decision Making*. The Human Factors of Decision Making in Complex system Conference, Dublane, Scotland

Modarres, M. (1993). *What every engineer should know about Reliability and Risk Analysis*. Marcel Dekker, N.Y., USA

NATO (2002). *NATO code of best practice for command and control assessment*, rev. ed., DoD Command and Control Research Program, Washington, D.C., USA

Njå, O., and Rake, E. L. (2003). *Risk Based Decision Making on Accident Scenes. In Emergency Management in a Changing World*. Paper presented at the The International Emergency Management Society 10th Annual Conference.

Orasanu, J., and Connolly, T. (1993). The Reinvention of Decision Making. In G. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds.), *Decision Making in Action: Models and Methods*. Norwood, N.J.: Ablex Publishing Corporation.

Perrow, C. (1999). *Normal accidents: Living with high-risk technologies*. Princeton, N.J.: Princeton University Press.

Rake, E. L. (2003). Emergency management and decision making on accident scenes: taxonomy, models and future research. *International Journal of Emergency Management, 1*(4), 397-409

Rake, E. L. (2004). *A Risk-Informed Approach to Decision Making in Rescue Operations.* Paper presented at the International Conference on Probabilistic Safety Assessment and Management (PSAM7/ESREL04), Berlin

Rake, E.L. and Njå, O. (2009) Perceptions and performances of experienced incident commanders. *Journal of Risk Research*, 12(5), 665-685

Rijpma, J.A. and van Duim, M.J. (2001). The Response to the Hercules Crash. In Rosenthal, U., Boin, A. and Comfort, L.K. (eds.). *Managing crises; Threats, dilemmas opportunities.* Charles C Thomas, Springfield, Illinois, USA. pp 143-155

Rosenthal, U., Boin, R. A., and Comfort, L. K. (Eds.). (2001). *Managing crises: Threats, dilemmas, opportunities*. Springfield, Ill.: Charles C. Thomas.

Rosenthal, U., Hart, P. t., and Charles, M. T. (1989). The World of Crises and Crisis Management. In U. Rosenthal, M. T. Charles & P. t. Hart (Eds.), *Coping with Crises: The Management of Disasters, Riots, and Terrorism*. Springfield, Ill., U.S.A.: C.C. Thomas.

Shattuk, L.G and Miller, N.L. (2006). *Extending naturalistic decision making to complex organisation: A dynamic model of situated cognition*. Organization studies 27(7), pp. 989-1009

Simon, H. A. (1955). *A behavioural and organizational choice*. Quarterly Journal of Economics, 69, pp. 99-118

SOUS (2005). *Sverige och tsunami- granskning och förslag.* Statens Offentliga Utredningar, Stockholm, Sweden

Vose, D. (2000). *Risk Analysis*. John Wiley & Sons, LTD, New York, USA

Weingart, S. and Wyer, P. (2006). *Emergency medicine decision making: critical choices in chaotic environments*, McGraw-Hill, New York, USA

Yates, J. F. (2001). "Outsider:" Impressions of Naturalistic Decision making. In E. Salas & G. Klein (Eds.), *Linking Expertise and Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates.

Section 2

Engineering

8

Uncertainties and Risk Analysis Related to Geohazards: From Practical Applications to Research Trends

Olivier Deck and Thierry Verdel
LAEGO – Laboratoire Environnement Géomécanique et Ouvrages, Université de Lorraine,
Ecole des Mines de Nancy – Parc de Saurupt – F 54042 – Nancy Cedex -
France

1. Introduction

Geohazards correspond to hazards that involve geological or geotechnical phenomena like earthquake, landslide, subsidence… Such hazards are generally classed into natural hazards even if their origin is not always natural, as for mining subsidences that are the consequence of industrial underground excavations. Geohazards are mostly investigated for the purpose of risk analysis. Studies first concern hazard (HAZUS®MH MR4, Romeo et al. 2000, Wahlström and Grünthal 2000) and vulnerability assessment (Zhai *et al.* 2005, McGuire 2004, Hazus 1999, Spence *et al.* 2005, Ronald *et al.* 2008), and secondly risk assessment and management (Karmakar *et al.* 2010, Merad *et al.* 2004). However, risk management must deal with many uncertainties that concern different aspects of risk assessment. This chapter aims to clarify the interactions between risk management and uncertainties within the context of geohazards.

Uncertainties may first be semantic when applied to the definition of the vocabulary used in risk analysis related to geohazards. Risk is generally synthesised as the conjunction of hazard and vulnerability (Karimi and Hüllermeier 2007), but many definitions are available for both terms. Moreover, themselves use terms that are mostly discussed in the literature, such as resilience (Klein *et al.* 2003). As a consequence, a precise definition of risk is necessary, in a given context, to avoid many misunderstandings amongst public authorities, scientists and citizens that may arise from semantic problems. Such a precise definition should not be interpreted as better than others but as a consensual definition adapted to the specific context of each study. This goal is addressed in Section 2 of this chapter.

The third section addresses the question of the relationship between risk and uncertainties and the identification or classification of the different possible uncertainties. While many authors consider aleatoric and semantic uncertainties (Bogardi 2004, Adger 2006, Ezell 2007) as the two main groups, this part focuses on other classifications and highlights definitions of uncertainties.

Finally, this chapter focuses on two specific aspects of the uncertainties and risk analysis related to geohazards: risk prioritisation and vulnerability assessment. These two aspects are illustrated with recent trends developed in the field of risk management within the context of mining subsidence hazards. Some final remarks are offered in Section 4.

2. Risk, hazard and vulnerability

Before any risk analysis, a precise definition of the risk is required to avoid misunderstandings amongst public authorities, scientists and citizens, the main actors involved in or concerned with such an analysis.

Many definitions are available, and none of them should be considered "the best". Of greatest importance is being aware of the complexity of risk and the fact that each definition focuses on different aspects of risk. In this chapter, definitions given by the United Nations (UN/ISDR 2004) are first presented as a reference and then discussed by comparison with other definitions.

2.1 About risk

The International Strategy for Disaster Reduction defines Risk as "the probability of harmful consequences or expected losses (deaths, injuries, property, livelihoods, economic activity disrupted or environment damaged) resulting from interactions between natural or human-induced hazards and vulnerable conditions" (UN/ISDR 2004). UN/ISDR then adopts the classical definition expressed by the formula: Risk = Hazards x Vulnerability. A note specifies that "some disciplines also include the concept of exposure to refer particularly to the physical aspects of vulnerability".

Risk is also a mathematical concept, mostly used in economics and engineering and defined as the mathematical expectation of losses or gains. This definition takes into account the occurrence probability of each possible consequence in terms of gains or losses. Therefore, it is usually necessary to estimate the occurrence probability of every event responsible for the consequences.

However, in natural risk assessment or industrial risk assessment, the mathematical expectation of losses is rarely computable because of uncertainties. One usually prefers to work with a risk matrix based on simple formulas, such as Risk = Frequency × Severity for industrially generated risks (Cox 2009, Suddle, 2009, Aven 2010) or Risk = Hazard x Vulnerability for naturally generated risks (Karimi and Hüllermeier 2007), where Frequency, Severity, Hazard or Vulnerability are defined using simple scales with few levels.

Even for these simpler definitions, risk assessment involves many difficulties. Consequently, most practical applications for risk analysis related to geohazards aim to identify different geographical areas associated with a reference event, i.e., the largest event in intensity that is likely to occur over a specific area during a fixed period of time. In many cases, the risk analysis is limited to the hazard assessment, and the vulnerability is basically investigated with an identification of the assets only. These assessments and the associated zoning are used to define legal specifications as standards for civil engineering projects.

On the whole, the definition of risk is a challenge, but it is necessary to avoid confusion and misunderstandings. In this chapter, we adopt the definition suggested by the (UN/ISDR 2004), where a risk is defined by the two components of hazard and vulnerability. In this definition, hazard is a notion that includes both the probability of an event and its intensity, while vulnerability characterises the assets' susceptibility to damage.

2.2 About hazard

The term hazard is primarily used to designate natural feared events as opposed to industrial accidents. This common designation is a consequence of the chosen risk definitions, which do not consider the same components: hazard x vulnerability for natural risks and frequency x severity for industrial risks. Natural hazards include geo-environmental hazards, such as earthquakes, floods, ground movements (e.g., landslides or subsidence) and fire. The International Strategy for Disaster Reduction defines a Hazard as *"a potentially damaging physical event, phenomenon or human activity that may cause the loss of life or injury, property damage, social and economic disruption or environmental degradation"*.

The definition of hazard is, on the whole, clearly stated as the combination of both the probability of occurrence of an event (e.g., an earthquake and flood) and the intensity of the event. Assessment of these two components is based on different methods, which can be empirically, theoretically or statistically defined. Whatever the method used, assessment of both the intensity and the probability are spoilt by uncertainties because of poor knowledge.

When natural hazards are cyclic, i.e., when phenomena are assumed to be stationary, a Poisson's statistical process can be considered. If historical data are available and sufficient in number, statistical analysis can then be used to assess a return period of the hazard in a given place or the probability of occurrence during a fixed period of time. For extreme events, historical data may be insufficient in number, and assumptions are necessary with respect to, for instance, the Gutemberg-Richter law for seismic activity or the Weibull statistical distribution for floods.

The advantage of a statistical analysis is that predictions can be validated by observations at a certain level of confidence. When a very large number of historical data are available, they can be used to assess both the trend of a hazard (probability and intensity) and the sensitivity to each parameter to understand the physical and mechanical phenomena that lead to an event. When historical data are few in numbers, as is often the case for ground movements such as subsidence and landslides, only the trend can be studied. However, risk analysis may require a good understanding of the influence of a set of parameters to go deeper into the analysis and the prediction of occurrence. Statistical analyses can then be combined with analytical models to capture the influence of the studied parameters.

Finally, the choice of the intensity parameter is difficult because it corresponds to a simplification of the phenomenon, as in the following two examples:

- The height of water for flooding situations might be chosen, even though the duration of the flood and the speed of the stream may also be important;
- The peak ground acceleration for earthquakes might be chosen, even though the amount of damage also depends on the duration of seismic activity and its spectra; the volume of unstable rock masses for landslides and other conditions.

When different parameters are available, it is necessary to select the one that is the most appropriate to assess the damage. Indeed, the two concepts of hazard and vulnerability are not completely independent. For example, in the case of ground subsidence, the maximal vertical ground subsidence is appropriated to assess the intensity in regard to the modification of rivers or pipe flow, but the horizontal ground strain that is not maximum in the same area as the vertical subsidence is appropriated to assess the intensity in regard to the building resistance.

On the whole, hazard is the component of risk that is the most investigated and the easiest to assess and to plot on a map for risk management purposes. When hazard consists into a range of possible intensities associated to different occurrence probabilities, engineering studies generally simplify analysis and also consider a maximum reasonable event. This event may correspond to the maximum intensity that already occurred when such data are available, or to a selected occurrence probability or return period (a century return period for flooding for example). On the contrary, vulnerability analysis involve more difficulties due to its more complex definition.

2.3 About vulnerability

The concept of vulnerability is used in many definitions of risk, but its definition still raises discussions. The United Nations, through the International Strategy for Disaster Reduction, defines vulnerability as *"the conditions determined by physical, social, economic, and environmental factors or processes, which increase the susceptibility of a community to the impact of hazards"* (UN/ISDR 2004).

A comparison of definitions is useful to grasp the notions included in the term "vulnerability". Ezell (2007) summarises 14 definitions of vulnerability, and (Griots and Ayral 2001) make an inventory of 17 definitions and split the vulnerability concept into two elementary notions: 1) the notion of sensibility, susceptibility, weakness and predisposition, and 2) the notion of damage, impact, consequences and losses. To study vulnerability, then, it is necessary to assess the susceptibility (first concept) of exposed assets to the considered hazard and the potential consequences (second concept) that may occur as a result.

The susceptibility (first concept) depends on different factors that are not only physical but also social. Schmidtleim et al. (2008) give an example that proves the importance that should be given to the social characteristics of the concerned community in the case of a hurricane. Other definitions lead to the same conclusion. For example, (Haimes 2006) defines vulnerability as *"the manifestation of the inherent states of the system (e.g., physical, technical, organizational, cultural) that can be exploited to adversely affect (cause harm or damage to) that system"* to measure the risks to critical infrastructures from terrorist attacks and natural disasters.

In France, the Ministry in charge of the Environment defines vulnerability as the "level of foreseeable consequences of one natural phenomenon upon assets" in which assets are "people, goods, activities, means, heritage... likely to be affected by a natural hazard" (MATE, 1997). This definition highlights that damage and its consequences (concept 2) may involve a large number of different assets.

For human-caused security threats, (McGill *et al.* 2007) describe five dimensions of the asset-level consequences: fatalities that take into account the number of deaths and injuries; repair costs measured in dollars; value of assets lost (e.g., goods, property, and information) measured in dollars; time to recuperate measured in units of time; and environmental damage. This definition highlights the fact that consequences may be of different types and that it is simplistic to reduce them to a cost. Mining subsidence events in Lorraine, France, showed, for instance, that social damage occurred in addition to direct costs associated with damage.

Schmidtleim et al. (2008) define vulnerability as "the likelihood of sustaining losses from some actual or potential hazard event, as well as the ability to recover from those losses". This definition highlights the importance of resilience. Resilience is a concept that is largely discussed by Klein et al. (2003). Based on several definitions, they suggest restricting this term to describe "the amount of disturbance a system can absorb and still remain within the same state or domain of attraction; the degree to which the system is capable of self-organisation; the degree to which the system can build and increase the capacity for learning and adaptation".

Based on similar considerations, Bogardi (2004) still reveals uncertainties due to several points: the question of "how far should vulnerability be seen as the 'susceptibility' alone or being rather the product of hazard exposure and that very susceptibility?"; the question of the "proper scale (national, regional, community, household or individual) to capture and to quantify vulnerability"; and "whether (social) vulnerability can adequately be characterised without considering simultaneously the response (coping) capacity of the same social entity".

From a theoretical point of view, the susceptibility of assets to damage may be dependent on the intensity of the hazard. Considering that this intensity may differ widely in its probability, a study of the vulnerability might lead to as many elementary studies as the number of various potential hazard intensities. Because of the number of studies that this theoretical point of view would lead to, engineers used to consider a reference event to make a single assessment of the vulnerability for a specific value of the hazard intensity.

The second question is also considered by (Balandier 2004), and it highlights the fact that the same risk element does not have the same importance depending on the type of hazard and the surface area concerned (e.g., country, city, district) because of their relative importance.

The third question refers to the resilience concept that has already been discussed.

In summary, the vulnerability term has many different meanings. It is of greatest importance, then, to clearly define the held meaning before engaging in any study. Figure 1 shows a possible synthesis, where the vulnerability is split into three components:

1. Weakness includes the physical vulnerability and is linked to the strength of assets (buildings and facilities in particular). The vulnerability increases as the value of the weakness increases.
2. The stakes value includes the functional vulnerability and is linked to losses associated with functional damage. The vulnerability increases with the increase of the stakes value.
3. Resilience as defined by Klein *et al.* (2003). The vulnerability decreases with the increase of the resilience.

In this synthesis, the assets are used to define elements that may be damaged (e.g., people, buildings, infrastructures, goods, and activities). The stakes value is used to define the importance of these elements according to the cost of repairs or the possible other consequences of damage (e.g., functional damage and social damage).

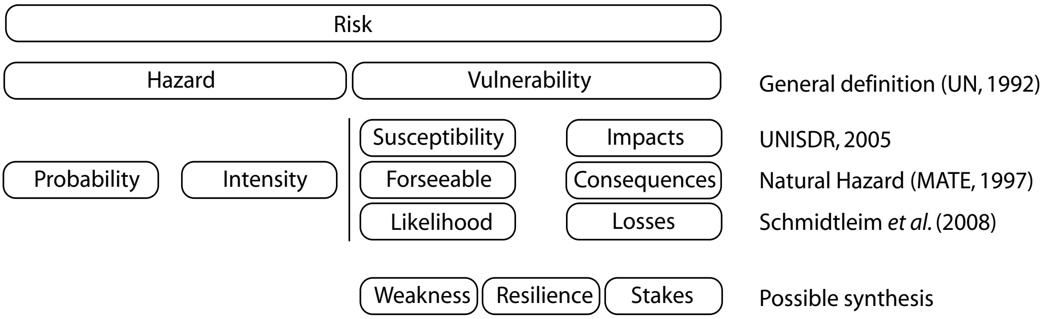


Fig. 1. Synthesis of different definitions for the Vulnerability term.

We have so far discussed risk, hazards and vulnerability, but without taking into account uncertainty. Next, we will investigate uncertainty as opposed to risk.

3. Uncertainty versus risk

A simple illustration of the notion of uncertainty is given by (Chacko 1990): when you throw a coin or a die, you know what results are possible, but it is unclear which of these outcomes will be achieved. For Chacko, uncertainty is not the ignorance of the possible outcomes (there is even a certainty on this issue in the case of the dice game), but the indeterminacy of the result that will happen. This uncertainty results, in particular, from the equi-probability of possible outcomes that prevents any rational estimate of the result. If the possible outcomes were not equally probable, the player would be well advised to bet on the number with the highest probability of occurrence. In making this choice, the player would certainly take a risk (that a different number falls anyway) but a calculated risk.

We see in this example, if a little simplistic and formal, that the concepts of uncertainty and risk are related. In common parlance, these concepts are often confused, but they refer to different situations that (Knight 1921) proposed be defined as follows: a "risky" situation is defined as a situation in which random outcomes can be attached to an objective probability distribution, while the outcomes of an "uncertain" situation cannot be associated to any of such probability distributions. This distinction has been discussed at length, particularly by economists. The notion of risk as understood by Knight is indeed "*only valid for repetitive-type decisions taken within a relatively stable economy*" (Galesne, 1996), while for (Hoskins 1973), this definition would in fact "*eliminate any reference to the notion of risk as part of the business life*" because the different possible states of the world are, in practice, difficult to define objectively. Moreover, economic practice has repeatedly shown that one can assign subjective probabilities, the fruits of experience, expertise or beliefs, to improve forecasting and decision making. Therefore, many authors have preferred to define a situation at risk as a situation for which a probability distribution, whatever its nature, objective or subjective, could be associated with its possible states. Inversely, an uncertain situation would be a situation in which no probability distribution can be assigned (Galesne, 1996). Within the same idea, (Callon *et al.* 2001) suggested a simple definition: "*we know we do not know, but that's about all we know: there is no better definition of uncertainty*". In contrast, the risky situation is a situation where we know something, whether the knowledge is probabilistic or not (objectively or subjectively), provided that this knowledge can assist in making a decision.

In summary, we can state that the notion of risk is associated with a rational decision based on the knowledge, which is possibly limited, of the states of the world, while uncertainty refers to a difficulty in describing, in deciding or in assessing the consequences of possible decisions. Risk is something for engineers or managers, while uncertainty would be more for the researcher.

However, this differentiation is mainly that of economists, decision theorists or even sociologists. Physicists and engineers are more tolerant of speaking about uncertainty and even of describing probabilistic states, and we suggest introducing now the different types of uncertainty that they primarily consider.

As shown in (Hacking 1975), from its emergence as a field (in approximately 1660), probability has been considered to have two faces ("Janus-faced" nature of probability). On one side, probability is statistical and applies to stochastic distributions of random processes; on the other side, it is epistemic and may express a degree of belief in the truth of propositions having no statistical nature. It is remarkable that a single term could continue, to this day, to designate these two radically different concepts: frequency and belief, objective and subjective probability, *a posteriori* and *a priori* probability, random and epistemic probability.

On this basis, different authors have tried to qualify the different types of uncertainties, such as (Haimes 2004), who suggested distinguishing between aleatoric uncertainty and epistemic uncertainty.

In geoengineering or geotechnical engineering, Benjamin and Cornell (1970), Ang and Tang (1984), Veneziano (1994), Paté-Cornell (1996), Hofer (1996) and, more recently, Baecher and Christian (2000, 2003) have also discussed the meanings of uncertainty in their field.

Aleatoric uncertainty is sometimes called the natural variability, being inherent in nature and not reducible (Gilbert and Tang, 1995). It has also been called objective or external uncertainty (NRC, 1996). This uncertainty is the result of the variability observed in known populations (Paté-Cornell, 1996), and it has a relationship with a long series of similar events (Baecher and Christian, 2003). Aleatoric uncertainty can also be described as a measurable uncertainty (as Keynes did in 1921 in his Treatise on Probability) by statistical methods. For simplicity, we can say that this uncertainty refers primarily to uncertainties in the data resulting from their random nature in space and time and from their small number, their inconsistency, their poor handling, transcription errors or low representativeness (in samples) (Baecher and Christian, 2002).

Epistemic uncertainty is itself often referred to as model or parameter uncertainty reflecting a lack of knowledge (Baecher and Christian, 2003; Gilbert and Tang, 1995; NRC, 1996) or a subjective or internal uncertainty (NRC, 1996). Some authors use the term ambiguous uncertainty (Pate-Cornell, 1996; Smith, 2000). This type of uncertainty reflects the inability of a model to translate the physical reality modelled or the impossibility to choose a model or the fact that a model cannot fit in time or be able to integrate new data (Baecher and Christian, 2002), which also refers to the robustness of a model. When this uncertainty is the value of a parameter, it may result from the uncertainty of the data (random uncertainty). Epistemic uncertainties may also reflect the uncertainty regarding the veracity of a used scientific theory or a belief about the occurrence of an event. These uncertainties do not rely (or rely very little) on experimental data; they are subjective and typically vary from one person to another (Baecher and Christian, 2002).

Hofer (1996) provided a good illustration of the difference between aleatoric and epistemic uncertainty. Consider two dice, one of which (A) is covered and left untouched, and it is unknown which side is up, while the other (B) is being cast continuously. The uncertainty about the number shown is then epistemic in the case of (A) because there is a lack of knowledge about it, while it is aleatoric in the case of (B). We can estimate the likelihood (objective probability) of each number in case (B), while for (A) we can only assess subjective probabilities based on what we know about the preferences of the person who put the dice on the table.

However, this distinction is sometimes a modeller's choice. For instance, in a calculation with a numerical model, it is much different to consider the cohesion of material as a space random variable (aleatoric uncertainty) or as an ordinary random variable with a unique and constant unknown value to be selected (epistemic uncertainty).

As suggested by (Haimes 2004), the aleatoric uncertainty category refers to temporal or spatial variability and to individual heterogeneities, while epistemic uncertainty comprises model uncertainty, parameter uncertainty in used models, and uncertainty in decision making from modelling.

While no misunderstanding arises from these definitions, epistemic uncertainty encompasses too many aspects to be practically used as a unique concept, which is why (Baecher and Christian 2003) suggested considering the uncertainty in decision making as a 3rd specific category of uncertainty excluded.

Similarly, for the purposes of a risk analysis, we have suggested (Cauvin *et al.,* 2008) distinguishing between not only two or three categories but four classes of uncertainty, ranging from a very general to a very specific uncertainty attributed to (1) the scientific, economical, and politic context of the risk study; (2) the expertise applying to deterministic human choices; (3) the use of models; and (4) the randomness and/or the lack of knowledge on data.

Figure 2 illustrates these 4 categories.

- *Resources uncertainty* deals with knowledge about both the general scientific context of the study and its local particularities. More specifically, it concerns the existence of information about the processes being investigated and the objects being studied.
- *Expertise uncertainty* concerns all of the choices, actions or decisions that can be made by the expert while realising the risk study. It mainly relies on his particular experience as an individual, on his subjectivity and on the way he represents and interprets the information he has gathered.
- *Model uncertainty* is basically induced by the use of tools to represent reality. Finally,
- *data uncertainty* represents both the natural variability existing in the data, the lack of knowledge about their exact values and the difficulty of clearly evaluating them.

The first 3 categories clearly cover the previously defined epistemic uncertainty, while the last one refers partly to the aleatoric uncertainty (natural variability) and partly to the epistemic uncertainty (lack of knowledge).

Moreover, we suggest in Table 1 some methods that can be carried out to deal with these uncertainties depending on their types.

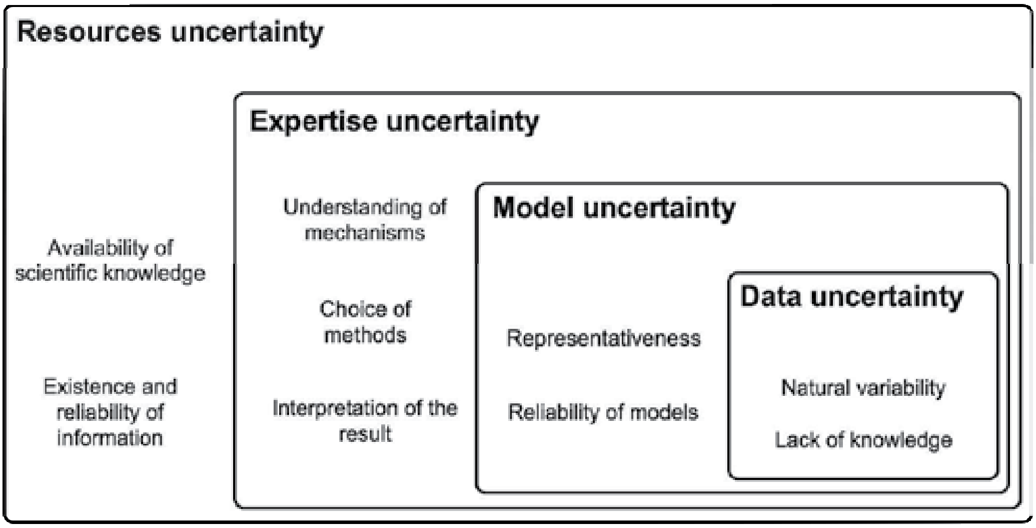


Fig. 2. Four categories of uncertainty in risk analysis

Type of uncertainty	Strategy
Data	- Use of confidence intervals or safety margins - Use of probability functions
Model	- Comparison between model outputs and reality - Integration of various models into the analysis
Expertise	- Comparison between outputs of various models - Use of a decision-aided approach - Use of counter-evaluation - Use of technical methodological guides
Resources	- "*as if*" strategy (qualitative assessment about the existence of a source of danger; we act "as if" it exists, for instance) - Index of existence (quantitative assessment of the degree of confidence in the existence of a source of danger; we try to quantify how confident the experts are about the existence of a danger, for instance)

Table 1. Some methods to deal with uncertainty, depending on the type of uncertainty (after Cauvin *et al.,* 2008)

4. Research trends for risk analysis

Research trends concern all of the features of the risk. It is always possible to go further in the analysis, the assessment of the occurrence probability and the intensity of a hazard. A large number of researchers deal with this matter. In this section, two different topics are specifically investigated.

The first is the use of multi-criteria methods for risk assessment. These methods display different advantages that can be useful to improve risk management and to take into account expert or data uncertainties, while they have been widely used for decision-aid problems.

The second topic is the assessment of the physical vulnerability, i.e., the assessment of the physical expected losses, such as damages to buildings and infrastructures. An interesting approach is the use of vulnerability and fragility functions. These functions are already widely used in seismic risk assessment and are now being developed for other hazards. These functions offer an effective response to the problematic uncertainties regarding damage assessment because they are probabilistic in nature. Moreover, these functions are easy to compile into software, and they can then be used to develop a probabilistic assessment of the vulnerability.

4.1 Multi-criteria methods for risk assessment

Multi-criteria decision-aid methods that have been used and developed in and for economics are now more widely used in all types of problems involving a decision process with uncertainties and the knowledge of the preferences of decision makers.

We are introducing here an example of the use of a multi-criteria method, more specifically, a method named ELECTRE, to deal with a risk zoning problem regarding a mining hazard in Lorraine.

4.1.1 Mining hazard in Lorraine (France)

The extensive mining activity in the French “Lorraine” region has created a large number of underground abandoned cavities that are now responsible for mining subsidence events, i.e., significant movements at the surface. These events result in serious damage to housing and other buildings in the area of influence of such movements. Mining subsidence is of a highly accidental nature when it takes place over mines that use the abandoned rooms and pillars method, even though this method should have allowed endless ground stability. Recent cases of mining subsidence (1996, 1997 and 1999) that have taken place in the Lorraine iron mining area highlight the hazard of such mining works when left abandoned.

The subsidence events that have happened in Lorraine have led public authorities to carry out investigations over the entire Lorraine iron-mining field to assess the hazard, vulnerability and risk of the whole territory. The first investigations highlighted the existence of approximately 20 km^2 of urbanised areas undermined by abandoned works consisting of rooms and pillars.

Different strategies and prioritisations have been put forward since 1996 (see Figure 3). There was an initial prioritisation based on a very simple assessment of the hazard (section 4.1.2). Then, a second prioritisation based on a multi-criteria analysis was carried out (section 4.1.3) to handle more sophisticated considerations regarding both the hazard and the vulnerability (Deck *et al.* 2009)

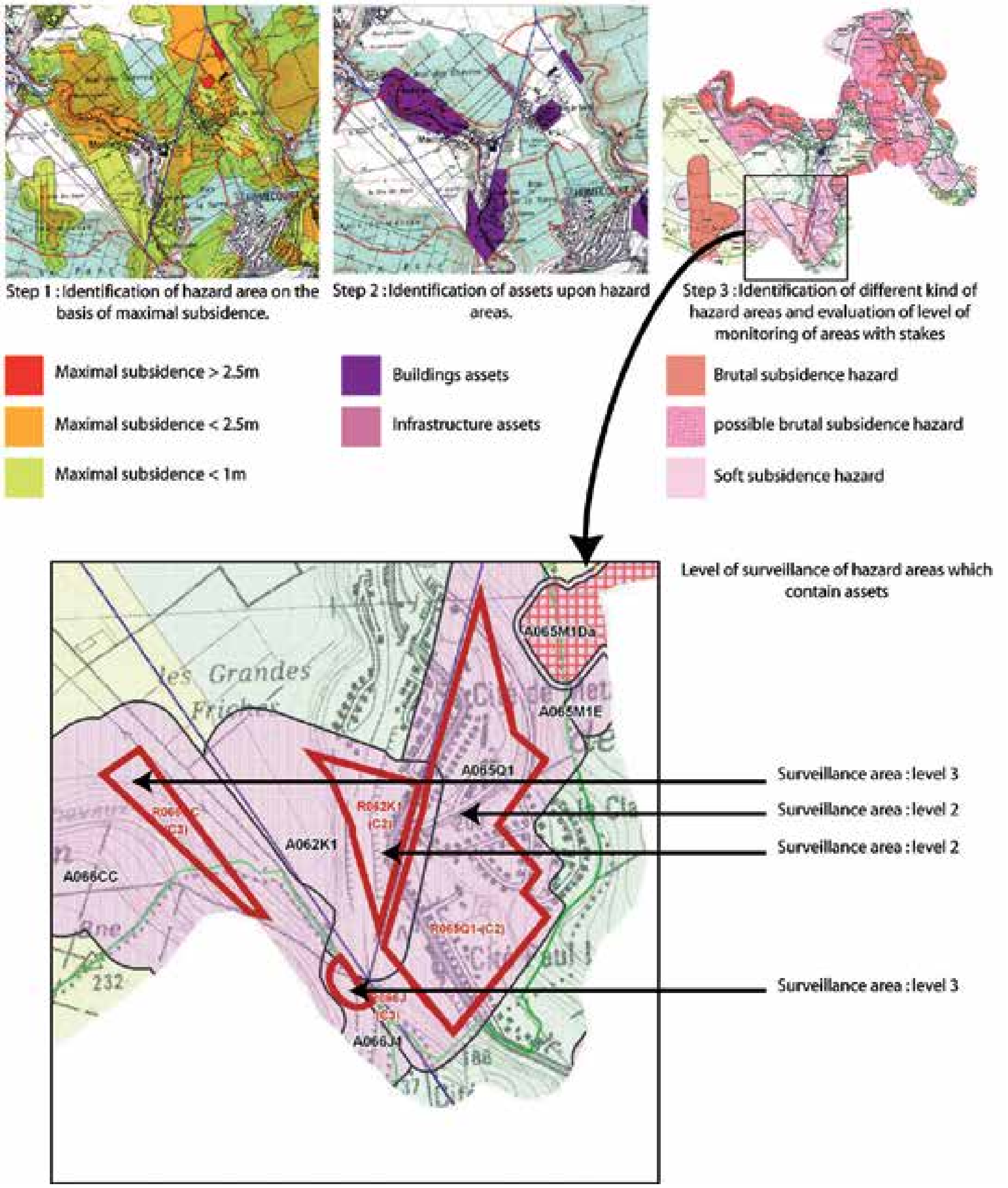

Step 1 : Identification of hazard area on the basis of maximal subsidence.
Step 2 : Identification of assets upon hazard areas.
Step 3 : Identification of different kind of hazard areas and evaluation of level of monitoring of areas with stakes
Maximal subsidence > 2.5m
Maximal subsidence < 2.5m
Maximal subsidence < 1m
Buildings assets
Infrastructure assets
Brutal subsidence hazard
possible brutal subsidence hazard
Soft subsidence hazard
Level of surveillance of hazard areas which contain assets
A065M1Da
A065M1E
A065Q1
A062K1
A066CC
A066JJ
R062K1 (C2)
R065Q1-(C2)
les Grandes Friches
232
Surveillance area : level 3
Surveillance area : level 2
Surveillance area : level 2
Surveillance area : level 3

Fig. 3. Development of risk assessment in the Lorraine iron-ore field.

4.1.2 First ranking system

As a result of the subsidence events in Lorraine in 1996 (Auboué), 1997 (Moutiers), and 1999 (Roncourt), as well as sinkholes in 1998 (Moyeuvre), public authorities ordered investigations to understand the extent of the problem.

Because of the lack of knowledge regarding subsidence phenomena, the first ranking system was based upon two main considerations:

- - the subsidence probability based on the value of the extraction ratio;
- - the subsidence intensity based on the value of the maximum expected vertical subsidence.

The first ranking system was in agreement with the regulations of the French urban code (articles R111-2 and R111-3). The system defines three types of areas depending on the maximum possible subsidence and the associated recommendations for building projects, see Table 2.

Maximum subsidence < 1 m	Surface of building < 400 m^2, Maximum length < 25 m, Number of floors ≤ ground floor + 3
Maximum subsidence < 2.5 m	Surface of building < 150 m^2, Maximum length < 15 m, Number of floors ≤ ground floor + 1
Maximum subsidence > 2.5 m	Forbidden

Table 2. The first ranking system used in Lorraine regarding the mining subsidence hazard and corresponding recommendations.

This first ranking system is called step 1 in Figure 3. This step deals mainly with the urban side of the vulnerability and is not suitable for the other aspects of vulnerability, especially human, social, and economic vulnerability.

4.1.3 Multi-criteria ranking system

Supplementary investigations were necessary to perform a more detailed hazard assessment and for the human vulnerability assessment. We developed (Merad *et al.*, 2004) a method based upon a multi-criteria analysis using the ELECTRE methodology (Roy, 1985; Roy and Bouyssou, 1993). The mathematical functions included in the method allow the management of a "*complex decision-making problem where the available information is uncertain and imprecise and where knowledge is incomplete*" (Merad *et al.*, 2004). This method uses weighted factors for all criteria, and it stresses their relative importance in risk assessment.

In this method, each studied zone is described by a set of criteria that characterises the hazard (probability and intensity of the subsidence) and the vulnerability. Each criterion is expressed in its own unit, and weighting factors are taken into consideration. What is unique in the multi-criteria ELECTRE methods is the specific way used to combine the criteria in the global assessment (here risk assessment) process. Indeed, the global level of risk does not rely on a mean value computed over the criteria but on the comparison of the studied zones in pairs or with predetermined virtual zones characterising the limits between the four risk classes or levels.

More practically, in the case of the ELECTRE III method, all zones are compared with each other based on all criteria. A zone is then considered as more risky than another if a majority of criteria (weighted factors taken into account) give a higher risk for this zone compared with the other. Such a comparison in pairs leads to a global order of the zones from the most risky to the least. The main disadvantage of this method is that any new analysed zone may change the global order already obtained, which is not easy to manage in practice when the objective is to decide actions to be taken on the most risky zones first.

Therefore, in the Lorraine case, the ELECTRE TRI method has been used. This method consists of comparing each studied zone to 3 predefined virtual zones that a group of experts has considered as the limiting zones between each level (group) of risk. Practically, as illustrated in Figure 4, if you call Zr1 the virtual zone that experts consider to be the limit for all criteria between the level of risk 2 and the level of risk 1 (the most risky), and if you want to classify a new zone Zi (dashed line in the Figure) relative to Zr1 (resp. Zr2, Zr3), you have to compare Zi with Zr1 (Zr2 and Zr3, respectively) regarding all criteria, and you decide that Zi is more risky than Zr1 (Zr2 and Zr1, respectively) if a majority of criteria confirm that decision.

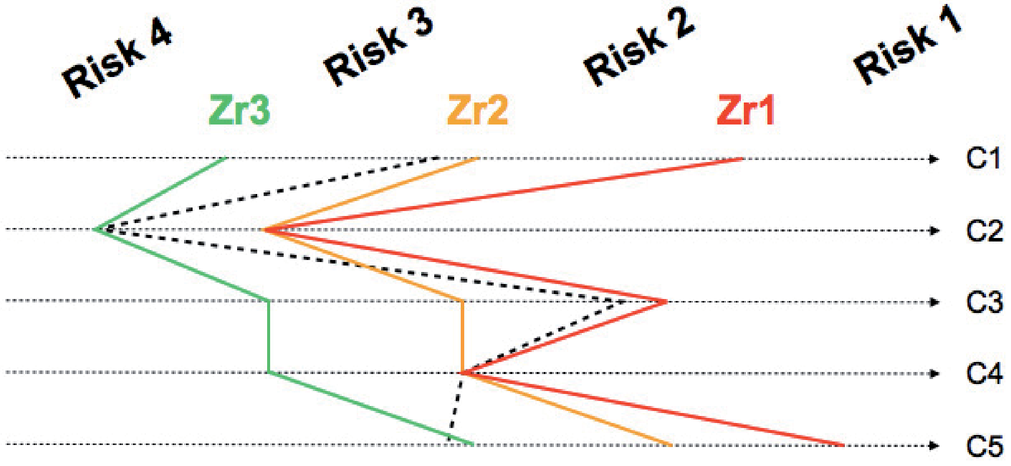


Fig. 4. Principle of virtual limiting Zones (Zr1, Zr2 and Zr3) between classes of risk and comparison of a zone Zi (dashed line) with them

This explanation of the ELECTRE method is a bit simplistic. In fact, the method offers many nuances, which can be found in (Merad *et al*. 2004). One is the way the method deals with uncertainties.

Because of uncertainties, it is not always easy in practice to say whether the value of a criterion for the studied zone Zi is really higher than the value of the same criterion for the reference zone, Zri. ELECTRE methods deal with such problems by the use of fuzzy logic so that the limits between the risk levels become fuzzy intervals (Figure 5, top).

Figure 5 illustrates the principles of dealing with uncertainties in the ELECTRE TRI method. At the bottom of the figure, we can see how a zone Zi is compared with a virtual zone Zri regarding the criterion j. The value c_g(Zi S Zri), called the local agreement index, is a value between 0 and 1 that indicates to what degree we can consider that Zi is more risky than Zri regarding the selected criterion j (x-axis). The figure shows that we consider a certain uncertainty characterised by both p and q so that Zi is considered more risky than Zri when the value of the selected criterion on Zi (called g_j(Zi)) starts to be higher than g_j(Zri)-p. Then, the weighted mean of the local agreement indices over all criteria provide a global agreement index (value between 0 and 1), which indicates whether Zi is globally more risky than Zri. The same calculations have to be performed to reversely compare Zri and Zi because this comparison is not symmetrical (i.e., Zi can be considered as more risky than Zri, and Zri can also be considered as more risky than Zi because of uncertainties). Finally, the methods can state the risk group to which any zone Zi belongs.

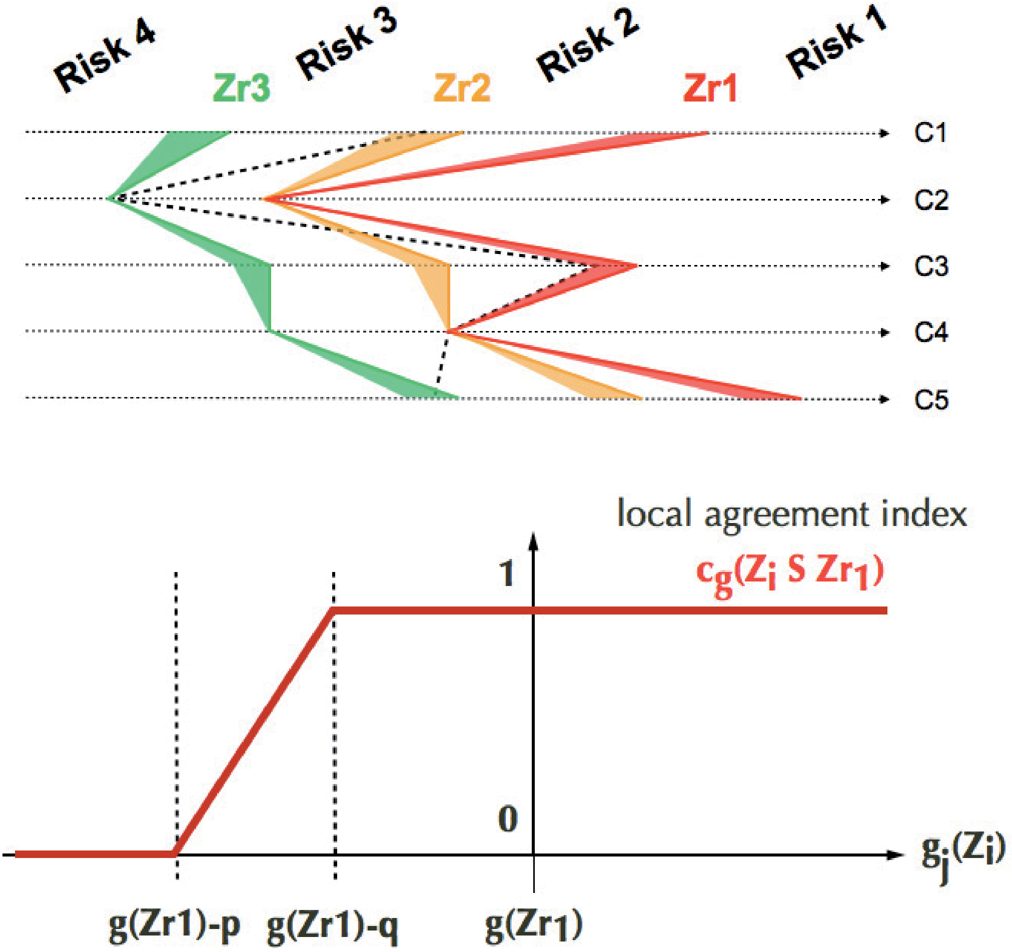


Fig. 5. How the ELECTRE TRI method deals with uncertainties. g_j is the value on criterion j, q is the indifference threshold and p is the preference threshold. The top chart shows the uncertainty on each criterion as a band at the left of each Zri profile.

This methodology has been applied to constructed zones (civil security objectives), where the main objective was to identify zones requiring specific surveillance because of their high-risk level. Therefore, the studied zones have been placed in one of the risk groups given in Table 3. An extension has then been applied to non-urbanised zones from a land-use planning perspective.

Level 1: Very high risk	Real-time monitoring
Level 2: High risk	Periodic monitoring, which will become real-time at the first forewarning.
Level 3: Medium risk	Supplementary investigations are required to assess the need for periodic monitoring.
Level 4: Slight risk	No monitoring is required. Only levelling measurements are made.

Table 3. Second ranking system used in Lorraine regarding the mining subsidence hazard and the corresponding recommendations in terms of monitoring.

Regarding the vulnerability assessment within the risk analysis method, two kinds of assets were considered: buildings and infrastructures. Figure 6 shows the used criteria and their weight factors in the analysis.

In the case of buildings, no other asset is taken into account. The buildings' assets are then assessed with one criterion that may have 5 values, from "business park", which corresponds to a small vulnerability level because of its single daily activity, to "city", which corresponds to the highest vulnerability level because of its daily and nightly activities and the potential number of affected people. This classification reflects the population vulnerability and, to a lesser degree, the economic or structural vulnerability, as these kinds of vulnerabilities are indirectly taken into account because they increase with population.

The weight factors linked with the probability, intensity and vulnerability criteria raise a question related to the previously given definitions of risk. If risk is the product of hazard and vulnerability, does the sum of the weight factors for each component have to be equal? In a multi-criteria ranking system, the sum of the weight factors related to the hazard criteria reach the value of 46, while the sum of the weight factors related to the vulnerability criteria reaches "only" a value between 2 and 14, depending on the assets in the area. This difference produces results that are more dependent on hazard than on vulnerability. Consequently, this multi-criteria ranking system may be criticised because it focuses on hazard assessment rather than on risk assessment.

Apart from this problem, we are convinced that multi-criteria methods are exceptionally well suited to risk analysis because they easily accommodate uncertainties and inaccuracies, and they can easily take into account experts' or decision makers' opinions in the evaluation process.

This section shows how the uncertainties influence may be taken into account with multi-criteria methods in order to develop operational tools for experts or decision makers. However, the final results are mostly dependent on the uncertainties about each component of the risk assessment. For instance, the vulnerability assessment is very simple in the presented case and strictly restricted to the presence of assets in the studied area. For this reason, the next section focuses on the vulnerability assessment and the development of vulnerability functions that take into account some uncertainties.

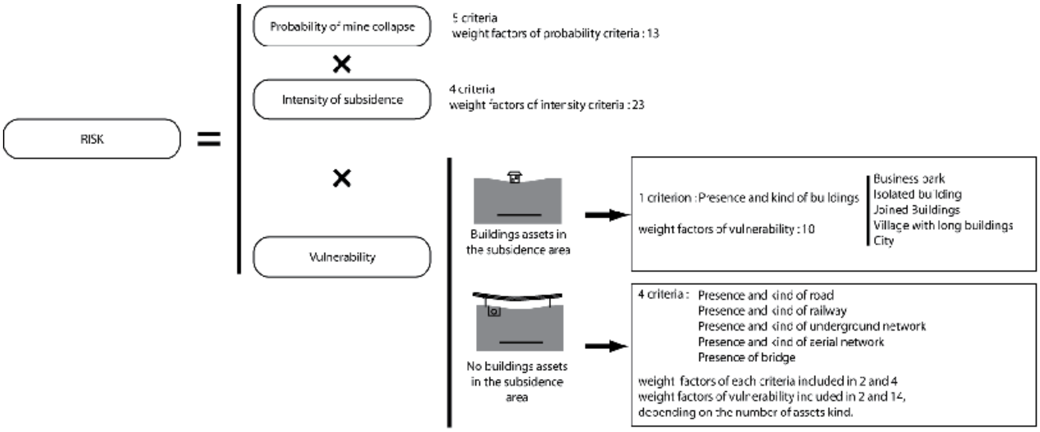

RISK
=
Probability of mine collapse
5 criteria
weight factors of probability criteria : 13
×
Intensity of subsidence
4 criteria
weight factors of intensity criteria : 23
×
Vulnerability
Buildings assets in the subsidence area
1 criterion : Presence and kind of buildings
weight factors of vulnerability : 10
Business park
Isolated building
Joined Buildings
Village with long buildings
City
No buildings assets in the subsidence area
4 criteria : Presence and kind of road
Presence and kind of railway
Presence and kind of underground network
Presence and kind of aerial network
Presence of bridge
weight factors of each criteria included in 2 and 4
weight factors of vulnerability included in 2 and 14, depending on the number of assets kind.

Fig. 6. Criteria and weight factors used in the actual assessment of risk, hazard, and vulnerability in the Lorraine region regarding the iron field mining subsidence hazards.

4.2 Building vulnerability

Today, building vulnerability is primarily investigated with vulnerability and fragility curves. These curves are a powerful way to manage data and model uncertainties associated with the assessment of building damage. Many curves have been empirically developed for different hazards based on case studies and statistical analyses of existing damage. However, each curve may be specific for a given place and a given event, and their use in another context may be contestable. An interesting way to strengthen empirical curves and reduce uncertainties associated with the choice of vulnerability curves is to develop theoretical curves that may be calibrated for any specific context. Moreover, such theoretical curves can then be used for sensitivity analysis and assessment of the dependency on data and model uncertainties. The present chapter illustrates a methodology to develop theoretical vulnerability curves based on Monte-Carlo simulations within the context of mining subsidence hazard.

4.2.1 Concept of vulnerability curves

The vulnerability of buildings and territories to natural hazards is often studied with vulnerability and fragility curves that allow assessment of the damage distribution for a given number of building types in relation to the event intensity, see e.g., earthquakes (HAZUS 1999, McGuire 2004), flooding (USACE 1996, Jonkman *et al.* 2008, HAZUS), volcanoes (Spence, 2005), and tsunamis (Ronald and Hope 2008). Fragility and vulnerability curves are thus developed for a given building type, and they allow quick and realistic damage assessment of all buildings grouped into the same type.

Vulnerability and fragility curves use the following three main types of input data:

1. A damage scale
2. A building typology
3. An intensity criterion

For example, the EMS (Grunthal, 1998) considers a six-level damage scale that consists of no damage (D0), slight damage (D1), and so on, up to very heavy damage (D5). Most of the existing methods define an equivalent number (four levels of damage in the HAZUS and six levels in volcanic risk assessment (Spence *et al.* 2005)).

The building typology must be defined according to the most important parameters relevant to the resistance of the buildings against the considered hazard. For instance, the building materials (e.g., concrete, wood, and masonry), the quality of the construction, the type of foundations, and the global stiffness of the building are important in earthquake engineering. For example, the EMS (Grunthal, 1998) considers 11 main building types.

The criterion for the event intensity may be a physical parameter (height or speed for a tsunami or acceleration for an earthquake) or an empirical one (earthquake intensity in EMS).

Fragility curves provide the probability of reaching or exceeding a given damage state as a function of the intensity of the natural event (see Figure 7b), and they are usually modelled by lognormal functions. A crucial point is that fragility curves clearly take into account that not all buildings of the same type will suffer the same level of damage for a given event intensity, which can be interpreted as the consequences of the uncertainties about the data of both the building characteristics and the hazard intensity.

Vulnerability curves are relationships between the mean amount of damage for a given type of building and the value of the event intensity (see Figure 7c). Vulnerability curves may be deduced from fragility curves with Eq. 1:

$$\mu_D = \sum P_k \cdot D_k \tag{1}$$

where μ_{D} is the mean damage for a given intensity, P_{k} is the probability of a damage grade D_{k}, and k is the range of damage category (from 0 to 5 in the EMS damage scale, for instance).

The example shown in Figure 7 concerns a massive stone masonry building (type M4), according to the EMS-98. Figure 7a shows the damage distribution for this type of building during an earthquake of intensity 11. This distribution can be plotted in Figure 7b, where each dot on the figure corresponds to the different fragility curves of this type of building. By calculating the mean of the damages (Eq. 1), it is then possible to plot one point of the vulnerability curve, as shown in Figure 7c. Fragility and vulnerability curves may then be modelled by fitted mathematical functions.

In practical terms, when developed and validated, fragility and vulnerability curves are both efficient and accurate. Vulnerability curves are used to obtain a synthetic result of the mean damage to buildings in a selected territory. When applied to a single building, fragility curves may be used to assess the probability of reaching a particular damage level. When applied to a set of buildings, fragility curves may be used to assess the damage distribution of all buildings.

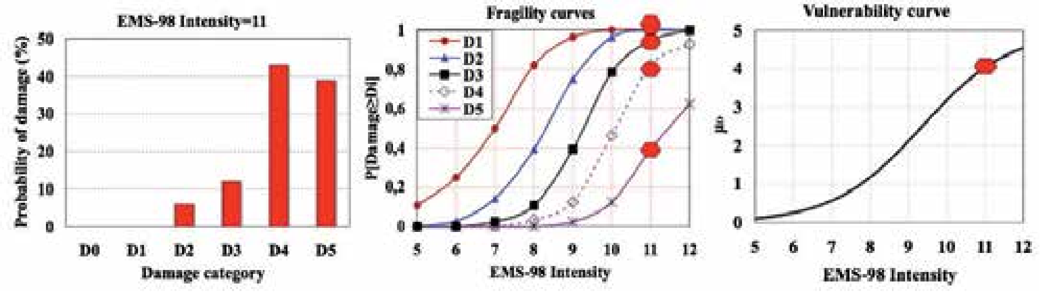


Fig. 7. Damage distribution (a), fragility curves (b), and vulnerability curves (c) for the M4 building type, according to EMS-98 for the assessment of earthquake building damage.

4.2.2 Methodology to define vulnerability and fragility curves

The methodology adopted to develop vulnerability and fragility curves where data are not sufficient in number (no previous event or no statistical data available) is based on the damage assessment of a set of theoretical buildings whose characteristics are consistent with a particular building type but are also variable to take into account the variability in the type and uncertainties (Saeidi *et al.* 2009, 2011). The method is based on five steps (see Figure 8).

The first step consists of making preliminary choices regarding a damage scale, an intensity criterion of the considered hazard and a method for the building damage evaluation.

The second step consists of defining a building typology and choosing the representative characteristics of each type. Each characteristic is supposed to be deterministic or probabilistic, (defined as a range of possible values with an attached probability distribution) to take into account variability into a given building type and uncertainty about its evaluation.

The third and fourth steps consist of a Monte Carlo simulation. For each type, the third step consists of simulating a database of 1000 virtual buildings whose characteristics (e.g., height, length, materials, and mechanical properties) are consistent with the studied building type. The fourth step consists of evaluating the damage of the 1000 simulated buildings for one value of the intensity criterion and counting the number of buildings in each damage class. The results may then be used to plot a set of points for both the fragility curve (probability of reaching or exceeding a given damage class) and the vulnerability curve (mean damage). Finally, by repeating this step for all of the values of the intensity criterion, both the vulnerability and fragility curves can be drawn.

The fifth step consists of fitting a mathematical model to the results to express the fragility and vulnerability as mathematical functions. A tangent hyperbolic function can be used for the vulnerability functions and a lognormal function for the fragility functions.

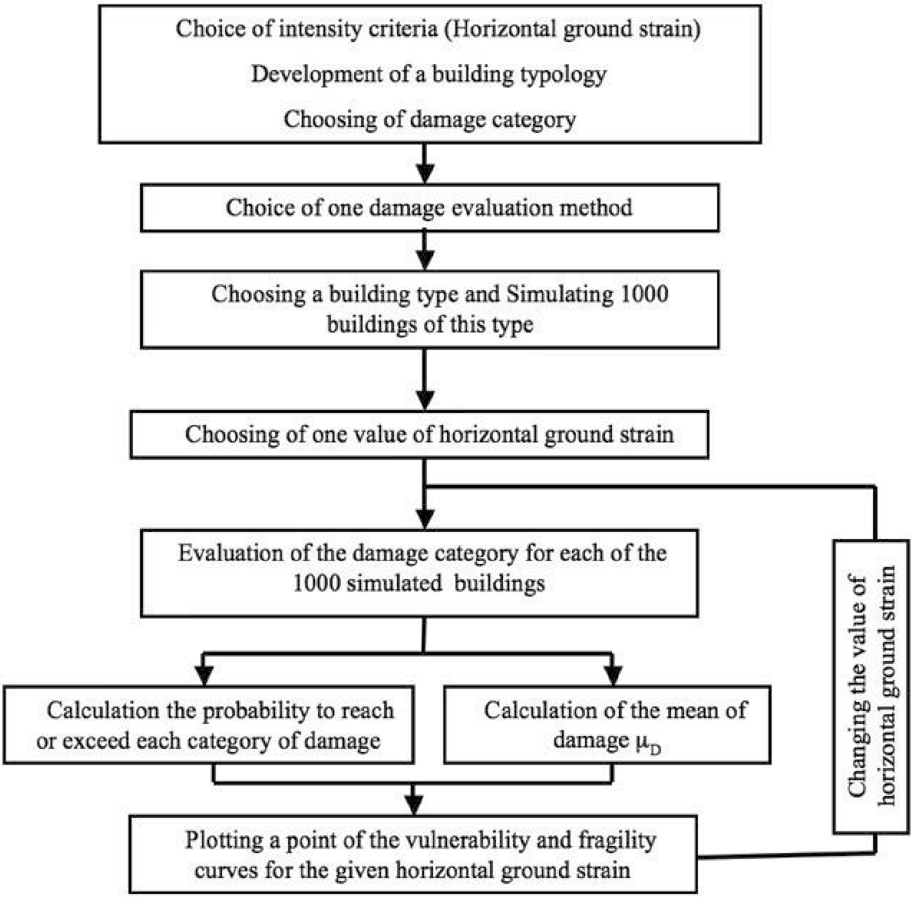

Choice of intensity criteria (Horizontal ground strain)
Development of a building typology
Choosing of damage category
Choice of one damage evaluation method
Choosing a building type and Simulating 1000 buildings of this type
Choosing of one value of horizontal ground strain
Evaluation of the damage category for each of the 1000 simulated buildings
Calculation the probability to reach or exceed each category of damage
Calculation of the mean of damage μ_D
Plotting a point of the vulnerability and fragility curves for the given horizontal ground strain
Changing the value of horizontal ground strain

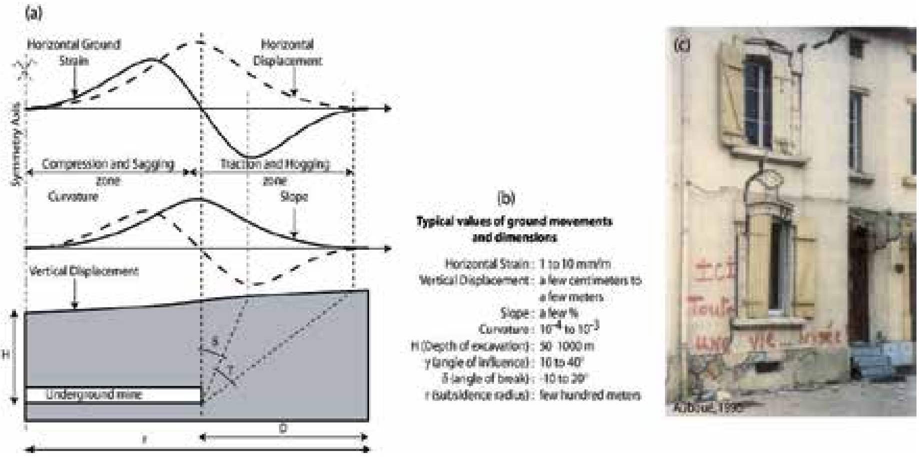
Fig. 8. Methodology for the determination of vulnerability and fragility curves in the subsidence zone.

Next, we illustrate this approach on the Lorraine case.

4.2.3 Application

The Lorraine region features large quantities of iron, salt, and coal deposits that were heavily extracted until the beginning of the 1990s (salt continues to be mined here). The presence of former iron mines raises many issues, including that of building vulnerability. Five subsidence events occurred between 1996 and 1999 (two in the city of Auboué in 1996, two in Moutiers in 1997, and the last in Roncourt in 1999), which caused damage to more than 500 dwellings (see Figure 9). Many other cities and villages in this area may still be affected by this phenomenon. The described methodology is applied to develop vulnerability curves within the context of mining subsidence hazard.

Mining subsidence is a consequence of the collapse of underground mines and quarries. It produces significant horizontal and vertical movements at the ground surface. The maximum value "Sm" of the vertical subsidence is usually considered as a characteristic of the trough, but the horizontal ground strain "ε" is mostly correlated to the associated structural damage. Consequently, the horizontal ground strain that involves from an horizontal compression near the centre of the subsidence to an horizontal extension near the edges is used to assess the hazard intensity and the associated potential damage. The assessment of ε is possible with different methods (Whittaker and Reddish 1989).



Typical values of ground movements and dimensions

Horizontal Strain :	1 to 10 mm/m
Vertical Displacement :	a few centimeters to a few meters
Slope :	a few %
Curvature :	10^{-4} to 10^{-3}
H (Depth of excavation) :	50 1000 m
γ (angle of influence) :	10 to 40°
δ (angle of break) :	-10 to 20°
r (subsidence radius) :	few hundred meters

Fig. 9. Description of the main characteristics involved in mining subsidence and associated consequences (Saeidi *et al.* 2009). a) Typical profiles of the ground displacements and locations of the compression/sagging and traction/hogging areas. b) Typical values of the subsidence dimension and grounds movements. c) Typical damage due to mining subsidence in the city of Auboué, France.

The methodology for the development of vulnerability curves requires the use of a damage assessment method. Many methods have been developed within the context of mining subsidence hazards. Some methods are empirical (NCB, 1975; Wagner and Schumann, 1991; Yu *et al.*, 1988; Bhattacharya and Singh, 1984; Dzegniuk and Hejmanowski, 2000; Kwiatek, 1998) and based on retro-analysis, while others are analytical and based on mechanical models (Boscardin and Cording, 1989; Boone, 1996; Burland, 1997; Boone, 2001; Finno *et al.* 2005). Most of these methods use a four- or five-level damage scale, and they consider the horizontal ground strain for the hazard intensity.

In the following, the methodology is illustrated with the most simple but popular method of the (NCB 1975), which links building damage (five levels) to the building length and the horizontal ground strain. Other methods can also be used (Saeidi *et al.*, 2009, 2011), and the results will be compared and discussed.

The second step requires defining a building typology. Most of the buildings in the Lorraine region are small masonry buildings with or without reinforcements and can be classified into five types (four masonry building types MR1 to 3 and MC1 and one concrete frame building CF1). For example, the MR2 building type consists of a 10 to 20 metres long masonry of rubble stones, a height of 5 to 8 metres, poor quality mortar without protection against mining subsidence effects, a cellar and concrete slab and a simple external shape with good symmetry of the bearing walls.

The third step consists of the simulation of a database with 1000 theoretical buildings. To complete this database, the variability of each criteria used in the different methods of damage assessment is considered to be in agreement with the building type. A uniform distribution is used to define the final value of each building. This variability within a building type may be interpreted both as a real physical and observed difference between the buildings and as uncertainties concerning their real characteristics.

For example, when the NCB method is considered, the 1000 theoretical buildings only differ in length, which is randomly chosen between 10 and 20 m. Therefore, the development of vulnerability functions requires a proper definition of the variability of each parameter used by the considered method. Preliminary tests, with a number of buildings between 200 and 2000, showed that 1000 buildings provided acceptably accurate results.

$$P(D_i) = \frac{N(D_i)}{n} \tag{2}$$

where $N(D_\mathrm{i})$ is the number of buildings in the damage class "D_i" and "n" is the total number of buildings (1000 in this example).

The vulnerability curve is the relationship between the mean damage and the horizontal ground strain and is calculated with Eq. 3.

$$\mu_D(\varepsilon) = \sum_{i=1}^{4} P(D_i) \cdot D_i \tag{3}$$

where $\mu_\mathrm{D}(\varepsilon)$ is the mean of damages for the value "ε" of horizontal ground strain and $P(D\mathrm{i})$ is the probability of damage in the class "Di", as calculated by Eq. 2.

The plot of mean damages is given in Figure 10 and shows a discontinuous curve that is the consequence of using threshold values in all of the empirical methods. This result is hardly compatible with reality because damage should continuously increase with increasing horizontal ground strain. This assumption is also corroborated by the shape of all vulnerability functions developed in other fields, where a tangent hyperbolic function is often used (Lagomarsino *et al.*, 2006). To determine a continuous building vulnerability curve in agreement with the discontinuous curve previously plotted in Figure 10, a tangent hyperbolic function may be fitted on data according to Eq. 4

$$\mu_D(\varepsilon) = a[b + Tanh(c \cdot \varepsilon + d)] \tag{4}$$

where $\mu_\mathrm{D}(\varepsilon)$ is the mean of damages for a value "ε" of the horizontal ground strain and *a, b, c,* and *d* are four coefficients that must be determined for each building type.

These parameters are not independent; two relationships exist between them. According to Table 3, for a horizontal ground strain equal to zero, there is no damage to buildings, and for a horizontal ground strain greater than 9 mm/m, the mean damage to buildings is maximum and equal to four (greatest level in the damage scale). Therefore, this leads to the two boundary conditions detailed in Eq. 5, and only two parameters must still be determined. We used a nonlinear regression method to find the best values of these two parameters. The final continuous vulnerability curve for the "CF1" building type is shown in Figure 8.

$$\begin{cases} \mu_D(0) = 1 \\ \mu_D(9) = 4 \end{cases} \Rightarrow \begin{cases} a = \dfrac{1}{[(\dfrac{Tanh(d+9\cdot c)-4\cdot Tanh(d)}{3}) + Tanh(d)]} \\ b = \dfrac{Tanh(d+9\cdot c)-4\cdot Tanh(d)}{3} \end{cases} \quad (5)$$

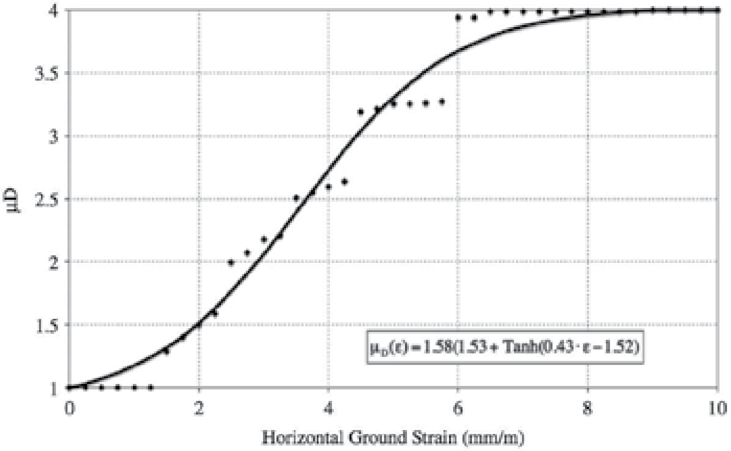


Fig. 10. Vulnerability function and curve for the CF1 building type, built from Table 7.

The next step in determining vulnerability and fragility curves is the damage assessment for all theoretical buildings for different values of the horizontal ground strain between 0 and 10 mm/m.

The damage level of each building is calculated for the different values of horizontal ground strain, and the probability of damage in each damage class "P(D_i)" is calculated with Eq. 3.

The influence of the damage assessment method is investigated in Figure 11, and it is a good way to assess the model uncertainty if no arguments are available to privilege one method. The results of the vulnerability curves for the MR2 building type (Table 5) obtained with different methods show significant differences. In particular, the NCB method gives less damage than the other methods; consequently, this method is considered less conservative. A mean vulnerability curve can be calculated. Unless the user has scientific arguments for justifying one method by considering the special features of the studied case, it may be concluded that the mean method MD(ε) gives the most probable damage assessment.

The results are then compared with empirical data of damaged buildings in Lorraine (see Figure 11). This comparison shows that there is a good agreement between the observations and the calculated vulnerability curves. Nevertheless, some differences remain for the lowest and greatest values of the horizontal ground strain. The vulnerability functions underestimate the damage for the lowest values of the horizontal ground strain (less than 3 mm/m), and they overestimate the damage for the greatest values (greater than 9 mm/m). One possible explanation for these differences is the existence of preliminary building damage due to building aging and other building pathologies (e.g., settlement during construction). For the greatest values of the horizontal ground strain, the overestimation of the damage shows that there are always some building types that are stronger than the predicted resistance.

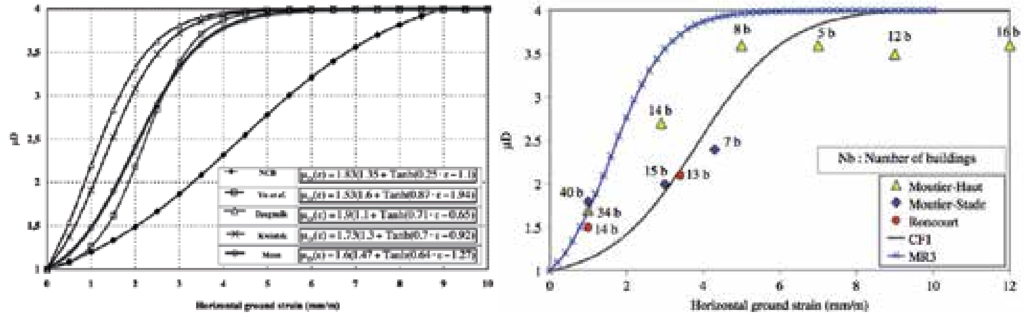


Fig. 11. Vulnerability curves for MR2 type by all methods (left) and comparison with observed damaged buildings in Lorraine (right).

In conclusion, vulnerability functions are a powerful and operational way to assess building vulnerability over a large territory. If statistical analysis of previous damage due to a specific hazard is the easiest method to get and validated such an approach, other methods must be developed when such data are missing. This example presents an innovative way to develop such functions. Based on Monte Carlo simulations, the method allows taking into account both variability and uncertainties. However, some steps of the methodology can be questioned, in particular when the 1000 theoretical buildings are simulated. First the choice of the statistical distribution for each parameter is not evidence; secondly the method actually doesn’t consider correlations between parameters (e.g. long buildings may also be the highest for instance). This should be addressed in future work.

5. Closure and future work

Risk analysis is a complex field because of the numerous different forces that interact to assess the risk and to define solutions for the risk management. Public authorities, scientists and citizens may have different objectives, and they seldom use the same definitions. Consequently, a clear definition of the terms used in risk analysis is a preliminary step before analysing other sources of uncertainties.

Uncertainties may then be classified into various types that highlight different possible features, such as the possibility to quantify them or not with statistical or probabilistic distributions, fuzzy logic or any other method. Finally, many sources of uncertainties may be identified, for which different solutions can be suggested.

Two examples have been investigated. The first deals with a decision-aid method that is an interesting approach to managing expertise uncertainties. Based on multi-criteria analysis, initially used and developed in and for economics, this decision-aid method has been developed for the risk management of areas vulnerable to mining subsidence hazards. The second example deals with vulnerability assessment and, more specifically, with building damage assessment. The use and development of vulnerability functions are an interesting way to manage both model and data uncertainties.

In sum, these two examples illustrate how uncertainties can be taken into account for risk management related to geo-hazards. However, it is obvious that uncertainties may complicate the risk assessment that must be understood and discussed amongst public authorities, scientists and citizens. A difficulty, then, is the ability of scientists to both manage uncertainties, with more or less complex methods, and communicate the results to allow for good understanding and use by public authorities and citizens without oversimplifications.

6. References

Ang A. H-S. & Tang W.H., (1984) Probability concepts in Engineering Planning and Design, Vol2 – Decision, Risk and Reliability, John Wiley & Sons, New York.

Aven T. , On how to define, understand and describe risk, Reliability Engineering & System Safety, Volume 95, Issue 6, June 2010, Pages 623-631.

Baecher G.B. & Christian J.T. (2000), Natural Variation, Limited Knowledge, and the Nature of Uncertainty in Risk Analysis. Presentation to The 25th Anniversary Engineering Foundation Conference on Risk and Uncertainty in Water Resources Engineering, Santa Barbara, October, 2000.

Baecher G.B., Christian J.T. (2002), The concept of uncertainty in geotechnical reliability, Not published.

Baecher G.B. & Christian J.T. (2003), Reliability and Statistics in Geotechnical Engineering, John Wiley & Sons, Ltd.

Balandier P. (2004). Urbanisme et Aménagements – Objectifs et problématiques. Collection Conception Parasismique

Benjamin J.R. & Cornell C.A., (1970) Probability, Statistics and Decision for Civil Engineers, Mc Graw Hill, New York.

Bhattacharya S, Singh MM. Proposed criteria for subsidence damage to buildings. Rock mechanics in productivity and protection, 25th Symposium on rock mechanics 1984;747-755.

Boone, 2001

Bogardi J.J. (2004). Hazards, risks and vulnerabilities in a changing environment: the unexpected onslaught on human security. Global Environmental Change, 14, pp. 361-365.

Boscardin MD, Cording EJ. Building response to excavation – induced settlement. J. of Geotechnical Engineering 1989 ; 115: 1-21.

Burland JB. Assessment of risk of damage to buildings due to tunnelling and excavation. Earthquake geotechnical engineering, editions Ishihara, Balkema, 1997 ; 1189-1201.

Callon M., Lascoumes P., Barthe Y. (2001), Agir dans un monde incertain - Essai sur la démocratie technique. Seuil, Paris.

Cauvin M., Salmon R., Verdel T. (2008), Dealing with uncertainties in the context of post mining hazard evaluation, Post-Mining 2008, Nancy, February 6-8.

Chacko G.K. (1990), Decision-Making under Uncertainty : An Applied Statistics Approach, Praeger Publishers.

Cox. (2009). What's Wrong with Hazard-Ranking Systems? An Expository Note, Risk Analysis, Vol. 29, No. 7, pp. 940 - 948.

Deck O., Verdel T., Salmon R. (2009). Vulnerability assesment for mining subsidence hazard. International Journal on Risk Analysis – Risk Analysis, Vol. 29, No. 10, 1380-1394.

Dzegniuk B, Hejmanowski R, Sroka A. Evaluation of the damage hazard to building objects on the mining areas considering the deformation course in time. Proceedings of Xth International Congress of the International Society for Mine Surveying, 2-6 November 1997, Fremantle, Western Australia.

Ezell B. C. (2007). Infrastructure vulnerability assessment model (I-VAM). Risk Analysis, Vol. 27, No. 3.

Finno R.J. , Voss F.T., Rossow E., Tanner B. (2005). Evaluating Damage Potential in Buildings Affected by Excavations, Journal of Geotechnical and Geoenvironmental Engineering 131, No. 10, 1199-1210.

Galesne Alain (1996), Choix des investissements dans l'entreprise, Rennes:Cerefia.

Gilbert R.B. & Tang W.H. (1995), Model uncertainty in geotechnical reliability and decision analyses, Proceedings of Applications of Statistics and Probability, Lemaire, Favre & Mébarki (eds). Balkema. pp 109-114.

Griot C. et Ayral P.A. (2001). Terminologie en science du risque, d'après le document de travail du colloque international "Dire le risque : le risque mis en examen", Mèze, 18-20 mai 2001.

Grunthal G. European Macroseismic Scale. Centre Européen de Géodynamique et de Séismologie, Luxembourg. 1998; Vol. 15.

Hacking Ian (1975), L'émergence de la probabilité, (Traduction de l'anglais, Seuil, 2002).

Haimes Y. (2006). On the definition of vulnerabilities in measuring risks to infrastructures. Risk Analysis, Vol. 26, No. 2.

Haimes Y. (2004), Risk Modeling, Assessment, and Management, 2nd revised ed., John Wiley & Sons Inc.

HAZUS. Multi-hazard Loss Estimation Methodology Earthquake Model, Technical and User Manuals. Federal Emergency Management Agency, Washington, DC. 1999, chapter 2.

HAZUS®MH MR4, Multi-hazard Loss Estimation Methodology - Flood Model - Technical Manual

Hofer E. (1996) When to separate uncertainties and when not to separate, Reliability Engineering and System Safety 54, 113-118.

Hoskins C.G. (1973), Distinctions between Risk and Uncertainty. Journal of Business Finance, Spring.

Jonkman S.N., Bockarjova M., Kok M. and Bernardini P. Integrated hydrodynamic and economic modelling of flood damage in the Netherlands, Ecological economics 66 (2008), 77–90.

Karimi I. , Hüllermeier E. . Risk assessment system of natural hazards: A new approach based on fuzzy probability, Fuzzy Sets and Systems 158 (2007) 987 – 999.

Karmakar S., Simonovic S. P., Peck A., Black J. An Information System for Risk-Vulnerability Assessment to Flood, Journal of Geographic Information System, 2010, 2, 129-146 doi:10.4236/jgis.2010.23020.

Klein R.J.T., Nicholls R.J., Thomalla F. (2003). Resilience to natural hazards: how useful is this concept? Enviromental Hazards, 5, pp. 35-45.

Knight F.H. (1921) Risk, Uncertainty, and Profit. University of Chicago Press.

Kwiatek J. Protection of constructions on surface ground mine. (Traduction in poland "Ochrona objektow budowlanych na terenach gorniczyych"). GIG, Katowice, 1998.

Lagomarsino S, Giovinazzi S. (2006). Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings. *Earthquake Engineering* , 4, 415–443.

MATE (1997). Plans de prévention des risques naturels prévisibles – Guide général. La documentation Française

McGill W. L., AYYUB B. M. and Kaminskiy M. (2007). Risk analysis for critical asset protection. Risk Analysis, Vol. 27, No. 5.

McGuire RK. Seismic Hazard and risk analysis. EERI Earthquake Engineering research Institute, 2004.

Merad M.M, Verdel T., Roy B., Kouniali S. (2004), Use of multi-criteria decision-aids for risk zoning and management of large area subjected to mining-induced hazards, Tunneling and Underground Space Technology 19 (2004) 125–138.

National Coal Board. Subsidence engineering handbook. 1975; 45-56 chapter 6.

Neil Adger W. N., Vulnerability, Global Environmental Change 16 (2006) 268–281

NRC (National Research Council, 1996), Understanding risk: Informing decision in a democratic society. Washington, DC, National Academies Press.

Paté-Cornell M.E. (1996) Uncertainties in risk analysis: Six levels of treatment, Reliability Engineering and System Safety 54, 95-111

Romeoa R., Paciellob A., Rinaldis D.. Seismic hazard maps of Italy including site effects, Soil Dynamics and Earthquake Engineering, Volume 20, Issues 1-4, 6 October 2000, Pages 85-92

Ronald TE, Hope AS. Loss Estimation models and metrics Risk Assessment, modeling and Decision Support. 2008;135-157

Roy B. (1985), Méthodologie multicritère d'aide à la décision, Paris, Economica.

Roy B., Bouyssou D. (1993), Aide multicritère à la décision : Méthodes et cas, Paris, Economica.

Saeidi A., Deck O., Verdel T. (2009) - Development of buildings vulnerability functions in subsidence regions from empirical methods – Engineering Structures 31 (2009), 2275-2286.

Saeidi A., Deck O., Verdel T. (2011) - Development of buildings vulnerability functions in subsidence regions from analytical methods – Geotechnique, A paraître

Schmidtlein M. C., Deutsch R. C., Piegorsch W. W. and Cutter S. L. (2008). A sensitivity analysis of the social vulnerability index. Risk Analysis, Vol. 28, No. 4.

Smith R.P. (2000), Risk, Uncertainty and Ambiguity in Engineering Design Decision Making, The Open Workshop on Decision-Based Design: Origin, Status, Promise, and Future, NSF, State University of New York at Buffalo. (http://dbd.eng.buffalo.edu/papers/rpsmith.html).

Spence RJS, Kelman I, Baxter PJ, Zuccaro G, Petrazzuoli S. Residential building and occupant vulnerability to tephra fall. Journal of Natural Hazards and Earth System Sciences 2005 ; 5 : 477-494.

Suddle S. , The weighted risk analysis, Safety Science, Volume 47, Issue 5, May 2009, Pages 668-679.

UN/ISDR (2004). Living with Risk - A global review of disaster reduction initiatives (2004 version).

USACE (US Army Corps of Engineers), 1996. Risk-based analysis for flood damage reduction studies. Report EM 1110-2-1619, 63.

Veneziano, D. (1994). "Uncertainty and Expert Opinion in Geologic Hazards." The Earth, Engi-neers, and Education: A symposium in Honor of Robert V. Whitman, Cambridge, Department of Civil and Environmental Engineering, MIT: 102-124

Wagner H, Schumann EHR. Surface effect of total coal seam extractions by underground mining methods. J.S.Afr.Inst.Min.Metal 1991 ; 91:221-231.

Wahlström R. Grünthal G., Probabilistic seismic hazard assessment (horizontal PGA) for Sweden, Finland and Denmark using different logic tree approaches, Soil Dynamics and Earthquake Engineering, Volume 20, Issues 1-4, 6 October 2000, 45-58.

Whittaker B.N. et Reddish D.J. (1989). Subsidences : Occurrence, Prediction, Control. Editions Elsevier.

Yu Z, Karmis M, Jarosz A, Haycocks, C. Development of damage criteria for buildings affected by mining subsidence. 6th annual workshop generic mineral technology centre mine system design and ground control 1988; 83-92.

Zhai G., Fukuzono T. and Ikeda S. Modeling flood damage: case of Tokai flood 2000. Journal of the American Water Resources Association, february 2005, 77-92.

9

A Monte Carlo Simulation and Fuzzy Delphi-Based Approach to Valuing Real Options in Engineering Fields

Roberta Pellegrino and Nicola Costantino
Politecnico di Bari – Dipartimento di Ingegneria Meccanica e Gestionale
Italy

1. Introduction

The success of a firm depends on its ability to manage uncertainty of investment projects and strategies it decides to make or develop. During the management of new projects, routines and technologies, its constant objective should be to earn increasing returns by exploiting opportunities and limiting losses that could be created by uncertainty. Thus, firms must carefully recognize and evaluate the actions to be taken to respond to uncertainty. To help managers in their decision-making process in uncertain environments, new techniques and theories have been developed. One of them is the real option theory, where a real option is the right, not the obligation, to take some action in the future (Dixit and Pindyck, 1995). The formal approach, which originated from financial models, deals with future uncertainty and opportunities a firm can seize, and aims at valuing the flexibility that often managers have to "react" to uncertainty. In this sense, the real option potential to estimate the value of this flexibility is appealing for managers. As Leslie and Michaels (1997) report, in fact, over the past years, the theory has drawn a growing body of literature and has gathered support across the business world in academia, consulting, and the corporation. Copeland and Weiner (1990) of McKinsey observe that the "use of options methodology gives managers a better handle on uncertainty".

Despite the growing support the real option theory has been attracting in academia and its apparent relevance in business decisions, few corporate managers and practitioners have truly recognized or applied the power of real options in managing their businesses (Leslie and Michaels, 1997; Lander and Pinches, 1998). In other words, the application of real options to managerial practice is poor, and is often limited to a conceptual level. Several reasons could explain why real options are not widely used in practice, as some studies analyzed (Lander and Pinches, 1998; Borison, 2003; and others). Anyway, all these reasons could be traced back to a fundamental issue, that is, the "financial" origin of the real option theory and their evaluation models.

From a practical standpoint, real option modeling learned from the financial world is not easy to use or implement as real world cases are often reasonably sophisticated and complex. The inputs required for the application of these models (with financial features) are often not defined or easy to define in the real world. Thus, in order to adapt financial option-based models to the real world, the complexity of these models must be reduced so that they become mathematically tractable. This implies the need for making a variety of simplifying assumptions. For example, option-based models are severely limited if there are more than one or two sources of uncertainty. In the presence of two or more uncertainties, therefore, they are usually modeled in such a way that they are combined and then treated as a single uncertainty. Furthermore, in real world cases these uncertainties have technical (engineering) - economical features rather than financial ones. These and other "adaptation assumptions" make Real Option Pricing (ROP) models a black box for managers, and *de facto* limit their transparency.

The present chapter presents the results of a research aimed to develop a methodological approach to analyze and assess real options in real world investment opportunities. It combines in a consistent and original way two well-known techniques, namely the Monte Carlo simulation for real option pricing and the fuzzy-Delphi method for eliciting, when historical data miss, probabilistic input parameters from the knowledge of even more than one expert in a consistent, structured and transparent way.

The chapter is organized as follows. The next section presents an overview on real options, their origin, characteristics as well as the challenges in their use by practitioners. Section 3 presents the new approach based on the Monte Carlo simulation proposed for evaluating managerial flexibility. Section 4 explains how to elicit uncertain parameters from experts when historical data are unavailable. Section 5 provides an application of the proposed approach, and finally conclusions and future works end the chapter.

2. Real option theory: State of art

All projects are subject to uncertainty, arising from diverse sources (including technical, management and commercial issues, both internal and external to the project). Project managers widely recognize that successful management of uncertainty is intimately associated with project success, since the proactive project manager constantly seeks to steer the project towards the achievement of the desired objectives (Hillson, 2002). Investment decisions that are at the core of any development strategy imply - to some extent - taking the hard choice to sink economic resources now, in the hope of future benefits, betting on the distant and uncertain future horizon (European Commission, 2008). Traditional investment decisions are characterized by irreversibility and uncertainty about their future rewards. Once money is spent, it cannot be recovered if the payoffs hoped for do not materialize. These decisions are typically made by using traditional project evaluation approaches, such as those based on Discounted Cash Flows (DCF) analysis. They assume implicitly that a project will be undertaken now and operated on continuously at a set time scale, until the end of its expected useful life, even though the future is uncertain. Therefore, they are "static" and underestimate the upside value of investment (Kogut and Kulatilaka, 1994) by assuming management's passive and inflexible commitment to a certain "operating strategy". They are also "deterministic" since they make implicit assumptions concerning a certain "expected scenario" of cash flows. In the real world, because of uncertainty and competitive interactions, the realization of cash flows will probably differ from what management originally expected.

As new information is available and uncertainty about the market conditions and future cash flows is gradually resolved, management should revise the operating strategy it originally anticipated in order to achieve the initial desired goals (Boute *et al.*, 2004). However, unexpected events during the project management can have a range of effects on achievement of project objectives, from the total disaster to the unexpected welcome surprise. Thus the role of the managerial flexibility to adapt in response to new information is not only limiting downside losses relative to the initial expectations under passive management (uncertainty with negative effects), but also improving the upside potential of the investment by the exploitation of favorable events (uncertainty with positive effects). In a competitive market/environment like the global one, it is unlikely that a project or, more in general, a firm can succeed by formulating and following a detailed long-term plan and operating strategy rigidly (Yeo and Qiu, 2003). As Coy (1999) reported, Hewlett-Packard Co. CEO Lewis E. Platt said, at the end of the last millennium, "anyone who tells you they have a 5- or 10-year plan is probably crazy. This is the age of scenario planning. You need not only speed but agility". Hence, traditional managerial techniques arisen from stable environments are in crisis since they cannot capture the value of the flexibility in changing the operating strategy to capitalize from favorable opportunities or to cut losses in the case of adverse development (Olafsson, 2003). A better valuation approach to support decision making in uncertain environments, indeed, should incorporate the uncertainty and the active decision making required for a strategy to succeed (Luehrman, 1998). Hence, it is essential that flexibility be quantified. Any attempt to quantify this flexibility leads almost naturally to the concept of options (Trigeorgis, 1996), i.e., each source of flexibility is, in technical terms, a "real option".

A lot of progress that have been done in the real option literature have changed the way of thinking about an investment opportunity. During project management, managers may make several choices about project characteristics every time new information from market is available. A real option is the way to respond to market changes. This possibility that managers have to adapt their decisions to the change of market has value that must be considered during the decision making process. In other words, this flexibility creates "options" that increase the value of the project and determines the failure of the traditional techniques. Real options originate from the concept of financial options, since they are defined as options written on real assets. Typical real options are the option to hold or abandon a project, the option to decide the timing of investment, the option to choose the production technology, inputs and outputs, the option to reduce or expand the production capacity (Amram and Kulatilaka, 1999). The two next sections discuss these two kinds of options. Of course, the financial world is not as the real one. Therefore, when the theory is applied to the real world, this difference should be considered. Sections 2.3-2.5 discuss some practical issues concerning the application of techniques traditionally used for financial options to model and value the real ones.

2.1 Financial options

The concept of real option has its roots in the theory of financial options. An option is a contract that gives the holder the right - not the obligation - to buy or sell a predefined quantity of an *underlying asset* at a specific price, called *strike price* or *exercise price*, at or before the expiration date of the option (*maturity*). The holder (buyer) pays a price for this right (Damodaran, 2001). There are two types of options: *call options* and *put options* that respectively give the holder the right to buy or put the underlying asset. There are also *American options* and *European options*: the first can be exercised every time before the expiration date, while the latter can be exercised only at the expiration date. The option (call) will be exercised only if the value of the asset is higher than the strike price, on the contrary the option will never be exercised and will expire worthless.

If the option is exercised, the buyer buys the stock at the strike price; hence, the gross profit of the option is the difference between the value of the underlying asset and the exercise price. The net profit on the investment is the difference between the gross profit and the cost to have this right, i.e., that one initially paid for the call (namely *option premium* (Yeo and Qiu, 2003)). Therefore if K is the strike price and S_T is the asset value, the payoff of a call option is max(S_T - K; 0), while the payoff of a put option is max(K - S_T; 0). The Figure 1 illustrates a payoff diagram for call and put options.

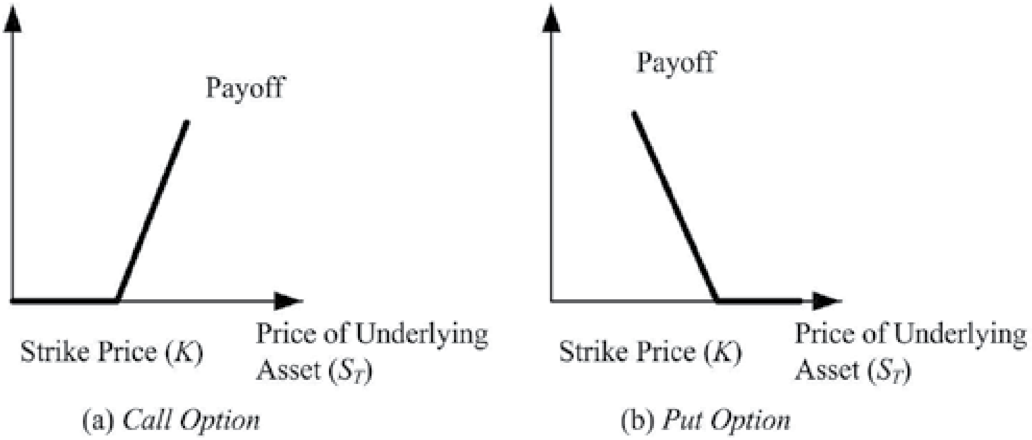


Fig. 1. Payoff of call and put options

2.2 Real options

In the seventies, some authors started to observe that the concept of option and, consequently, the methods for valuing financial options could be applied to evaluate flexibility associated with physical investments. The first who recognized the analogy between financial options and the flexibility of real world investments was Myers (1984). For this relationship, he coined the expression real option, that is, an option “written” on a real asset. He stated that real options are options on real assets while financial options are options on financial assets. Real options in option thinking are based on the same principles as financial options. Having a real option means to have “the possibility for a certain period to either choose for or against something, without binding oneself up-front” (Yeo and Qiu, 2003). However, the similarity between real and financial options does not mean that they are the same. The major difference is that real options are applicable to real assets (Kumar, 1996). A real asset is usually something tangible, such as a factory, machinery, engineering system, etc., while a financial asset typically consists of stocks, bonds, currency, etc. There are also other important differences that must be taken into consideration when the option pricing tools, built *ad hoc* for financial options, are used for evaluating real options:

1. Financial options have a shorter life (less than 1 year to expiration), while real options last more (even some options never expire).
2. Financial options are options on underlying assets that can be traded. Contrarily, real options often are on non-traded assets. Of course, the fact that financial options relate to traded assets simplifies the parameters' estimation. For instance, it is possible to calculate the variance of stock return rate from historical data or other options on the same asset.
3. Financial options are simple since they involve a simple option with only one constant exercise price. Contrarily, the exercise price of a real option can vary randomly. Furthermore, more than one option can be on the same underlying asset.
4. Both kinds of options assume that risk (i.e., the uncertainty of underlying asset) is exogenous. This is true in the financial world, since the uncertainty about the rate of return of a share cannot be influenced by who negotiate options on that asset. In the case of real options, indeed, this is not completely true. The behavior of a company, holder of a real option, can influence competitors and, then, the uncertainty of the specific project.
5. A financial option is a contract, that is, an agreement between two parties (respectively, holder and writer) to be executed later according to the ways set by it. In particular, the holder, paying today an amount of money called option premium, acquires the possibility to take some actions in the future (e.g., the right to buy (or sell) the asset on which the call (or put) option is written at a predefined price, called exercise price). The writer has the obligation to meet the holder's requirement. In the case of real options, the situation is deeply different. A real option, in contrast, is not a "contract" between two parties, rather than the possibility that management has to take some actions during project management (e.g., delay or abandon the project, etc.). In other words, it is the higher cost paid to increase the future flexibility of the investment project. Real options represent, in fact, the opportunity that uncertainty offers to the manager to "actively" manage the projects, reacting to changing competitive scenario. During project management, several situations that differ from those expected can arise, new information can be acquired, etc. The uncertainty resolution allows managers to adapt their choices to new circumstances. This is what makes decision-making process flexible and eliminates the irreversibility that characterizes traditional "static" now-or- never decision-making (e.g., if the company does not make the investment now, it will lose the opportunity forever). This flexibility of the decision-making process can be included into the evaluation of a project investment through real options analysis. Thus it is clear that, contrary to financial options, state that a real option is a contract "written" on an asset has a pure analogical meaning.

2.3 Evaluating real options: Empirical evidence

Hartmann and Hassan (2006), in a study on the application of real option analysis for pharmaceutical R&D project valuation, identify two different ways of using the Real Option Analysis (ROA). The first modus of usage can take place in a pure conceptual manner (Real Options Reasoning, ROR). In this case, emphasis is attributed to the innovative management philosophy rather than new calculation methods. "This 'application as a concept' aims to provide a more holistic analysis of the project features from an option's perspective" (Hartmann and Hassan, 2006). The second utilization is based on the first one. After identifying all relevant options in the project, the real option methodology is employed for concrete valuation procedure (ROP).

As for ROR, there are no problems since lots of studies clearly defined what a real option is and the associated managerial flexibility. As far as ROP is concerned, because of the importance of the theme and the need for including real options into the decision-making process, a huge study by researchers has been made during the years. It aims at identifying the best way of modeling and valuing the managerial flexibility (or real options) and, consequently, the right investment strategy. The Black-Scholes formula is one of the most known of these options valuation techniques. It seems very simple to use (it is just a formula application), even though the underlying assumptions are not understood or applicable to the real cases. The binomial tree proposed by Cox *et al.* (1979) is another financial approach largely used in the real option world. Maybe it is more emphasized than the Black-Scholes formula since it can be used to price American options. Another approach for assessing financial options is that one proposed by Longstaff and Shwartz (2001).

Unfortunately, however, the assumptions underlying most of these approaches and the appropriate conditions for their application to real cases are often not spelled out or differ widely from approach to approach and are even contradictory. Furthermore, the difficulties in implementing such approaches as well as their pros and cons are not explained.

Other approaches were developed to overcome the limits of the assumption underlying traditional financial approaches. The so-called "revised classic approach" (Borison, 2003) states explicitly that the assumptions underlying real options are restrictive. They suggest, therefore, that "the classic finance-based real options approach can be applied where these assumptions apply, while management science-based approaches, such as dynamic programming and decision analysis, be applied where they do not" (Borison, 2003). In particular, real options should be used where investments present only public risks, while dynamic programming or decision analysis where investments are dominated by private risks (Dixit and Pindyck, 1994; Amram and Kulatilaka, 2000). It is clear that "the primary, conceptual difficulty with the revised classical approach is separating all investments, in various shades of grey, into black and white, namely 'all market risk' and 'all private risk"' (Borison, 2003). Nau and McCardle (1991) and Smith and Nau (1995) study the relationship between option pricing theory and decision analysis and demonstrate that the two approaches yield the same results when applied correctly. Smith and Nau (1995) and Smith and McCardle (1998, 1999) proposed a method that integrates the two approaches. Such method bases on the assumption that there are two types of risk associated with most real investments: public or market risks, which can be hedged by trading securities and valued using option pricing theory, and private risks, which are project-specific risks and can be valued using decision analysis techniques. However, while the McCardle-Nau-Smith approach has a natural appeal in contexts where the distinction between markets risks and project specific risks is very natural (e.g., oil and gas exploration projects), there are several situations or industries where the distinction between market risks and project-specific risks is not as sharp (Brandão *et al.,* 2005).

2.4 Challenges to the practical implementation of modeling and valuing real options

As previously discussed, during the last decades, a growing body of literature has been dedicated to real options because of the importance of this theme and the weakness of traditional approaches (such as DCF) in an uncertain environment. Despite the growing attention to real options in theory, there is a little application of real options in practice. Companies and managers do not or rarely use real options in managing their businesses (Hartmann and Hassan, 2006). In their 2001 book Real Options: A Practitioner's Guide, co-authors Copeland and Antikarov predicted that real options would supplant NPV in just 10 years. "That prospect seems even more unlikely today" (Teach, 2003). In fact, in 2000 Bain & Co. conducted a survey about the use of 25 management tools by senior executives across more than 30 industries. This survey reveals that only 9% out of 451 participants use ROA and 32% of them, who had used real-options until then, abandoned this tool in 2000 (Teach, 2003). Another survey conducted by Ryan and Ryan (2002) found that in a sample of 205 Fortune 1000 firms only 11.4% use ROA as an auxiliary method compared with 85.1% for sensitivity analysis and 66.8% for scenario analysis. As far as "basic" capital budgeting tools are concerned, NPV is on the top of the list with 96% (Teach, 2003). The reason for this small use of real options probably may be that real option approach does not describe how manager has to take a decision (Lander and Pinches, 1998). It is necessary to specify that any decision making tool cannot give the best solution because it cannot substitute for managerial experience, knowledge or critical reasoning.

Lander and Pinches (1998) state that there are three primary reasons explaining why current models to evaluate real options are not used in practice. First, the models for evaluating real options are arcane. They are not well known or understood by managers and practitioners. Managers are not conversant with the mathematics of these models. Second, the models currently used originate from the financial world. Consequently, many of the required assumptions are often violated in the real world. This limits the use of real option approach and the reliability of the results. For example, option pricing models - typical of the financial world - have been extended to real investments. Since it is difficult to predict the future value of underlying asset, they assume that this value follows a stochastic process. Binomial model assumes that it is a binomial multiplicative process, but it is no sure that this hypothesis is correct and realistic. Furthermore, if this hypothesis is assumed to be valid, it is necessary to estimate the (financial) parameters of the model which is a difficult task. Third, these models are very complex from a mathematical perspective. Such complexity increases when the number of options embedded in the project increases, and their reciprocal dependence too. Thus a series of simplifying assumptions is usually made. As a consequence, the use of these models and the reliability of the results are limited. For all these reasons it is probably difficult for the managers or practitioners to model real investment opportunities which present real options with these techniques. There should be more attention to the practical side of modeling decision making process so that even those who are not too skilled can use these methods (Lander and Pinches, 1998).

2.5 Development of new approaches

The limits of the traditional financial techniques for evaluating real options have encouraged lots of academics to develop and explore new techniques to be used in the real world applications and engineering systems.

One of these is an approach based on a combination of Monte Carlo analysis and genetic algorithm. It can be used to find solutions in cases with a very large number of possible futures and system designs (Zhang *et al.,* 2008). For example, Lazo *et al.* (2003) use this approach to finding an optimal decision rule for oil field development alternatives that may help decision-making with regard to developing a field immediately or waiting for better market conditions. This optimal decision rule is formed by three mutually exclusive alternatives which describe three exercise regions along time, up to the expiration of the concession of the field. This approach is well suited to the design problem of complex engineering system where there are a lot of design alternatives and the benefit and costs of each of them are known. Genetic algorithms, in fact, provide an effective means to solve very large, non-complex, path dependent problems (Hassan and de Neufville, 2006). However, one of the major limits of this approach is the difficulty that mangers can have in handling with tools like genetic algorithm. This maybe justifies the little use in practice of it.

Johnson *et al.* (2006) proposes system dynamics models for applying real options to practice in the oil and gas industry. They discuss why they can be considered useful for evaluating projects with real options. System dynamics is able to realistically model many systems that use real options while relaxing common assumptions of perfection used by traditional real option models. They state that "modeling realistic managerial behavior in development projects is an area where system dynamics is well suited and has seen extensive application. [. . .] This capability can be used to develop models of managerial behavior when evaluating the purchase or exercise of a real option. [. . .] In addition to the basic decision rule for exercising the option, system dynamics can model other influences on managerial exercise choices, such as incentives, delays and biases, and nonlinearities". However, while the authors believe that system dynamics can help real options transition from theory to practice, they think that this is not the single solution to the challenges facing real options use. While the use of system dynamics has been increasing over the past several years (especially in the academic world), its use has not become common place. This can be in large part due to the unfamiliarity of project managers with system dynamics.

de Neufville *et al.* (2006) present a spreadsheet approach to valuing "real options" in projects. The model avoids complex financial procedures and, therefore, should be readily accessible to practicing professionals responsible for engineering design and management. It involves three steps:

- Step 1 It consists in determining the most likely projections of future costs and revenues of the project, and calculating the economic value. The design that maximizes the NPV is the base case against which flexible solutions are compared.
- Step 2 It consists in simulating possible scenarios in order to explore the effect of uncertainty. The collection of NPV obtained provides the distribution of possible outcomes for a project as well as an expected NPV.
- Step 3 The effects of various ways to provide flexibility are analyzed by changing the costs and revenues to reflect the design alternatives. The difference between the resulting best expected NPV and that of the base case is the value of flexibility.

The authors demonstrate the ease use of the method through a practical application to the design of a parking garage. Designers can take advantages of possible growth by building expansion flexibility into design. For example, they can make the columns big enough to support additional levels, should demand justify expansion of the parking garage in later years. In this case, the decision to construct an extra floor was made if the capacity was less than the demand for two consecutive years. This approach is surely valuable due to its easiness of applications. At the same time, however, it does not include the essence of the real options that is the exercise of the option or flexibility whether it is convenient according to the benefits and costs generated by this decision. Furthermore, the decision-making could be more complex than just observing the level of demand, and it should also cope with several uncertainty sources and variables. The next section discusses how Monte Carlo methods can be applied to overcome the problems of the traditional financial option pricing models.

3. Monte Carlo simulation for real option pricing

This section proposes an approach based on the Monte Carlo simulation as a possible way of evaluating managerial flexibility while accomplishing this objective. In the following subsections, a conceptual framework to apply it to real world cases is presented, and the pros and cons of this approach are discussed.

3.1 Monte Carlo simulation

The idea of Monte Carlo calculation is a lot older than the [digital] computer (Newman and Barkema, 1999). Stanistaw Ulam, a Polish-born mathematician who worked for John von Neumann on the Manhattan Project in 1944 and Edward Teller on the hydrogen bomb in 1951, is credited with inventing the modern Monte Carlo technique in 1946 (Parr and Smith, 2005). The first paper published using the term “Monte Carlo” was written by Metropolis and Ulam, “The Monte Carlo Method”, Journal of the American Statistical Association, in 1949. It is considered a powerful and flexible tool, even though quite old; hence, it has been applied to a variety of fields.

The Monte Carlo method can solve a problem by simulating directly the physical process, and is not necessary to write down the equations that describe the behavior of the system (Figure 2). This technique involves the random sampling of each probability distribution within the model to produce hundreds or even thousands of scenarios (also called iterations or trials). Each probability distribution is sampled in a manner that reproduces the distribution’s shape.

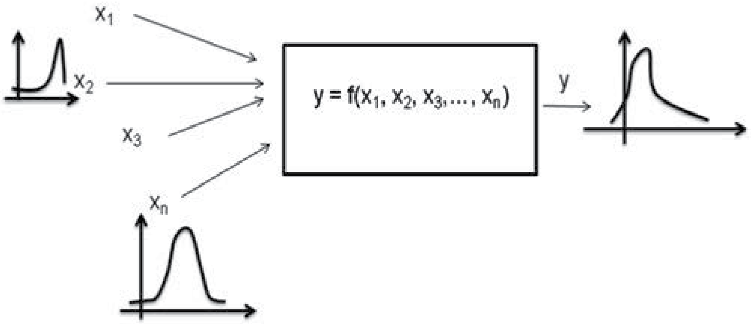


Fig. 2. Monte Carlo simulation

The distribution of the values calculated for the model outcome therefore reflects the probability of the values that could occur (Vose, 1996). The main elements of the techniques are briefly the following:

1. Input data: the inputs of the model that should be identified by the analyst. They can be known and fixed, namely deterministic, or uncertain, namely probabilistic.
2. Output variables: the results or objectives of the simulation. In other words, they represent what the analyst would like to estimate.
3. Model: the set of equations that describe the relationships between input and output. They translate the working of the system into "mathematical terms". Figure 3 illustrates key elements of the simulation.

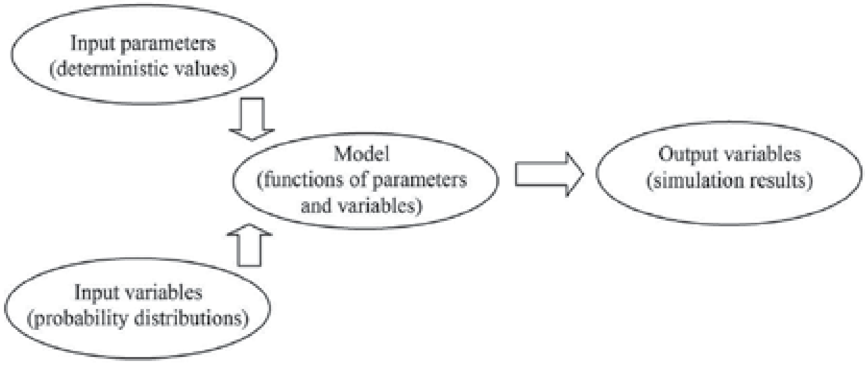


Fig. 3. Elements of the Monte Carlo simulation

The Monte Carlo method builds on the consideration that a direct analytical solution of the problem could be too burdensome or even impossible. The problem is then solved numerically: a "quite high" number *N* of possible combinations of values that input variables can assume is produced. Then the output is calculated by using model's equations. In order to have each of these combinations, a value for each input variable is generated randomly according to its probability distribution and the possible correlation among variables. Such procedure is repeated *N* times, with a quite high value of *N* to have sense from a statistical point of view. These values represent a sample of values for the output variables, and therefore can be used to build their probability distributions.

3.2 Real Option pricing with Monte Carlo simulation

Real options can be categorized as those that are either "on" or "in" projects (de Neufville, 2002). "Real options 'on' projects are financial options taken on technical things, treating technology itself as a 'black box'. Real options 'in' projects are options created by changing the actual design of the technical system" (Wang and de Neufville, 2005). Generally speaking, in financial problems that involve options, it is sufficient to correctly model the (stochastic) evolution of the underlying activity over time (e.g., the share price) and decide whether to exercise the option just looking at its value. The determination of the stopping rule is relatively simple. Such assumption could make sense for real options "on" projects (Wang and de Neufville, 2005), even though real options cases present always more sources of uncertainty than financial options.

Contrarily, real options "in" projects are much more complex and difficult to define in real projects or strategies. In this case, it is not sufficient to model the underlying asset as a whole or a black box. Instead, each source of uncertainty and the relationships among them should be described, as well as the decisional rules that lead to the decision of exercising or not the option whether some events occur, or some variables reach specific levels or values which make convenient the option exercise, and so on. To model simply the evolution of the whole underlying activity over time as a black box has no more sense. It is necessary, instead, to model the system with its real features, and include the decision maker's behavior and reaction to changing situations. To understand the nature of real options "in" project deeply, a simple example from Wang and de Neufville (2005) can be considered. A spare tire on a car "gives the driver the 'right, but not the obligation' to change a tire at any time, but this right will only rationally be used when the car has a flat". It is evident the difference between this kind of option and, for example, the option to abandon or delay a project. In the first case, we should model the system and decision maker's behavior that will decide to change the tire whether there is really a flat, the change in that place and moment is economically advantageous, and so on. In the second case, it would make sense, in order to use a financial approach, modeling the (stochastic) evolution of the project value (as a black box) and deciding to exercise the abandonment option, for example, when the salvage value is higher than the project value. Sometimes, however, this could be a too strong assumption also for real options "on" projects.

Real options in general, and particularly those "in" projects, require, therefore, a deep understanding of the project with its own characteristics. Such knowledge is not readily available among options analysts. This has determined few analyses of real options "in" projects, despite the managerial flexibility is very often more similar to a real option "in" rather than "on" projects. Moreover, since the data available for real options "in" projects are of much poorer quality than that of financial options, real options "in" projects are different and need an appropriate analysis framework. Furthermore, there are typically several uncertainties in a real option evaluation. They are of different kind, such as technical, economic and not only financial. They are often correlated and not independent.

For all these reasons, this research investigates and proposes an original approach based on the Monte Carlo simulation (MCS) for real option pricing. It starts from the traditional concepts of option pricing, but, by using simulation, it takes into consideration multifarious uncertainty sources (as they are in the real world).

As discussed, most of the traditional methods are one/two factors models; they assume that only one/two (financial) inputs are uncertain. Given the characteristics of financial investment decisions, these methods generally are limited to evaluating price uncertainty only. Differently, the MCS is a powerful tool that considers the flexibility in the number and specification of the uncertainties in the decision problem (Triatis and Borison, 2001). MCS is able to include different (theoretically "endless") sources of uncertainty in the evaluation of investment projects. A probability distribution, resulting from historical data (if available) or managers' subjective experience, is assigned to each uncertainty or variable model input. Instead of a single "deterministic" value, the MCS gives a "more realistic" probabilistic representation of the model outputs that managers can use with other considerations to support their decision-making process, and, of course, justify the greater initial investment required by this kind of "flexible" projects. The MCS-based approach can model the decision maker's behavior and system through simple equations and decisional rules.

The Monte Carlo method, in fact, is a numerical approximation technique that can be used to solve any kind of problem regardless of the type of uncertainty or description. As results of simulation, it produces a data sample which can be analyzed using standard statistical tools facilitated by the software performing the MCS. Moreover, compared to other techniques, such as what-if analysis, scenario planning, etc., the MCS handles correlations among variables correctly and easily.

The potential of this method is its flexibility for the user modeling, reducing the problem of "curse of modeling", permitting a large popularization of real options modeling for people without a comprehensive finance knowledge. This method, in addition, can give (probabilistic) results, even when the system complexity does not have closed-form solutions. Table 1 reports a comparison between the Monte Carlo simulation-based approach and traditional financial techniques for option pricing.

Monte Carlo simulation-based approach	**Financial models for real options**
Several (theoretically "endless") sources of uncertainty	One / two factors models: only one/two (financial) inputs are uncertain
Simple modeling of the decision making behavior and decisional rules for exercising options according to each source of uncertainty	Exercise options according to the value of the whole underlying activity modeled as a 'black box'
Probabilistic output: range of values corresponding to several possible scenarios weighted for their probability	Deterministic output: one single value corresponding to the single scenario analyzed
'Transparency' and flexibility of modeling (through a simple spreadsheet)	Black box: too many underlying mathematical assumptions
. . .	. . .

Table 1. MCS-based approach versus traditional financial models for real option pricing

Even if MCS is a well known approach, its use in assessing the value of flexibility given by real options embedded in some projects or engineering systems is relatively new and considered only in few fields. The MCS flexibility has fostered some authors to enhance its potentiality in real options evaluation (Wang and de Neufville, 2005; Triantis and Borison, 2001; Chiara *et al.*, 2007), even if it is rarely applied or, in any case, applied in a way that is different from the approach developed by this research. Triantis and Borison (2001) state that the MCS is able to generate a large number of possible scenarios for the underlying project cash flows or value, based on assumed probability distributions for each uncertainty. The real option value can be then calculated for each of these scenarios, and the average of these values is discounted back to the present. They emphasize the MCS potential, but do not explain how managers should apply this technique. In other cases, such as in Copeland and Antikarov (2001), the MCS is used to generate a distribution for the value of an underlying "developed" project using distributions of variables such as price, market share, and market size that determine the project's cash flows and value. This volatility is then used to generate a binomial tree for the evolution of the project's value over time, and the value of an option on the project is calculated by using the binomial option pricing model. Chiara *et.al.* (2007) apply least-squares Monte Carlo method by Longstaff and Shwartz

(2001) to quantify the value of a revenue guarantee in BOT (Build-Operate-Transfer) projects. Only recently, de Neufville *et al.* (2006) have proposed an approach to valuing real options that is no more based on the traditional financial techniques, but on the modeling of the system through the use of a spreadsheet and simulation. It has been developed for including flexibility in the design of engineering systems. For example, the case example used in their paper is the design of a parking garage. In particular, they analyze whether to design footings and columns of the original building of a parking garage so that additional levels of parking can be added. For this case, the decision to construct an extra floor was made if the capacity was less than the demand for two consecutive years. Of course, real options, as way of thinking, are not only concerned with the design of engineering systems, but with any managerial flexibility that could be created by uncertainty. Therefore, this chapter presents an engineering approach to valuating real option "in" projects based on the modeling of the system. At the same time, however, this approach needs to be generalized in order to be applied to all or the most part of real cases with embedded real options.

3.3 MCS-based approach: Methodological framework

This section presents the methodological framework of the developed approach. It consists in three basic modules, as shown in Figure 4. The first module (Module 1) refers to the modeling of the considered projects in terms of real options. In other words, the specific project should be analyzed and possible real options (managerial flexibility) identified. This stage is equivalent to the so-called Real Option Reasoning (ROR). This "application as a concept", or as "a way of thinking", aims to provide a more holistic analysis of the project or strategy features from an option's perspective.

Module 1 is basically structured in two elements: (i) the identification of managerial flexibility (real option "on" and "in" projects) and (ii) the definition of the decisional rules for the option exercise (i.e., the conditions that lead the decision maker to consider whether to exercise the option) or, in other words, the construction of the conceptual framework of the option exercise.

The second module (Module 2) is related to the parameters calculation and uncertainty modeling. In particular, it consists in (i) identifying the inputs of the problem, and (ii) categorizing them into deterministic/static (fixed and known with certainty) parameters and uncertain variables whose uncertainty can be modeled by a probability distribution. As for uncertain variables, the probability distribution can be determined from the historical data when they are available (frequency interpretation of probability), otherwise from expert's opinion (subjective probability or degrees of beliefs (possibilistic reasoning)). Moreover, (iii) the possible correlation among variables should be captured in order to obtain a realistic output. The next section discusses the issue of input modeling and the related techniques, and presents the method developed by this research, namely the fuzzy Delphi method[1]. It aims at eliciting information from more experts, while coping with "vagueness" of human language.

[1] Note that fuzzy logic is not compatible with probabilistic logic, which is why it is important to realize that MSC solves the problem numerically and thereby makes these fine distinctions irrelevant as long as the uncertainty can be modeled in one way or the other.

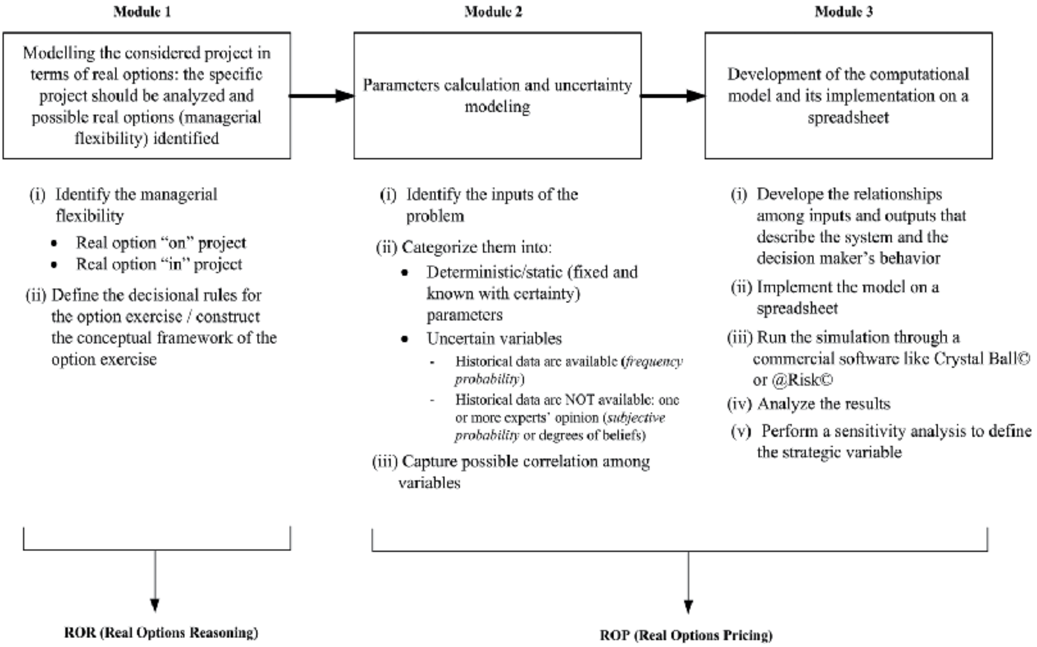

Module 1
Modelling the considered project in terms of real options: the specific project should be analyzed and possible real options (managerial flexibility) identified
Module 2
Parameters calculation and uncertainty modeling
Module 3
Development of the computational model and its implementation on a spreadsheet
(i) Identify the managerial flexibility
• Real option "on" project
• Real option "in" project
(ii) Define the decisional rules for the option exercise / construct the conceptual framework of the option exercise
(i) Identify the inputs of the problem
(ii) Categorize them into:
• Deterministic/static (fixed and known with certainty) parameters
• Uncertain variables
- Historical data are available (frequency probability)
- Historical data are NOT available: one or more experts' opinion (subjective probability or degrees of beliefs)
(iii) Capture possible correlation among variables
(i) Develope the relationships among inputs and outputs that describe the system and the decision maker's behavior
(ii) Implement the model on a spreadsheet
(iii) Run the simulation through a commercial software like Crystal Ball© or @Risk©
(iv) Analyze the results
(v) Perform a sensitivity analysis to define the strategic variable
ROR (Real Options Reasoning)
ROP (Real Options Pricing)

Fig. 4. Methodological framework of modeling

The third module (Module 3) is on the development of the computational model and its implementation on a spreadsheet (or any other environment able to run Monte Carlo simulations). It consists in five basic activities: (i) developing the relationships among inputs and outputs that describe the system and the decision maker's behavior, (ii) implementing the model on a spreadsheet, (iii) running the simulation through a commercial software like Crystal Ball® or @Risk®, (iv) analyzing the results and (v) performing a sensitivity analysis to define the strategic variable.

Module 2 and 3 refers to the Real Option Pricing (ROP). In these last stages, real options have a role of "an analytical tool" whose focus is represented by the concrete evaluation procedures.

3.4 MCS-based approach for American and European options: Limits and opportunities

Monte Carlo simulation does not have as many assumptions as the traditional financial real options model. It is possible to specify the probability distributions of the uncertain input data, and describe the function between the input uncertain variables and the output payoff. Then, simulator can do the "brute force" work (Wang and de Neufville, 2005). However, there are some "real or outward" issues about the use of MCS that need to be understood and discussed.

First, if there are multiple sources of uncertainty, it could be prohibitive to calculate the option value at required accuracy. This problem is known as "the curse of dimensionality". It refers to the number of samples per variable which increases exponentially with the number of variables to maintain a given level of accuracy. In reality, this problem has been overcome by the high capacity of the modern simulators and the sampling procedures used. Contrarily, this is an "open problem" for the financial real options models. For example, for more than three or four state variables, both lattice and finite-difference methods face several difficulties and are not practical or even not applicable.

Second, one of the main drawbacks of MCS is its "forward" approach rather than "backward". This aspect makes the MCS successful only for European options, while for American option an optimization technique like that developed by Longstaff and Shwartz (2001) should be used. In reality, as this research shows, this limit of the MCS is only outward, and this approach can be used for both European and American real options. European options are past dependent in the sense that cash flows depend only on past information. In this case, MCS values a function by unfolding uncertainty as it evolves from the past (forward induction) (Cortazar, 2001). In the case of American options, the situation is different. American options may be exercised at any of several dates. Thus, cash flows on a given date depend not only on past information, but also on expectations of future events. The value of a security is traditionally obtained by some kind of backward induction. Examples of backward induction procedures have been proposed for valuing assets, from dynamic programming, to binomial and multinomial trees, to finite difference procedures for solving partial differential equations. All these procedures start from some boundary conditions and solve simultaneously for asset value and the optimal exercise policy, determining the shape of the cash flow function in such a way as to maximize asset value. Thus, several studies recognized that the MCS fails in addressing this issue. This research demonstrates that this is not true. In fact, the first aspect to be considered is that in real investment projects, options are often American, in the sense that they can be exercised in any date. Most probably, their exercise is conditioned to the occurrence of some particular events (see the example of the flag in the tire). In other words, they will be exercised whether some events occur, and for some aspects this makes these options equivalent to European options. Moreover, in most real cases American real options can be exercised at any date, but only once during the lifetime of the options. These and other cases can be simply modeled and estimated by MCS. At each trial of the simulation, the simulator chooses randomly input values from their probability distributions. Thus, each trial is a possible scenario that can happen. As really happens, since all input parameters of the problem are known for the scenario, the decision maker will decide to exercise the option if, according to the data, it is allowed and advantageous.

The decisional rule and exercise policy is applied to each scenario (as if it was "deterministic"). Since it is a probabilistic problem, each trial of the simulation will repeat this procedure with the same decisional rule (for the option exercise). The probability distribution of the output payoff will be defined by all outputs of the various scenarios. Thus, for example, one trial will give that the option exercise is, say, at year 3 with its related payoff; another trial will return that the option exercise is at year 5 with its payoff; another will return that the option exercise is not advantageous; and so on. Therefore, it is clear that the simulation output (typically, a probability distribution of the option payoff) is a realistic estimation of the possible option values, and takes into account the "American characteristic" of the option.

Third, last but not least "concern" about simulation is related to the input parameters of the models. MCS needs to have sound probability distributions and stochastic models for the underlying uncertain variables. If the model inputs, or the model itself, are wrong, the simulation can only serve the role of "garbage in and garbage out" (Wang and de Neufville, 2005). Thus, a correct and reasonable procedure to determine the input data from historical data (if available) or experts' option should be designed.

4. Uncertainty modeling and input calculation

The previous section concluded with the discussion about the limits and opportunities of the Monte Carlo simulation-based approach to valuing real options. One of the most critical aspects of the Monte Carlo technique is the determination of the input variables and of their probability distributions. If the distributions of model inputs are wrong, the simulation can only serve the role of "garbage in and garbage out", since the results can be unreliable. Using incorrect or wrong input nullifies, *de facto*, the advantages of adopting the MCS. This issue sketches the need for developing a method that can be used along with the MCS approach for the calculation and modeling of uncertain input variables. This section accomplishes this objective.

The discussion of uncertainty is closely related to probability, and different conceptions of probability underpin the different ways in which uncertainty has been expressed (Dequech, 2004; Lawson, 1988). One important distinction is that between the theories in which probability is a property of the way one thinks about the world, a degree of belief, and those theories where probability is a property of the real world. Perlman and McCann (1996) use the terms "epistemic" to denote the first type of probability (a degree of belief) and "aleatory" to denote the second type of probability (a feature of reality). The subjective probability theory of de Finetti is example of the former, while frequency theory belongs to the second category (Dequech, 2004).

There are essentially two sources of information used to quantify the uncertainty of the variables within the problem under consideration. The first is available data and the second is experts' opinion. Figure 5 sketches the ways of defining probability distributions of uncertain variables.

The next section presents how to derive distribution probability from data as well as from experts. Then fuzzy logic is introduced the approach used to cope with the vagueness of humane language, even in the presence of more than one experts.

4.1 Deriving distributions from data

When observed data for an uncertain variable is available, a distribution that realistically models the true uncertainty can be derived from it. The key assumption is that the observed data can be thought of as a random sample from a probability distribution that should be identified in order to be reproduced in the simulation. In other words, this method is based on the definition of probability based on frequency (frequency probability). The data may come from several sources, such as historical data (time series), surveys, scientific experiments, and so on. The basic steps which should be followed in order to derive a probability distribution are the following:

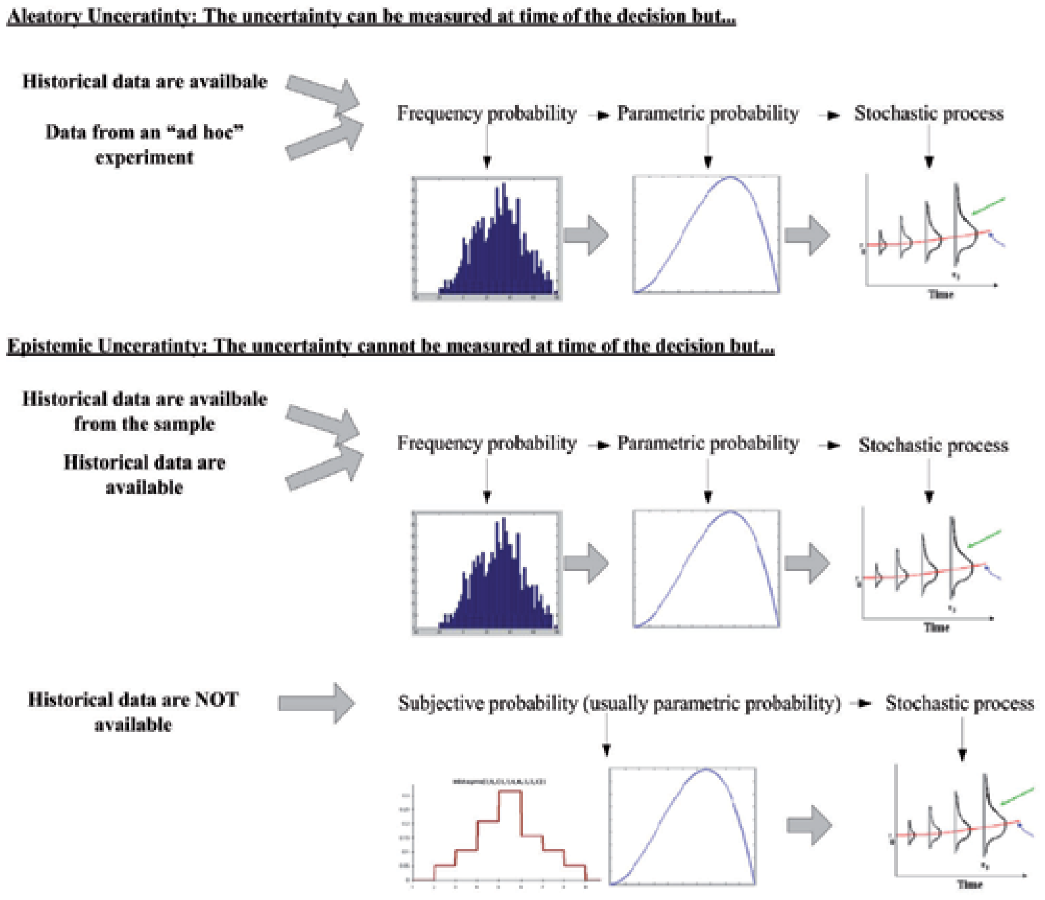

Aleatory Unceratinty: The uncertainty can be measured at time of the decision but...
Historical data are availbale
Data from an "ad hoc" experiment
Frequency probability
Parametric probability
Stochastic process
Time
Epistemic Unceratinty: The uncertainty cannot be measured at time of the decision but...
Historical data are availbale from the sample
Historical data are available
Frequency probability
Parametric probability
Stochastic process
Time
Historical data are NOT available
Subjective probability (usually parametric probability)
Stochastic process
Time

Fig. 5. Project uncertainty and probability distributions (Source: Chiara (2008))

a. Analyzing the available data. The first step should be the analysis of the available data. In particular, (1) it is necessary to understand the nature of the variable or, in other words, if it is discrete or continuous. A variable that is discrete in nature is usually, but not always, best fitted to a discrete distribution. In reality, in certain circumstances, a continuous distribution can describe better the variable. In this case, of course, the value sampled from the distribution should be rounded off in order to be acceptable. It is important to clarify that the reverse never occurs. A continuous variable is always described by a continuous distribution and never by a discrete distribution (Vose, 1996). Then, (2) the range of the variable should be defined. (3) The possible correlations of the variable with another variable within or outside the model should be studied.
b. Fitting a distribution to the available data. The available data can be fitted to an empirical or theoretical distribution. In the first case, starting from the available data the distribution is defined empirically. Usually, a cumulative frequency derived from data is sufficient to define a probability distribution. If the number of data is very large, it is usually easier to arrange the data into histogram form and then define a cumulative frequency. In reality, the availability of software able to automatically attempt fits to several distribution types makes this activity very ease. However, it is also possible to determine the theoretical

distribution also without this software. The distribution parameters that make a distribution type best fit the available data can be determined in several ways.

c. Goodness-of-fit statistics. The goodness-of-fit-statistics most commonly used are the Chi squared X^2 and Kolmogorov-Smirnoff (*K-S*) statistics. They are generally used for discrete and continuous distributions respectively. The lower the value of these statistics, the better the theoretical distributions fit the data. The goodness-of-fit-statistics do not represent the true probability measure that data actually come from the fitted distribution. Instead, they represent the probability that random data generated from the fitted distribution would have produced a goodness-of-fit statistic value as low as that calculated for the observed data. In particular, the Chi squared X^2 statistic measures the degree of comparability between the expected frequency of the fitted distribution with the frequency of a histogram of the observed data. The *K-S* statistic measures the maximum vertical distance between the cumulative distribution function of the fitted distribution and the cumulative distribution of the data.

4.2 Defining distributions from expert opinion

In most cases, insufficient data is available to completely specify the uncertainty of a variable; hence, one or more experts will usually be consulted to provide their opinion of the variable's uncertainty. This approach is clearly based on the subjective probability definition. Probability distribution functions are typically divided into two categories: nonparametric and parametric distributions. A parametric distribution is based on a mathematical function whose shape and range is determined by one or more distribution parameters. Sometimes, these parameters do not have a direct relationship to the distribution shapes. Examples are Normal, Beta, Lognormal, and so on. Nonparametric distribution, instead, have a direct and intuitive relationship between their parameters and their shape or range. Examples are Uniform, Triangle, Discrete, and so on. Therefore, "as a rule, non-parametric distributions are far more reliable and flexible for modeling expert opinion" (Vose, 1996). However, when the expert is very familiar with using the parameters that define the distribution or the parameters of a parametric distribution are intuitive (such as for Binomial distribution), the use of parametric distributions is preferred to the non-parametric ones in modeling expert opinion.

One of the distributions mostly used to model expert opinion is the BetaPERT distribution. It is useful to point out this distribution since it is relatively new and especially used in risk analysis. Strictly speaking, it is a parametric distribution, but it has been adapted so that the expert should only provide estimates of the minimum, most likely and maximum value for the variable. Examples of BetaPERT are shown in Figure 6.

The equation of the BetaPERT distribution is related to the Beta distribution, as in (1).

$$BetaPERT(a,b,c) = Beta(a_1, a_1)(c-a) + a \tag{1}$$

where:

$$a_1 = [(\mu-a)(2b-a-c)]/[(b-\mu)(c-a)] \tag{2}$$

$$a_2 = a_1\ (c-\mu)/(\mu-a) \tag{3}$$

$$The\ mean\ \mu = (a+4b+c)/6 \tag{4}$$

As shown in (4), the mean for the BetaPERT distribution is four times more sensitive to the most likely value than to the minimum and maximum values. Consequently, the expected value is less sensitive to minimum and maximum values than triangular's expectation.

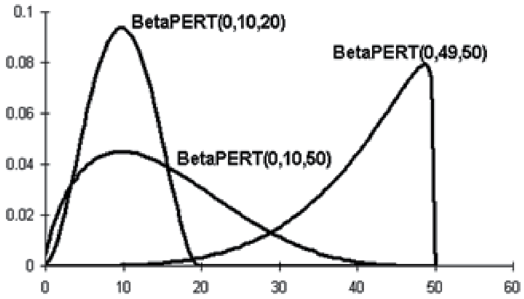


Fig. 6. Examples of BetaPERT distributions

4.3 Fuzzy set theory and fuzzy logic

When the value of a variable has to be determined by experts' opinions, it is necessary to consider that human propositions and reasoning can be affected by "vagueness and nonspecificity" (Klir, 1995). In fact, "an economic agent does not generally manifest a perfect aptitude to clearly discriminate between the alternatives he prefers and those he does not prefer. Even if he is well informed, it does not follow that his behavior obeys a binary logic of the type: preference-non preference, as in the classical theory" (Ponsard, 1988). Human reasoning and decision making is very often based on genuine uncertainty embedded in natural language. Economic agent, in fact, expresses his experience or predictions in "imprecise" linguistic terms, such as: "The price of oil is *not likely* to *increase substantially* in the *near future*" (Klir, 1995). Thus classical two-valued logic often is not able to model economic reality and human reasoning. In this sense, fuzzy logic provides a natural framework to express and deal with this kind of vagueness.

Fuzzy set theory and fuzzy logic, therefore, address a fundamentally different class of problems from that addressed by probability theory. Probability theory deals with propositions that are mutually exclusive. There is simply uncertainty as to which is true. The real number attached to the proposition is a measure of how much we believe a certain proposition to be true. Conversely, fuzzy set theory deals with proposition that have "vague meaning". In fuzzy logic, a proposition may be true or false up to a certain degree: the interval between 0 (false) and 1 (true) describes human reasoning. For example, if we consider the previous statement "the price of oil is *not likely* to *increase substantially* in the *near future*" (Klir, 1995), the expression "not likely" is a statement of uncertainty and can be represented by a probability. However, although the terms "increase substantially" or "near future" are not precise, they are not an expression of uncertainty. We assume that we can measure the increase and know how large it is. The question is how much the increase's particular size fits the economic agent's vague description of "increase substantially". Zadeh (Yager *et al.*, 1987) addresses such problems with *fuzzy set theory*. Fuzzy set theory associates a real number between 0 and 1 with the membership of a particular element in a set. Let U be a collection of objects denoted generically by u, which could be discrete or continuous. U is called the universe of discourse and u represents the generic element of U.

A *fuzzy set F* in a universe of discourse *U* is characterized by a membership function μ_F which takes values in the interval [0, 1], namely μ_F: $U \rightarrow [0, 1]$. A fuzzy set may be viewed as a generalization of the concept of an ordinary set whose membership function only takes two values 0, 1. Thus a fuzzy set *F* in *U* may be represented as a set of ordered pairs of a generic element *u* and its grade of membership function: F = (u, μ_F (u))|$u \in U$. In this sense, "the use of fuzzy sets provides a basis for a systematic way for the manipulation of vague and imprecise concepts" (Lee, 1990).

Fuzzy sets can be employed to represent linguistic variables. A linguistic variable can be regarded either as a variable whose value is a fuzzy number or as a variable whose values are defined in linguistic terms. A linguistic variable is characterized by a quintuple (x, $T(x)$,U, G,M) where x is the name of the variable; $T(x)$ is the term set of x, i.e., the set of names of linguistic values of x with each value being a fuzzy number defined in U; G is a syntactic rule for generating the names of values of x; and M is a semantic rule for associating with each value its meaning. For example, if speed is interpreted as a linguistic variable, then its term set *T*(speed) could be *T*(speed) = slow, moderate, fast, very slow, more or less fast, ..., where each term in *T*(speed) is characterized by a fuzzy set in a universe of discourse U = [0, 100]. The term "slow" can be interpreted as "a speed below about 40 mph", "moderate" as "a speed close to 55 mph", "fast" as "a speed above about 70 mph" (Lee, 1990). These terms can be characterized as fuzzy sets whose membership functions are shown in Figure 7.

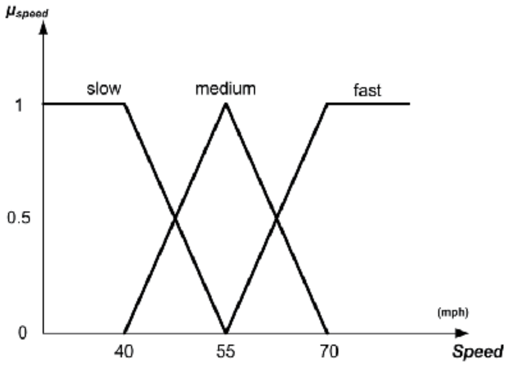


Fig. 7. Diagrammatic representation of fuzzy speeds. "Speed" is linguistic variable with three terms: "slow", "medium" and "fast" (Source: Lee (1990))

4.4 From one to more experts: The Delphi method

The Delphi technique is a method of eliciting and refining group judgments. It was developed in the 1950s by the Rand Corporation at Santa Monica, CA. It is a widely used and accepted method for achieving convergence of opinions concerning real-world knowledge by using a series of questionnaires to collect data from a panel of selected subjects (Hsu and Sandford, 2007). The rationale for the procedure is that "two heads are better than one" when the exact knowledge is not available (Dalkey, 1969). This procedure has three key features. First, the response is anonymous, because the opinions of members of the group are obtained by a formal questionnaire. Second, it employs multiple iterations that refer to the feedback process. This process consists of a series of rounds. In each round every participant works through a questionnaire which is sent to the coordinator. Then a summation of collected opinions is returned to each participant. Therefore based on the position of the whole group they can modify their initial judgments about the information provided in previous iterations. Third, there is a statistical group response defined by aggregating the individual opinions in the final round. All these features are designed to minimize the biasing effects of dominant individuals, irrelevant communications (noise), and group pressure toward conformity (Dalkey, 1969).

4.5 The fuzzy Delphi method

As discussed in previous sections, when insufficient data is available to completely specify the uncertainty of a variable, experts will usually be consulted to provide their opinions of the variable's uncertainty (subjective probability or degree of believes). This implies to cope with the vagueness of experts' judgments and the subjectivity of each opinion. In order to determine the uncertain parameters required by the MCS-based real options model, an approach using both the Delphi technique and the fuzzy logic was developed (Costantino *et al.*, 2009; Costantino and Pellegrino, 2010b; Pellegrino and Costantino, 2011). It is able to take into consideration the vagueness of opinions and, at the same time, to reach a certain degree of "objectivity" by interviewing more experts. The proposed approach modifies that introduced in 1988 by Kaufmann and Gupta and applied by Cheng and Lin (2002) to the decision making. The procedure adopted by the fuzzy Delphi approach identifies two kinds of information that should be elicited from experts:

i. Estimates of uncertain parameters of the model (quantitative variables). As discussed, to model the expert opinion it is preferable to use non-parametric distributions rather than parametric distributions (whose shape and range are determined by one or more distribution parameters that often do not have a direct relationship with distribution's shape). Moreover, one of the probability distributions most commonly used is the BetaPERT distribution. Three-point estimates, therefore, are usually quite sufficient to model an expert's opinion about some uncertain variables (Vose, 1996). In this fuzzy Delphi approach the uncertain quantitative variables were mostly modeled by BetaPERT distribution. Therefore, experts are asked to express their opinion about the three input parameters of this distribution. The procedure can be, of course, easily generalized for all kinds of distributions. Experts are asked to express each of these three values by a "trapezoidal" fuzzy number (one pessimistic, one optimistic and two values that define the "most plausible" interval of values, as illustrated in Figure 8) instead of a crisp number. A trapezoidal fuzzy number was chosen since its membership function is intuitive and easy to understand. However, the membership function can be easily generalized. Therefore, we will have one fuzzy number which describes the minimum, one for the maximum and another one for the most likely value. Each fuzzy number represents a "possibility distribution" (Zadeh, 1978) of the value under consideration (i.e., min, max and most likely).

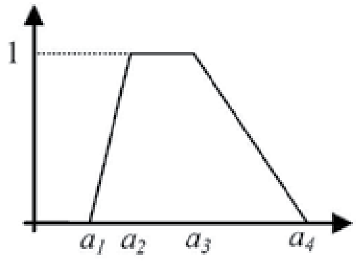


Fig. 8. Example of a trapezoidal fuzzy number $A = (a_1, a_2, a_3, a_4)$

ii. Subjective estimation of discrete probabilities (linguistic variables). An expert will sometimes be asked to provide an estimate of the probability of occurrence of a discrete event. This is a difficult task for the expert. He will have to rely on his feel based on his understanding and perception of this event. This task is often complicated by the difficulty in having to visualize the difference between 30% and 40%. A way to avoid this problem is to offer the expert a list of probability phrases (Vose, 1996). In the proposed fuzzy Delphi approach each expert is asked to select a phrase that best fits his understanding of the probability of each event that has to be considered, as in Table 2. Using the language of fuzzy number, these probability phrases represent the term set of the linguistic variable (probability of a discrete event). Each value of the term set, i.e., each probability phrase, is a fuzzy number whose membership function is again a trapezoidal one for its simplicity.

Probability phrase	**Linguistic Variable**	**Fuzzy number**
Almost certain	Very high (VH)	(0.8, 0.9, 1, 1)
Highly likely	High (H)	(0.7, 0.8, 0.8, 0.9)
Fairly likely	Medium high (MH)	(0.5, 0.6, 0.7, 0.8)
Even chance	Medium (M)	(0.4, 0.5, 0.5, 0.6)
Fairly unlikely	Medium low (ML)	(0.2, 0.3, 0.4, 0.5)
Highly unlikely	Low (L)	(0.1, 0.2, 0.2, 0.3)
Almost impossible	Very low (VL)	(0, 0, 0.1, 0.2)

Table 2. Probability phrases and fuzzy linguistic variables

The fuzzy Delphi method proposed to determine the final values of the input variables in the model consists of the following steps.

Step 1. Experts E_i, $i = 1, ..., n$ are asked to provide the possible values of the three-point estimates of the considered quantitative variables (i). Each estimate is expressed by a trapezoidal fuzzy number (5): the pessimistic value $a_1^{(i)}$, the most plausible interval of values ($a_2^{(i)}$, $a_3^{(i)}$), and the optimistic value $a_4^{(i)}$.

$$A^{(i)} = (a_1^{(i)},\ a_2^{(i)},\ a_3^{(i)},\ a_4^{(i)}) \qquad i = 1, ..., n \tag{5}$$

The expert is also asked to select a probability phrase for each discrete event he wants to estimate (ii). Each probability phrase is associated with a linguistic variable that corresponds to a predefined trapezoidal fuzzy number, as shown in Table 2. For each question the expert should also give an explanation of his answer, in order to let the coordinator know the main reasons of the choice.

Step 2. The average value A_m of all $A^{(i)}$, $i = 1, ..., n$, is computed as calculated by (6).

$$A_m = [(a_{1m}^{min},\ a_{1m}^{mode},\ a_{1m}^{max}),\ (a_{2m}^{min},\ a_{2m}^{mode},\ a_{2m}^{max}),\ (a_{3m}^{min},\ a_{3m}^{mode},\ a_{3m}^{max}),\ (a_{4m}^{min},\ a_{4m}^{mode},\ a_{4m}^{max}) \tag{6}$$

where:

$$a_{jm}^{min} = min_{\ i=1,\ldots,n}\ a^{(i)}{}_j,\ a_{jm}^{mode} = mode_{\ i=1,\ldots,n}\ a^{(i)}{}_j,\ a_{jm}^{max} = max_{\ i=1,\ldots,n}\ a^{(i)}{}_j \qquad j = 1, 2, 3, 4 \tag{7}$$

Then the average position of the group on the quantitative variable (expressed as "according to the other interviewed experts, the variable x varies in a range between x_{min} and x_{max}, with a peak in x_{mode}") along with the opinions of other experts on the probability phrases and the main explanations of their answers are sent back to the expert E_i for reexamination.

Step 3. In the third round, each expert E_i can revise his judgments based on the information discussed at the end of step 2, i.e. he can present a revised trapezoidal fuzzy number for the quantitative variables (8) and modify the probability phrase (linguistic variable) according to the position of the whole group.

$$B^{(i)} = (b^{(i)}{}_1,\ b^{(i)}{}_2,\ b^{(i)}{}_3,\ b^{(i)}{}_4) \qquad i = 1, ..., n \tag{8}$$

This process starting with Step 2 is repeated. The new average position of the group B_m is calculated according to equation (6), where $a_1^{(i)}$, $a_2^{(i)}$, $a_3^{(i)}$, $a_4^{(i)}$ are substituted by $b_1^{(i)}$, $b_2^{(i)}$, $b_3^{(i)}$, $b_4^{(i)}$. The process could be repeated again and again until to successive means A_m, B_m . . . become reasonably close and the group almost agrees on the value of the linguistic variables (or, in the worst cases, there are no more changes in the experts' evaluations).

Step 4. If the final opinions are not coincident, the mean can be considered for both the quantitative variables and linguistic ones expressed in terms of fuzzy numbers (9).

$$\mathrm{A_m} = \left(m_{1m}, m_{2m}, m_{3m}, m_{4m}\right) = \left(\frac{1}{n}\sum_{i=1}^{n} a_1^{(i)}, \frac{1}{n}\sum_{i=1}^{n} a_2^{(i)}, \frac{1}{n}\sum_{i=1}^{n} a_3^{(i)}, \frac{1}{n}\sum_{i=1}^{n} a_4^{(i)}\right) \tag{9}$$

The output of the fuzzy Delphi process so far is a fuzzy set, specifying a possibility distribution of the uncertain variable parameters. In defining the probability distributions in MCS model, a nonfuzzy (crisp) value is required. Consequently, one must defuzzify the fuzzy output obtained from the fuzzy Delphi process, namely:

$$z_0 = defuzzifier(z) \tag{10}$$

where z_0 is the nonfuzzy output and defuzzifier is the defuzzification operator. The defuzzification method used in this research is the centroid calculation, which returns as a crisp number the center of the area under the curve (fuzzy number). Therefore, the defuzzification value of the trapezoidal fuzzy number for each variable (quantitative or linguistic) is used as input in the model.

5. Case study applications

The described approach was applied to some real cases on different engineering fields (Costantino and Pellegrino, 2010b; 2011; Costantino *et al.*, 2009) in order to show and test its potential in evaluating investment projects that present managerial flexibilities.

One example is represented by the use of the MCS approach for estimating the risks in project finance transactions where contractual structures for allocating risks during the operating period are adopted. Examples are put-or-pay, take-or-pay or take-and-pay, and throughput contracts (Razavi, 1996).

Take-and-pay and take-or-pay are two different off-take agreements: the first requires the buyer to take and pay for the good or service only if delivered, whereas the second requires a payment unconditionally, even if no good or service is provided or producible by the seller. The commitment of all or an agreed percentage of the output capacity of a project through these contracts gives a purchaser supply certainty and, at the same time, provides sales predictability to the project company (Hoffman, 2007). Lenders can also rely upon off-take agreements for repayment of their loans. However, there are also some drawbacks, especially when buyers have highly variable and uncertain cash flows. Even though a purchaser may have obligated itself to purchase a certain capacity of the project, there may be circumstances under which it does not want to take the contracted product or service. According to these contracts, the purchaser should pay the standby charge in any case. In order to reduce this kind of "stiffness" and mitigate the buyer's risk, a source of flexibility is often added in these contracts, that is, the clause named "option capacity". Before an off-take obligor wishing be excused from its purchase obligation is required to pay a standby charge for not taking such product, the withdrawing purchaser's capacity is offered to other purchasers. To the extent that other purchasers exercise their option to pick up and purchase the capacity, the initial purchaser is relieved of a standby charge (Sullivan, 2004). The opportunity, given by the option capacity, to mitigate the buyer's damages when he decides to "abandon" the contract can be worth and should be evaluated by the buyer in order to determine the fair value of the contract.

In order to evaluate the flexibility added by the "option capacity" clause in the take-and-pay and take-or-pay contracts, the problem can be modeled using an approach based on Real Options. According to this approach, the buyer purchases the good for the price P and holds a put option on the good with the exercise price x. For example, if the buyer's promise is enforced by damages, then d is the damage and x is the difference between the contract price and this damage (or $P = d + x$). In other words, x is the "recovered" value.

Thus the possibility of the buyer to decline to take up its allocation of the product can be modeled as an abandonment (put) option where the damage is the standby charge. Therefore, the difference between the total cost on the purchased amount, as indicated in the contract terms, and the standby charge is the exercise price of the put option, or the "avoided loss". The cost of the other supply source (with the assumption that the business carries on as usual after breaching the contract) represent the underlying asset, by using the real option terminology. In this case the exercise price is not fixed a priori. In fact, with the option capacity clause, the standby charge may not be paid by the initial off-taker if other purchasers decide to exercise their (call) option to pick up the available capacity. This process is illustrated in Figure 8.

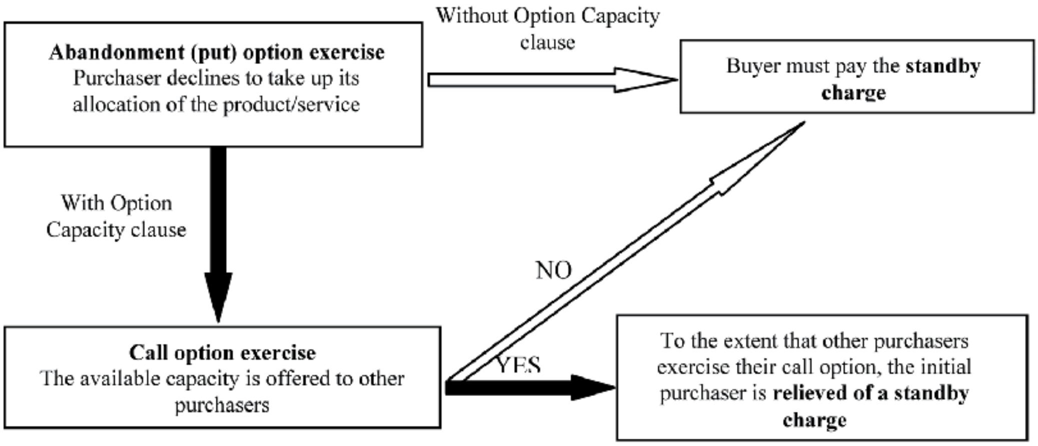

Abandonment (put) option exercise
Purchaser declines to take up its allocation of the product/service
Without Option Capacity clause
Buyer must pay the standby charge
With Option Capacity clause
NO
YES
Call option exercise
The available capacity is offered to other purchasers
To the extent that other purchasers exercise their call option, the initial purchaser is relieved of a standby charge

Fig. 9. Model of off-take contracts based on real options

This way, the damage of the purchaser, and therefore its risk, is mitigated and the off-take contract has more value. A computational model was developed for estimating the value of the flexibility given to the off-taker by the option capacity clause in take-or-pay and take-and-pay contracts[2].

The probabilistic input variables of the model are of two kinds: quantitative variables which were modeled as BetaPert distributions and subjective probabilities of some discrete events.

The experts were asked to give three-point estimates for the variables of the first group in terms of trapezoidal fuzzy numbers rather than crisp numbers, as the example in Table 3 shows.

	minimum	most likely	maximum
Probabilistic Variable 1	(a_1, a_2, a_3, a_4)	$(\beta_1, \beta_2, \beta_3, \beta_4)$	$(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$
…			
Probabilistic Variable n	$(\delta_1, \delta_2, \delta_3, \delta_4)$	$(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$	$(\eta_1, \eta_2, \eta_3, \eta_4)$

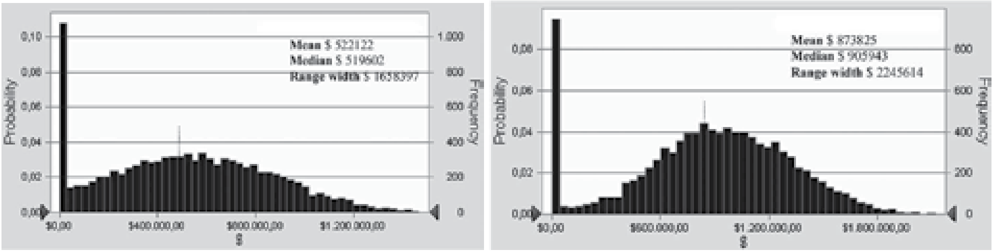
Table 3. The fuzzy three-point estimates of quantitative variables

The second group includes variables related to the subjective probability of some discrete events. The experts were asked to select for each of these events the probability phrase that best fits their perception. The linguistic variables associated with these phrases are expressed in terms of trapezoidal fuzzy numbers rather than crisp numbers (see Table 2).

The fuzzy Delphi procedure was used for eliciting the required probabilistic variables from experts.

The outputs of the model are probability distributions. In this case they represent the values of the (put) option to breach the contract with and without the option capacity clause as well as the value of the flexibility added to the contract by this clause. They are shown in Figure 10 and 11.

[2] Details about the model as well as its equations are described in (Costantino and Pellegrino 2011).



(a) (b)

Fig. 10. Comparison of the value of the option to breach the contract without (a) and with (b) the option capacity clause

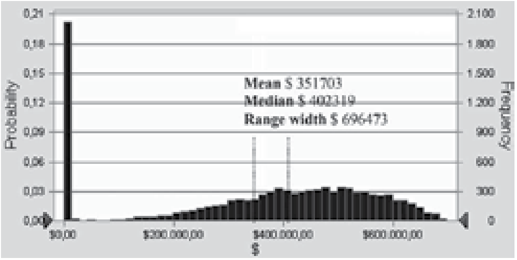


Fig. 11. Benefits given by the option capacity clause

The case shows some limitations of the developed approach. On one hand, in fact, the modeling of the decision making process gives more reliable results than traditional "black-box" approaches. On the other hand, it could seem to be more sophisticated than the application of a simple formula like the Black-Scholes one, and discourage its use by managers. For this reason, it could be important to apply the approach to several cases in order to give practitioners examples of how to use it.

6. Future work and closure

This research has developed a methodological approach to analyzing and assessing the value of real options in real world investments and strategies. The developed approach can be applied to estimating the value of the managerial flexibility (i.e., real options) in any engineering fields characterized by issues and sources of uncertainty that differ substantially from those typical of the financial world where option theory originated. In other words, the approach can be used any time money is spent in order to create a flexible project or strategy able to react to future uncertain events and, consequently, exploit opportunities and limit losses generated by uncertainty.

The proposed approach combines in a consistent and original way two well-known techniques, namely the Monte Carlo simulation for real option pricing and the fuzzy-Delphi method for eliciting, when historical data miss, probabilistic input variables from the knowledge of more than one expert in a consistent, structured and transparent way.

The proposed approach to modeling and valuing real options in real world problems (1) is more practically oriented than the traditional financial techniques, (2) is easier to understand and implement and, therefore, more likely to be used, (3) can model and value real cases having multiple uncertainties with multifarious features (such as technical, economic, and so on), (4) can model and simulate decision maker's behavior about the real option (i.e., managerial flexibilities) exercise, (5) can make real option valuation accessible to corporate managers or practitioners, strategists, and other decision makers, which could not have a financial background, (6) can offer a guide or a tool to support the decision maker rather than an exact numerical valuation as in the case of financial methods.

The proposed research has important and practical merits. Foremost, the potential of the methodology developed is its flexibility for the user modeling, reducing the problem of "curse of modelling" and permitting a large popularization of real options modeling for people without a comprehensive finance knowledge. The "transparency" of the approach would make managers able to use real options for estimating the value of the managerial flexibility. Managers can consider and manage the implication of the uncertainty, avoiding "myopic" decisions due to traditional DCF techniques which ignore the managerial flexibility. Findings of this research, and particularly the methodological framework of the MCS and fuzzy Delphi-based approach, are applicable to any problem, which implies real options, encountered in any engineering fields, such as natural resources, manufacturing, business investment, R&D, and so on. As a consequence, the application of real options to engineering issues, which is now still "confined" to a pure conceptual and theoretical level, just as "way of thinking" (due to the complexity of most of the traditional techniques of real option pricing), could be addressed to a practical level by using this "practically oriented" approach.

Future work will focus on applying the developed approach on several cases from real world in order to tests it and give practitioners examples of how to use it.

7. References

Amram, M. & Kulatilaka, N. (1999). Disciplined decisions: aligning strategy with the financial markets. *Harvard Business Review*, Vol.77, No.1, pp. 95–104.

Amram, M. & Kulatilaka, N. (2000). Strategy and Shareholder Value Creation: The Real Options Frontier. *Journal of Applied Corporate Finance*, Vol.15, No.2, pp. 15–28.

Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, Vol.81, pp. 637–659.

Borison, A. (2003). Real options analysis: where are the Emperor's clothes? *Proceedings of Real Options Seventh Annual International Conference*, Washington DC, July 10-12, 2003.

Boute, R.; Demeulemeester, E. & Herroelen, W. (2004). A real options approach to project management. *International Journal of Production Research*, Vol.42, No.9, pp. 1715–1725.

Brandão, L.E.; Dyer, J.S. & Hahn, W.J. (2005). Using Binomial Decision Trees to Solve Real-Option Valuation Problems. Decision Analysis, Vol.2, No.2, pp. 69–88.

Cheng, C.H. & Lin, Y. (2002). Evaluating the best main battle tank using fuzzy decision theory with linguistic criteria evaluation. *European Journal of Operational Research*, Vol.142, pp. 174–186.

Cheng, C.H. & Lin, Y. (2002). Evaluating the best main battle tank using fuzzy decision theory with linguistic criteria evaluation. *European Journal of Operational Research*, Vol.142, pp. 174–186.

Chiara, N. (2008). *Class Notes for Manage Civil Infrastructure Systems*. Columbia University - Civil Engineering and Engineering Mechanics.

Chiara, N.; Garvin, M.J. & Vecer, J. (2007). Valuing Simple Multiple-Exercise Real Options in Infrastructure Projects. *Journal of Infrastructure Systems*, Vol.13, No.2, pp. 97–104.

Copeland, T. & Antikarov, V. (2001). *Real Options*. Texere LLC, New York.

Copeland, T. & Weiner, J. (1990). Proactive management of uncertainty. *The McKinsey Quarterly*, Vol. 4, pp. 133–152.

Cortazar, G. (2001). *Simulation and numerical methods in real options valuation*. Pontificia Universidad Catolica de Chile-General.

Costantino, N. & Pellegrino, R. (2010a). Choosing between single and multiple sourcing based on supplier default risk: A real options approach. *Journal of Purchasing & Supply Management*, Vol.16, pp. 27–40.

Costantino, N. & Pellegrino, R. (2010b). Evaluating Risk in Put-or-Pay Contracts: An Application in Waste Management Using Fuzzy Delphi Method. *Journal of Applied Operational Research*, Vol. 2, No. 1, pp. 62-70.

Costantino, N. & Pellegrino, R. (2011). Risk mitigation in take or pay and take and pay contracts in project financing: the purchaser's perspective. *International Journal of Project Organisation and Management*, Vol. 3, Nos. 3/4, pp. 258-272.

Costantino, N.; d'Amato, M. & Pellegrino, R. (2009). A Real Options and fuzzy Delphi-based approach for appraising the effect of an urban infrastructure on surrounding lands. *Fuzzy Economic Review*, Vol.14, No. 2, pp. 3-16.

Cox, J.; Ross, S. & Rubinstein, M. (1979). Option pricing: a simplified approach. *Journal of Financial Economics*, Vol.7, pp. 229–264.

Coy, P. (1999). Exploiting uncertainty: the 'real options' revolution in decision-making. *Business Week*, Vol.7, No.118.

Dalkey, N.C. (1969). *The Delphi method: an experimental study of group opinion*. The Rand Corporation, Santa Monica CA.

Damodaran, A. (2001). The promise and peril of Real Options. Stern School of Business, 2001.

de Finetti, B. (1931). *Sul significato soggettivo della probabilità*. Fundamenta Mathematicae, Vol.17, pp. 298–329.

de Finetti, B. (1968). Probability: the Subjectivistic Approach. In: *La philosophie contemporaine*, R. Klibansky, (Ed.), La Nuova Italia, Florence, pp. 45-53.

de Finetti, B. (1972). *Probability, Induction, and Statistics*. Wiley, New York.

de Neufville, R. (2002). *Class notes for Engineering Systems Analysis for Design*. MIT engineering school-wide elective - Cambridge, MA.

de Neufville, R.; Scholtes, S. & Wang, T. (2006). Real options by spreadsheet: Parking garage case example. *Journal of Infrastructure Systems*, Vol.12, No.2, pp. 107–111.

Dequech, D. (2004). Uncertainty: individuals, institutions and technology. *Cambridge Journal of Economics*, Vol.28, pp. 365–378.

Dixit, A. & Pindyck, R. (1994). *Investment Under Uncertainty*. Princeton University Press, Princeton (NJ).

Dixit, A.K. & Pindyck, R.S. (1995). The options approach to capital investment. *Harvard Business Review*, pp. 105–115.

European Commission (2008). Guide to Cost Benefit Analysis of Investment Projects. *Directorate General Regional Policy*, July 2008.

Harmantzis, F.C. & Tanguturi, V.P. (2007). Investment decisions in the wireless industry applying real options. *Telecommunications Policy*, Vol.31, pp. 107–123.

Hartmann, M. & Hassan, V. (2006). Application of real options analysis for pharmaceutical R&D project valuation - Empirical results from a survey. *Research Policy*, Vol.35, pp. 343–354.

Hassan, R. & de Neufville, R. (2006). Design of Engineering Systems Under Uncertainty via Real Options and Heuristic Optimization. In: *Proceedings of Real Options Conference*, New York, June, 2006.

Hassan, R. & de Neufville, R. (2006). Design of Engineering Systems Under Uncertainty via Real Options and Heuristic Optimization. In: *Proceedings of Real Options Conference*, New York, June, 2006.

Hillson, D. (2002). Extending the risk process to manage opportunities. *International Journal of Project Management*, Vol.20, pp. 235–240.

Hoffman, S. (2007). *The Law and Business of International Project Finance*. Cambridge University Press, Third Edition.

Hsu, C.C. & Sandford, B.A. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research & Evaluation*, Vol.12, No.10, pp. 1–7.

Johnson, S.; Taylor, T. & Ford, D.N. (2006). Using System Dynamics to Extend Real Options Use: Insights from the Oil & Gas Industry. In: *Proceedings of 2006 International System Dynamics Conference*, Nijmegan, The Netherlands, July 23-27, 2006.

Klir, G. (1995). Fuzzy *Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, Upper Saddle River, N.J..

Kogut, B. & Kulatilaka, N. (1994). Option thinking and platform investment: investing in opportunity. *California Management Review*, Vol.36, No. 20, pp. 52–71.

Kumar, R.L. (1996). A note on project risk and option values of investment in information technologies. *Journal of Management Information Systems*, Vol.13, No.1, pp. 187–193.

Lander, D.M. & Pinches, G.E. (1998). Challenges to the Practical Implementation of Modeling and Valuing Real Options. *The Quarterly Review of Economics and Finance*, Vol.38, pp. 537–567.

Lawson, T. (1988). Probability and uncertainty in economic analysis. *Journal of Post Keynesian Economics*, Vol.11, No.1, pp. 38–65.

Lazo, G.J.; Pacheco, M.A. & Vellasco, M.B.R. (2003). Real Option Decision Rules for Oil Field Development Under Market Uncertainty Using Genetic Algorithms and Monte Carlo Simulation. In: *Proceedings of Seventh Annual Real Options Conference*, Washington, DC, USA, July 10-12, 2003.

Lee, C.C. (1990). Fuzzy Logic in Control Systems: Fuzzy Logic Controller - Part I. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.20, No.2, pp.404–418.

Leslie, K.J. & Michaels, M.P. (1997) The real power of real options. *The McKinsey Quarterly*, Vol.3, pp. 4–22,.

Longstaff, F.A. & Schwartz, V (2001). Valuing American Options by Simulation: A Simple Least-Squares Approach. *The Review of Financial Studies*, Vol.14, No.1, pp. 113–147.

Luehrman, T.A. (1998). Investment opportunities as real options: getting started on the numbers. *Harvard Business Review*.

Myers, S.C. (1984). Financial theory and financial strategy. *Interfaces*, pp. 126–137.

Newman, M.E.J. & Barkema, G.T. (1999). *Monte Carlo methods in statistical physics*. Oxford University Press, New York.

Olafsson, S. (2003). Making decisions under uncertainty - implications for high technology investments. *BT Technology Journal*, Vol.21, No.2, pp. 170–183.

Parr, R.L. & Smith, G.V. (2005). *Intellectual property: valuation, exploitation, and infringement damages*, Wiley, New Jersey.

Pellegrino, R. & Costantino, N. (2011). Risk Mitigation in Take or Pay and Take and Pay Contracts in Project Financing: the Purchaser's Perspective. *International Journal of Project Organization and Management – Special issue on Risk Management in Projects and Enterprises*, Vol. 3, Nos. ¾, pp. 258-272.

Perlman, M. & Jr McCann, C. (1996). Varieties of uncertainty, In: *Uncertainty in Economic Thought,* C. Schmidt, (Ed.), Edward Elgar, Cheltenham, U.K.

Ponsard, C. (1988). Fuzzy mathematical models in economics. *Fuzzy Sets and Systems*, Vol.28, No.3, pp. 273–283.

Razavi, H. (1996). *Financing Energy Projects in Emerging Economies*. Pennwell Books.

Ryan, P.A. & Ryan, G.P. (2002). Capital budgeting practices of the Fortune 1000: How have things changed?. *Journal of Business and Management*, Vol.8.

Smith, J.E. & McCardle, K.F. (1998). Valuing Oil Properties: Integrating Option Pricing and Decision Analysis Approaches. Operations Research, Vol. 46, No.2, pp. 198–217.

Smith, J.E. & McCardle, K.F.(1999). Options in the real world: Lesson learned in evaluating oil and gas investments. *Operations Research*, Vol.47, pp. 1–15.

Smith, J.E. & Nau, R.F. (1995). Valuing Risky Projects: Option Princing Theory and Decision Analysis. *Management Science*, Vol.41, No.5, pp. 795–816.

Sullivan, R. (2004). *International Project Financing*, 4th Edition, Juris Publishing, New York.

Teach, E. (2003). Will options take root?. *CFO online*, July 1st, pp. 73–76.

Triantis, A. & Borison, A. (2001). Real Options: State of Practice. *Journal of Applied Corporate Finance*, Vol.14, No.2, pp. 8–24.

Trigeorgis, L. (1996). *Real Options*. The MIT Press.

Vose, D. (1996). *Quantitative Risk Analysis. A Guide to Monte Carlo Simulation Modelling*. John Wiley Sons.

Wang, T. & de Neufville, R. (2005). Real Options 'in' Projects, *Proceedings of Real Options Nineth Annual International Conference*, Prise, France, June, 2005.

Yager, R.R.; Ovchinnikov, S.; Tong, R.M. & Nguyen, H.T. (1987). *Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh*. Wiley, New York,.

Yeo, K.T. & Qiu, F. (2003). The value of management flexibility a real option approach to investment evaluation. *International Journal of Project Management*, Vol.21, pp. 243–250.

Yeo, K.T. & Qiu, F. (2003). The value of management flexibility a real option approach to investment evaluation. *International Journal of Project Management*, Vol.21, pp. 243–250.

Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, Vol.1, pp. 3–28.

Zhang, S.; Buurman, J. & Babovic, V. (2008). Design of a Maritime Security System under Uncertainty Using an Evolutionary Real Options Approach, *Proceedings of 12th Annual Real Options Conference*, Rio de Janeiro, Brazil, July 10-12, 2008.

10

Fire Analysis and Production of Fire Risk Maps: The Trabzon Experience

Recep Nisanci, Volkan Yildirim and Yasar Selcuk Erbas
University of KTU, Trabzon
Turkey

1. Introduction

Today fire has an important place among the causes of loss of life and property. To intervene as soon as possible is of great importance, to eliminate or reduce the destructive effects of fires. Effectiveness of the intervention is directly linked with instant access to the fire, but the topography of the Black Sea region makes it difficult to intervene in fires on time due to poor access. Therefore, in order to fulfill the needs of the fire service and get fires under control, systems supported by information technologies can be used.

Information Technologies (IT) that have developed rapidly in recent years are also used in fire management. Parallel to the development of Geographic Information Systems (GIS), it is possible to achieve efficient results in applications based on spatial information. Since GIS can analyse intensive data volumes and is highly effective in responding to spatial queries, it can be used in the analysis of data concerning urban fires. When the cost-benefit results of GIS are compared with economic losses that occurred as a result of fire, it can be argued that GIS provides a relatively economic solution. In practice, the formation of fire analysis for urban areas is not common. Similarly, in the literature, the studies based on GIS supported fire analysis are very limited, although there are several empirical models (Itoigawa *et al.*, 1989), represented by the model proposed by Zhao (2010). However, these models focus mainly on simulating the overall behaviours of urban fire (e.g. spreading speed of fire, size of the burned out area) and the physical aspects of an urban fire that spread from one building to another (Himoto and Tanaka, 2003). Urban fire is an important problem, not only for developing countries, but also for developed countries. In the United States in 2006, one person died in a fire accident approximately every 162 minutes on average, and one person was injured every 32 minutes (Karter and Stein, 2008). The threat of urban fire is a significant problem in the United States, with building fires being responsible for over 3,000 deaths, 15,000 injuries and $9.2 billion in fire-related property damage in 2005 (Yamashita, 2000). Each year, fire causes about 300,000 deaths globally and most of these occur in the home (Zhang *et al.*, 2006). This situation is no different for Turkey; in 2002, there were 42,367 cases of fire and the figure reached 93,601, in 2008, the loss of lives rose from 224 to 402 in that period (Table 1). In 2007, the total value of goods damaged due to fire was around $350 million (GDCD, 2009).

Year	Number	Increase (%)	Deaths
2002	42.367	-	224
2003	56.482	33.3	505
2004	60.801	7.6	330
2005	57.293	-5.8	290
2006	81.149	41.6	349
2007	94.353	16.3	358
2008	93.601	0.0	402
Total	**443.679**	**220**	**2234**

Table 1. Number of urban fires and deaths per year, from 2002 to 2008 in Turkey (GDCD, 2009).

In Turkey the Fire Brigade operates under the local municipal administration. Fire records are written on printed materials, for example, the notification form before the fire and the report issued after the fire. Sometimes, some municipalities also record this information in databases and these are shared with other institutions in a common database. One of the main problems showing in the records is the accurate determination of the fire's location. The General Directorate of Civil Defense collects the fire records and publishes them annually, however, this information is generally not supported with a fire risk map indicating the locations of urban fires.

For effective fire management, conventional fire records must be supported with maps and there should be dynamic integral spatial data, such as the location of hydrant access routes together with limitations and information on risk areas. For effective firefighting it is of crucial importance that there are sufficient fire hydrants and they are appropriately maintained. In the fire reports there is evidence of a shortage of hydrants in addition to non-functioning hydrants and this has serious implications for the successful extinguishing of fires. In urban areas, especially in old settlements, many of the streets are narrow which can prevent the entry of fire trucks and first aid teams. Furthermore, the close proximity of buildings results in the rapid spread of fire. In many urban areas in Turkey there are old wooden houses and dilapidated unoccupied properties, and these can be more seriously damaged than other buildings.

In order to effectively fight fires and potentially prevent fires, a large amount of data needs to be collected. This should include: 1) a full description and location of the properties in the area, 2) occupants of the buildings with special attention to the children, the elderly, physically and mentally disabled, 3) location of hydrants and other water sources and 4) any sources of risk in the district or area. This data needs to be in a common database and GIS can play a significant role in the accumulation and maintenance of information. In preparation for fire management, GIS can help to determine the best distribution of hydrants, location of fire stations, classification of fire regions according to fire type and the creation of region specific early intervention plans (Nisanci, 2010).

Significantly, to eliminate or reduce the destructive effects of fires, fire risk management is a vital element for those who make decisions (e.g. local governments and municipalities) concerning fire. For this purpose, GIS - through effective spatial data storage and query - can produce dynamic fire maps. In the study presented in this chapter, the city centre of

Trabzon in Turkey was selected as the pilot area for the establishment of a sample fire database based on GIS and as the basis of sample spatial queries in support of fire management. For this pilot application it was necessary to perform a number of tasks. First, according to months, the fire records, fire types and fire distribution of the city centre between the years 2005 to 2008 were examined and analysed. Second, related fire data were imported into a computer environment as a GIS supported database. Later, the data in the database was visualized, analysed and queried in order to demonstrate the capabilities of the system. Specifically, an analysis of fire hydrant location was carried out and the related needs were analysed. The local fire occurences showed that 51.50% of the area was under low risk, 34.40% of the area was under moderate risk and 14.10% of the area was under high risk. Additionally, spatial queries were performed and pixel-based risk maps were produced.

2. Materials and methods

The GIS-based fire analysis method comprises some procedural steps, such as optimal location modelling, time modelling, incident trend modelling data collecting, data modelling and analysis/queries.

2.1 Fire risk analysis with GIS techniques

Although there are many formal definitions of GIS, for practical purposes GIS can be defined as a computer-based system to aid in the collection, maintenance, storage, analysis, output and distribution of spatial data information (Bolstad, 2005). GIS is an important and efficient tool that can be used by local administrations to minimize natural disasters. Thus, GIS technologies have been used in fire analysis related to the optimum location of fire stations, for example, Habibi *et al.* (2008) has made spatial analysis of urban fire stations in Tehran, using an Analytical Hierarchy Process (AHP) and GIS. The authors stated that using models and software in urban planning has become prevalent in response to the complex dimension of urban issues and the role of many different indicators in this field. Yang *et al.* (2004) also carried out studies concerning the selection of fire station locations using GIS. Jasso *et al.* (2009) have stated that location information of fires from 911 emergency calls could not be determined accurately. GIS by matching address information with coordinate information directly helps in the determination of places of fires or accidents in the shortest time. The literature reveals an increasing use of GIS in the fire service in the last decade (Corcoran *et al.*, 2007). Fire station locations, the shortest path and incident density analysis are important steps for effective fire analysis. In this context, forming optimal location models is very important and this is discussed in the next section.

2.2 Optimal location model

Optimal locations were primarily modelled by Launhardt and Weber (Weber, 1909) to determine the optimal locations for industry plants, according to access to markets and raw materials within a continuous area of possibilities, and in order to minimize the transport costs to serve a spatially distributed demand. Optimal location theory and modelling of facilities has seen important developments since the 1980s and allow analysing the spatial organisation of human activities (Carreras and Serra, 1997). The transport network often plays an important role in these models (Thomas, 2002).Within the large variety of location problems reviewed by ReVelle and Eiselt (2005), one can distinguish between two types of models, those dealing with d-dimensional space and those dealing with networks. Network-based measures are determined by summing the lengths of links or other criteria related to the length (e.g. distance-time, distance-cost). Better than coordinate-based measures (like straight distance), it takes into account the heterogeneity of space between starting and ending points along the network (Smith, 2003). The characteristics of the transport infrastructures enables accurate modelling of the cost of travel between locations, which is likely not to be uniform along the road network.

Within network facility location problems, network-based plant and warehouse location studies have been developed since the 1960s, which mainly emphasises the minimum distance location of facilities (called p-facilities) on a network of nodes, known as the p-median problems (Daskin, 1995; Thomas, 2002; Torsten and Densham, 2003; ReVelle and Eiselt, 2005). This has lead to the development of a variety of algorithms to address specific p-median problems (e.g. Kariv and Hakimi, 1979; Peters, 1996). More recently, particularly in retail location problems, the p-median model is typically applied when locating p facilities to serve demand, generally disperse (e.g. multiple points customer) in order to minimize the aggregate transportation cost (Daskin, 1995; Ghosh et al., 1995).

2.3 Time modelling

Utilising a fire station layer and a street layer, response time analysis can be performed. A street layer is often represented in GIS as a series of lines that intersect on the map, creating a GIS street network. Each street line segment between intersections contains attribute information, such as road type, distance and travel speeds (miles or kilometres per hour). This allows users to identify a station location, specify a travel time and run a network analysis. The result will be displayed by an irregular polygon around the station that illustrates where the fire apparatus could travel in any direction for the specified time. This type of analysis can be performed on a single station or simultaneously on all stations to analyse gaps in coverage, establish run orders and more (ESRI, 2006).

2.4 Incident trend modelling

Incident trend analysis can be done quickly, displayed logically and understood easily. For example, a GIS user could request to see arson fires that occurred between the hours of 1:00 a.m. and 5:00 a.m. on Saturdays in fire districts 1 and 2. GIS will interrogate the record database and place points on the map that match the request. These types of analyses provide decision support for issues related to fire prevention, staffing requirements and apparatus placement/deployment (ESRI, 2007).

2.5 Data collecting

GIS software adds intelligence to spatial data, whether the data is generated in the field with GPS or remotely with lidar and photogrammetry. You can enter raw data, measurements and field sketches directly into the GIS, enabling you to efficiently manage your data in a geodatabase with other spatial information. You can use GIS technology for collecting, importing, converting and storing spatial measurement and computational fabrics. You can integrate computations, such as COGO (COGO is a ArcGIS extension used in civil engineering or surveying for solving cooridante geometry problems). The COGO extension provides a dialogue for setting common properties for the coordinating geometry tools and traverse least squares, and pre-existing networks, as well as importing spatial data feature classes and relationships (ESRI, 2006).

2.6 Data modelling

Data models are the rules the GIS follows, such as "county lines do not overlap", and are essential for defining what is in the GIS, as well as supporting the use of GIS software. All spatial data models fall into two basic categories:

1. Vector data model: discrete features, such as customer locations and data summarised by area, are usually represented using the vector model.
2. Raster data model: continuous numeric values, such as elevation, and continuous categories, such as vegetation types, are represented using the raster model (URL1, 2011).

2.7 The importance of a fire information system (FIS) in urban areas

Unlike a flat paper map, a GIS-generated map can represent many layers of different information. This representation provides a unique way of thinking about geographic space. By linking map databases, GIS enables users to visualize, manipulate, analyse and display spatial data. GIS technology can create cost-effect and accurate solutions in an expanding range of applications. GIS displays geographic data as spatial data layers. The data layers used in this study in the design of GIS for fire are given in Figure 1.These data layers and their attribute data can be expanded according to needs analyses and user requests. These needs analyses should include queries and the requirements for effective fire fighting before, during and after the fire (see Table 2).

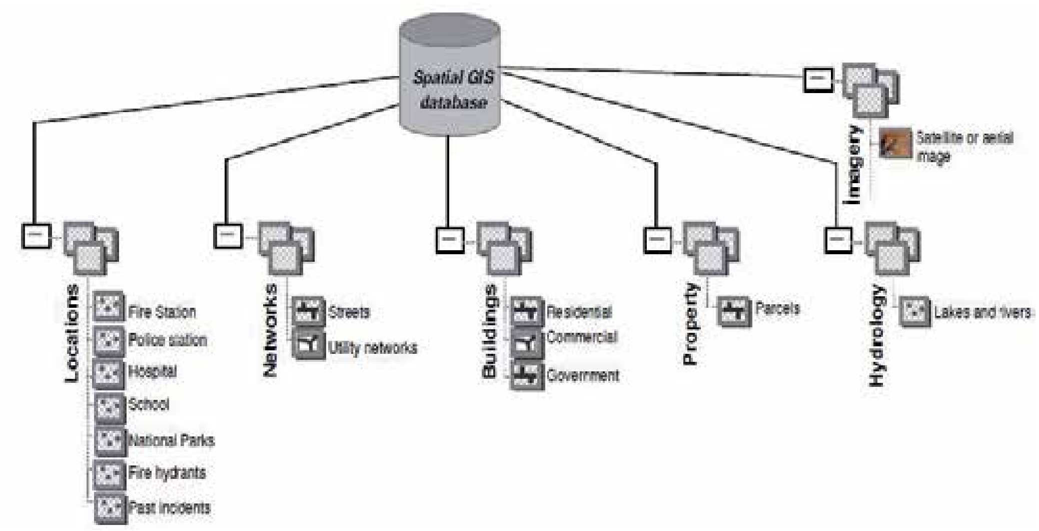


Fig. 1. Main spatial data layers used in GIS related to fire incidents (Nisanci, 2010).

A Fire Information System (FIS) which includes all spatial and non-spatial data related to fire (see Figure 2) should be connected to the GIS database so that when needed any query can be acted upon to receive the required information from the GIS database. For example, a request for information about the inhabitants living in the building that is on fire can be obtained immediately. The number of elderly people who have died in fires in many countries is not to be underestimated. If we can gain information about the number of residents and their physical or mental condition via a social health database, this can save many lives. Additional information on the residents should be obtained from the civil registry, social health organisations and amalgamated into the FIS database. Furthermore, the FIS should be able to access information in the GIS database at the moment of the fire about the building (architectural details including fire escapes, building structural projects, the location of gas and electricity supplies). In this way, fire teams will be able to act quickly and effectively, and more importantly, with greater safety when dealing with a fire. In addition, the address of the building on fire or the address of the reporter should be acquired by integrating GIS database for fire management with the database of the Telecommunications Directorate and Civil Registry as shown in Figure 2. In this way, through this distributed data model, the FIS will be optimized and work more efficiently.

In order to maximize the fire fighting ability of the Fire Brigade and thus save lives, prevent injury and reduce economic losses, the Fire Directorate should ensure that the records are regularly updated and undertake such studies as necessary to maintain an accurate database that can be shared with relevant parties. GIS is a complement of systems which can produce solutions by finding answer to questions such as how to prepare for fires in areas where fire truck access is restricted (or prevented), ascertaining where there is a need for more fire hydrants and whether maintenance is required in existing hydrants, if the current distribution of fire stations is adequate and if they are appropriately located and where the areas of particular risk are. This paper describes the development of the FIS model for Trabzon city using GIS technology and shows how appropriate databases can be created to respond to queries and provide information for analysis (Nisanci, 2010).

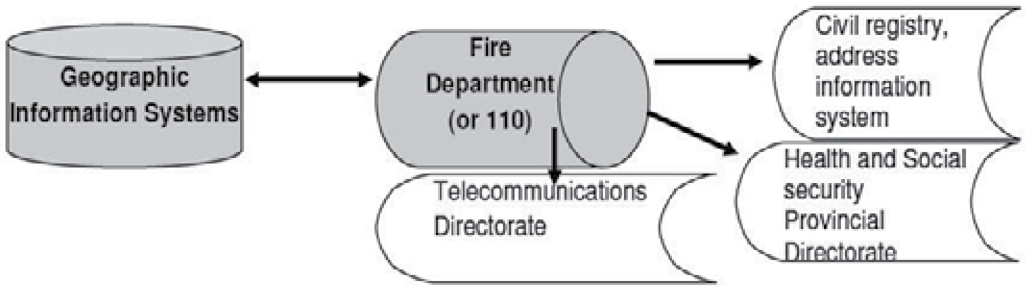

Geographic Information Systems
Fire Department (or 110)
Civil registry, address information system
Health and Social security Provincial Directorate
Telecommunications Directorate

Fig. 2. Basic structure of distributed data model for FIS.

Before	At The Incident Place	Before and After Fire
Location and spatial distribution of hydrants and maintenance status.	Where is the nearest hydrant?	Check on maintenance status.
Are fire stations located appropriately?		Are fire stations located appropriately?
Areas that cannot be accessed by fire trucks and other emergency vehicles.		Areas that cannot be accessed by fire trucks and other emergency vehicles.
Road information.		Update road information.
Where are tall buildings requiring special fire fighting techniques?	Where is fire exist?	
Building danger classification (high, medium and little danger).	Are there available building plans (architecture, machine, etc.)?	Revise building danger classification.
Building address and database.	Are there vulnerable people living in house (children, elderly, disabled etc.)?	Update database information.
	Location of the nearest first aid centre (hospital, health centre)?	
Where are the water depots and other water resources in the urban environment?	Where is the nearest water source?	Revise information on water depots and other water resources, and their current condition.
Analysis of fire risk regions.		
Where are areas that have a risk of explosion (petroleum stations, transformers)?		Update information about areas that have a risk of explosion.
Fire database to be maintained (type, date, time, dead-causally, economic damage, etc).		Fire database to be updated.

Table 2. Fire information system needs analysis (Nisanci, 2010).

2.8 Fire analysis for Trabzon city

The city centre of Trabzon in Turkey (see Figure 3) was selected as the study area and the records pertaining to fire fighting and prevention held by the administration of Trabzon municipality were the main data source for this study. In the urban information system that was previously created through a joint project undertaken by Karadeniz Technical University and Trabzon municipality, the fire records were added to a spatial database. With this study, in the first stage, through a specially developed interface all the existing paper records from 2000 to 2008 from the Fire Directorate records were transferred to the database. This information included: fire type, number of dead/casualties and time taken to access the fire. In the second stage, fire address records were matched with the addresses in the database. As can be seen in Table 3, the number of fires has increased since 2002 and although the number of deaths was only two over the whole period, the number of people injured has increased (Nisanci, 2010).

From fire records in Trabzon province between the years 2002 to 2008, the incidents were classified according to the basic cause of the fire. The number of cases and their causes were as follows: 641 – solid fuel cookers and stoves, 537 – chimneys, 224 – electricity, 80 – Liquefied Petroleum Gas (LPG), 40 – central heating, 6 – fuel oil, 3 – chemicals and 1 – gas compression.

Year	Number of Fires	Increase on Previous Year(%)	Deaths	Injuries
2002	151	-	-	2
2003	159	5.2	-	3
2004	227	42.8	1	3
2005	256	12.8	-	7
2006	233	-9.0	-	8
2007	241	3.4	1	6
2008	265	10.0	-	7
Total	**1532**	**75.5**	**2**	**36**

Table 3. Number of fires in Trabzon city centre from 2002 to 2008 (Nisanci, 2010).

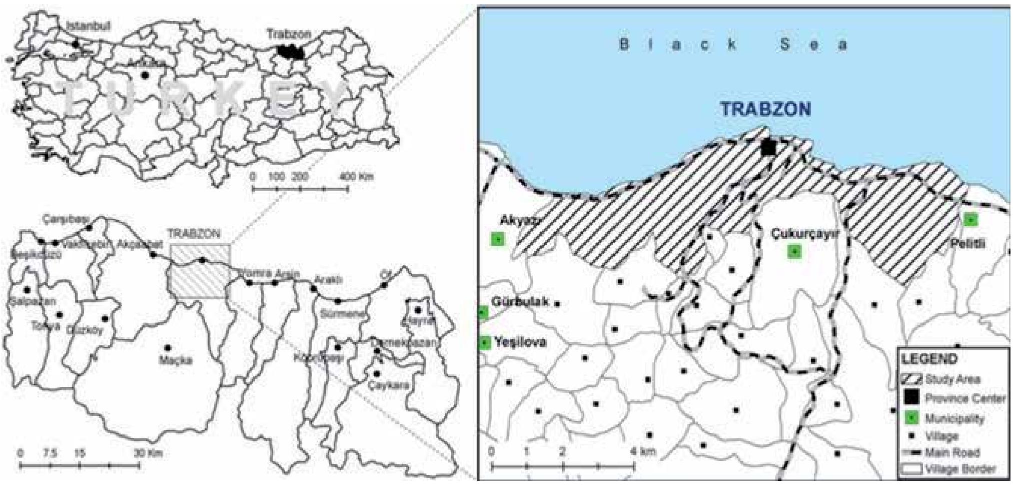


Fig. 3. Study area.

3. Case study and results

In Turkey, access roads to buildings for fire trucks are defined in Article 22 of the regulation on protection of buildings from fire. According to this, "inner access roads are the road providing access to a building from a main road. The normal width of inner access roads must be at least 4 m and in the case of a dead end street, it must be 8 m". Based on this Article of the regulation and the width of a fire truck being 3 m, the roads that cannot be entered were determined using road and building data from the database. For this analysis, buffer areas of 2 m around each building were created in the database as a buffer layer. Then, the road data was intersected with the buffer layer formed and finally, roads were determined where a fire truck cannot enter or where there is insufficient space to manoeuvre (Figure 4).

3.1 Spatial distribution analysis of hydrants

An important layer of an FIS is placement of hydrants. In the study area, the total number and location of hydrants were not known. Therefore in this study, using the Topcon GMS 2 DPGS device it's static accuracy is 3mm horizontal and 4mm vertical, the locations of known hydrants was found and transferred to the spatial database. It was also detected in the study that some hydrants are surrounded by plants (Figure 5), thus, they are not visible at night. In addition to the work on hydrants, the locations of water mains and depots within the topological structure were recorded in the database.

In clause 3 and 11 of Article 95 of the regulation on protection of buildings from fire, hydrants are to be placed within the system "to cover the immediate and nearby buildings and for fires not extinguished in the first intervention fire trucks should be able to gain access easily and use the hydrants to protect other buildings from the fire. The distance between hydrants is 50 m in very high risk regions, 100 m in high risk regions, 125 m in medium risk regions and 150 m in low risk regions."

From the fire intensity analysis produced as part of the study the Kermekaya district of Trabzon city was determined to be in very high risk area, it is therefore concluded that the existing four hydrants are insufficient and the distance between these fire taps is much greater than the required distance for high risk fire regions as stated in the regulations. Therefore, it is determined that in the Kemerkaya district, 50 additional hydrants are required, the locations of the existing and proposed hydrants are shown in Figure 6.

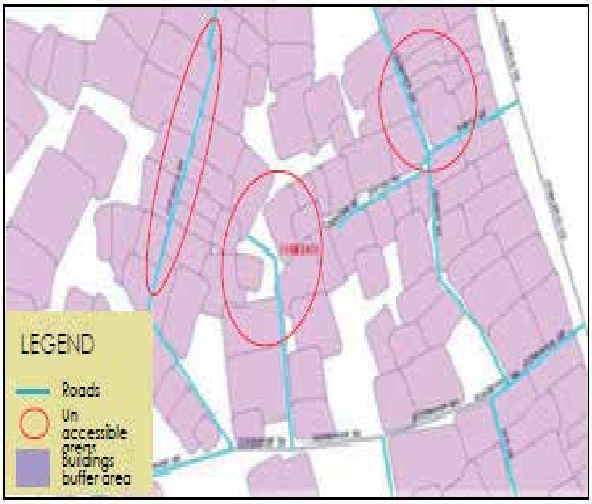


Fig. 4. Sample for streets that a fire truck cannot enter (Cömlekci district, Ortanca and Cemiyet Street) (Nisanci, 2010).

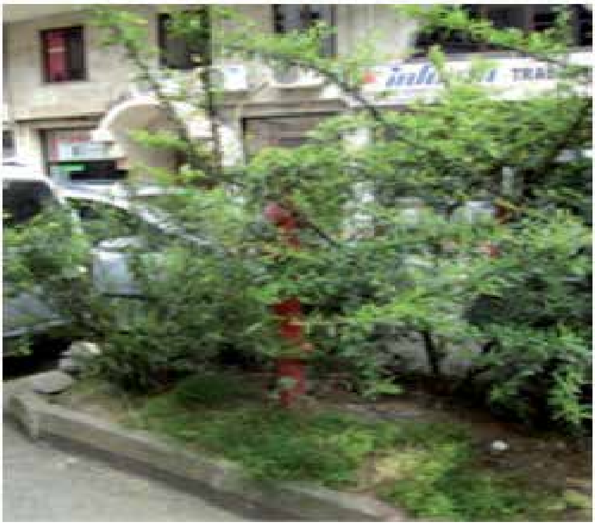


Fig. 5. Example of the lack of visibility of some hydrants covered by plan (Nisanci, 2010).

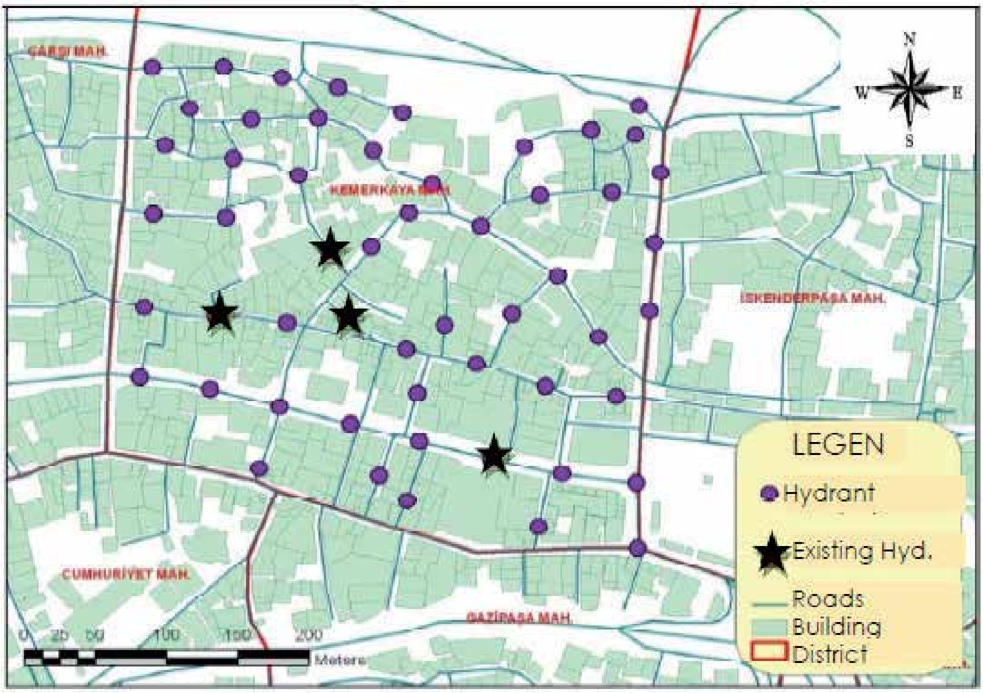


Fig. 6. Spatial distribution of fire hydrants in the Kemerkaya district in Trabzon city (Nisanci, 2010).

3.2 Analysis of the fire number according to months

Between 2002-2008 fires per month were analysed and a 'Fire Distribution According to Months' database was created (Table 4).

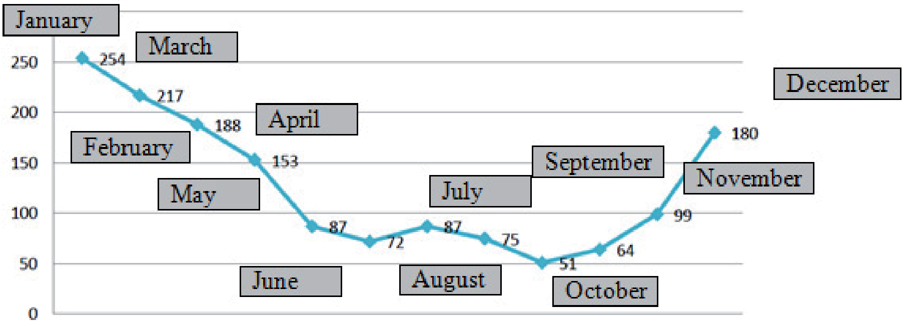


Table 4. The results of this analysis show fires were seen more in the winter months, such as December-January-February.

3.3 Analysis of accessibility to fire in the shortest time

Access to a fire incident and making an intervention in the shortest time has great importance for extinguishing a fire. In this analysis, the areas that a fire truck can reach according to certain time intervals and speed were determined (Figure 7). In this analysis, the speed of a fire truck, its access time to the incident site (obtained from previous records in the fire database) and the distance of fire site to the fire station were examined. The areas to be accessed in 3, 5 and 7 min were determined using a truck traveling at 45 km/h. Accordingly created maps are shown below (see Figures 8-9-10). The speed of the fire vehicle was calculated as an average value from the access time of the fires that occurred in previous years. It was discovered that in particular, the east of the city has the most serious risk in terms of the lack of rapid accessibility to the fire location.

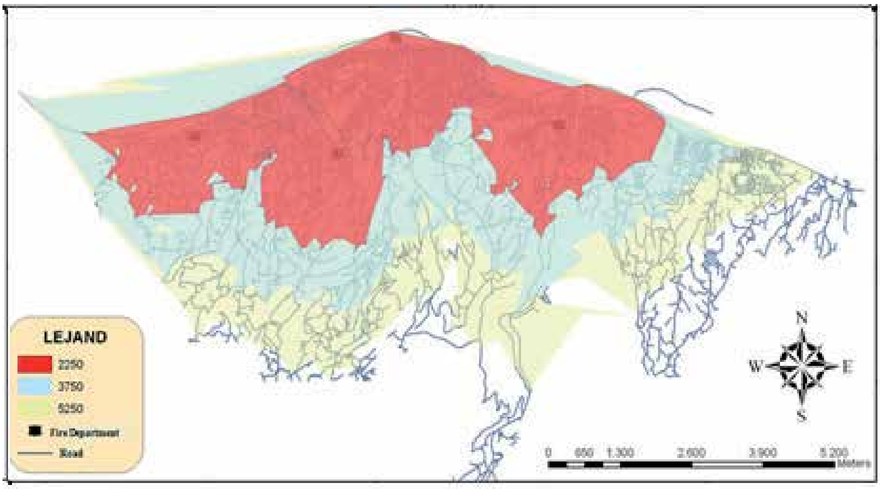


Fig. 7. A regional map of fire truck access according to time.

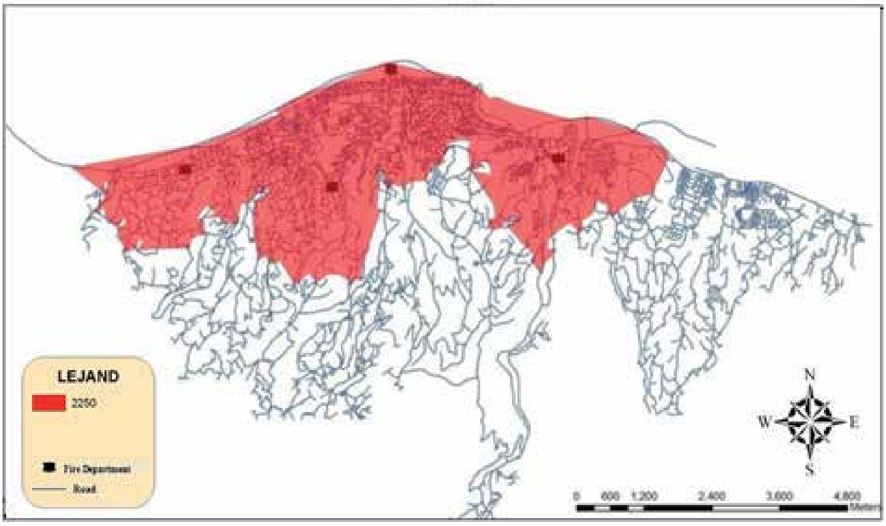


Fig. 8. The areas that a fire truck can access within 3 min in Trabzon city.

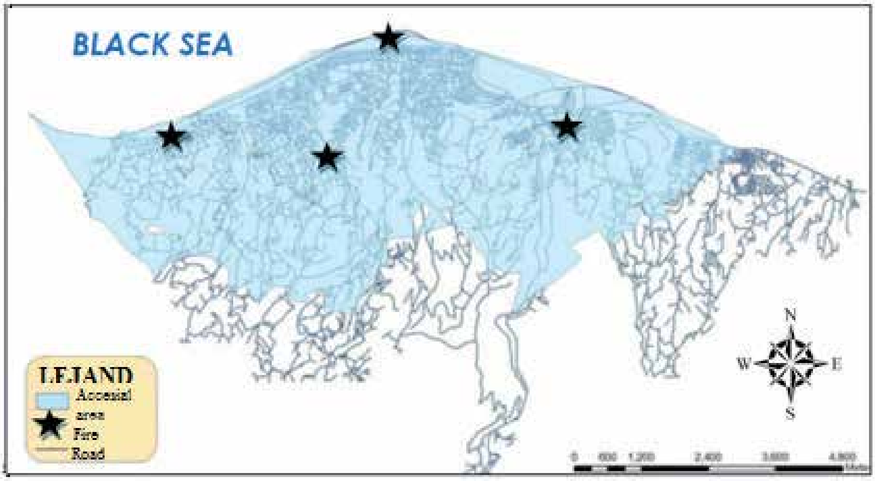


Fig. 9. The areas that a fire truck can access within 5 min in Trabzon city.

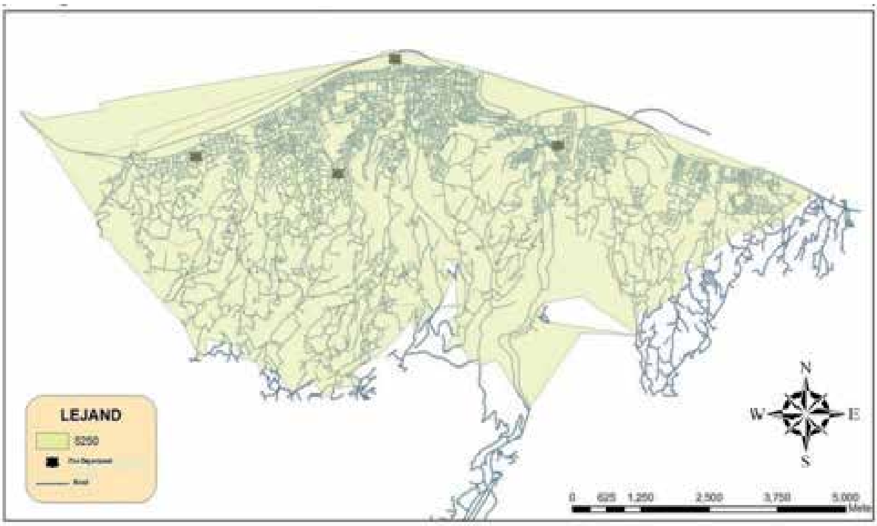


Fig. 10. The areas that a fire truck can access within 7 min in Trabzon city.

3.4 Analysis of fire types

Fire types between 2002-2008 were analysed by creating a database from fire reports as shown in Table 5. According to this fire types created map are shown below (Figure 11).

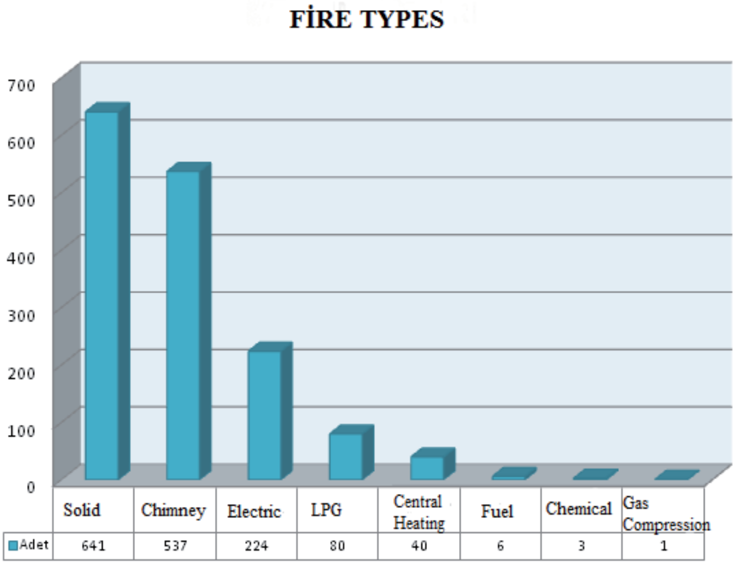


Table 5. Types of fire that occurred between 2002-2008 (Nisanci, 2010).

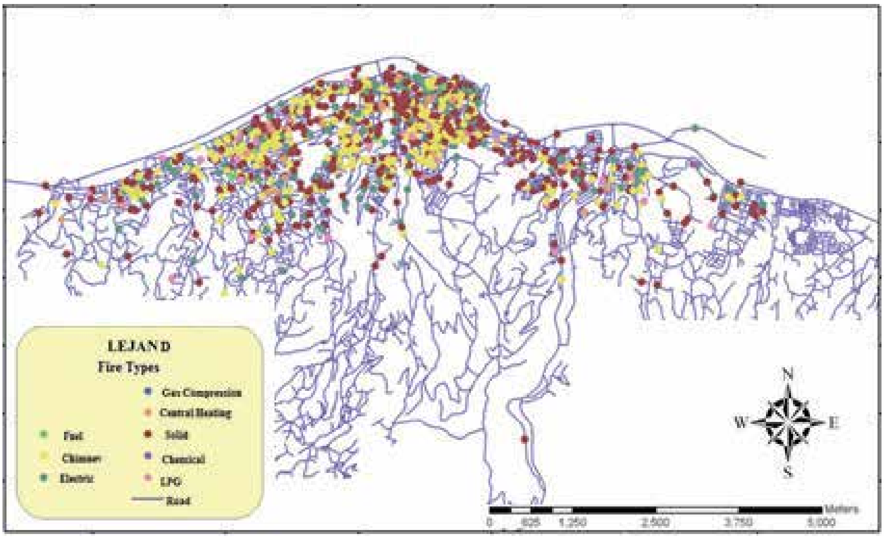


Fig. 11. Fire types map.

3.5 Determination of high risk areas

Article 17 of the regulation on the protection of buildings from fire evaluates high risk sites at the following places:

a. Where flammable and detonable gases are stored, where the transport loading and unloading sales operations of LPG, natural gas and similar gases are carried out.
b. Where flammable materials that can explode easily with the effect of heat and pressure are found, where munitions, gunpowder, dynamite and similar materials are produced, stored and sold.
c. Where inflammable liquids are produced, stored and sold. In consideration of these criteria, together with the analyses and queries that were conducted, 99 high risk areas in Trabzon were determined in accordance with the regulatory information given above. The distribution of these sites is shown in (Figure 12).

3.6 Determination of fire density

Fire density is determined according to the numbers of fires that occurred in a given period of time. The required data for this was obtained from the database created for Trabzon city. For this analysis, first, a grid network of 500 by 500 m formed to cover all the city was overlaid with the QuickBird satellite image of the city (Figure 13). Then this network was transformed into ArcGIS topological data, fire density of each pixel was detected according to the number of fires within each pixel. This process provides a visible display of the fire density in the city. When the fire density map was examined, a total of 369 fires were seen to be concentrated on the city centre with the density decreasing from the city centre outwards (Figure 13). In other words, there is an inverse proportion between the distance to the city centre and number of fires and the density. The reason for this distribution arises from the city centre being an old settlement, with old buildings and heating being provided by coal or gas burning stoves.

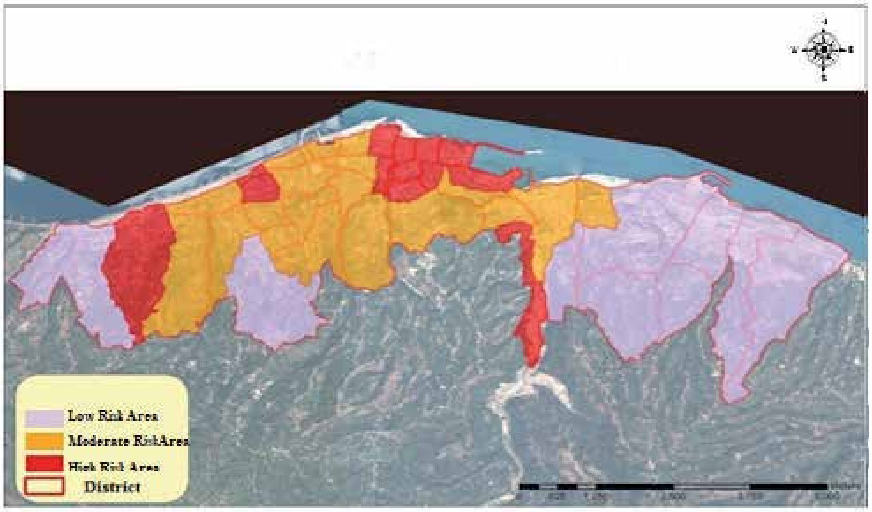


Fig. 12. High risk places found in and around Trabzon city centre (Nisanci, 2010).

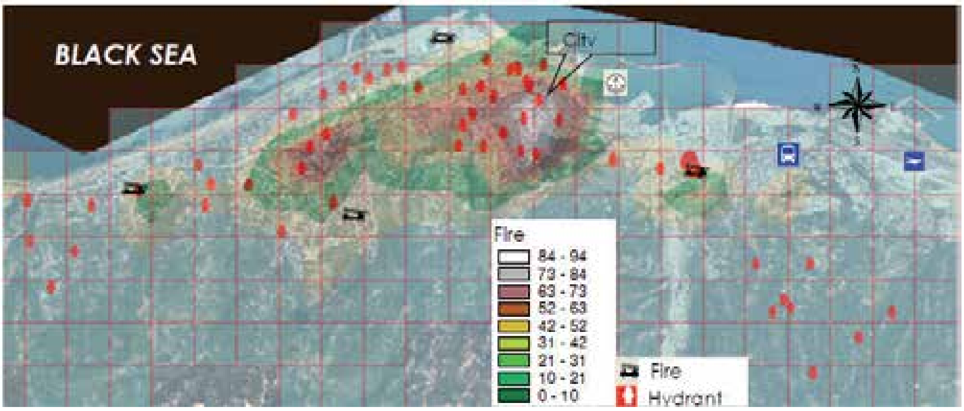


Fig. 13. Fire density risk map of Trabzon city and vicinity (Nisanci, 2010).

4. Conclusion and future work

In conclusion, the fires that occurred in Trabzon city were recorded according to the numbering system and on a street basis in a GIS database. It is now possible to facilitate a more advanced analysis by associating the fire site database with the building and cadastral parcel database. The data comprising graphical database within the FIS pilot application for Trabzon city consists of spatial data such as fire stations and hydrant locations, roads, satellite images with high resolution, written and verbal details relating to buildings (such as residential, official facilities, factories, depots). Furthermore, a verbal database from attribute data was created, such as fire types, fire number etc. It was found that the spatial information of hydrants is deficient and current data on their usability is not available. It was also observed that there were administrative problems in terms of the purchase, maintenance and repair of hydrants. Furthermore, the distribution of hydrants is not compatible with the regulations, there are insufficient numbers of hydrants in the areas where fires are considered to be intensive. There are errors in the existing records of fire reports, such as address, fire occurrence and access time to the fire, however, these errors can be reduced to a minimum using the prepared interface. Furthermore, it has also been observed that a standard for spatial and attribute data with respect to fire records has not been developed.

In this study, the GIS analysis showed that were local fire occurences in 51.50% of the area under low risk, 34.40% of the area under moderate risk and 14.10% of the area under high risk. Problems will continue to be experienced in the access to many streets and it is important that more hydrants are installed in these areas to allow effective fire fighting. It has been concluded that the measures required had not taken in the intensive fire areas found at the end of the GIS analysis, although it is evident from fire records that the intensity of fires in these regions are high.

Today, fire is an important risk factor for urban areas. For this reason, firstly fire risk areas are established by using spatial databases in cities. The use of traditional statistical methods with address data, the number of fires and so on, is not adequate as a modern solution. Determination of the optimum route, accessibility analysis, emergency response management applications should be made by performing GIS-based fire scenarios on the risk areas. Especially in risk areas, fire hydrant areas should be optimized, and associated with a spatial fire database. In the event of fire, in order to minimize loss of life and property, it should be ensured that areas such as shopping centres, hospitals, schools, hotels and convention centres have fire escape plans and floor plans added to the fire databases.

5. Acknowledgments

The author would like to acknowledge the contributions of the Black Sea Technical University Scientific Research Project Unit.

6. References

Bolstad P (2005). GIS Fundamentals: A First Text on Geographic Information Systems, Eider Press, White Bear Lake, Minnesota.

Corcoran J, Higgs G, Brunsdon C, Ware A, Norman P (2007). The use of spatial analytical techniques to explore patterns of fire incidence: A South Wales case study, Computers, Environ. Urban Syst. 31: 623-647.

Daskin, M (1995). Network and Discrete Location. Models, Algorithns and Methods, New York, Wiley Interscience. 498 p.

Ghosh A, McLafferty S., Samuel Craig C. (1995) multifacility retail networks, in Drezner, Z (ed.) Facility Location: A Survey of Applications and Methods, Heidelberg, Springer Verlag, 571p.

General Directorate of Civil Defense (GDCD) (2009). http://www.ssgm.gov.tr.

Habibi K, Lotfi S, Koohsari MJ (2008). Spatial Analysis of Urban Fire Station Location by Integrating AHP Model and IO Logic Using GIS (A Case Study of Zone 6 of Tehran), J. Appl. Sci. 8(19): 3302-3315.

Himoto K (2003). A Physically-Based Model for Urban Fire Spread. Fire Safety Science 7: 129-140. DOI:10.3801/IAFSS.FSS. pp. 7-129.

Itoigawa E, Iwami T, Kaji T, Kawanaka T, Kumagai Y, Tukagoshi I, Masuyama T (1989). Study of Real Time System for Information Processing of Post-earthquake Fire, Fire Spread Prediction and Evacuation Control, Report of The Building Research Institute, No: 120.

Jasso H, Hodgkiss W, Baru C, Fountain T, Reich D, Warner K (2009). Using 9-1-1 call data and the space–time permutation scan statistic for emergency event detection, Government Information Quarterly, 26: 265-274.

Karter MJ, Stein JGP (2008). U.S. Fire Department Profile Through 2007, Fire Analysis and Research Division National Fire Protection Association, http://www.doi.idaho.gov/SFM/FDProfile_2007.pdf.

Nisanci R(2010). GIS based fire analysis and production of fire-risk maps: The Trabzon experience, Scientific Research and Essays Vol. 5(9), pp. 970-977.

URL1, (2011). Data Types and Models, http://www.gis.com/content/data-types-and-models (access date 04/08/2011).

Yamashita K (2000). Understanding Urban Fire Modeling Fire Incidence Using Classical and Geographically Weighted Regression, B.A., Western Washington University, Master of Arts, Department of Geography, USA.

Yang B, Viswanathan K, Lertworawanich P, Kumar S (2004). Fire Station Districting Using Simulation: Case Study in Centre Region, Pennsylvania, J. Urban Plan. Dev. 130(3): 117-124.

Zhang G, Lee AH, Lee HC, Clinton M (2006). Fire safety among the elderly in Western Australia, Fire Safety J. 41: 57-61.

Zhao S (2010). GisFFE—an integrated software system for the dynamic simulation of fires following an earthquake based on GIS, Fire Safety J. 45: 83-97.

11

Flood Risk Management in Rivers and Torrents

Luca Franzi
Regione Piemonte
Italy

1. Introduction

Among the meteorology-related risks, flooding is one of the first risks which human communities had to cope with in the past. From a historical point of view, the floodplains have been the preferred places for settlement of anthropogenic activities, because of the availability of richer soils, waters supplies, ways for transportation. Even if when speaking of floods and their effects the emphasis is generally placed on destructive forces, historically floods have also been considered as beneficial processes, providing the recharging of water sources, the fertility of soil, carrying nutrients and sediments, and assuring the rejuvenation of the river ecosystem (WMO, 2006). Even if the colonization of mountain regions is more difficult and generally offers less natural resources, evidence shows that also more impervious regions are urbanized, such as the torrent fans. This is as a reflection of a sharp population increase, an expanding economical growth, a great investment in infrastructures. As a consequence, structural control countermeasures have been traditionally placed along rivers and torrents, in the effort to reduce the risks and to assure a total protection against flooding and inundations.

Experience shows that the total protection is a myth (WMO, 2004), either along floodplains or torrents. Actually, due to an inadequate understanding of flood processes, to an increasing quantity and value of the settlements at risk, Flood Risk Management (FRM) strategies which aimed at keeping the floods inside the channels, proved to be unsuccessful.

The main purpose of this chapter is to focus on the different strategies which have been proposed by scientists, and applied by practitioners and decision makers to cope with floods and to manage risks due to inundations by rivers and torrents. The subject is challenging as it involves different disciplines, ranging from hydrology, structural engineering, geology, economy, politics and social sciences. A continuous interchange among the different disciplines is generally needed to find what strategies are best to follow, considering either the positive or the negative effects of flooding, balancing benefits and losses, structural and non-structural countermeasures.

The interest in this subject is double. First of all, flood risk management is a topical subject. Actually the European Union Commission has recently approved the directive on the assessment and management of flood risks, which obliges member States to establish flood risk management plans, coordinated at the level of the river basin districts, "with a view to avoiding and reducing the adverse impacts of floods" and addressing "…all aspects of flood risk management focusing on prevention, protection, preparedness, including flood forecasts and early warning systems and taking into account the characteristics of the particular river basin or sub-basin" (European Commission, 2007). The directive is an important step for enhancing the flood risk management practices in European countries.

Secondarily, as mentioned before, FRM requires a strong collaboration of many professionals, which generally work independently. Therefore researchers, technicians, engineers, geologists, sociologists, and politicians are equally involved, together with people. Experience shows that without a strong collaboration and interaction among them, the benefits for endangered populations are not equally distributed, the economical repayment is less, the engineering solutions show to be only moderately effective, politics in risk management prove to be inefficient. The collaboration implies a shared understanding about the risk concept (What's flood risk? Which are its components?) and a shared methodology to manage risk (How can flood risk be assessed, evaluated and treated?).

Therefore, in this chapter, following the approaches proposed in literature, a concept of *risk management* is proposed, with a presentation of some strategies for its management (Section 2). The aim is to show and discuss what is the state-of-the–art (Section 3), by virtually addressing to the professionals involved in flood risk management, comparing different approaches in FRM, showing the necessity to bring the flood risk management in a common context of scientific, political and public debate, as well as proposing a common understanding of concepts. In particular the concept of flood risk management is introduced as the main concept, instead of the earlier and narrower paradigms of flood defence and flood control. A more in-depth description of the FRM followed in the Northern part of Italy is given (Section 4). It will be shown, in particular, that the effectiveness of the applied strategies strongly depends on the uncertainties in the flood risk assessment. As a consequence, FRM strategies should be enough flexible to adapt to new circumstances and evidences, taking into account a good balance between planning and civil protection strategies.

2. Flood risk management strategies in literature and history

Flood risk management (FRM) strategies can be very different. The term itself "strategy" seems to be a simple term and can be defined as the set of activities that aim at influencing the world around us. Notwithstanding strategy research has developed different concepts of strategy. On this subject, history itself shows some representative examples of strategies (Kersting, 2008), often derived from practice, common sense and intuition and many handbooks have been recently published (Pettigrew *et al.* 2002, Easterby-Smith 2003, Pool et a., 2004) in scientific literature, with a more structured and less empirical approach.

Historically, the first FRM strategy had a style of acceptance of the flooding processes, as something people had just to live with (Kersting, 2008). This is the case of the populations which lived along the Euphtrates, the Tigris and the Nile. People simply built houses on piles to keep safe during floods, or look for safe higher grounds.

A first historical Italian example of FRM strategy debate, can be read in Tacitus (Tacitus, 115 AD), who reports that after the Tevere inundations (15 AD) a debate arose in the Roman senate on the best practices that should have been followed to reduce flood risks. Two strategies were proposed, that is the "do-nothing" one, implying no interventions, and the "structural" one, which implied the construction of dams and Tevere diversions. The latter approach prevailed (Tacitus, 115), after a discussion with land owners, which opposed against the damming of the Velino lake, which would have caused the permanent inundation of the surrounding territories. Tacitus reported that a participatory approach in the choice of the best solution was followed by Tiberius emperor, who rejected the proposals based on superstition or sibylline responses.

Historically the early FRM strategies focused on the implementation of flood control countermeasures, and are generally addressed to as "defence strategies" (Stalenberg and Vrijling, 2006). These flood control practices are widely documented during the Renaissance. For example, Coccapani's opera entitled "Trattato del modo di ridurre il fiume Arno in un canale" (Tractate on how to transform the Arno river into a channel, Coccapani, 1610) focused on structural controls of the river, either for reducing the inundation risk in the surrounding agricultural areas or for navigation purposes. In that opera, Coccapani harshly criticized the intervention already made, stating that "fluvial accidents [in Arno river] are not due to natural defects, but mainly to improper interventions, made by inexperienced architects". Also Galileo Galilei (Acanfora, 1990) confirmed that the structural measures to canalise Arno river have been extensively used for river "corrections", for navigation purposes (see Figure 1).

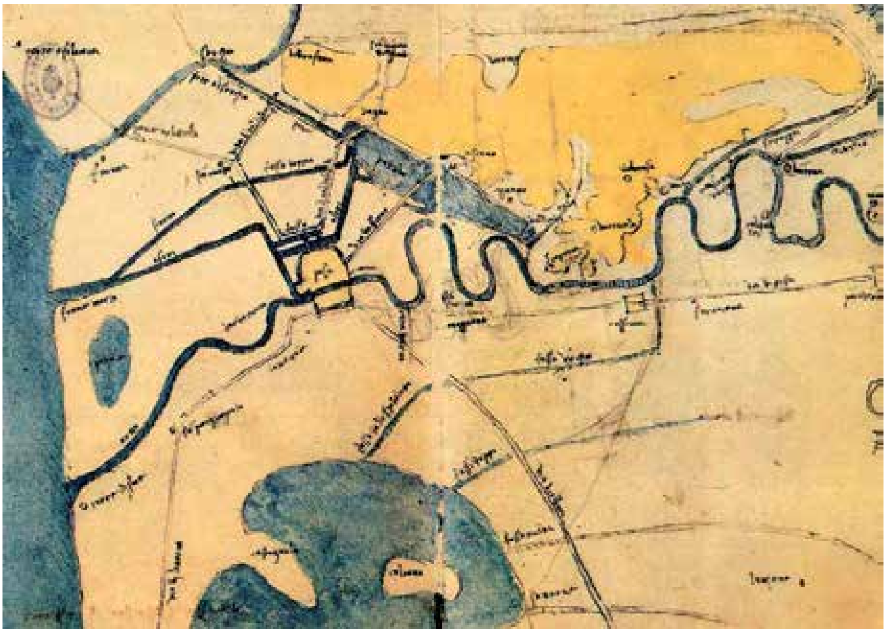


Fig. 1. Map of Leonardo da Vinci (1503), representing the Pisa surroundings and the project of diversions.

In more recent times, the role played in Northern Italy by the protection strategies is documented by the creation of the "*Magistrato civile per lavori generali che riguardano il grande sistema del Po*" (The civil magistrate for the general works of the Po river system) in 1806, that is a corps of technicians and engineers which were in charge of the maintenance and construction of structural interventions.

Some hints to the necessity of dealing with trade-off, considering either the hydraulic system, or the social and local administration context, can be found in Paleocapa (Paleocapa, 1868). He was an engineer, member of the Italian Parliament, who proposed a participatory approach to assess the effectiveness of the proposed countermeasures. In particular, Paleocapa claimed that the diversion of Brenta, Bacchiglione and Sile rivers away from the Venice lagoon should have been the result of a trade-off, balancing the necessity to avoid the sedimentation in the Venice lagoon, the expenses to maintain the artificial embankment system, and the reduction of damages in the Provinces which were crossed by the diverted river courses. Paleocapa claimed that those Provinces could not support the economical expenses for the maintenance of the diversion system. Actually, according to him, it was unfair that the Venice lagoon enjoyed the benefits of interventions, while the Provinces had to suffer for the inundations of the diverted rivers and to pay for the maintenance of the embankments.

The strategies of *flood defence* and of flood control were widely applied even after the Second World War, in Europe, in the 1950s to the 1980s, imposing a strong engineering approach to keep floods inside the river channel or inside lateral embankments. These concepts revealed to be captious, suggesting incorrectly that humans can control nature, influencing processing, according to their finalities.

In the 1990s the concept of *flood risk management* was introduced in Italy (Po Basin District Authority [PBDA], 2001a; Italian Parliament, 1989), implying that floods are natural phenomena which cannot be prevented and pointing out that some human activities can significantly contribute to increase the adverse impacts of flood events (EC, 2007).

In recent years, probably as a consequence of political debate about floods, the scientific literature gets richer in papers dealing with FRM strategies. The classic definition of strategy given in business economics (Chandler 1962) has been considered to be ineffective (Floodsite, 2005), because FRM strategies generally have to be implemented under conditions of increasing uncertainty (see section 4). Uncertainty itself is a relevant topic for modelling and managing (Sayers *et al.*, 2002; Hall *et al.*, 2003). Therefore other definitions have been proposed, by Universities and private or public Associations and Authorities.

One of the early definition describes *strategy* as "*a consistent set of measures, aiming to influence developments in a specific way*"(Hooijer *et al.*, 2004, p.346), a definition that focuses on the content of strategies (i.e., the countermeasures, the general aims and targets, the specific alternatives and so on), without hints to the way the strategy is implemented, to the societal context, to the way the alternatives are balanced and evaluated.

Often the FRM strategy definitions refer to simple daily live expressions, which can be interpreted as an empirical approach, more based on intuition than on systematic investigation. These definitions generally refer to the way of using structural measures, like "do-nothing" strategy, "do-minimum" strategy, "as-low-as-reasonably possible" strategy, "as-more-natural-as-possible" strategy.

By following (Hutter, 2006) and (FloodSite, 2007) and the (PBDA, 2001), strategy is here defined as a "*constant combination of long-term goals, aims and measures, as well as a process that is continuously aligned with the social context*". At present, either scientific literature (Hutter

2006, Floodsite 2007) or practice (PBDA, 2001a), consider that a multidimensional understanding of strategy, is needed in FRM, encompassing (Pettigrew and Wipp, 1991):

1. the dimension of the *content (deciding what to do)* which refers to complex hierarchy of priorities, targets and combination of countermeasures (structural and non-structural) and alternatives, in order to manage risk;
2. the dimension of the *context (what the initial and boundary conditions are)* where the floods occur; this implies the understanding of conditions which are inside the societal texture (human resources, responsibility, culture etc.) or outside (economics, politics, legal framework);
3. the dimension of the *processes (deciding how to do it)*, which describes how strategies are formulated and how they are implemented; process is about learning how to deal with diverse political interests, cultural attitudes, how unexpected conditions or demands are considered.

It should be also considered that strategic choices are not easily reversible in time. Many resources are needed to change a strategy that demonstrated to be ineffective, especially if it has been applied for a long time. The changes in strategies require fresh resources (human and economical), time, changing power structures, changing habits and way of thinking.

On the other hand, a strategy has to be flexible enough to adapt to new situations and conditions.

3. Essentials in flood risk management – A discussion of the state-of-the-art

As mentioned before, the Countries in the European Union are now requested to address to flood risk in a more systematic way than in the past. This implies a different strategy in FRM, as mentioned, from a defense approach to a multidimensional one. In this frame the definitions of concepts is the first step for a shared understanding. Setting common definitions is an important tool to avoid the scientist debate to remain isolated from public and political debates. Without a shared understanding of what we mean by words we use, we are in danger of being misunderstood (Klijne *et al.* , 2008).

3.1 Flood, risk and risk-management concepts

Defining the three terms, i.e. flood, risk and management, is an extensively discussed subject. Many authors and associations propose definitions of their own, (PWDA, 2001; WMO, 2009; FloodSite, 2009; EEA, 2007) but a uniformity in terms use has not been yet reached. Therefore the definitions of the terms used in the following are indicated.

In the EU floods directive, the term "flood" has a precise meaning (EC, 2007):

> *"the temporary covering by water of land not normally covered by water. This shall include floods from rivers, mountain torrents, Mediterranean ephemeral water courses, and floods from the sea in coastal areas, and may exclude floods from sewerage systems".*

In spite of the simplicity of this definition, experience shows that different types of floods can be recognised and categorized in an heuristic way, on the basis of their more recurrent aspects, related to their physical source (rainfalls or storms), the geographic area where they occur (valleys and floodplains, or mountains), their dynamic characteristics and the speed of onset.

For the sake of simplicity, and for reasons of symmetry with the study cases that are shown in Section 4, two different processes are here described, that are generally addressed to as *river floods* and *alluvial flooding due to flash floods*.

As it is well known, river floods origin in plane areas, in valleys, and are characterised by a low velocity of onset. The flood is mainly caused by water, over-topping the lateral protections or gradually inundating the valley, with a concentration of fine-grained sediment, which is mainly transported in water suspension and, to a less extent, on the river bed itself. Lowland floods tend to inundate large areas than floods in upland areas and generally are more lasting. They result from prolonged rainfalls over larger areas, carried by advective clouds and associated with warm or cold weather fronts.

Flash floods can occur in mountainous regions during intense rainstorms, in small catchments up to several square kilometres with steep slopes, impermeable surfaces or saturated soil; they are characterised by high flood water velocity and by a rapid onset, causing floods within a few hours or less. The onset velocity leaves little time for warning and evacuation, especially in the fan areas, which are generally the most urbanised and where there is the highest concentration of receptors. Therefore, the timely prediction of flash floods is the main challenge. Flash floods and, consequently, the alluvial fan inundations, are difficult to forecast, and the risk management is often a very difficult task.

Floods are natural processes, except for the cases where they are man-induced (like in the case of floods caused by dam failures), and do not necessarily imply risks. Intuitively, as far as rivers and torrents are concerned, risk can be defined as the potential loss, due to hazardous phenomena and processes, which generally can be forecast in real time conditions, or within a short time.

By a more technical viewpoint, the following definition is assumed (WMO, 2009) for flood risk:

> *"potential losses associated with a hazard or an extreme event to a given place within a given period of time, which can be defined in terms of the adverse consequences (damage/losses) and the probability of occurrence"* .

Therefore the concept of risk necessarily implies the concept of loss, of the probability of flood occurrence, the intensity of the phenomena, the damages that can be produced by the natural event and the vulnerability of the anthropogenic context.

As far as losses are concerned, it is a useful approach to distinguish between direct and indirect losses (Penning et a., 2000). Direct flood losses are mainly due to the immediate physical interaction of flood with anthropic elements, humans, property and the environment. Indirect flood losses are damages caused by disruption of physical and economic linkages of the economy and the extra costs of emergency, such as the emergency countermeasures that are taken for civil protection aims. At least six different components should be taken into account:

1. the loss of lives, meanly of people living in the floodplains or on in the alluvial fans;
2. the physical damage, that is the loss of functionality of the anthropogenic structures, such as houses, bridges, levees, roads, dams, etc.;
3. the psycho-social impact, that is the psychological effects on people affected directly or indirectly by the flood due to the loss of property or of livelihoods , the displacement from one's home , the disruption of economical, family and social affairs;
4. the functioning disruption, that is the interruption of the interconnections among people, services and webs, so that also people and economical activities that are far from the place where the flood event occurred, suffer for effects of the breaking of interconnections; this is the case, for example, of oil pipelines, water pipelines, railways;
5. the economical impact, that is the hindering of economical growth and development, that is due to the high cost of relief and recover, which may adversely impact investments in infrastructure and the development activities in the area; generally either private and public sectors are discouraged to investments in high recurrent flooding conditions;
6. the economical cost for the emergency countermeasures taken for civil protection aims and other actions taken to prevent flood damage and other losses.

In terms of logical understanding of risk, according to the definition given above, risk can be considered as the combination of three factors, according to the logical formula (UNDRO, 1980; Varnes, 1984; PBDA 2001a, Klijn *et al.*, 2008):

$$\text{Risk} = H \times V \times E \tag{1}$$

where H indicates the hazard, V the vulnerability and E the exposure.

This expression indicates that risk, in natural disasters due to flood, can be expressed as the non-linear combination of three factors, that are hazard, vulnerability and exposure (figure 2).

The expression (1) has not to be strictly considered as a mathematical formula. It expresses the idea that, in risk assessment, three factors superpose in a non linear way. Risk factors can be defined in the following way:

- hazard: *"the probability of occurrence , within a specific period of time in a given area, of a potentially damage natural process"* (UNDRO, 1984), with a specific intensity; the intensity is expressed by referring to a specific scalar or vector quantity or to a graduated scale; in floods generally engineers and practitioners refer to the return time period, that is the is the inverse of the probability that the event will be exceeded in any one year; flood velocities and water depths are generally the physical quantities by means of which the hazard is expressed, while, when sediment transport processes are more relevant (like in the case of alluvial flooding on debris fan), the total deposited sediments or the maximum energy of the water sediment flows are generally addressed to;
- vulnerability: "*the degree of a loss to a given element at risk, or set of such elements resulting from the occurrence of a flood with a given intensity*", (UNDRO, 1984); vulnerability is a function of the hazard level;
- exposure: "*elements at risk, or receptors, that is people, properties and goods that can be lost, injured or damaged during an event*" (UNDRO, 1984); also in this case the numbers and types of the elements at risk vary according to hazard level.

As stated before, the analysis of the exposure should also consider all the indirect factors that contribute to amplify the total losses, i.e. the direct and indirect effects on society, economy and psychology. Obviously, reduction or increase of each factor (Figure 2), implies reduction or increase of risk itself (WMO, 2009).

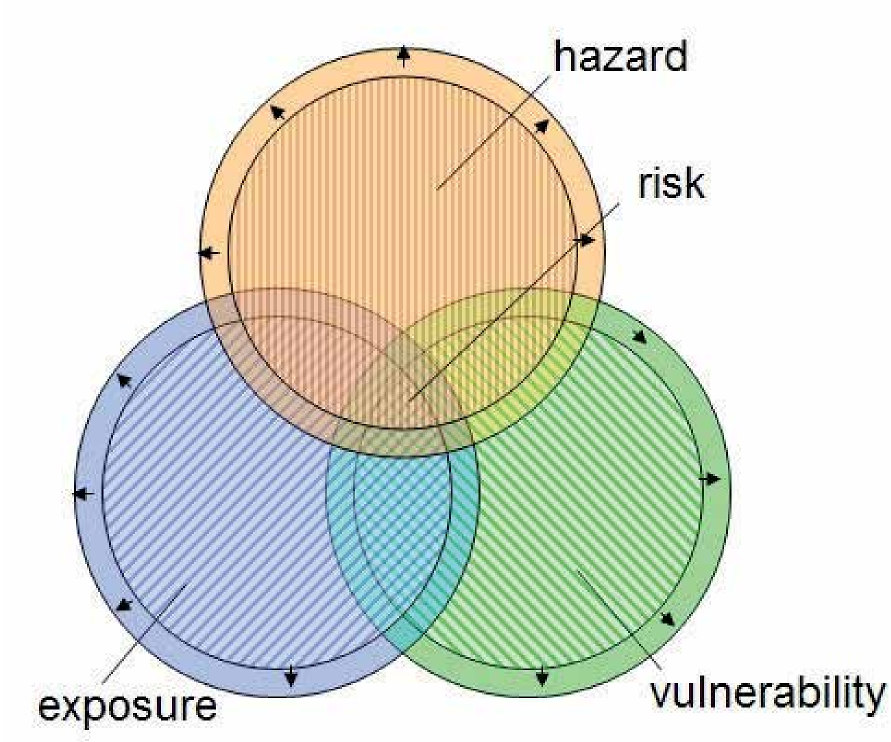


Fig. 2. Elements composing the risk, defined as the superposition of hazard, exposure and vulnerability. The arrows indicate the effect on risk due to an increase in hazard, vulnerability, exposure (WMO, 2009).

Flood risk has dual dimensions, that is objective (physically measured) and subjective (socially-evaluated). The former, which lay in domain of the scientific investigation, shows a high variability in the natural and physical processes which occur in rivers and torrents, in different climatic conditions. The latter is due to the fact that (i) risk affects different people differently; (ii) impacts of risk may cross the territory; (iii) risk perception and concern is different from person to person and from community to community, due to their different thinking, feeling and action; (iv) needs for and capacity of risk reduction are different from person to person.

These topics have to be carefully taken into account by decision makers when they discuss what measures are best to take in the risk management. This latter concept - risk management - can be defined as the "*Coordinated activities to direct and control an organization with regard to risk*" (ISO, 2007). In this chapter, we consider that risk management includes risk assessment, risk evaluation and risk treatment, defined as indicated in table 1. These three components are discussed in the following.

Concept	Definition and reference	Other references
Risk Assessment	"Risk Assessment means a process or method for evaluating risk associated with a specific hazard and defined in terms of probability and frequency of occurrence, magnitude and severity, exposure, and consequences" (FEMA, 1997)	(Environmental protection Agency 1986; DOT, 2005)
Risk Evaluation	"Establishment of a qualitative or quantitative relationship between risks and benefits, involving the complex process of determining the significance of the identified hazards and estimated risks to those organisms or people concerned with or affected by them." (EEA, 2007)	(European Environment Agency, 2007; DRJ & DRII, 2007; WMO 2009)
Risk Treatment:	"Process of selection and implementation of measures to modify risk." (ISO, 2007)	(WMO, 2009)

Table 1. Definitions of the different phases of risk management.

3.2 Risk assessment

In risk assessment, all the components which are intrinsically and technically connected to risk are considered, that is the flood hazards, the vulnerability and the exposure. A principle of risk assessment is that *it is better to be roughly right than exactly wrong*. In particular this is evident in the flood processes which can only be roughly predicted, like in the case of alluvial fan inundations (see also Section 4). Risk assessment includes different basic steps, that can be summarised as follows:

- estimation of the hazard, according to its technical definition, and including the location, frequency and severity of the flood;
- estimation of the exposure, evaluating the number of people, buildings, factories, cultivations etc. exposed to the hazard; these are generally called *elements at risk* or *receptors*;
- estimation of the vulnerability of the elements at risk and receptors, which is usually expressed as percentage losses of people, buildings, cultivations, etc.;
- superposition of the hazard, exposure and vulnerability.

Estimation of the different factors affecting risks is generally a challenge for technicians, practitioners, public administration and researchers alike. A complete discussion of the risk components would imply a deep understanding of the natural hazardous processes, of statistics, economy and of geology, engineering matters, so that just some hints can be given here as far as the estimation of hazard, exposure and vulnerability are concerned.

3.2.1 Estimation of the hazard

According to the definition above, flood hazard should be defined by means of a complex system of probabilistic, or deterministic, modelling approaches. As far as the flood hazards in rivers are concerned, the main steps are the following (PBDA, 2001a; Klaus *et al.*, 1994):

- determination of the design flood at a give location, by means of the regression of available data (direct methods) or by means of modelling the hydrological processes (indirect methods);
- hazard mapping, that is the mapping of the extent of areas that can be flooded by the design discharge;
- hazard ranking, that is the mapping of the areas with different hazard levels; this can be done by addressing to depth, duration of the floods or to the velocity of water.

When the analysis is performed to assess the hazard in areas protected by embankments or levees, it should also be assessed the reliability of the defence structures in relation to different loading conditions, particularly from the pressure or impact of the flood on the defence structures. At its turn, this reliability is strongly dependent on the maintenance state, the structure age, and it is in general a source for uncertainty (Vrijling and Gelderr, 1997; Merkel and Westrich, 2008).

The described methodology for risk assessment, which appears to be simple in principle, suffers from many uncertainties that generally stem from different sources. Hazard assessment uncertainties can be divided into the following main groups.

- *Knowledge uncertainties.* These uncertainties that come from basic lack of knowledge, so that many phenomena can be only roughly understood and simulated; this is the case of the triggering mechanics of flash floods, or the morphological changes induced by floods during an event.
- *Input and parametric uncertainties* in modelling. When mathematical models are applied, it should take into account that the computed results are obtained for particular simulation conditions, i.e. for given inputs and set of parameters. A different choice of inputs or a different set of parameters (Kuczera and Mroczkowski, 1998; Beven and Freer, 2001) influences the results. For example, as far as design discharge is considered, it should be considered that flood probability estimation is generally based on a statistical regression of available data, which often show to be statistically inconsistent (see Section 4 for the Dora Baltea study case); moreover, the statistical methods generally assume the s*tationary* (figure 3) of the flood formation processes (see below).
- *Calibration uncertainty in modelling.* It is generally due to the criteria which is adopted for model calibration, especially in flood modelling; generally the "optimization" criteria are arbitrarily assumed (Beven and Freer, 2001).
- *Structural uncertainty.* It is an inherent feature of the applied model: it is a consequence of the simplifying assumptions made in approximating the actual environmental system with a mathematical hypothesis.

As hinted above, flood hazard assessment is generally made by assuming *stationarity,* that is by assuming that the river basin remains the same in time, while, in reality, the physical characteristics of the basin or of the river valley change in time (Figure 3). Variations in time of soil use (man-induced or natural, like vegetation) of topography and of rainfall regime (even without considering relevant climate changes; Arcilla, 2007) can alter the way in which floods originate (Ducrocq, 2008) or are conveyed through the valley. The key message that should be taken into mind by decision makers is that the past cannot be the sole guide to the future.

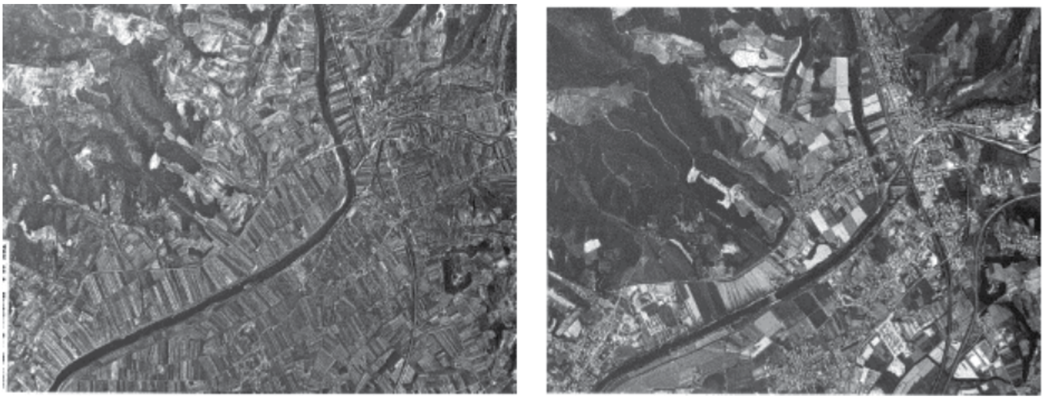


Fig. 3. Aerial pictures of Arno river (Italy) near Montelupo Fiorentino in the period 1954 (left)-1993 (right). Urbanisation that took place especially during the fifties and the sixties, substantially increased the number of receptors along rivers and altered the processes of formation and propagation of floods along watercourses (Autorità di Bacino del fiume Arno, 1997).

When the modelling uncertainties do not allow us to obtain results within reasonable tolerances (Figure 4), generally more heuristic approaches are followed for example with the help of empirical evidences. In particular, risk assessment in alluvial fan flooding is affected by stronger uncertainties (see Section 4). In these cases a detailed back-analysis is a basic step to estimate (even roughly) the order of magnitude of the hazards, by means of the reports recorded in newspapers, geomorphological or statistical approaches (Franzi and Bianco, 2000).

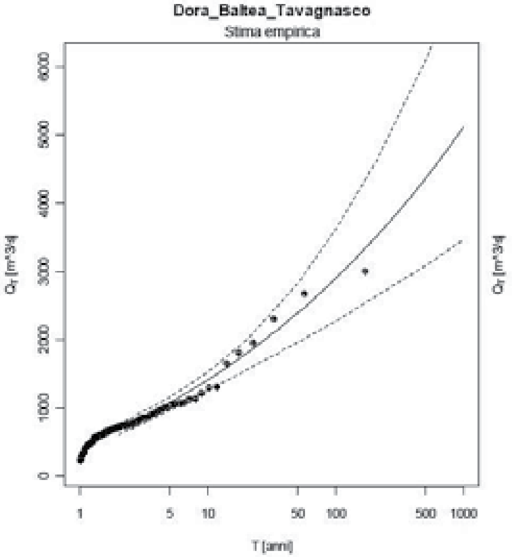


Fig. 4. Confidence intervals of quantiles. The data refer to Dora Baltea river (Italy). The variability in the assessment of discharge (dotted lines) which corresponds to a given return period can be easily read on the figure. (Claps *et al.*, 2008).

3.2.2 Estimation of the vulnerability

Generally speaking, vulnerability is the proneness of structures, goods, humans, communities to be impacted by flooding (UNDRO, 1984). Therefore vulnerability represents the inadequacy of structures, or the incapacity or inability of a community to resist and/or recover from the impacts. This inadequacy or inability can play an important role in the transformation of a hazard into a disaster.

According to the element at risk (inhabitants, structures, communities), a different vulnerabilities should be considered, as every element can react to and recover from floods in a different way. The total vulnerability is the combination of the complex interconnections among communities, individuals and structures and their capabilities to cope with disasters. The factors influencing the vulnerability are mainly the following:

- physical/material conditions: related to the kind and quality of material of constructions and to the physical state of people that can be involved in the floods;
- organizational: related to the way in which communities are managed, structured, interconnected;
- attitudinal: related to culture, awareness.

In table 2, these factors are briefly listed an analysed (WMO, 2006).

Vulnerability factors	*Vulnerable elements*	*Source for vulnerability*
Physical/material	Infrastructures, houses, farms, humans,	Weak infrastructures / Wrong design Materials inadequate for the design floods Lack of maintenance works Physical degradation of materials Unpreparedness / Unawareness Malnutrition, disease, handicaps, age
Organizational	Communities	Lack of leadership, or organizational structures Lack of representation Spot aids without a central management Scarce preparedness of civil protection Lack of information to people
Attitudinal	Humans, communities	Lack of the awareness of rights and obligations Lack of autonomy Heavy dependence on external support

Table 2. Vulnerability sources, for different kinds of receptors.

Considering that vulnerability factor can be quantified by values between 0 (invulnerable receptors) and 1 (totally vulnerable receptors), some technical detailed approaches have been proposed in literature for vulnerability quantification (Green *et al.*, 1994). Due to the complexity of the different factors and scales on which vulnerability depends, the attempt to include also structural, durability, resilience, robustness factors to model all the vulnerability components, generally results in a higher complication of methods. In the flood management plan that is now in force in Italy, the vulnerability has been simply ranked into classes, as shown in table 3 (PBDA, 2001a). This approach is an oversimplification of the reality, but it proves to useful in many cases.

Vulnerability classes	Items can be damaged
V1	Elements at risk can undergo minor level functional damages
V2	Elements at risk can undergo medium level functional damages
V3	Elements at risk can be destroyed or severely damaged
V4	Elements at risk are certainly destroyed or severely damaged
V5	Elements at risk are certainly destroyed or severely damaged and there is the possibility that human life can be lost

Table 3. Vulnerability classification according to Italian legislation (*PMC, 1990*).

3.2.3 Estimation of the exposure

The concept of exposure refers to all elements at risk (receptors), that is people and properties that can be lost, injured or threatened by a hazard. Assessment of exposure has to consider its dependence on the hazard levels, the variability of exposure in time (especially in urbanized areas, where the soil use destination can change rapidly in time) and the uncertainty in its determination.

The most useful procedure consists in establishing the total number and kind of receptors that may be hit by a flood, that is the degree of exposure.

Identification of receptors, for different hazard levels, is therefore a central point, and one common way of achieving this is to produce an overlay of the receptors, eventually with the help of Geographic Information Systems (GIS).

As far as *people* are concerned, data must be collected on the number of people who reside, work or travel in the area liable to flooding, together with their demographic characteristics, as these affect their personal vulnerability. These data can often be obtained from national censuses, municipalities, local administrations databases. Field surveys and data on industrial and commercial enterprises can help to have a better estimation of people working in the exposed areas. Assessing the number of people who travel through the area liable to flooding may need special surveys or data from state transport departments.

As far as *property* is concerned, data are needed on the number and location of different types of property (houses; factories; etc), as well as on their value and their susceptibility to flooding. Moreover all kinds of properties should be collected and surveyed, that is economical properties and activities, State properties, like bridges roads, but also aqueducts, drinkable water wells, pumping stations, and cultural properties, like national libraries, museums. Cultural receptors should not be underestimated, as the social capacity to recover after a flood also depends on personal losses and on the loss a cultural heritage (Figure 5).

Fig. 5. Florence flooding in 1966, November 3rd-4th(left) and the so-called "mud angels" (right), that are students and volunteers that helped to move the artistic works of art and to recover books during and after flooding.

Similarly, as environment can be considered a property, the exposure of habitats and species should also be considered, especially when dispersed pollutants can adversely affect floodplain ecology and ecosystems.

There are two principal ways to obtain this required land use information: by carrying out field surveys (primary data) or, more usually, by relying on existing land use data (secondary data).

The advantage of *primary data* is that all required land use and property information can be collected at the level of detail that is needed. On the contrary, field surveys are time-consuming and costly, whereas secondary data, such as national censuses of land use, are often readily available. The main disadvantage of secondary data is that they are not produced for the purpose of flood risk assessment (often they are collected for local property taxation purposes) and, therefore, they probably not contain all necessary information at the required level of detail. Moreover secondary data can reveal to not be updated.

Obviously, the total number of receptors which can be hit by flooding varies according to the hazard level, including the probability of failure of the structural countermeasures which should avoid flooding.

3.3 Risk evaluation

Risks assessment is generally the product of the collaboration among scientists, practitioners, engineering, economics and people, and its contents are more technical than decisional. After risk assessment, risk management should focus on establish what are the *acceptable and tolerable* risks levels. By considering that total protection against risk or the total elimination of hazard is impracticable from a economical point of view and scientifically unfeasible, decision makers should establish what is or what are the risk levels at which risks are still acceptable or tolerable. This implies subjective and objective assessment.

In the flood management plan in Italy (PBDA, 2001a) this concept has been efficiently expressed throughout the notion of "compatible risk" which is here reported, as follows:

> *The compatible risk specifies and plans out what are the hydraulic and geological risk conditions that should be residual at a basin scale. Its assessment depends on the social and economical expectations for protection(…).The gap between present risk and compatible risk implies the necessity of the implementation of countermeasures.*

Tolerability and acceptability are the two main concepts which have to guide the evaluation of the flood risk. The key idea is that some risks can result to be unacceptable, but can be still tolerable by society, by a local community or by individuals. Predictably, each alternative solution will present some internal conflicts between locally acceptable levels of risk and socio-economic (UNDRO, 1979).

The approach to tolerable/acceptable risks is known as ALARP principle (Floodsite, 2009), and involves the definition of (Figure 6):

- an upper bound risk level above which the risk is no more acceptable (level of maximum tolerability);
- a lower bound risk level (individual or societal) below which risk is not a concern (level of acceptability);
- an intermediate region, i.e. a tolerability region, where societal and individual risk reduction is required to achieve a level "as low as reasonable practice".

From a social science point of view, the upper and lower limits of tolerability region may differ significantly between persons, among individuals and society. A public consensus on risk acceptance may not exist.

Moreover the determination of the tolerability and acceptability of risk implies the determination of:

- the individual acceptance – that is an analysis of the socio-economical concerns and of the risks perceptions, which are mainly linked to the attitude of people in endangered areas, to their past or recent experience, to the past or recent psychological stress caused by flood events; (UNDP- DHA 1994);
- of the expert acceptance - that is an analysis of the effective situation, evidencing the effective level of risk and the possibility to recover from a flood; the expert acceptance has to be the result of the application of economic principles, laws and safety norms.

All people, whether aware or unaware of the actual flood risk they run individually, have their own rationality, and this is not necessarily the same as that of the scientist or the flood risk manager. In general, a correct approach should not rank individual acceptances as "correct" or "not correct", as individuals perceptions are different.

However it should be clear that either in "expert" or in "individual" perception of risk, humans tend to believe and think about risks with less precision than they really have.

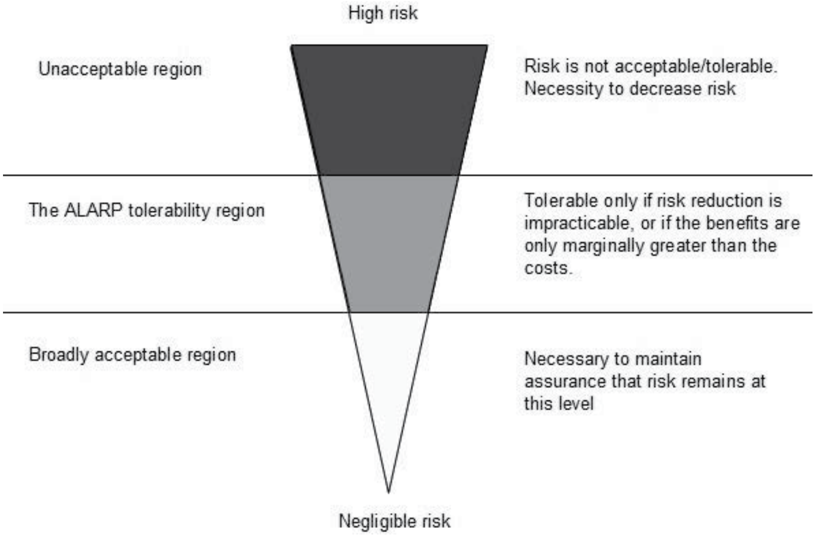

High risk
Unacceptable region
Risk is not acceptable/tolerable. Necessity to decrease risk
The ALARP tolerability region
Tolerable only if risk reduction is impracticable, or if the benefits are only marginally greater than the costs.
Broadly acceptable region
Necessary to maintain assurance that risk remains at this level
Negligible risk

Fig. 6. Sketch of the ALARP principle (Floodsite, 2009).

In order to allow a comprehensive assessment of risks, risk evaluation should elaborate *risks scenarios*, that is individuate and represent the more significant situations for every level of risks. In this frame, the capacity to simulate different situations, and the availability of flexible analysis instruments can strongly influence the capacity to let a better understanding of the consequences of flooding.

For example in the Po river basin, it is well known that floods peaks reduction can be strongly influenced by the extension of flooding in the upper parts of the watershed basin, that is in the North-western areas of Italy. In general, the larger is the flooding extension in the upper part of the Po basin, the stronger is the reduction of flood peaks of downstream discharges. The physical process which causes a decrease in hazard level, is known in literature as peak discharge *lamination* and it is a positive effect of flooding. In order to understand what should be the optimum in the flooding extension, different scenarios have to be theoretically developed, between to following extreme conditions, that, for the sake of simplicity, can be described as follows:

- it is maintained and enhance a high protection levels for existing receptors in the upper part of watershed, by means of levees, embankments and floodwalls, limiting lamination of the flood; as a consequence, receptors in the lower part of the watershed could experience a risk increase, due to higher peak discharges and flood levels;
- it is maintained and enhanced a high protection levels for existing receptors in the lower part of the watershed, by means of enhancing large inundations in the upper part, in order to increase flood lamination; receptors in the upper part should experience a risk increase do to the lack of defences and more frequent flooding.

What should be the "best" solution between the two, that is the optimal extension of inundations, is a subject of risk evaluation through different scenarios.

In risk evaluation a social and economical *cost effectiveness* analysis should be performed, in order to separate what is feasible from an economical and technical point of view from what it is not. This allows the establishment of priorities and, implicitly, the risk level that is residual after the implementation of countermeasures.

Some of the most used methods in cost-effectiveness analysis are the *cost-benefit analysis* (which is indicated in literature with CBA) or the *multi criteria analysis* (MCA). Some indications about these methods can be found in literature (WMO, 2007) and are not considered here.

3.4 Risk treatment

The strategies in flood risk treatment can range from risk reduction (prevention and protection), risk transfer and risk retention, that can be described as follows (Table 4).

- Risk reduction (including prevention and protection). It is a common way to cope with risk, and includes all the measures that reduce the three factors which risk is made of, that are the hazard, the exposure and the vulnerability, mainly by means of flood protection and flood prevention countermeasures. Risk protection generally refers to the hazard reduction, and to the traditional methods in civil engineering interventions. Risk prevention generally includes all the actions that aim at the vulnerability and exposure reduction, without influencing the flood dynamics. Risk reduction countermeasures can be classified into structural and non structural, as it is discussed below (PBDA, 2001a).
- Risk transfer (or sharing). In this case risk in transferred to assurance companies, by means of a insurance policy; in risk transfer people can temporarily suffer for the economical loss, but they can recover in the medium period by means of the monetary refunding. Obviously some losses cannot be totally transferred, so that risk transfer in these cases is just a monetary compensation. The basic principle is to spread the risks over time, and among individuals, organisations or government, which pay insurance premium against a specific flood risk level (WMO, 2009).
- Risk retention. In this case people live together with risks, with or without preparedness, being or not aware of the presence of risk. Since risk cannot be completely eliminated and the total protection is unfeasible, residual risk forces people to live in conditions where the probability to have losses is not equal to zero. Residual risk can be defined as "*the risk that remains after risk management and mitigation*". As residual risk is retained, people should be preliminary informed about the risks they can experience and should be informed how to eventually cope (individually or collectively) with risks, especially in collaboration to civil protection agencies.

Risk reduction countermeasures can be classified into structural and non structural (PBDA, 2001a).

Structural countermeasures have been extensively used in the past, and they are the most traditional tools to cope with floods (WMO, 2004, 2005; PBDA, 2001a). The approach is that of tradition civil engineering, that is based on construction of permanent concrete/steel/stony structures to protect from floods. Generally structural measures focus on the reduction of flood hazard, by reducing the flood magnitude, the flood extension or decreasing the vulnerability of receptors. A classification of structural countermeasures can be made by referring to their extension, that is intensive, if located along or across rivers, or extensive, if diffusely spread all over the basin (Table 4).

Risk treatment strategy	*Instruments*	*Example of countermeasures*
Risk reduction (prevention and protection)	Structural and non structural measures.	See table 5.
Risk transfer	Insurance policy	Policy for mandatory/voluntary insurances against natural disasters
Risk retention	Living with risk Emergency plans Civil protection countermeasures	*Preliminary to flood:* Information about the kind of risks;; implementation of water proof defence implementation of civil protection plans *When flooding is imminent* Detection of flood formation Forecasting of flood discharges by means of modeling Early warning/warnings dissemination Warning confirmation Response (Closure of roads and bridges, operation of barriers, provision of temporary flood protection measures), evacuation, rescue *Post-flood actions*

Table 4. Strategies in risk treatment.

At present, the necessity to take structural countermeasures in flood management, is often a consequence of urbanization and of the flood management strategies followed in the past. Historically the enhancement of protection against floods favored the occupation of floodplains, increasing the total number of receptors potentially at risk. At the present state, the necessity to maintain the economical activities on floodplains as well as the need to protect urban areas, force local authorities to implement, and even expand and extend the present systems of structural protection countermeasures.

Structural countermeasures have to be technically feasible and economically reasonable, cost-effective, and sustainable.

Non structural countermeasures do not aim at affect directly the physics of flood process, but they influence the vulnerability and the exposure to flood. The implementation of non structural countermeasures should be the consequence of regulation, of the application of laws and directives.

Classification of countermeasures	Functionality of the countermeasures	Type of countermeasures	Protect-ion	Prev-ention
Intensive structural countermeasures	Transversal protections, against bed erosion	Channel stabilisation works, weirs, dikes	X	
	Longitudinal protections containing flooding	Levees, flood walls, embankments	X	
	Stabilization of riverbanks, against lateral erosion	Rock, concrete, composite revetments, gabions or geotextiles revetments, ripraps, groins.	X	
	Maintenance of bed river profile as and conveyance	Sediment excavation, artificial aggradations, river training (straightening, widening deepening, hard-lining), removal of structural operas with negative impacts / are incompatible /show to be anomalous with the flood management plan		X
	Diverge or reduce flood discharge	Sluices and flood control channels, detention ponds, dams,	X	
	Reducing the flood peaks	Reservoirs, retention polders, creation of temporary storage areas.		X
Extensive structural countermeasures	Interventions aiming at influencing the flood formation mechanics	Renaturation		*X*
		Maintaining or increasing the total areas of the natural flooding areas		*X*
		Reforestation of hill slopes, soil use to reduce the total runoffs or increasing the duration of the rainfalls runoffs processes, increasing the infiltration and retention capacity of the soils, river rehabilitation;		*X*
Non structural countermeasures	Real time flood prevision and communication	Evacuation of the total number of people at risk; roads and bridges closure,	X	X
	Regulation of soil use	Regulation, laws and acts, Flood Hazard Zoning, building regulations on constructions, technical layout of installations; regulations on timely evacuation.		X
	Flood surveillance,	Real time control of the functionality of the defence system, including levees		X
	Ordinary maintenance			X

Table 5. Structural and non structural countermeasures.

4. Planning in flood risk management

The necessity to implement FRM plans, stems from the definition which has previously given in section 2, abut the three dimensions of strategy. As it has been widely debated in Section 3, risk assessment, evaluation and treatment phases can allow an appropriate understanding:

- of the actions needed to manage the risk (the *content* dimension of the strategy),
- of the human resources, of the economical, societal conditions of the endangered areas (the *context* dimension of the strategy).

Plans are defined in the following way (DHS, 2008): "a plan is a continuous, evolving instrument of anticipated actions that maximize opportunities and guide response operations. Since planning is an ongoing process, a plan is an interim product based on information and understanding at the moment, and is subject to revision". In this frame, the plan can be roughly considered as a tool to describe how the contents of the strategy are formulated and implemented in a given watershed basin. According to the European directive, the aim in planning is the "reduction of potential adverse consequences of flooding for human health, the environment, cultural heritage and economic activity, and, if considered appropriate, with non-structural initiatives and/or on the reduction of the likelihood of flooding" (EC, 2007).

At framework level, we think planning can be fashioned in different ways (see also Floodsite, 2007; PBDA, 2001):

1. the classic mode of *programming*, which is appropriate under conditions which are highly predictable; this is the case when a plan indicates the amount of economical investments and their chronology, to implement countermeasures, in a given area, in a given time interval;
2. the *scenario based planning*, which considers different possible futures; it is appropriate when the flood risk conditions can be predicted reasonably well, in a given watershed;
3. the *preparedness strategies*, increasing organizational activities for coping with the unexpected conditions; this strategy is predominant when the uncertainties do not allow a reasonable and appropriate understanding of the risk.

By referring to the content of plans, they should address all phases of the flood risk management cycle (Figure 7) but focusing particularly on (see also 2007/60 EU directive):

- programming the preventing countermeasures, including maintenance and watch (long term);
- flood event management, containing all the activities that should be carried out to reduce the impact of the floods, when flooding is imminent or already taking place;
- civil protection measures, that is the activities that should be carried out after flooding; the extent of these activities strongly depend on the resilience of the anthropic system, that is the *"ability to resist, absorb, recover from or successfully adapt to adversity or a change in conditions"* (UNDRO, 1980).

Designing the flood risk management plan, the political authorities should improve and facilitate people participation, at several degrees, from empowerment, to ownership, to consultation, to simple information, improving the awareness of people in endangered areas (WMO, 2006).

The results of flood risk management should be monitored by authorities, in order to better calibrate the planned interventions. Collected data should improve the informative systems and databases, increasing the total number of recordings of past events (Bovo et a. 2007). Monitoring should also include the after flood debriefing, in order to indicate to managers the lesson learned, and suggest what enhancements should be implemented. In this frame flood protection plans result to be an ongoing process, as they change in time, enhancing the territorial understanding of the flood risks and allowing a constant adaptation to changing circumstances and changing societal requirements. A good example of Italian experience in the mitigation-preparadness-recovery cyclic process can be found in Bovo et al. 2007, as far as Regione Piemonte (Italy) is concerned.

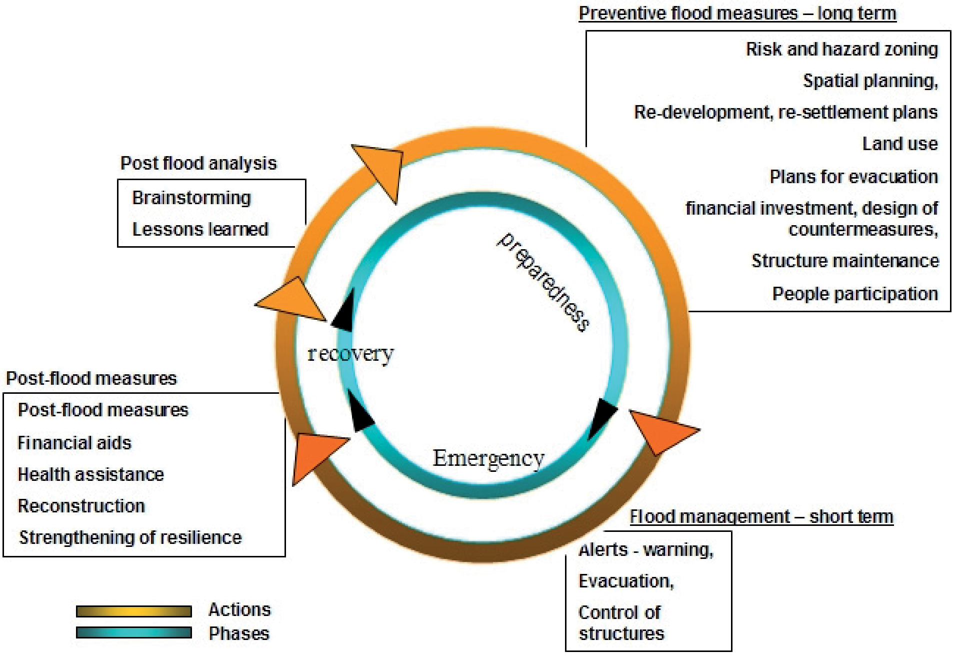


Fig. 7. Cyclic process in flood risk management.

From what has been above discussed, a discussion of the effectiveness of planning is necessary to understand what are the best ways to ameliorate the approaches described herein.

By referring to the FRM plan developed by the Po Basin District Authority (indicated with PAI), in the next section, the effectiveness of the FRM strategies is discussed, by referring to either to floods in main rivers or to flash floods in torrents. In particular the role played by the uncertainties in the risk assessment are discussed and put into evidence.

It will be put into evidence that the scenario-based planning is preferable when the uncertainties in risk assessment are less. The preparedness strategy has to be preferable for highly uncertain conditions, like in the case of alluvial fan flooding.

In the next section, the practices followed by the PBDA are described and discussed, as far as the main rivers in the Po basin are concerned. In particular, the effectiveness of the plan during a recent flood occurred in Northern Italy, in the Dora Baltea watershed is discussed.

4.1 Effectiveness of flood risk management in rivers (Po basin district)

As anticipated in Section 2, different FRM strategies have been applied, in the course of the centuries, in the Northern part of Italy. The changes in strategies were either due to the political changes in the Governments, or to the evidence of the ineffectiveness of the practices that were followed till that time. It can be fairly stated that since 1989, the Italian legislation imposed a more advanced approach in FRM, well before the approval of the 2007/60 directive. Nevertheless the experience shows that risk management is an ongoing process, so that FRM plans have to be continuously adapted to new circumstances or physical evidences.

Therefore, in the following, the practices followed after the approval of the 183/89 act are described and their effectiveness is discussed, by referring to the 2000 flood in Dora Baltea river.

4.1.1 The practices in risk assessment, evaluation and treatment

Large inundations in Po watershed basin, either along the Po River or along its tributaries, are generally connected to extended rainfalls, generally linked to large meteorological perturbations. The soil moisture conditions play one of the major roles in the rainfalls-runoffs transformation, together with the rainfalls intensities, their geographical distributions, and, in some cases, the snowmelt. For instance the 2000 (October 13rd - 16th) flood would have been probably more severe if the drop in temperature did not transform the rainfalls precipitations into snow (Ratto *et al.*, 2003), when the thermometric zero dropped from about 2800 m a.s.l. to about 1500 m a.s.l., in alpine regions.

From a methodological point of view, the amount of available data on floods occurred in the past, either quantitative or qualitative, shows a strong heterogeneity over the basin, so that the estimation of the hazard requires the application of hydrological and hydraulic models, probabilistic or deterministic. Historical data on floods are available either in a structured (like in hydrological annals, which were published by the hydrographic Italian service, up o the nineties, or in databases and informative systems organized by regional authorities (Bovo et a. 2007, Regione Piemonte, 2012)) or in a non structured (newspapers, witnesses, reports) way. Probabilistic estimation of floods is a very important step because, for the aims of PAI, flood risk assessment in Po basin refers to a protection level that is defined by the project flood discharge, estimated by means of the methods proposed in literature, which can be classified as direct or indirect methods.

Direct methods have been applied when the total number of locally recorded flood discharges allows a statistical inference of data. When data are scarce, or when the total amount of data did not allow a reliable statistical inference, indirect methods have been applied to estimate the flood discharges or the flood water levels, for the considered cross sections. Generally the complexity of the indirect methods spreads from the most empirical to the most detailed.

Determination of the project discharge for main rivers is in charge of the Po basin districts authority. The procedure, which was adopted in PAI and published in 1998 (PBDA, 1998), allowed the estimation of the design discharge for the most significant cross sections of the whole basin, by means of:

- the regression analysis of the recordings in the annals;
- the estimation of the peak flood discharges which were not measured but that caused inundations; as far as Regione Piemonte (Italy) is concerned, an efficient and updated informative system for risk prevention, (named *Banca Dati Geologica)* includes hundreds descriptions of the effects of past inundations (Regione Piemonte, 2012), covering a period that extends from the 17th century to now.
- the collections of the estimates of the peak discharges in papers, reports etc..

In this activity, a collaboration with the Regional Authorities, Research National Council and Universities demonstrated to be important to achieve a better interpretation of the available data and a better understanding of the possible magnitude of floods.

In the design discharge determination, the central concept is the flood return period, also known as recurrent interval.

The theoretical return period is the inverse of the probability that the event will be exceeded in one year. For example, a 10-year flood has a 1 / 10 = 0.1 or 10% chance of being exceeded in any one year and a 50-year flood has a 0.02 or 2% chance of being exceeded in any one year. Correspondently, if a discharge Q has the return time T, it means that it can be exceeded, on average, once in T years.

For the main rivers, the hazard assessment in the Po watershed basin has been made by referring to different return periods T, that is T=200 years and T=500 years, and therefore two different design discharges, which are respectively Q_{200} and Q_{500}, have been estimated for a set of relevant river cross sections.

The estimations of the design discharges have been published by the District authority and are available for the public, together with the estimations of the water levels for different return periods (PBDA, 1998).

By applying hydraulic and geomorphologic models, the PBDA proposed a classification of hazards that is mainly based on the concept of the river corridors (table 6, fig.8), which are indicated as corridors A, B and C. The corridors are defined by considering two complementary components, that are the hydraulic component and the geomorphologic component. As far as the hydraulic component is concerned, the corridors are defined as follows:

- B corridor: it corresponds to the areas which can be flooded by a design discharge of 200 years, Q_{200}, with velocities less than 0.4 m/s;
- A corridor: it corresponds to the areas where al least the 80% of the total design discharge Q_{200} flows;
- C corridor: it corresponds to the areas which can be flooded by a design discharge of 500 years, Q_{500}

These corridors have been mapped on 1:25.000 scale, all over the Po watershed basin. (PBDA, 2001).

When the actual defence system is not sufficient to avoid the flooding of the Q_{200} discharge in areas that should be protected, the designed limit of inundation is represented on maps by means of a different graphic item , that is called "B design limit". This limit indicates that the present protection system shows a deficit and that appropriate countermeasures to avoid the flooding of Q_{200} discharge should be implemented, such as new levees or embankments. As long as the risk situation remains unchanged, stricter soil use regulation are in force in the areas behind the B design limit. The approach of fluvial corridor has been followed also elsewhere (OPW, 2008).

	Hydraulic component	Geomorphologic component
B corridor	Areas flooded by the design discharge having T=200 years, with velocities less than 0.4 m/s	Reactivation of recently abandoned watercourses
A corridor	Areas where al least the 80% of the total design discharge Q_{200} flows	Reactivation of recently abandoned watercourses
C corridor	Areas flooded by the design discharge having T=500 years	Reactivation of abandoned watercourses
B design limit	The actual defence system does not avoid the flooding of the 200 years design discharge beyond the limit. Protection works are necessary to protect the 200y-flood prone areas.	(see left)

Table 6. Technical definition of the fluvial corridors approach.

Risk vulnerability and exposure components are estimated by means of qualitative scales (Table 3). Risk classes are ranked by intersecting the different exposure/vulnerability classes with hazards, obtaining four risk classes. At the present implementation state of PAI, risks are not mapped on topographic layers but refer to administration units. In other words, for each municipality, a risk class has been assigned, without mapping the risks on floodplains. As to the receptors at risk, they are indirectly pointed out in the implementation norms (PBDA, 2001b), which refer to the most relevant categories of receptors which are inside each flooding corridor, discriminating the actions needed, the regulation of activities, prescriptions and prohibitions (Table 7).

Existing structures		New structures	
Allowed	Not allowed	Allowed	Not allowed
Ordinary manutention work, with implementation of vulnerability mitigation devices. Demolition without reconstruction.	Enlarge cubage. Enlarge total occupied area.	Public works, bridges.	Drinkable treatment implants. Garbage treatment. Campings. New settlements.

Table 7. Example of regulations in A corridor. The table is a simplification of the art.29 of the implementation norms (PBDA, 2001b).

One of the limitation of the followed approach is that the hazard is defined by referring to the inundation of the design discharges, without considering that locally the real hazard conditions can be very different. For example, some urbanized areas are in the same corridor but can experience different water levels during the same flood, or can be affected by floods with a different frequency.

As far as this procedure is concerned, it is remarkable that the reference to return periods does not consider river geographic location, so that the approach it is uniform all over the river basin. Some local exceptions can be found for rivers where the Q_{100} design discharge has been adopted, instead of the Q_{200}.

As indicated before, the implementation norms represent an important legal component to detect receptors at risk and to treat risk in PAI, and are complementary to the following other actions in PAI for risk treatment, which are:

- the implementation of a program of interventions, founded by the national government; standard designs of the typical structural countermeasures are provided (PBDA, 2001c);
- involvement of local administrations, which have to adapt local planning to hazards zonation and implementation norms; the general strategy is to avoid to increase the total risks on flood prone areas, reducing the total exposure or the vulnerability;
- risk transfer, which does not refer to insurance obligations, but it is related to a physical transfer of flood risks; as mentioned before, in Po basin, the floods effects in the valley part (namely the areas next to the Po delta) strongly depend on the lamination processes that take place in upper part of the watercourse, either spontaneous (flooding in alluvional areas) or man-induced (retention ponds); therefore large flooding areas for peak flood lamination are designed and maintained, in order to make the floods less severe for downstream areas; in this frame, flood risk is retained in the upper part of the basin.

The PAI allows an efficient participatory approach for hazard mapping and risk management. Participation of local administrations, stakeholders and people was already allowed in the implementation phase of PAI, so that private and public remarks could be submitted to and discussed with PBDA to modify, change or adapt the PAI to local situations. As an ongoing process, after the adoption in 2001, some local revisions to PAI have been proposed and adopted by the PBDA, in order to enlarge the risk management approach of "river corridors" to more rivers. When recent floods showed that the risk management had to be revised, the PAI has been updated accordingly. This is the case of 2000 flood, which occurred in the North Western part of Italy and showed to be an exceptional event, from a statistical point of view. At present, the approaches followed in PAI are being integrated and completed according to the requirements of the 2007/60 European directive.

4.1.2 Discussion on the effectiveness of FRM – The 2000 flood in Dora Baltea river

In decision making about flood risks, it should be considered that risk assessment is generally based on estimations, which are affected by several uncertainties and can show to be inaccurate if compared to real flood effects. Comparison among depicted scenarios and reality is a useful tool to learn useful lessons about scientific and management limitations (Table 8), as it can be deduced by the study case of the 2000 flood in Italy.

On October 13–16th, 2000 heavy rainfalls interested the North-western Italian Alps from the upstream reach of the Po river to the Ticino river, causing huge flooding and landslides with significant damages to houses and infrastructures and several life losses (Ratto *et al.*, 2003).

The whole Valle d'Aosta region (excluding the North-Western sector where less damages have occurred) was interested by extensive flooding, landslides, soil slips and debris/earth flows on the alluvial fans with damages to houses and infrastructures for more than 500 million euro and several life losses in the population (14 persons died). From a hydrological point of view, the discharges data (recorded or reconstructed) showed to be exceptional, as they were higher than those previously recorded in annals since 1925 by the hydrographic Italian service. In Piemonte the areas near Ivrea were seriously affected by the flood, an abandoned course of the Dora Baltea (named Rio Ribes) was reactivated and discharges by far higher than the estimated ones were convoyed downstream. Several bridges broke down and the road connections were interrupted for many hours.

On October 18–19th the areas involved by the event were surveyed through aerial photography. The survey was focused also on the Dora Baltea river (watershed basin area: 3920 km2)valley line, on the adjoining mountainsides, and on the valley lines of the main tributaries. Dora Baltea is a tributary of the Po river, and therefore is under the administrative competence of the PBDA.

The surveys were carried out through the following steps:

- through aerial photography interpretation techniques a thematic cartography in scale of 1 : 5000 and 1 : 10 000 has been produced, highlighting the main typologies of hydrogeological phenomena identified;
- the produced thematic cartography has been digitized and georeferenced in order to allow an integrated use with information of different type and provenance;
- the thematic cartography has been updated and corrected as a result of numerous direct surveys.

Since a risk assessment plan was already in force since 1998 all over the Po basin (named PSFF - plan of fluvial corridors, adopted by PBDA; PBDA, 1998), the effects of the 2000 flood allowed the district authority to check the effectiveness of the risk assessment procedures which had been adopted.

In particular (Table 8), as far as flooding in Dora Baltea is concerned, the comparison of risk assessment and reality showed the situations where the system of fluvial corridors demonstrated to be effective, and the areas where it had to be revised. Nevertheless the philosophy of the fluvial corridors systems was maintained and integrated.

A hydrological and a hydraulic model were applied to revise the risk analyses previously made (Table 8). The major discrepancies in risk assessment were due to the underestimation of the design discharges in the old PSFF plan. Thanks to a more detailed topographic support available for technicians a more precise analysis of the inundation areas was carried out.

Indications on the acceptable risk levels for different kinds of receptors in corridors have been given, in order to guide the municipalities to a major awareness in planning. This was a step towards a more detailed management of risks in floodplains, and a substantial improvement of the approach that was followed in PAI. In particular, for the receptors on floodplains, it has been performed an analysis of the actual protection level, which has been classified into *deficit* (when the actual protection level is less than that acceptable) and *surplus*, and ranked (Figure 9).

Moreover it was planned to apply and calibrate, on the whole Dora Baltea river basin, numerical models for the rainfall-runoff and flood propagation simulation, in order to forecast critical events along the hydrographical network starting from the forecast and observed data. This improves the data management during the events through the consecutive issue of bulletins to the civil protection offices, updating the measured data and their trend, allowing the decision-making processes based on the more probable scenario (Ratto *et al.* 2003). However, it has been observed that the alert system, organized by Regione Piemonte Authority, and people preparedness hugely contributed to reduce the total number of victims. Dissemination of information of hazard alerts proved to important for a timely decision making.

After analysing the discrepancy between risk floods scenarios made before the flood and real flood effects, it should be noted that the reasons for such discrepancy have to be substantially found in the uncertainties which affected technical evaluations in the 1998 plan.

	What has been observed	What has been done
Risk assessment hydrological hazard	Discharges (measured at gauges stations or estimated by means of back analysis) during the flood revealed to be *outliers*. The flood peaks data, available at the date of the flood, had to be integrated.	Revision of hydrology Revision of the design discharge adopted by PBDA; the highest increment of the design discharge are about 20% (Tavagnasco cross section).
Risk assessment	Flooding areas resulted to be outside the hazard areas (see Figure 8). Geomorphological processes showed to have been underestimated.	Corridors A, B, and C have been revised (Figure 8) and the new PAI version was adopted again by PBDA, after discussion with regions, stakeholders, municipalities, people.
Risk evaluation	Some anthropic activities are inconsistent with respect to the risk assessment.	PBDA adopted a list of activities which are considered to be acceptable or not acceptable, depending on the risk level.
Risk treatment	The actual structural defence system revealed not to be effective	Risks and hazard maps have been updated, and non structural countermeasures have been implemented *Surplus* and *deficit* in the protection levels have been mapped on local scale (figure 9). Programming structural and non structural interventions (total about 100 M€) Reconstruction of damaged infrastructures

Table 8. Post flood analysis after 2000 flood event in Dora Baltea.

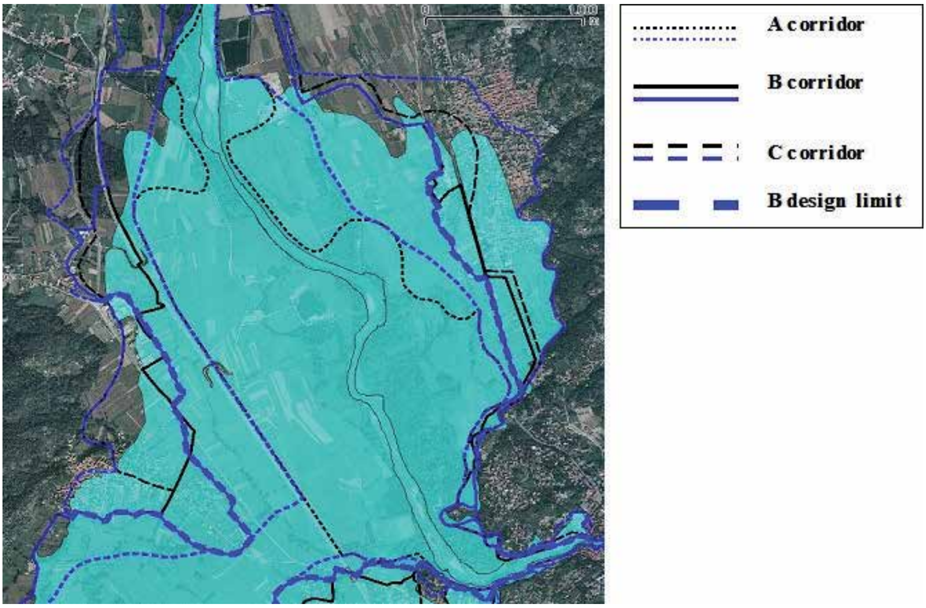


Fig. 8. Flooded areas near Ivrea, with the superposition of the fluvial corridors (before the event, black colour) and after revision of the PAI (in blue colour).

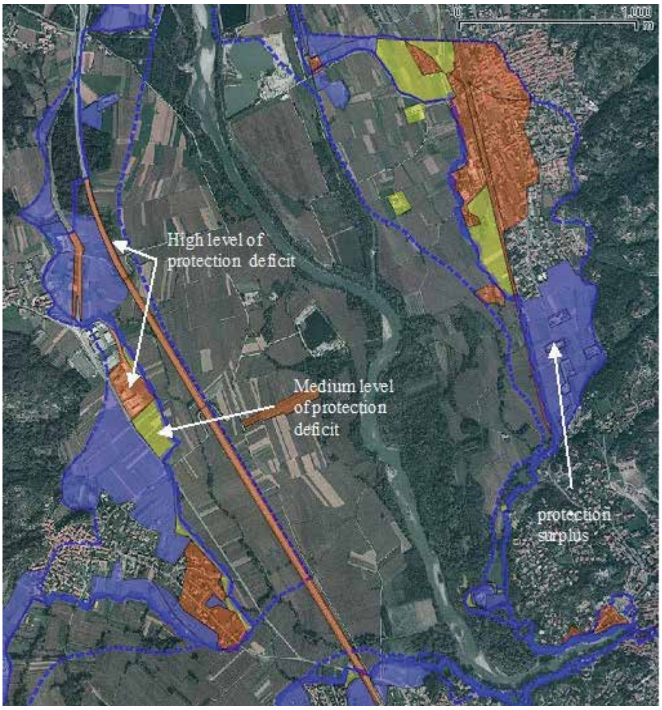

0
1,000 m
High level of protection deficit
Medium level of protection deficit
protection surplus

Fig. 9. Zoning of the areas near Ivrea (PBDA, 2008). Each colour corresponds to a different protection level. In yellow and orange areas there is a deficit in the protection level; in blue areas there is a surplus.

The 2000 flood put into evidence two different categories of uncertainty in technical assessments, which can be described as knowledge uncertainty and structural uncertainty (Floodsite, 2009).

As to the first source of uncertainty, after 2000 flood, the estimated peak discharge datum (reconstructed by means of a hydrological and hydraulic back analyses) resulted to be higher than the maximum historical recorded value, so that the hydrological assessment made before the flood event resulted to underestimated. Such a discrepancy between assessments and reality is typical when new data are available (Gribbs, 1969). The evidence of a new datum (the 2000 flood peak) improved the basic knowledge of the hydrological regime of Dora Baltea river, described by a set of data available in more than 70 years of observations (Claps, 2008). This induced the PBDA to revise past hydrological assessment and to consider more sever design discharge floods.

As to structural uncertainty, the geomorphologic dynamics of Dora Baltea river during the flood showed the necessity to improve the modelling of flooding. In particular the reactivation processes in Rio Ribes had to be modelled with a higher accuracy, and a higher detail. Actually, in 2000 flood the reactivation of the secondary river, Rio Ribes, near Ivrea occurred in a way that was not completely predicted, although reactivation processes were already taken into account in the PSFF plan adopted in 1998 (PBDA 1998), on the base of the reactivation which already occurred in in 1920, 1993 and 1775, 1834. Actually the overtopping of the 2000 flows occurred in two cross sections where, during the 1993 recent flood event, no flooding was experienced (Turin Province, 2006). A highway and a national road were seriously damaged, and urbanized areas were heavily flooded. The magnitude of reactivation has been a subject of analysis, resulting that the design discharges in Dora Baltea upstream the Rio Ribes is about 2800 m^3/s and that about the 40% of the total discharge flows in the secondary river.

In order to control the reactivation processes, a large weir has been design in the areas where the fluvial reactivation takes place, substantially controlling the way in which the discharges enter the Rio Ribes, and controlling therefore the total discharges that flow downstream.

The risk management plan in Dora Baltea is now in force since the 2008 (Deliberation of the institutional committee, n.4/2008; PBDA, 2008).

4.2 Effectiveness of flood risk management in torrents (Regione Piemonte, Po basin district)

From a normative point of view, following the indications given by the PBDA in the PAI plan, FRM approaches can vary from region to region, especially as far as the hazard and risk assessment methodologies are concerned.

Therefore, in the following, the practices followed in Piemonte region (Deliberation n. 2-11830) are described and their effectiveness is discussed, by referring to a 2008 flash flood in Pellice catchment.

Note that from a geographical point of view, it should be mentioned that Piemonte region is at the head of the Po river catchment and it is prevalently mountainous. Actually about 43% of the total territory lays in the Alps, or in Apennines, and abut 30 % is hilly. As a consequence, due to the total number of torrents and the relatively small dimensions of the catchments, the assessment of the flood risks has to be very detailed.

4.2.1 The practices in risk assessment, evaluation and treatment

As far as torrents are concerned, in principle, flood hazards can be assessed by means of the methodology previously described, but it should be clear that the related phenomena can be very different in reality.

What is more challenging is that, in Alpine catchments, different kinds of currents can alternate in time in the same torrent (and its fan), from those which are more related to landslide liquefactions, to debris flows, to immature debris flows (Takahashi 1991) to ordinary water currents. Experience in some instrumented basin (e.g. rio Moscardo, Italy) show that the flood processes in alpine torrents can have mechanics that are intermediate between those of water flows (showing a Newtonian rheology) to those of debris flows. Back analysis confirms this behaviour also from a phenomenological point of view, even if it is very difficult to understand the inner nature of the process, as both phenomena are impulsive, the onset is very fast and the transported sediment carried by the flow is not negligible (Arattano and Franzi, 2004, 2006). Moreover, during the same event, different processes can alternate in time or superpose. Some of the most critical situations are represented by floods on alluvial fan, which are due to the spreading of the currents on debris cone.

As far floods on alluvial fan are considered, the frequency of occurrence is widely debated in literature, as the flood intensities are strongly dependent on the sediment and debris transported by the flow. Therefore the statistical quantification of the severity of the process is generally very problematic. Also the recourse to modelling is very problematic, as the mechanical behaviour of the flood currents cannot be determined *a priori* and the rheology is in general non Newtonian. Consequently hazard mapping on alluvial fans necessarily considers different scenarios, the frequency of which is generally only roughly estimated. To this aim the historical data can help the decision makers to propose solutions for the endangered areas (PBDA, 2001a).

Simplification and classification is needed and regulation in planning generally helps the practitioners. As far as torrents, the PBDA classifies two different processes that are equally related to meteorological causes:

- the flooding processes along the torrents with or without the deposition or erosion of sediments; these are indicated as linear processes, since they take place along a watercourse; according to PAI implementation guidelines in Regione Piemonte (act n. 2-11830, 2009), flood risk assessment relies on the concept of the return period (Table 9);
- the flooding on alluvial fan, due to the water sediment flows entering the alluvial cone, with or without the deposition of sediments; these are addressed to fan processes; since the reference to a statistical flooding recurrence is not attainable (and the estimations of the recurrence periods are no more realistic) the PAI risk assessment in made by referring to *"fan reactivation"* concept (Table 9). This concept matches quite well with the definition given by FEMA (2007) about alluvial fan flooding: *"Flooding occurring on the surface of an alluvial fan or similar landform which originates at the apex and is characterized by high-velocity flows; active processes of erosion, sediment transport, and deposition; and unpredictable flowpaths" (FEMA, 2007).*

Torrent floods	Alluvial fan flooding
Very high hazard corridor – *Ee corridor*: areas flooded by a design discharge having a return period between T=20-50 years	Very high hazard (*Ca*): alluvial fans where torrents showed water-sediment flooding, with sediments deposition, in the last 30 years; the fan areas are not presently protected.
High hazard corridor - *Eb corridor*: it is outside the Ee corridor; areas flooded by a design discharge having a return period between T=100-200 years;	High hazard (*Cp*): alluvial fans where torrents showed water-sediment floodings, with sediments deposition, in the last 30 years; the fan areas can be protected, but the works are not still sufficient to maintain an acceptable level of risk.
Medium-moderate hazard corridor – *Em corridor*: outside the Eb corridor; areas flooded by the design discharge having a return period between T=200- 500 years;	Mean or moderate hazard (*Cn*): alluvial fan where torrents did not showed any flooding process in the last 30 year or where flooding areas in the fan are protected (protection works along the torrent or located on the alluvial fan)

Table 9. Criteria for hazard zoning in Piemonte region, following the indications given in PAI (PAI, Relazione Generale, p.213). The criteria have been modified by the regional deliberation DGR2-11830, published on the official regional bulletin (Note: the indication in the table is an oversimplification of the deliberation, which consists on tens of pages).

Obviously the two kinds of processes are strongly connected, since the only distinction between the two is the topographic location (along a torrent or o the torrent fan) but there is a physical continuity in the water-sediment discharges that run torrents, deposit on debris fans, and flow downstream, connecting, as tributaries, to other rivers or torrents.

In this frame, the risk assessment in alluvial fan flooding is affected by higher uncertainties, mainly due to the fact that the estimation of the recurrence period of such processes remains unknown. It is generally false that the recurrence of the alluvial fan flooding processes is the same as that of hydrological triggering conditions, even when mathematical formulations of hydrological thresholds are proposed in literature. Moreover the intensity of the processes can strongly vary from one event to another, so that the total volume of deposited sediments or the morphological changes or the channel cutting on the fan, can strongly change according to the “boundary” conditions which are unknown.

For this reasons, the strategy that is generally followed, is based on civil protection measures.

What is more challenging for civil protection agencies is what should be done in real time, in critical hydrological conditions. Actually considering the total number of fans that are potentially endangered (in Piemonte the total number of alluvial fans is about 2400, Arattano *et al.* 2010), risks cannot be generally managed by adopting *generalized* civil protection countermeasures (i.e. by ordering evacuations, which would result in false alarms) in real time conditions; the strong spatial variability of rainfall intensities and the strong uncertainties in threshold rainfall conditions cause the decision making to be very uncertain too.

For instance, 24 hours before the fan flooding in Villar Pellice in 2008 (Figure 10), the regional environmental agency, by applying some meteorological models, established that in the Pellice valley the risk level, for the next 36 hours, was rated "3" ("high criticality" level) in a scale that ranges from 1 to 3. This level corresponds to the possible occurrence of "soil slips, debris flows and floodings".

However the bulletin was based on a rainfall forecasting at a regional scale and it could not give more precise indications on the exact location where the highest rainfall intensities might have occurred. Therefore the issue of the determination of the possible receptors at risk was just "downscaled" from a total 2400 fan areas, to about 100 fan areas of the Pellice river basin (293 km^2). What is more impressive is that some hours before the event, there were no rainfall evidences about a possible flooding on Villar Pellice. Risk assessment in real time conditions showed to have strong limitations.

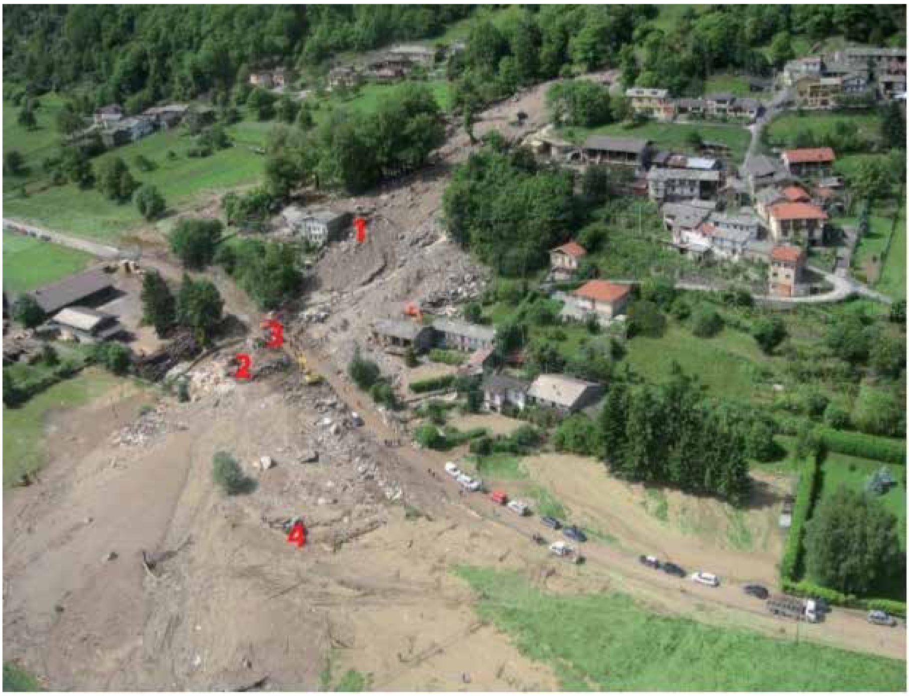


Fig. 10. Aerial view of Garin, affected by the debris flow of Cassarot creek occurred on 29 May 2008. 1 – destroyed house; 2 – more destroyed building; 3 – location where one of the cars hit by the flow was found (the fourth victim); and 4 – location where the ambulance and the tractor, hit by the flow, were found.

Uncertainties effect risk assessment, risk evaluation and also risk treatment. Decision making on the priority of interventions ("if" and "which" alluvial fans should be sheltered) on a total of 2400 fan areas is therefore a very hard task, that lead local administration to ask for higher economical budgets, in order to have the chance to treat the highest possible number of situations. Consequently the implementation norms (PAI, 2001b) are more restrictive as far as the activities are concerned. In Table 10 an example regarding the Eb corridor and the Ca zones is provided.

	Existing structures		New structures	
	Allowed	Not allowed	Allowed	Not allowed
Eb corridor	Enlargement for sanitary and functional adaptations. Manutentions. Renovations, without enlargement in area and cubage	Enlarging cubage. Enlarging total occupied area.	Drinkable treatment implants	New settlements.
Ca	Demolition without reconstruction. Ordinary and extraordinary maintenance work. Vulnerability mitigation devices.	Enlarging cubage. Enlarging total occupied area. Agricultural changes in cultivation types.	Defence works. Enlargement of implants for sever water treatment.	New settlements.

Table 10. Example of regulation norms in Eb corridor and in the Ca alluvial fan. The table is a simplification of the art.9 of the PAI implementation norms (PAI, implementation norms, 2001).

4.2.2 Discussion on the effectiveness of FRM – The 2008 fan flooding in Villar Pellice

The 2008 event occurred in Villar Pellice can be considered representative of the many high uncertainties which practitioners and decision makers have to deal with, when the alluvial fan risk assessment is considered.

Indeterminacy of the flood occurrence was mainly due to the fact that the processes, that is the spreading of water and sediment on the alluvial fan, could not (or only partly could) be determined from the data available, i.e. real-time rainfalls and historical background.

During the May 28-30 2008 meteorological event, many soil slips originated in the Pellice valley, generally triggered above 1500 m a.s.l.. The alluvial fan flooding in Villar Pellice hit the inhabited areas at 10.25 a.m. and was preceded by the activation of a series of soil slips that interrupted in different points the viability, making the area inaccessible. The deposits left by the flow that run the Cassarot creek covered an area of about 28,000 m^2, with a maximum thickness up to 3 m. The estimated total transported volume was about 40,000 m^3 (Lollino *et al.,* 2008).

The May 29, 2008 disaster caused tragic effects: four casualties, 24 four buildings and three cars (among which an ambulance) hit by the debris flow (Figure 10). Four people died, three in the house destroyed by the current and another in a car (number 3 in Figure 10). Another house was also destroyed (number 2 in Figure 10) without losses.

As witnessed by the Villar Pellice inhabitants, the current consisted probably of a single debris flow wave that was preceded by a loud noise. The witnesses were not aware of what was going on, even if the water discharge of the Cassarot creek had already flooded the street. A technician of the municipality carried out a survey about 40 minutes before the event. According to this survey, the Cassarot creek already showed a very high discharge that was however contained within the channel, there were no flood evidences and the road crossings were all practicable. Before the debris flow arrival, there were no evident signs of the imminence of the event. Rainfalls intensities in the hours before the flooding showed a strong spatial variability but anyhow they cannot considered as exceptional. At Bobbio Pellice rain-gauge station (the nearest to the flooding area) the rainfalls intensities maintained for several hours less than 10 mm/h (an intensity which is not exceptional for alpine areas), and increased suddenly from about 12 mm/h to more than 45mm/h, when probably the debris flood was triggered (Figure 11).

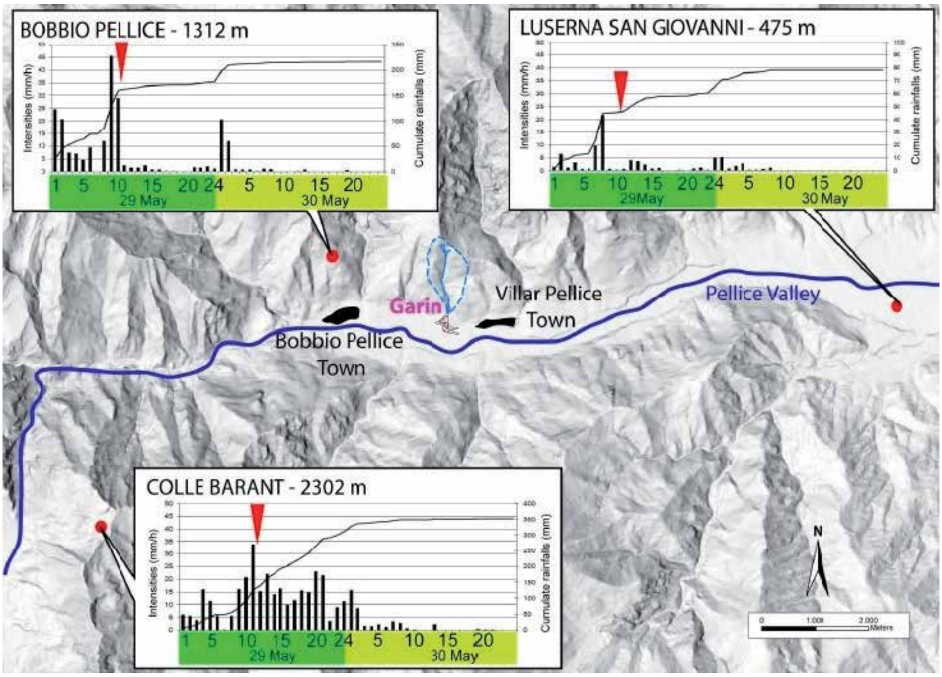


Fig. 11. Location of the raingauges near Villar Pellice with the intensities and cumulated rainfalls on 29–30 May 2009 (the start time for all of the graphs was midnight on 29 May). The red arrow shows the debris flow time occurrence.

From the viewpoint of territorial planning, the available data and the field surveys carried out on the Cassarot torrent, had allowed a hazard mapping on the fan at a local scale, as indicated in figure 12. The hazard areas mapped in Villar Pellice, which were proposed to the local administration but were not still in force, were partially confirmed by the 2008 event. In particular, one of the destroyed houses (house A in Figure 12) was in an area where the debris flow hazard was classified as "*high*" and "*partially*" protected (*Cp* areas, figure 12), while in the adjacent areas the hazard was classified as "medium" (*Cn* areas). The other destroyed house (B in figure 12) was in an area where the flood hazard was classified as "*very high*" (*Ee* areas).

The comparison among the risk assessment and the reality allowed us to reconsider risk assessment on the fan.

As far as zone mapping on alluvial fan is considered, the zoning of high hazard matches quite well with the flooding areas. According to hazard maps, in Ee areas (which have been considered outside the possible fan flooding processes), the prevailing hazards were due to valley processes, that is the flooding of Pellice river which flows downstream the fan. Risk assessment and evaluation was probably biased by historical information about antecedent processes, a source of information that guide technical decision making in high uncertainties. Actually, the comparison of different images of the Garin alluvial fan (Figure 13), shows that, after the 1977 event, the processes affected a part of the fan without causing any damage to the house destroyed in 2008. In 1998 the image shows the house destroyed in 2008 as it was until the day before the 2008 event; the image presents a very different situation compared to the following and it is particularly important because it allows to evaluate which was the morphological context in which those who carried out an evaluation of the hazard of the area had to operate. The 2008 picture shows the post-event situation in all its drama.

The experience gained in the alluvial flooding allowed the regional authorities to make more steps towards real time assessment of hazards and risks. In particular it was observed (Table 11) that the assessment of the risks could be improved by increasing the risk perception of people living in endangered areas, and helping majors of small municipalities to take decisions in real time by experts (Booker et a., 2009). Moreover a collaboration among weather forecasting centres and experts is also necessary, so that the real time estimations of local rainfalls obtained by radars can allow to take decisions more rapidly and to improve the performance of civil protection strategies and protocols (Borga *et al.*, 2006).

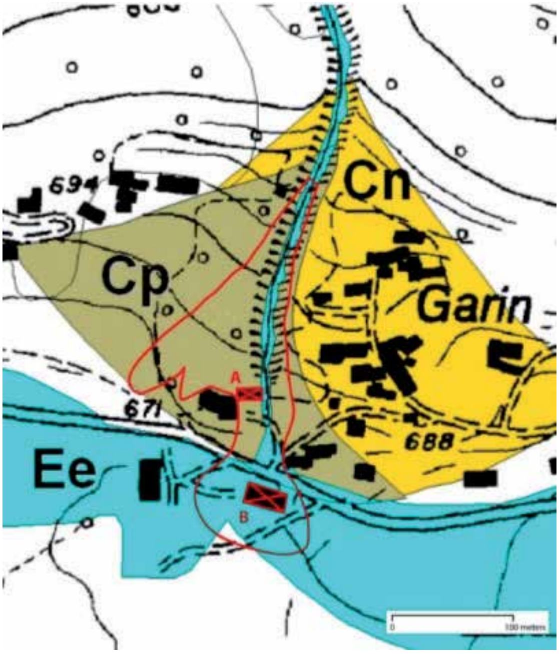


Fig. 12. Hazard mapping on the Villar Pellice fan. In this picture, the areas that are shown are classified by areas at risk: CP – area characterised by high debris flow hazard and partially protected; Cn – area characterised by medium debris flow hazard; Ee – area characterised by very high water flood hazard. The red line shows the limit of debris flow deposits.

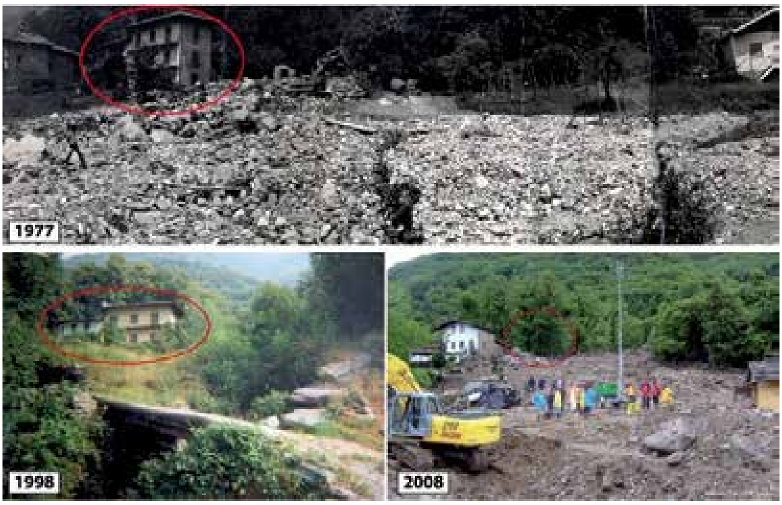


Fig. 13. Comparison of different images of the Garin alluvial fan: after the 1977 event – as the image shows, the debris flow affected a part of the fan without causing any damage to the house destroyed in 2008. In 1998 – the image shows the house destroyed in 2008 as it was until the day before the 2008 event.

	What has been observed	What has been done
Risk assessment	Rainfall conditions showed a high spatial variability (in the range between 10 mm/h to 45 mm/h. The processes showed high onset velocity and very high energy. Rainfalls , some hours before the onset, could not be considered as exceptional. There was no evidence, in real time conditions, about the onset of the process. Flooding areas resulted to partially match with those mapped in the hazards maps;	Threshold rainfall conditions for the process onset are going to be determined. Risk assessment has been revised locally; no return periods for this kind of events have been assessed. A different hazard zoning has been proposed
Risk evaluation	The presence of houses resulted to be inconsistent with respect to the evidences of risk. The high onset velocity did not leave enough time for evacuation.	Risk evaluation was revised, locally. People living in areas where the risk is not tolerable, have been relocated.
Risk treatment	Mitigation of risk by means of structural countermeasure is not probably effective. The energy of the process on the fan area shows that the vulnerability of the houses cannot be properly reduced.	Some houses are going to be relocated. More effective civil protection countermeasures are in study. An improvement in spatial and temporal forecasting, action protocols, dissemination of hazard alerts and timely decision making is needed. Real-time meteorological alerts through the analysis of radar images are in study.

Table 11. Post flood analysis after 2008 disaster in Villar Pellice.

5. Closure and future work

Experience in risk management let decision makers to consider that floods are generally the result of complex interaction among different factors, mainly anthropogenic and natural, random and deterministic. Many factors are considered to be constant in time (assumption of *stationarity*) without considering any dynamic changes. If, on one hand, this is a first order approximation for the description of natural phenomena, on the other hand it should be clear that the watershed characteristics undergo continuous changes in time. Plans should be enough flexible to adapt to these changes.

Together with the uncertainties due to the *non-stationarity* of the factors, a reliable FRM strategy should also take into account the different sources of uncertainties, which can be very representative, as in the case of the Dora Baltea river or of the Cassarot torrent. It is therefore confirmed that risk planning is an on-going process that is an interim product based on information and understanding at the moment, and is (and should be) subject to revision. That is why plans are best described as "living" documents.

From this point of view, research on *how* territorial and FRM planning can efficiently and rapidly adapt and change, according to a changing context, is strongly desirable and advantageous. Actually there are many local risk situations where the present norms and regulations do not perfectly match with the actual flood risk conditions. Therefore a tighter collaboration between practitioners and planners would be an important step for bridging the gap between real risk situations and planning.

Moreover a plan should efficiently balance different strategies, according to the uncertainties that affect our understanding of the processes.

The lower is the total uncertainty, the most consistent are the estimation of the risks. Consequently, in these cases, the prevailing strategy in FRM should consider a scenario-based planning, where the risk conditions can be predicted with reliable approximations. This is the case of Dora Baltea river, where either the available data and the understanding of the physical, phenomenological and social *context* allow a strategy that can be based on flooding scenarios. Revisions of these scenarios are needed, especially after the evidences of recent floods, and the *contents* of the prevailing strategy (and in particular the types of countermeasures), remain the same.

On the contrary, the higher is the total uncertainty, the less consistent are the results in risk assessment. The prevailing strategy, on its turn, should be based on civil protection countermeasures, including the countermeasures in emergency conditions and after emergence. The 2008 flooding on Garin fan showed that the processes were not perfectly understood. Both the social *context* (inhabited areas on the fan) and the physical *context* (the onset conditions and the energies involved) strongly advice the decision maker to consider the limitations of a *scenario-based*-strategy.

This distinction between these two cases, which are two boundary conditions, is simple in theory but it very difficult in practical applications. Actually, in theory, the two strategies are technically feasible, and can be implemented by decision makers. In practice the researchers, the practitioners, the professionals involved in FRM do not exactly know (or perceive) "how far" their perception/estimation/assessment of risk is from real risk. This cognitive bias is a strong limitation in practical cases, as operators cannot move from preconceptions, but instead anchor to them even in light of new data/information. Consequently, in practice, the choice of the prevailing strategy that should be adopted is not always straightforward.

Future works should focus on the application field of the two strategies, considering the relationship between the (theoretical and practical) effectiveness of the FRM strategy vs the uncertainty in the description of the hazards. Also scientific investigation should pay more attention on identification (and quantification) of uncertainty, instead of celebrating the results obtained in very particular cases. Actually, in many scientific investigations, the perfect match between computed results and observed data sometimes can give the false impression to be "perfectly right", causing an overestimation of the available scientific instruments: software, measurement systems, laboratory investigations. On the contrary, it should be taken into account that often, in a changing context, our knowledge can reveal to be inappropriate.

Finally we should think of dealing with uncertainty of flood risk in long-term planning as a process prone to interruption, irrelevance for ongoing decision making, and post-disaster politics. This is efficiently reported by (Floodsite, 2005): "*Long-term planning, as strategic planning (Bryson 2004), is prone to be interrupted because of decision makers shifting their attention to pressing problems of the day (…). Usually, elected politicians and citizens have much more on their agenda than long-term planning in flood risk management*". This is a non-scientific, but crucial aspect in FRM strategies, which can also imply the necessity to a cultural change in risk management.

6. Acknowledgments

It is a great pleasure to thank prof. Bianco of the Politecnico of Turin for his continuous suggestions and hints in this wide and interesting filed of research. I am grateful to Dr. Villacis of UNDP, Dr. Abrate and Dr. Teruggi of the WMO, for the precious documentation and the kind invitations in their activities and debates; to Dr .Arattano, Dr. Luino and Dr. Giordan of the CNR-IRPI of Turin for the documentation of the 2008 flash-flood and the useful discussions. Institutional thanks to our Department director Coccolo, who honoured this chapter with his scientific advices, to ing.Ercole for his precious hints and to our supervisor ing. Masoero. Special thanks to my daughter Chiara, who hinted me how to look beyond the appearance of things.

7. References

Acanfora, E. (1990). Sigismondo Coccapani un artista equivocato, *Antichità viva*, Vol.29.

Arattano, M. & Franzi, L. (2004). Analysis of different water-sediment flow processes in a mountain torrent, *Natural Hazards Earth System Science*, Vol.4, pp. 783-791, available from:
http://www.nat-hazards-earth-syst-sci.net/4/783/2004/nhess-4-783-2004.html

Arattano, M., Conte, R., Franzi, L., Giordan, D., Lazzari, A. & and Luino, F. (2010). Risk management on an alluvial fan: a case study of the 2008 debris-flow event at Villar Pellice (Piedmont, N-W Italy), *Natural Hazards Earth System Science*, Vol.10, pp.999–1008, available from:
www.nat-hazards-earth-syst-sci.net/10/999/2010

Arattano, M., Franzi, L. & Marchi, L. (2006). On the influence of rheology on debris flow mathematical simulation: a real case, *Natural Hazards Earth System Science*, Vol.6, pp. 519–528, available from: http://www.nat-hazards-earth-syst-sci.net/6/519/2006/nhess-6-519-2006.pdf

Arcilla, A., Jimi´Nez, J.A. Valdemoro, H.I. & Gracia, V. (2007). Implications of climatic change on Spanish Mediterranean low-lying coasts: the Ebro delta case, *Journal of Coastal Research,* 24, ISSN 0749-0208.

Autorità di Bacino dell'Arno (1997) *Trasformazione del territorio e sviluppo dell'edificato lungo il corso dell'Arno e degli affluenti (1954, 1993 e 1995),* quaderno n.7, available from: http://www.arno.autoritadibacino.it/cont/testo.php?id=37&biblio=2

Autorità di distretto idrografico del fiume Po (2008). *Adozione di Variante del Piano stralcio per l'Assetto idrogeologico - Variante delle fasce fluviali del fiume Dora baltea*; deliberazione n.4/2008, available from:

Autorità di distretto idrografico del fiume Po (2006). *Caratteristiche del basino del po Autorità di distretto del bacino del fiume Po,* available from: http://www.adbpo.it/download/bacino_Po/AdbPo_Caratteristiche-bacino-Po_2006.pdf

Autorità di distretto idrografico del fiume Po (1998). *Direttiva piena di progetto,* Institutional committee acts, Italy (in Italian)

Autorità di distretto idrografico del fiume Po (2001a). *Piano di assetto idrogeologico - Relazione generale,* Institutional committee acts, Italy (in Italian)

Autorità di distretto idrografico del fiume Po (2001b). *Piano di assetto idrogeologico - Norme di attuazione*, Institutional committee acts, Italy (in Italian)

Autorità di distretto idrografico del fiume Po (2001c). *Piano di assetto idrogeologico - Quaderno delle opere tipo*, Institutional committee acts, Italy (in Italian

Beven, K. & Freer, J. (2001) Equifinality, data assimilation, and data uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *Journal of hydrology*, Vol.249, pp.11–29.

Booker, J.M., Anderson, M.C.n Meyer M.A. (2009). The role of expert knowledge in uncertainty quantification, *Proceedings of U.S. Army Conference on Applied Statistics*, 2009.

Borga, M., Degli Esposti, S. & Norbiat, D. (2006). Influence of errors in radar rainfall estimates on hydrological modelling prediction uncertainty, *Water Resources Research*, Vol. 42

Bovo, S., Forlati, F, Campus, S., Barbero S. (2007) *Evaluation and Prevention of natural risks".* London, Taylor & Francis/Balkema

Bryson, J. M. (2004), *Strategic Planning for Public and Nonprofit Organizations. A Guide to Strengthening and Sustaining Organizational Achievement*, Jossey-Bass, San Francisco.

Claps, P. & Laio F. (2008). *Aggiornamento delle procedure delle procedure di valutazione delle piene in Piemonte,con particolare riferimento ai bacini sottesi da invasi artificiali. VOLUME I:Costruzione e applicazione delle procedure di stima delle portate al colmo di piena,* Report of the Dipartimento di Idraulica,Trasporti ed Infrastrutture Civili,Politecnico di Torino,306 pp., 2008, available from: http://www.idrologia.polito.it/web2/persone/team-members/claps/pubblicazioni/

Chandler, A. D. (1962). *Strategy and Structure. Chapters in the History of the Industrial Enterprise,* The MIT Press, Cambridge/Massachusetts.

Coccapani, S. (1610) *Trattato del modo di ridurre il fiume Arno in canale*, National Biblioteque, Florence.

Ducrocq, V., Nuissier, O., Ricard, D., Lebeaupin, C. & Thouvenin, T. (2008). A numerical study of three catastrophic precipitating events over southern France. Mesoscale triggering and stationarity factors, *Quarterly journal of the royal meteorological society,* Vol.134, pp. 131–145.

Department of Transportation (2005). *Risk Management Definitions,* Washington, DC: DOT, Office of Hazardous Materials Safety, 2005.

Department of homeland security (2008) *National response plan.* available from: http://www.dhs.gov

Disaster Recovery Journal and DRI International. *Generally Accepted Practices For Business Continuity Practitioners*. 20 Aug 2007

Easterby, S. & Lysles, M. (2003). *Handbook of Organizational Learning and Knowledge Management*, Malden/USA, Basil Blackwell.

Environmental Protection Agency (1986). *Emergency Planning & Community Right to Know Act (42U.S.C. 11001 et seq., 1986).*

European Commission (2007), *Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks*, available from:
http://ec.europa.eu/environment/water/flood_risk/index.htm

European Environmental Agency (2007). *EEA Multilingual Environmental Glossary,* Copenhagen, Denmark. Accessed at:
http://glossary.eea.europa.eu/EEAGlossary

Federal Emergency Management Agency (2007) *Guide to emergency management and related terms, definitions, concepts, acronyms, organizations, programs, guidance, executive orders & legislation* available from:
http://training.fema.gov/EMIWeb/edu/docs/terms%20and%20definitions/Terms%20and%20Definitions.pdf

Federal Emergency Management Agency (2007) *Alluvial Fan Flooding*, available from:
http://www.fema.gov/plan/prevent/floodplain/nfipkeywords/alluvial_fan_flooding.shtm

Federal Emergency Management Agency (1997). *Multi Hazard Identification and Risk Assessment – ACornerstone of the National Mitigation Strategy*. Washington.

Floodsite (2005). *Strategies for pre-flood Risk management,* Report Number T13-07-04.

Floodsite (2009). *Language of risk project definitions (second edition),* Report Number T32-04-01

Franzi, L. & Bianco, G. (2001). A statistical method to predict debris flow volumes deposited on a debris fan, *Physics And Chemistry Of The Earth Part C-Solar-Terrestial And Planetary Science,* Vol. 26, No.9, pp. 683-688, ISSN: 1464-1917.

Green, C., Van der Veen, A., Wierstra E., & Penning-Roswell, E. (1994). Vulnerability Refined: Analysing Full Flood Impacts. In *Floods Across Europe: Flood Hazard Assessment, Modelling and Management.* Eds. Edmund Penning-Roswell and Maureen Fordham. London: Middlesex University Press: 32-68.

Grubbs, F. E. (1969) Procedures for detecting outlying observations in samples. *Technometrics, Vol.*11, pp. 1–21.

Hall, J.W., Weadowcroft, I.C., Sayers, P. & Bramley, M. (2003), Integrated Flood Risk Management in England and Wales, *Natural Hazards Review*, Vol. 4, No. 3, pp. 126-135.

Hooijer, A., Klijn, F., Pedroli B, & Van Os, A. (2004). Towards Sustainable Flood Risk Management in the Rhine and Meuse River Basins: Synopsis of the Findings of IRMASPONGE, *River Research and Applications*, Vol. 20, pp. 343-357.

Hutter, G. (2006), Strategies for Flood Risk Management. A Process Perspective, In: *Flood Risk Management . Hazards, Vulnerability and Mitigation Measures*, Schanze J., Zeman E., Marsalek J. (eds.), pp. 229-246, Springer, Berlin,.

Kersting, N.F. (2008). *Changes in flood management strategies over time.* Unesco-IHE website, available from:

http://www.unesco-ihe.org/Flood-Management-Education-Platform/Flood-Modelling-for-Management2/Changes-in-Flood-Management-Strategies-over-Time

Klaus J, Pflügner W, Schmidtke R F, Wind H., & Green C (1994). Models for Flood Hazard Assessment and Management, In: *Floods across Europe. Hazard assessment, modelling and management,* pp.67-106, Penning-Rowsell E C, Fordham M (Hrsg.).

Klijn, F., Samuels, P., Van Os A., (2008). Towards Flood Risk Management in the EU: State of affairs with examples from various European countries, *Journal of River Basin Management,* Vol. 6, No. 4, pp. 307–321.

Kuczera, G. & Mroczkowski, M. (1998) Assessment of hydrologic parameter uncertainty and the worth of multi-response data, *Water Resources Research,* Vol.34, No.6, pp. 1481-1490.

ItalianParliament (1989) *act 183/89,* published on Gazzetta Ufficiale on 18 may 1989.

International Organization for Standardization (2007). Societal Security – Guideline for Incident Preparedness and Operational Continuity Management, Nr. 22399.

Lollino, G., Mortara, G., Luino, F., & Giordan, D. (2008). *Colata detritica torrentizia in localit`a Garin (Comune di Villar Pellice – TO) CNR report of event,* available from:
http://www.irpi.to.cnr.it/documenti/CNR-RPI Rapporto% 20Garin 9%20giugno%202008.pdf

Merkel, U. & Westrich, B. (2008). Pc-river - probabilistic reliability analysis for river dikes, *Proceedings 4th International Symposium on Flood Defence: Managing Flood Risk, Reliability and Vulnerability,* Toronto, Ontario, Canada, May 6-8, 2008.

Office for public works, Environment heritage and local government (2008) *The planning system and flood risk management, Guidelines for planning authorities,* rep. n. 20.

Paleocapa, P. (1868). *Osservazioni sulla parte idraulica della Legge 20 marzo 1865 per l'ordinamento dei lavori pubblici,* Verona; available at:
http://archivio.camera.it

Penning-Rowsell, M.A. & Green, M. (2000). New Insights into the Appraisal of Flood-Alleviation Benefits: Flood Damage and Flood Loss Information, *Water and Environment Journal,* Vol.14, No.5, pp. 347–353.

Pettigrew, A., Thomas, H, Whittington R (eds.) (2002). *Handbook of Strategy and Management,* London, Sage.

Pettigrew, A., & Whipp, R. (1991). *Managing Change for Competitive Success,* Oxford/UK, Blackwell.

Poole M S, Van De Ven A (eds.) (2004), *Handbook of Organizational Change andInnovation,* New York, Oxford University Press.

President of the Ministries Council (1990). Decreto published on *Gazzetta Ufficiale* 4 aprile 1990, n. 79.0

Regione Piemonte (2009), Deliberation n. 2-11830, R*egional official bulletin n.34*, available from: http://www.regione.piemonte.it/governo/bollettino/abbonati/2009/34/siste/00000512.htm (in Italian)

Regione Piemonte (2012). *Banca Data geologica,* available from: http://webgis.arpa.piemonte.it/bdge/index.php

Ratto, S., Bonetto, F. & Comoglio, C. (2003). The October 2000 flooding in Valle d'Aosta (Italy): Event description and land planning measures for the risk mitigation, *Journal of River Basin Management,* Vol. 1, No. 2, pp. 105–116.

Sayers, P. B., Hall, J.W, Meadowcroft, I. C. (2002). *Towards Risk-based Flood HazardManagement in the UK*, Proceedings of Institution of Civil Engineers, pp. 36-42.

Stalenberg, B. & Vrijling, J.K. (2006*).* Interaction between Dutch flood protection and urbanisation, *Proceeding of the International Symposium of Lowland Technology*, Saga, Japan, September, 2006

Tacitus, P.C. (115). *Annales*, Vol.1, available at: http://www.progettovidio.it/dettagli1.asp?id=1962&opera=Annali&libro=Libro%20I

Takahashi, T. (1991). *Debris flow,* IAHR Monograph Series, Balkema,Rotterdam, The Netherlands, 165.

UNDP-DHA (1994). *Vulnerability and risk assessement*, Disaster management training program, 2nd edition.

UNDRO (1980). *Natural disasters and vulnerability analysis,* Office of the United Nations Disaster relief Co-ordinator /UNDRO), report of expert goup meeting, 1980.

Varnes (1984). *Landslide hazard zonation, a review of principles and practice,* UNESCO ed.

Vrijling, J.K., & Van Gelder, P.H.A.J.M., (1997). The Effect of Inherent Uncertainty in Time and Space on the Reliability of Vertical Breakwaters, *Proceedings of the ESReDA Seminar on Industrial Application of Structural Reliability Theory*, pp.223-239, Paris, France, October 2-3, 1997

Weather meteorological organization (2007) *Economic aspects of integrate flood management*, Flood management policy series. APFM, 2007

Weather meteorological organization (2002) *Guide on improving public understanding of and response to warnings,* report WMO/RTD, no,1139, 2002

Weather meteorological organization (2004) *Integrated flood management*, Associated program in flood management 2004

Weather meteorological organization (2005) *Overview situation paper on flood management practices,* Associated program in flood management

Weather meteorological organization (2009) *Risk sharing in flood management*, Associated program in flood management.

Weather meteorological organization (2006) *Social aspects and stakeholders involvement in integrated flood management,* Associated program in flood management

Analysis of Historical River Floods – A Contribution Towards Modern Flood Risk Management

Jochen Seidel[1], Paul Dostal[2] and Florian Imbery[3]

[1]Institute of Hydraulic Engineering,
Department of Hydrology and Geohydrology, University of Stuttgart, Stuttgart
[2]German Aerospace Center, Bonn
[3]German Meteorological Service (DWD),
Department Climate and Environment Consultancy, Offenbach am Main
Germany

1. Introduction

The occurrence of several extreme flood events in Central Europe in the last two decades, in particular the flood along the Elbe River in 2002, and the resulting damage have shown shortcomings in the field of flood protection and have raised discussions on how to deal with flood risk in the future. As a consequence, the German Federal Ministry of Education and Research funded the "Risk Management of Extreme Flood Events" (RIMAX) research programme, which aimed at the development and implementation of improved instruments for flood risk management (www.rimax-hochwasser.de).

A first step towards modern flood risk management is the assessment of the flood risk. Many installations for flood protection are usually designed for a specific flood discharge, commonly the 100 year return period is used. Therefore, extreme value statistics and the calculation of return times are an important and wide spread tool to assess the flood risk in river catchments, which is a fundamental part of flood risk management. One major drawback of extreme value statistics is the underlying data series, which often comprise only a few decades. The occurrence of flood events frequently shows a cyclic behaviour. Therefore, extreme value statistics based on short data series often do not reflect or capture the behaviour of a specific catchment sufficiently.

Models for risk assessment are only as good as the data used in their development. The development of detailed, calibrated and verified flood modelling is therefore essential in order to accurately assess exposure. This requires good records of data such as rainfall, stream gauge and tidal information as well as historical flood levels and especially historical flood reports including historical hydrometeorological data (Middelmann, 2007). The analysis of historical flood events can shed some light on flood triggering mechanisms and can be applied to extend data series into the past in order to improve the validity of return times. Therefore, there's a need to assess extreme flood events with a high magnitude and a low probability of occurrence. Hence, one of the fundamental deficiencies of flood risk management often is the lack of information or knowledge of extreme flood events which occurred in the past. Even if there is some information of historical flood events (e.g. flood marks on old buildings) in a specific river basin, it is still difficult to convert this qualitative historical information into a quantity which can be used to evaluate extreme discharges, for example. This deficit can be overcome by a detailed reconstruction of past flood events.

In this chapter, we present a case study for the Neckar catchment in south west Germany, where different independent methods are applied to reconstruct and analyse two historical flood events in 1824 and 1882. The weather conditions which led to these floods were reconstructed by using the information from various historical sources (e.g. archive records, qualitative and quantitative meteorological observations, contemporary newspaper reports, chronicles, etc.). Discharges in the Neckar catchment for these two flood events were simulated with a water-balance model and compared with the estimated discharges on the basis of historical river morphology data such as cross profiles and water surface slopes. These results were then used to extend the data series for a gauging station in the Neckar River, where modern discharge data exits from 1921 onwards. By including the discharges from the extreme floods in the 19th century the data series for this site could be extended by 100 years into the past. This information leads to a better assessment of the river discharge characteristics and to a more stable calculation of return times.

This chapter presents a geographic description of the study area followed by an overview about state of the art research in this particular subject. Afterwards, the data basis for reconstructing the extreme flood events and the methods applied for this case study are described. Finally, the advantages and limitations of the case study are discussed.

2. Study area

The Neckar River basin (area: 14,000 km^2) is located in the south west part of Germany and is a tributary of the Upper Rhine Valley (Fig. 1). The elevation in the Neckar catchment ranges from 1020 m above sea level (a.s.l.) in the Black Forest to 78 m a.s.l. at Mannheim; the mean value is 435 m a.s.l.

The precipitation in the region is strongly modified by the local orography. The highest mean annual precipitation values of 2000 mm are recorded in the mountain ranges of the northern Black Forest. Towards the eastern part of the study site, the mean annual precipitation decreases to values of approximately 800-1000 mm. The Neckar is the principal tributary of the Upper Rhine, rises in the eastern Black Forest and is 367 km long. It is the river with the largest catchment in the Federal State of Baden-Württemberg, south west Germany. The Neckar passes through the cities of Tübingen, Stuttgart and Heidelberg and flows into the River Rhine at Mannheim (Fig. 1). The Neckar is navigable from Plochingen at the inflow of the Fils (km 202.5) to Mannheim (at the river mouth) and beside the rivers Rhine and Main, it is one of the three main waterways in Baden-Württemberg. Several industrial centres are situated in the Neckar basin around Stuttgart and Mannheim. These areas, therefore, constitute a high flood damage potential.

One of the most extreme and devastating floods in south west Germany was the catastrophic flood event in October 1824, which also affected the adjacent areas in southern Germany and eastern France. A further severe flood event occurred in December 1882. The first gauging station at the Neckar River was installed in Heilbronn in 1827 (Centralbureau für Meteorologie und Hydrographie, 1889). Starting in 1881, continuous and systematic

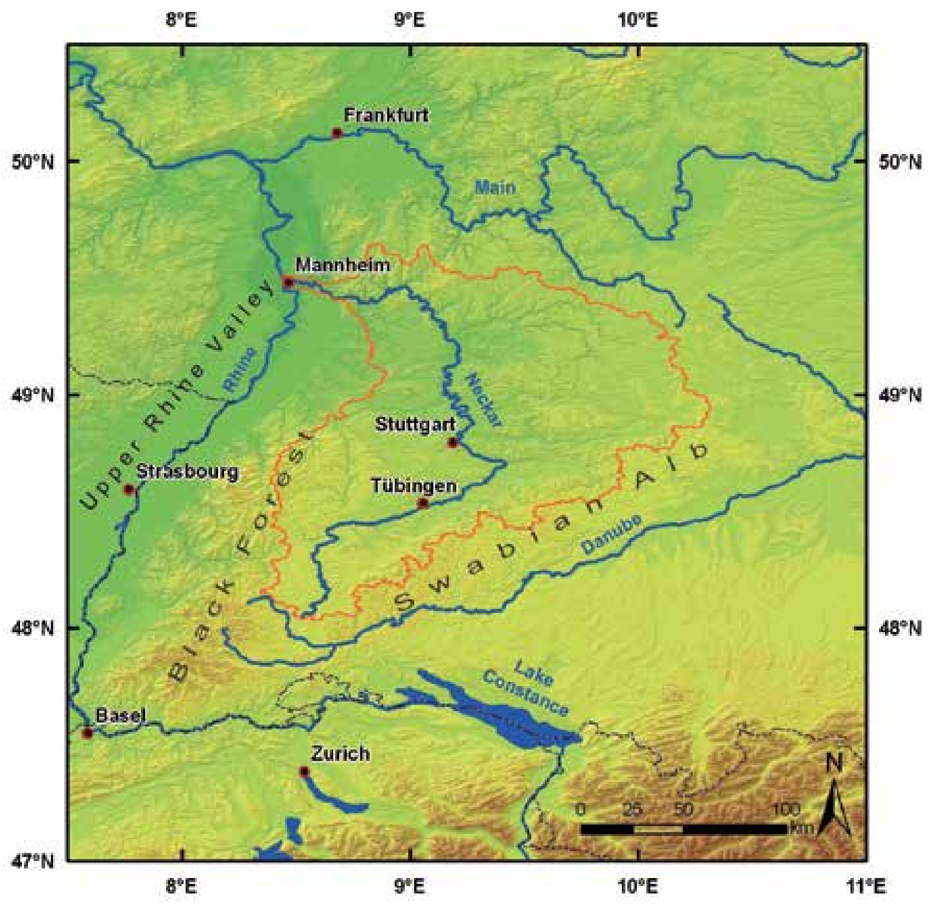


Fig. 1. Overview map of the study area. The orange line depicts the catchment of the Neckar River

measurements of the water levels were carried out at six gauging stations along the Neckar River (Statistisch-Topographisches Bureau, 1883). At present, 17 gauging stations are installed along the main channel of the Neckar.

3. Approaches and methods for reconstructing historical flood events

As stated above, the information from historical documents has a great potential for the reconstruction of floods in the past and can be informatively helpful for contemporary flood risk management. Therefore, a lot of research activities in the field of historical hydrology have been carried out in the past years, e.g. in the RIMAX research programme or the EU funded project "Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation" (SPHERE). A review of scientific methods using historical data for improving flood risk estimation is given by Benito et al. (2004). Furthermore, Kundzewicz & Brázdil (2006) published a special issue on the field of historical hydrology.

One approach to reconstruct historical flood events, which was used in this case study, is based on the analysis of a flood triggering meteorological situation on a regional scale. This approach is reliant on information from historical data, such as meteorological and hydrological measurements, to reconstruct past flood events, to identify their hydrometeorological causes and to quantify the areal precipitation in a specific river catchment. The reconstruction of the meteorological situation allows the analysis of a river provided that a run-off model for it is available to implement meteorological data. Restricting such events to rainfall makes the flood reconstruction independent from river profiles and discharge values. This avoids also uncertainties arising from the instationarities of the river characteristics caused by river morphology changes during past centuries. One of the first steps to analyse the 1824 and 1882 flood events was to find the mostly entire material in archives and libraries as well as in private estates.

3.1 Information from regional historical meteorological data

For the analysis of the 1824 and 1882 flood events in south west Germany, a variety of different data sets were consulted. A detailed overview of the collected and analysed data was given by Seidel et al. (2009), Bürger et al. (2006), Sudhaus et al. (2008), Dostal et al. (2007) and Straub (2007). The main difference between the datasets of 1824 and 1882 is the different quantity and quality of the meteorological measurements. In 1824, most meteorological observations were locally conducted by scientists or other people, but these measurements were not necessarily linked or standardized. However, one local meteorological network was already founded in south west Germany by 1824. These meteorological observations were the result of the first standardized measurements in south west Germany and already used in contemporary scientific publications related to the flood catastrophe in 1824. Schübler has published the daily precipitation amounts from five meteorological stations for the extreme flood event at the end of October and the beginning of November 1824 (Schübler, 1825). These data are still only fractional or unpublished and were firstly used by Bürger et al. (2006). The location of the sites with meteorological measurements in 1824 is depicted in Figure 2 .

In 1882, the data source is fundamentally different. During the second half of the 19th century, meteorological observation networks were founded and the data were regularly exchanged and published in newspapers or other chronicles. The first meteorological congress in 1875 led to the establishment of standards, and the same time meteorological data were regularly published by various national weather observation agencies, e.g. the Meteorologische Beobachtungen in Deutschland (Germany), Annalen der Schweizerischen Meteorologischen Zentralanstalt (Switzerland) and monthly reports of the forest meteorological Stations in Alsace-Lorraine (now France). For south west Germany and the adjacent regions in northern Switzerland and eastern France, daily observations from 46 stations were available for the reconstruction of the meteorological conditions in 1882. The historical meteorological data for 1824 and 1882 usually contain three daily measurements of air pressure, air temperature, relative humidity, wind direction and strength and degree of cloudiness. Mean daily temperatures (T$\bar{x}$) were derived according to equation 1

$$T\bar{x} = \frac{(T_7 + T_{14} + 2 \cdot T_{21})}{4} \tag{1}$$

where T_7 is the temperature reading at 7 a.m., T_{14} ist the reading at 2 p.m. and T_{21} is the reading at 9 p.m. This formula was commonly used at meteorological stations in Germany

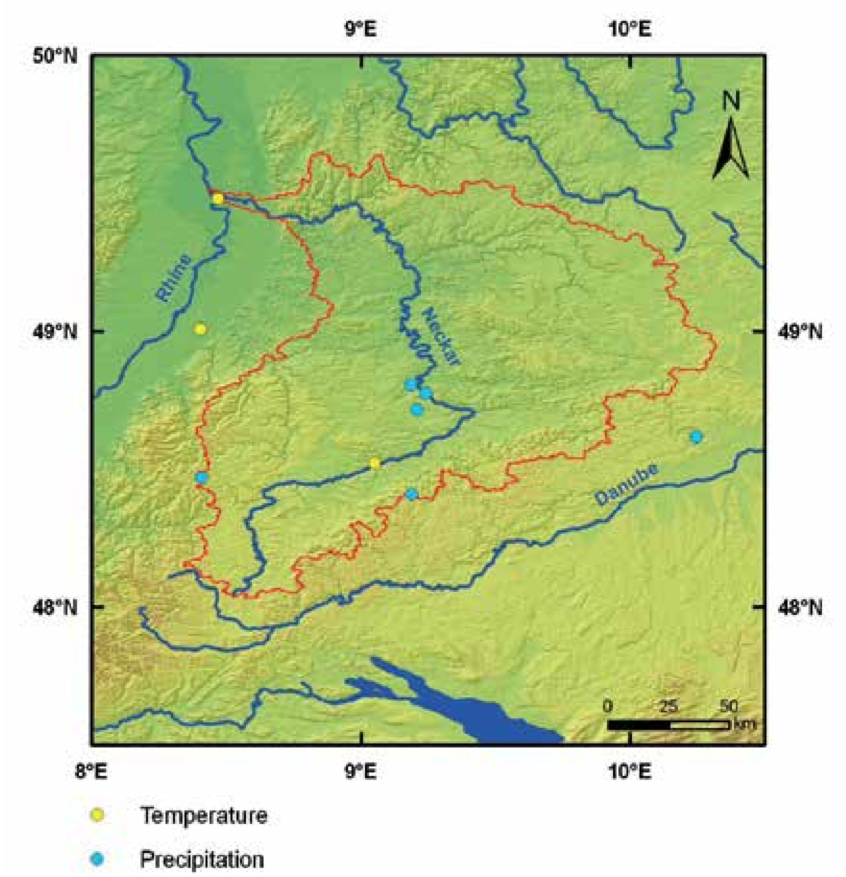


Fig. 2. Locations with meteorological observations in 1824

before hourly automatic readings were established. Additional information was often supplied by weather symbols indicating the occurrence and duration of phenomena such as precipitation, fog, etc. Beside air pressure and air temperature, quantitative precipitation values were also available. Daily precipitation amounts for various sites from December 25 to 28 1882 were published by Honsell and Tein (1891). Daily precipitation amounts for the complete year 1882 were also available for 26 stations in the Alsace-Lorraine Region from the French Weather Service Méteo-France. All in all, daily precipitation amounts during the flood event of December 1882 were available for more than 100 stations in southern Germany and adjacent areas in northern Switzerland and eastern France. For the interpolation of daily areal precipitation, only the data from stations within or in the vicinity of the Neckar catchment were used (c.f. Figure 3).

Additionally to the observed meteorological data, contemporary newspaper reports and documentary sources were also evaluated for both flood events. These documents revealed various important and detailed information like local weather conditions before and during the flood events and height of water levels at different locations along the rivers and damage reports.

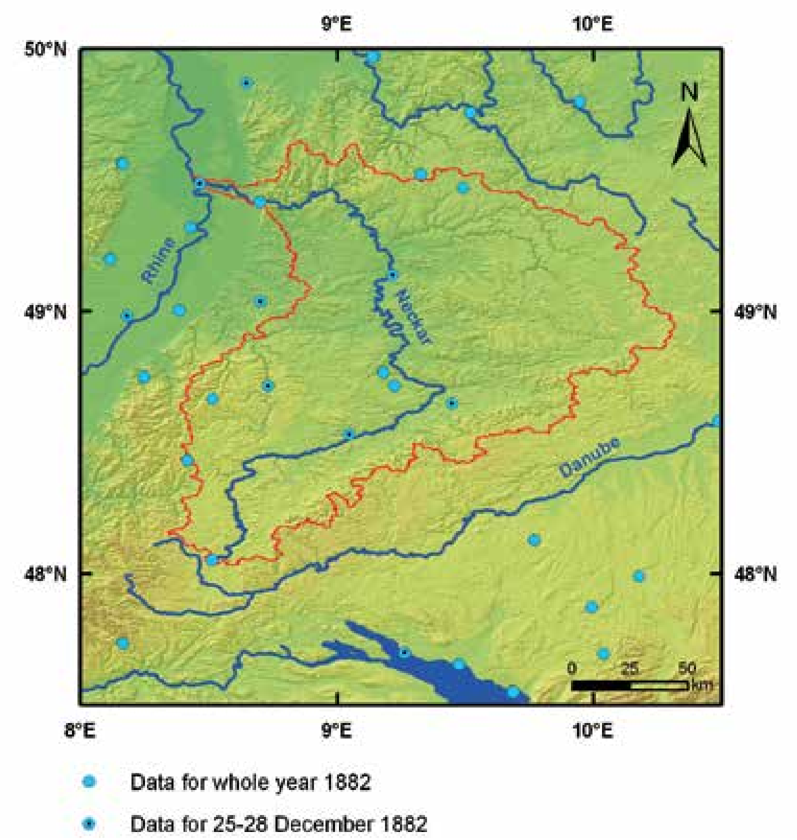


Fig. 3. Locations with meteorological observations in 1882

3.2 Information from large scale meteorological data

The atmospheric circulation pattern before and during the both flood events was derived from two different datasets. A historical meteorological dataset with climate data distributed over Europe (Barriendos et al., 2003) was used for determining the circulation patterns for the 1824 flood event. For the second flood event in 1882, the EMULATE data set (Ansell et al., 2006) was used, which comprises daily SLP-Grids with a 5° resolution for Europe and the North Atlantic.

4. Regionalisation of historical data

4.1 Discharge simulation: prerequisites and assumptions

The discharges for the 1824 and 1882 flood events were simulated with the water balance model LARSIM (Ludwig & Bremicker, 2006). Due to the fact that the historical station network was less dense than the station density currently implemented in the LARSIM model, it was necessary to create daily temperature and precipitation grids for the study area. The procedures on how these grids were derived are described in the following subsections. From these gridded data sets, the corresponding meteorological parameters (temperature, precipitation) were derived for the station sites used in the LARSIM run-off model.

The LARSIM model is based on a 1 km grid and incorporates different meteorological parameters (e.g. air temperature, precipitation, air pressure, wind direction, wind velocity) as well as 16 land-use classes and current river profiles. Due to the circumstance that only a few of these parameters could be reconstructed on the basis of historical data, some simplifications regarding the model parametrisation were made. The land use was disregarded because several studies (e.g. Bronstert et al. (2003); Haag et al. (2005); Ott & Uhlenbrook (2004)) have shown that this factor had only a small influence on discharge values during advective hydrometeorological extreme events. Therefore, the historical land use from 1824 and 1882 was not considered for the LARSIM simulation. Some hydrometeorological parameters such as air humidity and evapotranspiration were not considered either because of their minor influence on discharges during such an extreme event. For other parameters like wind speed and sunshine duration, there was no or only fragmentary information available from the historical meteorological observations. Therefore, the LARSIM simulation was carried out with a parametrisation for the year 1993 and only the meteorological parameters temperature and precipitation were modified accordingly with the historical values from 1824 and 1882. The flood event of 1882 was caused by rain and snow melt. The incorporated snow module in LARSIM, the threshold for air temperature below which precipitation falls as snow, was set to +1.5°C (daily mean value).

4.2 Interpolation of historical data for the 1824 flood event

The most important parameter for the simulation of discharges with hydrologicals models is precipitation. One of the major difficulties in this regard is the determination of the areal precipitation. As the meteorological observations network of the 19$^{\text{th}}$ century was by far not as dense as today, a straightforward interpolation of precipitation from the historical data would not incorporate the high spatial variability of this parameter. Therefore, similar weather patterns from modern data were analysed and compared with the historical situation during the floods of 1824 and 1882.

The idea is to find best matches at stations with historical and modern observation and to use the spatial pattern of the dense modern data as additional information for the interpolation of the historical precipitation data. The data sets were taken from German Weather Service (DWD) for the years 1934-2006. Due to the different data quality and quantity from 1824 and 1882, two different approaches were used for finding similar precipitation patterns and interpolation of the data.

For the case of the 1824 flood event, the best match regarding the precipitation pattern could be determined for 27 and 28 October 1998. This precipitation pattern correlates highly with the meteorological conditions of 28-29 October 1824. The weather course of October 1998 also caused heavy precipitation and floods in the Neckar catchment (LfU, 2000).

Assuming that the regional distributions of precipitation in October 1824 and 1998 have a high similarity, the regional precipitation pattern in the Neckar catchment for the flood event of 1824 was modelled in several steps (Fig. 4). Using the historical measurements and weather descriptions of 1824, the regional distribution for the strong precipitation of 28-29 October 1824 was determined by means of a regression model between seven historical precipitation measurements conducted in October 1824 by Schübler (1825) and measured precipitation values from 1998 at the same sites. The best result was achieved with a linear-logarithmic

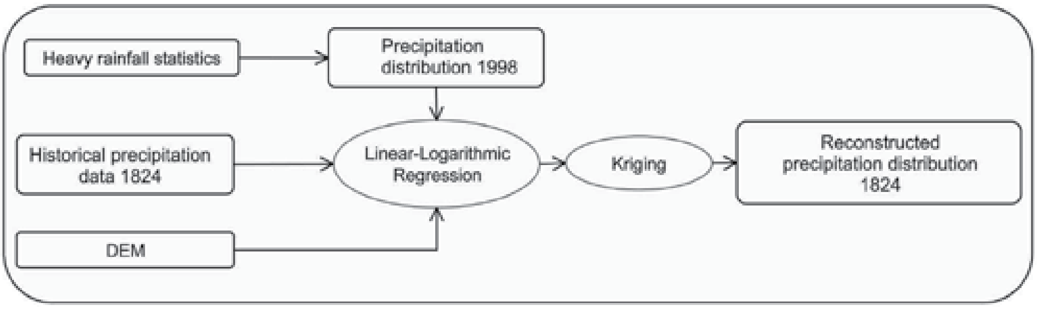

Heavy rainfall statistics
Precipitation distribution 1998
Historical precipitation data 1824
Linear-Logarithmic Regression
Kriging
Reconstructed precipitation distribution 1824
DEM

Fig. 4. Method for obtaining areal precipitation data for the 1824 flood event

regression using equation 2

$$N_{1824} = (1.5 + \frac{4}{ln(DEM)^4}) \cdot N_{1998} \quad (2)$$

where N_{1824} are the precipitation measurements in 1824, N_{1998} are the precipitation measurements at the same sites in 1998 and DEM is the elevation of these sites. This regression model yielded a R^2 of 0.88 and was used to modify the dense precipitation data for 1998 (i.e. 220 stations in the study area) to obtain enough values for the 1824 precipitation event for a reliable spatial interpolation. This was done using kriging interpolation to obtain a quantitative precipitation distribution (1km grid) for the 36h rainfall event of 28-29 October 1824.

4.3 Interpolation of historical data for the 1882 flood event

In comparison to the flood event of 1824, a dataset with daily temperature and precipitation measurements for the whole year of 1882 was available. Daily temperature grids from historical meteorological observations were derived by establishing a linear regression between altitude and daily temperature means for each station during the year 1882 on the basis of 22 reference stations in the study area. Afterwards, the digital elevation model (DEM) was multiplied in the form of

$$T_{(x,y)i} = a_i \cdot DEM_{(x,y)} + b_i \quad (3)$$

where $T_{(x,y)i}$ is the temperature at the DEM raster cell with x as latitude and y as longitude; a_i and b_i are the derived daily regression coefficients and i is the corresponding day of the year (DOY) in 1882. The reconstruction of a daily temperature grid was necessary because the 1882 flood event was caused by a combination of snow melt and precipitation and this had to be considered in the simulation of discharges.

Although the network of meteorological stations in south west Germany and the adjacent areas in France and Switzerland was relatively dense in the 1880s, the Neckar catchment had only data from eight stations from which daily precipitation and temperature readings were available. As a first step, a time series of daily precipitation patterns for the period 1 January to 24 December 1882 was created on the basis of the historical precipitation measurements, which was used as input data to simulate the discharge in the Neckar until the flood event. In order to capture the spatial variability of the flood triggering precipitation after 24 December 1824, a similar approach as in the the case of the 1824 flood event was applied. The aim was to find a comparable 3-day qualitative precipitation pattern for the period of 25-27 December 1882 but with a higher density of precipitation measurements. Five representative (i.e. spatially well distributed over the study area) stations were selected where daily precipitation data for 1882 was available. These stations are located in Freudenstadt, Villingen, Stuttgart, Buchen and Ansbach. A 5x3 matrix of precipitation data was created for the five stations and for the corresponding three days (25-27 December 1882). These data were normalized at the upper left value. In an analogical way, the time series of recent precipitation data for these five observation sites were compiled for the period 1958-2005, where data was available from the German Weather Service DWD for these stations. Each matrix of the moving window was normalized as described above and compared with the normalized historical matrix with a Kendall Rank Correlation test (c.f. Figure 5). With this non-parametric test, the degree of correlation of two samples, transferred to ranks, can be described. The Kendall Rank Correlation is calculated as

$$\tau = \frac{n_c - n_d}{n(n-1)/2} \tag{4}$$

where τ = Kendall's Rank Correlation value, n_{c} is the number of concordant pairs, n_{d} is the number of discordant pairs and n is the sample size. If Kendall's τ is 1, the agreement of the two rankings is perfect; τ = 0 means complete independence of the rankings.

The flow chart in Figure 6 gives an overview on how the input data for the run-off simulation were derived.

5. Results of the case study

5.1 Reconstruction of atmospheric circulation patterns and weather conditions in 1824

The circulation pattern for 26-30 October 1824 was identified as a cyclonic west situation (Wz), according to the weather classification of Hess & Brezowski (Gerstengabe & Werner, 1999). In October 1824, a low-pressure area over the British Isles, opposed to a high-pressure area over the Mediterranean, caused a strong pressure gradient that carried warm and humid air masses from the Atlantic into Central Europe, causing gales and thunderstorms with heavy precipitation over south west Germany (Fig. 7). The low-pressure system triggered heavy rainstorms in the northern and southern parts of the Black Forest.

5.2 Reconstruction of atmospheric circulation patterns and weather conditions in 1882

In the year 1882, large parts of western Germany were hit by two large flood events. The first one, which is not part of this study, occurred at the end of November 1882 and affected mainly the lower course of the River Rhine. This flood event was caused by extraordinarily high rainfall amounts in the November of 1882, a maximum being between 23-26 November 1882. During the second half of December 1882, the weather was characterized by a strong high pressure area over Russia and Eastern Europe, which led to calm weather conditions in the study area. This is also well documented by the meteorological observations in the region. In some parts of south west Germany, this led to an atmospheric inversion with sunshine and higher temperatures in the mountain ranges and fog and lower temperatures e.g. in the Upper Rhine Valley and the area around Lake Constance. From the 21 December 1882 onwards, the circulation changed towards a meridional pattern, which brought cold air masses from the Northern Atlantic Ocean into Central Europe. This led to a decrease in air temperature

	Freudenstadt (s_1)	Villingen (s_2)	Stuttgart (s_3)	Buchen (s_4)	Ansbach (s_5)
25 Dec (d_1)	35.0 mm	18.3 mm	7.1 mm	30.3 mm	0.2 mm
26 Dec (d_2)	55.9 mm	32.5 mm	10.5 mm	33.7 mm	21.4 mm
27 Dec (d_3)	74.8 mm	20.3 mm	5.3 mm	33.2 mm	16.1 mm

$$\begin{array}{cc} & \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\ \begin{array}{c} d_1 \\ d_2 \\ d_3 \end{array} & \left(\begin{array}{ccccc} 35.0 & 18.3 & 7.1 & 30.3 & 0.2 \\ 55.9 & 32.5 & 10.5 & 33.7 & 21.4 \\ 74.8 & 20.3 & 5.3 & 33.2 & 16.1 \end{array}\right) \end{array} \tag{1}$$

$$\begin{array}{cc} & \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\ \begin{array}{c} d_1 \\ d_2 \\ d_3 \end{array} & \left(\begin{array}{ccccc} 1.00 & 0.52 & 0.20 & 0.87 & 0.01 \\ 1.60 & 0.93 & 0.30 & 0.96 & 0.61 \\ 2.14 & 0.58 & 0.15 & 0.95 & 0.46 \end{array}\right) \end{array} \tag{2}$$

$$\begin{array}{cc} & \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\ \begin{array}{c} d_1 \\ d_2 \\ d_3 \end{array} & \left(\begin{array}{ccccc} 1.00 & 0.42 & 0.17 & 0.22 & 0.04 \\ 1.55 & 0.80 & 0.17 & 0.36 & 0.77 \\ 1.91 & 0.75 & 0.15 & 0.81 & 0.27 \end{array}\right) \end{array} \tag{3}$$

Fig. 5. Measured precipitation (25-27 December 1882) compared with modern precipitation data with the help of the Kendall Test

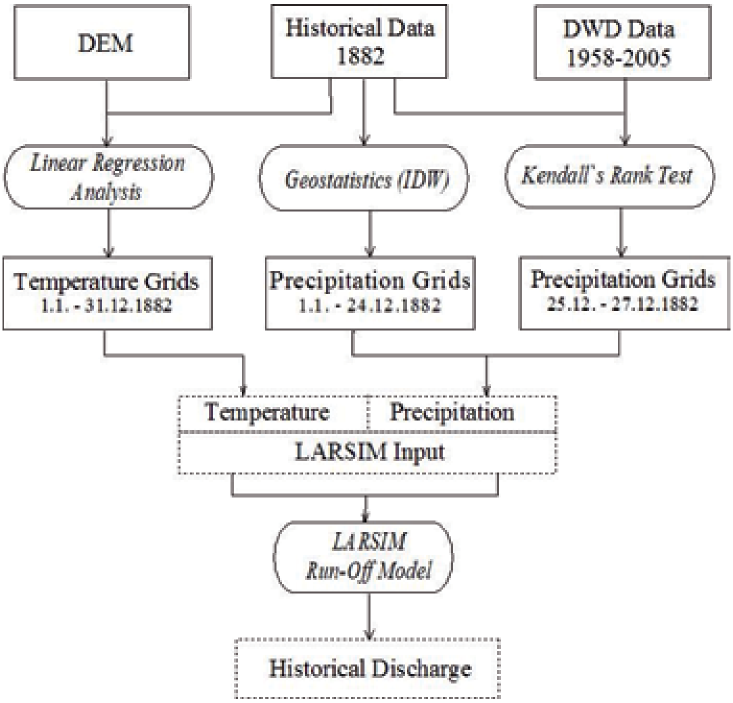


Fig. 6. Method for obtaining daily temperature and precipitation data for the 1882 flood event

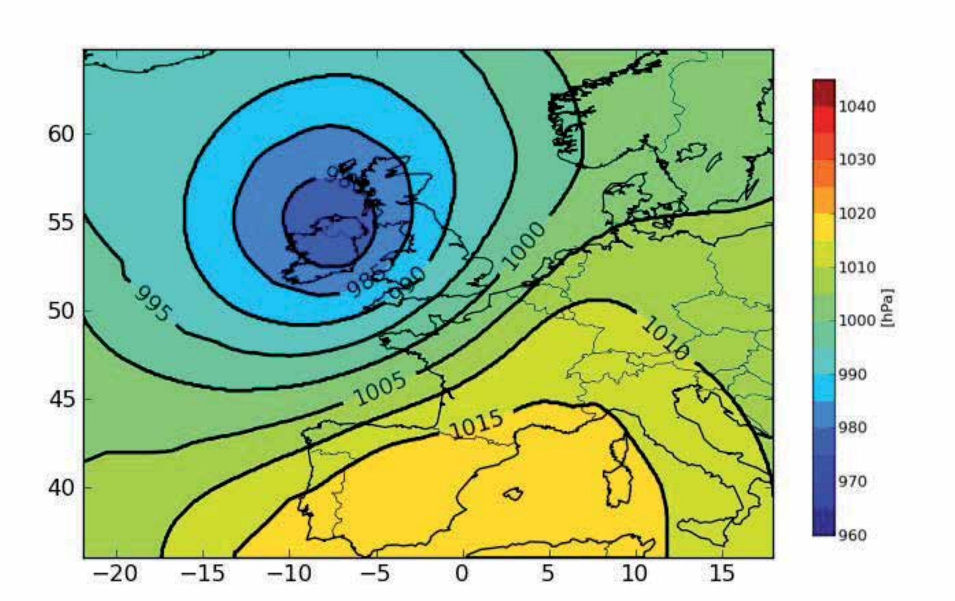


Fig. 7. Atmospheric circulation pattern over Central Europe from 26-28 October 1824

and heavy snowfall in most parts of the study area. According to the meteorological data and the qualitative data from historical newspapers, the onset of the snowfall in south west Germany was between December 22 and 23. From 25 December 1882 onwards, the meridional circulation changed towards a zonal westerly pattern which brought warmer air masses into Central Europe (Fig. 8). This led to a rapid and strong temperature increase and long lasting rainfall. This caused the snow that had fallen the days before to melt. The combination of rainfall and snow melt led to devastating floods, especially in tributaries of the Rhine River. In some parts of the Neckar catchment, the water levels were only exceeded by the extreme flood event of 1824 (Bürger et al., 2006).

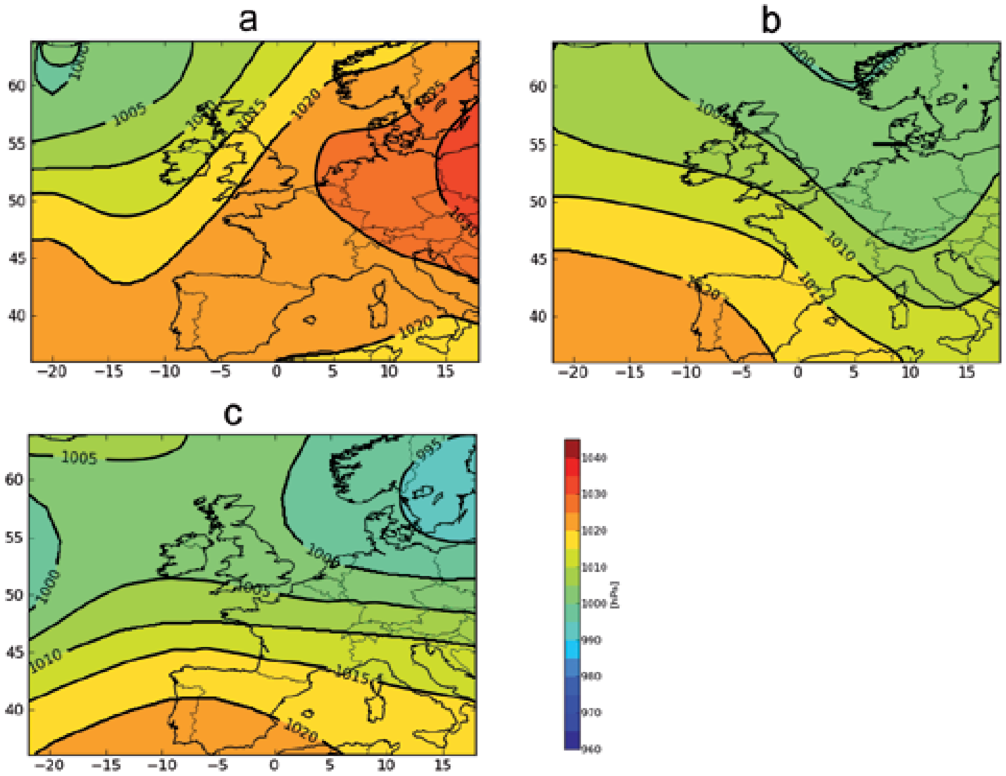


Fig. 8. Circulation patterns over Central Europe on (a) 20 December, (b) 23 December and (c) 26 December 1882

5.3 Reconstructed precipitation pattern for 1824

With the information of the historical data a distinct zonal weather condition with strong cyclonic characteristic could be detected for October 1824, which led to long lasting rainfall in Central Europe. For the detailed reconstruction of precipitation in the Neckar catchment for October 1824, the best corresponding weather situation for the years 1934 to 2004 was detected using the monthly weather forecasts of the German Weather Service (DWD) and the heavy precipitation statistics for southwest Germany (DWD, 2002).

The weather pattern for the period 27 and 28 October 1998 correlates highly with the meteorological conditions from 26 to 28 October 1824. Assuming that the spatial distribution of the precipitation in October 1824 and 1998 has a high similarity, the regional precipitation pattern in the Neckar catchment for the flood event of 1824 was modelled in several steps.

With the historical measurements and weather descriptions of 1824, the spatial distribution of the precipitation event from 28 to 30 October 1824 was determined using the derived regression model. For the spatial interpolation of the precipitation, the measurements from 220 weather stations of the German Weather Service were linked in a geostatistical model with a digital elevation model (DEM) using Kriging Interpolation. The calculated values were used for geostatistical modelling of the quantitative rainfall distribution (1 km grid) for the 36 hour rainfall event from 28 to 30 October 1824.

From the modelled precipitation pattern for the Neckar catchment and its neighbouring areas (Fig. 9), it can be seen that the highest precipitation with values up to 230 mm in 36 hours occurred in the western crest of the northern parts of the Black Forest. This is approximately twice as much as the average precipitation amount for October in the climate standard period 1961-1990 (DWD, 2002). The lowest precipitation amount was determined for the area in the Upper Neckar Valley. This corresponds very well to the descriptive weather observations from various historical records in this area. The upper areas of the Neckar catchment were not so severely affected by the flood in 1824 as the areas in the Black Forest nor in the northern and eastern part of the Neckar catchment. The rainfall distribution can be explained by the atmospheric circulation pattern and local orographic features. Southwesterly air flows are typical for the region. In particular, the northern parts of the Black Forest frequently receive relatively high precipitation during such weather conditions due to the orographic features found on the western side of the Upper Rhine Valley.

5.4 Reconstructed temperature and precipitation patterns for 1882

Using Kendall's Rank coefficient test, the best match for a similar precipitation pattern was found for the period between 29-31 January 1983, which yielded a τ of 0.86 (c.f. Fig 5). For this period, the linear regression between the historical data and the identified recent pattern resulted in $R^2 = 0.93$ and a regression coefficient of 2.16. The 1983 data comprises more than 400 precipitation measurements within the study area. Precipitation grids were interpolated from this data and then multiplied with the regression coefficient of 2.16 in order to adjust the precipitation amounts for the situation for 25-27 December 1882. A correction of +30 % was applied for all historical precipitation data from 20 December 1882 onwards. This was necessary in order to account for measurement errors during snowfall and strong wind (Rapp & Schönwiese, 1996). Information from contemporary newspaper reports states that the snow which had fallen from 22 December onwards had completely melted by the evening of 26 December. This information is also confirmed by the temperature data. Due to a change in the circulation pattern, warm air masses were transported to southern Germany from 26 December onwards. This led to a strong temperature increase in the study area of up to 10K in 24 hours. As a consequence, the snow which had fallen from 22 December onwards melted completely, even in higher elevations above 1000m in the Black Forest.

The total precipitation from December 25 to 27 1882 summed up to 180 mm in the Black Forest (c.f. Fig. 10). During these three days, the central parts of the Neckar catchment received relatively small precipitation amounts since this area is situated on the leeward side of the

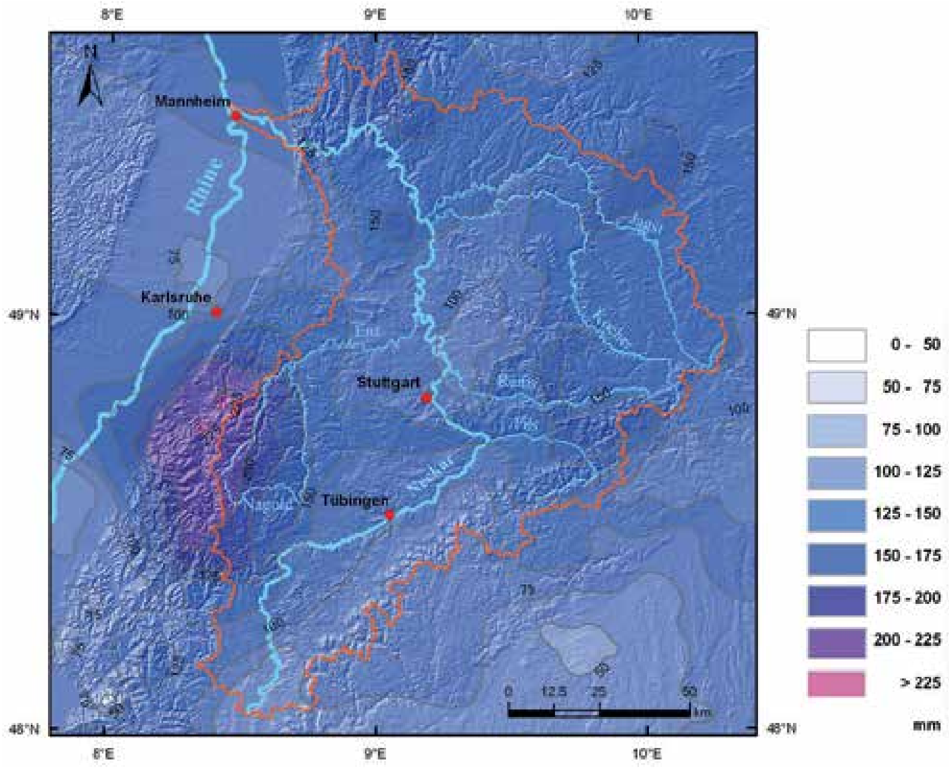


Fig. 9. Reconstructed areal precipitation for 28-29 October 1824

Black Forest and thus in the precipitation shadow during westerly circulation patterns. The combination of snow melt and the strong precipitation led to this extreme flood at the end of December 1882.

5.5 Simulated discharges for the 1824 and 1882 flood events

The reconstructed areal precipitation patterns were used as input for the water-balance model LARSIM to simulate the historical discharges. The simulated discharges along the course of Neckar during the flood events of 1824 and 1882 are shown in Figure 11 in comparison with the results from other sources like historical documents (Königliches Ministerium des Innern, 1896) and historical river cross profiles. These discharge values were derived from an analysis of water levels from historical river cross profiles carried out by Sudhaus et al. (2008) and were used to verify the results from the run-off modelling using the historical meteorological data. Furthermore, the values for the current design flood with a return time of 100 years (HQ_{100}) and the extreme flood scenario ($HQ_{extreme}$) also depicted (LfU, 2005).

The results for 1824 correspond well to the few discharge values which were found in historical documents. For example, in Bad Cannstatt (district of Stuttgart at river km 183), a discharge of 1320 $m^3 s^{-1}$ was recorded in a historical record for the 1824 flood. Local historical administrative reports recorded discharges between 4560 and 4800 $m^3 s^{-1}$ for Offenau (near Heilbronn at river km 98) and Heidelberg (river km 24) during the 1824 flood event (Sudhaus et al., 2008). The discharge values of 1824 in the Lower Neckar clearly exceed the values for

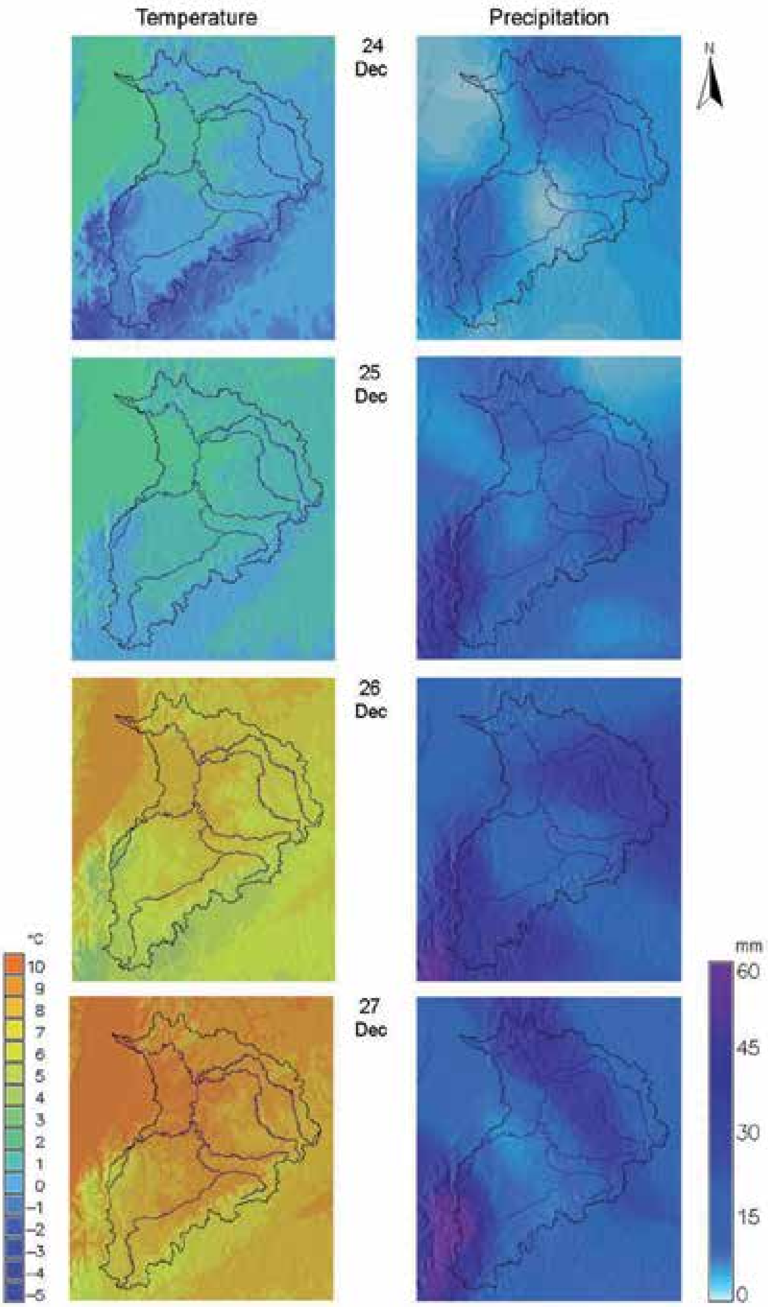


Fig. 10. Reconstructed temperature and areal precipitation for the flood event in December 1882

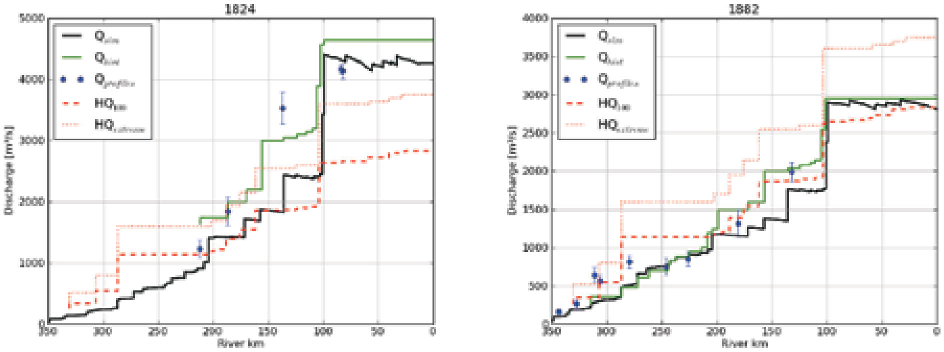


Fig. 11. Peak Discharges of the flood events in 1824 and 1882 along the Neckar River based on different data. Q_{sim}: Simulation with LARSIM-Model, Q_{hist}: Reported discharge values from historical sources, Q_{prof}: Discharge based on historical river cross profiles (Sudhaus et al., 2008), HQ_{100}: Design flood with 100 year return period, $HQ_{extreme}$: Extreme flood scenario.

HQ_{100} and $HQ_{extreme}$. A flood event with the dimension of 1824 would have devastating consequences for the riparian owners. The flood event of 1882 in the lower course of the Neckar River is comparable with a 100 year flood (c.f. fig. 11).

5.6 An example of using historical information for extreme value statistics

The usual practice for determining return times of floods is to calculate these values with the help of extreme value statistics. A major drawback of this method is that the underlying discharge data (e.g. annual peak discharges) often comprise only a few decades, which do not necessarily represent the flood characteristics of a river. Apart from this, the choice of a distribution function and the time span covered by discharge data can have a major influence on the results of extreme value calculations. In the following example, we show how information from historical flood events can be included in statistical extreme value analysis by using the maximum likelihood method and by making assumptions about the maximum annual peak discharges for the years in between the historical events. For the Plochingen gauging station situated in the middle course of the Neckar River, annual peak discharge data are available from the observation period 1921-2006. The highest observed value in this time span was a flood event in 1978 with a peak discharge of 1150 $m^3 s^{-1}$. This data series was extended by incorporating peak discharge values obtained from the reconstruction of the historical flood events in 1824 and 1882, which yielded 1650 $m^3 s^{-1}$ and 1200 $m^3 s^{-1}$, respectively. Furthermore, the historical sources revealed information about a flood event in 1851 at this site, for which the peak discharge could be narrowed down to a range between 1200 and 1500 $m^3 s^{-1}$. These results individually show that all three major historical flood events in the 19th century had higher peak discharge values than those of the observation period from 1921-2006. Furthermore, it can be concluded that for all other flood events in the time between 1820 and 1920, the annual peak discharge did not exceed the threshold value of 1190 $m^3 s^{-1}$ (otherwise such events would have been found in the historical records). Therefore, the annual peak discharge data can be extended back to the year 1800 by using the peak discharges from the three historical flood events (in case of the 1851 event also with the uncertainty range) and by assuming that a certain threshold level was not exceeded in all the other years in between (Fig. 12).

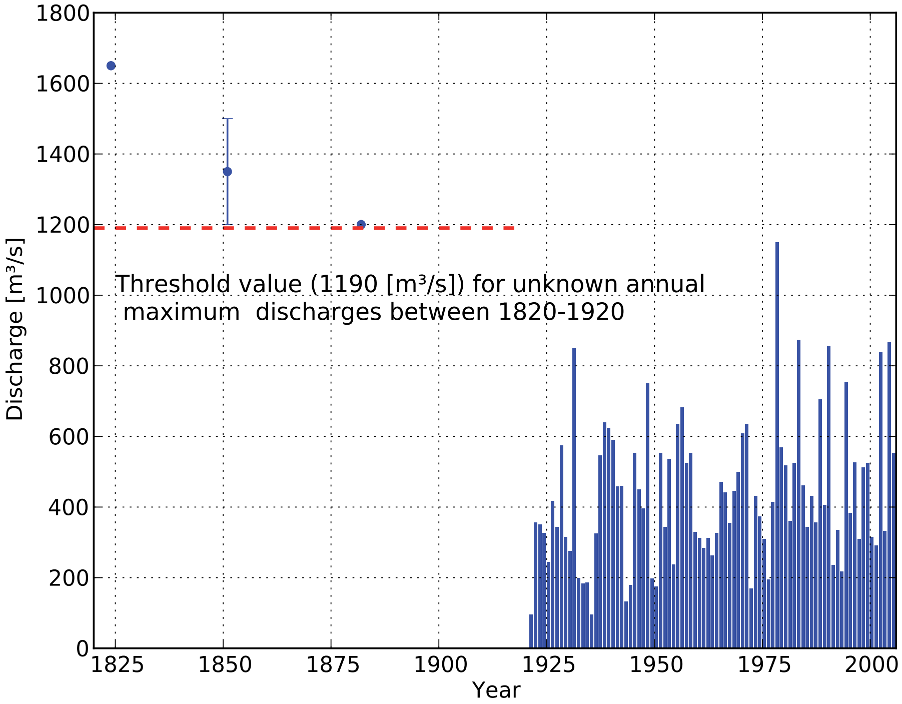


Fig. 12. Maximum Annual Discharges at the Plochingen Gauging Station (Neckar) for the Period 1820-2006

For calculation of the extreme value statistics based on the extended data series, the 2 parameter Gumbel-distribution was used. The location and scale parameters μ and β were estimated using the maximum likelihood method. This method has the advantage that data with an uncertainty range and threshold assumptions can be used

$$L(x,\mu,\beta) = \sum_{i}^{n} \ln f(x_i,\mu,\beta) + k\ln(F(T) + \ln(F(x_{max} - F(x_{min})) \rightarrow max \qquad (5)$$

where T is the threshold value (here 1190 $m^3\,s^{-1}$) which was not exceeded within k years and x_{max} and x_{min} are the upper and lower boundaries (i.e. the uncertainty range of the peak discharge) of a specific event. The changes in return times when incorporating the historical information at the Plochingen gauging station are shown in Table 1.

Data Basis	Peak Discharges [$m^3\,s^{-1}$]				
	HQ_{10}	HQ_{50}	HQ_{100}	HQ_{200}	HQ_{1000}
1921-2006	695	958	1069	1180	1436
1820-2006	744	1033	1156	1278	1561

Table 1. Peak annual discharges for return times at the gauging station Plochingen (Neckar) based on different data series

6. Discussion and conclusions

The results for these two historical flood events have shown that there is no universal tool or method for reconstructing past flood events. The presented research is a closer view on single extreme events. An approach for reconstructing a historical flood event is always a process of dealing with the specific meteorological and hydrological circumstances as well as specific contemporary data and other relevant information for the river catchment which is under investigation. Analysing historical flood events has to be evaluated against reliable data and reliable references.

The results presented in this study show the potential of historical data for the reconstruction of flood events. A detailed analysis of flood events is generally possible if certain historical meteorological and hydrological data are available, which is usually the case for the past 150 to 200 years. This applies also to flood events where the distribution of meteorological stations is not very dense and unfavourable. With the modern analogue approach, it was possible to convert the historical data into gridded data sets which can be used for run-off modelling.

One deficit of the presented study is the need for hydrometeorological data in order to conduct such run-off simulations. In our presented case study, the data gaps could be closed using modern analogues. A very successful approach without such data or with a much lower number of direct data input is shown in the work of Barriendos et al. (2003) and Lobanova (2002), for example. They used indirect data like clerical reports concerning extreme flood or weather events to assess the intensity of hydrological extremes. But this approach is attended with time-consuming research in historical archives and still has low acceptance among decision makers.

The analysis of the flood triggering hydrometeorological conditions also revealed some interesting information. The flood events of 1824 and 1882 were quite different regarding their causes, intensity and characteristics. The extreme rainfall event at the end of October 1824 occurred when the soils were already more or less saturated by the previous weather conditions and was caused only by rainfall. The flood event in 1882 occurred when a combination of temperature increase and high precipitation at the end of December lead to snow melt and subsequently to high discharges. This weather situation is quite common in southwest Germany and caused a severe flood in 1993 in the Neckar catchment, for example. Unlike other severe historical flood events in the Little Ice Age which occurred after extraordinary winters (e.g. in 1784 or 1845), the flood events of 1824 and 1882 occurred under conditions which can also be expected nowadays or in foreseeable climate change scenarios.

The results of this study have also raised some questions regarding future work. One potential field of research is the development of new methods for the interpolation of precipitation, especially when only few data are available. This could be achieved by taking the local orography and the wind direction during the rainfall event into account. A second perspective would be to reconstruct long discharge times series based on large scale data such as atmospheric pressure fields, which are available on a daily basis from 1850 onwards. For this purpose, a downscaling approach can be used by establishing a connection between large scale atmospheric circulation patterns and discharge increments in the river catchment, as it has been done by Bárdossy & Filiz (2005). With this method, phases of high and low flood frequencies can be detected which reveal information about the temporal and spatial flood variability in a certain river catchment. Furthermore, it would be interesting to investigate the transferability of the methods used in this case study to other river catchments in Central Europe.

7. Acknowledgements

The authors would like to thank the BMBF (German Federal Ministry of Education and Research) and the Federal State Ministry of Science, Research and Arts in Baden-Württemberg for funding this research work.

8. References

Ansell, T. J., Jones, P. D., Allan, R. J., Lister, D., Parker, D. E., Brunet, M., Moberg, A., Jacobeit, J., Brohan, P., Rayner, N. A., Aguilar, E., Alexandersson, H., Barriendos, M., Brandsma, T., Cox, N. J., Della-Marta, P. M., Drebs, A., Founda, D., Gerstengarbe, F., Hickey, K., Jónsson, T., Luterbacher, J., Nordli, Ø., Oesterle, H., Petrakis, M., Philipp, A., Rodwell, M. J., Saladie, O., Sigro, J., Slonosky, V., Srnec, L., Swail, V., García-Suárez, A. M., Tuomenvirta, H., Wang, X., Wanner, H., Werner, P., Wheeler, D., & Xoplaki, E. (2006). Daily Mean Sea Level Pressure Reconstructions for the European North Atlantic Region for the Period 1850-2003. *Journal of Climate* Vol. 19, No. 12, 2717-2742.

Bárdossy, A. & Filiz, F. (2005). Identification of flood producing atmospheric circulation patterns, *Journal of Hydrology*, Vol. 313, No. 1-2, 48-57.

Barriendos, M., Coeur, D., Lang, M., Llasat, M.C., Naulet, R., Lemaitre F. & Barrera, A. (2003). Stationarity analysis of historical flood series in France and Spain (14th-20th centuries). *Natural Hazards and Earth System Sciences* , Vol. 3, 583-592.

Benito, G., Lang, M., Barriendos, M., Llasat, M.C., Francés, F., Ouarda, T., Thorndycraft, V.R., Enzel, Y., Bárdossy, A., Coeur, D., & Bobée, B. (2004). Use of Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation. Review of Scientific Methods. *Natural Hazards*, Vol. 31, No. 3, 623-643.

Bronstert, A., Niehoff, D., and Fritsch, U.: Auswirkungen von Landnutzungsänderungen auf die Hochwasserentstehung. *Petermanns Geographische Mitteilungen*, Vol. 147, No. 6, 24-33.

Bürger K., Dostal P., Seidel J., Imbery F., Barriendos M., Mayer H., & Glaser R. (2006). Hydrometeorological reconstruction of the 1824 flood event in the Neckar River basin (southwest Germany), *Hydrological Sciences Journal*, Vol 51, No. 5, 864-877.

Centralbureau für Meteorologie und Hydrographie (ed.) (1889). *Die Wassermengen der fliessenden Gewässer im Großherzogthum Baden*. Beiträge zur Hydrographie des Großherzogthums Baden, 8. Heft, Berlin, Germany.

Dostal, P., Bürger, K., Seidel, J., Imbery, F., & Sudhaus, D. (2007). Lernen aus der Vergangenheit. Historische Hochwasseranalyse. Ein Beitrag für den heutigen Hochwasserschutz. *Berichte zur deutschen Landeskunde*, Vol. 81, No. 3, 233-245.

Deutscher Wetterdienst. (2002). *Deutsches Meteorologisches Jahrbuch 2002*. DWD, Offenbach.

Gerstengabe, F.-W. & Werner, P.C. (1999). *Katalog der Großwetterlagen Europas (1881 - 1998). Nach Paul Hess und Helmuth Brezowsky*. 5th edition, Potsdam & Offenbach.

Haag, I., Gerlinger, K., & Kolokotronis, V. (2005). Auswirkungen von Windwurfschäden auf Hochwasserabflüsse am Beispiel des Enz-Nagold-Gebiets. *Wasserwirtschaft* Vol. 10, 8-14.

Honsell, M. & Tein, M. (1891). Auftreten und Verlauf der Hochwasser von 1824, 1845, 1852, 1876 und 1882-83, In: *Ergebnisse der Untersuchung der Hochwasserverhältnisse im deutschen Rheingebiet. Auf Veranlassung der Reichskommission zur Untersuchung der Stromverhältnisse des Rheins und seiner wichtigsten Nebenflüsse und auf Grund der von den Wasserbaubehörden der Rheingebietsstaaten gelieferten Aufzeichnungen*, Centralbureau für Meteorologie und Hydrographie im Grossherzogthum Baden (ed.), Ernst, Berlin.

Königliches Ministerium des Innern (ed.) (1896). *Verwaltungsbericht der Königlichen Ministerialabteilung für den Strassen- und Wasserbau für die Rechnungsjahre vom 1. Februar 1893/94 und 1894/95*, Stuttgart.

Kundzewicz, Z.W. & Brázdil, R. (eds.) (2006). Special Issue: Historical Hydrology. *Hydrological Sciences Journal*, Vol. 51, No. 5.

Lobanova, M. (2002). Application of past information for reducing flood risk (the case of Lensk city), In: *Palaeofloods, Historical Data and Climatic Variability: Applications in Flood Risk Assessment*, Thorndycraft, V.R., Benito, G., Barriendos, M. & Llasat, M.C. (eds.), 231-236. CSIC, Madrid, Spain.

Ludwig, K. and Bremicker, M. (eds.) (2006). The Water Balance Model LARSIM - Design, Content and Applications. *Freiburger Schriften zur Hydrologie 22*, Freiburg, Germany.

LfU, Landesanstalt für Umweltschutz Baden-Württemberg. (2000). *Das Hochwasser vom Oktober/November 1998 in Baden-Württemberg. Oberirdische Gewässer/Gewässerökologie 65*, Karlsruhe, Germany.

LfU, Landesanstalt für Umweltschutz Baden-Württemberg. (2005). *Abflusskennwerte in Baden-Württemberg, Oberirdische Gewässer/Gewässerökologie 94*, Karlsruhe, Germany.

Middelmann, M. (ed.) (2007). *Natural Hazards in Australia: Identifying Risk analysis Requirements*, Geoscience Australia, Canberra.

Ott, B. & Uhlenbrook, S. (2004). Quantifying the impact of land-use changes at the event and seasonal time scale using a process-orientated catchment model. *Hydrology and Earth System Science*, Vol. 8, No. 1, 62-78.

Rapp, J. & Schönwiese, C.D. (1996). Niederschlag- und Temperaturtrends in Baden-Württemberg 1955-1994 und 1895-1994, In: *Wasser - Die elementare Ressource: Leitlinien einer nachhaltigen Nutzung*, Lehn, H., Steiner, M., & Mohr, H.(eds.), Springer, Berlin, Heidelberg, Germany, 113 - 170.

Seidel, J., Imbery, F., Dostal, P., Bürger, K., & Sudhaus, D. (2009). Potential of historical meteorological and hydrological data for the reconstruction of historical flood events - the example of the 1882 flood in Southwest Germany. *Natural Hazards and Earth System Sciences*, Vol .9, 175 - 183.

Schübler, G. (1825). Über die ungewöhnliche Überschwemmung zu Ende Octobers des vorherigen Jahres und die dabei in verschiedenen Gegenden Württembergs Gefallene Regenmenge. *Correspondenzblatt des Württembergischen Landwirthschaftlichen Vereins*, Vol. 7, 191-198.

Sudhaus, D., Seidel, J., Bürger, K., Dostal, P., Imbery, F., Mayer, H., Glaser, R., & Konold, W (2008). Discharges of Past Flood Events Based on Historical River Profiles. *Hydrology and Earth System Sciences*, Vol 12, 1201 - 1209.

Statistisch-Topographisches Bureau (ed.) (1883). *Württembergische Jahrbücher für Statistik und Landeskunde*, Stuttgart, Germany.

Straub, H. Historische Hochwasserinformationen und deren Nutzung. Arbeitskreis KLIWA (ed.), Kliwa-Berichte, Heft 10, 113-130, 2007.

Section 3

Information Management

13

Understanding Components of IT Risks and Enterprise Risk Management

Abdul Rahman Ahlan and Yusri Arshad
Department of Information Systems,
Kulliyyah of Information and Communication Technology
International Islamic University Malaysia,
Malaysia

1. Introduction

There is no doubt that information technology (IT) or information system (IS) improves the efficiency and efficacy of our daily lives. IT derives much of its usefulness from the ability to link systems together to improve functionality and communications (Ahlan, 2005). Inherent in these links are interdependence, interoperability and interconnectedness (O'Brien, 1996). Traditionally, IT is perceived to take the role of back-end support system to an organisation and thus, has little strategic value. Nowadays, this perception has changed primarily due to the potentials that pervasive IT can provide to all aspects of daily profitable organisations', communities' or individuals' efficiencies and efficacies and ultimately to achieve strategies and objectives. IT innovations facilitate all these ever increasing sophistication of IT users (Ahlan, 2005).

Nonetheless, the rapid adoption of IT poses organisations particularly to increasing and excruciating complex and sophisticated risks whether inherent or external. IT security, or risk, has been a highlight of every organisation since the inception of computer systems. Different organisations bear different sensitivity to particularly data and information risks and exposures to technical, organisational, project and human's risks (Wei *et al.*, 2010; Ahlan *et al.*, 2011). Manufacturing environment, for example, is less sensitive to information risk compared to healthcare and education sectors which in turn less sensitive compared to banking and finance sector. Universities data and information are highly sensitive and the risks are high. The more IT-laden organisations the more IT risks they are subjected to. Moreover, IT hardware, software and systems are becoming more sophisticated and expensive. Likewise, hackers or computer intruders and fraudsters are also becoming more sophisticated and constantly one step ahead of technology (Gerace and Cavusoglu, 2009). Hence, this puts pressure on manufacturers and service providers as well as IT managers to continuously increase the quality and security of their products and services.

Hence, this study aims to synthesise the risk factors associated with IT/IS and categorise or classify them into a few main major themes to guide the IT management in their risk management exercises. This chapter is organised into five main sections. First, the chapter begins with introduction to IT and risk in general. Second is the description of methodological approach, review of literature on and description of IT risk, factors and enterprise risk management. Third is the result and discussion of IT risk classification identified from the reviewed articles. Finally, the chapter ends with a brief description of future work.

2. Literature review

Extant literature shows that IT/IS has improved significantly compared to twenty years ago. As technology and systems become more complex and sophisticated, the risks associated to them are also increasingly growing and sometimes more difficult to detect. Different organisations bear different sensitivity to data and information risks and exposures to technical, organisational, project and human's risks (Wei *et al*., 2010; Ahlan *et al.,* 2011).

To ensure a systematic review of the state of the art literature, we follow the approach suggested by Webster and Watson (2002). In a first step, we searched the online database Proquest or ABI/INFORM, ScienceDirect, Emerald and the ACM Digital Library using the search terms "IT risk", "IT security" and "IT risk management" in the abstract, title and keywords. We had to limit to "information technology" in the search in order to reduce the number of articles found which are not relevant to "IT". The articles selected were published from 2001 to 2011. However, a review of the articles revealed that not many articles focused specifically on IT risk factors and enterprise risk management. In a second step we filtered the identified articles according to those in Association for Information System (AIS) journal rankings and book publications. In addition, we also included a few important relevant articles published earlier than 2001 and those from other field such as business, management, operation research journals and conference proceedings. Hence, we reviewed in total 46 relevant articles directly related to IT risks which are tabulated in Table 2 in Appendix. The summary of IT risks categories are tabulated in Table 1. Next sections briefly present review on IT/IS risks and management from literature. The sections are organized according to topics found in the literature.

2.1 Information system and technology roles and risk exposures

Information system plays an important role in any modern organisations to support its strategic, tactical and operational levels activities. These systems are at the core of the information management of the organisations and allow them to operate efficiently and maintain their competitive advantage. Three vital roles of IS, but not limited to, include (i) Support of business operations; (ii) Support of managerial decision making; and (iii) Support of strategic competitive advantage. According to (O'Brien 1996, p.7), ''if IS do not properly support the strategic objectives, business operations or management needs of an enterprise, they can seriously damage its prospects for survival and success''.

Recently, advances in IT have exposed the IT departments, infrastructures, functions and services to more threats from internal and external risks. These threats can be detrimental to not only technical aspects but also the data and information in the organisations which can be costly and even cause terminal loss or bankruptcy. Hence, recognising IT as a technology with the fastest rate of development and application in all branches of business, requires adequate protection to provide high security and quality products and services. The aim of the safety analysis applied on an IS is to identify and evaluate threats, vulnerabilities and safety characteristics. Moreover, IT assets are exposed to risk of damages or losses. In addition, IT/IS security also involves protecting information stored electronically. That protection implies data integrity, availability and confidentiality.

For example, numerous government reports in United States published over the last few years indicate that federal automated operations and electronic data are inadequately protected against information risks. These reports show that poor security program management is one of the major underlying problems (GAO, 1999). A principal challenge many agencies face is identifying and ranking the information security risks to their operations which is the first step in developing and managing an effective security program. Taking this step helps ensure that organisations identify the most significant risks and determines what actions are appropriate to mitigate them (Ahlan *et al.*, 2011).

In addition to information risks, most security incidents today are caused by flaws in software, called vulnerabilities. It is estimated that there are as many as 20 flaws per thousand lines of software code. Computer Emergency Response Team/Coordination Center (CERT/CC), United States statistics reveal that the number of vulnerabilities reported has increased dramatically over the years, from only 171 in 1995 to 8,064 in 2006. Along with vulnerabilities, the sophistication of attack tools has also advanced over time. Using the interconnected nature of the Internet and automated attack tools, attackers exploit software vulnerabilities at an alarming rate to cause serious damage to organisations (Gerace and Cavusoglu, 2009).

Nowadays, there are many types of computer crimes reported in the United States such as money theft (44%), damage of software (16%), theft of information (16%), alteration of data (12%), theft of services (10%) and trespass (2%) (Boran, 2003). This is also happening in other countries. Hence, in order to minimise losses, it is necessary to introduce risk management and risk assessment in the areas of IT and operational risks. The objective of IT/IS risk management is to protect IT/IS assets such as data, hardware, software, personnel and facilities from all external (e.g. natural disasters) and internal (e.g. technical failures, sabotage and unauthorised access) threats so that the costs of losses resulting from the realisation of such threats are minimised (Gottfried, 1989).

There are myriad dimensions to the complexity associated with protecting our interconnected IS from the technical, managerial, organisational, institutional, cultural, and international political perspectives. This reality makes it difficult to understand the complex interconnectedness of these IS (Longstaff *et al.*, 2000). Modelling and subsequently assessing and managing the risks that face these infrastructures are thus a formidable task. Each dimension is important and must be addressed. However, only when we analyse all the important aspects and perspectives in a complete vision can we make appreciative progress towards the infrastructures' protection and sustained operation. According to (Longstaff *et al.* 2000), we can broadly categorise the complexity of interconnected infrastructures as structural-based, which includes hardware, structures and facilities, and human-based, which includes institutions, organizations, culture and language. There is a dangerous disconnect among the professionals from the multiple disciplines that conceive, plan, design, construct, operate, maintain, and manage these complex infrastructures.

2.2 IT risk and management

IT risk management (RM) and risk assessment (RA) are the most important parts of Information Security Management (ISM). The important step in risk management cycle is risk identification which is to be done comprehensively and iteratively. This chapter, therefore, aims to synthesise the risk factors associated with IT and categorise or classify them into a few main major themes to guide the IT management in their risk management exercises.

2.2.1 IT risk definitions

Various fields such as IT, Engineering, Banking, Insurance, Economics, Management, Medicine and Operations Research have studied risk and risk management in their own domains. Nonetheless, each field addresses risk in a fashion relevant to its object of analysis and, hence, adopts a particular lens of viewpoint. Therefore, the authors will present here some of the risk definitions used in the different fields and relate them to IT risk used in this study.

- Generally, risk occurs in a situation when decisions are made knowing the probability of a risk event which shows that the decision maker has more information available than if he did not (Frame, 2003).
- Furthermore, risk, a measure of the probability and severity of adverse effects, is a quantitative entity and in order to manage it we must be able to quantify it. However, quantifying the efficacy of risk assessment and management for software and information assurance in a well-defined metric (one that others can apply, duplicate and compare) has proven difficult. We have made great progress in quantifying all kinds of risk but not in quantifying the true value of risk to information integrity or to infrastructure protection (Longstaff *et al.*, 2000). In other words, risk is also taken to be a negative outcome or event that has a known or estimated probability of occurrence based on experience or some theory (see, for example, Charette, 1991;Willcocks and Margetts, 1994).
- For example, medicine often focuses solely on the probability of a disease's occurrence (e.g., heart attack), since the negative consequence is death in many cases. It would be useless to focus on the consequence itself since it is irreversible. Odds of occurrence are the key element. Data is used to determine which factors can influence those probabilities (heredity, smoking habits, cholesterol level and others). In its definition of sentinel events (occurrence involving death or serious injury), the Joint Commission on the Accreditation of Healthcare Organisations uses "risk" as "the chance of serious adverse outcome" (Kobs, 1998 as cited by Longstaff *et al.*, 2000). Life insurance adopts this approach and uses mortality tables to estimate probabilities. In this context, a "good risk" will be a person with a low probability of dying within a given period (and hence, for the insurance company, a low probability of having to pay a compensation) and a "bad risk" would be a person with a high probability of dying within the period.
- Levin and Schneider (1997 as cited by Aubert *et al.*, 2005) define risks as "… events that, if they occur, represent a material threat to an entity's fortune" (p.38). Using this definition, risks are the multiple undesirable events that may occur. Applied in a management context, the "entity" would be the organisation. Given this perspective, risks can be managed using insurance, therefore compensating the entity if the event

occurs. They can also be managed using contingency planning, thus providing a path to follow if an undesirable event occurs. This definition of risk is analogous to the concept of risk as a possible reduction of utility discussed by (Arrow 1983).

- On the other hand, finance field adopts a different perspective of risk. They view risk as equated to the variance of the distribution of outcomes. The extent of the variability in results (whether positive or negative) is the measure of risk (Aubert *et al.*, 2005). Risk is defined here as the volatility of a portfolio's value (Levine, 2000). Risk management means arbitrating between risk and returns. For a given rate of return, managers will prefer lower volatility but would be likely to tolerate higher volatility if the expected return was thought to be superior. Portfolio managers therefore aim to build a portfolio that is on the efficient frontier, meaning it has the highest expected return for a given level of risk, and the lowest level of risk for a given expected return (Schirripa and Tecotzky, 2000).
- Other fields, such as casualty insurance, adopt a perspective of risk as expected loss. They define risk as the product of two functions: a loss function and a probability function (Aubert *et al.*, 2005). Car insurance is a good example. In the eventuality of an accident, there is a loss function that represents the extent of the damages to the car, which can range from very little damage to the total loss of the car. There is also a probability function that represents the odds that an incident will occur. The expected loss (risk) is the product of these two functions (Bowers *et al.*, 1986).
- Another important distinction in risk analysis is the notion of endogenous versus exogenous risk. Exogenous (or external) risks are risks over which we have no control and which are not affected by our actions. Earthquakes or hurricanes are good examples of exogenous risks. Although we have some control over the extent of damage by selecting construction standards, we have no control over the occurrence of such natural events. Endogenous (internal) risks, on the other hand, are risks that are dependent on our actions. A car accident is an example of risk where a strong portion is endogenous. While a driver has no control over other drivers (the exogenous portion), the probability of an accident is strongly influenced by the driver's behaviour and ability (endogenous). The driver also controls part of the loss function, by deciding to drive an expensive car or a cheap car. This could explain why there is always a deductible amount with car insurance, to ensure that the driver will behave in a way that will minimize the endogenous portion of the risk. By being made responsible for a portion of the damages, the driver is enticed to act with caution (Aubert *et al.*, 2005).

In IT/IS studies, risk has been heavily researched in the areas of software development (see, for example, Boehm 1991; Charette, 1991; Griffiths and Newman, 1996; Lyytinen *et al.*, 1998; Ropponen, 1999) and project management (as examples only see Keil, 1995; Morris, 1996; Willcocks and Griffiths, 1996).

Bahli and Rivard (2003) propose a scenario-based conceptualisation of the IT outsourcing (ITO) risk, wherein risk is defined as a quadruplet comprising a scenario, the likelihood of that scenario, its consequences and the risk mitigation mechanisms that can attenuate or help avoid the occurrence of a scenario. This definition draws on and extends a risk assessment framework that is widely used in engineering. The proposed conceptualisation of risk is then applied to the specific context of ITO using previous research on ITO as well as transaction cost and agency theory as a point of departure. Agency theory and transaction cost theory suggest four main risk scenarios that can be associated with outsourcing: (1) lock-in, (2) contractual amendments, (3) unexpected transition and management costs and (4) disputes and litigation. Resource based view theory identify risks on competences and capabilities of stakeholders while social exchange theory looks from service receiver-provider relationship exchange during ITO project arrangements (Arshad, 2011).

IT risks are perceived to culminate from the potentials that any undesirable events which can bring losses, threats to privacy and security of data and information and life of organisations and individuals. Raftery (1994) suggests that risk can be quantifiable, and proposes that risk is the actual outcome of an activity deviating from its estimate or forecast value. Risk may, therefore, be expressed as an exposure to economic loss and gain. As can be seen, the differences between risk and uncertainty events lie in the (in)ability to know their probability and to quantify their attributes.

In other words, IS has long been at some risk from malicious actions or inadvertent user errors and from natural and man-made disasters. In recent years, systems have become more susceptible to these threats because computers have become more interconnected and, thus, more interdependent and accessible to a larger number of individuals. In addition, the number of individuals with computer skills is increasing, and intrusion, or "hacking," techniques are becoming more widely known via the Internet and other media (GAO, 1999).

Let us consider a technology selection scenario. In a study, (Cochran 2006) suggests that when consumers are confronted with technology decisions, these technology attributes (interdependence, interoperability, and interconnectedness) must be considered. As the numbers and types of information technologies continue to multiply every year, selecting the "right" product is getting more difficult. Thus, for academics, for instance, trying to understand the factors motivating particular technology selection decisions, this becomes a significant yet complex issue.

Cochran (2006) asserts that there are three high level assessment areas in making technology decisions: "standalone" product assessment, technical compatibility assessment, and technology survivability assessment. This is shown in Figure 1. This is because practitioners making technology selection decisions cannot afford to make selection decisions based on the product alone. They must be concerned with whether the product will be compatible with or disrupt existing technologies already in place in the organisation. For example, the "best" technology according to its features and functionality may be extremely expensive to implement if it has incompatibilities. Decision makers must also worry about the survivability of the technology in the marketplace in order to avoid being "stranded" without support. An implemented technology could lose much of its value if the vendor folds or is acquired by another company. Furthermore, there are switching costs inherent in these technology decisions that must be considered. The model, however, does not focus on IT risk criteria or risk theories.

Furthermore, (Cochran 2006) differentiates between technical- and social compatibility. Technical compatibility refers to the capability of multiple products to work together. For example, "will this software package operate on the computer systems we have?" Social compatibility refers to "the degree which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters." For example, "will this software alter the way that the organisation orders supplies?" or "will this software be compatible with the existing knowledge of the end user?"

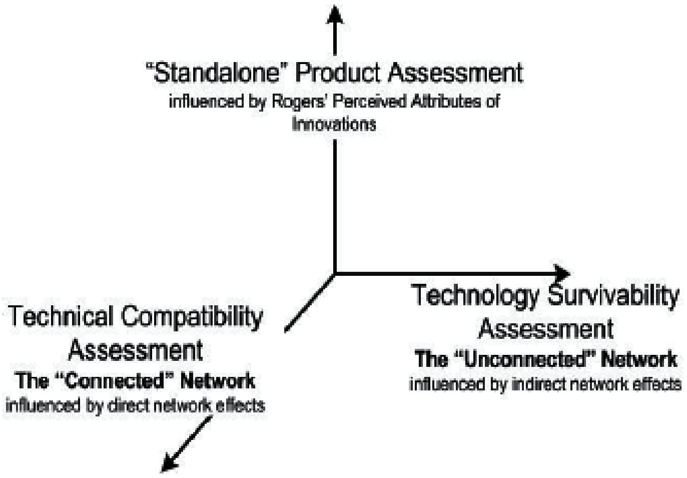

"Standalone" Product Assessment
influenced by Rogers' Perceived Attributes of Innovations
Technical Compatibility Assessment
The "Connected" Network
influenced by direct network effects
Technology Survivability Assessment
The "Unconnected" Network
influenced by indirect network effects

Fig. 1. Technology Evaluation Axis

Based on open-ended interviews of six Chief Information Officers (CIOs), project managers and similar positions, the informants have already indicated the complexities that these decisions entail, as well as the areas that are most difficult to assess. Notably, they have stressed the difficulty in assessing the full impact of the compatibility issue, as well as the difficulty in predicting the future of technologies. Furthermore, they discussed the consequential effects that previous infrastructure decisions can have on current and future decisions.

Thus, without thorough understanding of the factors that must be considered, outcomes of the selection decisions are more uncertain. Once the factors are understood, strategies for better assessment and mitigating risk can be developed. The following section describes the risk management application in IT field decision making process.

2.2.2 IT Risk management decision making

Decision-making takes place in an environment which has three components – certainty, uncertainty and risk (Flanagan and Norman, 1993). While certainty can be thought of as a situation in which all the factors causing a possible event can be exactly specified and known by a decision-maker, uncertainty entails the exact opposite, making an uncertain situation impossible to describe in terms of its probability of occurrence.

Risk management tools take into account whether risk is endogenous or exogenous. In finance, for example, risk is considered exogenous. The methods used to manage risk are concerned with diversification, insurance and allocation of assets. There is no direct action that managers can take to reduce the probability of a given event. In engineering or medicine, a portion of the risk is always endogenous. Risk management takes this into account. Patients are informed of the portion they control and are proposed healthier diets and lifestyles; employees are provided with security guidelines and actions are taken to reduce directly the probability of undesirable consequences. In IT field, generally, risk management involves analysis or risk identification, planning, implementation, control and monitoring of implemented measurements. Risk Assessment, as part of Risk Management, consists of several processes: (1) Risk identification; (2) Relevant risk analysis; and (3) Risk evaluation. In addition, (Rosman 2008) asserts that the four important aspects of risk management processes include: (1) understanding risk and risk management; (2) risk identification; (3) risk analysis and assessment; and (4) risk monitoring.

Risk management recognises risk, accesses risk and takes measures to reduce risk, as well as measures for risk maintenance on an acceptable level. The main aim of risk assessment, however, is to make a decision whether a system is acceptable and which measures would provide its acceptability. For every organisation using IT in its business processes, it is important to conduct the risk assessment exercise. Numerous threats and vulnerabilities are presented and their identification, analysis, and evaluation enable evaluation of risk impact, and proposing of suitable measures and controls for its mitigation on the acceptable level (Nikolić and Ružić-Dimitrijević, 2009).

In the process of risk identification, its sources are distinguished by a certain event or incident. In that process, the knowledge about the organisation, both internal and external, has an important role. Besides that, past experiences from this or a similar organisation about risk issues are also very useful. There are many techniques for identifying risks available such as checklists, experienced judgments, flowcharts, brainstorming, Hazard and Operability studies, scenario analysis and others (Nikolić and Ružić-Dimitrijević, 2009). In order to assess the level of risks, likelihood and the impact of incidental occurrences could be estimated. This estimation can be based on experience, standards, experiments, expert advice and others. Since every event has various and probably multiple consequences, the level of risk is calculated as a combination of likelihood and impact. Risk analysis or assessment can be either or a mix of quantitative, semi-quantitative or qualitative approaches (Macdonald, 2004).

There are numerous methods applied in risk assessment. In different countries, there are different methods. Even in the same area, there are various methods and applying each depends on a particular occasion. However, the methodology is similar that is system characterisation and description, threat and vulnerability identification, risk assessment, recommended measures and others. The differences in methods are due to the level of development of methodology items. All methods should present common descriptions of threats, vulnerabilities, assets groups and finally, a classification of risks. In that way, they can be compared and in order to achieve the best results, it is useful to apply the combination and optimization of methods. ISO standards for IT security (13335, 17799, and 27001) are general guidelines for implementing the IT security management process but there are no solutions provided on how to conduct it specifically (Nikolić and Ružić-Dimitrijević, 2009). In addition, Sarbanes Oxley (SOX) also requires organisations to assess their IT compliance for reporting purposes. COSO and COBIT are commonly used IT control assessment guidelines in organisations nowadays. Solms (2005) suggest that COBIT (2000) and ISO 17799 (ISO/IEC 17799, 2000) frameworks are complementary and, therefore, are actually very good choices as reference frameworks for Information Security governance. Used together, they provide a synergy which can be very beneficial to organisations.

Thus, implementing a proper risk management approach or technique to manage risks are necessary in today's organisations. The process of risk management is usually divided into risk identification, risk analysis, risk response planning and risk monitoring and control (Hillson, 2002). These steps are sometimes iterative and not always taken in sequence. Generally, it is necessary to express these steps in terms of activities and methods undertaken in the organisations. Once these activities are identified, it is then possible to assess the risk management practices implemented.

The effective management of risk lies in understanding the probability of a risk occurring, and if it does occur, how severe the adverse effect of the risk is likely to be. Between these two domains, risk may therefore be mitigated, accepted, avoided or transferred. In the context of construction, risks may affect cost, quality, safety, environment and time, among others.

External, or global risk, is the risk that falls outside an organisation's control because they arise outside the realm of the organisation's operations (Frame, 2003). Although external risks arise from sources that are different from internal risks, the same risk management principles can be applied to manage them. The management decision pertaining to risks would be dependent on the severity and probability of each particular case of risk. In this context, some risks may be extremely severe if these occur but the probability of their occurrence could be very remote. Consequently, risks may be mitigated, accepted, avoided or transferred as the case may be. In some instances, all aspects of a risk management framework may apply; while in other instances, only selected risk management principles within a framework would suffice. For this reason, it is not possible to tabulate responses to risk management because the spectrum of risks encountered in real life is too diverse and wide ranging to make any tabulation meaningful and succinct. Risk management decisions should therefore be determined on the facts and circumstances of each particular case.

The Project Risk Analysis and Management Guide (PRAM) compiled by the members of the Special Interest Group on Risk Management (APM, 2007) states that implementing a risk management system helps the formulation of more realistic plans in terms of both cost and time estimates. An increased understanding of the risks that might occur and their possible impact which can lead to the minimization of such risks and/or the allocation of these risks to the party best able to handle them is also possible. In addition, an independent view of the risks which can help to justify the decisions and enable the more efficient and effective management of risks are facilitated. Finally, a contribution to the building up of statistical data of historical risks that will assist in such future operations and the facilitation of greater but more rational risk taking and thus increasing the benefits that can be gained from doing so. Sadgrove (1996) adds that risk management helps a company avoid additional costs and disruptions to their operations and identify the risks that are worth pursuing and those that should be shunned. External risk management is especially important also because the firm's operations are now exposed to a dynamic environment influenced by macro-economic, political and social factors.

In any organisations nowadays, IT risk management is enforced at different stage of criticality. In medium to large organisations, enterprise risk management is normally practised in order to mitigate the organisational exposures related to IT risks. This is further explained in the following section.

2.3 Enterprise risk management

The earlier sections elaborated on the importance and steps of IT risk assessment and management in organisations. IT risks are avoidable and unavoidable and therefore, must be managed to minimise the risks. In any organisations, this is known as enterprise risk management (ERM). According to COSO (2004), it is:

- A process, on-going and flowing through an entity;
- Effected by people at every level of an organisation;
- Applied in strategy setting;
- Applied across the enterprise, at every level and unit, and includes taking an entity-level portfolio view of risk;
- Designed to identify potential events that, if they occur, will affect the entity and to manage risk within its risk appetite;
- Able to provide reasonable assurance to an entity's management and board of directors; and
- Geared towards achievement of objectives in one or more separate but overlapping categories

The underlying premise of ERM is that every entity exists to provide value for its stakeholders. All entities face uncertainty and the challenge for management is to determine how much uncertainty to accept as it strives to grow stakeholder value. Uncertainty presents both risk and opportunity with the potential to erode or enhance value. ERM enables management to effectively deal with uncertainty and associated risk and opportunity and thereby enhancing the capacity to build value (COSO, 2004).

Furthermore, according to COSO (2004, p.1), ERM encompasses:

- Aligning risk appetite and strategy – Management considers the entity's risk appetite in evaluating strategic alternatives, setting related objectives, and developing mechanisms to manage related risks.
- Enhancing risk response decisions – Enterprise risk management provides the rigor to identify and select among alternative risk responses – risk avoidance, reduction, sharing, and acceptance.
- Reducing operational surprises and losses – Entities gain enhanced capability to identify potential events and establish responses, reducing surprises and associated costs or losses.
- Identifying and managing multiple and cross-enterprise risks – Every enterprise faces a myriad of risks affecting different parts of the organization, and enterprise risk management facilitates effective response to the interrelated impacts, and integrated responses to multiple risks.
- Seizing opportunities – By considering a full range of potential events, management is positioned to identify and proactively realize opportunities.
- Improving deployment of capital – Obtaining robust risk information allows management to effectively assess overall capital needs and enhance capital allocation.

ERM deals with risks and opportunities affecting value creation or preservation. It is defined as a process effected by an entity's board of directors, management and other personnel and applied in strategy setting and across the enterprise, designed to identify potential events that may affect the entity, and manage risk to be within its risk appetite, to provide reasonable assurance regarding the achievement of entity objectives. Value is then maximised when management sets strategy and objectives to strike an optimal balance between growth and return goals and related risks and efficiently and effectively deploys resources in pursuit of the entity's objectives.

These capabilities inherent in ERM help management achieve the entity's performance and profitability targets and prevent loss of resources. ERM helps ensure effective reporting and compliance with laws and regulations and helps avoid damage to the entity's reputation and associated consequences. In summary, ERM helps an entity get to where it wants to go and avoid pitfalls and surprises along the way (COSO, 2004).

For example, risk management is performed at three levels within Department of Education and Training (DET) (NSW DEC, 2011). These include:

1. Strategic – this relates to risks associated with DET carrying out its business objectives as articulated in the DET Corporate Plan. These risks are identified, documented and managed in the organisation's business plans down to the business unit level (Regions and Directorates). Existing reporting systems are used to report achievement of objectives and management of identified risks.
2. Operational – this relates to the management of risks associated with the DET business units (Regions and Directorates) meeting their specific objectives. These risks are identified, documented and managed in the unit's operational plans. Existing reporting systems are used to report achievement of objectives and management of identified risks.
3. Specialist Areas – to support both Strategic and Operational risk management, DET has established specific policies, procedures and guidelines to ensure effective management of risks relating to:

- occupational health and safety
- child Protection
- serious incidents
- safety and security
- corruption prevention
- business continuity
- environmental management

Section 3 presents the result from the IT risk categorisation and elaborates on each risk categories and examples of situations which the risks might occur.

3. Results and discussion

From the literature analysis, we attempt to provide comprehensive IT risk factors into major IT risk categories. The findings suggest that IT risks generally originate from (I) technical or operational (hardware, software and systems); (II) data and information security; and (III) organisation, project, legal and human or people sides. This is further elaborated under each category in the following sections. Due to a large number of relevant literatures available, we only provide a non-exhaustive list of selected literature for the categorical risk example which is shown in Table 1 below.

Author(s)	Journal/Book (Year)	Risk types/issues	Categories
L.P. Willcocks, M.C. Lacity and T. Kern	Journal of Strategic Information Systems (1999)	Type and scope of outsourcing, vendor selection criteria and process, the role of the contract, retained capabilities and management processes, and partnering and relationship dimensions	II, III
Bunmi Cynthia Adeleye, Fenio Annansingh and Miguel Baptista Nunes	International Journal of Information Management (2004)	Strategic and operational risks which may have considerable financial and reputation costs.	I, III
Ward and Griffiths	Book (2001)	Not achieving the planned benefits, not meeting agreed deadlines, using more resources than initially foreseen, change in functional an procedural requirements, budget overrun and deficient change over of systems and problems associated with the operation and maintenance of these systems.	I, II, III
Kweku-Muata Osei-Bryson and Ojelanki K. Ngwenyama	European Journal of Operational Research (2006)	IS outsourcing contracts	I, II, III
Tafti, M	Industrial Management & Data Systems (2005)	Contracts, privacy and security, technical returns, loss of IT expertise, hidden costs and outsourcing decision process.	I, II, III
Kroenke	Book (2009)	Incorrect data modification, data disclosure and technological security	I, II, III
Rao	EDPACS (2004)	Technological security and legal/political issues	I, III
Ramanujan & Jane	Journal of American Academy of Business (2006)	Incorrect data modification, data disclosure and legal/political issues	II, III
Bouchaib Bahli and Suzanne Rivard	Omega (2005)	Transaction,client and supplier sources	II, III
Kakoli Bandyopadhyay, Peter P. Mykytyn and Kathleen Mykytyn	Management Decision (1999)	Application level, organizational level and interorganizational level.	I, III
Melinda Cline, Carl S. Guynes and Andrew Nyanoga	Journal of business and economic research (2010)	Environmental conditions and changes, organisational conditions and changes, managerial cognition, managerial actions, changes in the content of strategy and organisational outcomes.	III

Author(s)	Journal/Book (Year)	Risk types/issues	Categories
Fariborz Farahmand, Shamkant B. Navathe, Gunter P. Sharp and Philip H. Enslow	Information Technology & Management (2005)	Network system threats [(1)Threat agent: Environmental Factors, Authorized users and Unauthorized users and (2) Penetration technique: Physical, Personnel, Hardware, Software and Procedural].	I, III
Benoit A. Aubert, Michel Patry and Suzanne Rivard	The DATA BASE for Advances in Information Systems (2005)	Principal, agent and transaction categories.	II, III
Barki, H., Rivard, S., and Talbot, J.	Journal of Management Information Systems (1993)	Technological newness (need for new hardware, software), application size (project scope, number of users, team diversity), expertise (lack of development expertise, task of application-specific expertise, lack of user experience), application complexity (technical complexity, links to existing legacy systems), organizational environment (task complexity, extent of changes, resource insufficiency, and magnitude of potential loss).	I, II, III
Boehm, B.W.	IEEE Software (1991)	Software risk factors, including personnel shortfalls, unrealistic schedules and budgets, developing the wrong functions, developing the wrong user interface, "gold-plating," a continuing stream of changes in requirements, shortfalls in externally furnished components, shortfalls in externally performed tasks, performance shortfalls, and strained technical capabilities.	I, III
McFarlan, F.W.	Harvard Business Review (1981)	Dimensions of project risk based upon project size, experience with the technology, and project structure.	I, III
Keil, Mark., Cule, Paul E., Lyytinen, Kalle and Schmidt, Roy C.	Communications of the ACM (1998)	Four quadrants of risks including risks associated with customer mandate, scope and requirements, execution, and environment.	II, III
Thomas A. Longstaff., Clyde Chittister, Rich Pethia and Yacov Y. Haimes	Journal of Computer (2000)	Risk in systems integration: software development, temporal, leadership, environment, acquisition, quality and technology	I, II, III

Table 1. Risk types/factors (includes IT/IS outsourcing, investment, project management)[1]

[1] Note: For Summary of Risk Factors in Information Systems Projects (1983-1997), see Mary Sumner (2000), 'Risk Factors in Enterprise Wide Information Management Systems Projects'. Association of Computing Machinery (ACM).

3.1 Technical and operational risks

- Large IT risks originate from technical or operational risks in hardware, software and systems. In hardware, this can be in terms of faulty or defect products that can affect other hardware and systems within the same or networked environment. Even though manufacturing warranties do cover products defects after purchases, electrical short circuit in the hardware, for instance, could pose threats to other hardware, software and systems as well as data and information.
- Furthermore, the complexity of our technological organisation and society has forced us to deal with coupled and interconnected systems of systems whose likelihood of failure is ever increasing. The dominance of IT in our business and commerce has also created an almost critical-path dependency across our interconnected IS and critical infrastructures. For example, banking and finance institutions depend on the information infrastructure to operate their systems, reliable telecommunications depend on electricity and the electric utilities depend on a reliable source of energy. This networked systems and environments apply to most organisations nowadays even to small businesses with peer to peer or client-server and shared computers and peripherals.
- Therefore, computer security has become an important issue in this networked environment. The proliferation of personal computers, local area networks and distributed processing has drastically changed the way we manage and control information resources. Internal controls that were effective in the centralised, batch-oriented mainframe environment of yesteryears are inadequate in the distributed computing environment of today. Attacks on computer systems and networks are on the rise and the sophistication of these attacks continues to escalate to alarming levels. As more organizations share information electronically and autonomous computer networks work their way into our everyday lives, a common understanding of what is needed and expected in securing information technology resources is required.
- This is because the world of computers has changed dramatically over the decades. Twenty years ago, most computers were centralised and managed by data centers. Computers were kept in locked rooms and staffs of people made sure they were carefully managed and physically secured. However, in the computing world of today, autonomous network communications are setting the standards on how we interact with one another in a global environment. An effective security plan can successfully provide adequate safeguards to protect an organization's vital resources and assets.
- An ineffective security plan increases the economic costs associated with software vulnerabilities. It decreases the efficiency of an organisation and does not protect the resources and assets of the organisation. Inadequate protection of system resources compromises information obtained through email, research data and configuration data, services obtained via IS and applications and equipment such as computers and networking components. In addition, components vital to an organisation such as confidentiality, integrity, authenticity and availability are also compromised.
- Hence, an effective computer security plan protects an organisation's valuable resources, such as information, hardware and software. Furthermore, it also strengthens the aforementioned vital components of an organisation. Through the selection and application of appropriate safeguards, a security plan helps the organization's mission by protecting its physical and financial resources, reputation, legal position, employees

and other tangible and intangible assets. An effective security procedure reduces the economic costs associated with software vulnerabilities.

- For instance, the common threats to IS and computer networks can be classified into the Accidental, Intentional, Passive and Active categories. Accidental threats are losses due to malfunctions or errors. Some examples of accidental threats are power failures, hardware vulnerabilities in network switches, routers and other hardware components, software failures and natural threats such as fires and flooding.
- Intentional threats cause damage or corruption to computer assets. Sabotage is a type of intentional threat that uses small virus programs often propagated by unsuspecting users. Denial of Service (DoS) is another form of intentional threat that causes loss of availability of service. Some examples of DoS include e-mail spamming and network packet attacks aimed at host vulnerabilities.
- Passive threats do not change the state of the system. They may include loss of confidentiality but not the loss of integrity or availability. An example of a passive threat is traffic analysis, a form of eavesdropping in which an analysis of traffic patterns is used to infer information that is not explicit. Another instance of a passive threat is replay which is the repetition of valid messages in order to gain unauthorised access and masquerade as another entity.
- Unlike passive threats, active threats change the state of the system. These include changes to the data and software. Some examples of active threats are Trojan horses and trapdoor software, both of which alter parts of the system to allow unauthorised access. Security threats that are common today differ from those in earlier times. With worldwide Internet connections, anyone can gain access into an organisation's computer system from anywhere in the world and steal passwords although the building may be physically secured.
- Thus, even though physical security accomplished its objective in this scenario, the network is still not secure. Viruses and worms can be passed from machine to machine. Global autonomous networks provide an opportunity for "electronic thieves" to open windows and doors in the computer system's architecture. This "virtual thief" can detect and then exploit vulnerabilities in hundreds of machines in a matter of hours.

3.2 Data and information security risks

- In this information and knowledge era, organisational and individual data and information are available in digital forms. In many instances they are available on networked environement. Thus, they are susceptible to theft, misuse, abuse, modification, improper disclosure, fraud and others. It is, therefore, important that this risk is minimised in any organisation. One important method to curb this risk is through digital certificates and signatures whereby only certified authorised names are allowed to access any particular privileged authorised data and information. Moreover, most organisations nowadays also impose access level security controls on their networks and enterprise resource planning or other systems such as accounting, operations, human resource, marketing and management. Data administrator levels are also controlled between higher, middle and lower level staff. Nevertheless, sophisticated hackers, spyware and other sniffing tools are always on the lookout for data and information intrusions. Thus, IT managers must be constantly alert on any unusual logging activities in their organisations' systems and servers. Any irregularities must be reported and taken action immediately to avoid foreseeable losses due to data and information theft and intrusions either from inside or outside the organisations.

- Hence, organisations must follow some acceptable international standards and compliance regulations on IT risks and security controls such as ISO, COBIT, COSO and SOX. The purpose of any IT standard is, for example, to provide steps that employees must take to avoid inappropriate release of private and confidential organisational information. The focus of the standard is on the sensitive information that exists in a digital form, whether stored in a database, used in an application, transmitted over a network, or used in a report. Organisations and individuals information must be protected from any inappropriate sharing, releasing or use. When the information exists in a digital or electronic format, additional steps must be taken to ensure the protection of the information from loss, corruption, or inappropriate disclosure.
- Understanding the risks involved in handling information in digital form includes an appreciation of the greatly increased vulnerability made possible by technological conveniences that offer portability, easy copying, and wide—potentially global—distribution. The lack of reliable and current data often precludes precise determinations of which information security risks are the most significant and comparisons of which controls are the most cost-effective. Because of these limitations, it is important that organizations identify and employ methods that efficiently achieve the benefits of risk assessment while avoiding costly attempts to develop seemingly precise results that are of questionable reliability.
- Thus, all organisations and individuals information must be handled with appropriate security and access controls, and with attention to safeguarding confidentiality. No information should be exposed inappropriately. Many data elements and other types of information are protected by each country's current statute or regulation or in Malaysia case such as Malaysia legal acts, Malaysia Communication and Multimedia Commission (MCMC), university acts and others. Information that is not protected by law or regulation should, nonetheless, be protected against inappropriate exposure.

3.3 Organisation, project and human or people risks

- These types of risks originate from or within the organisations, projects and people. In an organisational environment, the policies, procedures, regulations, cultures and others, if not carefully designed, can pose risks to IT environment. Building security, access controls, electrical fittings, for example, can become sources of threats to IT hardware and software. Organisational type, vertical or hierachical, sizes, structures and building occupational health and safety implementation can result in different level of risks. In many organisations that create a proper IT division or department, the risks are minimised by the hands of professionally-trained staff. It is important for all staff to adhere to all IT security and controls policies and guidelines imposed by the management. Therefore, many small organisations are at risk of having their computer systems, hardware and software misused, abused, fraud, improperly installed and others.
- On project risk, the sources of risks can originate from any sources in the project cycles or processes. Project panel and stakeholders must carry out due diligence exercise on

feasibility of projects to reduce risks. IT project management consists of several important stages as stated in Project Management Body of Knowledge (PMBOK® Guide), which is the standard put forward by the Project Management Institute (PMI). They include Initiation, Planning, Execution/Managing and Closing. The guide lists nine elements of project management encompassing Project Integration Management, Scope Management, Time Management, Cost Management, Quality Management, Human Resources Management, Communications Planning, Risk Management and Procurement Management. This stages and processes also apply in any ITO projects.

- While ITO is associated with significant benefits, it can also be a risky endeavour. Researchers and practitioners also recognise that, in some circumstances, ITO entails risk, and that it sometimes leads to undesirable consequences that are the opposite of the expected benefits. In ITO projects, either onshore or offshore, more risks are posed depending on the nature of ITO projects themselves. Among the major risks are in selecting the right providers, win-win terms and conditions in contractual documents, access to organisational buildings and information privileges and project management service deliveries. In many ITO literatures, many projects failed due to poor project management, contract clauses and cultural differences in the case of offshoring. Relationship between service receiver and provider is also crucial for ITO success. ITO failure is not much attributed to technology but more on human competence, capabilities and relationship. Moreover, risk in systems integration, including software development, temporal, leadership, environment, acquisition, quality and technology, could become major sources of risks in IT and ITO projects (Arshad, 2011).
- Furthermore, since ITO projects involve relevant stakeholders within and outside the organisations, these human or people risks add more to inherent IT risks in the projects. Lack of commitments, understanding, competence and capabilities and communications, for example, can increase ITO project risks. In addition, staff within an organisation can also involve in stealing private and confidential information, hardware and software, improper usage, maltreatments, carelessness and other damages to IT hardare, software, systems and information. (See Sumner (2000) for further reading on IS project risk factors).
- Finally, IT risks originating from human or people could be attributed to human errors and misbehaviours. Competence and capabilities distinguish each staff in their work professionalism. As in ITO projects, human or service provider-receiver relationship is crucial for ITO project success. These can be found in many ITO literature. Human's attitude such as greedy, carelessness, selfish and others can increase IT risks in any organisations.

The previous section divides and elaborates IT/IS risks into three types: 1) Technical and operational risks; 2) Data and information security risks; and 3) Organisation, project and human or people risks. IT risk nature depends largely on types of assets or projects. Each IT hardware, software, system or project has its own inherent and incidental risk associated to it. Uncertainty presents both risk and opportunity, with the potential to erode or enhance value. Therefore, any organisation must undertake a risk assessment and management initiative to minimise risks that could result in big potential losses. Enterprise risk management enables management to effectively deal with uncertainty and associated risk and opportunity, enhancing the capacity to build value.

4. Future work and closure

Risk is a common terminology adopted in every field including IT/IS. While many definitions offered from different perspectives, IT risk in this study adopt the definition of IT risk being the uncertainty that a foreseeable loss or damage can result for such uncertain probabilistic events.

IT risk nature depends largely on types of assets or projects. Each IT hardware, software, system or project has its own inherent and incidental risk associated to it. This chapter classifies IT risk into three types, namely: 1) technical and operational risk; 2) data and information security risk; and 3) organisation, project and human risk.

Uncertainty presents both risk and opportunity, with the potential to erode or enhance value. Therefore, any organisation must undertake a risk assessment and management initiative to minimise risks that could result in big potential losses. Enterprise risk management enables management to effectively deal with uncertainty and associated risk and opportunity, enhancing the capacity to build value.

In short, aligning organisational strategy with IT strategy could help manage IT risks in the organisation. Organisational strategy must acknowledge the potential IT risks associated to all organisational assets that can increase its liabilities. Continuous IT risk assessment and management exercise, not only identify and minimise IT risks at an early stage and manage them, but also facilitate and help achieve the overall organisational short, medium or long-term goals and strategies.

Finally, organisations may opt to undertake ERM exercise for better control IT risks and security. The sooner, continuous and consistent applications of ERM can significantly minimise calculated risks and increase the profitability of organisations which in turn will be ploughed back into the organisations, staff, communities and stakeholders.

This study provides the theoretical foundation on IT risk components. While IT risk studies have been carried out based on a researcher’s theoretical framework, our next initiative is to perform multiple case study using mixed and multi method research in Malaysian organisations context to explore on the IT risks in practices. Another possible study is to perform comparative practices between developing and developed world organisational contexts.

5. Acknowledgment

This study is funded by Malaysia Ministry of Higher Education (MOHE) under Fundamental research grant scheme FRGS 10-029-0148. We would like to thank to International Islamic University Malaysia, Research Management Centre (RMC) for their kind assistance. Finally, our sincere gratitude goes to the organisations and respondents who contribute to the study directly or indirectly.

6. Appendix

Table 2 below lists the relevant articles reviewed in this study. While most of them represent articles found in the AIS journals, the authors, however, also include other relevant articles found in books and conference proceedings in order to enrich the sources for the literature review. The articles were published from 1991 to 2011.

Journal / Book / Proceedings	Year
Book - Course Technology, Cengage Learning	2011
Wireless Network	2011
ACM Computing Surveys	2011
Information Privacy & Security	2010
Business & Economics Research	2010
Proceeding of ACM New Security Paradigms Workshop (NSPW)	2010
IEEE Security and Privacy	2009
Consortium for Computing Sciences in Colleges	2009
Proceeding of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops	2009
Systems and Software	2008
Communications of the ACM	2008
Information security technical report	2008
The VLDB	2008
Consortium for Computing Sciences in Colleges	2008
Assocation of Computing Machinery	2006
Consortium for Computing Sciences in Colleges	2006
Proceeding of ACM International Conference on Privacy, Security and Trust	2006
ACM Special Interest Group on Management Information Systems	2006
Proceeding of InfoSecCD Conference	2006
European Journal of Operational Research	2006
Strategic Information Systems	2005
Computers & Security	2005
Omega	2005
Information Technology and Management	2005
Information Security Curriculum Development (InfoSecCD) Conference	2005
Proceeding of 7th International Conference on Electronic Commerce, ICEC	2005
The DATA BASE for Advances in Information Systems	2005
Computers & Security	2004
International Journal of Information Management	2004
Sixth International Conference on Electronic Commerce, ICEC	2004
International Journal of Information Management	2004
Information Management & Computer Security	2003
Pers Ubiquit Computing	2003
Consortium for Computing Sciences in Colleges	2003
Information Technology	2000
IEEE Computer	2000
ACM SIGCPR Computer Personnel	2000
Supply Chain Management: An International Journal	2000
Information Management & Computer Security	1999
Management Decision	1999
Strategic Information Systems	1999
Supply Chain Management: An International Journal	1999
MIS Quarterly	1998
ACM Computing Surveys	1993
Management Information Systems	1991
IEEE Software	1991

Table 2. Related articles under review

7. References

Ahlan, A. R. (2005). Information technology implementations: Managing IT innovation in the Malaysian commercial banking industry. Unpublished doctoral dissertation, University of Cardiff, United Kingdom.

Ahlan, A. R., Arshad, Y. & Lubis, M. (2011). Implication of Human Attitude Factors Toward Information Security Awareness in Malaysia Public University. Proceedings in International Conference on Innovation and Management (IAM2011), Kuala Lumpur, Malaysia.

APM (2007). Project Risk Analysis and Management Guide, second edition. Association for project management (APM).

Arshad, Y. (2011). IT Outsourcing decisions and implementations in Malaysia public healthcare sector agencies: Grounding an ITO relationship model using qualitative approach. Unpublished doctoral dissertation. International Islamic University Malaysia.

Aubert, B. A., Patry, M & Rivard, S. (2005). A Framework for Information Technology Outsourcing Risk Management. The DATA BASE for Advances in Information Systems, 36,4.

Bahli, B. & Rivard, S. (2003). The information technology outsourcing risk: a transaction cost and agency theory-based perspective. Journal of Information Technology, 18, pp.211–221.

Bahli, B. & Rivard, S. (2005). Validating measures of information technology outsourcing risk factors. Omega, 33, pp.175 – 187

Barki, H., Rivard, S., & Talbot, J. (1993). Toward an assessment of software development risk. Journal of Management Information Systems, 10(2), pp.203-225.

Benoit A. Aubert, Michel Patry & Suzanne Rivard (2005) A Framework for Information Technology Outsourcing Risk Management. The DATA BASE for Advances in Information Systems, 36(4), pp.9-28.

Boehm, B.W. (1991). Software Risk Management: Principles and Practices. IEEE Software, 12, pp.32–41.

Boran, S., (2003). IT security cookbook. Boran Consulting.

Bouchaib Bahli & Suzanne Rivard (2005) Validating measures of information technology outsourcing risk factors. Omega, 33, pp.175 – 187.

Bunmi Cynthia Adeleye, Fenio Annansingh & Miguel Baptista Nunes (2004) Risk management practices in IS outsourcing:an investigation into commercial banks in Nigeria. International Journal of Information Management, 24, pp.167–180.

Cochran, J. (2006). A Comprehensive Model for Understanding Technology Selection Decisions of Interconnected Information Technologies, Proceedings of SIGMIS-CPR'06, April 13–15, 2006, Claremont, California, USA.

COSO (2004). Enterprise Risk Management — Integrated Framework: Executive Summary. By Committee of Sponsoring Organizations of the Treadway Commission.

Fariborz Farahmand, Shamkant B. Navathe, Gunter P. Sharp & Philip H. Enslow (2005) A Management Perspective on Risk of Security Threats to Information Systems. Information Technology and Management, 6, pp.203–225.

Flanagan, R. & Norman, G. (1993). Risk management and construction, Blackwell Scientific Publications, London.

Frame, J.D. (2003). Managing risk in organizations: A guide for managers, Jossey-Bass, NY, US.

GAO/AIMD-00-33 (1999). Information Security Risk Assessment, Practices of Leading Organizations. A Supplement to GAO's May 1998 Executive Guide on Information Security Management. United States General Accounting Office, 1999 [ai00033.pdf]

Gerace, T. & Cavusoglu, H. (2009). The Critical Elements of the Patch Management Process. Communications of the ACM, 52(8).

Gottfried, I.S. (1989). When disaster strikes. Journal of Information Systems Management, pp.86-9.

Hillson, D. (2002). Extending the risk process to manage opportunities. International Journal of Project Management, 20, pp.235–240

June Wei, Jason O'Connell, & Meiga Loho-Noya (2010) Information Technology Offshore Outsourcing Security Risks and Safeguards. Journal of Information Privacy & Security, 6(3), pp.29-46.

Kakoli Bandyopadhyay, Peter P. Mykytyn & Kathleen Mykytyn (1999) A framework for integrated risk management in information technology. Management Decision, 37(5), pp.437-444.

Keil, Mark., Cule, Paul E., Lyytinen, Kalle & Schmidt, Roy C. (1998). A framework for identifying software project risks. Communications of the ACM, 41(11), pp.76–83.

Kroenke, D. (2009). Using MIS. Upper saddle river: Pearson prentice hall.

Kweku-Muata Osei-Bryson & Ojelanki K. Ngwenyama (2006) Managing risks in information systems outsourcing: An approach to analyzing outsourcing risks and structuring incentive contracts. European Journal of Operational Research, 174, pp.245–264.

Longstaff, T.A., Chittister, C, Pethia, R. & Haimes, Y.Y. (2000). Are We Forgetting the Risks of Information Technology? Journal of Computer.

Macdonald, D. (2004). Practical machinery safety. Pondicherry, India: Integra Software Services.

McFarlan, F.W. (1981). Portfolio approach to information systems. Harvard Business Review, 59(5), pp.142–50.

Melinda Cline, Carl S. Guynes & Andrew Nyanoga (2010) The impact of organisational change on information systems security. Journal of business and economic research, 8(1), pp.59-64.

Nikolić, B. & Ružić-Dimitrijević, L. (2009). Risk Assessment of Information Technology Systems. Issues in Informing Science and Information Technology. 6, pp.595-615.

NSW DEC (2011). Enterprise Risk Management in the Department of Education and Communities.
https://www.det.nsw.edu.au/policies/general_man/erm/PD20040036.shtml
accessed on 31st October 2011.

O'Brien, J. A. (1996). Management information systems: Managing information technology in the networked enterprise. Boston: McGraw-Hill.

Raftery, J. (1994). Risk analysis in project management, E & FN Spon, London.

Ramanujan, S. & Jane, S. (2006). A legal perspective on outsourcing and offshoring. Journal of American academy of business, 8(2).

Rao, M. (2004). Key issues for global IT sourcing: country and individual factors. EDPACS, 32(4), pp.1-12.

Rosman, R. (2008). Risk Management and Performances of Islamic Banks: A Proposed Conceptual Framework. 2008 EABR & TLC Conferences Proceedings.

Sadgrove, K. (1996). The Complete Guide to Business Risk Management. Aldershot: Gower.

Solms, B. V. (2005). Information Security governance: COBIT or ISO 17799 or both? Computers & Security, 24, pp.99-104.

Straub, D.W., & Welke, R.J. (1998). Coping with Systems Risks: Security Planning Models for Management Decision Making, MIS Quarterly, 22(4), pp.441-469.

Tafti, M. (2005). Risks factors associated with offshore IT outsourcing. Industrial management and data systems, 105(5), pp.549-560.

Thomas Gerace & Huseyin Cavusoglu (2009). The Critical Elements of the Patch Management Process. Communications of the ACM, 52(8).

Ward, J., & Griffiths, P. (2001). Strategic planning for information systems. Chichester:Wiley.

Webster, J., & Watson, R. T. (2002). Analysing the past to prepare for the future: Writing a literature review. MIS Quarterly, 26(2), pp.13-23.

Whitman, M. E., & Mattord, H. J. (2004). Management of Information Security. Boston: Thompson Course Technology.

Willcocks, L.P., Lacity, M.C. & Kern, T. (1999). Risk mitigation in IT outsourcing strategy revisited: longitudinal case research at LISA, Journal of Strategic Information Systems, 8, pp.285–314.

14

Enterprise Cyber Risk Management

Patrick L. Brockett, Linda L. Golden and Whitley Wolman
University of Texas at Austin
USA

1. Introduction

Cyber risk represents an ever-growing threat to public and private institutions alike due to its potentially disastrous effects on organizational information systems, reputational risk, and potential loss of consumer- and stakeholder's confidence. With the advent of the internet and the corresponding proliferation of information technology, firms, non-profits, and governmental entities were generally unprepared for identifying and addressing this risk, but the threat has increased in both frequency and severity over time, and the nature of attacks has also changed. In many early cases, the perpetrators of cyber attacks and information disruption campaigns interrupted business operations simply for their own amusement, or viewed breaking into the corporate information technology (IT) infrastructure as a challenge. They would deface websites or take down servers in order to aggravate or simply to challenge other cyber professionals in order to prove they could do it, not to profit (Hallam-Baker, 2008). However, as the Internet has grown and e-commerce has blossomed, employee access to company data has increased, and remote access to internal computer systems has become commonplace, cyber attackers have evolved, becoming more sophisticated and their effects becoming more devastating (Rhemann, 2011). Current cyber threats and attackers are increasingly focused on profiting from the consequences of their attack actions and either exploit the data they illicitly obtain for private gain or require payments from the victimized enterprise to restore service, access, or websites back to operational functionality (Maillart & Sornette, 2010).

The focus of this chapter will be on enterprise cyber risk management and risk mitigation (as opposed to individual consumer cyber risk, an extensive connected topic which is of interest in its own right but not addressed here). In this chapter we will investigate cyber risk of importance to enterprise to include information theft, compromise of consumer information, and the interruption of e-commerce. This chapter will focus on several important aspects of cyber risk and how these affect the economics and security of organizations. Cyber risk is unique among other operational enterprise risks due to its mobile location, scope of threat, and high-profile impact. With the proliferation of business services, systems, and data accessible via the Internet, cyber threats to enterprises (public and private) have grown immensely. Furthermore, companies have now realized that they run the risk of creating liabilities from any cyber event that could affect related services and products (e.g. the theft of email addresses from Epsilon on April 2011 also affected customers of numerous other businesses, such as Hilton Hotels, Citibank, etc., and has cast reputational risk on not only the original company but also on their clients, and has created increased potential legal liability as well.).

The ripple effect that cyber attacks can engender can influence suppliers, end users, and the organization itself, and could even have the potential to destabilize large swaths of the economy if the target of the cyber attacks were systemically important (such as a systemically important financial institutions, a utility operators, a water treatment facility, a transportation network, etc.). Additionally, cyber espionage techniques are developing rapidly, making enterprise trade secrets also vulnerable to competitor theft.

As with many other hazards faced by businesses, insurance companies who specialize in risk assumption and risk pooling saw a potential financial opportunity in filling the cyber risk hazard management needs of enterprises by providing insurance policies designed to protect or indemnify against the financial consequences of these Internet-related threats. Several insurance companies have started to offer connectivity-related policies that cover cyber information and security breaches. Early on, as the insurers tentatively waded into this new market, it was difficult to generate data on electronic losses. Although Internet-related insurance coverage is still in its infancy (as compared to other insurance classes), insurance companies over the past few years have now improved their ability to more accurately price policies and predict potential losses. These companies, including AIG, Chubb, Fidelity, Marsh, and Lloyds of London, have written policies that can hedge or transfer varying aspects of cyberspace risk (Gordon *et al.*, 2003). Further aspects of cyber-related insurance will be discussed in detail later in this chapter.

The chapter starts with a general discussion of cyber risk threats to organizations including trends and costs. These threats will be dichotomized into those cyber risk threats that arise internal to the organization (e.g., employee cyber-based financial theft, employee data theft, identity theft using internal company data, etc.), and those risk that arise external to the organization (e.g., hackers stealing data, money or trade secrets, or adversaries shutting down or disabling internal information technology, vulnerability of IT systems to external power surges, blackouts, etc.). Next in the chapter we shall discuss emerging cyber risk threats and trends, with particular attention on risk consequences of the emerging trend of organizations and individuals to use wireless mobile technology (smart phones, iPads, etc.) to conduct business to business transactions, access enterprise networks, do banking, and accomplish retail consumer purchases.

Having identified these major cyber risks, in this chapter we subsequently investigate the underlying economic considerations (and theory) related to cyber risk, including the extent to which these threat costs are internalized in stock prices and who bears the costs for such risks. One of the most important risk financing mechanisms utilized by enterprises for all the risks they face is insurance risk transfer. This is also the case to a certain (but more limited) degree with cyber risks as well. Consequently, after discussing the economic aspects of cyber threats, we next discuss cyber risk insurance, its availability, coverage and the economic issues related to cyber risk insurance such as moral hazard/adverse selection and issues concerning systemic risk causing correlation in the insurer's portfolio, such as the dominant use of particular software by multiple users (e.g., Microsoft Windows, Adobe Reader) so that hackers exploiting vulnerabilities in a single software product can cause losses for numerous insured clients. This reduces the risk pooling and diversification benefits that insurers depend upon when pricing their products (i.e., their estimates of aggregate loss probabilities based on independent loss occurrences which may differ substantially from those actually experienced when risks are highly correlated). The chapter concludes with comments about future trends and research.

2. Cyber risk threats

The cyber risk threat to enterprises is large and growing (Hallam-Baker, 2008; Rhemann, 2011). The Federal Bureau of Investigation (FBI) in the USA, universities, and other research organizations have delved deeply into the issues surrounding cyber security as a threat to public governments as well as private corporations. A 2002 Computer Security Institute/FBI joint study on cyber risk found that 90 percent of respondents had detected computer breaches within the past year, with an average loss of over $2M per organization (Power, 2002). In the then relatively new age of information technology and the Internet, most companies were not adequately prepared to face these types of costly losses. By 2008, however, the Computer Security Institute/FBI study found that the average loss had decreased to approximately $300,000, suggesting that companies and the security software they use have become more sophisticated in an effort to deal with the increasing threat of criminal cyber activity (Computer Security Institute, 2008). The 2008 CSI/FBI survey also found that companies significantly boosted their internal budgets associated with cyber security, which further implies that companies are spending more money, time, and manpower to mitigate these risks (Computer Security Institute, 2008). Cyber threats can shut down power grids, steal information and intellectual property, uncover competitors' bids, and disable web sites needed for business activity, causing substantial financial harm to unprepared enterprises. Accordingly, it is likely that companies will need to continue to focus on these cyber risk security issues as hackers continue to get more sophisticated causing more losses to business, on-line services, and operations, and especially as companies become more dependent on the Internet for e-commerce, mobile (or m-) commerce, or simply for daily operations, administration, and field contact with employees.

As the proliferation of information technology, the increasing facilitation of remote access to enterprise computers, and the corresponding risk of cyber threats have increased, so has the attention paid to these issues also increased. When focusing on these threat issues, for this chapter it is useful to broadly dichotomize cyber risk into 1) cyber risks that arise internal to the organization and 2) those that arise external to the organization. While certain risk mitigation techniques are common to both sources of cyber risk threats (e.g., securitization and password protection of sensitive information or technology, segmentation of information and its access within an organization, etc.), other techniques are more appropriate for one risk source rather than the other. The threats that each source of risk poses can be different and may often require different approaches. Moreover, as developing nations fight for parity with the more developed countries in terms of electronic Internet access and technological and industrial development, firms, non-profits and governmental entities and institutions are likely to see an increase in cyber threats from these sources outside the control or jurisdiction of the enterprise's host country. We shall discuss each risk source in turn.

2.1 Internal cyber risk threats

Ironically, a very high risk of cyber crimes comes from within rather than outside the organization. While an employee can be a company's greatest asset, employees are constantly exposed to vast amounts of confidential information and are, by necessity, trusted with proprietary company information, inventory and property. Sometimes the temptation for individual gain can be too great. Or, an employee who spent time developing the important proprietary company information can feel they have a right to this company intelligence as a result of their time spent in research and development, product development, or technology transfer activities. Consequently, a company can be exposed to data or intellectual property theft from within rather than without.

Data theft is the term used when information is illegally copied or taken from a business or other individual. Employee theft of data, formulae, and process information can compromise the enterprise as readily as an external data theft attack, however because of their privileged position, the employee has more ability to act as the perpetrator since they already have trusted permission or password admittance into the cyber system of the enterprise for legitimate reasons, a permission that they may then turn against their employer. In fact, the FBI reports that employee theft is the fastest growing crime in America. The US Chamber of Commerce estimates about 75 percent of employees steal from their employer, with approximately 30 percent of corporate bankruptcies being the direct result of employee theft. The majority of involved individuals are higher level employees, and, on average, the time until discovery is approximately 18 months, giving substantial time for financial damage (Burke & Cooper 2010 p.433). Enterprises must be as vigilant against internal cyber threats as they are to external threats.

Removable media devices are the number one internal cyber security threat vehicle. Research conducted by Centennial Software in May 2007 found IT managers believe removable media devices now pose a larger security threat than either malware or viruses. In this 2007 survey quoted in Feig (2007), 38.4 percent of more than 370 respondents listed portable devices as their number one risk, up from the 25.7 percent in 2006. Due to this reality, in "IT Acceptability Policies" manuals, more organizations are now including security considerations of removable media devices in their risk management endeavors. Eighty percent of respondents reported that their organizations now dictate protocols for unauthorized use of removable media devices, with some prohibiting their use entirely. Other enterprises have modified their IT systems to either disable the USB ports or have installed software to prohibit downloading or uploading data without authorization via USB ports. The survey also found that 67 percent of IT staff use some form of removable media device on a daily basis and that the most popular type of device (65 percent) is the USB flash drive (Feign 2007). Never-the-less, despite the low cost, ease of use, ready availability to employees, and small size of USB devices, these devises were only used in 9 percent of data theft cases (Patel & Mischon De Reya 2011). Other larger, more sophisticated or faster devises are now also being used, such as iPods and MP-3 players.

Colorful names are often given by security professionals to the ingenious use of removable media drives for data theft. "Thumbsucking", for example, is the name given to data theft using a USB mass storage device, such as a USB flash (or thumb) drive to download confidential network information, literally "sucking" the data out of the network and onto the USB drive (Walsh 2011). This type of internal data theft threat has increased over the years. Whereas a previous limitation to the use of USB flash (or thumb) drives was one of memory space on the USB drive, this has been largely removed with modern USB drives. Price constraints have also been significantly alleviated on USB flash drives. Moreover flash drives are highly portable, compact, easily concealed, and instillation does not require the user to restart the computer system, making it a cheap and convenient tool for cyber theft (Walsh 2011).

Another fanciful name for a different, serious cyber theft risk is “pod-slurping”. This involves using an iPod or MP3 type player to rapidly steal gigabytes of information from an enterprise's computer system (Giannoulis 2011). iPods are widely used by employees and often played (with approval) while attached to enterprise or office computers. However, they also can be used to download massive amounts of confidential company information. According to Mello (2005), a 2004 research report on security risks (conducted by Gartner technology research and advisory company) stated that portable devices posed serious threats for companies, and this report inspired a security engineer named Abe Usher to write a “proof of concept” program called “slurp.exe” that allows an iPod or other removable device to be used to “suck” 100 MB worth of data from the Windows "Documents and Settings" directory in a matter of minutes (Giannoulis 2011). Mello (2005) reports that Usher wrote in his blog (www.sharp-ideas.net) “Using slurp.exe on my iPod, it took me 65 seconds to copy all document files (*.doc, *.xls, *.htm, *.url, *.xml, *.txt, etc.) off of my computer as a logged in user. Without a username and password, I was able to use a boot CD-ROM to bypass the login password and copy the document files from my hard drive to my iPod in about 3 minutes, 15 seconds." While this “proof of concept” program illustrates the potential for data theft using these devices (virtually the entire set of business records of a small to medium sized company could be downloaded in minutes), there is also empirical evidence of its actual use. In 2005 a Chinese spy sought asylum from the Australian security forces saying that over a period of several months he had stolen confidential data using his MP3 player (Hughes & Allard 2005). Also, there have been several court cases involving using such devices to steal confidential company data.

Identification theft (ID-theft) is another well publicized cyber risk vulnerability of enterprises, from both internal and external sources. Employees (and successful external hackers) can obtain access to customer records such as names, phone numbers, addresses, usernames, passwords and PINs, credit card and other account numbers, as well as Social Security numbers (Miller 2008). This information can then be sold on the Internet or used by the intruder him/herself to commit identity fraud (or for blackmail or extortion purposes to the enterprise via threatening data exposure). For example, when an employee, in the normal course of business, gains access to credit card numbers they may be tempted to use this information to make purchases or obtain other lines of credit for their own benefit (Stroup 2011). The risk is not small. According to the identity theft research center (ITRC 2011), in 2011 there have been a total of 112 breaches and 5,460,925 records exposed, as of April 5, 2011. Moreover, using data from various sources such as social media sites (Facebook, etc.) the identity thief can gather sufficient information to match data records allowing them to break password cryptology security, obtain credit card numbers and make purchases using another’s identity or credit cards. As a concrete illustration of the above mentioned internal cyber risk threat, we relate that in January 2009, Johns Hopkins University began receiving reports of identity theft activities in the Baltimore area surrounding their University (McMillan 2009). Ultimately, Johns Hopkins Hospital ended up having to warn over 10,000 patients about a woman that worked for the hospital who had access to Social Security numbers, names, addresses, dates of birth, telephone numbers, parents’ names, and medical insurance information and who had used this information to commit fraud. Yet another example of this internal source of cyber risk is a Wells Fargo Bank employee (Roberta Dunsworth) convicted in federal court for identity fraud. She was charged on December 1, 2010 with ten counts of bank fraud, two counts of aggravated identity theft, and two counts of fraudulent use of unauthorized access devices. Her fraudulent activities occurred while employed at the Wells Fargo Bank where she used the identity of a bank customer to obtain a credit card and a debit card and to open bank accounts (Admin 2011). Such employee related cyber risks can pose great financial, as well as and legal problems for employers if not adequately addressed preemptively.

An important internal covariate of internally perpetrated cyber risk is having disgruntled employees. These individuals may be motivated by revenge and will attempt to sabotage or destroy enterprise software or databases, thus depriving the enterprise of their property or costing the enterprise money. This more malevolent form of cyber risk should be addressed proactively by implementing a well-designed corporate security initiative, that includes policies such as, changing passwords prior to termination of employees with Internet access to enterprise computers and enforcing robust password strength requirements. Maintaining a log of user access to all corporate systems can be a preventative measure against cyber crime as well.

Internal data theft could also involve the employee as a victim of data theft. For example, New York Police Department employees became data theft victims in early 2009, when the personal records of approximately 80,000 police officers in a pension fund were stolen when an employee gained entry into a disaster recovery facility in Staten Island (InfoSecurity 2009). Similarly, in 2006 the McCombs School of Business at the University of Texas at Austin experienced a data breach with more than 197,000 personal records of faculty, staff, students, alumni and donors stolen. These examples illustrate the complex nature of cyber crime where inside intrusions include multiple types of data breaches such as customer data, employee data, and sensitive corporate data. Additionally, external components also may impact cyber security risk. We discuss the external components of cyber risk next.

2.2 External cyber risk threats

As developing nations fight for parity with the United States or other developed nations, American and other enterprises are likely to see an increase in targeted cyber attacks attempting to access their secret information or competitive knowledge (e.g., stealing competitive bids for strategic purposes). However, the risk from cyber threats goes beyond even the private sector into the public and governmental sectors. In 2010, President Obama declared threats to cyber-security a national security issue identifying "America's digital infrastructure [as] a 'strategic national asset' and [appointing] Howard Schmidt, the former head of security at Microsoft, as his cyber-security tsar" (The White House, 2010). In 2010, the President also directed the Pentagon to establish the U.S. military's Cyber Command to utilize a "full-spectrum" of operations in cyberspace (Economist, 2010). Additionally, the White House has directed several studies and initiated a national cyberspace strategy which stipulates the scope, process and development for using cyberspace (The White House, 2010).

The above pronouncement provides a further illustration of how flash drives, discussed previously as a threat in the internal risk section, can also pose cyber risk threats externally, even to the most secured enterprises. Flatley (2010) relates how an infected USB thumb drive was placed by a foreign intelligence agency in the parking lot of a Department of Defense facility in the Middle East. As might be expected by human nature, the person who found it put it in their computer. The computer was connected to the US Central Command and, according to Deputy Defense Secretary William J. Lynn in *Foreign Affairs*, the malware that was embedded on the device was able to spread and pass "undetected on both classified and unclassified systems, establishing what amounted to a digital beachhead, from which data could be transferred to servers under foreign control." The same could occur with corporate or other enterprise computers as it exploits the natural human curiosity that would occur if one were to find a "lost" flash drive. It was the above incident that in part, motivated the White House to establish a Cyber Command. Corporations should also take note.

There also have been well-reported cases of cyber espionage by Chinese and Russian sources that have infected military networks and corporate networks alike. Certain elements in China have targeted military networks in order to gain access to military secrets and information on U.S. operations. Experts are unclear about whether the attackers are officially sanctioned by the Chinese government, but their locations have been traced back to mainland China.

National Journal reported,

> …Brenner, who works for Director of National Intelligence Mike McConnell, looks for vulnerabilities in the government's information networks. He pointed to China as a source of attacks against U.S. interests. 'Some [attacks], we have high confidence, are coming from government-sponsored sites,' Brenner said. 'The Chinese operate both through government agencies, as we do, but they also operate through sponsoring other organizations that are engaging in this kind of international hacking, whether or not under specific direction. It's a kind of cyber-militia.… It's coming in volumes that are just staggering' (Harris, 2008).

Cyber threats from Russia tend to come from organized crime groups that prey on unsuspecting websites or servers and even demand a ransom in order to restore functionality. Russian attackers have also sought to influence political movements through coordinated denial of service attacks that hinder political organizations. In a Wired magazine article, Don Jackson of Internet security firm SecureWorks is quoted as saying:

> "… the denial-of-service attacks managed to shut down more than 80 percent of Kyrgyzstan's bandwidth. While both sites now seem to be up and running, several commentators have speculated that the attack is meant to thwart Kyrgyzstan's embattled political opposition — which depends on the Internet to organize — or to pressure Kyrgyzstan's government, which hosts a U.S. airbase outside of the capital, Bishkek"(Hodge, 2009).

Another example of governments using cyber tactics to control political events was the recent revolution in Egypt in spring 2011. The Mubarak regime in Egypt essentially "pulled the plug" on the Internet throughout Egypt to control protests and political gatherings. If the Internet can be shut down, or severely disrupted, then most enterprises throughout the entire country will be put at risk. Proactive data back-up, record keeping, and other plans must be made to mitigate this damage should such an event occur.

Identity and personal information theft discussed previously from an internal risk perspective is also a substantial risk from external sources, and poses a growing cyber threat to consumers and retailers using Internet technology. As recently as April 2011, computer database breaches affected millions of people. Epsilon, an online marketer, announced to customers that its email database had been hacked and intruders took millions of email addresses (Spicer & Aspan, 2011). The resulting damage not only affects the online marketer, but also the companies it services and the end users of their services. In the end, costs could grow into the millions if the company is found liable for negligence.

The Brookings Institution recently released a paper discussing the perils of identity theft and the need for consumer trust in online commerce (Friedman *et al.,* 2011). With the growth of social networking, online retail sites, and business services utilizing cyberspace, personal identity protection will continue to play a key role in the future of information technology development. Increasingly, a person's identity is used as the key to unlock various online portals of information, including secured corporate and governmental websites. Social security and bank account numbers are needed for online account creation, as well as government services. The potential for theft and fraud is greatly increased by the numbers of online users and the amount of information stored in electronic databases. Similarly, it is noted that users tend to use the same passwords for multiple Internet accounts, which could cause the risk to spread to other unsuspecting companies. Furthermore, websites are collecting larger amounts of data through "history sniffing" and the identities created from user-specific plug-ins (Friedman *et al*., 2011). This may be especially problematic because using smart phones to conduct business, (m-commerce) is substantially less secure, with employees possibly storing passwords and account numbers on the smart phones, and because smart phones (possibly containing account information and passwords) are lost or stolen at an alarming rate (Brockett *et al* 2011)

Whereas the traditional cyber risk threats discussed previously can be construed as exacerbated extensions of already existing risk control problems (physical risk perimeter securitization, employee theft risk control, corporate spying and intellectual property theft, etc.), the newly developing wireless or mobile technology is creating new (as opposed to merely enhanced) enterprise risk vulnerabilities that need enterprise wide attention. This is discussed in the next section.

3. Emerging cyber risk threats: Mobile Internet access, spear phishing and pharming

The ability to conduct business wirelessly (known as mobile or m-commerce) is revolutionary and growing in importance. Twenty two percent of consumers in 2010 used smart phones for price checking, 21 percent for doing product research, and 13 percent for making purchases using their phones (Schwartz, 2010), and enterprises including governmental entities, are changing web representations to accommodate this new interface modality. Reportedly 74 percent of online businesses have a mobile commerce strategy in place or are developing one (Marcus 2010 quoting a study by the National Retail Federation). Of those retailers not already involved in mobile commerce, one quarter say that they intend to begin within the next year (Siwicki 2010). Traveling sales people often access corporate data remotely using mobile smart devices, often smart phones, so the newly evolving wireless security risk is important to all enterprises.

There are, however, important distinctions between the risks associated with mobile Internet access and more familiar e-commerce or cyber risks. There are unique business challenges, cyber threats and data theft vulnerabilities embedded in this new modality. Mobile Internet connectivity uses a different communication channel (wireless) making business interface transactions more accessible at more times and places but also a different physical mode of communication than e-commerce. Due to the mobility of the communication, an interface between the accessing and accessed devises is more anonymous, making it more difficult to validate a transaction, and making it more difficult to secure the Internet transmission (eavesdropping on wireless communications is relatively easy if one is motivated to do so). Thefts are also more difficult to trace back to the perpetrator.

The Brookings Institution report "Online Identity and Consumer Trust: Assessing Online Risk" summarizes the difference in the physical characteristics of cyber risk and how those threats affect users:

> When communicating via a wireline, it is intuitive to most that the data traffic is leaving the computer through a data cord to an interface that connects with the Internet Service Provider…. It is quite difficult for a typical cyber criminal to intercept data … that he had not physically tapped. The same cannot be said of wireless network communications. Malicious actors can learn a great deal from unencrypted Wi-Fi links in their vicinity. …if the wireless connection is not itself encrypted using a modern standard …, then any nearby attacker can listen to all unencrypted traffic traveling between the computer and the wireless router. The data are being broadcast to the surrounding area by both the computer and the router in the same way that noise from a conversation is vulnerable to eavesdroppers. Thus, information that is not encrypted at the end points of the transaction can be intercepted. Tools to capture this traffic and reassemble the data packets into web pages are widely available, and usable to any moderately sophisticated computer user (Friedman *et al.*, 2011).

Mobile communication differs in the types of devices used, the development languages, communication protocols, and even the technologies used (Coursaris and Hassanein 2002). These differences make mobile communication subject to new threats. A mobile business operating system provides the infrastructure for running smart phone applications and mobile access to company computers but, according to (Ghosh and Swaminatha 2001), the platforms and languages being developed for wireless devices have failed to utilize even the basic security concepts already present in hardwired desktops. Without a secure infrastructure for mobile devices, achieving secure mobile Internet communication or secure employee to employer communication may not be possible. Additionally, the nature of the software applications developed for mobile devices are important to overall security since logical flaws or oversights in these applications can present exploitable security loopholes allowing a point of entry by the malicious hacker, and consequently making enterprises (and individual mobile communication device users) vulnerable. Unfortunately, the devices' physical limitations often force application developers to make security and performance trade-offs. Limited power, processing cycles, memory, and bandwidth can force developers to give up security features like encryption in order to improve online performance (Ghosh and Swaminatha., 2001). The use of lower-level languages for phone communication development, as well as their often lacking built in non-functional security requirements ensures continuation of software vulnerabilities.

With mobile devices, there is also the potential to remotely access data on "always on" mobile phones, including passwords, contact lists and other information. The phone hacking scandal involving News International's paper *News of the World*, a subsidiary of Rupert Murdoch's international News Corporation organization in the UK in 2011 illustrates this threat is real. These news organizations hacked into the voice mail of such well protected people as members of the Royal Family. They also allegedly hacked into the phones of a murdered English schoolgirl, relatives of British soldiers who had died, and even some of the victims of the terrorist bombings in London on 7/7/2005. The police have a list of approximately 4,000 people they are contacting who may have been hacked, including celebrities, politicians, and sports stars (BBC 2011). This can be a substantial concern for commercial enterprises as well since executives and sales persons use their mobile phones for negotiations, contract bids, and other purposes requiring secrecy.

Given their size and portability, phones are also at risk of physical theft and loss, with an estimated two million phones lost or stolen each year (Siciliano, 2011b). Some of the data stored on such devices may be proprietary business data, and, additionally, there is an increased risk that someone that finds or intentionally steals these smart phones can use the stolen device to access internal corporate systems, including servers and file systems using passwords stored on the smart phone. An estimated 52 percent of smart phone users store passwords on their phones; and 87.5 million people do banking using their phones (Siciliano 2011a). Moreover, since most people (even employees with secured access) do not use PIN codes to lock their cell phones, and since most people use the same password for multiple "secured" sites, a great vulnerability is created by mobile access that is not present in otherwise Internet accessible enterprises. One problem with current mobile phones is there is no readily available mechanism to authenticate that a particular user belongs to phone being used (Ghosh and Swaminatha., 2001), so access to an employee's phone may open the entire enterprise to potential cyber risk.

Two other emerging cyber threats are spear phishing and pharming. Phishing occurs when a potential thief sends emails to people which masquerades as a legitimate request from the organization whose letterhead and logo they have hijacked. They ask the recipient to click on a link to supply information (account number, password, social security number, etc.) in order to verify some aspect of their account. The link (as well as the email) are phony, and once the recipient puts in their information, it is captured and used for accessing bank accounts, credit cards, identify theft, or downloading malware which can take over the computer remotely to access all information on the individual's computer (such as other passwords or corporate information). Spear Phishing is a refinement which is even more difficult counter. Using some inside information gathered by other means (e.g., hacking a corporation's computer, social media sites, etc.) they do a controlled and targeted phishing expedition rather than simply sending random emails. Using information particular relevant to this well defined smaller group of people they construct an email which is more specific and has an enhanced air of being a legitimate request by a supervisor or trusted superior (often it is designed to come from a higher-up in the organization so compliance is enhanced). Again, if any one of the many targeted individuals within the company responds and logs on, security is breached and computer programs can be downloaded that allow full access to company computers for espionage purposes, identity theft, malicious destruction of data, extortion, or financial thievery (FBI 2009).

Pharming is the name given to a different type of technique used to direct the unsuspecting victim to a malicious website where malware can be downloaded or password and account numbers can be harvested in bulk numbers. Unlike phishing where the individual phishing lines (emails) are set out to catch fish, in pharming a network node is hijacked and all traffic going through this node which thinks it is going, for example to CibiBank.com, will instead be directed to another website controlled by the criminal. Essentially the criminal harvests multiple users' information at once, and need never even contact the user or need the user to respond to an email. The way is works is this. On the Internet, the website addresses are a sequence of numbers representing the site (e.g., 123.456.7.8 might represent the web site for XYZCompany.com). There is a translation mechanism built into DNS servers that converts words we write (say XYZCompany.com) into numeric address of the web site we want to access (123.456.7.8). A pharmer hacks into the DNS server and changes the translation book so that when you (or anyone else using this server) types in XYZCompany.com, it automatically sends the communication to another site (say 987.654.3.2) instead of the real site (123.456.7.8). As the phony site looks the same as the original, the unsuspecting user logs in without knowing that their account information and password are compromised. Malware can be downloaded onto the requesting computer, compromising many business activities and trade secrets (Norton 2011). Companies involved in Internet commerce have major concerns with pharming and the consequent fraud as their clients get scammed. Online banking sites are particularity sensitive to this threat. Moreover, adware and spyware removal software and antivirus software is ineffective in protecting against this threat since the hijacking occurs on the DNS server away from the requesting or responding computers, and hence is not detected by either side of the transaction which was hijacked. With the growth of wireless routers (both in businesses and homes) and in public access wi-fi availability, the potential to hijack data and transactions in mass quantities via pharming is an increasing threat that requires very specific anti-pharming defenses by the enterprises involved.

Having delineated important cyber risks we now turn to an investigation of the financial and economic consequences of such risks. In many ways the development of the Internet (and the consequent development of cyber risk threats) has been generated by economic considerations. Mobile Internet devises are rapidly replacing hard wired desk top computers as the Internet devices of choice, and the economics of this transition is impactful. Additionally, economic theory can alert us to possible ways of handling cyber risk problems, such as the economic research into moral hazard associated with "free goods" or public goods. We shall next discuss the economic aspects of cyber risks.

4. The economics of cyber risk

Motivated by efficient market theory, (Garg *et al.*, 2003) used event study methodology to indirectly measure the economic impact of Internet security breaches on stock process of breached firms (and also on Internet security providers). According to the efficient market hypothesis, all information about past and future events within a company (or industry) should be reflected in the stock price, which itself reflects investors' beliefs about future cash flows to investors. A security breach can cause a rethinking about future vulnerability as well as future legal risk, and hence reflect market assessment of impact on cash flows different from the reported financial loss (which may be biased or underreported).

Consequently, they reason, by looking at the abnormal (negative) return of a breached company's stock price, they can get an exogenous estimate of the permanent market assessed financial effect of the breach, different from the stated breach cost information. Moreover, such actual breach cost information, they argue, is not generally available to the market. Expenditures (and capability) in IT are not often reported. Also, most firms tend to underreport negative information containing security breach events simply because there is no incentive to correctly volunteer this information. Companies prefer not to seem vulnerable to their customers and competitors or to other potential predators. Why would these firms give an edge to the competition or a green flag to the cyber criminal if they do not have to? Additionally, in the age of management by stock price, firms also withhold this information to avoid lowering the price of the company's stock and falling out of favor with investors (Garg *et al.*, 2003). Another reason for not releasing such information, of course, is pride. No one likes to have their shortcomings and failures broadcast to the media. Once the breach is public, markets can react, and their reaction reflects the decreased valuation of the enterprise. Using their event study methodology applied to 22 cyber security breach events Garg, et al (2003) found the lasting effects on stock prices of security breaches is an order of magnitude larger than other reported loss costs ($17-28 million as opposed to other reported estimates of $50K to $2 million per incident). Thus, the economic effects on breached firms are quite significant.

Economists have also used other microeconomics tools in order to price in certain aspects of information security. Bohme (2005) argues that insecure software technologies such as public access wi-fi availability in some cities are economically underpriced by the market due to costs of their negative externalities not being valued. Thus, public access wi-fi are similar to a public good since insecure nodes not only affect their own systems but those systems and users that are connected to it. This enables viruses and other attacks to proliferate much faster and more freely than other physical attacks. Since responsibility for preventing the attack is uncertain, no one user has an incentive to spend heavily on securing one's own infrastructure (Bohme, 2005) which would also benefit others. Consequently, Internet users are unlikely to procure expensive protection software that protects the next user. This implies that the incentives are misaligned and showcases the risks of these interdependent information networks. Economic theory suggests that the bearer of risk should be the entity that has best control of the risk. In the case of cyber risk, this is often the manufacturer of the devices used (smart phones, for example) and the provider of the free wi-fi. Regulators and governments can assist in this endeavor.

Having investigated the economic aspects of cyber risk, it is natural next to turn to the most common financial mechanism available for indemnifying enterprises against the potentially disastrous financial consequences of a successful cyber crime perpetrated against the enterprise. This mechanism is insurance, and the next section discusses the growing area of cyber risk insurance.

5. Cyber risk insurance

Once companies evaluate their current conventional insurance coverage, many firms then evaluate and purchase Internet and information security insurance to cover their specific insurable cyber risks. Along with determining their budgets for cyber insurance, firms must choose between a blanket coverage policy and a more expensive, yet highly customized policy tailored to their business needs.

Anderson and Moore (2006) argue that cyber insurance is extremely difficult to price given the interconnectedness of the information security infrastructure and the interdependence on one piece of popular software (i.e. Microsoft Windows) whereby a general vulnerability in one product may expose every firm using that software to cyber threats (Anderson and Moore, 2006). Accordingly, an insurer insuring company X for cyber risk is also insuring against another company Y (e.g. Microsoft) who is not a client, and is not paying premiums, having created an exposure to cyber risk. Why should the insurance company pay for damages caused by another firm that infected or caused the infection of the covered firm? It seems that the law of large numbers which is often used to justify insurance company coverage could be defeated by the breadth and scope of the damages (and correlated risks of other insured companies using the same software) thus leading to the insurer's inability to pay claims and insolvency. If one flaw in a very common software system affects millions of users and propagates through several firms, the insurer might have difficulty paying all the resulting correlated damages for the sustained losses. It seems here that the network effect, typically lauded in economics, would have detrimental effects on the insured and insurers alike as it defeats the "independent identically distributed" (or at least uncorrelated) assumption which underlies insurance risk pooling and the general benefits of diversification. As our global interconnectedness grows, we must monitor the potential ripple effects that common interconnectedness can leverage on the global economy.

Firms have increasingly externalized the financial consequences of cyber risk by purchasing insurance to transfer that risk outside the company. Initially, insurance companies, given their lack of experience and practice associated with cyber risk insurance, have offered smaller coverage policies and packages. As firms and insurance companies develop more sophisticated analysis of cyber threats, the market for cyber insurance will likely grow. Indeed, many firms are pushing for the development of new markets and products with information security growth as a potential target (Gordon *et. al.*, 2003). As mentioned previously, several aspects of cyber-related insurance including pricing, information asymmetry, and correlations continue to influence the insurance market.

Traditionally, insurance premiums for commercial general liability are based on a firm's general features such as industry area, sales revenues, number of employees, and other similar characteristics (Baranoff *et al.*, 2010). Consequently, premiums typically do not reflect the firm's security activities, whether good or bad (Schwartz *et. al.*, 2010). If firms were more likely to demonstrate strict security practices, cyber risk coverage related premiums could be lowered, similar to the effect that a built-in safe or fire sprinkler system would have on a homeowner's policy, or a theft security system for an automobile would have on automobile insurance. For that reason, insurance companies find it necessary to separately assess and monitor these security precautions in order to verify the strength and level of protection.

Gordon, *et. al.* (2003) discuss their research on three aspects of cyber risk insurance including policy coverage pricing, adverse selection, and moral hazard. Since pricing depends heavily on actuarial estimates and historical data, pricing policies for Internet-related coverage are more uncertain than conventional insurance where data on claims have been gathered by firms such as the Insurance Services Office (ISO) and where coverages are more standardized (and the risks do not change over time) making actuarial loss estimates more credible. However, some insurance firms have established insurance policies and quantified this difficult to assess risk (although critics argue over the accuracy of their projections). Also, (Gordon *et al*. 2003) discuss adverse selection for firms in terms of their "likelihood of a breach." If a firm has a higher likelihood of facing cyber threats, that firm may have an increased likelihood of purchasing insurance to transfer this risk, similar to a smoker or someone in poor health that would buy more health insurance because they know that they have a higher than average (for their risk pool) chance of loss, but are being charged the average risk pool premium. When someone else is paying part of the risk cost, it is economically rational to buy more insurance. Insurance companies can mitigate the risks associated with the above mentioned adverse selection by requiring a security audit of the firm.They might also be able to differentiate premiums based on a firm's current security profile (creating a separating equilibrium solution to the adverse selection problem).

Moral hazard is another economic problem that occurs with cyber insurance products. Moral hazard occurs when the actions the firm takes are different simply because they have insurance indemnification. Why spend money on cyber security when losses are (in large part) indemnified by the insurer and hence shared by the risk pool ex ante, but premium savings are entirely captured by the owners? Gordon *et al*. (2003) argue that the moral hazard problem faced by insurance companies offering cyber risk policies could be eased by offering premium reductions to firms that take appropriate security measures on their own. The firm should be given financial incentives that influence its decisions to mitigate the risk on its own (similar to what occurs in workers compensation and other insurance). The firm's expense on risk reducing processes and behavior would help the firm mitigate cyber risk and would potentially reduce the impact of a cyber event, thus lowering premiums. Additionally, deductibles, policy limits, and coinsurance are standard tools used by insurance companies when information asymmetry is present (Schwartz *et al*., 2010). This puts a higher financial burden on the insured party to mitigate the effects of adverse selection resulting from information asymmetry, and inaction caused by the insurer being unable to adequately monitor behavior (moral hazard).

Cyber insurance coverage typically involves both first party and third party coverage for potential damages from Internet-related activities. Retail names in the cyber insurance market include Chubb's Cyber Security, Lloyd's e-Comprehensive, and Marsh's NetSecure. As described previously, however, insurance companies are often reluctant to underwrite large amounts of damages due to the relative newness of this specific type of insurance, the degree to which the insured has control over the frequency and severity of losses, and the lack of well verified loss data upon which to make actuarial estimates (and the potential sizes of risk and correlation with other risks). Lloyd's of London, however, offers a $50,000,000 limit under its e-Comprehensive policy but will write a custom policy for up to $200,000,000 (Gordon *et. al*., 2003). As more actuarial and damages data becomes available and cyber risk protection protocols become more standardized, it is likely that firms will be able to compete more broadly on coverage and premiums.

In all risk situations, even including potentially catastrophic risk scenarios, the best (and most cost effective) approach is to act to avoid the risk (risk prevention) or to reduce the consequences (risk mitigation) of the risk even before the risk has been materialized and potentially ruinous losses have been incurred. Risk mitigation- or risk prevention techniques can enhance the defenses of enterprises, and lower the cyber risk insurance premiums an enterprise pays to be indemnified after a loss event. In the next two sections, we discuss several risk prevention and risk mitigation methods for cyber risks.

6. Cyber threat risk prevention techniques

The adage "An ounce of prevention is worth a pound of cure" is especially true when dealing with cyber threats. If, for example, an enterprise's financial transaction over the Internet is hijacked and funds or information are stolen, it may be quite some time (if at all) before the theft is noticed. Additionally, it is likely that the proceeds will never be recovered nor will the thieves be apprehended (Clarke, 2008). It is much better to prevent the theft or cyber crime in the first place, and the first line of defense is purchasing a good suite of security software including, anti-spyware, adware detection, malware and antivirus protection that has been obtained from a reputable vendor. An automated update feature together with an automated routine scan of the system is also a must, and software patches should be installed when available. It is also good business practice to seek advice from advisors — cyber risk insurers, lawyers, accountants, and risk managers. For example, the cyber risk insurer Crum and Forster makes a private web portal that provides their clients with technical resources geared toward assisting them in preventing both network and private cyber losses, and provides support recovery if a cyber loss should occur. (Insurance Journal, 2011).

Concerning internal cyber theft of money, there are fundamental sound practices enterprises should follow to reduce the cyber risk associated with financial accounts, including implementing procedures to password protect checking accounts, accounts receivable checks, vendor and payroll checks and credit card receipts. Since many cyber breaches go undetected for long periods of time, there are additional procedures that can prevent ongoing cyber theft including separating the duties of check writing from reconciling checking accounts, as well as performing unannounced periodic audits of accounts payable and checks paid. Over a certain amount, the enterprise should also establish a dual signature requirement for checks made out, and establish limits on the credit card spending on employee credit cards. This prevents (or mitigates) large losses if a cyber thief enters the system as checks or fund transfers cannot be routinely done in secrecy. Similar controls also should be used to protect intellectual property and valuable information such as databases by restricting access or needing verification to obtain copies, or keeping an automated log of who has accessed a particular record or data set. Commercial and non-profit enterprises do not have the same legal protection against cyber thievery of bank accounts that individuals do (the bank must reimburse the individual but not the company) so proactive diligence is especially warranted by enterprises for transactions involving financial transfers over the Internet (Johnson 2011). While cyber theft insurance can provide a loss control mechanism against such risks, it will generally be subject to a deductible and hence still contain a loss potential for the enterprise

Additionally, many instances of internal cyber (or just plain employee) theft could have been avoided had employees, prospective employees (and even board members and trustees) undergone a criminal background check. Unwillingness to agree to such checks should be a red flag. Also it may be worthwhile to have employees (regardless of their tenure), undergo a criminal background and credit check every five years or so, especially if they have access to financial accounts or check signing authority. As mentioned previously, disgruntled employees should be particularly scrutinized if they have sensitive information access. According to the 2012 Global State of Information Security Survey by Pricewaterhouse Cooper (PwC 2012), 15 percent of respondents (from over 9,600 CEOs, CFOs, CISOs, CIOs, CSOs, VPs and directors of IT and information security from 138 countries.) strongly agreed that the risks to company data had increased due to employee layoffs.

Additional prevention measures can include encrypting signals at both ends of the communication channel, and higher level authentication of identity before allowing entrance into cyber locations having potential for breach and information loss. Some banks, for example will, each time, use a secondary verification method before allowing access to accounts. In this process the individual is sent a text or email message with a specialized code which must be entered along with the password when attempting to log in to the account. Similar methods can prevent many forms of unauthorized access into enterprise computer systems, and thus prevent losses before they occur.

7. Cyber risk mitigation techniques

While not all risks can be prevented, the damaging effects can be mitigated by judicious planning. Typically, firms that are looking to counter their cyber risk will utilize risk management frameworks and techniques that identify information security vulnerabilities. The first step is a security audit performed by the firm (or a third party) which identifies risks and vulnerabilities within the company's systems. This step usually involves inspecting the physical computing environment for external risk threats, as well as examining electronic networks (including offsite access by employees and customers). Additionally, companies gather information on current risk profiles by interviewing IT managers and determining the financial costs of the risk management process. In many cases, firms take the recommended steps to coordinate their own in-house response by setting up access controls and enabling firewalls before they consult externally with insurers or security experts (Siegel *et. al.*, 2002).

A very important risk mitigation technique for enterprises to implement is the use of data encryption, essentially coding each document so that it cannot be read, even if stolen or hijacked in mobile transmission. Encrypting transmission and/or documents makes it almost impossible for third parties to productively hack into databases or mobile devices (Ohlhorst 2010). There are many ways to use encryption. Single files can be encrypted or entire archives can be encrypted. There are several different types of encryption. The two main leading types of encryption are private key cryptography and public key cryptography.

Private key encryption has a single key that is used for encryption and decryption. According to (Ohlhorst 2010) "Private key algorithms are generally very fast and easily implemented in hardware, so they are commonly used for bulk data encryption." Private key encryption is mainly used for file, directory, and partition encryption that is only known by the owner of the data. There are two general categories of private key algorithms: stream ciphers and block ciphers. A stream cipher individually encrypts every byte of the data and is commonly used for wireless communications. Alternatively, block ciphers encrypt one block of data at a time and are used mainly for data encryption (Ohlhorst 2010).

Public key cryptography involves the use of two distinct but related keys: a public key and a private key. The public key can be shared with anyone and is used to encrypt data meant for the holder of the private key. The private key cannot be shared and is used to decrypt any data encrypted by the public key. Public key cryptography is primarily used for e-mail messages, file attachments, digital signatures and other transaction-related processes (Ohlhorst 2010).

Monitoring and detection is also a critical step in avoiding cyber risk. Many times firms are unaware of, or provide an inadequate response to, a possible breach that could have been thwarted. If the threatened firms used updated monitoring and intrusion techniques to detect attacks or threats in real-time, their performance rate would increase significantly. Security consultants can help both in delineating risk, outlining risk mitigation techniques, assessing the financial consequences of such risks, and performing and monitoring (as the environment is constantly changing) risk audits and assessing cyber vulnerabilities.

8. Some comments about future trends and research

Regulation and controls in cyber technology have developed at a much slower pace than actual growth and progress in the technology itself, thereby causing a lag in enforcement and justice. As mentioned previously, many of the cyber risk threats originate from countries different from the host country, and regulating or enforcing laws against such trans-border criminals can be difficult or even impossible. Governments and international regulatory bodies, such as the United Nations, are now trying to develop stricter regulations in order to deter these types of illicit cross national cyber risk threat activities. However, until there is broad consensus on enforcement and retribution, companies will be forced to tackle these risks on their own. Any risk manager looking into the future must be able to plan for these unique threats and their growing sophistication. Since the Internet allows potential access from anywhere, firms and governmental enterprises must be prepared to address both internal and external cyber risk threats.

Proactively, regulators and governments can also intervene to help reduce the risk of cyber theft or crime. As mentioned previously, currently software makers for mobile Internet devices (smart phones, iPads, etc.) do not adhere to security requirements that hardwired, Internet-connected computers use due to tradeoffs between security and the storage size, and speed of performing tasks on these devices. Security has taken a back seat in this trade-off, and most users are unaware. Regulators can impose standards, which make cyber theft via such mobile devices more difficult and have mobile Internet device manufacturers make software patches available as vulnerabilities become known. The exact form of such regulations is an important topic for future research.

The rather rapid emergence of the Internet and information technology over the last decade has contributed to more efficient communication, which has allowed companies to reach broader markets in the new global economy. As companies and organizations continue to rely increasingly on cyberspace for communications, on-line services, and electronic databases, and as employees continue their trend toward mobile or remote access to enterprise infrastructure and assets, the importance of mitigating cyber risk for enterprises will continue to rise. Insurers have already begun to offer cyber-related coverage but what remains to be seen is how effective those policies will be in transferring the risk. Insurers will be concerned about moral hazard problems wherein an enterprise, which has cyber risk insurance takes less protective action because they have insurance and do not bear the entire risk costs.

Currently, individual firms are able to mitigate their risks through risk management processes tailored to mitigate or control cyber threats. If firms utilize firewalls, pin and password access systems, encryption, and secure ids, they have a greater (but non-zero) chance of avoiding a large-scale Internet-related attack and financial losses. As a result of such proactive policies, they are also likely to obtain lower premiums for cyber risk insurance coverage, and better security audits by insurers, which can reduce insurance costs. Never-the-less, the potentially widespread impact of an organized or coordinated cyber attack or information security breach could overwhelm any insurer with claims (and a cyber attack on infrastructure could also overwhelm governmental enterprises as well). Therefore, it is critical that enterprises establish policies that aid in pre-loss financing of these potential damages in ways that avoid insolvency.

It is clear that public and private institutions will increasingly feel the effects of cyber risk from their own actions or those of a connected supplier, distributor, or end user. And, as developing countries push for greater access to the global market, competing companies may see cyber information control as a means of accomplishing that goal (e.g., cyber information theft of such things as intellectual property, competitive bids, etc.). This vulnerability may be especially pertinent to those companies who engage in outsourcing in a manner that necessitates data access by foreign companies to host company computers. Even while the host country firm may have installed cyber threat protection the outsourced foreign firm may not have equivalent cyber protection, and their access vulnerability together with their permitted access to host computers may pose risks for the host enterprise. Due to substantial interconnectedness, enterprises must be cognizant of cyber risk management plans of their suppliers and their downstream distributors who have access to the enterprise's computer accounts.

As information technology evolves, enabling enterprises to utilize the benefits advancing technology provides, from enhanced marketing facilitation, to enhanced employee access to important enterprise information during negotiation processes, and to allowing customers to access personal records from anywhere, any time, we must remain vigilant and protect our enterprises from the cyber vulnerabilities that this new technology brings. As Dr. Martin Luther King, Jr. (1963) said, "All progress is precarious, and the solution of one problem brings us face to face with another problem." Even as new technology is creating opportunities for enhanced efficiency in enterprise activities via such advances as immediate employee access, targeted marketing ability, better customer service, and enhanced governmental transparency, Internet users should be aware that these advances create the new problem of enterprises becoming increasingly susceptible to cyber threats.

An important area of future research is how to enjoy the benefits of the information explosion without succumbing to the perils of cyber risk from inside and outside the organization. Benchmarking, state-of-the-art risk mitigation techniques, and proactive management will be necessities in the forthcoming world of interconnected commerce.

Cyber risk considerations should rise to the level of Boards of Directors in the near future as the consequences of failure are simply too large to ignore or do otherwise.

9. References

Admin[at]databreaches.net. (April 2011). Former Wells Fargo employee sentenced for ID theft, In: *Office of Inadequate Security,* September 19, 2011, Available from http://www.databreaches.net/?p=17654

Anderson, R., & Moore, T. (2006). The Economics of Information Security, *Science,* Vol. 314, No. 5799 October 2006 pp. 610-613, Available from http://www.sciencemag.org/content/314/5799/610.full.pdf

Baranoff, E., Brockett, P., & Kahane, Y. (June 2009). Risk Management for Enterprises and Individuals, In: *Flatworld Knowledge,* Available from http://www.flatworldknowledge.com/printed-book/1635

BBC News. (2011). *Q&A: News of the World phone-hacking scandal, BBC Mobile News UK,* (August 4, 2011), Available from http://www.bbc.co.uk/news/uk-11195407

Bohme, R. (2005). Cyber-Insurance Revisited, *Proceedings of the Workshop on the Economics of Information,* The Heartland Institute, Policy Documents, Chicago, Illinois, USA January 1, 2005, Available from http://heartland.org/policy-documents/cyber-insurance-revisited

Brockett, P. Golden, L., Manika, D & Song, A. (2011). Developments in Mobile Commerce: Economic Opportunities, Risk Analysis and Risk Management, *Working paper,* Center for Risk Management and Insurance, University of Texas at Austin, USA

Burke, R. & Cooper, C. (2010). *Risky Business, Psychological, Physical and Financial Costs of High Risk Behavior in Organizations,* p.433, Gower Publishing, Ltd., ISBN 978-0-566-08915-2, Surrey, England.

Clarke, R. (June 15-18, 2008). A Risk Assessment Framework for Mobile Payments, *21st Bled eConference e Collaboration: Overcoming Boundaries through Multi-Channel Interaction* Bled, Slovenia, April 29, 2011, Available from http://domino.fov.uni-mb.si/proceedings.nsf/Proceedings/FC5AA5C853A1CF3DC1257481003D0293/$File/06Clarke.pdf

Coursaris, C. & Hassanein, K. (2002). Understanding M-commerce - A consumer centric model, *Quarterly Journal of Electronic Commerce,* Vol. 3, No. 3 pp. 247-271

FBI (2009). Spear Phishers: Angling to Steal Your Financial Info, (April 1, 2009), *The FBI, Federal Bureau of Investigation,* Available from http://www.fbi.gov/news/stories/2009/april/spearphishing_040109

Feig, N. (2007). Banks Aren't Securing USB Ports, Study Reports, In: *Bank Systems and Technology,* June 17, 2007, Available from http://www.banktech.com/risk-management/201000516

Flatley, J. (2010). Thumb drive-based malware attack led to formation of US Cyber Command, *Engadget, AOL Tech,* August 26, 2010, Available from http://www.engadget.com/2010/08/26/thumb-drive-based-malware-attack-led-to-formation-of-us-cyber-co/

Friedman, A., Crowley, P., & West, D. (2011). Online Identity and Consumer Trust: Assessing Online Risk, *The Brookings Institution,* January 11, 2011, Available from http://www.brookings.edu/papers/2011/0111_online_identity_trust.aspx

Garg, A., Curtis, J., & Halper, H. (2003). Quantifying the financial impact of IT security breaches, *Information Management and Computer Security,* Vol. 11, No. 2 pp. 74-83, Available from http://www.emeraldinsight.com/journals.htm?articleid=862842&show=abstract

Ghosh, A. & Swaminatha, T. (2011). Software Security and Privacy Risks in Mobile-Commerce, *Communications of the ACM,* Vol. 44, No. 2 (2001), pp. 51-57

Giannoulis, P. (2011). Pod slurping: The latest data threat, In, *SearchMidmarketSecurity.com*. April 11, 2011, Available from http://www.searchmidmarketsecurity.techtarget.com/tip/Pod-slurping-The-latest-data-threat

Gordon, L., Loeb, M., & Sohail, T. (2003). A Framework for Using Insurance for Cyber Risk Management, *Communications of the Association of Computing Machinery,* Vol. 46, No. 3, (March 2003), pp. 81-85, ISSN 0001-0782, Available from http://portal.acm.org/citation.cfm?id=636774

Hallam-Baker, P. (February 21, 2008). Famous for Fifteen Minutes: A History of Hacking Culture, In: *CSO Online-Security and Risk*, September, 11 2011, Available from http://www.csoonline.com/article/217058/famous-for-fifteen-minutes-a-history-of-hacking-culture

Harris, S. (2008). China's Cyber-Militia, *National Journal,* May 31, 2008, National Journal Group, Inc. 2011, Available from http://www.nationaljournal.com/magazine/china-s-cyber-militia-20080531

Hodge, N. (2009). Russian 'Cyber Militia' Takes Kyrgyzstan Offline? *Wired,* (January 28, 2009), Available from http://www.wired.com/dangerroom/2009/01/cyber-militia-t/

Hughes, G. & Allard, T. (2005). Fresh from the Secret Force, a spy downloads on China, *The Sydney Morning Herald*, (June 9, 2005), Available from http://www.smh.com.au/news/National/Fresh-from-the-Secret-Force-a-spy-downloads-on-China/2005/06/08/1118123901298.html

Identity Theft Resource Center. (2011). *Identity Theft Resource Center A Nonprofit Organization,* March 21, 2011, Available from http://www.idtheftcenter.org/artman2/publish/lib_survey/ITRC_2008_Breach_List.shtml

InfoSecurity. (March 5, 2009). NYPD victim of data theft, In: *infoSecurity.com,* March 30, 2011, Available from http://www.infosecurity-us.com/view/555/nypd-victim-of-data-theft-/

Insurance Journal. (June 8, 2011). Crum & Forster Launches New Service to Protect against Cyber Risk, *Insurance Journal* August 17, 2011, Available from http://www.insurancejournal.com/news/national/2011/06/08/201793.htm

Johnson, S. (2011). Cyber-theft bedevils businesses: Commercial enterprises don't enjoy the same protections as consumers from online bank heists, In:The Miami Herald, Business technology section, September 17, 2011, Available from http://www.miamiherald.com/2011/04/04/2150009/cyber-theft-bedevils-businesses.html

King, Martin Luther, Jr. (1963). Quote taken from *Strength to Love*

MacMillan, J. (2009). Johns Hopkins Tells Patients: Employee Stole Data for Fraud, In: *CSO Online - Security and Risk* . CXO Media Inc., April 12, 2011, Available from http://www.csoonline.com/article/492427/johns-hopkins-tells-patients-employee-stole-data-for-fraud

Marcus, S. (July 22, 2010). Top 5 Mobile Commerce Trends for 2010, April 30, 2011, Available from http://mashable.com/2010/07/22/2010-mobile-commerce-trends/

Maillart, T., Sornette, D. (2010). Heavy-tailed distribution of cyber-risks, *European Physical Journal B,* Vol. 75, No. 3 (June 2010), pp. 357–364, Available from http://www.springerlink.com/content/866j4814v275r582/fulltext.pdf

Mello, J. (September 29, 2005) Pod Slurping: Threat or Hype? In: *Welcome to TechNewsWorld,* March 27, 2011, Available from http://www.technewsworld.com/story/46417.html?wlc=1302480778

Miller, M. (June 30, 2008). Data Theft: How Big a Problem? In: *informIT,* Pearson Education, March 21, 2011, Available from . www.informit.com/articles/article.aspx?p=1220308

Ohlhorst, F. (February 10, 2010). Three encryption apps to keep your data safe - data encryption - PC World Business, In: *PC World Australia,* April 12, 2011, Available from http://www.pcworld.idg.com.au/article/335681/three_encryption_apps_keep_your_ data_safe/

Patel, H., Morrison, D. & Mischon De Reya, M. (n.d.). Information Theft: Are nervous employees sizing up your data? In: *KPMG: Cutting Through Complexity*. March 20, 2011), Available from http://www.datalossbarometer.com/14737.htm

Power, R. (2002). CSI/FBI computer crime and security survey. *Computer Security Journal,* Vol. 18, No. 2 (2002), pp. 7-30

PwC (2012). 2012 Global State of Information Security Survey, September 18, 2011, Available from http://www.pwc.com/gx/en/information-security-survey/key-findings.jhtml

Rhemann, M. (2011). "Cyber Trends" In: *Trends Digest,* September 11, 2011, Available from http://trendsdigeststore.com/CyberTrends.aspx

Richardson, R. (2008).CSI Computer Crime and Security Survey, *Computer Security Institute/Federal Bureau of Investigation* 2008, Available from http://gocsi.com/sites/default/files/uploads/CSIsurvey2008.pdf

Schwartz, M. (2010). The Mall in Your Pocket, *Gifts & Decorative Accessories Vol.* 111, No. 10 (2010) pp. 54-58

Schwartz, G., Shetty, N., & Walrand, J. (2010). Cyber-Insurance: Missing Market Driven by User Heterogeneity, *Submission to Workshop on the Economics of Information Security (WEIS),* February 2010, Available from http://www.eecs.berkeley.edu/~schwartz/missm2010.pdf

Siciliano, R. (February 15, 2011). Lost or stolen mobile devices can lead to identity theft, In: *McAfee Blog Central,* April 30, 2011, Available from http://blogs.mcafee.com/consumer/identity-theft/lost-or-stolen-mobile-devices-can-lead-to-identity-theft

Siciliano, R. (April 18, 2011). The Rise of Smartphones and Related Security Issues, In: *Infosec Island,* April 30, 2011, Available from https://www.infosecisland.com/blogview/13078-The-Rise-of-Smartphones-and-Related-Security-Issues.html

Siegel, C., Sagalow, T., & Serritella, P. (2002). Cyber-Risk Management: Technical and Insurance Controls for Enterprise-Level Security, *CRC Press, (*March 4, 2002), Available from http://www.eprivacy.com/lectures/cyber-risk.pdf

Siwicki, B. (September 28, 2010). Mobile raises new fraud risks for merchants, In: *Internet Retailer,* April 10, 2011, Available from http://www.Internetretailer.com/2010/09/28/mobile-raises-new-fraud-risks-merchants

Spicer, J., Aspan, M. (2011). More customers exposed as big data breach grows, *Reuters*, 3 April 2011, Available from http://news.yahoo.com/s/nm/20110403/bs_nm/us_citi_capitalone_data

Stroup, J. (n.d.). Business Identity Theft: Your Risks from Employees, In: *Identity Theft - What You Need to Know to Protect Yourself from Identity Theft,* March 21, 2011, Available from http://idtheft.about.com/od/businessidtheft/a/IDT_EEs.htm

The Economist. (2010). War in the fifth domain: Are the mouse and keyboard the new weapons of conflict? *The Economist Newspaper Limited, London,* July1, 2010, Available from http://www.economist.com/node/16478792

The White House. (2010). Fact Sheet for National Strategy for Trusted Identities in Cyberspace, *Office of the Press Secretary,* June 25, 2010, Available from http://www.whitehouse.gov/the-press-office/fact-sheet-national-strategy-trusted-identities-cyberspace

Walsh, J. (n.d.) What is data theft? In: *article pros.* March 20, 2011, Available from http://www.articlepros.com/computers_and_Internet/data_recovery/article-131141.html

15

Trust in an Asynchronous World: Can We Build More Secure Infrastructure?

Dragutin Vuković
INKUS Ltd.
Croatia

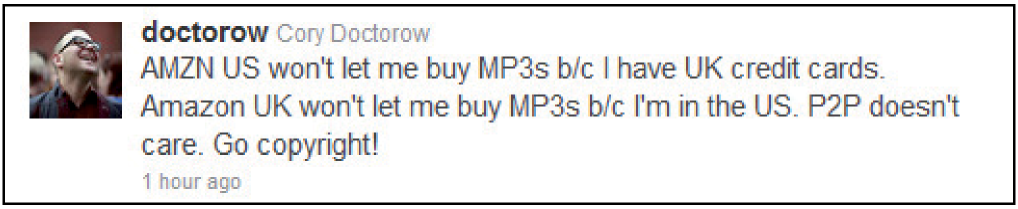
1. Introduction

Through history, many methods were designed and used for secure transfer of confidential information. During the WWII Allied Forces developed a method for securing phone conversation which included a pair of synchronized phonographs playing identical copies of white noise records. This method, called SIGSALLY, a.k.a. X System, Project X, Ciphony I or Green Hornet, used as secure speech system for the highest-level Allied voice communications (Fagen, 1978), is a very literal example of synchronicity because it would not work at all if there was no perfect synchronism maintained between phonographs at both ends of a voice channel. Other methods were devised for military communication by all participants, to mention only the famous German Enigma machine (Winkel, 2005). The common denominator of these historic methods is that all of them need a synchronicity in one way or another – synchronous keys, one-time-pads, etc. Therefore, when collaborating on matters that include a risk of information abuse, traditional approach calls for a synchronous procedure, that is, if there is a need to transfer some confidential information from one person to another, both persons will agree to meet at the same place, at agreed time. They will meet and authenticate each other by agreed procedure; third party may be involved to secure authentication. Information will be exchanged on agreed medium and participants will part assured that information is exchanged in a secure manner. This is an example of synchronous procedure, meaning that all participants have to be in contact at the same agreed time and the information is under control of a trusted party during the whole transaction. Synchronicity of the procedure is historically a prerequisite for establishing the trust relationship between participants of information exchange.

Procedures to transfer and store information in a secure manner are deployed in computing and communication networks in various forms and technologies, but they are basically all drawing on the same principle of synchronicity. With the development of Internet technologies, transfer and storage procedures are becoming more asynchronous. This means that not all parties involved are in contact at the same time, information could be, for a period of time, left into the custody of a party whose trust, and even authenticity, is not completely assured.

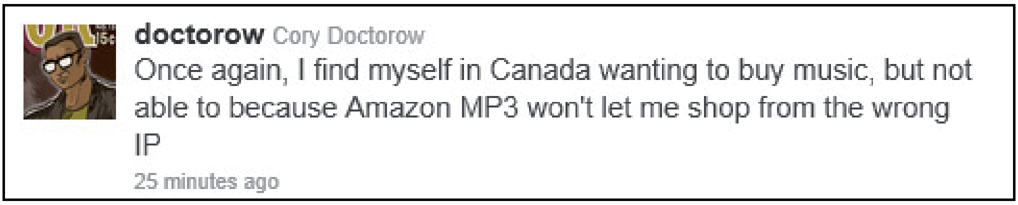
From this, we are witnessing many problems with exposed personal information such as stolen identity abuse, credit card fraud, not to mention the confidential information leakages from stolen computers and media not properly protected. Consequently, there is a diminution of customers' confidence in IT services and especially when provided through public network. Users want to be assured that their data are safe and secure with service provider. On the other side, service providers want to be assured that theirs services will be properly paid for. Some other confidence issues can be identified as well. Service providers are increasingly worried about risks their businesses are exposed in such environment, and they are building layer upon layer of information security technology to protect their customers' information as well as their shareholder value.

To illustrate this lack of confidence on the service providers' side, we will present recent 'tweets' by Cory Doctorow – famous writer, blogger and activist (see Figures 1 and 2).



doctorow Cory Doctorow
AMZN US won't let me buy MP3s b/c I have UK credit cards. Amazon UK won't let me buy MP3s b/c I'm in the US. P2P doesn't care. Go copyright!
1 hour ago

Fig. 1. Cory Doctorow's tweet on 2011-06-06T13:00



doctorow Cory Doctorow
Once again, I find myself in Canada wanting to buy music, but not able to because Amazon MP3 won't let me shop from the wrong IP
25 minutes ago

Fig. 2. Cory Doctorow's tweet on 2011-08-06T19:00

Cory Doctorow is a citizen of United Kingdom, born in Canada. During his travel to the USA, he wanted to buy some music from Amazon US, but Amazon US did not have confidence in Doctorow's UK credit cards. He then tried to buy music from Amazon UK, but Amazon UK did not have confidence in Doctorow's current location (i.e.: IP address) which happen to be in USA at the moment. Similar lack of confidence is observed again by Doctorow shortly later as he travelled to Canada.

Such lack of confidence, albeit frequent at current level of Internet technology development, is hindering further proliferation of on-line services, causing dissatisfaction of all parties involved, customers and service providers – customers because of inability to access services they are willing to consume and spend for, and providers because of potential markets being unutilised.

It is our belief that the cooperative behaviour in on-line services can be greatly improved by utilizing appropriate functions in the underlying infrastructure to foster trust and confidence between customers and on-line services providers. This belief is based on our conclusions inferred from the research in several fields spanning sociology (confidence, trust), technology (networking, communication, Internet technologies), and management (identity, information security):

- Confidence stems from trust, and trust is established between entities which are represented by their identities (Benantar, 2005), (Six, 2005).
- In computing and communications technology, the concept of trust is generally bound to reliability and dependability, which is a misconception in the sense of establishing the trust relationship between customers and on-line services regarding information security (Smith, 2005), (Serpanos, 2011).
- There are information technologies ready available and capable of supporting and enhancing trust relationship between various on-line systems but their deployment lacks systemic approach, i.e. they are utilized in specific services, thus multiplying non-trust boundaries which have to be handled on a per-case basis (Lerner *et al.*, 2002), (Mather, 2009).

These findings led us to the research question addressed in this chapter – could we envision a model for distributed computer system which would foster sociological notions of trust and confidence within the infrastructure? Model implementations would then utilize existing technologies and solutions in a systemic way to enforce establishment of stronger trust relationships between virtual digital entities, promoting confidence in on-line services regarding information security and enabling cooperation.

In this chapter we first discuss, in Section 2, the flow of information on the Internet and how it becomes more and more asynchronous with the proliferation of advanced technologies. Then we give an overview of risks involved due to asynchronous nature of data storage, transfer and processing in contemporary Internet technologies, in Section 3. Section 4 provides some insight about the notion of trust, its relation to confidence and its role in distributed computer systems. New and enhanced architecture of distributed computer systems, named *multilevel cell distributed computer system architecture*, to be utilized through the internet, is proposed in Section 5. Section 6 discusses, using trust-confidence-cooperation (TCC) model, how the cell architecture can provide enhancements to cooperation in on-line business. Some closing afterthoughts are given in Section 7.

2. Asynchronicity in modern communications

We infer from Section 1 that synchronicity is the underlying principle of security procedures in various areas, including information transfer and storage which is of main interest to us. Then, with new technologies in the IT age, synchronicity was sacrificed to achieve customer friendliness via speed, but this introduced asynchronous solutions, which is expanded on next.

Problem with synchronous procedures is that they spend time waiting for synchronization events to co-occur, adding to the overall length of the procedure. This was not the problem while the information transfer itself was comparably slow, so that overhead incurred by synchronization took only a small fraction of message duration. Shape of things has changed with the emergence of Internet. Nowadays, when high speed wired and wireless communication is omnipresent, everything is happening much quicker than before. Every part of world is accessible and digital information can be transferred with incredible speed. These speeds are made possible both by pure technological advances in electronic communication circuitry and by the fact that most data transfer technologies at physical, data link layers are asynchronous.

Our way of life changed accordingly. We learned to exchange large amount of information on a regular basis and we expect it to happen almost in real time. That is why we can send photos to our friends right from the field, and we expect their comments to come back to our smartphones in minutes. This can also be achieved despite geographical dispersion of people and varying frequency of when we meet. In order to enjoy this convenience we are ready to loosen our expectations regarding confidentiality and privacy of our information. Thus we entered asynchronous mode of operation, letting things happen more quickly, on the expense of some security issues.

On the other hand, when we come to information whose confidential nature we want to maintain during transfer and storage, we still at present stick to synchronous methods, having them proved successful before (perceived performance) and finding them acceptable among methods offered (value similarity). We will use cryptographic methods synchronized by exchanging keys, secure services bound by strong contracts, enforced by law. Or we may simply fall back to old-fashioned method of delivering information personally, possibly having it written only in messenger's memory, a communication channel with small bandwidth and large delay, having unreliable information storage.

Special category of information that we are interested in securing, are bound to transfer of financial value, such as electronic funds transfer, electronic money, credit cards, or anonymous debit cards (Androulaki, 2009), but this is a whole area on its own, which will not be discussed here although this area may also benefit from confidence and trust enhancements provided by architectural concepts discussed.

Reality is that we want more and more information to be transferred at higher and higher bandwidths, having smaller and smaller delays, with more and more confidence. The emergence of Web 2.0 applications and proliferation of cloud computing paradigm made our expectations only bigger, and asynchronicity more certain.

Cloud computing appears to have emerged very recently as a subject of substantial industrial and academic interest, though its meaning and scope, fit with respect to other paradigms is hotly debated. For some researchers, clouds are a natural evolution towards full commercialisation of grid systems, while for others they may be dismissed as a mere rebranding of the existing pay-per-use or pay-as-you-go technologies. From either perspective, it appears that 'cloud' has become the label of choice for accountable pay-per-use access to a wide variety of third-party applications and computational resources on a massive scale. Clouds are now supporting patterns of less-predictable resource use for applications, services across the IT spectrum, from online office applications to high-throughput transactional services and high-performance computations involving substantial quantities of processing cycles and storage. The current notion of clouds seems to blur the distinctions between grid services, web services, and data centres, amongst others, and brings considerations of lowering the cost for relatively bursty applications to the fore.

Cloud computing is a new way of delivering computing resources, not a new technology. Computing services ranging from data storage and processing to software, such as email handling, are now available instantly, commitment-free and on-demand. Since we are in a time of belt-tightening, this new economic model for computing has found fertile ground and is seeing massive global investment. According to IDC's analysis, the worldwide forecast for cloud services in 2009 will be in the order of $17.4bn (IDC 2009, as cited in

Catteddu, 2009). The estimation for 2013 amounts to $44.2bn, with the European market ranging from €971m in 2008 to €6,005m in 2013 (Bradshaw 2009, as cited in Catteddu, 2009).

There is probably no other field of research with as huge amount of literature than the field of modern data communication and Internet, especially when looking in publishing rate terms (number of titles per unit of time). Therefore it would be quite unwieldy to produce a thorough overview of literature in this field. Here are some pointers to titles that could provide good starting point for interested readers to investigate further in the field:

- Data transfer and communication technologies as building substance of Internet: (Lerner *et al.*, 2002), (Governor, 2009), (Sobh *et al.*, 2010), (Sorensen, 2010), (Preve, 2011), (Serpanos and Wolf, 2011).
- Middleware and Internet: (Lerner *et al.*, 2002), (Puder *et al.*, 2006), (Toninelli *et al.*, 2011), (Georgantas *et al.*, 2011).
- Contemporary technologies and paradigms in Internet – Web 2.0, grid and pervasive computing, cloud computing: (Mattern, 2006), (Puder *et al.*, 2006), (Reese, 2009), (Governor, 2009), (Rittinghouse and Ransome, 2010), (Zheng, 2010), (Zagalo *et al.*, 2011).
- Identity and privacy in distributed systems: (Benantar, 2005), (Windley, 2005), (Waldo *et al.*, 2007), (Nin and Herranz, 2010), (Sileo, 2010), (Papacharissi, 2011).

With this in mind, we are ready to probe the risks in the asynchronous world, which is done next in Section 3.

3. Risks in an asynchronous world

According to analyst firm Gartner, cloud computing is fraught with security risks (Brodkin, 2008). Smart customers will ask tough questions and consider getting a security assessment from a neutral third party before committing to a cloud vendor. Cloud computing has "unique attributes that require risk assessment in areas such as data integrity, recovery, and privacy, and an evaluation of legal issues in areas such as e-discovery, regulatory compliance, and auditing," Gartner says.

According to Gartner, before selecting a cloud vendor, customer should raise seven specific security issues (Brodkin, 2008):

1. **Privileged user access.** Sensitive data processed outside the enterprise brings with it an inherent level of risk, because outsourced services bypass the physical, logical and personnel controls IT shops exert over in-house programs. Get as much information as you can about the people who manage your data. Ask providers to supply specific information on the hiring, oversight of privileged administrators, and the controls over their access.
2. **Regulatory compliance.** Customers are ultimately responsible for the security and integrity of their own data, even when it is held by a service provider. Traditional service providers are subjected to external audits and security certifications. Cloud computing providers who refuse to undergo this scrutiny are signalling that customers can only use them for the most trivial functions.
3. **Data location.** When you use the cloud, you probably won't know exactly where your data is hosted. In fact, you might not even know what country it will be stored in. Ask providers if they will commit to storing and processing data in specific jurisdictions, and whether they will make a contractual commitment to obey local privacy requirements on behalf of their customers.

4. **Data segregation**. Data in the cloud is typically in a shared environment alongside data from other customers. Encryption is effective but isn't a cure-all. Find out what is done to segregate data at rest. The cloud provider should provide evidence that encryption schemes were designed and tested by experienced specialists. Encryption accidents can make data totally unusable, and even normal encryption can complicate availability.
5. **Recovery.** Even if you don't know where your data is, a cloud provider should tell you what will happen to your data and service in case of a disaster. Any offering that does not replicate the data and application infrastructure across multiple sites is vulnerable to a total failure. Ask your provider if it has the ability to do a complete restoration, and how long it will take.
6. **Investigative support.** Investigating inappropriate or illegal activity may be impossible in cloud computing. Cloud services are especially difficult to investigate, because logging and data for multiple customers may be co-located, may also be spread across an ever-changing set of hosts and data centres. If you cannot get a contractual commitment to support specific forms of investigation, along with evidence that the vendor has already successfully supported such activities, then your only safe assumption is that investigation and discovery requests will be impossible.
7. **Long-term viability.** Ideally, your cloud computing provider will never go broke or get acquired and swallowed up by a larger company. But you must be sure your data will remain available even after such an event. Ask potential providers how you would get your data back, if it would be in a format that you could import into a replacement application.

In the paper edited by Daniele Catteddu and Giles Hogben, (Catteddu, 2009) published by The European Network and Information Security Agency (ENISA) in the context of ENISA's Emerging and Future Risk programme, a group of selected industry, academic and government experts in the subject area, expressed their opinions about benefits, risks and recommendations for information security in cloud computing. Experts identified a number of cloud specific risks, the most important of which we will enumerate here:

1. **Loss of governance.** In using cloud infrastructures, the client necessarily cedes control to the cloud provider on a number of issues which may affect security. At the same time, service level agreements may not offer a commitment to provide such services on the part of the cloud provider, thus leaving a gap in security defences.
2. **Lock-in.** There is currently little on offer in the way of tools and procedures or standard data formats or services interfaces that could guarantee data, application and service portability. This can make it difficult for the customer to migrate from one provider to another, or migrate data and services back to an in-house IT environment. This introduces a dependency on a particular cloud provider for service provision, especially if data portability, as the most fundamental aspect, is not enabled.
3. **Isolation failure.** Multi-tenancy, shared resources are defining characteristics of cloud computing. This risk category covers the failure of mechanisms separating storage, memory, routing, even reputation between different tenants (e.g., so-called guest-hopping attacks). However it should be considered that attacks on resource isolation mechanisms (e.g., against hypervisors) are still less numerous and much more difficult for an attacker to put in practice compared to attacks on traditional OSs.

4. **Compliance risks.** Investment in achieving certification (e.g., industry standard or regulatory requirements) may be put at risk by migration to the cloud:
    a. if the cloud provider cannot provide evidence of their own compliance with the relevant requirements
    b. if the cloud provider does not permit audit by the cloud customer.

    In certain cases, it also means that using a public cloud infrastructure implies that certain kinds of compliance cannot be achieved.

5. **Management interface compromise.** Customer management interfaces of a public cloud provider are accessible through the Internet, mediate access to larger sets of resources (than traditional hosting providers), therefore pose an increased risk, especially when combined with remote access and web browser vulnerabilities.
6. **Data protection.** Cloud computing poses several data protection risks for cloud customers and providers. In some cases, it may be difficult for the cloud customer (in its role as data controller) to effectively check the data handling practices of the cloud provider, thus to be sure that the data is handled in a lawful way. This problem is exacerbated in cases of multiple transfers of data, e.g., between federated clouds. On the other hand, some cloud providers do provide information on their data handling practices. Some also offer certification summaries on their data processing and data security activities, the data controls they have in place, e.g., SAS70 certification.
7. **Insecure or incomplete data deletion.** When a request to delete a cloud resource is made, as with most operating systems, this may not result in true wiping of the data. Adequate or timely data deletion may also be impossible (or undesirable from a customer perspective), either because extra copies of data are stored but are not available, or because the disk to be destroyed also stores data from other clients. In the case of multiple tenancies, the reuse of hardware resources, this represents a higher risk to the customer than with dedicated hardware.
8. **Malicious insider.** While usually less likely, the damage which may be caused by malicious insiders is often far greater. Cloud architectures necessitate certain roles which are extremely high-risk. Examples include cloud provider system administrators and managed security service providers.

The risks listed above do not follow any specific order; they are just some of the most important cloud computing specific risks identified during the assessment. ENISA's report delves further into a more detailed analysis of specific risks in several categories such as policy, organizational risks, technical risks, legal risks and some risks not specific to the cloud.

ENISA's and Gartner's reports, while partly overlapping, also complement each other and together give a fair overview of risks that both, cloud service users and cloud service providers, may expect to face in the course of building a cloud economy.

Ironically, books on security sell poorly, whereas books on hacking into systems sell much better, a trend that is worrying when taking into account the increasing magnitude of these problems. Here is a sampling of titles that can be of use to interested reader who wants to extend his knowledge into the field of information security, especialy in areas discussed here:

- Challenges, approaches and solutions to risks of information security in computing systems: (Fagen, 1978), (Lerner *et al.*, 2002), (Cranor, 2005), (Winkel *et al.*, 2005), (Biskup, 2009), (Viega, 2009), (Wong and Yeung, 2009), (Arata, 2010), (Graham *et al.*, 2011), (Andress, 2011).
- Information security issues in contemporary distributed computing systems – Web 2.0., grid, cloud: (Mather, 2009), (Rittinghouse and Ransome, 2010), (Thuraisingham, 2011).
- Managing information security and associated risks: (Broder, 2006), (Tipton and Krause, 2008), (Arata, 2010), (Aven and Renn, 2010), (Vladimirov *et al.*, 2010), (Brotby, 2009), (Tiller, 2011).
- Identity, privacy and access control: (Benantar, 2005), (Windley, 2005), (Waldo *et al.*, 2007), (Mather, 2009), (Arata, 2010), (Nin, 2010), (Sileo, 2010), (Papacharissi, 2011).

While being aware of risks in itself is pointless unless we can do something about it, it follows that we must manage these risks. However, due to the asynchronous nature of the problem at hand we not only have to manage the risks, but we also have to build confidence in that the risks *are* being managed and that the transaction is trustworthy in itself if we follow certain rules including risk management. This, and more, is discussed in Section 4.

4. Trust and risk management

In his influential book (Fukuyama, 1995), Francis Fukuyama argued that public values, especially trust, shape the direction of national economies. Among other things, Fukuyama shows how trust reduces transactions costs, and ultimately, economic friction. Here, we will first give a brief overview of trust and its relation to confidence as it is seen from sociological standpoint. Then we will proceed to discuss the ways trust is seen in the realm of information technology. This would establish a background for thinking about a better use of trust and confidence concept to mitigate some risks in distributed computer architecture.

4.1 The importance of trust

Many authors have emphasized the importance of trust for achieving organizational success. The overview presented in (Six, 2005) shows that many see trust as necessary in contexts of high ambiguity and uncertainty, as well as in contexts of high complexity. Trust, on the one hand, can provide a sense of security that will help survival in these contexts, and on the other, it can help with the risk taking necessary for survival in complex environments. Trust, when present, is said to enhance the ability to change and to support, potentially radical, change. This is because trust is said to assist in learning, creativity and innovation. Furthermore, it is a lubricant for social relations which improves efficiency or, as John Locke declared, trust is 'the bond of society', the *vinculum societatis*.

Trust is also seen to foster and maintain cooperation, as it encourages information sharing, enriches relationships, increases openness and mutual acceptance, and enhances conflict resolution and integrative problem solving. The presence of trust, it has been argued, reduces the need for detailed contractual and monitoring devices, is thus important in governance and taking it one step further, in complex environments, detailed contracting and monitoring are often undesirable since they may constrain the scope and motivation for quality, for innovation based on individual variety and initiative. Trust can have extrinsic value, as a means to achieve social or economic goals, and it can have intrinsic value, as a dimension of relations that is valued for itself, as part of a broader notion of well-being or the quality of life. People may prefer, as an end in itself, to deal with each other on the basis of trust. One motive for doing this is to build confidence, which is discussed next.

4.2 Trust and confidence

We define trust as the willingness, in expectation of beneficial outcomes, to make one vulnerable to another based on a judgement of similarity of intentions or values, but here we want to emphasize that trust is based on social relations, group membership and shared values. Confidence is defined as the belief, based on experience or evidence, that certain future events will occur as expected. Both trust and confidence support cooperation. But whereas confidence has a specific performance criterion, trust is placed in the freedom of the other. In the case of trust, the other is free to act in ways that indicate shared values, regardless of whether specific acts are expected or not. In the case of confidence, the other must act in specifically expected ways.

The crucial point that is generally overlooked is the dependence of confidence on trust (O'Neill, 2004 as cited in Siegrist, 2007). We all describe within communities; we can't do otherwise. However, since our descriptions are linked to our communities, and accepted as justified only within them, we normally are not made aware of the dependence of the one upon the other – and the potential rejection of our descriptions within other communities. To take a very simple example, one might claim that one's confidence that the Earth will circle the sun is not based on a relation of trust. But one only has the idea of 'the Earth circling the sun' as a consequence of one's membership in a particular community. There is nothing given about that or any other description. This, of course, can become a serious matter when the descriptions one makes provoke more variable, contested effects on others than a description of the Earth's relation to the sun.

4.3 Trust and distributed computer system architecture

Within the realm of technology, trust and control have usually been associated with reliability, integrity (Smith, 2005) and correctness (Jayaswal and Patton, 2006) and were not seen as a separate issue until the arrival of complex computer-controlled systems. Computer science had initially approached trust, control from the perspective of security. Recognising that trust is not controllable, the security developed an elaborate structure of control, in an attempt to minimise elements of trust. However, more recently, the recognition of the fundamental nature of trust has been addressed in initiatives such as trusted computing, where individual devices are given assurance in their own configuration on the basis of a hardware-based root of trust. The need for a portable root of trust has also fuelled the creation and popularity of smart cards (Cofta, 2007).

In data communication, the understanding that trust precedes meaningful and secure communication has eventually led to the concept of trust management, the separate layer of interactions that lead to the creation and maintenance of trust relationships between communicating nodes, following e.g. business agreements, contractual dependence, personal relationship, etc. Pretty Good Privacy (PGP) has been exploring the area of peer-to-peer trust while Public Key Infrastructure (PKI) proposed the multi-stage model of trust (Biskup, 2009). More recently, Web Services Trust language (WS-Trust) has established itself as a standard within Service-Oriented Architecture (SOA), the potential foundation of Web 2.0 (Thuraisingham, 2010). Grid computing and pervasive computing environment have brought different challenges to trust management.

The need to effectively manage distributed computing systems has led to constructs such as trusted domains (several computers trusting each other's authentication capabilities) (Rittinghouse, 2010), trusted credentials (others' identities accepted without any further proof), trusted storage (storage space accessible only to selected users), trusted zones (privileged Internet address space) etc. In all these cases there is a notion of trust as essential yet different from actual cooperation or communication, something that requires special management practices. Usually, the ability to manage trust is granted to system administrators or users, in the expectation that the technical structure of trust will reflect trust in respective social relationships. Research on autonomous agents has liberated trust management from the need for an *a priori* trust, managed by the user or the administrator. Agents were vested with the ability to make and break the trust relationship (that can be more correctly called 'the relationship of confidence'), usually on the basis of past experience and through the process of learning, whether from direct interactions or from others' experience. Autonomous agents have brought the notion of imperfect trust (where trust is no longer a binary proposition), the problem of trust propagation and reasoning. The new approach to trust has also, unfortunately, revealed new threats to trust, usually in the form of attacks on reputation.

Interest in large systems, whether created by autonomous agents, *ad-hoc* networks or in any other way, required more specific instruments to discuss the reasoning about trust. Formalisation of trust proposes logical primitives, schemes that can be used in reasoning about trust. The formalisation of reasoning has led to the creation of several formal systems and supporting tools. Both reasoning and transitivity require trust and confidence to be qualified. The desire to measure trust and confidence generated significant amount of research.

From a more application-specific perspective, electronic commerce has used various metrics of trust to develop risk assessment, both for the seller, for the buyer. The commercial value of eBay's reputation system is widely known, and similar rating systems are used by other e-commerce sites. Collaborative filtering has been used to aid information search following the concept that trust is a qualified reliance on information, but as more automated systems moved into the area, collaborative filtering became the preferred solution for the recommendation. The needs of electronic commerce have stimulated the interdisciplinary approach to trust.

Another effect of the introduction of electronically mediated communication is the development of research in user trust in digital devices, e.g. in a form of web features that facilitate the creation of perceived trust, trust in information systems or in improvements of trust between people while communicating through a digital channel.

In a distributed computer system the establishment of trust typically includes specific administrative permissions and leverages cryptographically secure methods. These methods can establish identities, and provide various secure services to managed users. The full use of network services is reserved for managed users. These users have an identity on the network and are therefore trusted to interact with their piece of the network, which we will call 'cell', an organized collection of networked computers. By authenticating itself to the network and continually validating its authenticated status, this cell may become a trusted member of a network, which is also a cell, built on the same blueprint as the lower level cell.

Interested readers can find additional insight into the area of trust and confidence in the following literature:

- Intrinsic value of trust: (Blau, 1964), (Bradach, Eccles, 1989), (Gulati, 1995), (Nooteboom, 1996), (Powell, 1996), (Ryan, Oestreich, 1998), (Sako, 1998), (Marek, 2008), (Briggs, 2010).
- Trust is necessary in contexts of high ambiguity, uncertainty, and in contexts of high complexity: (Lewis, Weigert, 1985), (Shapiro, 1987), (Nooteboom, 1996), (Shaw, 1997), (Deering, Murphy, 1998), (Lane, 1998), (Nahapiet, Ghoshal, 1998), (Rousseau *et al.* , 1998), (Sako, 1998), (Senge *et al.*, 1999), (Costa, 2000), (Marek, 2008).
- How trust can provide a sense of security which will help survival in contexts of high ambiguity, uncertainty, and in contexts of high complexity: (McAllister, 1995), (Ellinor, Gerard, 1998), (Ryan, Oestreich, 1998), (Reina, Reina, 1999), (Senge *et al.*, 1999).
- How trust can help with risk taking necessary for survival in contexts of high ambiguity, uncertainty, and in contexts of high complexity: (Katzenbach *et al.*, 1995), (Shaw, 1997), (Lewis, 1999), (Reina, Reina, 1999), (Senge *et al.*, 1999), (Costa, 2000), (Marek, 2008).
- How trust enhances ability to change, supports radical change: (Argyris, 1970), (Katzenbach et al., 1995), (Shaw, 1997), (de Geus, 1997), (Deering, Murphy, 1998), (Ellinor, Gerard, 1998), (Ryan, Oestreich, 1998), (Reina, Reina, 1999), (Senge *et al.*, 1999), (Costa, 2000), (Marek, 2008).
- How trust assists in learning, creativity, innovation: (McAllister, 1997), (Shaw, 1997), (Zand, 1997), (Deering, Murphy, 1998), (Lane, 1998), (Lazaric, Lorenz, 1998), (Nahapiet, Ghoshal, 1998), (Rousseau *et al.*, 1998), (Ryan, Oestreich, 1998), (Sako, 1998), (Ghoshal, Bartlett, 1999), (Lewis, 1999), (Reina, Reina, 1999), (Senge *et al.*, 1999), (Costa, 2000), (Marek, 2008).
- Trust as a lubricant for social relations which improves efficiency: (Blau, 1964), (Fukuyama, 1995), (Hosmer, 1995), (Deering, Murphy, 1998), (Hollis, 1998).
- Trust fosters, maintains cooperation, as it encourages information sharing, enriches relationships, increases openness, mutual acceptance, enhances conflict resolution, integrative problem solving: (Shapiro, 1987), (Katzenbach *et al.*, 1995), (Mayer *et al.*, 1995), (Ross, LaCroix, 1996), (Wheatley, Kellner-Rogers, 1996), (Shaw, 1997), (Zand, 1997), (Deering, Murphy, 1998), (Elangovan, Shapirio, 1998), (Lane, 1998), (Rousseau *et al.*, 1998), (Ryan, Oestreich, 1998), (Tsai, Ghoshal, 1998), (Whitener *et al.*, 1998), (Zaheer *et al.*, 1998), (Ghoshal, Bartlett, 1999), (Lewis, 1999), (Reina, Reina, 1999), (Senge *et al.*, 1999), (Costa, 2000), (Marek, 2008), (Cofta, 2007).
- Trust applied to data transfer, storage in public networks and information systems: (Robinson *et al.*, 2005), (Lu and Tsudik, 2010), (Dong and Dulay, 2010), (Kellermann *et al.*, 2010), (Ma *et al.*, 2010), (Cornelis and De Cock, 2011).
- Trust, information security and risk management: (Benantar, 2005), (Siegrist, 2005), (Veeningen *et al.*, 2010), (Muller, 2010), (Khoury and Tawbi, 2010), (Crampton, 2010), (Ballardin and Merro, 2010), (Kamil and Lowe, 2010).

Next, we propose a concept of *multilevel cell distributed computer system* architecture capable of providing enhanced trust services and increased confidence to both cloud providers and cloud consumers, thus providing a foundation for future development of cooperation on global network, with better management and mitigation of risks.

5. Multilevel cell distributed computer system architecture

In the realm of distributed computer systems, such as Internet, trust and confidence are relations that could be established between digital, virtual entities inhabiting this realm. Virtual entities could represent real persons as well as technical resources (services). In order to make use of trust and confidence to promote better cooperation, we need a reliable identity management system capable of collecting, interpreting and representing social information about virtual entities, among other things. Social networks are building on this idea but they tend to centralize information. Every network builds its own set of information about entities, which gives rise to the kind of problems explained in the introduction of this chapter. Also, they lack appropriate interpreting functions that could provide measures of social trust needed to establish confidence.

To address such issues we devised a *cell* as a basic building block of distributed computer systems infrastructure. Cell is a self-contained computer system with clearly established boundaries, capable of communicating with other such systems. The smallest cell could be a single physical computer, although it can host several virtual entities, persons or services. A cell can be built from several computers dedicated to various cell functions. Several cells can be connected together to form a larger entity which represents itself as a cell to outer world, providing the same cell functions. It is thus possible to build a *multilevel cell distributed computer system*.

Cells need to provide many functions for their operation and cooperation within a distributed system. These functions have to be built upon certain design principles we established as a foundation to proposed architecture. We will here mention only those principles (Lerner *et al.*, 2002) that support our discussion regarding trust and confidence:

- computers associated with a cell direct all traffic between entities to flow unconditionally through its edge gateways (interconnection computers);
- all entities that establish a relationship with a cell must register with the cell;
- all traffic passing through the cell must be authenticated;
- the cell tracks all active entities.

Cell-based distributed system architecture is a development from the architectural concept based on communicating proxies, as described by (Lerner *et al.*, 2002). This is expanded on next.

5.1 Physical cell architecture

Although all functions can be implemented in single computer, it is recommended to implement different functional elements into different computers. Having this in mind we will call these computers, similarly: *access computer*, *interconnection computer*, *storage computer* and *service computer*. These computers comprise building blocks of the cell's internal architecture, as shown by Figure 3.

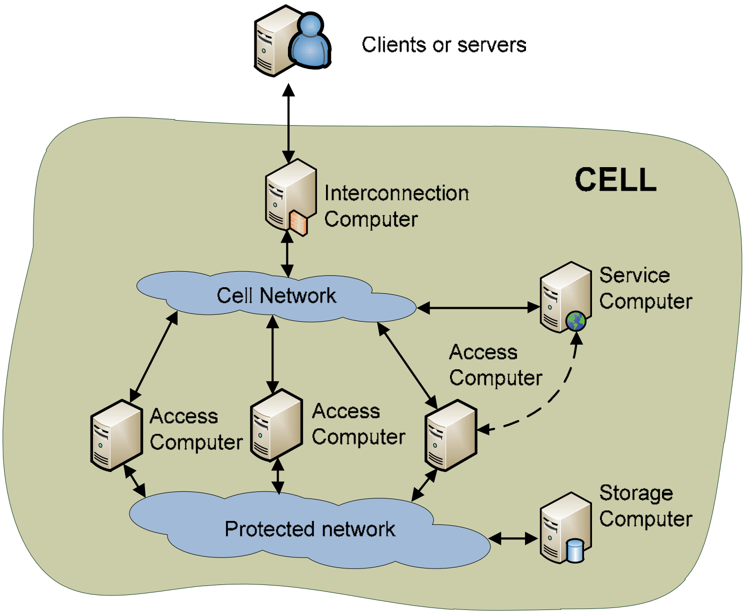


Fig. 3. Basic cell architecture

Interconnection computers enable communication with other networks and provide security boundary for the cell. They also run several other security related tasks such as registration, identification, authentication, accreditation, encryption/decryption, etc.

Service computers are, by their functions, general servers. They provide services to other computers, internal or external to the cell. Service computer can be connected to the cell network or directly to access computer.

Access computers are, by their function, proxy computers with some extended functionality. They are characterized by having at least two network interfaces. One interface connects access computer into the protected network that enables them to communicate with each other. Other interfaces connect access computers with external computers (clients, servers or other cells) through the cell network and interconnection computer.

Storage computers are essentially database servers. They provide services, related to databases stored on them, to other cell computers, but only through access computers. Therefore data stored in the cell are not directly reachable by the external computers.

Internally, cell has two networks: *cell network* and *protected network*. Cell network is semi-public network into which external computers (clients, servers or other cells) can enter through interconnection computer, so as to gain access to service computer or access computer. Interconnection computers and cell's security services protect this network from unauthorized access by external systems.

Protected network is unreachable for external computers, and only access computers can connect into it, in order to access the cell's data contained there on the storage computers which are also attached to the protected network. No traffic generated externally to the cell is allowed into this network.

We can also look at the physical cell architecture from a functional point of view. This view will show us four functional elements: interconnection, access, storage and service. Remember that a cell could be hosted on the single computer so these functional elements need not be implemented on physically different devices.

Functionally, we can have more views of the cell. Let's look at the cells placed at various levels of the distributed computer system hierarchy. Depending of their functional level, a cell will have different emphasis on various functions within it, as discussed in the next section.

5.2 Functional cell levels

Basic functional cell level is network. Cell in this level is called *network cell*. Network cell is based on locality of computers which constitute the cell. It is supposed that computers in the network cell all are contained in single building or several closely placed buildings, interconnected by private, physically secured network. Network cell centralizes functions such as traffic management, network management, quality of service management, messaging, etc.

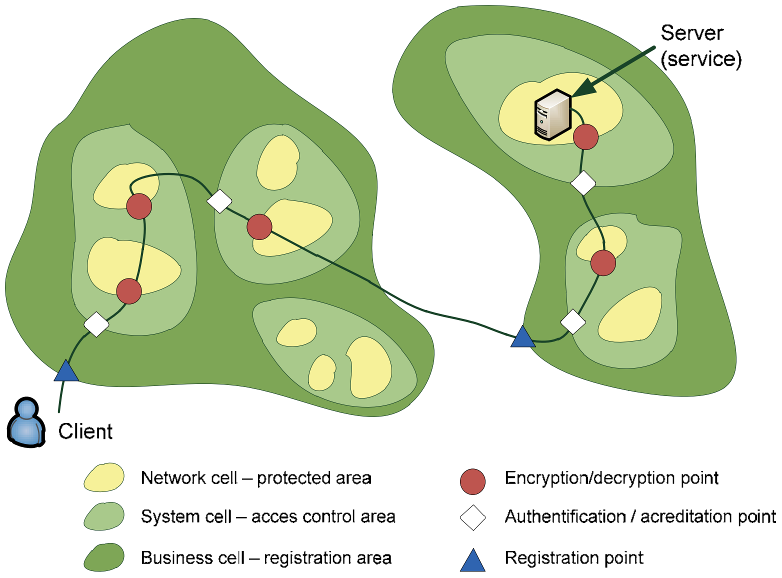

Server
(service)
Client
Network cell – protected area
System cell – acces control area
Business cell – registration area
Encryption/decryption point
Authentification / acreditation point
Registration point

Fig. 4. User query traversing a multilevel cell distributed computer system

Network cells are being connected together to form a *system cell*. If communication between network cells is not local, protected, there should be encryption/decryption function built into interconnecting computers. System cell centralizes functions of user identification, authentication, and accreditation.

To implement connection function efficiently, it should be founded on three general principles: *knowledge abstraction*, *lazy calculation*, and *multiplication of data about clients and servers*. All three principles are supported by cell architecture described here.

Knowledge abstraction supposes construction of abstract model for unified representation of servers and clients in our computer system. In the first step, the notion of server (physical entity providing a certain service) is replaced by more abstract term service (service itself) which cloaks physical characteristics of server computer. Thus, instead of client/server architecture, we are considering client/service architecture. This abstraction is especially convenient in modelling of distributed systems based on non-connection protocols. In such systems a client addresses a service, not a physical object. It has to be noted also that distinction between client and service is temporary and lasts only through single transaction. In the very next transaction roles could be reversed. Using knowledge abstraction we are able to disregard this difference at the model level.

Lazy calculation of object characteristics supposes that neither all objects, nor all their attributes, are at every moment present locally in all cells of distributed architecture. Only when objects' characteristic is explicitly referenced the system will contact other cells to retrieve the needed information. This principle is supported by institution of global catalogue which collects all objects but only with some subset of attributes.

Multiplication of data about clients and servers is achieved by partitioning and replicating objects into other cells. Replication topology and schedule should be designed carefully to optimize network traffic.

Cell architecture also supports efficient network traffic management, based on data replication as well as multiplication and distribution of functions. Network cell manages and monitors traffic, sends massages to computers or group of computers, measures and allocates bandwidth, etc. System cell replicates system and server data, and distributes control and administrative jobs. Business cell redirects network traffic, based on information in registration databases, internally or towards other cells.

While majority of connection, traffic management functions are performed by network and system cells, fundamental business cell purpose is to provide support to information system By executing many of monitoring, security and administrative functions, business cell simplifies server and client operation and makes their connection efficient.

5.3 Functional cell architecture

A cell is a fundamental building block of a distributed computer system. Single cell by itself also represents a distributed computer system, so it has to contain all functionalities of a whole distributed system. Thus by describing single cell functionality we are describing the functionality of the whole proposed distributed system.

Cell's basic functional structure is shown by Figure 5. In relation to Figure 3, showing types of networks and computers comprising a cell, Figure 5 emphasizes systems and services that implement cell functionality. Functions and services can be situated on one or more computers shown on Figure 3, and vice versa – several functions can be provided by single computer. Typical function and computer mapping between Figure 3 and Figure 5 might be as follows:

- *Interconnection computers* implement *security perimeter control* function (they are essentially firewalls);

- *Service computers* implement most of system functions – *identity management, certificate, keys management, business process orchestration, service publishing, message transfer, management, documents transfer, management, mobile services, cell management;*
- *Access computers* implement *application services;*
- *Storage computers* implement *database services*, as well as *legacy applications* where appropriate.

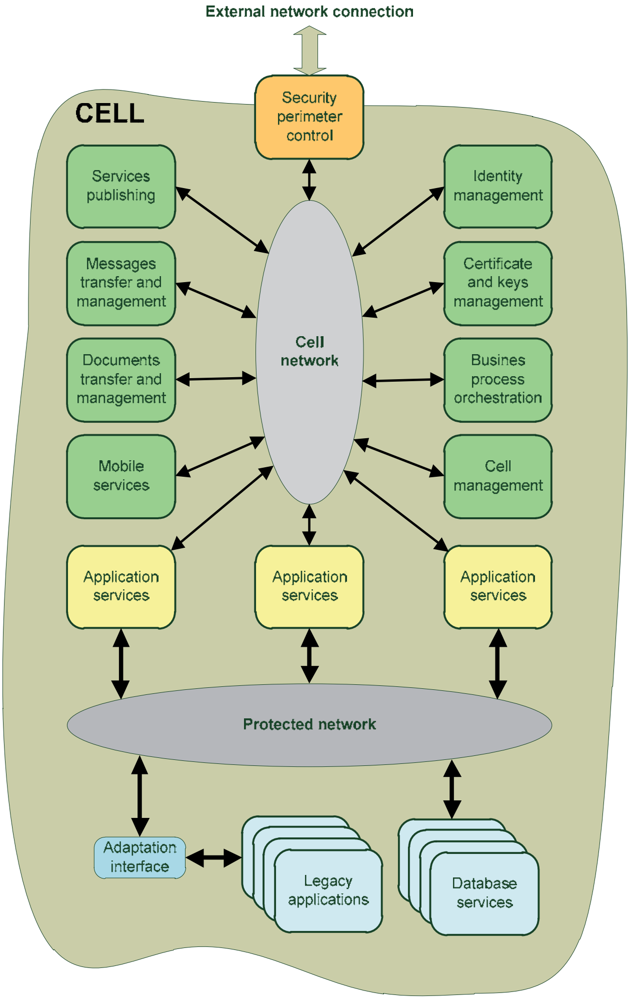


Fig. 5. Functional cell architecture

The key cell function providing support to promote trust in our proposed architecture is *identity management*. In a 'real world' an identity is any subset of attribute values of an individual person which sufficiently identifies this individual person within any set of persons. In cell architecture term 'identity' is extended to every identifiable object within distributed computer system, which may include persons as well as services. *Digital identity* is a digital representation of a 'real world' object. Digital identity is implemented as data structure characterizing the entity.

Besides basic data necessary for addressing (location and identification) identity can encompass data for other purposes such as authentication, accreditation, administration, brokerage, certification, provisioning, etc. In our model, identity data also include components that collect trust related information.

Cell's identity management is a framework within which digital identities are managed in the distributed computer system. Digital identity management comprises two parts: a) digital identity repository (directory) implemented as distributed, partitioned and replicated database, and b) set of operations on identity data. Full set of digital identity data is situated in a cell where there is an ownership relation on them, either original cell where identity was created or a cell where identity ownership was transferred.

Identity management subsystem is a hierarchical set of functions (Vukovic, 2011), consisting of:

- **Identity administration** – establishes digital identity's compliance with business processes:
  - *Existence* – digital identity establishment within the cell;
  - *Context* – managing information about digital identity's working environment; collection of trust related information belongs here;
  - *Provisioning* – connects digital identity dynamically with various tools administered to it through its life cycle.
- **Community management** – promotes cooperation between various groups of digital identities, i.e. digital communities:
  - *Authentication* – checks and confirms authenticity of digital identity's data, supports privacy and confidentiality;
  - *Authorization* – gives permission to access and utilize distributed resources and services; uses trust related data to determine level of confidence;
  - *Rendezvous* – establishes an appropriate level of cooperation between digital identities based on calculated trust and confidence information.
- **Identity integration** – makes a holistic view of identity data and establishes a consistent digital identity representation for other cells:
  - *Ownership* – while subsets of identity data may be replicated throughout distributed system, ownership of original data must be maintained and remain under sole control of primary owner;
  - *Brokerage* – exchanges information about digital identities between cells; collects trust related data when applicable;
  - *Connection* – establishes connections between cells for the purpose of identity data exchange.

Subset of digital identity data is available to other cells, depending on a level of trust needed for a digital identity to interact with cell's services. This subset of data is a digital identity's *virtual presentation*. Cell's identity management function keeps this subset at a necessary minimum to optimize privacy and minimize traffic.

6. Cooperation in multilevel cell distributed computer systems

In this section we will discuss the idea, stated in the introduction, that multilevel cell architecture can leverage trust relationship building between virtual entities on Internet, thus raising the level of confidence, resulting in more cooperation (i.e. business). For this purpose we will use the Trust, Confidence and Cooperation (TCC) model, described in (Siegrist, 2007).

The TCC model is designed to serve several useful purposes. The first is unification. It provides a framework within which all expressions of trust and confidence can be interpreted, related to one another. The second is specification. To a greater extent than available alternatives, it identifies the basic psychological processes involved in judgements of trust and confidence. The third is clarification. At the centre of the TCC model is an explicit account of the interaction between trust and confidence, a major source of confusion in other approaches. The final purpose is generation of new insights. By unifying and bringing more specificity and clarity to the understanding of trust and confidence, the TCC model points to potentially fruitful connections with other areas of social, psychological and applied research, and suggests novel research hypotheses.

The TCC model of cooperation postulates that trust is based on social relations and on shared values. Shared values can be measured in many different ways. In empirical studies, trust can be indicated variously by measures of in-group membership, morality, benevolence, integrity, inferred traits, intentions, fairness and caring. All of these, we will argue, can be taken to mean good intentions relative to those of the trusting person shared values.

As defined in (Siegrist, 2007) the model identifies constituent (in-coming) and product (out-going) elements, but does not specify how the former are combined to produce the latter. This allows for model to be mapped to functions of various systems, provide a basis for evaluation of how these functions contribute to cooperation. Elements of TCC model, as described in (Siegrist, 2007) are the following:

1. *Perceived amplitude of morality/performance information*: the judged degree to which the given information has morality/performance implications.
2. *Perceived valence of morality/performance information*: the judged degree of positivity/ negativity of the given information.
3. *Attributed values/performance*: the values/performance attributed by the observer to the other.
4. *Active values/active performance history*: in the case of values, these are the values that are currently active for the observer – which may be the product of existing social trust relations. In the case of performance, this is whatever history of relevant performance that is currently active for the observer.
5. *General trust/general confidence*: general trust is defined and discussed in previous sections. General confidence is the performance-based counterpart of the values-based general trust: the belief that things, in general, are under control, uncertainty is low and events will occur as expected.
6. *Value similarity/perceived performance*: value similarity is the judged similarity between the observer's currently active values and the values attributed to the other. Perceived performance is the observer's interpretation of the other's performance.
7. *Social trust/confidence*: these elements are defined and discussed in previous sections.
8. *Cooperation*: any form of cooperative behaviour between a object and another object or group of objects.

The application of the TCC model to evaluate the ability of cell architecture to enhance cooperation by utilizing trust and confidence information into identity management, is shown in Figure 6. The elements of the TCC model are aligned in parallel pairs for trust and confidence. Upper row of elements deals with trust, and lower row of elements deal with confidence. Both rows are combined at the right to produce a cooperation of a certain level.

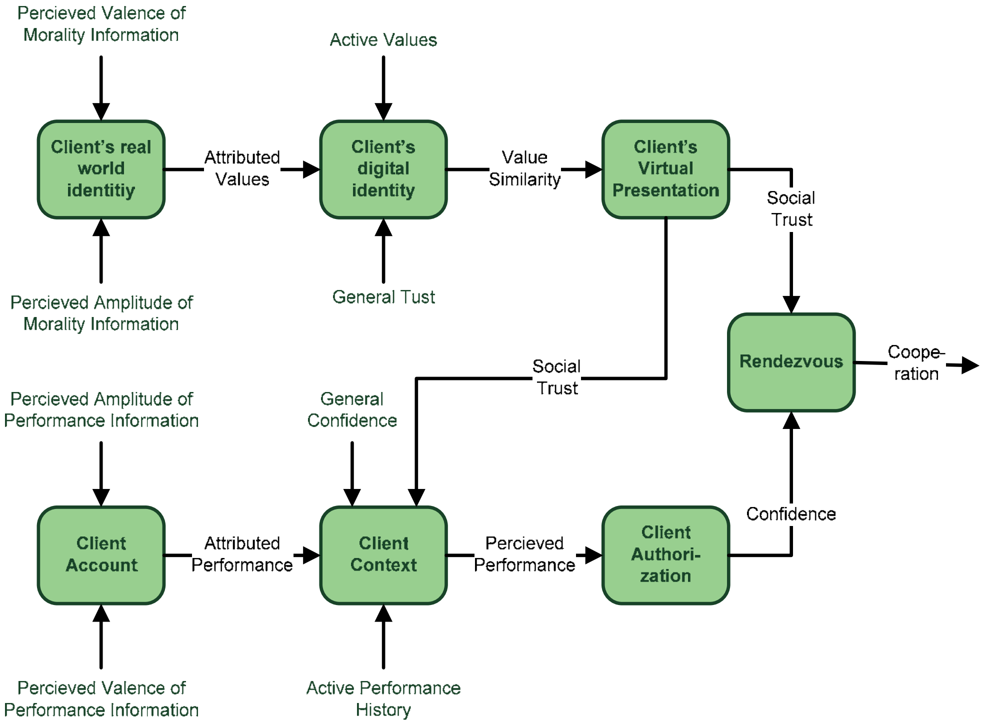

Percieved Valence of Morality Information
Active Values
Client's real world identitiy
Attributed Values
Client's digital identity
Value Similarity
Client's Virtual Presentation
Social Trust
Percieved Amplitude of Morality Information
General Tust
Rendezvous
Coope-ration
Social Trust
Percieved Amplitude of Performance Information
General Confidence
Confidence
Client Account
Attributed Performance
Client Context
Percieved Performance
Client Authori-zation
Percieved Valence of Performance Information
Active Performance History

Fig. 6. TCC model of cooperation in multilevel cell distributed system

Client's real world identity, client's physical and social 'self', is attributed with perceived valence and amplitude of morality information. These attributes represent client's morality as viewed by the community, and attributed values are maintained within client's digital identity, where they are combined with active values and general trust, to obtain a trust measure of digital identity. Based on the value similarity, depending on specific application needs, cell's identity management function will establish client's virtual presentation with corresponding social trust level in a domain of application.

At the service (server) cell there is an object called client account, a set of data containing information about client identity, as well as perceived valence and amplitude of performance information. This means that service provider is collecting past performance data about client as a basis of confidence, thus obtaining attributed performance info to the client context where it is combined with general confidence, social trust and active performance history. Resulting measure of perceived performance is fed to the client authorization where the appropriate level of confidence is established to the client to utilize specific resources.

At far right, value of social trust presented by client's virtual presentation and service provider's confidence expressed in client authorization are fed to rendezvous function to enable cooperative behaviour between client and service. Rendezvous function also checks the authenticity of cooperating parties.

Let's illustrate this process with the case of Cory Doctorow's inability to buy some music from Amazon, described in the introduction.

Trust: Cory Doctorow is a public person, known worldwide by his writing, public addresses, activism, etc. His morality information perceived valence is positive with quite high amplitude, though we will not try to give any exact measures here and now. From this information we can derive attributed values and input them into Doctorow's digital identity. Based on a value similarity between the cell's currently active values and the values attributed to the Doctorow's digital identity, system is able to create virtual presentation of Doctorow's digital identity, for this application of buying online music, as a man who will most likely pay all his expenses in time.

Confidence: Amazon's web shop keeps information about Doctorow's digital identity in form of a client account record. There it keeps various information about Doctorow's performance perceived by observing his digital identity's behaviour on Amazon web site, creating attributed performance as an input to the client context. Here the attributed performance data will be combined with Amazon's active performance history, general confidence and Doctorow's social trust level (according to lazy calculation principle, social trust will be included into calculation only when needed and available, e.g. when Doctorow signs into Amazon web shop). This yields the perceived performance information which will serve as a basis to authorize Doctorow to use certain resources, in this case to buy some music, thus expressing confidence that Doctorow's credit card information is valid and there is no risk of fraud.

Rendezvous: when Doctorow actually signs into the Amazon web shop and successfully authenticates his digital identity, rendezvous function combines his virtual presentation social trust information with confidence information on Amazon web site and enables transaction to proceed – Doctorow to download some music, and Amazon to charge his credit card with the proper amount.

Because the trust-confidence relation is established on the basis of perceived morality and performance information, instead of locality information (origin of credit card, IP address, etc.) cooperation may be independent of client or service location, i.e. Mr. Doctorow could buy music from any location on the planet, equally successfully from Amazon US, Amazon UK, or any other Amazon store.

7. Closure

Most of the functions of cells can be implemented with ready available software, although some adaptations may be needed such as, for example, inclusion of trust and confidence data with appropriate calculations into the identity management solution.

Conceptually, cell based architecture requires a global set of standards and compliance. Compliant and certified applications will enable network operators to achieve better account control and increased network traffic.

To enable cooperation of different components in a distributed computer system, unified interface between these components should be devised and implemented. Unified interface defines common data transfer formats and commands, using standardized protocols to achieve data independence. To support the asynchronous nature of the cell architecture, use of connectionless protocols is preferred.

Multilevel cell distributed computer system architecture has many implications to various technological and social aspects of distributed systems, especially Internet, which are not discussed here. We have discussed here only the ability of multilevel cell distributed architecture to leverage cooperation in on-line economy by including trust and confidence information into its identity management system. We do believe that cell based architecture, implemented into Internet, may offer an increased control over user identities yet support their mobility, thus reducing the risks of identity theft and related frauds, securing service providers against loss of confidential data, financial risks that may follow.

With cell architecture, network service providers can reduce cost, ease the entry into new markets, and be the vehicle for key partnerships with software vendors, content providers, and other businesses. But, the main development here is the ability to establish trust relationships not only among people represented by their digital identities, but also among digitally identified computing services in a distributed system. This could lead to a whole new practice of doing business online. We could envisage digital services trusted to negotiate with each other with confidence to reach the optimal agreement for all parties involved. For example, internet service provider infrastructure cells could negotiate optimal cost of bandwidth with cells holding long-haul data services. This could be done in seconds thus providing a very effective response to system dynamics, reducing risks due to interruptions of long-haul services.

Of course, this kind of infrastructure behaviour is yet to be extensively researched. Nevertheless, the interaction dynamics in presence of cells that intentionally change their behaviour based on trust relationship does appear as a promising and interesting research direction and worthy of further development.

8. Acknowledgment

This work has been supported, in more ways than just financially, by KING ICT Ltd., Croatia, http://www.king-ict.hr

9. References

Androulaki, E., Bellovin, S. (2009). *An Anonymous Credit Card System, in Trust, Privacy and Security in Digital Business, Proceedings of 6th International Conference, TrustBus 2009*, ISBN 978-3-642-03747-4, Linz, Austria, September 2009

Andress, J. (Ed.). (2011). *The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice*, Syngress, ISBN 978-1-59749-653-7, Waltham MA, USA

Arata Jr., M. J. (2010). *Identity Theft For Dummies*, Wiley Publishing Inc., ISBN: 978-0-470-56521-6, Hoboken NJ, USA

Aven, T., Renn, O. (2010). *Risk Management and Governance: Concepts, Guidelines and Applications*, Springer-Verlag, ISBN 978-3-642-13925-3, Berlin, Germany

Ballardin, F., Merro, M. (2010). *A Calculus for the Analysis of Wireless Network Security Protocols*, in *Formal Aspects of Security and Trust, Revised selected papers of 7th International Workshop, FAST 2010*, ISBN 978-3-642-19751-2, Pisa, Italy, September 16-17, 2010

Benantar, M. (2005). *Access Control Systems: Security, Identity Management and Trust Models*, Springer Science + Business Media, ISBN 978-0-387-27716-5, New York NY, USA

Biskup, J. (2009). *Security in Computing Systems: Challenges, Approaches and Solutions*, Springer Verlag, ISBN 978-3-540-78441-8, Berlin, Germany

Blau, P.M. (1964). *Exchange, Power in Social Life*, Transaction Publishers, ISBN 978-0-887-38628-2, New York NY, USA

Bradach, J.L., Eccles, R. G. (1989). Price, authority, trust: from ideal types to plural forms, *Annual Review of Sociology*, Vol. 15, Aug 1989, pp. 97–118, ISSN 0360-0572

Briggs, P., (2010). *The Evolution of Trust*, in *Trust Management V, Proceedings of 5th IFIP WG 11.11 International Conference, IFIPTM 2011*, ISBN , Copenhagen, Denmark, June 29 - July 1, 2011, (IFIP Advances in Information, Communication Technology)

Broder, J. F. (2006). *Risk Analysis, the Security Survey*, Butterworth-Heinemann, ISBN 978-0-7506-7922-0, Oxford, UK

Brodkin, J. (2008). *Gartner: Seven cloud-computing security risks*, In: InfoWorld, 2011-08-16, Available from http://www.infoworld.com/d/security-central/gartner-seven-cloud-computing-security-risks-853?page=0,1

Brotby, W. K. (2009). *Information security governance: a practical development and implementation approach*, John Wiley & Sons, ISBN 978-0-470-13118-3, Hoboken NJ, USA

Catteddu, D., Hogben, G. (Eds.). (2009). *Cloud Computing: Benefits, risks and recommendations for information security*, European Network, Information Security Agency (ENISA), Retrieved from http://www.enisa.europa.eu/act/rm/ files/deliverables/cloud-computing-risk-assessment/at_download/fullReport

Cofta, P. (2007). *Trust, Complexity and Control: Confidence in a Convergent World*, John Wiley & Sons Ltd, ISBN 978-0-470-06130-5, The Atrium, Southern Gate, Chichester, England

Costa, A.C. (2000). *A Matter of Trust: Effects on the Performance, Effectiveness of Teams in Organizations,* dissertation, University of Tilburg

Crampton, J. (2010). *Cryptographic Enforcement of Role-Based Access Control*, in *Formal Aspects of Security and Trust, Revised selected papers of 7th International Workshop, FAST 2010*, ISBN 978-3-642-19751-2, Pisa, Italy, September 16-17, 2010

Cranor, L. F., Garfinkel S. (2005). *Security and Usability: Designing Secure Systems That People Can Use*, O'Reilly Media, ISBN 0-596-00827-9, Sebastopol CA, USA

Deering, A., Murphy, A. (1998). *The Difference Engine: Achieving Powerful and Sustainable Partnering*, Gower, ISBN-13: 978-0-566-08048-7, Aldershot, England

Denison, D. R. (1996). What is the difference between organizational culture, organizational climate? A native's point of view on a decade of paradigm wars, *Academy of Management Review*, Vol. 21, No. 3, Oct 1996, pp. 619–54, ISSN 0363-7425

Dong, C., Dulay, N. (2010). *Longitude: A Privacy-Preserving Location Sharing Protocol for Mobile Applications*, in *Trust Management V, Proceedings of 5th IFIP WG 11.11 International Conference, IFIPTM 2011*, ISBN , Copenhagen, Denmark, June 29 - July 1, 2011, (IFIP Advances in Information, Communication Technology)

Elangovan, A.R., Shapiro, D.L. (1998). Betrayal of trust in organizations, *Academy of Management Review*, Vol. 23, No. 3, Jul 1998, pp. 547–566, ISSN 0363-7425

Ellinor, L., Gerard, G. (1998). *Dialogue: Rediscover the Transforming Power of Conversation*, John Wiley, ISBN 978-0-471-17466-0, New York NY, USA

Fagen, M. D. (Ed.). (1978). *A History of engineering and science in the Bell System: National Service in War, Peace (1925 - 1975)*, Bell Telephone Laboratories, ISBN 978-0-932-76400-3, New York NY, USA

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*, Free Press, ISBN 978-0-684-82525-0, New York NY, USA

Georgantas, N. et al. (2010). *A Coordination Middleware for Orchestrating Heterogeneous Distributed Systems*, in *Advances in Grid and Pervasive Computing, Proceedings of 6th International Conference, GPC 2011*, ISBN 978-3-642-20753-2, Oulu, Finland, May 11-13, 2011

Geus, A. de (1997). *The Living Company: Habits for Survival in a Turbulent Business Environment*, Harvard Business Press, ISBN-13: 978-1-578-51820-3, Cambridge MA, USA

Ghoshal, S., Bartlett, C.A. (1999). *The Individualized Corporation: A Fundamentally New Approach to Management*, Harper Paperbacks, ISBN 978-088-7-30831-4, New York NY, USA

Governor, J., Hinchcliffe, D., Nickull, D. (2009). *Web 2.0 Architectures*, O'Reilly Media, ISBN 978-0-596-51443-3, Sebastopol CA, USA

Graham, J., Howard, R., Olson, R. (2011). *Cyber Security Essentials*, Auerbach Publications, ISBN 978-1-4398-5126-5, Boca Raton FL, USA

Gulati, R. (1995). Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances, *Academy of Management Journal*, Vol. 38, No. 1, Feb 1995, pp. 85–112.

Hollis, M. (1998). *Trust within Reason*, Cambridge University Press, ISBN 978-0-521-58681-8, Cambridge, UK

Hosmer, L.T. (1995). Trust: the connecting link between organizational theory and philosophical ethics, *Academy of Management Review*, Vol. 20, No. 2, Apr 1995, pp. 379–403,

Jayaswal, B. K., Patton, P. C. (2006). *Design for Trustworthy Software: Tools, Techniques, and Methodology of Developing Robust Software*, Prentice Hall, ISBN 978-0-13-187250-9, Upper Saddle River NJ, USA

Kamil, A., Lowe, G. (2010). *Understanding Abstractions of Secure Channels*, in *Formal Aspects of Security and Trust, Revised selected papers of 7th International Workshop, FAST 2010*, ISBN 978-3-642-19751-2, Pisa, Italy, September 16-17, 2010

Katzenbach, J.R. (1995). *Real Change Leaders*, Crown Business, ISBN 978-0-812-92923-2, New York NY, USA

Kellermann, B., Potzsch, S., Steinbrecher, S. (2010). *Privacy-Respecting Reputation for Wiki Users*, in *Trust Management V, Proceedings of 5th IFIP WG 11.11 International Conference, IFIPTM 2011*, ISBN , Copenhagen, Denmark, June 29 - July 1, 2011, (IFIP Advances in Information, Communication Technology)

Khoury, R., Tawbi, N. (2010). *Corrective Enforcement of Security Policies*, in *Formal Aspects of Security and Trust, Revised selected papers of 7th International Workshop, FAST 2010*, ISBN 978-3-642-19751-2, Pisa, Italy, September 16-17, 2010

Lane, D., Maxfield, R. (1995). *Foresight, complexity and strategy*, Santa Fe Institute, Working Paper 95-12-106, Dec 1995, Sante Fe NM, USA

Lazaric, N., Lorenz, E. (1998). The learning dynamics of trust, reputation and confidence, in *Trust and Economic Learning*, Lazaric, N., Lorenz, E. (Eds.). pp. 1–20, Edward Elgar Publishing, ISBN 978-1-858-98460-5, Cheltenham, UK

Lerner, M.; Vanecek, G. ; Vidovic, N. & Vrsalovic, D. (2002). *Middleware Networks: Concept, Design and Deployment of Internet Infrastructure,* Kluwer Academic publishers, ISBN 0-792-3784-7, New York NY, USA

Lewis, J.D., Weigert, A. (1985). Trust as a social reality, *Social Forces*, Vol. 63, No. 4, June 1985, pp. 967–984, ISSN 0037-7732

Lu, Y., Tsudik, G. (2010). *Enhancing Data Privacy in the Cloud,* in *Trust Management V, Proceedings of 5th IFIP WG 11.11 International Conference, IFIPTM 2011*, ISBN , Copenhagen, Denmark, June 29 - July 1, 2011, (IFIP Advances in Information, Communication Technology)

Ma, Y., Abie, H., Skramstad, T., Nygard, M. (2010). *Assessment of the Trustworthiness of Digital Records*, in *Trust Management V, Proceedings of 5th IFIP WG 11.11 International Conference, IFIPTM 2011*, ISBN , Copenhagen, Denmark, June 29 - July 1, 2011, (IFIP Advances in Information, Communication Technology)

Marek, K. (2008). *Trust: Self-Interest and the Common Good*, Oxford University Press, ISBN 978–0–19–921791–5, New York NY, USA

Mather, T., Kumaraswamy, S., Latif, S. (2009). *Cloud Security and Privacy*, O'Reilly Media, ISBN 978-0-596-80276-9, Sebastopol CA, USA

Mattern, T., Woods, D. (2006). *Enterprise SOA: Designing IT for Business Innovation*, O'Reilly Media, ISBN 0-596-10238-0, Sebastopol CA, USA

Mayer, R. C., Davis, J. H., Schoorman, F. D. (1995). An integrative model of organizational trust, *Academy of Management Review*, Vol. 20, No. 3, Aug 1995, pp. 703–34, ISSN 0363-7425

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations, *Academy of Management Journal*, Vol. 38, No. 1, 1995, pp: 24–59, ISSN 0001-4273

McAllister, D. J. (1997). The second face of trust: reflections on the dark side of interpersonal trust in organizations, *Research on Negotiation in Organizations*, Vol. 6, (Jan 1997). pp. 87–111, ISSN 1040-9556

Muller, T., (2010). *Semantics of Trust*, in *Formal Aspects of Security and Trust, Revised selected papers of 7th International Workshop, FAST 2010*, ISBN 978-3-642-19751-2, Pisa, Italy, September 16-17, 2010

Nahapiet, J., Ghoshal, S. (1998). Social capital, intellectual capital, the organizational advantage, *Academy of Management Review*, Vol. 23, No. 2., May 1998, pp. 242–66, ISSN 0363-7425

Nin, J., Herranz, J. (Eds.). (2010). *Privacy and Anonymity in Information Management Systems*, Springer, ISBN 978-1-84996-237-7, London, UK

Nooteboom, B. (1996). Trust, opportunism and governance: a process, control model, *Organization Studies*, Vol. 17, No. 6, Nov 1996, pp. 985–1010, ISSN 0170-8406

Papacharissi, Z., (Ed.). (2011). *A Networked Self: Identity, Community and Culture on Social Network Sites*, Routledge, ISBN13: 978-0-415-80180-5, New York NY NY, USA

Powell, W. W. (1996). Trust-based forms of governance, in *Trust in Organizations: Frontiers of Theory, Research*, Kramer, R.M., Tyler, T.R. (Eds.). pp. 51–67, Sage Publications, ISBN 978-0 803-95740-4, Thousand Oaks CA, USA

Preve, N. P. (Ed.). (2011). *Grid Computing: Towards a Global Interconnected Infrastructure*, Springer-Verlag, ISBN 978-0-85729-675-7, London, UK

Puder, A., Römer, K., Pilhofer, F. (2006). Distributed systems architecture: a middleware approach, Morgan Kaufmann, ISBN 978-1-55860-648-7, San Francisco CA, USA

Reese, G. (2009). *Cloud Application Architectures: Building Applications, Infrastructure in the Cloud*, O'Reilly Media, ISBN 978-0-596-15636-7, Sebastopol CA, USA

Reina, D. S., Reina, M.L. (1999). *Trust, Betrayal in the Workplace: Building Effective Relationships in Your Organization*, Berrett-Koehler, ISBN 1-57675-377-8, San Francisco CA, USA

Rittinghouse, J. W., Ransome, J. F. (2010). *Cloud Computing: Implementation, Management, and Security*, CRC Press, ISBN 978-1-4398-0680-7, Boca Raton FL, USA

Robinson, P., Vogt, H., Wagealla, W. (Eds.). (2005). *Privacy, security and trust within the context of pervasive computing*, Springer Science + Business Media, ISBN 0-387-23462-4, Boston, USA

Rousseau, D. M., Sitkin, S. B., Burt, R. S., Camerer, C. (1998). Not so different after all: a cross-discipline view of trust, *Academy of Management Review*, Vol. 23, No. 3, Sep 1998). pp. 393–404.

Ryan, K., D.K. Oestreich (1998). *Driving Fear out of the Workplace: Creating the High-trust, High-performance Organization*, Jossey-Bass, ISBN 978-0-787-93968-7, San Francisco CA, USA

Sako, M. (1998). Does trust improve business performance?, in *Trust within and between Organizations: Conceptual Issues, Empirical Applications*, Lane, C., Bachmann, R. (Eds.). pp. 88–117, Oxford University Press, ISBN 978-0-199-24044-9, USA.

Senge, P. M., Kleiner, A., Roberts, C., Ross, R., Roth, G., Smith, B. (1999). *The Dance of Change*, Crown Business, ISBN 978-0-385-49322-2, New York NY, USA

Serpanos, D., Wolf, T. (2011). *Architecture of Network Systems*, Morgan Kaufmann, ISBN 978-0-12-374494-4, Burlington MA, USA

Shapiro, S. P. (1987). The social control of impersonal trust, *American Journal of Sociology*, Vol. 93, No. 3, pp. 623–658, ISSN 0002-9602

Shaw, R. B. (1997). *Trust in the Balance: Building Successful Organizations on Results, Integrity, and Concern*, Jossey-Bass, ISBN 978-0-787-90286-5, San Francisco CA, USA

Siegrist, M., Earle, T. C. & Gutscher, H. (Eds.). (2007). *Trust in Cooperative Risk Management: Uncertainty and Scepticism in the Public Mind*, Earthscan, ISBN 978-1-84971-106-7, London, UK

Sileo, J. (2010). *Privacy Means Profit: Prevent Identity Theft and Secure You and Your Bottom Line*, John Wiley & Sons, ISBN 978-0-470-58389-0, Hoboken NJ, USA

Six, F. (2005). *The Trouble with Trust: The Dynamics of Interpersonal Trust Building*, Edward Elgar Publishing Limited, ISBN 1-84542-290-2, Cheltenham, UK

Smith, S. W. (2005). *Trusted Computing Platforms: Design And Applications*, Springer Science + Business Media, Inc, ISBN 0-387-23916-2, Boston, USA

Sobh, T., Eleithy, K., Mahmood, A. (Eds.). (2010). *Novel Algorithms and Techniques in Telecommunications, Networking*, Springer Science+Business Media, ISBN 978-90-481-3661-2, Heidelberg, Germany

Sorensen, S. (2010). *The Sustainable Network: The Accidental Answer for a Troubled Planet*, O'Reilly Media, ISBN 978-0-596-15703-6, Sebastopol CA, USA

Thuraisingham, B. (2011). *Secure Semantic Service-Oriented Systems*, Auerbach Publications, ISBN 978-1-4200-7332-4, Boca Raton FL, USA

Tiller, J. S. (2011). *Adaptive Security Management Architecture*, Auerbach Publications, ISBN 978-0-8493-7052-6, Boca Raton FL, USA

Tipton, H. F., Krause, M. (2008). *Information Security Management Handbook, 6th Edition*, Auerbach Publications, ISBN 1-4200-6708-7, Boca Raton FL, USA

Toninelli, A., Pathak, A., Issarny, V. (2010). *Yarta: A Middleware for Managing Mobile Social Ecosystems*, in *Advances in Grid, Pervasive Computing, Proceedings of 6th International Conference, GPC 2011*, ISBN 978-3-642-20753-2, Oulu, Finland, May 11-13, 2011

Tsai, W., Ghoshal, S. (1998). Social capital, value creation: the role of intrafirm networks, *Academy of Management Journal*, Vol. 41, No. 4, Dec 1998, pp. 464–76, ISSN 0001-4273

Veeningen, M., de Weger, B., Zannone, N. (2010). *Modeling Identity-Related Properties, Their Privacy Strength*, in *Formal Aspects of Security and Trust, Revised selected papers of 7th International Workshop, FAST 2010*, ISBN 978-3-642-19751-2, Pisa, Italy, September 16-17, 2010

Victor, P., Cornelis, C., De Cock, M. (2011). *Trust Networks For Recommender Systems*, Atlantis Press, ISBN 978-94-91216-08-4, Paris, France

Viega, J. (2009). *The Myths of Security: What the Computer Security Industry Doesn't Want You to Know*, O'Reilly Media, ISBN 978-0-596-52302-2, Sebastopol CA, USA

Vladimirov, A., Gavrilenko, K., Michajlowski, A. (2010). *Assessing Information Security: Strategies, Tactics, Logic and Framework*, IT Governance Publishing, ISBN 978-1-84928-036-5, Cambridgeshire, UK

Vukovic, Dragutin, (2011). *In Cloud we Trust*, in *KOM 2011 – Electronic Communications Technologies And Standards in Informatics, Proceedings of 22nd Conference, KOM 2011*, ISSN 1334-4463, Opatija, Croatia, November 2011

Waldo, J., Lin, H., Millett, L. I. (Eds.). (2007). *Engaging Privacy and Information Technology in a Digital Age*, The National Academies Press, ISBN 978-0-309-10392-3, Washington D.C., USA

Wheatley, M. J., Kellner-Rogers, M. (1996). *A Simpler Way*, Berrett-Koehler Publishers, ISBN 978-1-576-75050-6, San Francisco CA, USA

Whitener, E. M., Brodt, S. E., Korsgaard M. A., Werner, J. M. (1998). Managers as initiators of trust: an exchange relationship framework for understanding managerial trustworthy behavior, *Academy of Management Review*, Vol. 23, No. 3, Jul 1998, pp. 513–30, ISSN 0363-7425

Windley, P. J. (2005). *Digital Identity*, O'Reilly Media, ISBN 0-596-00878-3, Sebastopol CA, USA

Winkel, B. J., Deavors, C., Kahn, D. (2005). *The German Enigma Cipher Machine: Beginnings, Success, and Ultimate Failure*, Artech House, ISBN 978-1-580-53996-8, Norwood MA, USA

Wong, A., Yeung, A. (2009). *Network Infrastructure Security*, Springer Science+Business Media, ISBN 978-1-4419-0165-1, New York NY, USA

Zagalo, N., Morgado, L., Boa-Ventura, A. (Eds.). (2011). *Virtual Worlds, Metaverse Platforms: New Communication, Identity Paradigms*, Information Science Reference (an imprint of IGI Global). ISBN 978-1-60960-854-5, Hershey PA, USA

Zaheer, A., McEvily, B., Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational, interpersonal trust on performance, *Organization Science*, Vol. 9, No. 2, May 1998, pp. 141–159, ISSN 1047-7039

Zand, D.E. (1997). *The Leadership Triad: Knowledge, Trust, and Power*, Oxford University Press, ISBN 978-0-195-09240-0, New York NY, USA

Zheng, L. et al. (2010). *A Scalable Multiprocessor Architecture for Pervasive Computing*, in *Advances in Grid, Pervasive Computing, Proceedings of 6th International Conference, GPC 2011*, ISBN 978-3-642-20753-2, Oulu, Finland, May 11-13, 2011

Adopting and Adapting Medical Approach in Risk Management Process for Analysing Information Security Risk

Ganthan Narayana Samy[1], Rabiah Ahmad[2] and Zuraini Ismail[1]
[1]*Universiti Teknologi Malaysia (UTM),*
[2]*Universiti Teknikal Malaysia Melaka (UTeM)*
Malaysia

1. Introduction

Risk management process is defined as a systematic application of management policies, procedures and practices to the tasks of establishing the context, identifying, analysing, evaluating, treating, monitoring and reviewing risk (AS/NZS ISO 31000:2009, 2009). In addition, precise security risk analysis method should provide two key advantages (Kim *et al.*, 2007). Firstly, effective monitoring of information security policies by protecting organisations critical assets and secondly, capacity to provide appropriate information for the purpose of future prediction and for the development secured information management. However in the real world, most of the organisations do not have proper data about security breaches because they typically fail to document and systematically record the threats incidents (Bojanc and Jerman-Blazic, 2008). According to (Baker *et al.*, 2007) stated that the lack of real data on risk factors is considered as one of the main problem in information security research. Therefore, most of the existing methods intended to estimate probability of an identified vulnerability of security breach is largely relied on guesswork or rough estimation (Baker *et al.*, 2007; Ekelhart *et al.*, 2009; Spears, 2006).

Moreover, the existing information security risk analysis methods have several shortcomings. First, only capable to identify specific threats such as a malicious attacks rather than various types of information security threats concurrency as stated in (Badr and Stephan, 2007; Kim *et al.*, 2007). Second, it is more focus on technology rather than emphasis on the people and process aspects of information systems (Spears, 2006). Third, lack of systematic methods to measure the value of information systems assets from the viewpoint of operational continuity (Suh and Han, 2003). The following limitation of the traditional method is the time-consuming factor and higher cost involved in conducting such analysis especially in medium to large organisations (Spears, 2006). The next limitation is that most of existing methods depends largely on IT professionals or risk analysis experts to conduct the risk analysis. Finally, an IT-centric approach to information security risk analysis indicated it does not involve business users or variety of field managers to understand the risks and threats in promoting security awareness throughout an organisation (Spears, 2006).

Therefore, this research attempt to introduce a new method for performing risk analysis by effectively adopting medical approach namely survival analysis and adapting the risk management process. Under survival analysis approach, a method which is known as Cox Proportional Hazards (PH) Model can be applied to identify significant information security threats. Basically, the risk management process will be based on (AS/NZS ISO 31000:2009, 2009) which provides a sequencing of the core part of the risk management process including establishing the context, risk identification, risk analysis, risk evaluation and risk treatment. Thus, this chapter will describe in greater detail the adoptions and adaptions of medical approach in risk management processes, suitable research method that can be applied and expected benefits from proposed method.

This chapter is organised as follows. The next section describes the previous studies related to this research. Section 3 explains adoptions and adaptions of survival analysis and Cox PH Model in risk management process. Section 4 presents the suggested research method that can be applied in this research. Section 5 presents the discussion, followed by closure and future work in Section 6. Moreover, the following section which is related studies will discuss the existing information security risk analysis methodologies, description about medical research design and approach that related to this research for better understanding of proposed method.

2. Related studies

Basically, there are applications of various risks assessment methods in survivability studies. However, none of the information security risk analysis methodology adopts survival analysis approach or a study has not been previously reported in the research literature as described in (Ma and Krings, 2008b) in order to identify a potential information security threats or factors. Moreover, researchers and information security practitioners or risk analysts can predict better results of events occurring and factors influencing these occurrences more precisely when they adapt medical approaches (Ryan and Ryan, 2008a, 2008b).

2.1 Information security risk analysis methods

Basically, information security risk analysis methods are classified into quantitative, semi-quantitative and qualitative types (AS/NZS ISO 31000:2009, 2009; Badr and Stephan, 2007).

Risk Analysis and Management Method (CRAMM) was developed by Central Computer and Telecommunication Agency (CCTA) by United Kingdom's government (Aime *et al.*, 2007). Current version of CRAMM is 5.1 which was released by Insight Consulting in 2005 based on existing best practices. Moreover, this current version complies with part two of the, BS7799 standard. CRAMM is divided into three stages namely; the first stage is asset identification and valuation, second stage is threat and vulnerability assessment and third stage is selection and recommendation. Basically, CRAMM provides well defined stages which cater for both the technical and the non-technical aspects of security. Furthermore, CRAMM will evaluate threats for both tangible and intangible assets. Moreover, this method also applies by combining the same kind of assets together in order to do a fast analysis. Thus, CRAMM contains a very large countermeasure library consisting of over

3000 detailed countermeasures organised into over 70 logical groupings (Siemens Enterprise, 2005). Therefore, this method can be used for reviewing security aspects, continuity and contingency planning, policy development and compliance, system development and compliance audits. CRAMM makes use of both the qualitative and quantitative elements. Fundamentally, CRAMM is a qualitative method, but the range of value can be transformed into quantitative values and subsequently combined to those values to produce values similar to "annual loss expectancy" value. Thus, the range of risks value is from one (low) to seven (high) (Bornman and Labuschagne, 2004; Siemens Enterprise, 2005).

According to (Maglogiannis and Zafiropoulos, 2006) presented a modelling approach for performing a risk analysis study of distributed HIS. This proposed method was based on basic features of the CRAMM risk analysis framework with the Bayesian Network modeling technique in order to identify assets, threats and vulnerabilities of healthcare information systems and present these interrelationships in a concise and flexible model. A case was applied to a healthcare information network operating in the North Aegean Region in Greece and the HIS assets, threats and vulnerabilities had been thoroughly analysed using the basic features of CRAMM. The findings of the analysis had been used to develop a Bayesian Network model to rank the threats with the highest risk, based on their posterior probability of occurrence in the case of the basic services and system failure (Maglogiannis and Zafiropoulos, 2006). Besides that, another study (Maglogiannis *et al.*, 2006) also had applied CRAMM methodology as a framework in order to identify healthcare information systems threats and the corresponding risks. The initial findings from CRAMM are used for the construction of a fault tree model for representing the logical interrelationships of failure events. Finally, the relationship was represented using an advanced bayesian network model that provides greater flexibility in modeling failure event scenarios and highlighting system critical areas. The proposed risk analysis framework had been applied to patient monitoring system for homecare telemedicine, namely the VITAL-Home System.

Besides that, Construct a platform for Risk Analysis of Security Critical Systems (CORAS) is another information security risk analysis method which was developed under the European Information Society Technologies (IST) program. Development of CORAS had several purposes namely, semi-formal methods for object oriented modeling, developing a framework for risk analysis method and computerised tools for critical systems. Basically, CORAS relies on The Unified Modeling Language (UML) methodology. Basically, CORAS framework has four main pillars (Aagedal *et al.*, 2002). The first pillar is risk documentation framework based on the Reference Model for Open Distributed Processing (RM-ODP). The second pillar is risk management process based on the Australian and New Zealand Standard (AS/NZS 4360:1999). The third pillar is an integrated risk management and development process based on Unified Process (UP). The fourth and last pillar is a platform for tool integration based on Extended Markup Language (XML).The CORAS risk management process is based on AS/NZS 4360:1999 Risk Management. It complements the Code of Practice for Information Security Management (ISO/IEC 17799:2000), Guidelines for the management of IT Security (ISO/IEC 13335:2001) and Functional Safety of Electrical/Electronic/Programmable Electronic Safety Related Systems (IEC 61508) (Aagedal *et al.*, 2002).

Furthermore, according to (Bones *et al.*, 2007) performed a risk analysis study to analyse the security challenges of an instant messaging (IM) service for healthcare industry based on CORAS risk management process framework. The methodology was based on the Australian and New Zealand standard for risk management (AS/NZS 4360/1999), which clearly sets out the risk analysis process in five main steps. The five main steps are namely, context identification, threat identification, impact and probability analysis, risk evaluation and risk treatment. The findings revealed a number of high risk or threats to instant messaging services used in the healthcare industry, namely, malicious software attacks due to unsecured network, intruder's attacks, hackers, power loss and failures on the device or programming errors.

The following information security risk analysis method is Operationally Critical Threat, Asset and Vulnerability Evaluation (OCTAVE). This method was introduced by the Carnegie Mellon Software Engineering Institute (SEI). This approach focuses on assets, threats and vulnerabilities. One of the main concepts of OCTAVE is self-direction. Primarily, OCTAVE relies on structured interviews as a tool in order to identify the critical for assets and it measures the level of risk. This means that, organisation's employees must lead the information security risk evaluation. Hence, an analysis team, consisting of staff from the organisation's business units as well as its IT department is responsible for leading the evaluation and recording of the results. The OCTAVE approach has three phases' namely organisational view, technological view and strategy and plan development phase (Bornman and Labuschagne, 2004). The three phases are build for asset-based threat profiles, identify infrastructure vulnerabilities and develop security strategy and plans. Each process has certain activities that must be completed, and within each of these activities, different steps must be taken in order to achieve the desired outputs. Finally, the result is based on threat profile of different assets. Each threat profile contains information on which mitigation decisions can it be based on. Basically OCTAVE method can be applied to large firms as well as for smaller firms which is known as OCTAVE-S (Caralli *et al.*, 2007).

On top of that, (Coleman, 2004) applied OCTAVE method at three healthcare organisations of different scale, complexity and geographic location in order to access a risk associated with biomedical systems. The evaluation of risks in terms of organisational impact was found to be critically important and was the most influential factor considered when prioritizing risks for mitigation. The differences and similarities observed during the three risk assessments periods, strongly support the concept of a decentralised decision making approach to information security in the healthcare industry. The similarities found between these organisations were in terms of threats that were found in biomedical systems namely related to network infrastructure failures that should have been taken into consideration. In summary, this method allows each organisation the freedom to consider their own unique circumstances, tailor the methodology to their needs and document their decisions accordingly.

The next information security risk analysis method is Information Security Risk Analysis Method (ISRAM). ISRAM is a quantitative based approach using quantitative measures. ISRAM was developed at the National Research Institute of Electronics and Cryptology and the Gebze Institute of Technology in Turkey (Karabacak & Sogukpinar, 2005). This approach allows the interaction between the managers and staffs of the organisation in order to do risk analysis. Basically, ISRAM is a survey-based model which has two main elements of risk namely, probability and consequence. ISRAM utilises numerical value between 1 and 25 as risk measure. This numerical value will act as qualitative measure for instance high, medium or low value. Finally, qualitative value will be used as a result for risk management decisions. The ISRAM methodology consists of seven steps (Karabacak and Sogukpinar, 2005).

Information Systems (IS) Risk Analysis Based on a Business Model is another risk analysis method which was developed by Korea Advanced Institute of Science and Technology (KAIST) in 2002 (Suh and Han, 2003). This model considers the replacement of assets and the disruption of operations together. Disruption of operations is a category of generic risks for example, consumer or buyer confidence, trust and goodwill of the company. This methodology has four stages namely starts with organisational investigation, asset identification and evaluation, threat and vulnerability assessment and finally annual loss expectancy calculation (Suh and Han, 2003).

2.2 Medical research design and approach

This section will presents the medical research and approach that going to applied in this research.

2.2.1 Retrospective cohort study

There are several different types of cohort studies (Euser *et al.,* 2009; Dunn and Clark, 2001). Type of cohort study that will be applied in this research is retrospective cohort study. Retrospective cohort study uses historical data to identify exposure level at some baseline in the past and respectively, follow-up for subsequent occurrences of disease between baseline and the present time (Friis and Sellers, 2010; Euser *et al.,* 2009).

2.2.2 Survival analysis approach

Survival analysis or failure time analysis is a specialised field of mathematical statistics as stated in (Ma and Krings, 2008c). Basically, survival analysis studies positive random variables with censored observations for describing times to events cases (Kleinbaum and Klein, 2005; Lee and Wang, 2003; Ma and Krings, 2008a, 2008b; Ricci, 2006). Events can be for example reaction from treatment or death and response of a disease. Therefore, "the study of survival data is focused on predicting the probability of response, survival, or mean lifetime, by comparing the survival distributions of experimental animals or of human patients and the identification of risk or prognostic factors related to response, survival, and the development of disease" as stated in (Lee and Wang, 2003).Examples of survival time are the lifetimes of organisms or survival times of cancer patients. There are several approaches to assess the associated risks with survival analysis namely, parametric, semi-parametric and non-parametric (Kleinbaum and Klein, 2005; Ma and Krings, 2008a, 2008b).

Generally, in survival analysis, collected data are subject to censoring (Kim *et al.*, 2010). Basically, censored observations can be defined as an exact survival time of the observed subjects is unknown (Clark *et al.*, 2003). Censoring may occur due to several reasons as stated in (Clark *et al.*, 2003; Kleinbaum and Klein, 2005) namely:

i. A person is lost of follow-up during the study period;
ii. A person does not experience the event before the study ends;
iii. A person withdraws from the study due to death (if death is not the event of interest) or some other reason (for instance, adverse drug reaction or other competing risk).

There are three types of censoring namely, right censoring, left censoring and interval censoring. Generally, the three situation stated above are known as right-censored data. Basically, for these situations the complete survival time intervals, which we do not really know, has been cut off (censored) at the right side of the observed survival time interval.

According to (Lee and Wang, 2003),"left censoring occurs when it is known that the event of interest occurred prior to certain time, but the exact time of occurrence is unknown". For instance, if we wish to know the age of diagnosis in a follow-up study of diabetic retinopathy. Thus, at the time of the examination, a 50 year old participant was found to have already developed retinopathy. However, there is no record of the exact time at which initial evidence was found. Therefore, age at examination that is 50 is a left-censored observation. Hence, it means that the age of diagnosis for this patient is at most 50 years. Interval censoring occurs when the event of interest is known to have occurred between times a and b. For example, according to medical records, it indicates that at the age of 45 years, the respondent did not have retinopathy, his age at diagnosis is between 45 and 50 years.

However, the survival analysis approach has unique mathematical models and methodologies that have been developed in order to extract the partial information from the censored observations without creating any unwanted biasness as mentioned in (Ma and Krings, 2008a) and considered as one of benefit which can be adopt by organisation to conduct risk analysis. Further description of survival analysis with censored data will be discussed in sub section 5.2.2. The following section will describe the Cox Proportional Hazards model which is one of the methods used in survival analysis for analysis data.

2.2.3 Cox proportional hazards (PH) model

Cox Proportional Hazards (PH) model, a popular mathematical model and is widely used for analysing survival time data in medical research. Basically, Cox PH model is a semi-parametric approach. In this research, Cox PH model will be used. Initially, Cox PH model proposed by Cox (1972, 1975) is treated as largely an empirical regression model, but later it was found that the framework of the model possesses exceeding flexibility to capture major hazards effects and failure mechanisms (Bradburn *et al.*, 2003; Kleinbaum and Klein, 2005; Lee and Wang, 2003; Ma and Krings, 2008a, 2008b, 2008c).

Basically, Cox PH model is a method for modeling time-to-event data in the presence of censored cases (uncompleted observation). However, Cox PH model allows inclusion of explanatory/predictor variables in the models. Cox PH model will handle the censored cases correctly, and it will provide estimated coefficients for each of the explanatory variables. Besides that, Cox PH model allows us to assess the impact of multiple explanatory variables in the same model. Thus, we can find out which explanatory variables or factors have significant impact on the event and forecast the survival probability according to the influence of factors. Cox (PH) model also can be used to examine the effect of continuous explanatory variables as mentioned in (Lee and Wang, 2003). Therefore, in survival analysis, event of the incidences will be presented in terms of hazard function and of covariates (Bradburn *et al.*, 2003).

The formula for the Cox PH model as expressed in (1), where the hazard function h(t) is dependent on (or determined by) a set of p covariates (x1, x2 ……xp), whose impact is measured by the size of the respective regression coefficients (b1, b2..…,bp). The term h_0 is called the baseline hazard, and is the value of the hazard if all the xi are equal to zero (the quantity exp(0) equals 1).

$$h(t) = h_0(t) \text{ x } \exp\{b_1x_1 + b_2x_2 + \ldots\ldots + b_px_p\}\ . \quad (1)$$

The Cox model is basically a multiple linear regression of the logarithm of the hazard on the variables x_i with baseline hazard being an 'intercept' term that varies with time. The covariates then acts multiplicatively on the hazard at any point in time and this provides the key assumption of the PH model that the hazard of the event in any group is a constant multiple of the hazard in any other group. Thus, this assumption shows that the hazard curves for the groups should be proportional and will not cross.

This proportionality implies that the quantities exp(b_i) are called hazard ratios. A value of b_i greater than zero, or equivalently a hazard ratio greater than one, shows that as the value of the ith covariates increases, the event hazard increases and, thus the length of survival decreases. Eventually, a hazard ratio above 1 indicates a covariate that is positively associated with the event probability and negatively associated with the length of survival. In this research, Cox PH model is defined h_0 (t) as the baseline hazard function. The covariates or explanatory variables in this research will be the potential threats that might affect the information systems which cause failure to system.

2.2.4 The application of survival analysis

Survival analysis has a track record for last two decades and has become the de facto standard in biomedical research (Ma and Krings, 2008a). However, today survival analysis has become a major tool for other fields of study for instance, in engineering reliability, networking and software reliability and survivability, machine learning, and prognostics and health management as stated in (Ma and Krings, 2008a, 2008b; Samrout *et al.*, 2009)

There are many examples of application of survival analysis approach in the medical field for example, which was done by (Ghazali *et al.*, 2010). The purpose of this study is to identify the five years of survival rate and prognostics factors for survival in patients with colorectal cancer. Therefore, in this research, Cox PH model was applied to model the prognostic factors for survival. In summary, factors such as Dukes staging, status of liver metastases and type of treatment are identified as an important independent predictors for survival in patients with colorectal cancer. Moreover, the results also further indicates that the patients with Dukes C staging together with the presence of liver metastases and who are treated with both chemotherapy and radiotherapy are at the greatest risk of death from colorectal cancer.

Another example of survival analysis application in medical domain is presented by (Maida *et al.*, 2009) and that study shows that certain selected wound may affect the survivability of cancer patients. Therefore, this study conducted a prospective observational study of 418 advanced cancer patients and derived hazard ratios (HRs) from Cox PH models. The result shows that the presence of pressure ulcers particularly in female cancer patients and 'other' type of wounds in all cancer patients contributes to reduce survival rate. Furthermore, this useful data can be used as an important measure in existing prognostic model in order to enhance prognostic accuracy.

Rabiah, (2006) combines Cox Proportional Hazard Regression and Genetic Algorithms (CoRGA) in order to select variables. In addition, CoRGA was applied to select best combination of risk factors for four-years, eight-years and fifteen-years, all-cause mortality in older people. In addition, CoRGA was used to identify risk factors for mortality in older men and women separately. The results show that CoRGA was able to select a variety of risk factors for short, medium and long-term, and all-cause mortality and also was able to identify new risk factors for mortality. In summary, CoRGA has the potential to complement traditional statistical methods for analysing survival data, particularly in the identification of putative risk factors for all-cause mortality in communities with older people.

The survival analysis approaches is also suitable to be applied in other fields such as in reliability engineering, social sciences and business. Examples of survival analysis in these fields are the life time of electronic devices, components or systems and workers compensation claims (insurance) and their various influencing risk or factors (Lee and Wang, 2003). Besides that, according to (Ma and Krings, 2008b, 2011) stated that other field of computer science and engineering also has great potential benefits by adopting survival analysis approach for example in network reliability and survivability as well as in prognostics and health management which merits further research. Recent evidence has reported that the application of Cox PH model in the reliability engineering field has increased (Samrout *et al.*, 2009). Moreover, according to (Samrout *et al.*, 2009) studies show that the Cox PH model is used as a modeling tool to integrate the effect of the corrective maintenance on the component's reliability through its relation on the component's age. Finally, the findings show that the corrective maintenance affects on the failure rate and has an effect on the adopted preventive maintenance policy.

Based on (Guo and Brooks, 2009) stated that adoptions of Cox PH model in order to analysis the duration from offering to listing using the data of Chinese A-share IPOs gives several benefits. For example, capability to incorporate information whether censored and uncensored observations to provide consistent parameter estimates and finally the results can be more precise to forecast and assess the listing hazard for a new offering. Therefore, Cox PH model is used in order to identify factors significantly related to the issuer's final listing. In summary, the findings are found to be significant for most of the endogenous factors that affect the issuing system, but factors for example offering price and floatation size reduce in favour of the effect of issuing year.

Besides that, survival analysis approach also has ability to demonstrate the characteristics of unemployment duration in Slovenia as stated by (Borsic and Kavkler, 2009). That study investigated the influence of several variables such as age, gender, level of education, and region using Cox PH model. Primarily, this study stated that Cox PH model has a potential to estimate the ratio of chances of employment for two selected groups of unemployment. Moreover, the results show that it takes a longer time for female and older unemployed persons to find a job and on the average the duration of unemployment decreases with increasing level of education. Also, the outcome of analysis stated that the results can help to identify potential unemployed target groups in order to improve the effectiveness of the employment policy.

According to (Ma and Krings, 2008c) presents a new dynamic hybrid fault models by extending the traditional hybrid fault models with survival analysis and evolutionary game theory. The application domain is Wireless Sensor Network (WSN). Basically, this research introduces survival analysis, which offers time and covariate dependent hazard or survivor functions. This research used Weibull survival distribution models to simulate the time-dependent survivor function of WSN nodes. The findings show that the new dynamic hybrid fault model which transforms hybrid fault model into time and covariate dependent models and make real-time prediction of reliability more realistic and also allows for real-time prediction of fault-tolerance. Furthermore, this research sets the foundations for integrating the hybrid fault models with reliability and survivability analysis by introducing the evolutionary game modelling and extends the evolutionary game theory in its modelling for the survivals of game players.

Section 2 and sub-sections discussed the related studies in this research in depth. Therefore, the next section will emphasise on how the proposed method been adopt and adapt into risk management process.

3. Adopting and adapting survival analysis and cox PH model into risk management process

This section is divided into two main sections. Section 3.1 and sub-sections presents the description of the proposed method according to risk management process in greater detail. Section 3.2 explains differences between general risk management processes with adoptions and adaptions of medical research design and approach in risk management processes.

3.1 Risk management processes in detail

Risk management process methodology is based on (AS/NZS ISO 31000:2009, 2009). This standard clearly sets out the risk management process into five main processes including, establishing the context, risk identification, risk analysis, risk evaluation and risk treatment. The Fig. 1 shows the main elements of risk management process and will be described in greater detail in the following sub-sections.

3.1.1 Communication and consultation

(AS/NZS ISO 31000:2009, 2009) define communication and consultation as "continual and iterative processes that an organisation conducts to provide, share or obtain information and to engage in dialogue with stakeholders regarding the management of risk". In addition, this component serves as a communication platform for stakeholders at all stages of risk management process (AS/NZS ISO 31000:2009, 2009).

3.1.2 Establishing the context

Establishing the context defines various types of context namely external, internal, risk management process context and sets the scope and risk criteria for following process with its organisational environment in order to evaluate risk (AS/NZS ISO 31000:2009, 2009). Furthermore, the measurement variables for risk criteria are regression coefficient (bi), hazard ratio [exp (bi)] and p-Value will be used as risk criteria in order to measure the level of risk and to determine which variables are significant information security threats.

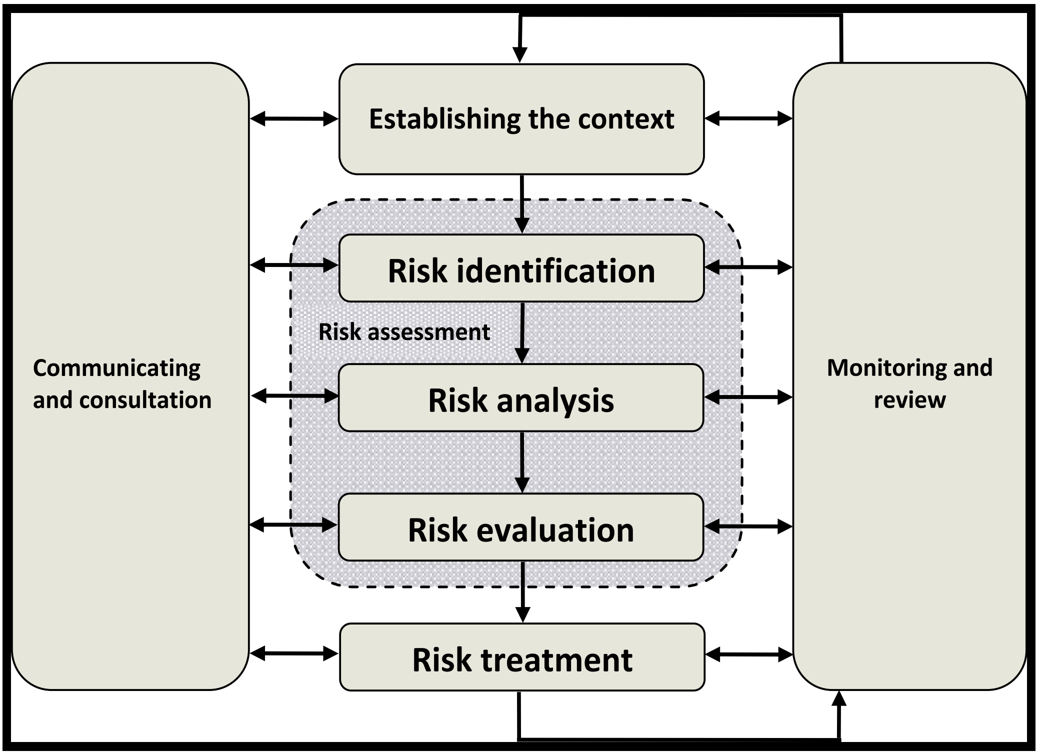


Fig. 1. General Overview of Risk Management Process according to (AS/NZS ISO 31000:2009, 2009)

3.1.3 Risk identification

The following component of risk management process is risk identification process. This process seeks to identify what, why and how the risks can arise and also as an input for further analysis (AS/NZS ISO 31000:2009, 2009). Further identification of appropriate tools and techniques used to identify risks will be carried out at this juncture. There are various methods used to identify risks for instance, brainstorming, scenario analysis, judgments based on experience and records and systems engineering techniques (AS/NZS ISO 31000:2009, 2009). However, the method that is going to be applied should represent or must be capable to identify types of risks correctly.

3.1.4 Risk analysis

Risk analysis process determines the level of the risk. By having identified the various kinds of potential threats from previous processes it will be used as a list of factors in order to analysis the risks. In addition, method of analysis will be dependent on the purpose of the analysis and also the availability of related information. Methods of analysis can be qualitative, semi-quantitative or quantitative or a combination of any two or three of the above depending on the situations. Furthermore, risk analysis process provides an input to risk evaluation process in order to select appropriate risk treatment strategies and methods to manage the risk (AS/NZS ISO 31000:2009, 2009). According to (AS/NZS ISO 31000:2009, 2009), the risk analysis process will focus to an estimate risk level which is derived from combination of likelihood and consequence.

However in this research, the proposed method namely, retrospective cohort study based on survival analysis will be applied to determine the level of risk. Under survival analysis approach, Cox PH model will be applied to identify which information security threats or independent variables such as technological obsolescence, hardware failures, software failures, malware attacks and power failure is most significant.

According to survival analysis perspectives, status variable (dependent variable) is coded in binary value such as 0 or 1. Therefore in this research, 1 is defined for failure if the event occurs during the study period, and 0 if the event does not happen. Finally, a number of systems will be analysed, for example 200 systems, the duration of the study period (for instance in years, months, weeks or days), its status variable together with its independent variables (explanatory variables) that might be a threats to the system failure will be analysed according to survival analysis which is using the Cox PH model as depicted in Fig. 2.

The final output from the Cox PH model will produce an estimate of hazard ratio of several explanatory variables. Basically, explanatory variable with positive regression coefficients are associated with decreased survival times (increased hazard), while variable with negative regression coefficients are associated with increased survival times (decreased hazard). Thus, based on the findings, the explanatory variables which has high hazard ratio will be consider for following process for evaluation purpose.

3.1.5 Risk evaluation

The purpose of this process is to compare the level of risk results found during the risk analysis process with risk acceptance criteria (AS/NZS ISO 31000:2009, 2009). The result of a risk evaluation is a prioritised list of risks for further analysis. A prioritised list of risks for risk treatment process will be diagnosis according to the medical perspectives namely; hazard ratio which is produced by Cox PH model results in order to determine the most predictor or significant variables. Where the most significant variables considered has a greater impact it, will be forwarded to the next process for treatment purposes.

3.1.6 Risk treatment

(AS/NZS ISO 31000:2009, 2009) stated that risk treatment as "involves selecting one or more options for modifying risks, and implementing those options and once implemented, treatments provide or modify the controls". The first step in risk treatment process is selecting the most appropriate risk treatment option by taking care of values and perceptions of organisation or stakeholders. Thus, in this research Cox PH model which produced hazard ratio will be used as an indicator to measure the level of risk for identified threats. The next step is to prepare and implement the chosen risk treatment plans. Consequently, the treatment options should be incorporated accordingly with the organisations management processes and also discussed with the appropriate stakeholders (AS/NZS ISO 31000:2009, 2009).

3.1.7 Monitoring and review

According to (AS/NZS ISO 31000:2009, 2009) monitoring is defines as "continual checking, supervising, critically observing or determining the status in order to identify change from the performance level required or expected". On the other hand, review is the "activity undertaken to determine the suitability, adequacy and effectiveness of the subject matter to achieve established objectives" (AS/NZS ISO 31000:2009, 2009). Besides this, any other factor which affects the treatment options also need to be considered in this process. Therefore, it is vital to repeat the risk management cycle on a regular basis in order to identify unwanted risks.

3.2 Medical research design and approach in risk management processes

Fig. 2. illustrates clearly the differences between general risk management processes with adaptions of medical research design and approach in risk management processes. There are three main differences in three different risk management processes. Firstly, under establishing the context process, one of the subsections of this process that is defining the risk criteria will be based on medical approach. According to (AS/NZS ISO 31000:2009, 2009) stated that organisation should define criteria or measure that going to used as an indicator to evaluate the level of risk. Therefore, in this research the establishment of measure will be based on medical approach namely, regression coefficient (b_i), hazard ratio [exp (b_i)] and p-value will be used as risk criteria in order to measure the level of risk and to determine which variable is significant threats based Cox Proportional Hazards (PH) Model output as shown in Fig. 2.

The second most significant dissimilarity can be observed in risk analysis process clearly. Basically, the general components of risk analysis process in order to measure the level of risk according to (AS/NZS ISO 31000:2009, 2009) is combination of livelihood and consequences. However in this research, the level to risk is determine completely based medical research design and approach as depicted in Fig. 2. above. Moreover, the interpretations the level of risk will be based on hazard ratio as an outcome of adopting the cohort study based survival analysis in this process. The following distinction can be noticed in risk evaluation process. In general risk management process, the evaluation of risk is based on outcomes of risk analysis process as mentioned in (AS/NZS ISO 31000:2009, 2009). Therefore in this research, the outcome from the risk analysis process will be evaluated based on medical perspectives as shown in Fig. 2.

4. Method

This section will discuss about suggested research method design that can be applied, justification for chosen research design, assumptions for collecting and analysis data and data analysis software that can used for analysis the collected data.

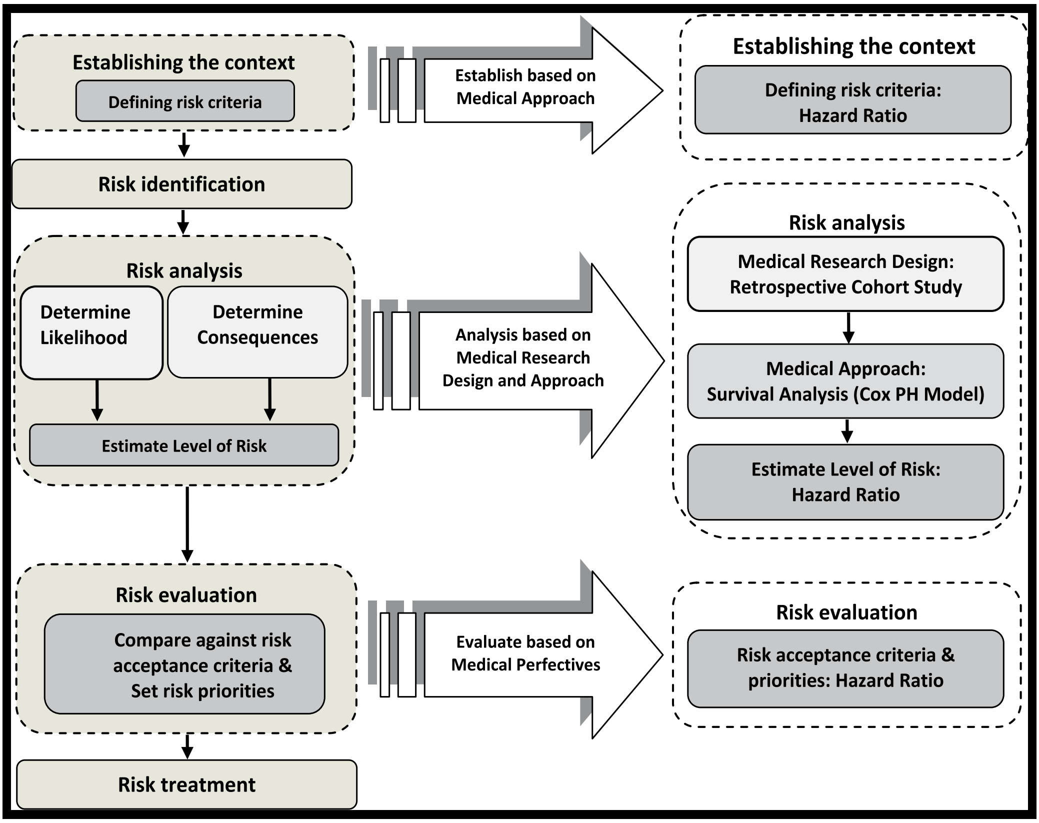

Establishing the context
Defining risk criteria
Risk identification
Risk analysis
Determine Likelihood
Determine Consequences
Estimate Level of Risk
Risk evaluation
Compare against risk acceptance criteria & Set risk priorities
Risk treatment
Establish based on Medical Approach
Analysis based on Medical Research Design and Approach
Evaluate based on Medical Perfectives
Establishing the context
Defining risk criteria: Hazard Ratio
Risk analysis
Medical Research Design: Retrospective Cohort Study
Medical Approach: Survival Analysis (Cox PH Model)
Estimate Level of Risk: Hazard Ratio
Risk evaluation
Risk acceptance criteria & priorities: Hazard Ratio

Fig. 2. Differences between General Risk Management Processes with Adoption and Adaption of Medical Research Design and Approach in Risk Management Process

4.1 The suggested research method design

The mixed method design was chosen for this research is termed as exploratory sequential mixed method design. Basically, the first phase of study will be a qualitative exploration (i.e. identification of potential information security threats). From this initial exploration, the qualitative findings (i.e. potential information security threats) will be used as an instrument for following phase. The second phase will be using quantitative approach as a follow up to the qualitative results in order to administer larger sample according to survival analysis approach in order to interpret the entire findings. Thus, in this research, exploratory sequential design is divided into two phases namely qualitative analysis which will take place in the risk identification process and this is followed by the quantitative analysis in the risk analysis process as illustrated in Fig. 3.

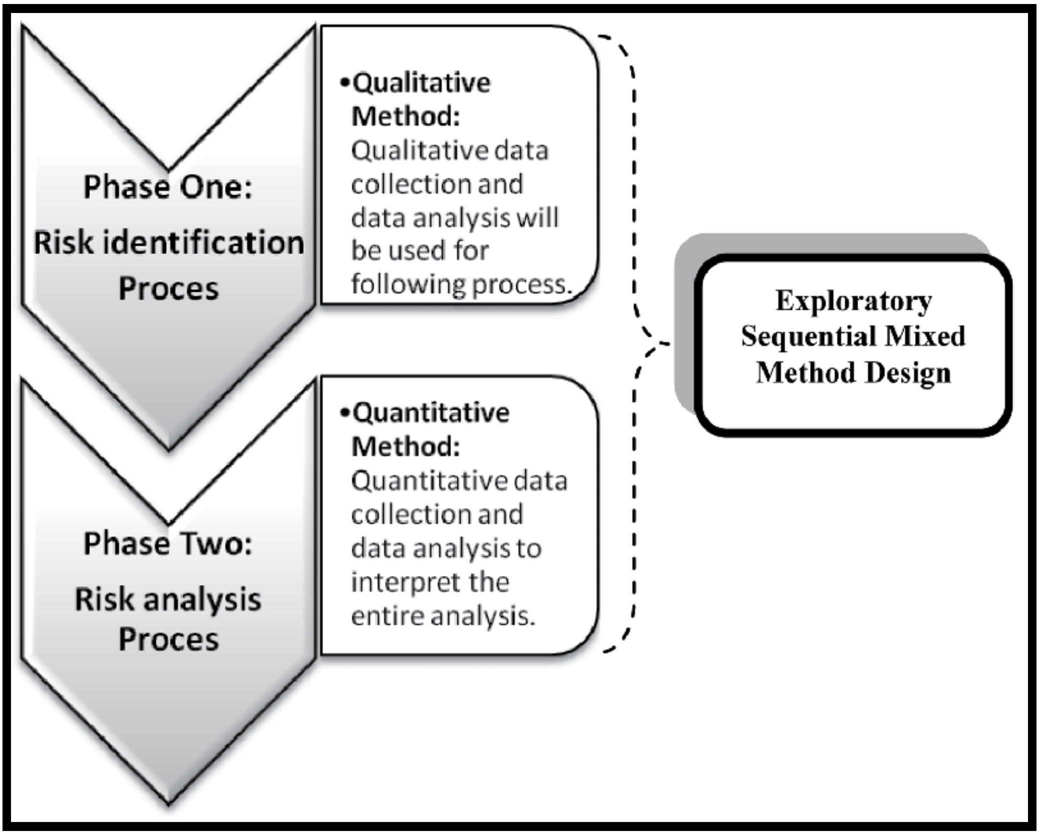


Fig. 3. The Research Design Overview

4.1.1 Justification for chosen research design

There are several key reasons for choosing this method. The reasons are as follows:-

i. The appropriateness to get the intended findings precisely. The exploratory sequential design is most suitable for this research as the results of the first phase, qualitative method basically can assist to develop the second phase, quantitative method (Creswell and Clark, 2011). Moreover, it has the ability to answer the proposed research questions correctly. The researchers cannot answer the proposed research questions using the qualitative method alone. Hence, the combinations of qualitative and quantitative methods are necessarily important in order to answer these research questions.

ii. The usefulness in order to identify important variables to study quantitatively when the variables are unknown (Creswell and Clark, 2007, 2011). There is a importance to conduct qualitative analysis at an early phase that is at the risk identification process in order to identify the potential information security threats and later to use it as a list for the following phase that is at the risk analysis process in order to select the most significant variable using the quantitative analysis.

iii. According to (Creswell and Clark, 2007, 2011) stated this kind of research design is suitable "when a researcher wants to generalise qualitative results to different groups, to test aspects of developing theory or classification, or to explore a phenomenon in depth and then to measure the prevalence of its dimensions". Therefore, this selected research design method is well suited to this research.

4.2 Assumptions for collecting and analysis data

This research has the following assumptions:-

i. In this research all the covariates or variables that is included in the PH Cox model as assumed as time independent variable. According to medical perspective, a time-independent variable is defined to be any variable whose whole value for given individual does not change over time (Kleinbaum and Klein, 2005). For instance, smoking status can change over the period, but for the purpose of research, the smoking covariate is believed as not change once it is measured and only one value per individual is used as stated in (Kleinbaum and Klein, 2005).
ii. Recurrent or repeated events of same variables during the follow-up time for given subject are not included in this research. Thus, the researcher in this research assumes that only first time occurrence of selected variables will be counted for the analysis.
iii. In general, cohort studies are said to be bias due to high rate of lost of follow-up or censored data. In addition, high amount of censored data will affect the significances of result. Thus in this research, the researcher assumes that censored data is low as provided by secondary sources in order to ovoid biasness from expected outcome.
iv. In this study, only the right censored data will be included for analysis.

4.3 Data analysis software

In this research, data analysis will be performed using Predictive Analytics SoftWare (PASW) Statistics, version 18.0. This software has been widely use in analysing statistical data in the medical domain and is well suited for the analysis survival data in this research.

The following section will explain the risk criteria definition and advantages of adopting and adapting medical research design and approach in risk management process.

5. Discussion

The following sub sections will discuss the risk criteria based on the medical interpretation in risk management, advantages adopting and adapting a retrospective cohort study, survival analysis approach and Cox PH model in depth.

5.1 Defining risk criteria

Risk criteria will be based on the medical perspective as described in Table 1. below. Therefore, risk criteria measures namely, regression coefficient (b_i), hazard ratio [exp (b_i)] and p-value will be used as an indicator in order to measure the level of risk in this research. Thus, the following table will describe each variable in more detail in terms of regression cofficient, hazard ratio and p-value with appropriate interpretation.

Measurement Variables	**Measurement Value**	**Medical Perspective Interpretation**	**Adaption in Risk Management Perspective & Interpretation**
Regression Coefficient (b_i)	Positive Coefficient	Associated with decreased survival times (increased hazard) that is risk of death is higher, and the prognosis worse	High level of risk
	Coefficient = 1	No relationship/does not influence survival times	Insignificant
	Negative Coefficient	Associated with increased survival times (decreased hazard) that is risk of death is low, and the better prognosis	Low level of risk
HR-Hazard Ratio [exp (b_i)]	HR > 1	Variable increase the odds of the event occurring (decreases survival times). Example: HR = 5, Exposed group has five times higher the hazard of the unexposed group	Exposed group has higher level of the hazard compare to unexposed group which is depends on HR value and b_i sign (positive/negative coefficient)
	HR = 1	No effect	Insignificant variable
	HR < 1	Variable less the odds of the event occurring (increasing survival times) Example: HR = 0.5, Exposed group has 0.5 times higher the hazard of the unexposed group	Exposed group has lower level of the hazard compare to unexposed group which is depends on HR value and b_i sign (positive/negative coefficient)
***p*-Value**	< 0.05	Variable is significant	Variable have significant impact /influence
	> 0.05	Insignificant variable	Variable does not have significant impact /influence

Table 1. Risk Criteria Measure

5.2 Advantages adopting and adapting medical research design and approach

The following subsections will explain the benefits adopting and adapting medical research and approach in this research.

5.2.1 The advantages adopting and adapting a retrospective cohort study

There are several advantages of using a retrospective cohort study. According to (Euser *et al.*, 2009; Friis and Sellers, 2010) stated that the major advantage of this cohort design is does not required long follow up period of observation like prospective cohort study. Therefore this cohort design is time efficient and also suitable to discover new findings based on exiting data as mentioned in (Euser *et al.*, 2009). Moreover, this cohort design also does not need a huge amount of expenses in order to conduct the study and at the same time capable to collect sufficient amount of date that required for research as stated in (Friis and Sellers, 2010). Thus based on advantages in term of time and cost factor lead to choose this cohort design in this research. Thus, this advantage can be adapted into risk management process for data collection purpose in order to identify information security threats.

5.2.2 The advantages of adopting and adapting survival analysis approach

Basically, there are many reasons or benefits using survival analysis approach. Firstly, researcher, information security practitioners or risk analyst can predict the study of events occurrence and factors influencing their occurrence more precisely. Further, this approach is more efficient, is a powerful tool and brings more benefits compared to other methods namely, the artificial neuron networks (ANN), evolutionary computing, fuzzy logic, decision tree approach and logistic regression as stated in (Chen *et al.*, 2009; Ma and Krings, 2008b; Ma, 2009).

In addition, certain independent variables in reliability engineering field which is related to failure time analysis cannot be analysed using multiple regression techniques. For example, multiple linear regression technique cannot be used for analysis of time-to-event data due to the limitation to handle censored observations or cases for which the event of interest has not yet occurred. Therefore, survival analysis approach such as the Cox PH model is recommended. Thus, survival analysis approach can generate more dynamic characteristics of the event, which were not found in traditional methods (Ma, 2009; Norusis, 2004; Yu and Bi, 2008).

Besides that, the flexibility of the survival analysis approach itself, i.e., information censoring. (Ma and Krings, 2008a) stated that information censoring as “observation of survival times is often incomplete”. Basically, the survival analysis approach has unique mathematical models and methodologies that have been developed in order to extract the partial information from the censored observations without creating any unwanted biasness as mentioned in (Ma and Krings, 2008a). Meanwhile, this approach also can be used to identify which factors have significant impact on the event and forecast the survival probability according to the influence of factors (Ma and Krings, 2008a, 2008b; Yu and Bi, 2008). Therefore, it is important to introduce survival analysis in risk management process. Therefore, this advantage can be used as a tool among the organisations that lack appropriate data to do risk analysis practice. Moreover, the organisation no need to worried about the reliability of result due to ability to handle incomplete data in appropriate manner is the most important and unique advantage of survival analysis approach (Ma and Krings, 2008a, 2008b).

5.2.3 The advantages of adopting and adapting cox proportional hazards (PH) model

In this sub section we are going to discuss a key reasons why we are choosing the Cox PH model for our research. Basically, Cox PH model is robust. This is because, even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard rations of interest and adjusted survival curves can be obtained for a wide variety of data situations. Thus, the results from using this Cox PH model will closely approximate the results for the correct parametric model as stated in (Kleinbaum and Klein, 2005).

Secondly, the Cox PH model is very reliable. Hence, when we may not be completely certain that a given parametric model is appropriate then the Cox PH model will give reliable enough results, so that, it is a safe choice of model to be used. Moreover, the researchers do not need to worry if a wrong parametric model is chosen (Kleinbaum and Klein, 2005). Moreover, the Cox PH model also is a powerful technique to examine the simultaneous relationship of the variables to survival as mentioned in (Lee and Wang, 2003). Basically, the identification of each variable can only identify which variable is significant. However, to determine the simultaneous effect of the variables, an appropriate multivariate statistical method is needed to apply (Lee and Wang, 2003). In this sense, the Cox PH model can be applied.

6. Closure and future research

Firstly, an important contribution of this proposed method is adopting a medical research design and approach and adapting them into risk management process in order to identify potential or influential information security threats, which previously were not undertaken. Therefore, this new way of conducting risk analysis studies holds significant impact among information security practitioners and risk analysis experts in order to identify and manage their information security breaches more effectively. Thus, this will be a new breakthrough in the field of risk management and information security in order to gain advantage using other domains approach.

The following contribution is flexibility of proposed method compared to existing information security risk analysis methods. The most important advantage adopting medical approach namely in survival analysis is information censoring. Information censoring referring to the observation of survival times is often incomplete. Basically, survival analysis approach has unique mathematical models and methodologies that have been developed in order to extract the partial information from the censored observations without creating unwanted bias. Therefore, existing organisation particularly small and medium size can conduct risk analysis studies in order to identify potential information security threats even though they does not have incomplete data about security breaches. Moreover, from the analysis the particular organisation can take appropriate security measure in order to prevent unwanted security breaches.

Lastly, this study provides guidelines for information security practitioners, risk analysts and for top level management in terms of making investment by focusing on the most significant threats based on a proposed method in order to manage their threats effectively. As a result, risk analysis outcomes will be used by organisations in order to identify the gaps in the existing security controls, policies and procedures.

Adopting and adapting medical research design and approach in risk management process in order to identify potential information security threats which are not previously been undertaken. Therefore this chapter discusses the proposed method according to medical perspective in risk management processes in detail. Moreover, the proposed method will be applied in selected government supported hospitals in Malaysia in order to identify potential information security threats in healthcare information systems. Thus, the expected results will be demonstrate the applicability of medical approach in risk management process and in information security domain.

7. References

Aagedal, J. O., den Braber, F., Dimitrakos, T., Gran, B. A., Raptis, D., & Stolen, K. (2002). Model-based risk assessment to improve enterprise security, *Proceedings of Sixth International Conference on Enterprise Distributed Object Computing, 2002 (EDOC'02),* pp. 51–62, 2002

Aime, M. D., Atzeni, A., & Pomi, P. C. (2007). AMBRA: automated model-based risk analysis, *Proceedings of the 2007 ACM workshop on Quality of protection*, pp. 43–48, 2007

Aschengrau, A., & Seage, G. R. (2003). *Essentials of epidemiology in public health*, Jones & Bartlett Learning

Badr, Y. & Stephan, J. (2007).Security and risk management in supply chains. *Journal of Information Assurance and Security*, Vol. 2 (4), pp. pp. 288-296

Baker, W. H., Rees, L. P., & Tippett, P. S. (2007). Necessary measures: metric-driven information security risk assessment and decision making. *Communications of the ACM, 50*(10), pp. 101–106

Bhopal, R. S. (2002). *Concepts of epidemiology: an integrated introduction to the ideas, theories, principles and methods of epidemiology*, Oxford University Press, New York

Bojanc, R., & Jerman-Blazic, B. (2008). Towards a standard approach for quantifying an ICT security investment. *Computer Standards & Interfaces*, *30*(4), pp. 216–222

Bones, E., Hasvold, P., Henriksen, E., & Strandenaes, T. (2007). Risk analysis of information security in a mobile instant messaging and presence system for healthcare. *International journal of medicl informatics*, *76*(9), pp. 677–687

Bonita, R., Beaglehole, R., & Kjellstrom, T. (2006). *Basic epidemiology*, WHO

Bornman, W. G., & Labuschagne, L. (2004). A comparative framework for evaluating information security risk management methods. *Proceedings of the ISSA 2004 enabling tomorrow Conference*, 2007

Borsic, D., & Kavkler, A. (2009). Modeling Unemployment Duration in Slovenia using Cox Regression Models. *Transition Studies Review*, *16*(1), pp. 145–156

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part II: Multivariate data analysis–an introduction to concepts and methods. *British journal of cancer*, *89*(3), pp. 431-436

Caralli, R. A., Stevens, J. F., Young, L. R., Wilson, W. R., & INST, C.-M. U. P. P. S. E. (2007). *Introducing octave allegro: Improving the information security risk assessment process,* Citeseer

Chen, Y., Zhang, H., & Zhu, P. (2009). Study of Customer Lifetime Value Model Based on Survival-Analysis Methods, *WRI World Congress on Computer Science and Information Engineering,* pp. 266–270, 2009

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. *British journal of cancer*, *89*(2), pp. 232-238

Coleman, J. (2004). Assessing information security risk in healthcare organizations of different scale, *International Congress Series*, pp. 125–130, 2004

Creswell, J. W, & Clark, V. L. P. (2007). *Designing and conducting mixed methods research*, Sage Publications, Inc.

Creswell, John W., & Clark, D. V. L. P. (2011). *Designing and Conducting Mixed Methods Research* (Second Edition.), Sage Publications, Inc., California

Dunn, OL., VA. Clark, VA. (2001). *Basic Statistics: A Primer for the Biomedical Sciences*, John Wiley & Sons, New York

Ekelhart, A., Fenz, S., & Neubauer, T. (2009). AURUM: A framework for information security risk management. *42nd Hawaii International Conference on System Sciences, 2009 (HICSS'09)*, pp. 1–10, 2009

Euser, A. M., Zoccali, C., Jager, K. J., & Dekker, F. W. (2009). Cohort studies: prospective versus retrospective. *Nephron Clinical Practice*, *113*(3), pp. 214–217

Friis, R. H., & Sellers, T. (2010). *Epidemiology For Public Health Practice* (4th ed.), Jones & Bartlett Learning

Ghazali, A. K., Musa, K. I., Naing, N. N., & Mahmood, Z. (2010). Prognostic Factors in Patients With Colorectal Cancer at Hospital Universiti Sains Malaysia. *Asian Journal of Surgery*, *33*(3), pp. 127–133

Guo, H., & Brooks, R. (2009). Duration of IPOs between offering and listing: Cox proportional hazard models–Evidence for Chinese A-share IPOs. *International Review of Financial Analysis*, *18*(5), pp. 239–249

Karabacak, B., & Sogukpinar, I. (2005). ISRAM: information security risk analysis method. *Computers & Security*, *24*(2), pp. 147–159

Kim, Y. G., Jeong, D., Park, S. H., Lim, J., & Baik, D. K. (2007). Modeling and Simulation for Security Risk Propagation in Critical Information Systems. *Computational Intelligence and Security*, pp. 858–868

Kim, Y., Kim, B., & Jang, W. (2010). Asymptotic properties of the maximum likelihood estimator for the proportional hazards model with doubly censored data. *Journal of Multivariate Analysis*, *101*(6), pp. 1339–1351

Kleinbaum, D. G., & Klein, M. (2005). *Survival analysis: a self-learning text* (2nd. edition), Springer, New York

Lee, E. T., & Wang, J. W. (2003). *Statistical methods for survival data analysis* (3rd. edition). Wiley-Interscience, New Jersey

Ma, Z., & Krings, A. W. (2011). Dynamic Hybrid Fault Modeling and Extended Evolutionary Game Theory for Reliability, Survivability and Fault Tolerance Analyses. *IEEE Transactions on Reliability*, *60*(1), pp. 180–196

Ma, Z. (2009). A new life system approach to the prognostic and health management (PHM) with survival analysis, dynamic hybrid fault models, evolutionary game theory, and three-layer survivability analysis, *Aerospace conference, p*p. 1–20, 2009

Ma, Z., & Krings, A. W. (2008a). Survival Analysis Approach to Reliability, Survivability and Prognostics and Health Management (PHM), *Aerospace Conference,* pp. 1–20, 2008

Ma, Z., & Krings, A. W. (2008b). Competing Risks Analysis of Reliability, Survivability, and Prognostics and Health Management (PHM), *Aerospace Conference,* pp. 1–21, 2008

Ma, Z. S., & Krings, A. W. (2008c). Dynamic hybrid fault models and the applications to wireless sensor networks (WSNs), *Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*, pp. 100–108, 2008

Maglogiannis, I., & Zafiropoulos, E. (2006). Modeling risk in distributed healthcare information systems. *28th Annual International Conference on Engineering in Medicine and Biology Society* (*EMBS'06)* pp. 5447–5450, 2006

Maglogiannis, I., Zafiropoulos, E., Platis, A., & Lambrinoudakis, C. (2006). Risk analysis of a patient monitoring system using Bayesian network modeling. *Journal of Biomedical Informatics*, *39*(6), pp. 637–647

Mikolajczyk, R. (2010). Methods and Concepts of Epidemiology. *Modern Infectious Disease Epidemiology*, pp. 193–208

Norusis, M. J. (2004). *SPSS 13.0: advanced statistical procedures companion*. Prentice-Hall, New Jersey

Rabiah Ahmad (2006). *Cox Proportional Hazard Regression and Genetic Algorithms (CoRGA) for Analysing Risk Factors for All-Cause Mortality in Community-Dwelling Older People*, Ph.D. Thesis, University of Sheffield, United Kingdom.

Ricci, P. F. (2006). *Environmental and health risk assessment and management: principles and practices* (Vol. 9), Kluwer Academic Publication

Rohrig, B., Du Prel, J. B., Wachtlin, D., & Blettner, M. (2009). Types of study in medical research: part 3 of a series on evaluation of scientific publications. *Deutsches Arzteblatt International*, *106*(15), pp. 262-268

Ryan, J. J. C. H., & Ryan, D. J. (2008a). Biological systems and models in information security. *Proceedings of the 12 Colloquium for Information Systems Security Education,* pp. 127-130, 2008

Ryan, J. J. C. H., & Ryan, D. J. (2008b). Performance metrics for information security risk management. *Security & Privacy, IEEE*, *6*(5), pp. 38–44

Samrout, M., Chatelet, E., Kouta, R., & Chebbo, N. (2009). Optimization of maintenance policy using the proportional hazard model. *Reliability Engineering & System Safety*, *94*(1), pp. 44–52

Spears, J. (2006). A Holistic Risk Analysis Method for Identifying Information Security Risks. *Security Management, Integrity, and Internal Control in Information Systems*, pp. 185–202

Standards Australia/Standards New Zealand. (2009). *Australian/New Zealand Standard for Risk management – Principles and guidelines (AS/NZS ISO 31000:2009)*, NSW and Wellington

Suh, B., & Han, I. (2003). The IS risk analysis based on a business model. *Information & Management*, *41*(2), pp. 149–158

Yu, C., & Bi, X. (2008). Survival Analysis on Information Technology Adoption of Chinese Enterprises, *4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM'08)*, pp. 1–5, 2008

Section 4

Finance and Economics

17

Risk, Return and Market Condition: From a Three-Beta to a Functional-Beta Capital Asset Pricing Model

Zudi Lu[1] and Yuchen Zhuang[2]
[1]*Section of Statistics, School of Mathematical Sciences, The University of Adelaide*
[2]*Department of Mathematics and Statistics, Curtin University of Technology, Perth*
Australia

1. Introduction

In this chapter, we are concerned with how a time-varying beta is linked to market condition in capital asset pricing model (CAPM). This is a question that is related to the recent interest in these large and unexpected swings in asset values, revived after the publication of Taleb's (2007) book, "The Black Swan: The Impact of the Highly Improbable", to explore the merits of beta in the presence of large market fluctuations (c.f., Estrada and Vargas 2011).

It is well known that capital asset pricing model due to (Sharpe 1964) and (Lintner 1965) conveys important information that individual securities are priced so that their expected return will compensate investors for their expected risk. Symbolically, CAPM can be expressed in a general non-expected form as

$$R_i = \alpha_i + \beta_i R_m + \varepsilon_i, \tag{1}$$

where R_i is the return on security i, R_m is the return on the market portfolio and β_i is the measure of security i's non-diversifiable risk relative to that of the market portfolio. Here the return on individual security R_i can be decomposed into the specific return, including expected specific return α_i and random specific return ε_i, and the systematic return, $\beta_i R_m$, owing to the common market return R_m. In this model, the quantity β_i is of particular importance, which is an alternative measure of the risk that an investor has to bear owing to the systematic market movement.

In the traditional CAPM, β_i is assumed to be constant. This assumption has been widely documented to be untrue in the literature. Blume (1971) was among the first to consider the time-varying beta market model, which showed that the estimated beta tended to regress toward the mean; see also (Blume 1975). Earlier studies that attempted to apply random coefficient model to beta include, among others, (Sunder 1980) and (Simonds, LaMotte and McWhorter 1986) who suggested a random-walk coefficient model, and (Ohlson and Rosenberg 1982) and (Collins, Ledolter and Rayburn 1987) who proposed an ARMA(1,1) model for the beta coefficient. More recent literature has widely recognized that the systematic risk of asset changing over time may be due to both the microeconomic factors in the level of the firm and the macroeconomic factors; see (Fabozzi and Francis 1978; Bos and Newbold

1984). Considerable empirical evidences have suggested that beta stability assumption is invalid. The literature is abundant, see, for example, (Kim 1993), (Bos and Ferson 1992, 1995), (Wells 1994), (Bos, Ferson, Martikainen and Perttunen 1995), (Brooks, Faff and Lee 1992) and (Cheng 1997).

The time-varying beta models have also been investigate by using Australian stock market data sets. Brooks, Faff and Lee (1992), and (Faff, Lee and Fry 1992) were among the first to investigate the time-varying beta models. Faff, Lee and Fry (1992) employed a locally best invariant test to study the hypothesis of stationary beta, with evident finding of nonstarionarity across all of their analysis. The random coefficient model was further suggested by (Brooks, Faff and Lee 1994) as the preferred model to best describe the systematic risk of both individual shares and portfolios. However, (Pope and Warrington 1996) reported that random coefficient model was appropriate only for a bit more than 10% companies in their studies. Faff, Lee and Fry (1992) investigated the links between beta's nonstationarity and the three firm characteristics: riskiness, size and industrial sector, without finding the strong pattern between firm size or industry sector and nonstationarity. Faff and Brooks (1998) modelled industrial betas by different regimes based on market returns and volatility of the risk-free interest rate, their univariate and multivariate tests providing mixed evidence concerning the applicability of a time-varying beta model which incorporates these variables. Groenewold and Fraser (1999) argued that the industrial sectors could be divided into two groups: one of them has volatile and non-stationary betas and the other group has relatively constant and generally stationary beta. Other recent studies include (Gangemi, Brooks and Faff 2001), (Josev, Brooks and Faff 2001), and others. An interesting study recently made by (Yao and Gao 2004) investigated the problem of choosing a best possible time-varying beta for each individual industrial index using the state-space framework, including the random walk models, random coefficient models and mean reverting models, which were examined in detail by using the Kalman filter approach.

When testing the validity of asset pricing models, many studies account for market movements, defined as up and down markets. For example, (Kim and Zumwalt 1979) used the average monthly market return, the average risk-free rate and zero as three threshold levels; when the realized market return is above (below) the threshold level the market is said to be in the up (down) market state. Crombez and Vennet (2000) conducted an extensive investigation into the risk-return relationship in the tails of the market return distribution; they defined up and down markets with two thresholds: zero and the risk-free rate. Further, to define three regimes for market movements, that is substantially upward moving, neutral and substantial bear, different threshold points were used, such as: the average positive (negative) market return, the average positive (negative) market return plus (less) half the standard deviation of positive (negative) market returns, and the average positive (negative) market return plus (less) three-quarters of the standard deviation of positive (negative) market returns. The conditional beta risk-return relation has been found to be stronger if the classification of up and down markets is more pronounced.

Galagedera and Faff (2005) has recently argued as in the finance literature and media that high volatility leads to high returns. High volatility in equity prices in many situations has been related to negative shocks to the real economy. On one hand, the volatility of macro-economic variables may partially explain the equity market price variation. On the other hand, the volatility in equity market prices may also be entrenched more in financial market disturbances. In particular, when the market volatility becomes extreme, it could have an impact on financial markets. Some securities are more susceptible to market volatility than others. Two interesting questions that arise in this setting were posed by (Galagedera and Faff 2005): (i) Does the beta risk-return relationship depend on the various market volatility regimes? (ii) Are the betas corresponding to these volatility regimes priced? There have been empirical evidences raising concern about the ability of a single beta to explain cross-sectional variation of security and portfolio returns. Security or portfolio systematic risk is known to vary considerably over time, as documented in the literature in the above. It is further well known that the volatility of financial time series, particularly in high frequency data, changes over time.

In their pioneering work of three-beta CAPM, (Galagedera and Faff 2005) made an assumption that the market conditions can play an important part in explaining a changing beta and could be divided into three states specified as “low", “neutral" or “high" market volatility. First, they fit a volatility model for daily market returns and obtain the estimates for conditional variance. Then, based on the magnitude of these estimates, (Galagedera and Faff 2005) classify daily market volatility σ^2_{Mt} into one of three market volatility regimes, using appropriately defined indicator functions:

$$I_{Lt} = \begin{cases} 1 & if \quad \sigma^2_{Mt} < \sigma^2_L \\ 0 & if \quad otherwise \end{cases} \tag{2}$$

$$I_{Nt} = \begin{cases} 1 & if \quad \sigma^2_L < \sigma^2_{Mt} < \sigma^2_H \\ 0 & if \quad otherwise \end{cases} \tag{3}$$

$$I_{Ht} = \begin{cases} 1 & if \quad \sigma^2_H < \sigma^2_{Mt} \\ 0 & if \quad otherwise. \end{cases} \tag{4}$$

Here σ^2_L and σ^2_H are constants; I_{Lt} represents the low market condition, I_{Nt} represents the neutral market condition, I_{Ht} represents high market condition. By investigating empirically on the single factor CAPM $R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it}$, to estimate the betas in the low, neutral and high volatility markets, Galagedera and Faff extended the market model given in (1), in the form:

$$R_{it} = \alpha_i + \beta_{iL}(I_{Lt}R_{Mt}) + \beta_{iN}(I_{Nt}R_{Mt}) + \beta_{iH}(I_{Ht}R_{Mt}) + \varepsilon_i, \tag{5}$$

where $\beta_{iL}, \beta_{iN}, \beta_{iH}$ are three constants defined as the systematic risks corresponding to the low, neutral and high market volatility regimes, respectively. This model is a richer specification than the traditional single factor CAPM. It is a three-state regime-switching model with the percentiles of market volatility used as threshold parameters. This model raises an interesting question: How is the beta risk linked to the market condition measured by the market volatility? *Our main objective in this chapter is to investigate whether and how the securities' responses to the market vary depending on the changing market volatility.*

We are making a careful investigation into this question in the following sections. In Section 2.1, we shall extend the three-beta model (5) to a more general functional-beta CAPM framework. Nonparametric estimation of the beta functional will be introduced in Section 2.2. We will use the data sets from the Australian stock market to empirically examine the evidences of functional-beta structure in CAPM. An introduction into the data will be provided in Section 3, where an estimation of the unobserved market volatility will be established. Section 4 will carefully examine the linkage of the beta to the market volatility by using both nonparametric and parametric approaches. Based on the findings from the nonparametric estimation, we can suggest reasonable regime switching thresholds, by which a regime-switching threshold CAPM will be proposed and investigated. We will conclude in Section 5.

2. Methodology: From a Three-Beta CAPM to a Functional-Beta CAPM

In this section, we will first propose a Functional-Beta CAPM that is a generalization of the Three-Beta CAPM suggested by (Galagedera and Faff 2005) in Subsection 2.1, and then introduce a nonparametric method to estimate the unknown functional beta in the Functional-Beta CAPM in Subsection 2.2.

2.1 Model

Following the idea of (Galagedera and Faff 2005), we consider a new more general structural framework to incorporate market movements into asset pricing models by including the changes in the conditional market volatility. We achieve this by noting that the model (5) can be expressed as

$$R_{it} = \alpha_i + (\beta_{iL} I_{Lt} + \beta_{iN} I_{Nt} + \beta_{iH} I_{Ht}) R_{Mt} + \varepsilon_{it} \equiv \alpha_i + \beta_{it} R_{Mt} + \varepsilon_{it}, \tag{6}$$

which is a time-varying beta model, with

$$\beta_{it} = \beta_{iL} I_{Lt} + \beta_{iN} I_{Nt} + \beta_{iH} I_{Ht}. \tag{7}$$

We note that the volatility of market returns is partitioned into three regimes in (2)–(4), which are the functions of the size of the conditional market volatility, say, σ^2_{Mt}. Therefore β_{it} is a simple functional of the market volatility σ^2_{Mt}, that is

$$\beta_{it} = \begin{cases} \beta_{iL} & if \quad \sigma^2_{Mt} < \sigma^2_L, \\ \beta_{iN} & if \quad \sigma^2_L \leq \sigma^2_{Mt} < \sigma^2_H, \\ \beta_{iH} & if \quad \sigma^2_{Mt} \geq \sigma^2_H. \end{cases} \tag{8}$$

So the three-beta CAPM proposed by (Galagedera and Faff 2005) is a simple functional beta model.

In this chapter, we will extend the model (5) that was suggested by (Galagedera and Faff 2005) and propose a general functional-beta model as follows:

$$R_{it} = \alpha_i + \beta_i(\sigma^2_{Mt}) R_{Mt} + \varepsilon_{it}, \tag{9}$$

where as before, R_{it} is the return of financial asset i at time t, R_{Mt} is the market return at time t, σ^2_{Mt} is the market volatility at time t, α_i is the conditional expected specific return, ε_{it} is random specific return, and β_i is the coefficient of the contribution due to the market factor, changing with the market volatility. Here $\beta_i(\cdot)$ is particularly important, which is the systematic risk functional, in capital asset pricing modelling. We may also treat α_i as varying with σ^2_{Mt} although its value is usually rather small to be often assumed as a constant. We call (9) the functional-beta CAPM. For our objective, we need to estimate the unknown α_i and $\beta_i(\cdot)$ in (9).

2.2 Estimation of functional-beta CAPM: nonparametric method

Given the historical observations (R_{it}, R_{Mt}), $t = 1, 2, \cdots, T$, we are concerned with how to estimate the unknown functional beta. First of all, we need some way to estimate the unobservable market volatility σ^2_{Mt}. Using the market returns R_{Mt}, $t = 1, 2, \cdots, T$, we can try to estimate σ^2_{Mt} in various ways. A simple way is to apply the econometric models of ARCH of (Engle 1982) or GARCH of (Bollersleve 1986), as done in (Galagedera and Faff 2005). More involved stochastic volatility models can also be applied (c.f., Gao, 2007, page 169). Alternatively, we can use realized market volatility as an estimate of σ^2_{Mt}; see (Allen *et al*. 2008) for a comprehensive review on realized volatility. In the following we assume the market volatility σ^2_{Mt} has been estimated, denoted by $\hat{\sigma}^2_{Mt}$, $t = 1, 2, \cdots, T$.

We will estimate the unknown functional $\beta(v)$ at the market volatility $\sigma^2_{Mt} = v$ by least squares local linear modelling technique (c.f. Fan and Gijbels, 1996). The basic idea of least squares local linear modelling technique with $\beta(\cdot)$ can be described as follows. When σ^2_{Mt} is equal or close to v, then $\beta(\sigma^2_{Mt})$ can be expressed or approximated by

$$\beta(v) + \beta'(v)(\sigma^2_{Mt} - v) \equiv \beta_0 + \beta_1(\sigma^2_{Mt} - v) \tag{10}$$

Locally at v, the model can then be approximately expressed as:

$$R_{it} \approx \alpha + (\beta_0 + \beta_1(\sigma^2_{Mt} - v))R_{Mt} + \varepsilon_t, \tag{11}$$

where though we can also assume α depending on σ^2_{Mt} in model (9) and apply local linear idea to $\alpha(\cdot)$, the estimation of $\alpha(\cdot)$ is of less interest in capital asset pricing modelling, which is very close to zero, therefore in (11) α is treated as a local constant to reduce the number of unknown local parameters.

Therefore, replacing σ^2_{Mt} by $\hat{\sigma}^2_{Mt}$, the least squares local linear estimate of α and $\beta(\cdot)$ in (9) can be made by setting $\widehat{\alpha(v)} = \widehat{\alpha}$ and $\widehat{\beta(v)} = \widehat{\beta_0}$, where $(\widehat{\alpha}, \widehat{\beta_0}, \widehat{\beta_1})$ minimizes:

$$L(\alpha, \beta_0, \beta_1) = \sum_{t=1}^{T} (R_{it} - [\alpha + (\beta_0 + \beta_1(\hat{\sigma}^2_{Mt} - v)R_{Mt})])^2 K(\frac{\hat{\sigma}^2_{Mt} - v}{h}), \tag{12}$$

where $h = h_T \to 0$ is a bandwidth that controls the length of the local neighborhood of v in which the observations locally used fall, *K(x)* is a kernel function, which may, for example, take

$$K(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{13}$$

Therefore, we have three unknown local parameters α, β_0 , β_1 . Applying partial differentiation, we get the expression of the estimators of the three unknown local parameters at v as follows:

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta_0} \\ \widehat{\beta_1} \end{pmatrix} = A_T^{-1} B_T, \tag{14}$$

where

$$A_T = \sum_{t=1}^{T} \begin{pmatrix} 1 & R_{Mt} & R_{Mt}(\hat{\sigma}^2_{Mt} - v) \\ R_{Mt} & R^2_{Mt} & R^2_{Mt}(\hat{\sigma}^2_{Mt} - v) \\ (\hat{\sigma}^2_{Mt} - v)R_{Mt} & (\hat{\sigma}^2_{Mt} - v)R^2_{Mt} & (\hat{\sigma}^2_{Mt} - v)^2 R^2_{Mt} \end{pmatrix} K\left(\frac{\hat{\sigma}^2_{Mt} - v}{h}\right),$$

$$B_T = \begin{pmatrix} \sum_{t=1}^{T} R_{it} K(\frac{\hat{\sigma}_{Mt}^2 - v}{h}) \\ \sum_{t=1}^{T} R_{Mt} R_{it} K(\frac{\hat{\sigma}_{Mt}^2 - v}{h}) \\ \sum_{t=1}^{T} (\hat{\sigma}_{Mt}^2 - v) R_{Mt} R_{it} K(\frac{\hat{\sigma}_{Mt}^2 - v}{h}) \end{pmatrix}.$$

It is well-known (c.f., Fan and Gijbels, 1996) that the bandwidth h plays an important role in the process of estimation. Therefore, how to choose the bandwidth becomes an important step in the estimation. A popular method is the cross-validation (CV) selection of bandwidth (cf., Stone 1974). We here apply a leave-one-out CV, defined below, which is relatively computationally less intensive in comparison with other more involved CV principles:

$$CV(h) = \sum_{s=1}^{T} \{R_{is} - \hat{\alpha}_{-s}(\hat{\sigma}_{Ms}^2) - \hat{\beta}_{-s}(\hat{\sigma}_{Ms}^2) R_{Ms}\}^2, \tag{15}$$

where $(\hat{\alpha}_{-s}(\cdot), \hat{\beta}_{-s}(\cdot))$ are the estimators of $(\alpha(\cdot), \beta(\cdot))$ obtained by minimizing (12) with the term $t = s$ removed from the sum of (12). We select the h_{opt} that minimizes $CV(\cdot)$ over $h \in [h_L, h_U]$, where $0 < h_L < h_U$ are appropriately given. To simplify the computation, a partition of $[h_L, h_U]$ into q points $h_1, h_2, \cdots, h_q$ is applied. We adapt an empirical rule for selecting a bandwidth by (Fan *et al.* 2003) to determining the bandwidth h; see also (Lu *et al.* 2009): Up to first order asymptotics, the optimal bandwidth is $h_{opt} = \{c_2/(4Tc_1)\}^{1/5}$, minimising $CV(h) = c_0 + c_1 h^4 + \frac{c_2}{Th} + o_P(h^4 + T^{-1}h^{-1})$. In practice, the coefficients c_0, c_1 and c_2 will be estimated from $CV(h_k)$, $k = 1, 2, \cdots, q$, via least squares regression,

$$\min_{c_0, c_1, c_2} \sum_{k=1}^{q} \{CV_k - c_0 - c_1 h_k^4 - c_2/(Th_k)\}^2, \tag{16}$$

where $CV_k = CV(h_k)$ obtained from (15). We let $h_{opt} = \{\hat{c}_2/(4T\hat{c}_1)\}^{1/5}$ when both $\hat{c}_1$ and $\hat{c}_2$, the estimators of c_1 and c_2, are positive. In the likely event that one of them is nonpositive, we let $h_{opt} = \arg\min_{1 \le k \le q} CV(h_k)$. This bandwidth selection procedure is computationally efficient as q is moderately small, i.e. we only need to compute q CV-values.

Before we can apply the estimation method introduced in this section to examine the empirical evidence of the functional-beta CAPM, we need to introduce the data sets that we will use, in next section.

3. Data sets from Australian stock market

We will use a set of the stocks data collected from Australian stock market to explore the evidence of functional-beta CAPM in this chapter. The reason why we use the Australian data is because we believe an Australian dataset is ideal for this task. Firstly, the Australian evidence regarding the CAPM is well studied by (Ball, Brown and Officer 1976); (Faff 1991); (Wood 1991); (Faff 1992); (Brailsford and Faff 1997); and (Faff and Lau 1997) as well as (Yao and Gao 2004). Ball, Brown and Officer (1976) may be the first published test of the CAPM using Australian data. They employed the basic univariate testing methodology and found evidence supporting their model. Therefore using the Australian data set will help us to better understand the varying-coefficient nature of CAPM. Secondly, it can be seen that a relatively few, very large companies dominate the Australian market. For example, around 40 per cent of market capitalization and trading value is produced by just 10 stocks, whereas a similar number of the largest US stocks constitute only about 15 per cent of the total US market. Moreover, there are typically prolonged periods in which many smaller Australian companies do not trade. Therefore the market risk may be a significant factor that impacts the risk of the individual stock or index measured by the beta in CAPM. Thirdly, despite the above argument, the Australian equity market belongs to the dominant group of developed markets. For instance, as at the end of 1996 it ranked tenth among all markets in terms of market capitalization at about $US312,000 million. Interestingly, this is not greatly dis-similar from the size of the Canadian market which ranked sixth (Faff, Brooks, Fan 2004). Therefore the Australian data may be of some typical properties that the other markets may share.

According to ASX Indices (including All Ordinaries Index, ASX 200 GICS Sectors Index), we take sample size 986, from August 2nd 2004 to August 8th 2008, for an illustration. The sectors indexes include ASX 200 GICS Energy, ASX 200 GICS Materials, ASX 200 GICS Health Care, ASX 200 GICS Financials, ASX 200 GICS Finance-v-property trusts. Moreover, we also take a group of individual stock data which is ANZ bank group limited as survey sample of individual stock analysis. An introduction to individual and market return series is presented in Subsection 3.1; estimation of the market volatility that we need in estimating functional-beta CAPM is detailed in Subsection 3.2.

3.1 Individual and market return series

At first we review the market return of Australia Index from August 2nd 2004 to August 8th 2008. The time series plot of the 5 sector daily indices and the ANZ stock daily closing price are depicted in Figure 1 together with their return series in Figure 2, where the daily return data denoted as R_t (for individual sector index or for individual security), can be expressed as:

$$R_t = (\log P_t - \log P_{t-1}) \times 100, \tag{17}$$

where P_t represents the closing price of individual sector index in day t. The daily market return data, R_{Mt} , can be expressed as:

$$R_{Mt} = (\log P_{Mt} - \log P_{M,t-1}) \times 100, \tag{18}$$

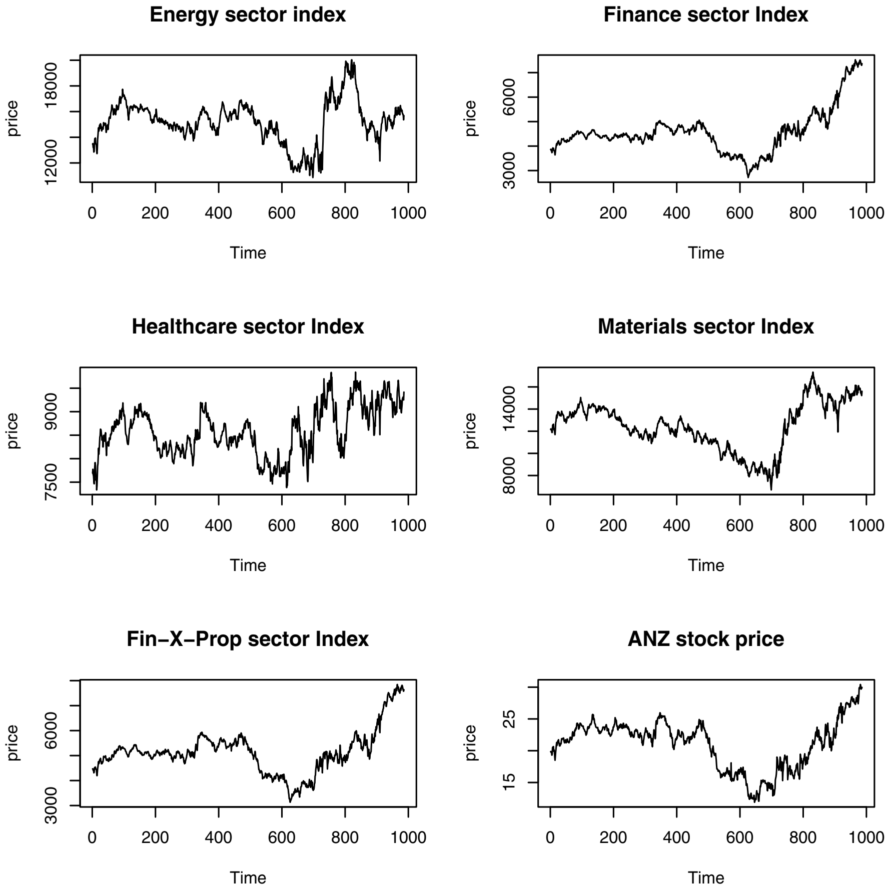
and P_{Mt} represents the closing price of all ordinaries index in day t, both of which are plotted in Figure 3.

In addition to market return data, R_{Mt}, we also need the daily market volatility, σ^2_{Mt}, which is unobservable directly. We discuss how to estimate σ^2_{Mt} based on the market return data, R_{Mt}, in the next subsection.

3.2 Market volatility

The market volatility that can not be observed directly needs to be estimated by using the market return series R_{Mt}. A popular method to estimate the market volatility in the literature is by using the family of GARCH models proposed by (Engle 1982) and (Bollersleve 1986); in particular, we produce the market volatility by the most popular GARCH(1,1) model applied to the return series of All Ordinaries Index. In the GARCH(1,1) model

$$\begin{cases} R_{Mt} = a_0 + a_1 R_{M,t-1} + \epsilon_t \\ \epsilon_t = e_t \sigma^2_{Mt} \\ \sigma^2_{Mt} = \alpha_0 + \alpha_1 \epsilon^2_{t-1} + \beta_1 \sigma^2_{Mt}, \end{cases} \tag{19}$$



Energy sector index

Finance sector Index

Healthcare sector Index

Materials sector Index

Fin−X−Prop sector Index

ANZ stock price

Fig. 1. The 5 ASX 200 GICS Sectors Indexes and ANZ daily closing price in Australia from August 2nd 2004 to August 8th 2008. Sample size =986.

we use R fGarch package to calculate the parameters under e_t satisfying $Ee_t = 0$ and $\text{var}(e_t) = 1$. In order to examine the impacts of distribution for e_t on the estimation of the volatility, we tried different distributions for e_t in the fGarch package, including normal, t and generalized error distribution and their skewed versions, and found that the non-skewed distributions are more acceptable according to their AIC values (Akaike, 1974), with the results listed in Table 1: Quite obviously it follows from the p-value in Table 1 that a_0 and a_1 in the AR part are close to zero while the GARCH parameters are all away from zero at the significance level of 5%. Also by the AIC values, these three GARCH models are quite close to each other and well fitted to the market return data set, with the GARCH-GED (i.e., with e_t of generalized error distribution) preferred. The estimated volatility series under GARCH-GED is plotted Figure 4, where the kernel density estimators of the estimated volatility series under different GARCH models are also displayed, confirming that the estimated volatility under different models are very close. In the following we will use the estimated volatility from the GARCH-GED model as the volatility series in the estimation of the functional-beta CAPM. The summary statistics

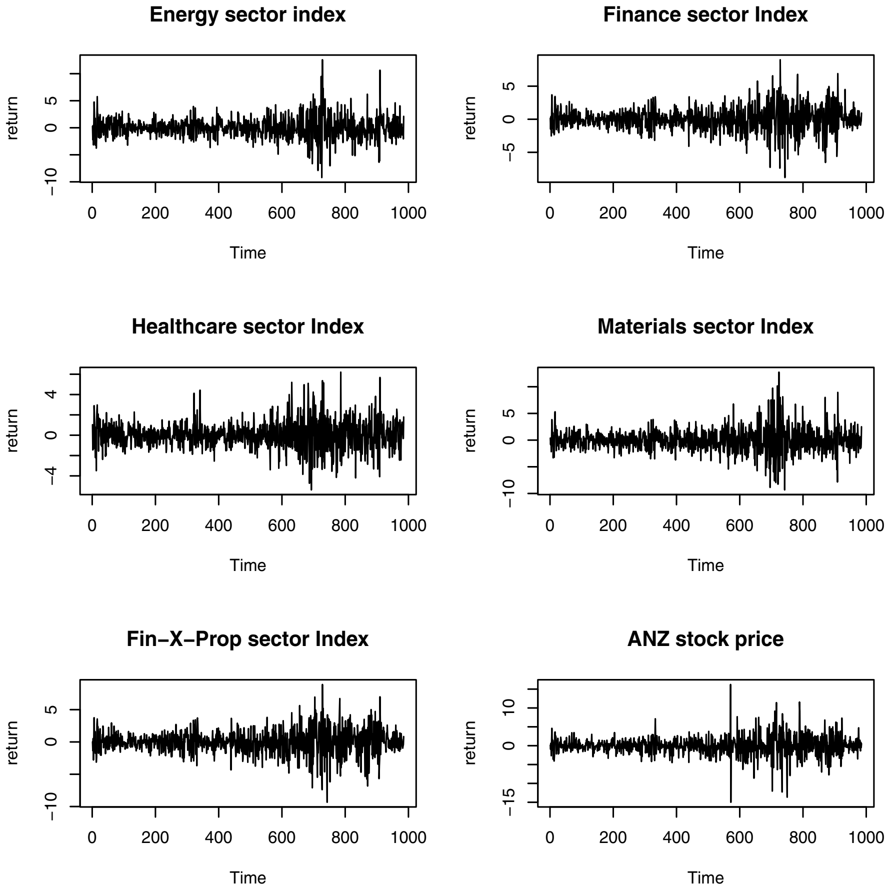


Fig. 2. The return series of the 5 ASX 200 GICS Sectors Indexes and ANZ daily closing price in Australia from August 2nd 2004 to August 8th 2008. Sample size =985.

	a_0	a_1	α_0	α_1	β_1	shape	AIC
garch-normal	-0.019310	-0.022673	0.033037	0.110890	0.871949		3.248890
(p-value)	(0.5731)	(0.4995)	(0.0146)	(4.53e-08)	(2e-16)		
garch-t	-0.032541	-0.015214	0.031242	0.114034	0.873597	10.00000	3.248322
(p-value)	(0.3363)	(0.6457)	(0.0360)	(1.14e-06)	(2e-16)	(7.50e-05)	
garch-ged	-0.031626	-0.020063	0.031619	0.111393	0.872692	1.682406	3.244669
(p-value)	(0.3536)	(0.5460)	(0.0267)	(3.94e-07)	(2e-16)	(2e-16)	

Table 1. Estimated GARCH models under different innovations for the return of all ordinaries index

on the market return and volatility are provided in Table 2, from which we can see that the market is quite volatile with large extreme values.

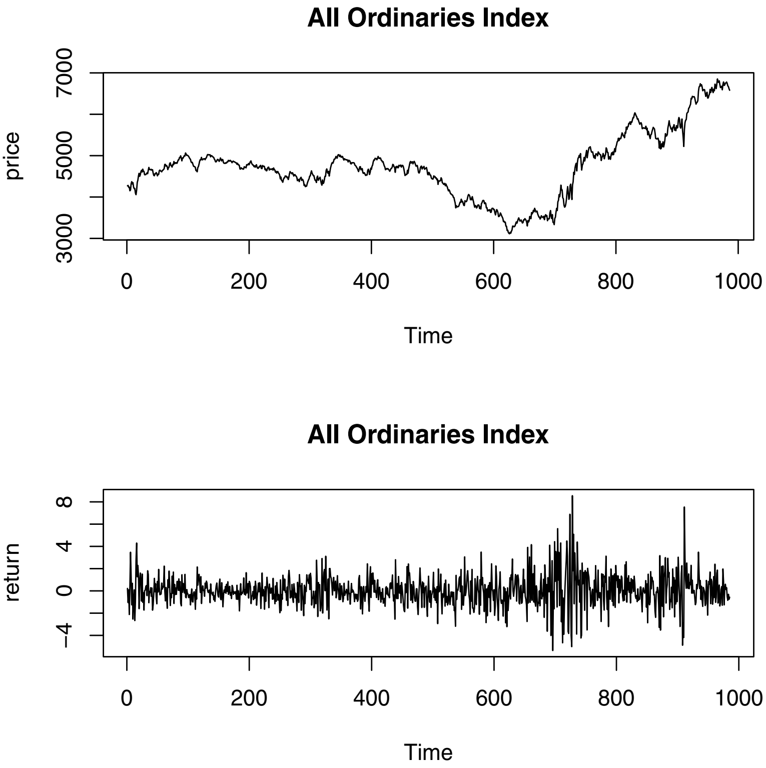


Fig. 3. All Ordinaries Index (sample size =986) and its return series (sample size=985) in Australia from August 2nd 2004 to August 8th 2008.

	mean	standard deviation	skewness	kurtosis	median	min	max
Market return	0.04388	2.0214	0.4596	6.5235	-0.0099	-5.3601	8.5536
Market volatility	1.3089	0.3169	2.1493	8.7173	1.1729	0.6023	4.3905

Table 2. Some statistics data for the Australia index

In the next section, we shall explore the functional form of the beta risk associated with the market volatility in CAPM with the stocks data sets from Australian stock market introduced in this section.

4. Functional-beta CAPM: Empirical evidences

To carefully examine the evidences of functional-beta model (9), we are using the methodology introduced in Section 2, under the semiparametric beta structure that is estimated by local linear fitting introduced in Section 2.2. The advantage of this semiparametric method lies in that the data will determine the relationship of the beta coefficient associated with the market volatility, without pre-specifying a parametric structure to avoid model mis-specification. Based on the findings from the semiparametric method,

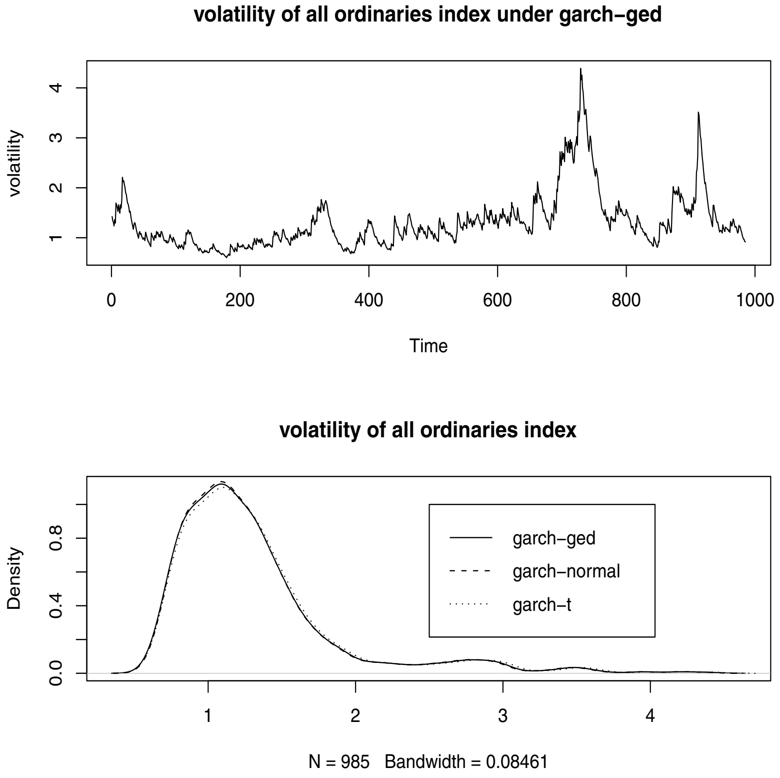


Fig. 4. The estimated volatility of All Ordinaries Index and its kernel probability density estimates under different GARCH models

we shall clearly see whether the beta coefficient associated with the market volatility is of three- or multi-beta structure or not. Differently from the three- or multi-beta CAPM in (Galagedera and Faff 2005), our new findings will indicate that the functional beta may probably be parameterized as threshold (regime-switching) stepwise linear functionals of the market volatility, rather than three or more simple constant beta's.

We will present our empirical evidence based nonparametric estimation method in Subsection 4.1 and further investigation into the parametric evidence in Subsection 4.2.

4.1 Nonparametric evidence

Referred to Section 2.2, when we apply local linear fitting method to estimate the unknown beta function in model (9), we need to use the real data to choose the ideal bandwidth for each Sector index. The values of CV are calculated against 40 points of the bandwidth h for eight groups of data (with bandwidth ranging from 0.2 to 0.6 with partition interval of length 0.01). Hence using the CV calculation procedure given in Section 2.2, we can have the chosen bandwidths as follows in Table 3:

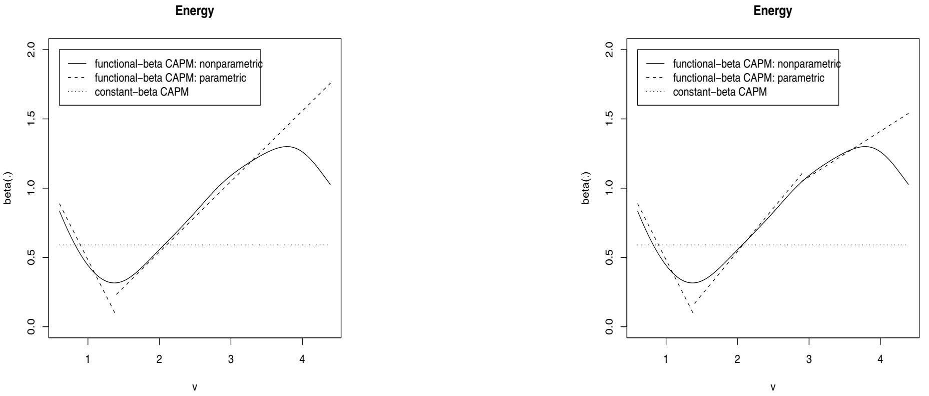


Fig. 5. For Energy sector index:

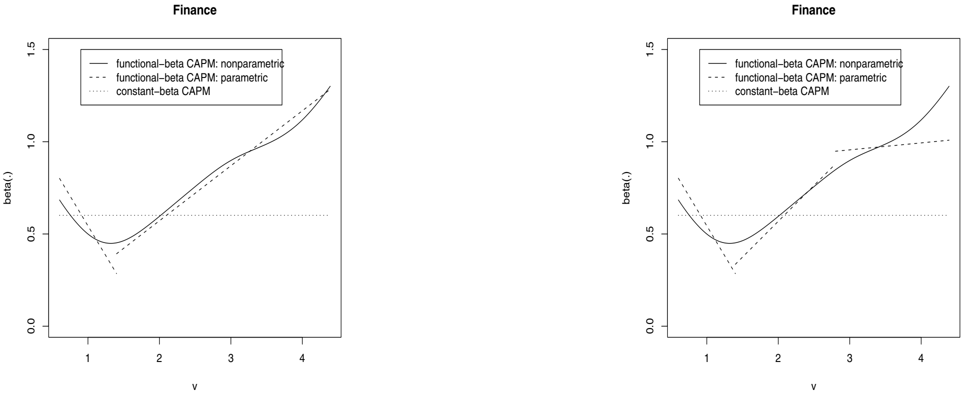


Fig. 6. For Finance sector index:

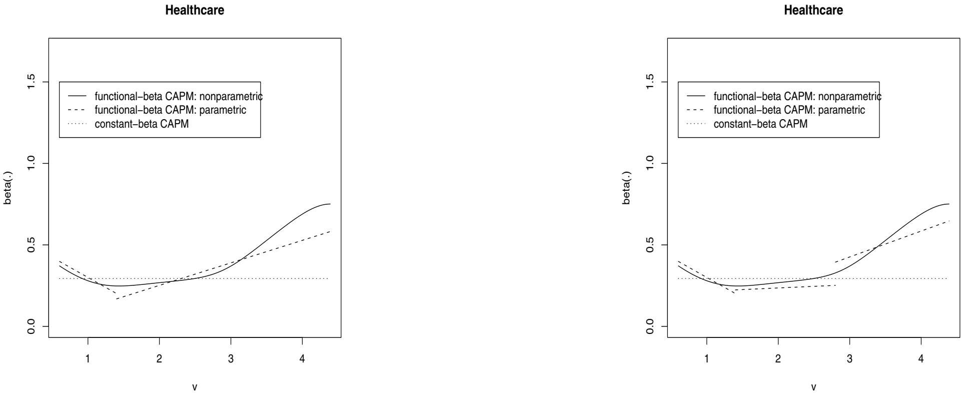


Fig. 7. For Health care sector index:

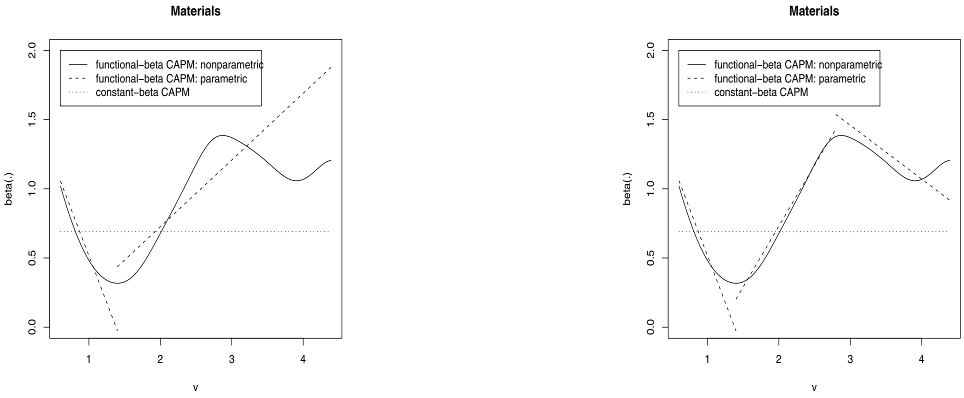


Fig. 8. For Materials sector index:

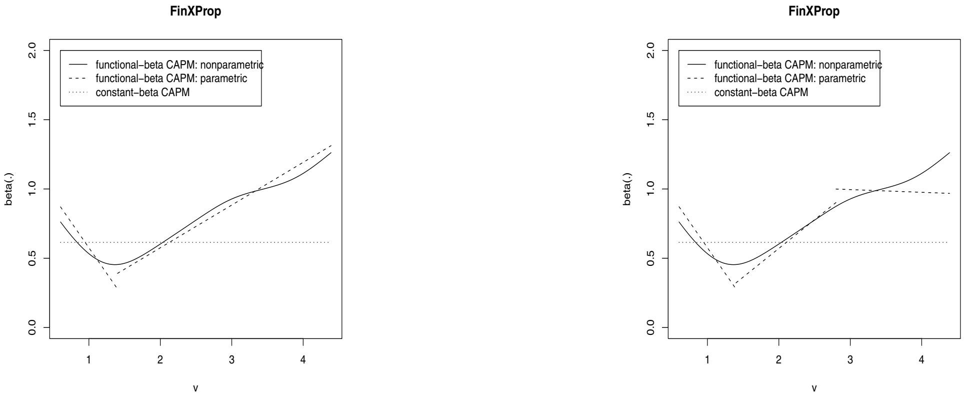


Fig. 9. For Financial-v-Properties Trusts sector index:

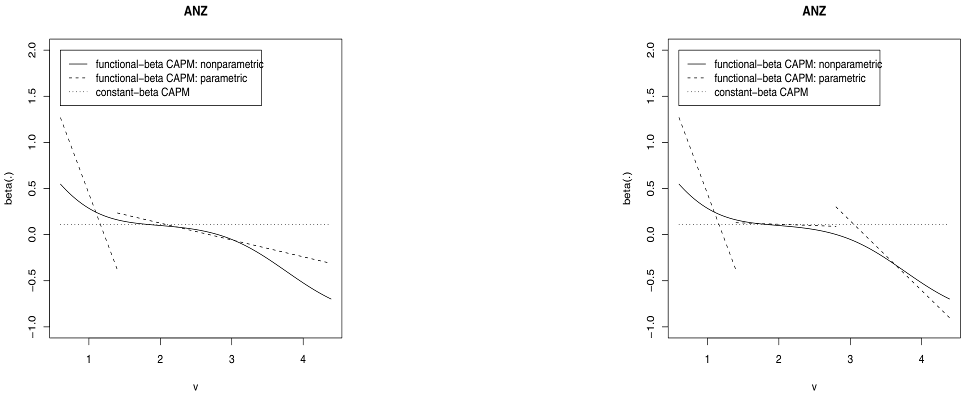


Fig. 10. For ANZ bank group limited:

	Energy	Finance	Healthcare	Materials	Fin-V-Prop	ANZ
Bandwidth	0.45	0.56	0.58	0.38	0.53	0.78

Table 3. Bandwidth Selection

Based on the chosen bandwidth in Table 3, the results of nonparametric estimation of beta functional can be plotted in graphs. For each of eight groups of data (mentioned in Section 3), we can have a curve of beta function plotted in the solid line in Figures 5–10, respectively. As most of the beta functions except for ANZ are positive, it means that the market return has positive effects on all individual sector return. Moreover, the time changing of the beta factor is obvious; it also shows that the individual returns are influenced by the market returns under conditions of market volatility at different levels. We will examine the findings of regime-switching phenomena from the nonparametric estimation more carefully by considering different parametric beta structures in CAPM, which are studied in the next subsection.

4.2 Parametric analysis

In this part we focus on further parametric investigation according to the previous work of nonparametric outcomes. How to specify parametric structures for functional beat? In a recent pioneering work of three-beta CAPM, (Galagedera and Faff 2005) made an assumption that the market conditions can play an important part in explaining a changing beta and could be divided into three states specified as -"low", "neutral" and "high". The nonparametric outcomes in Section 4.1 provide us with some possible ways of parametrization of the beta functional.

To capture the findings of regime-switching phenomena in functional-beta CAPM, we need to suitably specify the switching regimes of market condition in a parametric analysis of functional-beta. The difficult choice of the specific switching regimes of market condition can be suggested in accordance with the nonparametric outcomes of the functional beta, by which we can select reasonable changing points (thresholds) that are needed in parametric estimation. The problem of how many thresholds we should choose in the functional-beta model will be solved by Akaike's information criterion (AIC). This way, we shall have a general flexible functional-beta model which fits the financial data more adaptively.

With reference to the results of non-parametric estimates in Figures 5–10, the market volatility changing regime points σ_L^2 and σ_H^2 are quite amazingly very close to 1.4 and 2.8, respectively, for all the individual sector indexes and the ANZ stock, if we would apply a three-regime (two-threshold) CAPM as done in the three-beta CAPM of (Galagedera and Faff 2005). We therefore run the comparisons of the following five types of parametric models to examine which one appears more flexible and better fitted to the real data. (i) The first one is the traditional CAPM with a constant beta as coefficient. (ii) The second one is similar to the first one but it has a linear functional beta. (iii) In the third one, we divide the market volatility into two regimes and the beta functional can be parameterized as a two stepwise linear function. From the nonparametric estimates in Figures 5–10, it looks reasonable to set $\sigma_L^2 = 1.4$ as the threshold. (iv) In the fourth one, we divide the market volatility into three regimes and the beta functional can be parameterized as a three stepwise function, where we take $\sigma_L^2 = 1.4$ and $\sigma_H^2 = 2.8$. (v)The fifth model is the three-beta CAPM of (Galagedera and Faff 2005), with $\sigma_L^2 = 1.4$ and $\sigma_H^2 = 2.8$. Specifically,

(i) Traditional CAPM: $R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \qquad \beta_i = \beta_{i0}$
(ii)Linear-beta CAPM: $R_{it} = \alpha_i + \beta_{it} R_{mt} + \varepsilon_{it} \qquad \beta_{it} = \beta_{i0} + \beta_{i1}\sigma^2_{Mt}$
(iii) Two-regime (one-threshold) CAPM: $R_{it} = \alpha_i + \beta_{it,L}(I_{Lt}R_{Mt}) + \beta_{it,H}(I_{Ht}R_{Mt}) + \varepsilon_{it}$

$$\begin{cases} \beta_{it,L} = \beta_{i0} + \beta_{i1}\sigma^2_{Mt}, & \sigma^2_{Mt} \le \sigma^2_L \\ \beta_{it,H} = \beta_{i2} + \beta_{i3}\sigma^2_{Mt}, & \sigma^2_{Mt} > \sigma^2_L. \end{cases}$$

(iv) Three-regime (two-threshold) CAPM:
$R_{it} = \alpha_i + \beta_{it,L}(I_{Lt}R_{Mt}) + \beta_{it,N}(I_{Nt}R_{Mt}) + \beta_{it,H}(I_{Ht}R_{Mt}) + \varepsilon_{it}$

$$\begin{cases} \beta_{it,L} = \beta_{i0} + \beta_{i1}\sigma^2_{Mt}, & \sigma^2_{Mt} \le \sigma^2_L \\ \beta_{it,N} = \beta_{i2} + \beta_{i3}\sigma^2_{Mt}, & \sigma^2_L < \sigma^2_{Mt} \le \sigma^2_H \\ \beta_{it,H} = \beta_{i4} + \beta_{i5}\sigma^2_{Mt}, & \sigma^2_{Mt} > \sigma^2_H. \end{cases}$$

(v)Three-beta CAPM: $R_{it} = \alpha_i + \beta_{it,L}(I_{Lt}R_{Mt}) + \beta_{it,N}(I_{Nt}R_{Mt}) + \beta_{it,H}(I_{Ht}R_{Mt}) + \varepsilon_{it}$

$$\begin{cases} \beta_{it,L} = \beta_{i0}, & \sigma^2_{Mt} \le \sigma^2_L \\ \beta_{it,N} = \beta_{i2}, & \sigma^2_L < \sigma^2_{Mt} \le \sigma^2_H \\ \beta_{it,H} = \beta_{i4}, & \sigma^2_{Mt} > \sigma^2_H. \end{cases}$$

Here all the $\alpha_i, \beta_{i0}, \beta_{i1}, \beta_{i2}, \beta_{i3}, \beta_{i4}, \beta_{i5}$ are constants to be estimated by linear regression method.

One important problem in practice is the model selection, that is, which model is the best suitable for a real data set among Models (i)–(v). In order to verify which type of CAPM best suits each group of data respectively, Akaike's information criterion, AIC, is to be applied in this part by minimizing the value of $AIC(m)$, where m is the number of unknown parameters in the model. Note that all 5 models (i)–(v) can be expressed in a linear model in the form $R_i = (R_{i1}, \cdots, R_{iT})' = X\mathbf{b} + (\varepsilon_{i1}, \cdots, \varepsilon_{iT})'$ by suitably defining a $T \times m$ matrix X, where m is the number of parameters in each of the five CAPMs (See Table 4). Then we can define

$$\begin{aligned} AIC(m) &= T \log \widehat{\sigma}^2 + T(2m) \\ \widehat{R}_i &= HR_i \qquad H = X(X'X)^{-1}X' \end{aligned} \tag{20}$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(R_{it} - \widehat{R_{it}})'(R_{it} - \widehat{R_{it}}) = \frac{1}{n}R_i'(I-H)'(I-H)R_i. \tag{21}$$

The results in Table 4 show that except for the health care sector index selecting Model (v), all the other data sets select either Model (iii), Model (iv), which means the beta functional could be divided into two or three regimes. Models (i) and (ii) may be too simple to describe the relationship between market return and individual return.

The two-stepwise beta functional in Model (iii) estimated by using the common changing point σ^2_L for each data set is plotted in dashed line in the left panel of Figures 5–10, respectively, and the three-stepwise beta functional in Model (iv) in dashed line in the right panel of Figures 5–10, respectively. Obviously, due to the sparseness of highly extreme market volatility σ^2_{Mt} (c.f., Figure 4), the results of nonparametric estimation are poor and unreliable at

Model	(i): m=2	(ii): m=3	(iii): m=5	(iv): m=7	(v): m=4	chosen CAPM
Energy	3869.947	3799.963	3778.340	3779.698	3805.315	(iii)
Finance	3677.854	3650.706	3643.563	3645.087	3655.554	(iii)
HealthCare	3321.021	3318.320	3318.096	3320.025	3316.039	(v)
Materials	4137.633	4066.287	4044.999	4026.751	4077.107	(iv)
F-v-P Trusts	3778.767	3755.012	3746.462	3747.036	3758.679	(iii)
ANZ bank	4577.802	4571.992	4558.787	4558.482	4578.409	(iv)

Table 4. $AIC(m)$ and The Type of CAPM Chosen

extreme market volatility, while the parametric results of two-stepwise or three-stepwise beta functionals provide reasonable outcomes in Figures 5–10. Under moderate market volatility, both nonparametric and parametric outcomes are pretty consistent.

In (Galagedera and Faff 2005), the functional beta is assumed as three constants over three regimes, which is a special case of Model (iv) with $\beta_{i1} = \beta_{i3} = \beta_{i5} = 0$. To examine their work, here we test the significance of β_{i1}, β_{i3}, β_{i5} by applying T-statistics:

$$H_0 : \beta_{i1} = \beta_{i3} = \beta_{i5} = 0 \tag{22}$$

Applying linear regression method we get $\widehat{\mathbf{b}}$, and residuals $\widehat{r}$:

$$\hat{\mathbf{b}} = (X'X)^{-1}X'R_i, \ where\ \mathbf{b} = (\alpha_i, \beta_{i0}, \beta_{i1}, \beta_{i2}, \beta_{i3}, \beta_{i4}, \beta_{i5})', \tag{23}$$

and let δ stand for the standard deviation of $\widehat{r} = R_i - X\widehat{\mathbf{b}}$. Then the T-statistics value of each estimated parameter is $T_j = \widehat{b}_j / \{(XX^T)^{-1}_{jj}\delta^2\}^{1/2}$, where $\widehat{b}_j$ represents the jth element of $\widehat{\mathbf{b}}$. In a standard normal distribution, only 5% of the values fall outside the range plus-or-minus 2. Hence, as a rough rule of thumb, a t-statistic larger than 2 in absolute value would be significant at the significance level of 5%. The outcomes of the T-statistics with p-values for Models (iii), (iv) and (v) are listed in Tables 5–10, respectively, which indicate that the (Galagedera and Faff 2005)′ three-beta model is basically rejected except for the health care sector index.

	α_i	β_{i0}	β_{i1}	β_{i2}	β_{i3}	β_{i4}	β_{i5}	AIC
Threshold CAPM (2 step)								3778.340
Estimate	-0.0222	1.5001	-1.0176	-0.4794	0.5093			
T statistic	-0.4199	4.2287	-3.2583	-2.9629	7.9788			
p-value	0.6746	2.6e-05	0.0012	0.0031	4.1e-15			
Threshold CAPM (3 step)								3779.698
Estimate	-0.0192	1.5024	-1.0196	-0.5792	0.5527	0.5988	0.1933	
T statistic	-0.3625	4.2365	-3.2654	-2.2719	4.5061	0.8689	0.9408	
p-value	0.7171	2.5e-05	0.0011	0.0233	7.4e-06	0.3851	0.3470	
Three-beta CAPM (3 step)								3805.315
Estimate	-0.0336	0.3597		0.5410		1.2432		
T statistic	-0.6315	6.2018		9.2367		14.035		
p-value	0.5278	8.2e-10		1.5e-19		6.4e-41		

Table 5. Estimate, T-statistic and p-value for $\hat{\alpha}$ and each component of $\hat{\beta}$ in Models (iii), (iv) and (v) for Energy Sector Index

	α_i	β_{i0}	β_{i1}	β_{i2}	β_{i3}	β_{i4}	β_{i5}	AIC
Threshold CAPM (2 step)								3643.563
Estimate	0.0332	1.1911	-0.6470	-0.0248	0.2982			
T statistic	0.6728	3.5952	-2.2182	-0.1642	5.0018			
p-value	0.5013	0.0003	0.0268	0.8696	6.7e-07			
Threshold CAPM (3 step)								3645.087
Estimate	0.0355	1.1928	-0.6484	-0.2079	0.3883	0.8428	0.0377	
T statistic	0.7187	3.6012	-2.2237	-0.8730	3.3893	1.3094	0.1965	
p-value	0.4725	0.0003	0.0264	0.3829	0.0007	0.1907	0.8442	
Three-beta CAPM (3 step)								3655.554
Estimate	0.0255	0.4660		0.5790		0.9690		
T statistic	0.5174	8.6687		10.667		11.805		
p-value	0.6050	1.8e-17		3.3e-25		3.6e-30		

Table 6. Estimate, T-statistic and p-value for $\hat{\alpha}$ and each component of $\hat{\beta}$ in Models (iii), (iv) and (v) for Finance Sector Index

	α_i	β_{i0}	β_{i1}	β_{i2}	β_{i3}	β_{i4}	β_{i5}	AIC
Threshold CAPM (2 step)								3318.096
Estimate	0.0080	0.5473	-0.2468	-0.0246	0.1382			
T statistic	0.1909	1.9487	-0.9983	-0.1924	2.7342			
p-value	0.8486	0.0516	0.3183	0.8475	0.0064			
Threshold CAPM (3 step)								3320.025
Estimate	0.0082	0.5475	-0.2469	0.1961	0.0198	-0.0505	0.1589	
T statistic	0.1969	1.9494	-0.9990	0.9716	0.2043	-0.0926	0.9764	
p-value	0.8440	0.0515	0.3181	0.3315	0.8381	0.9263	0.3291	
Three-beta CAPM (3 step)								3316.039
Estimate	0.0054	0.2707		0.2365		0.4784		
T statistic	0.1308	5.9824		5.1772		6.9238		
p-value	0.8960	3.1e-09		2.7e-07		7.9e-12		

Table 7. Estimate, T-statistic and p-value for $\hat{\alpha}$ and each component of $\hat{\beta}$ in Models (iii), (iv) and (v) for Health care Sector Index

	α_i	β_{i0}	β_{i1}	β_{i2}	β_{i3}	β_{i4}	β_{i5}	AIC
Threshold CAPM (2 step)								4044.999
Estimate	-0.0213	1.8714	-1.3549	-0.2381	0.4821			
T statistic	-0.3518	4.6074	-3.7889	-1.2850	6.5962			
p-value	0.7251	4.6e-06	0.0002	0.1991	6.9e-11			
Threshold CAPM (3 step)								4026.751
Estimate	-0.0140	1.8769	-1.3594	-1.0294	0.8800	2.6250	-0.3886	
T statistic	-0.2341	4.6686	-3.8409	-3.5619	6.3288	3.3600	-1.6685	
p-value	0.8150	3.5e-06	0.0001	0.0004	3.8e-10	0.0008	0.0955	
Three-beta CAPM (3 step)								4077.107
Estimate	-0.0392	0.3529		0.7542		1.3335		
T statistic	-0.6415	5.3006		11.218		13.115		
p-value	0.5214	1.4e-07		1.5e-27		2.5e-36		

Table 8. Estimate, T-statistic and p-value for $\hat{\alpha}$ and each component of $\hat{\beta}$ in Models (iii), (iv) and (v) for Materials Sector Index

	α_i	β_{i0}	β_{i1}	β_{i2}	β_{i3}	β_{i4}	β_{i5}	AIC
Threshold CAPM (2 step)								3746.462
Estimate	0.0238	1.3180	-0.7423	-0.0411	0.3083			
T statistic	0.4574	3.7760	-2.4154	-0.2579	4.9091			
p-value	0.6475	0.0002	0.0159	0.7966	1.1e-06			
Threshold CAPM (3 step)								3747.036
Estimate	0.0267	1.3202	-0.7441	-0.2572	0.4141	1.0547	-0.0197	
T statistic	0.5133	3.7850	-2.4231	-1.0259	3.4323	1.5560	-0.0975	
p-value	0.6078	0.0002	0.0156	0.3052	0.0006	0.1200	0.9224	
Three-beta CAPM (3 step)								3758.679
Estimate	0.0146	0.4861		0.5820		0.9900		
T statistic	0.2780	8.5826		10.176		11.445		
p-value	0.7797	3.6e-17		3.4e-23		1.5e-28		

Table 9. Estimate, T-statistic and p-value for $\hat{\alpha}$ and each component of $\hat{\beta}$ in Models (iii), (iv) and (v) for Financial-x-trusts Sector Index

	α_i	β_{i0}	β_{i1}	β_{i2}	β_{i3}	β_{i4}	β_{i5}	AIC
Threshold CAPM (2 step)								4558.787
Estimate	0.0851	2.5144	-2.0707	0.4902	-0.1826			
T statistic	1.0836	4.7693	-4.4614	2.03844	-1.9251			
p-value	0.2788	2.1e-06	9.1e-06	0.0418	0.0545			
Threshold CAPM (3 step)								4558.482
Estimate	0.0903	2.5183	-2.0740	0.1687	-0.0282	2.4172	-0.7560	
T statistic	1.1504	4.7822	-4.4736	0.4455	-0.1547	2.3621	-2.4779	
p-value	0.2503	2.0e-06	8.6e-06	0.6560	0.8771	0.0184	0.0134	
Three-beta CAPM (3 step)								4578.409
Estimate	0.0453	0.1929		0.1149		-0.0954		
T statistic	0.5749	2.2468		1.3247		-0.7278		
p-value	0.5655	0.0249		0.1856		0.4669		

Table 10. Estimate, T-statistic and p-value for $\hat{\alpha}$ and each component of $\hat{\beta}$ in Models (iii), (iv) and (v) for ANZ bank group limited

5. Closure and future work

In this chapter, we have suggested a functional-beta single-index CAPM, extending the work of three-beta CAPM (Galagedera and Faff, 2004) that takes into account the condition of market volatility. Differently from the three-beta CAPM, we allow systematic risk β_i changing functionally with the market volatility σ_M^2, which is more flexible and adaptive to the changing structure of financial systems.

By using a set of stocks data sets collected from Australian stock market, empirical evidences of the functional-beta CAPM in Australia have been carefully examined under both nonparametric and parametric model structures. Differently from the three- or multi-beta (constant) CAPM in (Galagedera and Faff 2005), our new findings have convincingly demonstrated that the functional beta can be reasonably parameterized as threshold (regime-switching) linear functionals of market volatility with two or three regimes of market condition, taking as a special case the three-beta model of (Galagedera and Faff 2005) which was mostly rejected except for the health care sector index. In the condition of extreme market volatility, a parametric threshold functional-beta CAPM is found useful, which is of potential interest in exploring the Black Swan effect of the merits of beta in the presence of large market fluctuations.

The CAPM provides a usable measure of risk that helps investors determine what return they deserve for putting their money at risk. Our new model is no doubt helpful to better understand the relationship between risk and return under different market conditions. It can be potentially applied widely, for example, it may be useful both for market investors and financial risk managers in their investment/management decision-making, such as portfolio selection.

In addition, as done in (Galagedera and Faff 2005), it is interesting to investigate how the functional beta systematic risk is priced in the real financial assets.

We shall leave the above questions for future work.

6. Acknowledgements

The work of this research was partially supported by the Discovery Project and Future Fellowship grants from the Australian Research Council, which are acknowledged.

7. References

[1] Allen, D.E., McAleer, M. and Scharth, M. (2008). Realized Volatility Uncertainty. School of Accounting, Finance and Economics & FEMARC Working Paper Series, Edith Cowan University, August 2008 Working Paper 0807.

[2] Ball, R., Brown, P. and Officer, R., (1976). Asset Pricing in Australian Industry Equity Market. Australian Journal of Management 1, 1-32.

[3] Blume, M.E. 1971, On the assessment of risk, Journal of Finance, vol. 26, no. 4, pp. 275–88.

[4] Blume, M.E. 1975, Betas and the regression tendencies, Journal of Finance, vol. 30, no. 3, pp. 785–95.

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**, 307–327.

[6] Bos, T. & Ferson, T.A. 1992, Market model nonstationarity in the Korean Stock Market, in Pacific Basin Capital Market Research, vol. 3, Elsevier Science Publishers, Amsterdam.

[7] Bos, T. & Ferson, T.A. 1995, Nonstationarity of the market model, outliers, and the choice of market rate of return, in Advances in Pacific-Basin Financial Markets, vol. 1, eds. T. Bos & T.A. Ferson, JAI Press, Greenwich, CT.

[8] Bos, T., Ferson, T.A., Martikainen, T. & Perttunen, J. 1995, The international co-movements of Finnish stocks, The European Journal of Finance, vol. 1, pp. 95–111.

[9] Bos, T. & Newbold, P. 1984, An empirical investigation of the possibility of stochastic systematic risk in the market model, Journal of Business, vol. 57, no. 1, pp. 35–42.

[10] Brailsford, T. and Faff, R., (1997). Testing the conditional CAPM and the effect of Intervaling.*A Pacific-Basin Finance Journal* 5, 527-537.

[11] Brooks, R.D., Faff, R.W. & Lee, J.H.H. 1992, The form of time variation of systematic riskŮSome Australian evidence, Applied Financial Economics, vol. 2, pp. 191–8.

[12] Brooks, R.D., Faff, R.W. & Lee, J.H.H. 1994, Beta stability and portfolio formation, Pacific Basin Finance Journal, vol. 2, pp. 463–79.

[13] Brooks, R.D., Faff, R.W. & McKenzie, M.D. 1998, Time-varying beta risk of Australian industry portfolios: A comparison of modelling techniques, Australian Journal of Management, vol. 23, no. 1, June, pp. 1–22.

[14] Brooks, R.D., Faff, R.W., McKenzie, M.D. & Mitchell, H. 2000,ŠA multi-country study of power ARCH models and national stock market returns, Journal of International Money and Finance, vol. 19. pp. 377–9.

[15] Brooks, R.D. & King, M.L. 1994, Hypothesis testing of varying coefficient regression models: Procedures and applications, Pakistan Journal of Statistics, vol. 10, pp. 301–58.

[16] Cheng, J.W. 1997, A switching regression approach to the stationarity of systematic and non-systematic risks: The Hong Kong experience, Applied Financial Economics, vol. 7, no. 1, pp. 45–58.

[17] Carvalho, J.P., R. B. Durand and Hock Guan, (2002). Australian Internet Stocks: Were They Special? University of Western Australia Department of Accounting and Finance Working Paper 2002-151.

[18] Collins, D.W., Ledolter, J. & Rayburn, J. 1987, Some further evidence on the stochastic properties of systematic risk, Journal of Business, vol. 60, no. 3, pp. 425–48.

[19] J. Crombez en R. Vander Vennet (2000).The risk/return relationship conditional on market movements on the Brussels stock Exchange. *Tijdschrift voor Economie en Management*.

[20] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of United Kingdom inflation. *Econometrica* **50,** 987-1007.

Estrada, J. and Vargas, M. (2011). Black Swans, Beta, Risk, and Return. Available at: web.iese.edu/jestrada/PDF/Research/Others/BlackSwans-Beta.pdf

[21] Fabozzi, F. & Francis, J. 1978, Beta as random coefficient, Journal of Financial and Quantitative Analysis, vol. 13, no. 1, pp. 106–16.

[22] Faff, R.W., (1991). A Likelihood Ratio Test of the Zero-Beta CAPM in Australian Equity. Accounting and Finance 31, 88-95.

[23] Faff, R.W., (1992). A Multivariate Test of an Equilibrium APT with Time-Varying Risk Premia in Australian. Journal of Management 17, 233-258.

[24] Faff, R., (2001). ŞA Multivariate Test of a Dual Beta CAPM: Australian EvidenceŤ, Financial Review , Vol. 36, No. 4, pp. 157–174.

[25] Faff, R.W., Brooks, R. Fan and T.P.,(1992).*A TEST OF A NEW DYNAMIC CAPM*, http://www.sirca.org.au/Papers/1999039.pdf.

[26] Faff, R.W. & Brooks, R.D. 1998, Time-varying beta risk for Australian industry portfolios: An exploratory analysis, Journal of Business Finance & Accounting, vol. 25, no. 5 & 6 June/July, pp. 721–45.

[27] Faff, R.W. and Lau, S. (1997). A Generalised Method of Moments Test of Mean Variance Efficiency in the Australian Stock Market. Pacific Accounting Review 9, 2-16.

[28] Faff, R.W., Lee, J.H.H. & Fry, T.R.L. 1992, Time stationarity of systematic risk: Some Australian evidence, Journal of Business Finance and Accounting, vol. 19, pp. 253–70.

[29] Fama, E., and K. French. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.

[30] Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.

[31] Fan, J., Yao, Q. and Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. R. Statist. Soc. B*, **65**, 57–80.

[32] Galagedera, D.U. and Faff, R.W, (2005). Modeling the Risk and Return Relation Conditional on Market Volatility and Market Conditions. International Journal of Theoretical and Applied Finance, *Vol.8 ,No.1*, 75-94.

[33] Gangemi, M., Brooks, R.D. & Faff, R.W. 2000, Modelling AustraliaŠs country risk: A country beta approach, Journal of Economics and Business, vol. 52, pp. 259–76.

[34] Gao, J. (2007). Nonlinear time series: semiparametric and nonparametric methods. Chapman & Hall/CRC, Boca Raton.

[35] Groenewold, N. & Fraser, P. 1999, Time-varying estimates of CAPM betas, Mathematics and Computers in Simulation, vol. 48, pp. 531–9.

[36] Josev, T., Brooks, R.D. & Faff, R.W. 2001, Testing a two-factor APT model on Australian industry equity portfolios: The effect of intervaling, Applied Financial Economics, vol. 11, pp. 157–63.

[37] Kim, D. 1993, The extent of non-stationarity of beta, Review of Quantitative Finance and Accounting, vol. 3, pp. 241–54.

[38] Kim, M.K., and J.K. Zumwalt (1979).An Analysis of Risk in Bull and Bear Markets.*Journal of Financial and Quantitative Analysis*, 14, 1015-1025.

[39] Lintner, J (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets.*Review of Economics and Statistics* ,47, 13-37.

[40] Lu, Z., Steinskog, D.J., Tjøstheim, D. and Yao, Q. (2009). Adaptively varying-coefficient spatiotemporal models. *J. R. Statist. Soc. B*, **71**, Part 4, 859–880.

[41] Ohlson, J. & Rosenberg, B. 1982, Systematic risk of the CRSP equal-weighted common stock index: A history estimated by stochastic parameter regression, Journal of Business, vol. 55, no. 1, pp. 121–45.

[42] Pope, P. & Warrington, M. 1996, Time-varying properties of the market model coefficients, Accounting Research Journal, vol. 9, no. 2, pp. 5–20.

[43] Sharpe, William F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk.*Journal of Finance* 19, 425-442.

[44] Simonds, R.R., LaMotte, L.R. & McWhorter, A. Jr., 1986, Testing for non-stationarity of market risk: An exact test and power consideration, Journal of Financial and Quantitative Analysis, vol. 21, no. 2, pp. 209–20.

[45] Smith, L.,(1997). Risk Management and Financial Derivatives: A Guide to the Mathematics. 1-3.

[46] Stone, M. (1974). Cross-validation and multinomial prediction. Biometrika, 61, 509–515.

[47] Taleb, N. (2007). The Black Swan. The Impact of the Highly Improbable. Random House.

[48] Tong, H. (1990). *Nonlinear time series: a dynamical system approach*. Oxford University Press, Oxford.

[49] Wells, C. 1994, Variable betas on the Stockholm exchange 1971–1989, Applied Economics, vol. 4, pp. 75–92.

[50] Wood, J., (1991). A Cross-Sectional Regression Test of the Mean-Variance Efficiency of an Australian Value Weighted Market Portfolio. Accounting and Finance. **18**, 96-109.

[51] Yao,J. and Gao, J. (2004). Computer-Intensive Time-Varying Model Approach to the Systematic Risk of Australian Industrial Stock Returns. Australian Journal of Management, Vol. 29, No.1, 121–145.

18

Linking U.S. CDS Indexes with the U.S. Stock Market: A Multidimensional Analysis with the Market Price and Market Volatility Channels

Hayette Gatfaoui
Rouen Business School,
France

1. Introduction

The recent mortgage subprime crisis and the partly resulting global financial crisis shed light on the weaknesses and required enhancements of prevailing risk management practices (e.g. Basel 2 limitations and frailties). Among the most important enforcements, liquidity concerns, counterparty credit risk, the correlation between various risks (e.g. contagion risk) and model stress testing as well as related scenario analysis have been highlighted by the Basel Committee on Banking Supervision under the Basel 3 framework (Basel Committee on Banking Supervision, 2010). With regard to liquidity, various liquidity measures coupled sometimes with a stressed scenario analysis are proposed at both the level of financial assets and the bank level in in- and off-balance sheet prospects (Blundell-Wignall and Atkinson, 2010; Lumsdaine, 2011; Van Den End, 2011). On the correlation viewpoint, the risk of correlation between risks (e.g. impacts of liquidity risk on market risk and *vice versa*) refers to the linkages between asset classes, and linkages between banks/financial institutions among others (Embrechts *et al.*, 2002, 2003; Hull and White, 2006; Lumsdaine, 2011). On the stress testing and scenario viewpoints, the mitigation of potential model risk and measurement errors is targeted in risk assessment and risk forecast prospects (Basel Committee on Banking Supervision, 2010; Ferrari *et al.*, 2011).

Current research suggest that the credit market, as represented by credit default swaps or corporate bond spreads, is highly sensitive to the stock market trend and/or the corresponding market volatility as represented by equity volatility (Dupuis *et al.*, 2009; Gatfaoui, 2010; Scheicher, 2009; Vaz de Melo Mendes and Kolev, 2008). Under the Basel 3 setting, we focus concurrently on the correlation risk between credit default swap (CDS) spreads and referenced market risk components (Norden and Weber, 2009). Specifically, CDS spreads represent a credit risk proxy whereas market risk is envisioned with respect to two dimensions, namely a market price risk and a market volatility risk (Gatfaoui, 2010). The market price risk illustrates the impact of the global market trend (i.e. common general link within the stock market) whereas the market volatility risk represents the magnitude of global market moves (i.e. volatility feedback, liquidity concerns). In this lens, we study the asymmetric linkages between CDS spreads and both market price and market volatility risks, and we analyze the interaction between such linkages through a three-dimensional copula methodology. In particular, we assess the impact of the stock market trend on the credit market trend while describing also how the magnitude of stock market moves impacts the magnitude of credit market moves. Hence, we address simultaneously two dependent questions, which illustrate the general market behavior. First, does the stock market trend drive the directional moves of the credit market? Second, does the magnitude of stock market moves influence the magnitude of credit market moves?

The previous multivariate setting targets a sound assessment of credit risk in the light of the stock market's influence with respect to the curse of dimensionality puzzle. The curse of dimensionality refers to the trade-off between the number of parameters, the problem's dimension and the sample size in order to ensure a sound model estimation process (i.e. convenient and acceptable risk representation). Indeed, the sample size constrains the number of model parameters with respect to the dimensionality of the problem under consideration (Bellman, 1961; Liebscher, 2005; Weiss, 2011). Specifically, the accuracy of the estimation procedure relies on an exponential link between the problem dimensionality and the corresponding required number of data. Moreover, the observed multivariate dependence structures exhibit a negative link between CDS spreads and market price risk (i.e. trend impact) on one side, and a positive link between CDS spreads and market volatility risk (i.e. magnitude effect) on the other side. Taking into account simultaneously the dependence of CDS spreads relative to both market price and market volatility channels, allows for a better assessment of the correlation risk between the credit market and the stock market. Indeed, finer risk scenarios can be raised with respect to the market trend (i.e. bad or good market signal) and its corresponding volatility impact (i.e. strength of the signal).

In this lens, this chapter is organized as follows. Section 2 introduces the data as well as related statistical properties and stylized features. In particular, the directional impact of the stock market trend on CDS spread changes is quantified while the influence of the stock market volatility on the magnitude of CDS spread moves is also measured. Then, section 3 introduces a three-dimensional copula study, which measures the impact of the stock market trend and volatility on CDS spreads. As an extension, a scenario analysis is provided in section 4 to study the sensitivity of CDS spreads relative to the two stock market channels. Specifically, we assess the corresponding tail dependence, namely the extent to which extreme variations in both stock market trend and volatility impact extreme variations in CDS spreads. Finally, section 5 draws some concluding remarks and proposes some future research insights.

2. Data and stylized features

We introduce the stock market and credit market data under consideration as well as their corresponding statistical patterns.

2.1 Data

We consider two categories of data; - 1) U.S. stock market indexes and 2) credit default swap data, which focus on both North America and emerging markets. Our daily data consists of closing quotes extracted from Reuters, and ranging at most from September 28th 2007 to March 24th 2010, namely 618 observations per data series. With regard to the first category of data, we consider the logarithmic returns of the Standard & Poor's 500 stock market index in basis points and the level of the CBOE implied volatility index. Specifically, those two indexes are considered to be a proxy of the two complementary dimensions of market risk, namely the market price risk and the market volatility risk (see Gatfaoui, 2010). With regard to the second category of data, the credit default swap data come from credit derivatives indexes, which are delivered by Markit Corporation. In particular, we consider the spreads of Markit credit default swap indexes, or equivalently, the spreads of credit derivatives indexes that we name Markit CDX spreads. Those CDX indexes are split into two groups among which one set of spreads focuses on reference entities domiciled in North America while the other one relates to reference entities domiciled in emerging markets (see Table 1). In particular, the CDXEM index focuses only on sovereign entities whereas the CDXED relates to both corporate and sovereign entities. Moreover, the crossover index accounts for potential rating divergences between Standard & Poor's and Moody's rating agencies with respect to the lowest investment grade and highest speculative grade ratings (i.e. divergences relative to the frontier between investment and speculative grade borrowers). Furthermore, the CDX spreads under consideration are expressed in basis point and consist of the mid-market quotes on individual issuers.[1] Incidentally, CDXEM spread data range from February 1st 2008 to March 24th 2010, namely 538 observations per data series.

CDS label	Detail about reference entities and indices*
CDXEM	Emerging Market
CDXED	Emerging Market Diversified
CDXHY	North America Investment Grade High Yield
CDXHB	North America Investment Grade High Yield and B-rated
CDXBB	North America Investment Grade High Yield and BB-rated
CDXIG	North America Investment Grade
CDXIV	North America Investment Grade High Volatility
CDXXO	North America Crossover
SP500	Standard & Poor's 500 stock index
VIX	CBOE Implied Volatility Index

* Emerging market entities consist of Africa, Asia, Eastern Europe, Latin America and Middle East.

Table 1. Markit CDS indexes and stock market indices

2.2 Stylized features

We analyze the daily changes in CDX spreads, SP500 returns and VIX level. Incidentally, changes in CDX spreads reflect changes in credit markets and credit conditions. In particular, the direction of CDX spread moves (i.e. the sign of daily changes) illustrates the credit market status whereas the magnitude of CDX spread moves refers to the strength and health of the credit market. Indeed, increasing CDX spreads indicate a hazardous credit market so that lending becomes riskier (Fisher, 1959). Moreover, the larger the spread increase is, the riskier borrowing becomes and the higher the corresponding credit risk level is. In such a case, tougher credit conditions, which may

[1] The spreads are computed against corresponding LIBOR rates. The reader is invited to consult Markit Corporation's website at http://www.markit.com/en/ for further information.

illustrate a lack of funding liquidity or a deterioration in the corporate bond market liquidity among others (Brunnermeier, 2009; Das and Hanouna, 2009; Longstaff *et al.*, 2005), yield a weakened credit market because borrowing becomes more expensive in order to compensate for the increased risk of not refunding lenders (i.e. higher credit risk premium). Along with the referenced linkages between the credit and stock markets, we investigate the relationships between CDX spread changes and the variations in stock market conditions. Specifically, we focus on the link prevailing between CDX spreads changes on one side, and changes in both SP500 returns as well as VIX level. For this prospect, we first focus on the statistical properties of our time series (see Table 2). Changes in CDX spreads as well as market indexes exhibit mitigated skewness values and a positive excess kurtosis, underlining then their asymmetric and fat-tailed behavior over time. In particular, apart from CDXBB, CDXIV and CDXXO spreads, CDX spread changes exhibit a positive skewness, which underlines generally above average spread variations from one day to another. Moreover, we performed five different normality tests in order to cope with the asymmetric nature and the tail behavior of the considered financial times series.

All the statistics reject the Gaussian assumption at a five percent test level. Thus, CDX spread changes and variations in stock market indexes do not exhibit a Gaussian behavior. However, a complementary unit root test emphasizes the stationary behavior of the observed daily changes in both credit and market data as underlined by the Dickey-Fuller (DF) as well as the Z_τ and Z_ρ Phillips-Perron statistics for a unit lag (Dickey and Fuller, 1979; Fuller, 1996, Hamilton, 1994; MacKinnon, 1994; Newey and West, 1987; Phillips and Perron, 1988). Indeed, the stationary behavior is validated at both the one percent and the five percent test levels.[2]

Index	Mean	Std. Dev.	Excess kurtosis	Skewness	DF	Z_τ	Z_ρ	Normality tests*
CDXEM	0.0186	22.2388	33.2281	2.4708	-17.482	-17.546	-396.326	NO
CDXED	0.2042	22.1535	23.0788	2.2965	-23.693	-23.697	-590.169	NO
CDXHY	0.1426	27.8055	5.0361	0.0902	-20.399	-20.511	-513.335	NO
CDXHB	0.1361	25.1582	10.3125	0.3582	-20.090	-20.472	-537.111	NO
CDXBB	-0.0259	23.3170	26.5870	-0.0780	-27.804	-27.666	-724.143	NO
CDXIG	0.0681	6.1093	10.0182	0.0247	-21.775	-21.650	-504.261	NO
CDXIV	0.0340	11.4271	20.7487	-1.2384	-19.592	-19.658	-481.770	NO
CDXXO	0.0097	13.1560	15.5236	-0.6810	-31.220	-30.828	-796.120	NO
SP500	-0.0081	307.0674	4.5221	0.4861	-43.540	-753.123	-74.658	NO
VIX	0.0827	2.7742	8.9435	0.1618	-28.914	-624.136	-30.061	NO

* Jarque-Bera, Lilliefors, Cramer-Von-Mises, Watson, Anderson-Darling at a 5% level.

Table 2. Descriptive statistics of daily changes in CDX spreads and stock market factors

[2] The augmented Dickey-Fuller and Phillips-Perron unit root tests are performed without trend and constant terms, the Phillips-Perron test being robust to heteroskedasticity (e.g. serial correlation). The one- and five-percent critical levels are -2.5800 and -1.9500 for Dickey-Fuller test. Differently, the one- and five-percent critical levels correspond to -2.5800 and -1.9500 for Phillips-Perron Z_τ statistic whereas such critical levels consist of -13.8000 and -8.1000 for Phillips-Perron Z_ρ statistic.

Then, we control for an existing link based on the nonparametric Kendall and Spearman correlation coefficients (see Table 3). Indeed, the non-Gaussian behavior of our time series advocates the use of an appropriate correlation measure, which accounts for asymmetry and potential fat tails.

Spread	**Kendall correlation with**		**Spearman correlation with**	
	SP500	**VIX**	**SP500**	**VIX**
CDXEM	-0.2676	0.4200	-0.3634	0.5644
CDXED	-0.0916	0.2191	-0.1329	0.3114
CDXHY	-0.2423	0.4077	-0.3369	0.5627
CDXHB	-0.1382	0.3038	-0.1949	0.4246
CDXBB	-0.1409	0.2861	-0.1999	0.4042
CDXIG	-0.2887	0.4196	-0.4072	0.5779
CDXIV	-0.2107	0.3484	-0.2975	0.4887
CDXXO	-0.1772	0.2749	-0.2584	0.3912
SP500	1.0000	-0.4368	1.0000	-0.5533
VIX	-0.4368	1.0000	-0.5533	1.0000

Table 3. Kendall and Spearman correlations between CDX spread changes and changes in both SP500 and VIX

The obtained correlation estimates emphasize the significance of the correlation between the CDS market and the U.S. stock market. Indeed, all the correlation coefficients are significant at a five percent two-tailed Student t-test level. As expected, the link between CDS spreads and market price is negative whereas the link between CDS spreads and market volatility is positive. Moreover, the correlation between market prices and corresponding market volatility is negative. Such a pattern illustrates the well-known volatility feedback effect, which was formerly introduced by Black (1976). In particular, CDX spreads tend to increase when both stock market returns decrease and the corresponding stock market volatility augments. Conversely, CDX spreads tend to decrease when both SP500 returns increase and VIX level diminishes. Hence, a downward stock market trend and upward stock market volatility coincide both with upward CDS spreads. As a result, the statistical properties of stock- and credit market data support the investigation of a joint link between CDX spreads on one side, and both SP500 returns and VIX index on the other side (i.e. stock market price and volatility indexes). Such linkages are of high significance specifically when CDX spread moves are large (i.e. extreme variations and tail behaviors).

Importantly, the linkages prevailing between CDS spreads and the stock market volatility originate their full meaning from existing findings. In particular, market liquidity is strongly related to the corresponding market volatility (Brunnermeier and Pedersen, 2009) so that volatility encompasses liquidity features. Moreover, the interaction, or equivalently, the correlation between market risk and credit risk is well acknowledged nowadays (Brigo *et al.*, 2011; Hartmann, 2010; Liu *et al.*, 2006; Predescu *et al.*, 2009). Specifically, CDS spreads constitute a proxy of credit risk but also encompass a liquidity premium, which evolves both on a cross-sectional basis and over time (Predescu *et al.*, 2009). Hence, studying the linkages between CDX spreads and stock market volatility makes fully sense in terms of the impact or transmission of liquidity shocks from the stock market to the credit market. Basically, such linkages illustrate how aggregate CDS spreads react to modifications in funding means among others. The previous linkages are therefore emphasized by the significant correlation coefficients between CDX spread changes and VIX index variations.

Focusing on the dependencies between CDX spreads and market indexes, we then investigate graphically the existence of such links. For this prospect, we plot the CDX spread changes against changes in SP500 index on one side, and changes in VIX implied volatility index on the other side (see Fig. 1 and Fig. 2). The two-dimensional plots exhibit clearly linkages between CDX spreads and both market price and market volatility. However, such bivariate linkages are asymmetric (Gatfaoui, 2010). Moreover, we notice clearly differences between the dependence structures of CDX spread changes with respect to SP500 changes, and the dependence structures of CDX spread changes with respect to VIX changes. Furthermore, Fig. 2 exhibits flatter relationships with respect to CDXED spread changes.

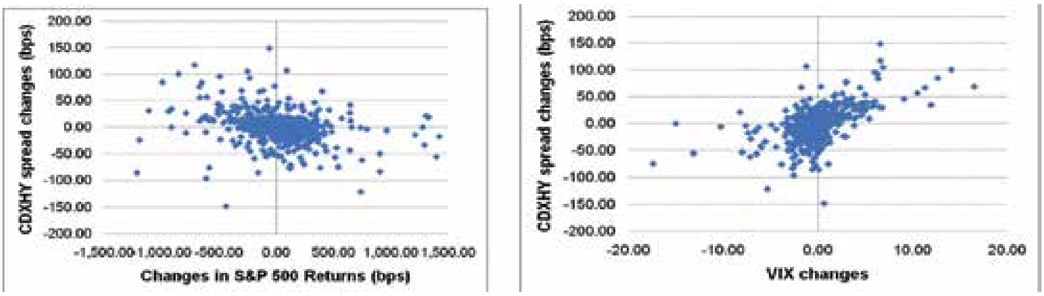


Fig. 1. Dependence structures of CDXHY spreads with both SP500 and VIX indexes

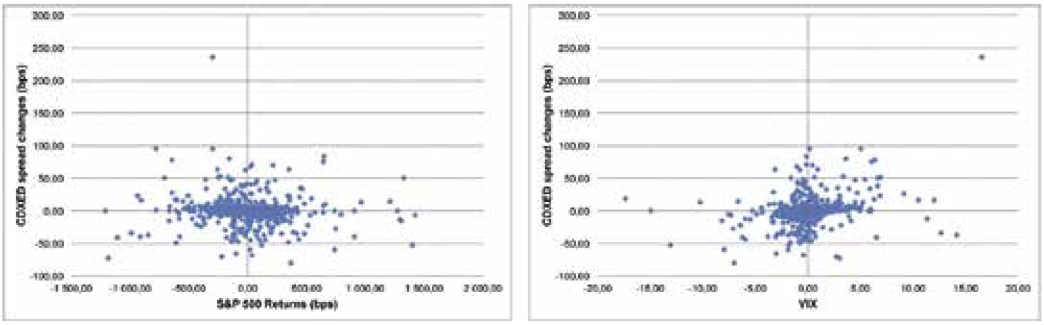


Fig. 2. Dependence structures of CDXED spreads with both SP500 and VIX indexes

As a complementary investigation, Fig. 3 aggregates the previous two-dimensional linkages while considering the three-dimensional dynamics of CDX spread, SP500 return and VIX level daily changes. Seemingly, the three-dimensional relationships reveal to be elliptical. Again, Fig. 3 exhibits flatter relationships with respect to CDXED spread variations.

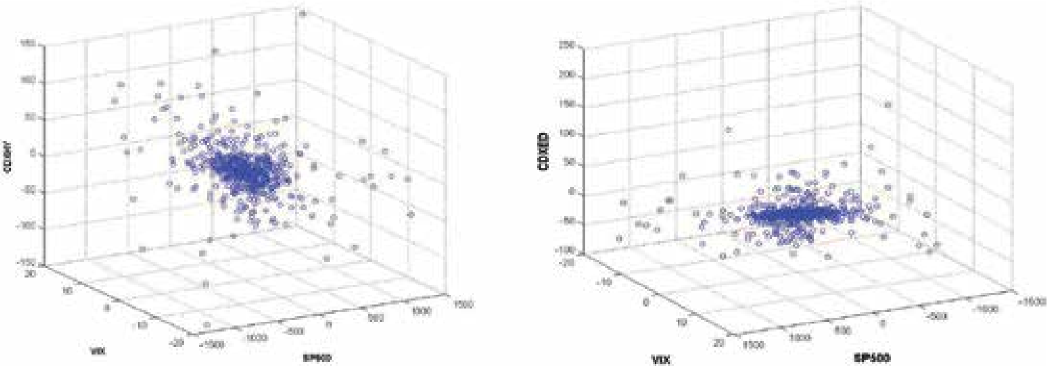


Fig. 3. Dependence structures of CDXHY and CDXED spreads with respect to both SP500 and VIX indexes

As a result, there exist negative linkages between CDX spread changes and SP500 return changes, and positive linkages between CDX spread changes and VIX changes. Such two-dimensional linkages reveal to be asymmetric and the two types of dependence structures of CDX spread changes relative to SP500 return changes and VIX changes respectively exhibit noticeable differences. However, according to Fig. 3, the three-dimensional joint behavior is rather elliptical, and exhibits some symmetric pattern from a tail prospect.

Such empirical findings are important for two main reasons. First, they confirm the well-known negative link between price and volatility, which was formerly referenced in the equity market (Black, 1976). Second, they underline competing effects with respect to the impact of equity prices, on the one hand, and equity volatility, on the other hand, on CDX spreads. With regard to the first reason, the negative link reflects an intuitive and straightforward situation. Indeed, equity volatility tends to increase during market disturbances and bear market times while prices are decreasing or low. Conversely, equity volatility tends to decrease during calm market periods and bull market trends while prices are generally increasing. One explanation for such an asymmetric volatility relies on the investors' behaviors. Basically, investors want all to get rid of their bad stocks when prices are decreasing, which generates a higher market activity and therefore a higher volatility of stock prices. Since many investors want to sell the same stocks at the same times, such common behaviors generate large price declines so that prices fluctuate importantly over time. That's also why volatility is often related to liquidity (Brunnermeier and Pedersen, 2009), namely the propensity of a stock (or an asset) to be transformed into cash. With regard to the second reason, the credit market is closely linked with the equity market (Brigo *et al.*, 2011; Fisher, 1959; Hartmann, 2010; Liu *et al.*, 2006; Predescu *et al.*, 2009). Therefore, two distinct linkages between the credit market and the equity market need to be considered. The first linkage is the negative impact of equity prices on CDX spreads whereas the second linkage is the positive impact of equity volatility on CDX spreads. As a result, the evolution of CDX spreads over time results from a tradeoff between the two previous market channels. In particular, an increase in CDX spreads is associated to both a decreasing market trend and an increasing market volatility. Hence, the first price channel gives information about the directional impact of the stock market on CDX spreads whereas the second market volatility channel indicates the magnitude of such an impact on CDX spread changes.

The decomposition of the two stock market channels helps identify more precisely the market risk and therefore its link with credit risk. Specifically, we target to assess the risk that CDX spreads cross upward a specific high threshold given that the stock market price and volatility channels also cross some corresponding downward and upward thresholds respectively. Recall that growing CDX spreads indicate a riskier lending activity and therefore a risk of not being refunded for lenders. The higher CDX spreads become or the more they grow, the riskier the credit risk is. In general, risk assessment and risk management techniques focus on the probability of occurrence of bad scenarios and the impact of such scenarios (e.g. credit losses and their consequences). In our case, a bad scenario is so that CDX spreads increase while the stock market trend is downward and the stock market volatility increases. The strength of each scenario depends on the chosen thresholds for credit and equity market data. Thus, the market price and volatility channels have a huge significance for risk management practitioners because a more accurate link between credit risk and market risk can be drawn.

As an example, we consider the changes in CDX spreads and corresponding changes in VIX and SP500 returns. We rank by ascending order those CDX spread changes and we also take the value, which indicates the beginning of the 10% largest CDX spread increases among the sample (i.e. the 10% biggest positive changes). Such a threshold value highlights the beginning of the 10% highest CDX spread sample, which is also called an upper tail. The upper tail stresses critical high values for the CDX spreads under consideration, and therefore describes high credit risk scenarios (e.g. extreme CDX spread values). Then, we count the number of cases when the sign of CDX spread changes coincides with the sign of the changes in both SP500 returns and VIX level (i.e. same directional moves). Such a counting process is achieved for the whole sample of CDX spreads and the corresponding 10% upper tail. The coincidence study between the changes across the whole sample illustrates the average correlation risk between credit risk and market risk factors whereas the coincidence study across the upper tail sample underlines some extreme correlation risk. We report in Table 4 the percentage of sign coincidence between CDX spread changes and changes in both SP500 returns and VIX level. Setting the 50% threshold as a discriminant value, CDX spreads evolve in the opposite direction of SP500 returns and in the same direction as VIX level across the whole sample and on an average basis. With respect to the upper tail sample of CDX spreads, high CDX spread changes generally evolve in the same way as the whole sample, and such behavior is even enforced. Indeed, the percentage of similar directional moves between high CDX spreads and SP500 returns diminishes whereas the percentage of matching directional moves between high CDX spreads and VIX level increases. Thus, we start considering some critical spread variations in relation with the corresponding changes in the two referenced stock market channels, namely SP500 returns and VIX level.

Therefore, managers can easily focus on three specific questions. First, can a critical threshold in stock market prices be linked to a critical threshold in CDX spreads? Second, can a critical threshold in stock market volatility be linked to a critical threshold in CDX spreads? Finally, can critical thresholds in both stock market price and stock market volatility be linked to a critical threshold in CDX spreads? Such management issues can be difficult to answer unless we understand the linkages between the credit and stock markets. Moreover, a better understanding also yields the use of more appropriate techniques for risk assessment prospects. Indeed, Fig. 1 and Fig. 2 exhibit clearly the non-linear nature of the previous linkages else we would clearly notice straight lines on those figures. Furthermore, the two stock market channels require considering a three-dimensional setting in order to assess the link between credit risk and market risk. As an extension, Fig. 3 also advocates the non-linear nature of the linkages between the two stock market channels and CDX spreads.

Index	Global coincidence		Upper tail coincidence*	
Spread	SP500	VIX	SP500	VIX
CDXEM	31.9287	63.3712	8.9552	34.3284
CDXED	39.0600	58.1848	38.7097	56.4516
CDXHY	35.6564	70.5024	25.8065	88.7097
CDXHB	41.0049	64.9919	30.6452	74.1935
CDXBB	41.3290	65.9643	33.8710	69.3548
CDXIG	35.4943	70.6645	17.7419	85.4839
CDXIV	38.8979	67.7472	22.5806	77.4194
CDXXO	41.1669	64.5057	17.7419	79.0323

* It is the sample of the 10% highest CDX spread increases.

Table 4. Percentage of sign coincidence between CDX spread changes and changes in both SP500 and VIX

3. A multivariate copula application

The previous stylized facts advocate the use of an appropriate statistical tool to handle simultaneously the dependence structure between the CDS market and the two components of market risk, namely market price and market volatility risks. For this purpose, we introduce the three-dimension copulas under consideration, the corresponding data fitting process and the selection criterion of the best copula model.

3.1 Copulas

Copulas are a useful tool to model multivariate dependence structures (Cherubini *et al.*, 2004; Durrleman *et al.*, 2000; Embrechts *et al.*, 2003; Genest *et al.*, 1995; Joe, 1997; McNeil *et al.*, 2005; Nelsen, 1999; Patton, 2009; Sklar, 1959, 1973). They present the advantage of not necessarily having to determine the distribution function of each of the variable under consideration. Hence, it is possible to specify the global dependence structure without knowing the margins (i.e. univariate distribution functions) of each variable under consideration. As a consequence, the corresponding model risk is minimized. As an example, Fig. 4 plots the empirical copula function, which describes the bivariate dependence structure of CDXHY spread changes with respect to SP500 changes on one side, and VIX changes on the other side (Deheuvels, 1979, 1980). The observed empirical behavior can easily be linked to the theoretical behavior of some well-known copula representations (Cherubini *et al.*, 2004; Joe, 1997; Nelsen, 1999).

Along with the previous framework, we focus on the three-dimensional representations of the dependence structures of CDX spread changes and the changes in the two market risk channels. Specifically, Sklar's theorem (1959) introduces a three-dimensional copula function as follows:

Theorem: Let F be a joint distribution function with margins F_X, F_Y and F_Z. Then there exists a copula C such that for all x, y, z in the real number set,

$$F(x,y,z) = \text{Prob}(\text{X} \le \text{x}, \text{Y} \le \text{y}, \text{Z} \le z) = C\left(F_X(x), F_Y(y), F_Z(z)\right) \tag{1}$$

If F_X, F_Y and F_Z are continuous, then C is unique; otherwise, C is uniquely determined on $\text{Ran}F_X \times \text{Ran}F_Y \times \text{Ran}F_Z$. Conversely, if C is a copula and F_X, F_Y and F_Z are distribution functions, then the function F defined by relation (1) is a joint distribution function with margins F_X, F_Y and F_Z.

Notice that the copula function can be rewritten as $C(u_1, u_2, u_3)$ where $U_1=F_X(X)$, $U_2=F_Y(Y)$ and $U_3=F_Z(Z)$ as well as $u_1=F_X(x)$, $u_2=F_Y(y)$ and $u_3=F_Z(z)$ take values in the [0,1] real subset. We label *ΔCDX*, *ΔSP500* and *ΔVIX* the daily changes in CDX spreads, SP500 returns and VIX level respectively. In particular, daily changes are computed as follows for any time *i* within the sample period under consideration:

$$\Delta CDX_i = X_i = CDX_i - CDX_{i-1} \tag{2}$$

$$\Delta SP500_i = Y_i = SP500_i - SP500_{i-1} \tag{3}$$

$$\Delta VIX_i = Z_i = VIX_i - VIX_{i-1} \tag{4}$$

Under such a setting, we face the well-known curse of dimensionality, which represents the trade-off between the dimension of our setting (i.e. a three-dimension setting), the number of parameters for each considered copula representation, and finally the number of available data points. Given that statistics often advocate parsimonious models, we'll focus on a specific set of Archimedean and Elliptical copulas (see Table 5). In particular, the Frank and Gaussian copulas exhibit no tail dependence, namely no link between the variables' extreme values. However, the Student T copula exhibits a symmetric left- and right-tail dependence. Differently, the remaining copulas exhibit asymmetric tail dependences. In particular, the Clayton copula exhibits lower tail dependence whereas the Gumbel copula exhibits upper tail dependence. Hence, we are able to capture various tail behaviors within financial markets.

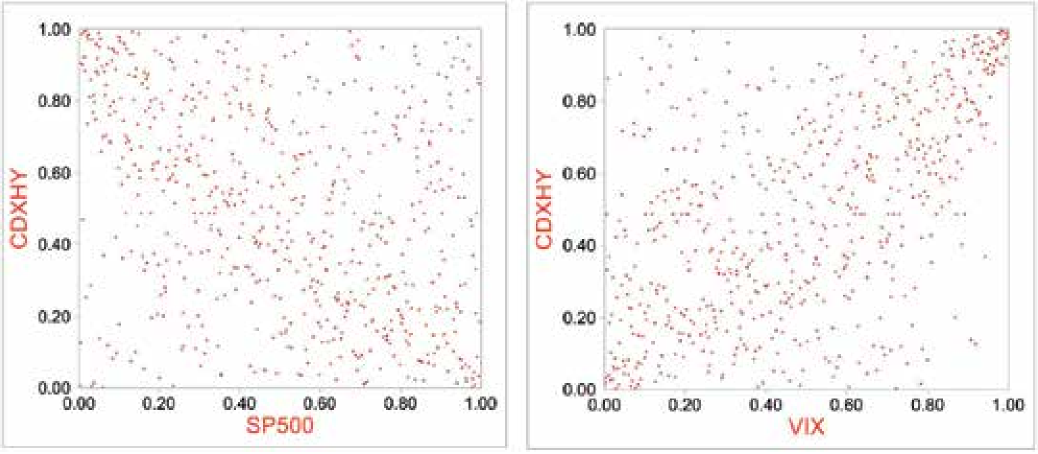


Fig. 4. Empirical bivariate copula functions of CDXHY spread changes

Copula	Attribute	Parameters
Clayton Frank Gumbel	Archimedean	Correlation θ
Gaussian	Elliptical	Correlation matrix
Student T		Degree of freedom v, Correlation matrix

Table 5. Three-dimension copulas and characteristics

Each of the three dimensions of our multivariate copula framework relates respectively to CDX spread changes (i.e. first dimension, or equivalently first variable), SP500 return changes (i.e. second dimension), and finally VIX changes (i.e. third dimension) from one day to another. This way, the relationship between CDX spreads and the two market dimensions are simultaneously accounted for. As an example, Fig. 4 exhibits the corresponding U_1 values for CDXHY spread changes on the vertical axis whereas it displays the corresponding U_2 and U_3 values of SP500 return and VIX level changes on the horizontal axis respectively.

For any positive correlation parameter θ and u_1, u_2, u_3 in [0,1], the *Clayton* copula writes:

$$C\left(u_1,u_2,u_3;\theta\right)=\frac{1}{\left(u_1^{-\theta}+u_2^{-\theta}+u_3^{-\theta}-2\right)^{1/\theta}} \tag{5}$$

For any correlation matrix ρ and u_1, u_2, u_3 in [0,1], the *Gaussian* copula writes:

$$C\left(u_1,u_2,u_3;\rho\right)=\frac{1}{|\rho|^{1/2}}\exp\left\{-\frac{1}{2}\xi^t\left(\rho^{-1}-\mathbf{1}\right)\xi\right\} \tag{6}$$

where ρ and ρ^{-1} are a three-dimension matrix and its inverse respectively, $|\rho|$ is the determinant of the correlation matrix, ξ is the vector of the inverse standard univariate Gaussian cumulative distribution function applied to each element u_1, u_2, u_3, and finally ξ^t is the transposed vector of ξ. We also introduce a three-dimension vector **1** of unit numbers.

The Archimedean and elliptical copulas under consideration satisfy specific assumptions as follows. For any positive correlation parameter θ and u_1, u_2, u_3 in [0,1], the *Frank* copula writes:

$$C\left(u_1,u_2,u_3;\theta\right)=-\frac{1}{\theta}\ln\left\{1-\frac{\left(e^{-\theta u_1}-1\right)\left(e^{-\theta u_2}-1\right)\left(e^{-\theta u_3}-1\right)}{\left(e^{-\theta}-1\right)^2}\right\} \tag{7}$$

For any positive correlation parameter θ and u_1, u_2, u_3 in [0,1], the *Gumbel* copula writes:

$$C\left(u_1,u_2,u_3;\theta\right)=\exp\left(-\left\{\left(-\ln u_1\right)^{\theta}+\left(-\ln u_2\right)^{\theta}+\left(-\ln u_3\right)^{\theta}\right\}^{1/\theta}\right) \tag{8}$$

For any correlation matrix ρ, degree of freedom ν and u_1, u_2, u_3 in [0,1], the *Student T* copula writes:

$$C(u_1,u_2,u_3;\rho,\nu)=\frac{1}{|\rho|^{1/2}}\frac{\Gamma\left(\frac{\nu+3}{2}\right)\left\{\Gamma\left(\frac{\nu}{2}\right)\right\}^3\left(1+\frac{1}{\nu}\xi^t\rho^{-1}\xi\right)^{-\frac{\nu+3}{2}}}{\left\{\Gamma\left(\frac{\nu+3}{2}\right)\right\}^3\Gamma\left(\frac{\nu}{2}\right)\prod_{n=1}^{3}\left(1+\frac{\xi_n^2}{\nu}\right)^{-\frac{\nu+1}{2}}} \tag{9}$$

where ρ and ρ^{-1} are a three-dimension matrix and its inverse respectively, $|\rho|$ is the determinant of the correlation matrix, Γ is the Gamma function, ξ is the vector (ξ_1, ξ_2, ξ_3) of the inverse univariate Student[3] cumulative distribution function applied to each element u_1, u_2, u_3, and finally ξ^t is the transposed vector of ξ. Charpentier *et al.* (2006) advocate a minimum number of two hundreds observations when the copula dimension is below or equal to three. According to those authors the length of our time series, or equivalently, the chosen sample size reduces thus the curse of dimensionality problem.

3.2 Estimation and selection

According to (Weiss 2011), the maximum likelihood method yields less biased and more stable parameter estimates for parametric copulas. Hence, we first estimate the copula parameters while running a Maximum Likelihood Estimation methodology (MLE). Specifically, we apply a semi-parametric MLE process, which is also known as canonical MLE. The semi-parametric MLE procedure relies on two stages so that the copula is specified while the univariate margins of the considered variables are not specified. The first stage computes the nonparametric cumulative distribution function (i.e. univariate margin) of each variable whereas the second stage maximizes the log-likelihood of the copula under consideration as a function of the corresponding copula parameters (Choros *et al.*, 2010; Genest *et al.*, 1995; Yan, 2006). However, we correct for possible parameters' uncertainty while applying a parametric bootstrapping technique in order to conform to the related MLE asymptotics (i.e. bootstrap MLE; Chen and Fan, 2006; Chernick, 1999; Davison and Hinkley, 2006; Efron, 1979; Simon, 1997; Varian, 2005).[4] The parametric bootstrap, which is also a resampling method, allows for assigning an accuracy measure to parameter estimates. Indeed, parameter uncertainty usually yields the under- or overestimation of model parameters when samples are not large enough. Correcting for uncertainty and sticking to MLE assumptions allow therefore for getting more accurate estimates and thus sounder risk assessment. Then, our selection process of the most appropriate copula representation relies on the information criterion principle (i.e. selection tool). In particular, we consider the Akaike, Schwarz and Hannan-Quinn information criteria. Those information criteria encompass two components, which are the forecast error committed by the model and number of estimated unconstrained parameters (Akaike, 1974; Lütkepohl, 2006; Hannan and Quinn, 1979; Schwarz, 1978). The model selection rule requires minimizing the information criterion. By doing so, the selection process targets an accurate and parsimonious model, which reduces the potential errors and estimation problems.

[3] This is a Student distribution with ν degree(s) of freedom.

[4] Recall that asymptotic principles rely on the infinite sample assumption.

The negative Kendall correlation between CDX spreads and SP500 return changes is incompatible with the Clayton copula representation. Moreover, the obtained parameter estimates for Frank copula are also incompatible with the corresponding theoretical specification. As a result, we display only the chosen information criteria for the remaining copulas, which consist of the Gumbel, Gaussian and Clayton copulas (see Tables 6, 7 and 8). Amongst the range of representations under consideration, the best copula or the optimal three-dimension copula estimation is that one which minimizes at least one (when not all) of the information criteria previously mentioned, namely Akaike, Schwarz and Hannan-Quinn information criteria. According to Tables 6 to 8, the optimal copula representation consists of the Student T copula for all CDX spreads under consideration, which implies a symmetric tail dependence of CDX spreads with respect to market risk channels.

Spread	Information criterion		
	Akaike	Schwarz	Hannan-Quinn
CDXEM	2.00747665	6.28599811	3.67664928
CDXED	2.00650408	6.42486904	3.72035251
CDXHY	2.00650411	6.42486907	3.72035253
CDXHB	2.00650411	6.42486907	3.72035254
CDXBB	2.00650407	6.42486903	3.72035250
CDXIG	2.00650409	6.42486905	3.72035252
CDXIV	2.00650410	6.42486906	3.72035253
CDXXO	2.00650415	6.42486911	3.72035258

Table 6. Information criterisa for Gumbel copula estimation

Spread	Information criterion		
	Akaike	Schwarz	Hannan-Quinn
CDXEM	-496.82	-484.01	-491.84
CDXED	-409.29	-396.05	-404.16
CDXHY	-592.34	-579.11	-587.22
CDXHB	-472.05	-458.82	-466.93
CDXBB	-444.55	-431.31	-439.43
CDXIG	-616.53	-603.30	-611.41
CDXIV	-517.10	-503.86	-511.98
CDXXO	-429.38	-416.15	-424.26

Table 7. Information criteria for the Gaussian copula estimation

Spread	Information criterion		
	Akaike	Schwarz	Hannan-Quinn
CDXEM	-709.90	-692.83	-703.27
CDXED	-505.90	-488.27	-505.90
CDXHY	-704.44	-686.81	-704.44
CDXHB	-578.60	-560.97	-578.60
CDXBB	-517.75	-500.11	-517.75
CDXIG	-735.21	-717.57	-735.21
CDXIV	-604.00	-586.37	-604.00
CDXXO	-486.81	-469.17	-486.81

Table 8. Information criteria for the Student T copula estimation

Further, Table 9 displays the corresponding Student T parameter estimates, namely the elements of the correlation matrix ρ and the corresponding number ν of degrees of freedom.

Spread	Correlation with		Correlation between	Degree of
	SP500	VIX	SP500 and VIX	freedom
CDXEM	-0.1185	0.3448	-0.6216	3
CDXED	0.0116	0.2182	-0.6072	4
CDXHY	-0.2580	0.5142	-0.6223	4
CDXHB	-0.1185	0.3448	-0.6216	3
CDXBB	-0.0589	0.2280	-0.5775	4
CDXIG	-0.3466	0.5648	-0.6061	3
CDXIV	-0.2893	0.4166	-0.5967	3
CDXXO	-0.1655	0.2379	-0.6068	5

Table 9. Parameter estimates of the three-dimension Student T copula

Apart from CDXED spreads, results conform to empirical facts so that:

- the correlation between the changes in CDX spreads and SP500 returns is negative,
- the correlation between the changes in CDX spreads and VIX levels is positive,
- the correlation between the changes in SP500 returns and VIX levels is negative.

The positive correlation between the changes in CDXED spreads and SP500 returns may result from the curse of dimensionality concern as well other data- and market-specific features of the emerging market diversified CDX index. However, we have only around 100 data points less as compared to the other CDX spread time series. Moreover, emerging corporate and sovereign credit markets require a specific attention and study (e.g. pricing issues, default events and correlations, quotation disruptions),[5] which is beyond the scope of the current research. Finally, the obtained correlation matrix elements are slightly different from the previous Kendall correlation estimates. Indeed, the average differences between the copula-based correlation and the Kendall counterparts are 0.0267 and 0.0237 with respect to SP500 returns and VIX level (i.e. average correlation differences between CDX spreads and referenced stock market benchmarks). In the same way, the average absolute differences between those two types of correlation estimates are 0.0648 and 0.0665 with respect to SP500 returns and VIX level.

Finally, the obtained results seem to conform to the theoretical behavior of the Student copula representation. As a rough guide, Fig. 5 plots the theoretical copula representation based on the simulation of pseudo-random numbers while using the estimated CDXHY copula parameters.[6] Strikingly, the similarity between the theoretical and empirical copula representations is obvious and noticeable.

[5] On the 18th July 2011, the CDXED spread series is still discontinued (see the Credit Index Rules published as of April 15, 2011 at http://www.markit.com/en/products/data/indices /credit-and-loan-indices/cdx/ cdx.page#).

[6] We simulated 1000 points of the theoretical dependence structure.

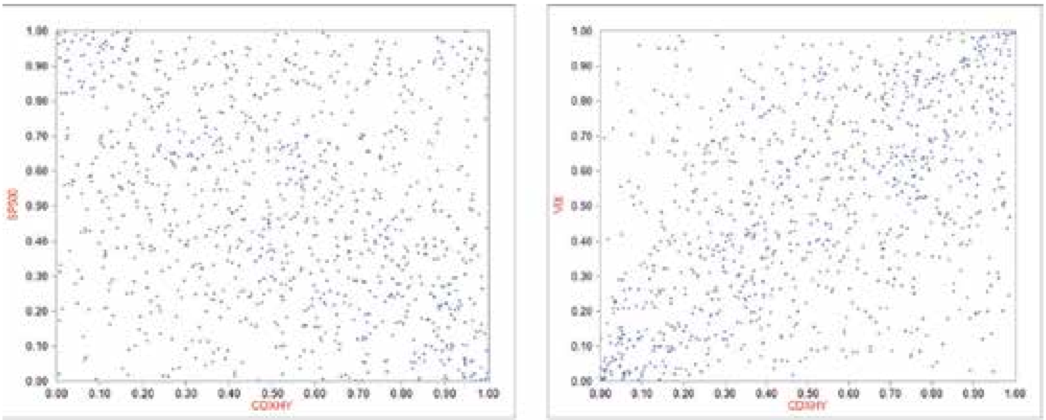


Fig. 5. Theoretical bivariate copula functions of CDXHY spread changes

4. Scenario analysis

We set up a stress testing methodology while considering two specific types of scenarios, namely a good scenario and a bad scenario.

4.1 Scenarios and conditional probabilities

We consider two extreme scenarios among which an extremely bad situation and a very good situation for the credit market. The bad situation refers to a deterioration of the credit market through the widening of CDS spreads while the good situation refers to an improvement of the credit market through a tightening of credit spreads.

The stress testing analysis is driven by the relationships between CDX spreads on one side, and both SP500 returns and VIX level on the other side. Remember that CDX spreads widen when either SP500 returns decrease (i.e. negative correlation) or VIX level increases (i.e. positive correlation). Conversely, CDX spreads diminish when either SP500 returns augment or VIX level declines. As a result a three-dimensional bad scenario is such that CDX spreads expand when both SP500 returns drop and VIX level rises. Conversely, a good scenario is such that CDX spreads tighten when both SP500 returns enhance and VIX level falls. Therefore, we perform a stress testing study (i.e., scenario analysis) based on the *Student T* optimal copula C^*, and compute the probability of occurrence of the bad situations (i.e. scenario 1) and the good situations (i.e. scenario 2). In particular, we assess the probability that CDX spreads increase (*decrease*) due to both a decrease (an *increase*) in SP500 returns and an increase (a *decrease*) in VIX level. Hence, we propose a three-dimensional tail-dependence analysis. With regard to the first scenario (scenario 1), the probability of occurrence of bad states is computed as follows:

$$\mathrm{Prob}\left(U_1 > 1-\mathrm{u}_\alpha \middle| U_2 \le \mathrm{u}_\alpha, U_3 > 1-\mathrm{u}_\alpha\right) = 1 + \frac{C^*\left(1-\mathrm{u}_\alpha, \mathrm{u}_\alpha, 1-\mathrm{u}_\alpha;\theta\right) - C^*\left(1-\mathrm{u}_\alpha, \mathrm{u}_\alpha, 1;\theta\right)}{\mathrm{u}_\alpha - C^*\left(1, \mathrm{u}_\alpha, 1-\mathrm{u}_\alpha;\theta\right)} = \alpha \qquad (10)$$

where α is the critical risk level under consideration (e.g., 5%, 10%), and u_α is the related quantile level. Recall also that we state U_1=$F_X(X)$, U_2=$F_Y(Y)$ and U_3=$F_Z(Z)$ with X=ΔCDX, *Y=ΔSP500* and *Z=ΔVIX*. Specifically, the probability of facing scenario 1 is a quantile-quantile dependence measure. Indeed, it measures to which extent the occurrence of both an extreme SP500 return decrease and an extreme increase in VIX level impacts the occurrence of an extreme CDX spread increase. Therefore, the higher the probability level α, the more correlated the 'bad' extremes are, or equivalently, the stronger the tail dependence is (Coles *et al.*, 1999). Moreover, a decreasing u_α value corresponds to an increasing value of (1- u_α), which highlights a worsening of the credit market through larger CDX spread increases. In practice, a very bad scenario is represented by a wide and positive CDX spread variation coupled with both a huge decrease in SP500 return (i.e., return variation with high magnitude and negative value) and a large positive variation in VIX level. Symmetrically, the probability of occurrence of good states as represented by scenario 2 is computed as follows:

$$\text{Prob}\left(U_1 \le \text{u}_\alpha \middle| U_2 > 1-\text{u}_\alpha, U_3 \le \text{u}_\alpha\right) = \frac{C^*\left(\text{u}_\alpha, 1, \text{u}_\alpha; \theta\right) - C^*\left(\text{u}_\alpha, 1-\text{u}_\alpha, \text{u}_\alpha; \theta\right)}{\text{u}_\alpha - C^*\left(1, 1-\text{u}_\alpha, \text{u}_\alpha; \theta\right)} = \alpha \qquad (11)$$

Therefore, the higher the probability level α, the more correlated the 'good' extremes are. On a practical viewpoint, a very good scenario is represented by a wide and negative CDX spread variation coupled with both a huge increase in SP500 return (i.e., return variation with high magnitude and positive value) and a large negative variation in VIX level (e.g., lower volatility regime). Indeed, the credit risk level is as low as u_α is small.

4.2 Estimation

Whatever the scenario under consideration, we fix a value for the critical risk level *a* and then solve for u_α in the conditional probabilities displayed in the previous section. For this purpose, we estimate the corresponding quantile u_α while solving numerically for a nonlinear optimization problem. We request a tolerance level of 10^{-5} for the gradient of the estimated coefficients. Basically, we consider successively two distinct stress levels, namely a 10 and 5 percent critical risk levels. Each obtained quantile u_α corresponds to a specific joint variation of CDX spreads, SP500 returns and VIX level.

Stating α = 5% and 10%, we consider the worst case (*scenario 1*), and the very good situation (*scenario 2*). We report the corresponding values for the dependence structure between CDX spreads and both SP500 and VIX in Tables 10 and 11. As a result, we are able to characterize critical thresholds for the variations in CDX spreads, SP500 returns and VIX level, which correspond to the relevant stress scenario. Notice that a positive value indicates an increase in the market data under consideration whereas a negative value illustrates a decrease from one day to another. For example, there is a 5 percent probability level that CDXEM spreads increase by 12.9600 basis points given that SP500 returns decrease by 306.0669 basis points and VIX index grows by 2.6000 between February 1st 2008 and March 24th 2010. Symmetrically, there is a 5 percent probability level that CXEM spreads drop by 24.4100 basis points given that SP500 returns increase by 402.1336 basis points and VIX index declines by 2.9300 over the sample time. Moreover, noticeable differences between the 5 and 10 percent risk thresholds can be listed.

With respect to scenario 1, CDX spreads and VIX level exhibit a larger increase under the 5 percent threshold whereas SP500 returns display a bigger decrease as compared to the 10 percent threshold. With respect to scenario 2, CDX spreads and VIX level exhibit a greater decline under the 5 percent threshold whereas SP500 returns display a higher increase as compared to the 10 percent threshold. Hence, results confirm the strength of the 5 percent probability setting as compared to the 10 percent probability threshold (i.e., tougher credit risk situation under the 5 percent probability level).

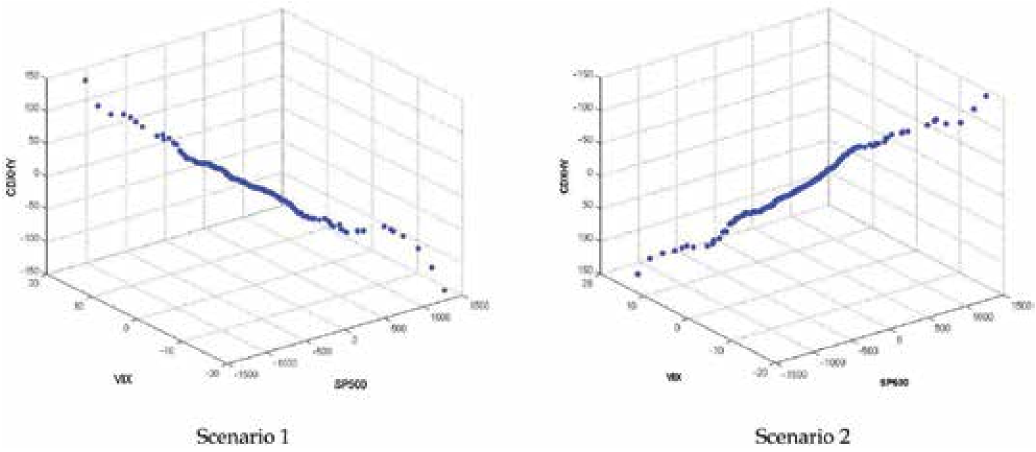
	Scenario 1			Scenario 2		
Index	**ΔCDX > x_{α}**	**ΔSP500 < y_{α}**	**ΔVIX > z_{α}**	**ΔCDX < x_{α}**	**ΔSP500 > y_{α}**	**ΔVIX< z_{α}**
CDXEM	12.9600	-306.0669	2.6000	-24.4100	402.1336	-2.9300
CDXED	17.2100	-309.5813	2.6100	-19.2000	300.4147	-2.1700
CDXHY	18.8900	-230.5572	1.7400	-15.7100	193.1290	-1.2800
CDXHB	24.1600	-298.1570	2.3200	-17.8400	241.8523	-1.6300
CDXBB	19.2399	-298.5321	2.3400	-14.0300	238.2396	-1.5400
CDXIG	6.0600	-306.0669	2.6000	-5.0100	256.4785	-1.8500
CDXIV	11.6300	-306.0669	2.6000	-7.7500	243.1394	-1.6400
CDXXO	12.0000	-309.5813	2.6100	-8.4100	240.9530	-1.6100

Table 10. CDX spread and market risk changes under a α=5% risk level

	Scenario 1			Scenario 2		
Index	**ΔCDX > x_{α}**	**ΔSP500 < y_{α}**	**ΔVIX > z_{α}**	**ΔCDX < x_{α}**	**ΔSP500 > y_{α}**	**ΔVIX< z_{α}**
CDXEM	8.5500	-230.9755	1.7400	-11.0100	256.4785	-1.8500
CDXED	8.7300	-230.5573	1.7400	-7.8100	205.3360	-1.3200
CDXHY	18.2300	-228.2211	1.7300	-10.9200	149.1986	-1.0400
CDXHB	16.7200	-225.9790	1.6400	-15.4500	216.3420	-1.4400
CDXBB	15.2400	-233.2346	1.7800	-13.6700	230.4725	-1.5300
CDXIG	4.5900	-229.2211	1.7300	-4.7100	240.9530	-1.6100
CDXIV	9.6900	-253.6435	1.9200	-7.6400	241.8523	-1.6900
CDXXO	10.0000	-230.9755	1.7500	-8.3400	238.5514	-1.5700

Table 11. CDX spread and market risk changes under a α=10% risk level

As a rough guide, we also plot the corresponding empirical CDX spread and market risk changes as a function of quantile u_{α} for CDXHY and CDXED indexes. Fig. 6 reports the scenarios 1 and 2 for CDXHY whereas Fig. 7 displays only scenario 1 for CDXED. In Fig. 6 under scenario 1, the top left corner corresponds to the highest CDX spread increase (i.e., lowest possible value of u_{α}), which matches both the biggest decrease in SP500 return and the largest increase in VIX level. Conversely, the bottom right corner of scenario 1 panel represents the largest CDX spread decrease (i.e., highest possible value of u_{α}), which matches both the biggest increase in SP500 return and the largest decrease in VIX level. Naturally, scenario 2 exhibits a symmetric behavior since it represents the mirror evolution of scenario 1. Strikingly, Fig. 7 displays some tail discontinuity with respect to the upper left corner as opposed to the bottom right corner. Such an outlier may generate estimation problems. Hence, such a pattern requires further investigation, which is left for future research.



Scenario 1

Scenario 2

Fig. 6. CDXHY spread and market risk changes for various α levels under scenarios 1 and 2

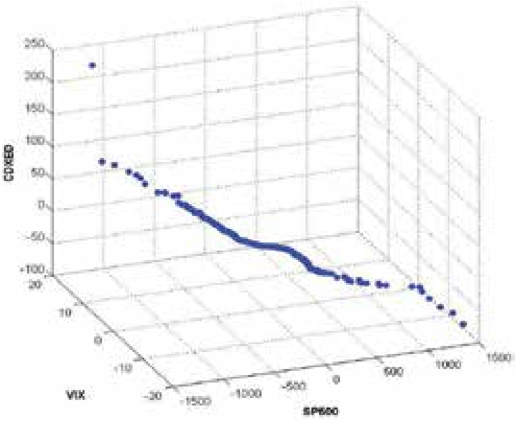


Fig. 7. CDXED spread and market risk changes for various α levels under scenario 1

5. Concluding remarks and future work

In this paper, we focused on the dependence structures between CDX spread changes on the one hand, and changes in both SP500 returns and VIX index, on the other hand. We empirically exhibited the asymmetric nature of each bivariate dependence structure, namely the dependence structures between CDX spreads and SP500 returns and between the CDX spreads and VIX index. We also emphasized the differences between those two types of bivariate dependence structures, which we handled simultaneously within a three-dimension copula analysis. Balancing the curse of dimensionality with a parsimonious modeling framework, we selected three Archimedean copulas and two classic elliptical copulas in order to test for various tail dependencies.

The estimation process and the selective information criterion statistics exhibited the Student T dependence structure as an optimal three-dimension copula representation. Therefore, we have to cope with symmetric tail dependencies between CDX spreads and the two stock market risk channels above-mentioned. Additionally, we are able to realize a more accurate and global credit risk scenario analysis in the light of both the stock market trend and stock market volatility levels. As an example, we quantified a 10 and 5 percent tail dependence levels in order to illustrate a stress testing analysis under two types of stock market stress. Such a scenario analysis helps identify thresholds for the variations in both stock market price and volatility, which impact variations in CDX spreads. In other words, we are able to characterize the impact of the stock market risk channels on the evolution of CDS spreads. Consequently, the three-dimension copula framework is a useful tool for risk monitoring/management and risk reporting prospects under Basel 3. In particular, the scenario analysis is of high significance for portfolio insurance when credit lines are involved in the investment portfolio under consideration. Indeed, such a framework is useful for value-at-risk or even stressed value-at-risk implementations as well as related scenario analysis (Embrechts and Höing, 2006). Further research should therefore focus on the dynamic implementation of the three-dimensional copula framework in a forecasting perspective.

6. Acknowledgement

We would like to thank participants at the 8th AFE conference (30th June-2nd July 2011, Samos, Greece) and the Quantitative Methods in Finance Conference (14th-17th December 2011, Sydney, Australia) for their helpful questions.

7. References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716–723.

Basel Committee on Banking Supervision (2010). Sound practices for Backtesting counterparty credit risk models, Bank For International Settlements Document, December 2010.

Bellman, R. E. (1961). *Adaptive Control Processes*. Princeton University Press, Princeton, NJ.

Black, F. (1976). *Studies of stock price volatility changes*. Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section, pp. 177–181.

Blundell-Wignall, A., & Atkinson, P. (2005). Thinking beyond Basel III: Necessary solutions for capital and liquidity. *OECD Journal: Financial Market Trends*, Vol. 2010, No. 1, pp. 1–23.

Brigo, D., Predescu, M., & Capponi, A. (2011). Liquidity modeling for credit default swaps: An overview. In: *Credit Risk Frontiers. The Subprime Crisis, Pricing and Hedging, CVA, MBS, Ratings and Liquidity*, Bielecki T., Brigo D. and Patras F., Chapter 19, John Wiley & Sons, Bloomberg Financial Series.

Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, Vol. 22, No. 6, pp. 2201-2238.

Brunnermeier, M. K. (2009). Deciphering the 2007-08 liquidity and credit crunch. *Journal of Economic Perspectives*, Vol. 23, No. 1, pp. 77-100.

Charpentier, A., Fermanian, J. D., & Scaillet, O. (2006). The estimation of copulas: Theory and practice. In: *Copulas-From Theory to Application in Finance*, J. Rank, Chapter 2, pp. 35-61, Risk Books.

Chen, X., & Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, Vol. 130, No. 2, pp. 307–335.

Chernick, M. R. (1999). *Bootstrap Methods, A practitioner's guide*. Wiley Series in Probability and Statistics.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula Methods in Finance*. Chichester: Wiley.

Choroś, B., Ibragimov, R., & Permiakova, E. (2010). Copula Estimation. In: *Copula Theory and Its Applications*, Lecture Notes in Statistics, Volume 198, Part 1, P. Jaworski, F. Durante, W. K. Härdle and T. Rychlik, Chapter 3, pp. 77-91, Springer Germany (Berlin Heidelbeg).

Coles, S., Currie, J., & Tawn, J. (1999). Dependence measures for extreme value analyses, Department of Mathematics and Statistics, Lancaster University, Working Paper.

Das, S. R., & Hanouna, P. (2009). Hedging credit: Equity liquidity matters. *Journal of Financial Intermediation*, Vol. 18, No. 1, pp. 112-123.

Davison, A. C., & Hinkley, D. (2006). *Bootstrap Methods and their Application* (8th edition), Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance. *Bulletin de l'Académie Royale de Belgique, Classe des Sciences*, Vol. 65, No. 6, pp. 274-292.

Deheuvels, P. (1980). Non-parametric tests of independence. In: *Statistique Non Paramétrique Asymptotique*, Lecture Notes in Mathematics 821, J. P. Raoult, pp. 95-107, Springer Verlag, Berlin.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, Vol. 74, No. 366, pp. 427–431.

Dupuis, D., Jacquier, E., Papageorgiou, N., & Rémillard, B. (2009). Empirical evidence on the dependence of credit default swaps and equity prices. *Journal of Futures Market*, Vol. 29, No. 8, pp. 695-712.

Durrleman, V., Nikeghbali, A., & Roncalli, T. (2000). *Which copula is the right one?* Technical Report, Operational Research Group of Crédit Lyonnais, Paris.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Annals of* Statistics, Vol. 7, No. 1, pp. 1-26.

Embrechts, P., McNeil, A.J., & Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In: *Risk Management: Value at Risk and Beyond*, M.A.H. Dempster, Cambridge University Press, Cambridge.

Embrechts, P., Lindskog, F., & McNeil, A.J. (2003). Modeling dependence with copulas and applications to risk management. In: *Handbook of Heavy Tailed Distributions in Finance*, S. Rachev, pp. 329–384, Elsevier, North-Holland.

Embrechts, P., & Höing, A. (2006). Extreme VaR scenarios in higher dimensions. *EXTREMES*, Vol. 9, No. 3, pp. 177-192.

Ferrari, S., Van Roy, P., & Vespro, C. (2011). Stress testing credit risk: Modelling issues. *Financial Stability Review*, Vol. 9, No. 1, pp. 105-120.

Fisher, L. (1959). Determinants of risk premiums on corporate bonds. *Journal of Political Economy*, Vol. 67, No. 3, pp. 217–237.

Fuller, W. A. (1996). *Introduction to Statistical Time Series* (2nd edition), Wiley, New York.

Gatfaoui, H. (2010). Investigating the dependence structure between credit default swap spreads and the U.S. financial market. *Annals of Finance*, Vol. 6, No. 4, pp. 511-535.

Genest, C., Ghoudi, K., & Rivets, L.-P. (1995). A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, Vol. 82, No. 3, pp. 543–552.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. Journal of the Royal Statistical Society, Vol. B 41, pp. 190–195.

Hartmann, P. (2010). Interaction of market and credit risk. *Journal of Banking and Finance*, Vol. 34, No. 4, pp. 697-702.

Hull, J. C., & White, A. D. (2006). Valuing credit derivatives using an implied copula approach. *Journal of Derivatives*, Vol. 14, No. 2, pp. 8–28.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability,Vol. 73, Chapmann & Hall, London.

Liebscher, E. (2005) Semiparametric density estimators using copulas. *Communications in Statistics - Theory and Methods*, Vol. 34, No. 1, pp. 59–71.

Liu, J., Longstaff, F. A., & Mandell, R. E. (2006). The market price risk of interest rate swaps: The roles of default and liquidity risks. *Journal of Business*, Vol. 79, No. 5, pp. 2337-2359.

Longstaff, F. A., Mithal, S., & Neis, E. (2005). Corporate yield spreads: default risk or liquidity? New evidence from the credit default swap market. *Journal of Finance*, Vol. 60, No. 5, pp. 2213-2253.

Lumsdaine, R. L. (2011). Correlations, models, and risk management in challenging times. *Journal of Financial Econometrics*, Vol. 7, No. 1, pp. 40-51.

Lütkepohl, H. (2006). Vector autoregressive models. In: *Palgrave Handbook of Econometrics*, Volume 1: *Econometric Theory*, T. C. Mills and K. Patterson, pp. 477–510, Palgrave Macmillan, Houndmills.

MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business and Economic Statistics*, Vol. 12, No. 2, pp. 167–176.

McNeil, A., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton.

Nelsen, R.B. (1999 *). An Introduction to Copulas*. Lectures Notes in Statistics, Vol. 139, Springer, New York.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, Vol. 55, No. 3, pp. 703–708.

Norden, L., & Weber, M. (2009). The co-movement of credit default swap, bond and stock markets: An empirical analysis. *European Financial Management*, Vol. 15, No. 3, pp. 529-562.

Patton, A. W. (2009). Copula–based models for financial time series. In: *Handbook of Financial Time Series*, Part 5, T. Mikosch, J.-P. Kreiß, R. A. Davis and T. G. Andersen, pp. 767-785, Springer, Berlin Heidelberg.

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, Vol. 75, No. 2, pp. 335–346.

Predescu, M., Thanawalla, R., Gupton, G., Liu, W., Kocagil, A., & Reyngold, A. (2009). Measuring CDS liquidity, Fitch Solutions presentation at the Bowles Symposium, Georgia State University.

Scheicher, M. (2009). The correlation of a firm's credit spread with its stock price: Evidence form credit default swaps. In: *Stock Market Volatility*, G. N. Gregoriou, Chapter 21, pp. 405-419, Chapman & Hall/CRC Finance, London.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, Vol. 6, No. 2, pp. 461–464.

Simon, J. L. (1997). *Resampling: The New Statistics* (2^{nd} Edition), Resampling Stats.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistiques de l'Université de Paris 8*, pp. 229 – 231.

Sklar, A. (1973). Random variables, joint distribution functions and copulas. *Kybernetika*, Vol. 9, No. 6, pp. 449–460.

Van Den End, J. W. (2011). Liquidity stress-tester: A model for stress-testing banks' liquidity risk. *CESifo Economic Studies*, Vol. 56, No. 1, pp. 38-69.

Varian, H. (2005). Bootstrap tutorial. *Mathematica Journal*, Vol. 9, No. 4, pp. 768-775.

Vaz de Melo Mendes, B., & Kolev, N. (2008). How long memory in volatility affects true dependence structure. *International Review of Financial Analysis*, Vol. 17, No. 5, pp. 1070-1086.

Weiss, G. (2011). Copula parameter estimation by maximum-likelihood and minimum-distance estimators: A simulation study. *Computational Statistics*, Vol. 26, No. 1, pp. 31-54.

Yan, J. (2006). Multivariate modeling with copulas and engineering applications. In: *Springer Handbook of Engineering and Statistics*, Part F, Pham H., Chapter 51, pp. 973-990, Springer, London.

19

Financial Risks: Cases Of Non-Financial Enterprises

Irina Voronova
Faculty of Engineering Economics and Management, Riga Technical University
Latvia

1. Introduction

From the anatomic point of view an integral system of risk management consists of risk measurement and management. This chapter is devoted to assessing the financial risks of small and medium-sized enterprises (SME) in the non-financial sphere as a constituent part of quantitative and qualitative risk measurement techniques.

In the assessment of financial risks of enterprises in the non-financial sphere, and in small and medium-sized enterprises in particular, it is recommended that the principles of Occam and KISS (keep it simple, stupid) be used as guidelines. The application of these principles in relation to the choice of the methods of financial risks assessment means that mainly simple methods - and those known by enterprise specialists - should be used. These sufficiently simple methods are the following: special ratios method and expert examination method.

The author evaluates the development of discriminant and conditional probability methods of financial risk assessment in nine East European countries. The usage of these methods enables small and medium-sized enterprises to assess, predict and manage risks related to liquidity, credit, decreasing financial stability and insolvency/bankruptcy.

In doing this research we applied an approach built on conditional probability models (logit and probit analysis) on source reviews and the author's own experience. Selection and assessment of the described models of insolvency/bankruptcy and provisions of financial risk assessment are based on the author's personal opinion.

The chapter proceeds as follows: after the introduction in Section 2 the financial risk definition is introduced and a three-level classification of financial risk is presented. Next, in Section 3 the development of financial ratios will be described and a review of using classic models for assessing insolvency/bankruptcy, multiple discriminate analysis (MDA) type models and their development in nine countries of Europe are given. Then in Section 4 a discussion will be conducted about the usage of express analysis of financial risks on the basis of the principles of the analysis of enterprise economic turnover balance sheet and quick tests as a simple instrument. Tests to assess the risks to the enterprises at different stages of the life cycle will also be discussed. At the end of chapter 4 a case study of the risk occurrence reasons is provided. It is possible to position these factors on a scale of risk probabilities by employing expert assessment. In Section 5 a critical evaluation of the methods of measuring the financial risks of enterprises in the non-financial sphere are discussed and future ideas are given. At the end of the chapter we offer some concluding ideas. In the appendix, we attach reference materials containing 31 models, developed in nine countries enabling determination of enterprises' and their partners' insolvency risk.

2. Introducing financial risks management for enterprises in the non-financial sphere

Financial risk has both objective and subjective bases. The objective basis of financial risk refers to an apriori uncertainty of the external environment related to an enterprise. The subjective basis of financial risk is based on the fact that risk is realised through the activity of an entrepreneur (an individual) for it is he/she who assesses risky situations, creates a number of outcomes and makes decision. Comprehension of the nature of financial risk and classification makes it possible to build up a system of integral risk management, relying on the existing elements of the organisation of enterprise management. We track the shift of the risk concept as hazard or "something that goes wrong" to the risk concept as uncertainty between entrepreneurs and their aims. These aspects are considered in Section 2.

2.1 The financial risk concept, its classification for enterprises in the non-financial sphere

There are different approaches that can be used when defining risk and financial risks. Holton (2004) described risk as composed of exposure and uncertainty. Financial risk definition can be conventionally divided into two groups. The first approach deals with risk as a hazard of potential losses and is related to the definition of financial risk. It is given in terms of investment risk (assets) and in terms of feasibility of making a structure of liabilities (Chapman, 2006; Kovalev, 2000). The second approach considers financial risk as probability of the occurrence of unfavourable financial consequences under the influence of various factors. It means that financial risk is a complex risk, since it focuses all enterprise risks and is used in monetary terms (Hawkins, 2003). We consider that such a limited comprehension of risk does not allow for a complete, complex technique of its assessment and analysis.

In the document ISO (2009) Risk Management – Principles and guidelines ISO 31000:2009 provides a new approach to risk definition. The definition of risk given by ISO 31000:2009 (taking into account two comments) is "*risk - effect of uncertainty on objectives. Objectives can have different aspects (such as financial, health and safety, and environmental goals) and can apply at different levels (such as strategic, organization-wide, project, product and process) (Note 2). Risk is often expressed in terms of a combination of the consequences of an event (including changes in circumstances) and the associated likelihood of occurrence. (Note 4)*". Taking into account the above, we define financial risk as the influence of uncertainty of occurrence of unfavourable financial consequences in the form of income or/and capital loss. This definition is the most reasonable and acceptable in practical activities.

One of the first risk classifications was suggested by J. M. Keynes (1936). Types of financial risk in the enterprise risk system are given in the works by J. Fraser and B. Simkins (2010), R. Moeller (2007), R.G. Picard (2004) and I.N. Dulova (2011). In the research by D. Luo and B. Sun (2010) a three-level classification of financial risks is introduced, though it may be applied only in the case of enterprise mergers and acquisition.

We consider that it is possible to use the three-level classification of financial risks. The three-level classification of financial risks of non-financial enterprises is given in fig. 1.

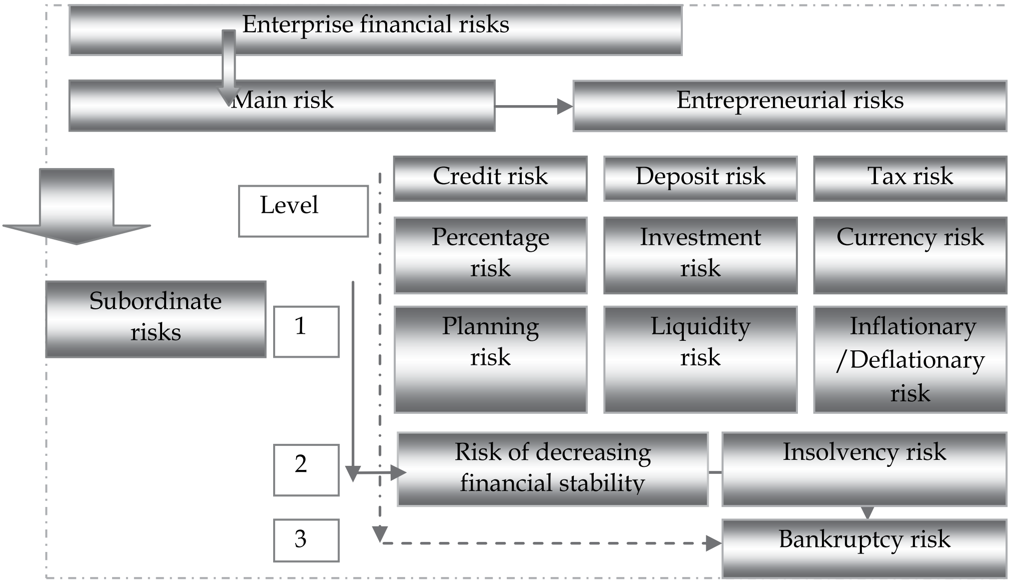


Fig. 1. Three-level classification of financial risks

The classification shown allows us to use a system of financial ratios which in our opinion makes it possible to integrate a system of risk management with a planning system of enterprise.

The first level of financial risk is represented by the concrete risks of an enterprise. The second level of risks refers to the risk of decreasing financial stability and insolvency risk. Selection of these two risks is connected with the fact that quantitative methods exist that enable assessment and prediction of the risk of decreasing financial stability and insolvency. The third level comprises bankruptcy risk which is assessed by financial sector representatives in connection with credit risk monitoring.

The risk of decreasing financial stability is the reflection of the reduction of monetary and merchandise flow balance, incomes and expenditure, and means and sources of their formation.

Solvency risk is an external expression of the financial state of an enterprise. The risk of the first level may trigger the risks of the second level which in their turn may lead to the risk of bankruptcy of an enterprise. In addition, credit risk may directly trigger bankruptcy risk (according to existing legislation).

Classification of financial risks is necessary for applying the most efficient methods, their assessment and management. In our opinion in establishing the system of financial risk management at an enterprise, which is adequate for the external and internal environment, we should rely on the following principles of its creation:

1. Compound approach to management. It is necessary to consider every risk not separately but in combination.
2. As a result of a compound approach it is reasonable to use an indicator, characterising the combined impact of all types of financial risks (indicator of common financial risk).

3. Introduction of a regular and compulsory procedure of identification, analysis and control of different types of financial risk at an enterprise. Activities in financial risk management should not be carried out periodically along with emerging problems but continually, regularly and the procedure of financial risk management should be one of the functions of the system of enterprise management.

2.2 System of risk management

The financial crisis actualised the interest in financial risk management but this interest is mainly connected with financial risk management in financial institutions (banks, insurance and investment companies) for which risk management is obligatory in compliance with existing laws (Basel II and Solvency II). For example, in the EU it is regulated on the basis of Basel II and Solvency II (see Basel Committee on Banking Supervision, 2003; Solvency II framework, 2009).

Much uncertainty in risk management of SMEs has been provoked by Basel II (Hensces, 2010). Risk management in large, small and medium-sized enterprises differs both according to the level of maturity and applied methods. Financial risk management of enterprises in the non-financial sphere should be considered as an integral part of enterprise risk management. The main countries of the European Union, the majority of participants in the non-financial sector of the economy, have small and medium-sized enterprises.

Bibliometric research of nine classes of risk management is applicable to five types of risk used in small and medium-sized enterprises (Brancia, 2011). The research undertaken by Hensces Thomas (2006, 2010a) in the field of practice of risk management in German small and medium-sized enterprises testifies to the fact that risks are strongly focused on the business owners. According to the study the most widespread method of risk management is considered to be business planning, though there is no strong link between planning and management. The same conclusion was made referring to balanced scorecard (BSC) as BSC combines successfully with risk management (Scholey, 2005). The results of the study lead to the conclusion that risk management is carried out in a rather rudimentary way. Similar conclusions can be attributed to the other countries of the EU and Russia (Netsymailo, 2009). In their research the authors K. Dumičić, M. Dumičić and R. Cukrov (2005) analysed protection instruments for different types of financial risk in Croatian companies.

The system of risk management should interact with other elements of the system of the management of entrepreneurial activity. The main elements of creating an integral system of managing strategy, quality and risk at the enterprises in the non-financial sphere are as follows: management by processes, strategic management, total quality management, audit (internal and external), system of organisation a production (BSC, theory of constraints [TOC]) and risk management. Depending on the combination of the main elements we consider four key approaches to the creation of an integrated system of risk management at an enterprise (Fig.2).

For example, the approach based on process is more widely used at medium-sized Latvian enterprises in the non-financial sphere. This is connected to the fact that a great number of Latvian enterprises, driven by a regard for their business competitiveness, have introduced a quality management system.

From the anatomic point of view, a system of risk management has two functions of the system of risk management and risk measurement. Fig. 2 shows the constituents of each of the functions. Let us consider the component of "Risk Measurement".

The existing methods and technologies of risk assessment can be conventionally divided into two large groups appropriate for the assessment of all types of risks and separate types of risks. The first group includes the following methods – BPEST, PESTLE, SWOT analysis, statistical analysis (VaR, Conditional Risk) etc. The second group involves the following methods – threat analysis, fault tree analysis, method of financial ratios etc.

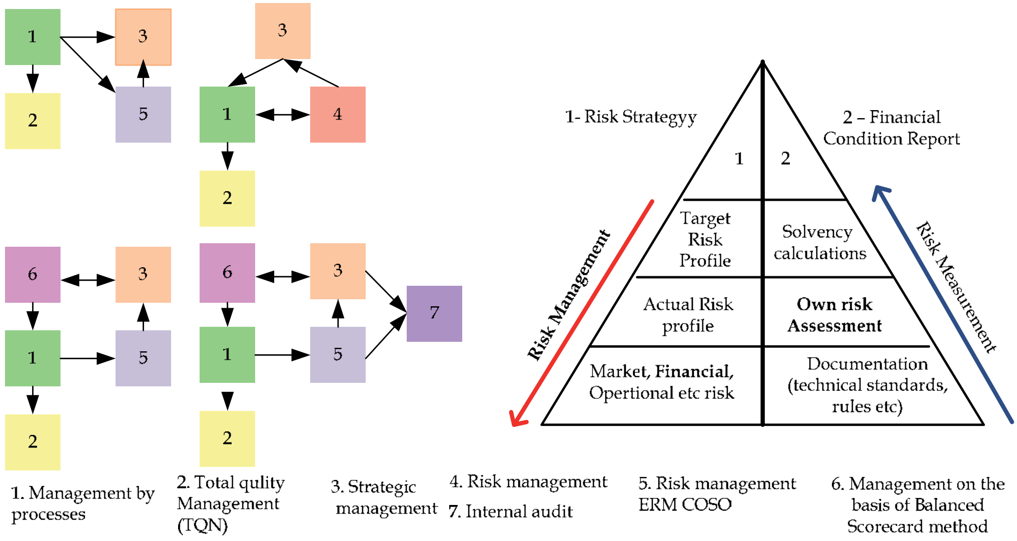

1
3
2
5
3
1
4
2
6
3
1
5
2
6
3
7
1
5
2
1- Risk Strategyy
2 – Financial Condition Report
1
2
Target Risk Profile
Solvency calculations
Actual Risk profile
Own risk Assessment
Market, Financial, Opertional etc risk
Documentation (technical standards, rules etc)
Risk Management
Risk Measurement
1. Management by processes
2. Total qulity Management (TQN)
3. Strategic management
4. Risk management
7. Internal audit
5. Risk management ERM COSO
6. Management on the basis of Balanced Scorecard method

Fig. 2. Elements of the creation of risk management integral system in enterprises in the non-financial sphere

3. Brief review of using a system of financial ratios and models of risks assessment

In this section we discuss the issues of the development of the system of financial rations with a view to the initial point of the development of the techniques of financial risks assessment. We will review the work of Altman (1968) and Ohlson (1980) as well as other articles using discriminant analysis and logistic regression. An evaluation of classical models, such as MDA, in nine countries in Eastern Europe will be at the end of this section.

3.1 Development of financial ratios as an instrument of financial risk assessment

Usage of financial ratios to analyse the activities of an enterprise started a century ago. The history of creating a ratio analysis can be considered as the history of the development of methods of financial risks assessment for the participants in the financial and non-financial sectors of the economy. It is possible to single out three periods in the development of the methods of financial risks assessment: initial stage (from 1891 to 1960), establishment (1960-2000) and improvement (from 2000 to present). This periodisation is conventional but may differ from the periodisation of other authors (Belovary, Giacomino & Akers, 2007).

The key trend of the research during the initial stage was to find financial ratios that could give predictions of an enterprise's solvency and bankruptcy in due time. In 1891 a liquidity coefficient was used for the first time. The name of the founder of the liquidity coefficient is not known but at present the liquidity coefficient and its normative values are used to assess liquidity risk in the short-term.

Crises periods in history are closely connected with the history of the development of ratio analysis. For instance, in 1905 there already existed a system of financial ratios (ten ratios, the author – James Cannon), in 1917 a system of normative values for financial ratios was created (the author - William Laugh) and in 1919 the dependence of an enterprise on its branches was studied and established (see Horrigan, 1968; Anjum, 2010).

Normative values of financial ratios in assessing financial risk serve as indicators for quantitative assessment of separate types of risks and are helpful in establishing a map of financial risk. The existence of differences in the value of financial ratios relating to branches supposes the necessity to track and collate indicators with average branch values. The data on average branch indicators are produced by statistical bodies and credit institutions.

R. Smith was one of the first researchers who in 1930 mentioned some financial ratios that testify to the solvency and financial problems of an enterprise. Smith studied different ratios of 29 bankrupt enterprises over ten years. Financial ratios were compared with the previous period indicators and their change trends were analysed compared with financial situations and changes in economic activities results.

Eight ratios were selected from the object of the research which, according to Smith, characterised the probability of bankruptcy. All these ratios were subdivided into two groups. Further R. Smith and A.H. Winakor (1935) (see Horrigan, 1968) carried out much broader research analysing the financial ratios of 183 bankrupt enterprises and the trends of their change. The findings of the study do not differ substantially from Smith's previous research. Many years later J.O. Horrigan (1968) came to the conclusion that Smith's method can be considered as the first attempt to use scientifically grounded methods to predict possible bankrupts and this is an important step in the development of financial analysis. Merwin's study (1942) can be viewed as the main turning point in the development of financial analysis. He reported that when comparing successful firms with failing ones, the failing firms displayed signs of weakness as early as four or five years before failure (Bellovary et al., 2007). Merwin (1942) found three ratios that were significant indicators of business failure - net working capital to total assets, the current ratio and net worth to total debt.

3.2 Classic models of assessment: For and against

Starting from the 1960s and 1970s the models of the assessment of enterprise insolvency risks were created. In 1962 Jackendoff researched the correlation of profitable and unprofitable firms. On the basis of his research Jackendoff came to the conclusion that the two correlations are the following highly profitable firms: ratio of current liquidity and net working capital to general assets. Moreover, profitable firms had lower debt-to-worth ratios than unprofitable firms. During that period models were created, named Z-assessment functions, which make it possible to carry out the assessment of future risks of company bankruptcy. These models were developed by using different methods - multiple discriminate analysis (MDA), logit analysis, the WILCOX method and others. These models are also used to assess creditworthiness and carry out a comparative analysis of different subjects. The first models were developed by using MDA. The founders of these models are W.H. Beaver (1967) and E.I. Altman (1968). From the period of the emergence of bankruptcy prediction models they have been subject to continuous analyses and critiques. These critical remarks boil down to the following:

- models do not take into account the seasonal factor and cyclic trends of the economy;
- models are developed on the basis of selection of an enterprise which is not always representative;
- models were developed on the sample of the statistical data of enterprises, representing a certain sector of the country's economy, but it is supposed to be used for other countries' enterprises.

Many studies (e.g. Bellovary et al., 2007) have shown limitations of practical application and the question is: "Why do we continue to develop new and different models for bankruptcy prediction?" The answer to the question is hidden in the desire to develop models with higher intellectual capacity at the expense of the increase of the number of indicators numbers. An increase of the number of indicators incorporated into the model of enterprise bankruptcy prediction does not mean an increase of its usefulness (Jones, 1987). The research by J. Bellovary et al. (2007) introduces convincing proof that the MDA type models have a relatively high accuracy of bankruptcy prediction (from the lowest - 32 % to 92% accuracy). The authors of bibliometric research (Genrih & Voronova, 2011) also confirm the conclusion about relatively high reliability of MDA type models.

A conducted analysis of the application of classic models of the MDA type demonstrated that Altman's models (1968, 1977, 2004, 2006 and 2010) and H. Fulmer's model are well known in different countries and many researchers carry out the monitoring of these models. Research by J. Mackavičus and A. Rakšteliené (2005) focuses on the application of Altman's models to predict company bankruptcy in Lithuania. Research by A. Stunžiené and V. Boguslauskas (2006) revealed that Altman's method, when applied to 56 Lithuanian joint stock companies, produced considerable errors. Research by R. Šneidere (2009) and Genriha & Voronova (2010) was undertaken according to the MDA type models and the monitoring of these models was carried out in Latvia. According to the data by P. Antonowicz (2007), monitoring of the application of 16 models of foreign authors in Poland demonstrated that models by Altman (1968, 1984) and Altman & Lavallee (1981) are placed within the framework of the best six models.

Table 1 gives the results of the analysis of the application and monitoring of four classic MDA type models in nine countries of East Europe conducted by the author. The findings express the author's personal opinion which is based on the bibliographic research of the sources of information.

During numerous studies (e.g. Belovary, 2007) of the models W.H. Beaver and E.I. Altman determined a number of substantial drawbacks. The most significant of these refer to the existence of the so-called "uncertainty zone" in the areas of decision-making. As a result, specialists in the field of financial management (in the first turn banking sector) completely refused to use the models of bankruptcy risk assessment, based on discriminant analysis, and began to pay more attention to other, more modern econometric tools, mainly the so-called logit models.

A number of researchers, in particular C. Lennox (1999), stated that in practice logit models enable more effective assessments of bankruptcy risks than can be provided theoretically by MDA. Moreover, the usage of logit regression model supposes wide opportunities for

Model	Model type and interpretation of results. Assessment of bankruptcy threat level	Country								
		1	2	3	4	5	6	7	8	9
Altman model (1968) public manufacturing companies	$Z = 1.2 \cdot WC / TA + 1.4 \cdot RE / TA +$ $+3.3 \cdot EBIT / TA + 0.6 \cdot MVE / TL + \cdot S / TA$ If $Z < 1.81$ a firm is not financially healthy and there is a high probability that it will go bankrupt within five years'; $1.81 < Z < 2.99$ - the "grey zone" – the area where companies are free of bankruptcy risk; If $Z > 2.99$ it concerns a completely financially healthy firm.	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{+}$	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{-}$
Altman $Z^{'}$ model (1983) for private companies (manufacturing)	$Z^{'} = 0.717 \cdot WC / TA + 0.847 \cdot RE / TA +$ $+3.107 \cdot EBIT / TA + 0.420 \cdot OC / TL +$ $+0.998 \cdot S / TA$ If $Z^{I} > 1.23$ bankruptcy is not likely; if $1.23 < Z^{\mathrm{I}} < 2.90$ - bankruptcy cannot be predicted (grey area); if $Z^{\mathrm{I}} > 2.9$ bankrupt cy is likely.	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{+}$	$\frac{+}{+}$	$\frac{+}{-}$	$\frac{+}{+}$	$\frac{+}{-}$	$\frac{+}{-}$
Altman $Z^{''}$ model (1993) for private companies	$Z^{''} = 6.56 \cdot WC / TA + 3.26 \cdot RE / TA +$ $+6.72 \cdot EBIT / TA + 1.05 \cdot OC / TL$ If $Z^{''} > 2.6$ bankruptcy is not likely; if 1.1< $Z^{''} < 2.6$ - bankruptcy cannot be predicted (grey area); if $Z^{''} < 2.6$ bankruptcy is likely.	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{+}$	$\frac{+}{+}$	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{-}$	$\frac{+}{-}$
Fulmer's H model	$H = 5.528 \cdot RE / TA + 0.212 \cdot S / TA +$ $+0.073 \cdot EBT / OC + 1.27 \cdot CF / TL -$ $-0.12 \cdot TL / TA + 2.335 \cdot CL / TA + 0.575 \cdot$ $\cdot Ln(TA) + 1.08 \cdot WC / TL + 0.894 \cdot$ $\cdot Ln(EBIT / I) - 6.075$ If $H > 0$ very low chance of bankruptcy; $H = 0$ critical point; If $H < 0$ - very large probability of bankruptcy.	$\frac{-}{-}$	$\frac{-}{-}$	$\frac{-}{-}$	$\frac{-}{-}$	$\frac{+}{+}$	$\frac{-}{-}$	$\frac{-}{-}$	$\frac{+}{-}$	$\frac{-}{-}$

Table 1. Assessment of the application of classic models of MDA type

Symbols adopted for table 1: CL - current liability; EBT - earnings before taxes; $EBIT$ - earnings before interest and taxes; I - interest; MVE – market value of equity; OC - own capital; RE - retained earnings; S – sales; STL - short-term liability; TA –total assets; TL - total liability; WC – working capital. 1 – Belarus; 2 – Estonia; 3 – Czech Republic; 4 - Poland; 5 – Latvia; 6 – Lithuania; 7 - Romania; 8 – Russia; 9 – Ukraine. In the numerator is indicated -if model is used (+) and if it is not used (-). In the denominator is indicated - if this model monitoring is conducted(+) and if it is not conducted (-) (carried out by the author according to the results of the given list of literature).

implementing manifold econometric tests, which make it possible to assess statistical value for both the model as a whole and the separate variables which form it. In addition, unlike

MDA, logit regression enables not only coming to a conclusion relating to the group of potential bankrupts (which testifies to the limits of the interpretation of the results of the accounts while using the models built on the basis of MDA) but also to assess enterprise bankruptcy risk on the quantitative scale.

A critique of Altman's models was undertaken in the works by Shumway (2001) and Chavan and Jarrow (2004) who employed a discrete hazard model or multiperiod dynamic logit model. These authors' hazard models were developed on the basis of the data of joint stock companies which are listed on NYSE, AMEX and NASDAQ stock exchanges. Their models have higher trustworthiness of bankruptcy prediction results than the simpler logit model.

Enterprises are more likely to default if they are less profitable and less liquid. It is banks that initiated the development of default models because they made use of the advantages of possessing a clients' base and developed these models in compliance with the requirements of the institutions that have supervision responsibilities over financial markets. Default modelling helps identify factors that influence the ability of enterprises to pay back borrowed money. Loan default is closely related to enterprise bankruptcy.

There are three features of the logit model: identify enterprise characteristics that determine financial health accounting ratios of the enterprise, identify appropriate weights to combine these factors into a single measure of financial health of the enterprise and this single measure of financial health is then mapped into a probability of default (PD). Publically available logit models are presented in table A1 (see appendix).

3.3 Adapted models: Experience of East European countries

When selecting models which allow carrying out an assessment of any impending crisis at an enterprise leading to insolvency/bankruptcy, it is recommended to use as a guideline already developed models of predicting bankruptcy, taking into account the conditions of specific country. Starting from the 1990s many East European countries saw a boom in adopting or/and creating models of enterprise insolvency/bankruptcy.

In tables A2, A3 and A4 the author introduced the summary of the models, mostly spread over nine countries of East Europe (see appendix). It is necessary to rely on the results of the tests about verification of accuracy/validity of the findings as a result of the usage of these models. These models can be used to assess their own enterprise financial risks, competitors' risks or business partners both in their own country and also foreign countries which eventually ensure the enterprise assessment and risk management. As a rule, all the given models (tables A2, A3 and A4) comprise publically accessible indicators. tables A2, A3 and A4 introduce the most popular models of discriminant type, making it possible to assess the existence/development of insolvency/bankruptcy risk in East European countries.

Model monitoring is employed in many countries, for example, Poland, Latvia un Russia (Antomowicz, 2007, 2011; Sneidere, 2007; Genriha&Voronova, 2010; Alekseeva, 2011). According to the data by Antonowicz (2007) over 34 models of Z score type were developed and monitored in Poland, whose range of accuracy is from 95% to 57%. Regardless of the great practical value of the assessment of enterprise insolvency/bankruptcy risk, the majority of studies conducted recently are of economic character and do not possess a revolutionary feature which was characteristic of the works by Altman and Ohlson.

The process of the development of discriminant models related to separate branches/regions can be used for individual countries. This trend is especially topical with reference to Russia and Ukraine (see table A3). For example, the static and dynamic models of the assessment of chemical and oil chemical enterprises' stability (Russia, Kramin T.V., 2003), bankruptcy risk of medium-sized enterprises of the printing and publishing industry (Russia, Leo Xao Suan, 1999), six-factor model of predicting the risk of losing solvency for the non-ferrous industry (Russia, Vishnjakov Ja., 2000) and two models of SNEs predicting bankruptcy (Lugovskaya, 2010). The treatment of the problem of bankruptcy assessment of Russian small enterprises was expressed in the works of detective A.E. Krioni (2009). In our opinion, this study causes some interest. He suggested using a financial integrity index instead of discriminant models. We argue that by creating an integrity index Krioni made a logical mistake in the denominator of the formula which may be removed by introducing a discounted capital value for the number of years of the enterprise's existence.

Besides the models, adapted to the conditions of Ukraine (shown in table A3), it is worth mentioning the model of the financial state of food and agricultural industry enterprises used by O.A. Smetanjuk (2007) and A.V. Chupis(1999) (. Among the works, highlighting the assessment of bankruptcy in Ukrainian enterprises the model developed by E.M. Andrushaka deserves special attention. This model (Andrushaka, 2004) can be considered as a hybrid model with elements of the taxonometric approach. In compliance with the accounting algorithm, an integral indicator of enterprise bankruptcy (Z) is determined

$$Z = \sqrt{\sum_{i=1}^{n}(1-N_i)^2 sign(1-N_i)} = \sqrt{U} \tag{1}$$

where

N_i - relation of the i indicator financial state to its normative value;
n - a number of indicators (in this model $n = 3$);
N_1 -a ratio of absolute liquidity divided by its normative value which equals 0.2;
N_2 - a ratio of own capital concentration divided by its normative value which equals 0.5;
N_3 - an indicator of own capital profitability, divided by its normative value which equals $0.1 + 1.1 \cdot b^3$ and b is annual inflation rate.

This model allows the calculation of Z on the basis of the value of the deviations of relevant indicators from normative ones. Deviations best side should reduce an indicator Z and at worst side increase it. The relationship described by equation (2) helps determine this signal:

$$sign(x) = \begin{cases} 1, X > 0 \\ 0, X = 0 \\ -1, X < 0 \end{cases} \tag{2}$$

When all indicators N_i are on the normative level and better than it, $Z = 0$, if $\sum_{i=1}^{n}(1-N_i)^2 sign(1-N_i) < 0$ then it is assumed that $U = 0$. The higher value of Z means the higher probability of bankruptcy.

Development of the model's insolvency/bankruptcy assessment was not so widespread in Estonia and Latvia (table A4). As for Lithuania, the number of developed discriminant models has been greater than in Estonia and Latvia, but they developed on a small statistical base and appropriate monitoring of accuracy of the employed models is not executed (table A4).

4. Composite instruments of the assessment of enterprise financial risks

In this section we will review the methods of risk assessment related to composite measure expanding upon the Tamari approach (1966) that are widespread in the Russian and German speaking countries. We also offer for usage the analysis of financial risks on the basis of the principles of the analysis of the balance sheet of enterprise economic turnover and the system of express tests of financial risks assessment for enterprises being at different stages of life cycle. Finally the chapter will provide a case study of risk occurrence reasons.

4.1 Composite measure of risk management

The third chapter is devoted to the analysis of the possibility of using classic and adopted models of financial risk assessment developed on the bases of multivariate discriminant analysis and conditional probability models. However, in our opinion, there are some other instruments of financial risk assessments which can also be more widely employed for these objectives. It is possible among the methods of financial risks assessment related to the composite measure group to single out firstly the risk index model introduced by Tamari (1966) and later extended by Moses and Liao (1987). A well known technique of detecting a crisis situation was developed by L.V. Doncova and N.A. Nikiforova (2009), with the changes introduced by G.V. Savickaja (2009), which can be used for identifying financial risks. The given technique is similar to Duran's technique by using the principles of creation, rather it is based on the score evaluation of six indicators than three indicators. The given models are static in the sense that they reflect the results of the previous period and to a large extent are more valid for a current assessment than for prediction of financial risk.

To assess insolvency risk cash flow is used rather rarely (Bellovary et al., 2007). There are different opinions on whether cash flow information can be used to predict the bankruptcy of a company. Some researchers (e.g. Zavgren, 1983; Watson, 1996) come to the conclusion that cash flow does not have sufficient proof in bankruptcy assessment, while others (e.g. Beaver, 1966; Aziz & Lawson, 1989; Foster & Ward, 1997; Sharma, 2001) prove the validity of using cash flow analysis in bankruptcy prediction. That is why it is worth noting the Quick Test (Peter Kralicek, 1990) being a one-dimensional test and used in German-speaking countries. The assessment of an enterprise's financial state is carried out on the basis of four indicators (stability, liquidity, profit/loss and profitability). Based on results of indicators, points are assigned. The final resulting value is the determined as a simple arithmetic mean of the points obtained for individual indicators. To determine two indicators the Quick Test is used on cash flow that in our opinion can contribute to the extension of an enterprise's possibilities in prediction of financial risk. The application of Kralicek's Quick Test (Kralicek, 1993) is expedient in dynamics to track the development trends.

The author's practical experience demonstrates that by using tests to discover financial risks, it is possible to employ Duran's technique (Voronova & Romanceviča, 2005). The given technique is based on the creation of integral value, applying the summing up of three main indicators characterising enterprise solvency with certain meaningful coefficients and further enterprise shifts from the first to the fifth classes: 1st class – an enterprise with good financial stability reserves, secure debt recovery; 2nd class – an enterprise with a steady level of debt risk not yet considered as problematic; 3rd class – problematic enterprise the financial state of which can be estimated as average. There is weakness of individual indicators; 4th class – an enterprise with an unstable financial state; 5th class – an enterprise with a critical financial state. The enterprise is completely unstable from the financial point of view and is loss-making.

Name of Indicators		Class Limits Appropriate Criteria				
		Class 1	Class 2	Class 3	Class 4	Class 5
Capital profitability	%	>30	29.9-20	19.9-10	9.9-1	<1
	point	50	49.9-35	34.9- 20	19.9-5	0
Liquidity ratio	value	>2.0	1.99-1.7	1.69-1.4	1.39-1.1	<1
	point	30	29.9-20	19.9-10	9.9-0	0
Financial independence ratio	value	>0.7	0.69-0.45	0.44-0.3	0.29-0.2	<0.2
	point	20	19.9-10	9.9-5	5-1	0
Class limits	point	100	99,9-65	64-35	34-0	0
Characteristic of an enterprise according to the risk level		below the accep-table risk	acceptable risk	high level risk	level risk of bankruptcy	actually insolvent

Table 2. Enterprise classification classes according to the level of solvency

Table 2 reflects the necessary indicators, scores and class limits of the appropriate criteria according to solvency level. Rating number (B) is determined as a sum of scores on each indicator $B(R_i)$. Determination of scores on a separate indicator is fulfilled by the method of linear interpolation. Linear interpolation (Meijering, 2002) is the simplest method of getting values at positions in between the data points, according to the following formula:

$$B(R_i) = B_{i\min} + \frac{B_{i\max} - B_{i\min}}{R_{i\max} - R_{i\min}} \cdot (R_i - R_{i\max}) \tag{3}$$

where

$B_{i\max}, B_{i\min}$ - being at positions in between the data points of a certain class;
$R_{i\max}, R_{i\min}$ - an interval of an indicator's value relevant to a certain class of i indicators.

The benefit of the introduced method is simplicity which allows this method to be applied as a constituent part of the test for evaluating enterprise solvency risk. For example, under the leadership of the author from 2005 to 2010 Duran's technique was used more than 300 times for express assessment of the financial risk of Latvian enterprises. In two cases the evaluated results did not conform to reality and the express analysis had to be supplemented with a deeper analysis of the stability of industrial activity.

4.2 Review of the financial stability indicators for risk assessment

The author considers it possible to implement express analysis of financial risks on the basis of the principles of the analysis of the balance sheet of enterprise economic turnover. The methodology was developed by M.S. Abrjutina and A.V. Grachev (1998) for the purpose of defining the critical point in the field of steady equilibrium in the system of structural ratios of equity and loan capital of an enterprise. N. Lace and Z. Sundukova (2010) researched different approaches for the evaluation of the level of financial stability including the evaluation of financial stability suggested by M. S. Abrjutina and A.V. Grachev (1998). The authors (Lace & Sundukova, 2010) came to the conclusion that the classification and structure of assets, as well as the financial policy of an enterprise, can be the most important factors for the evaluation of financial stability.

The given results confirm the fact that the concept of economic turnover balance sheet (ETB) is acceptable for the assessment of second level financial risks. Practical research on the sample of six Latvian branches (in terms of small, medium and large enterprises and a branch as a whole) in defining actual and sufficient ratios for financial stability indicators (current liquidity ratio and owner equity to total assets ratio) were executed by N. Lace and Z. Sundukova (2010).

Express assessment of financial risks can be fulfilled using the same methods which are applied for assessing enterprise financial soundness. Let us consider the methods of express analysis of financial risks, based on the analysis of the net assets value (NAV) and the principles of the analysis of an enterprise's ETB.

Assessment of enterprise financial soundness on the basis of NAV supposes the comparison of NAV with company capital (CC) and its minimum value ($CC_{\min}$) in compliance with the law, regulating the registration right of an enterprise. The following variants of NAV ratios CC and outcomes from the given comparisons are possible: NAV>CC- normal situation at an enterprise, financial risks are under control; NAV=CC – situation is critical, it is necessary to carry out the analysis of the financial situation, separate kinds of financial risks cause fears; NAV<CC, NAV> $CC_{\min}$ – situation is critical, eating away of own capital is evident. What is required is deep analysis of financial state. Introduction of anti-crisis management into operation is desirable, many kinds of financial risks cause fears. NAV< $CC_{\min}$ - crisis situation, extraordinary increase of own capital, threat of insolvency, pessimistic development of the scenario of enterprise existence threatening its winding up. NAV=0 and NAV<CC<0 – crisis state, own capital is used up inefficiently, the enterprise should be liquidated.

These six different situations have been considered and a diagnosis has been stated. Each diagnosis requires making certain managerial decisions for decreasing risks, having their own specifics for individual countries. This finding related to Russia can be found in the study by V. M. Voronina (2007). The technique ETB is based on the principle of national accounts. The construction of economic turnover balance sheet relies on the process of structuring enterprise property and combining enterprise property and the structure of enterprise assets with the capital structure.

This concept is based on the division of assets (A) of an enterprise into financial (FA) and non-financial assets (NFA). In its turn, financial assets are divided into mobile (MFA)[1] and non-mobile (NMFA)[2]. Non-financial assets are divided into long-term non-financial assets

[1] Mobile financial assets are highly liquid financial assets: cash and easily convertible short-term financial assets.

(LNFA)[3] and current non-financial assets (CNFA) (including inventories).

According to Abrjutina & Grachev (1999) the concept of financial balance and stability is reached, in the case of non-financial assets they are financed at the expense of own capital (OC), whereas financial assets - at the expense of debt capital (DC). The main principles of the technique are as follows:

1. Structuring of accounting balance sheet, separating economically homogeneous elements into assets and liabilities:

$$A = FA + NFA = (MFA + NMFA) + (LNFA + CNFA) \tag{4}$$

$$C = OC + DC \tag{5}$$

Table 3 shows classification of assets.

Classification of assets by operating cycle time				Classification of property by form			
Assets	Non-financial assets	Financial assets	In total	Property (P)	Own	Debt	In total
Non-Current (N)	NNA	NFA	NCA	Non – monetary form	OC_{NMF}	DC_{NMF}	$P - CFA$
Current (C)	CNA	CFA	CA	Monetary form	OC_{MF}	DC_{MF}	CFA
In total	NFA	FA	A	In total	OC	DC	P

Table 3. Classification of assets/property by operating cycle time and by form

Current assets (CA) can be divided into own current assets (OCA) and debt current assets (DCA). Own current assets (OCA) are provided by a part of own capital (OC), but debt current assets (DCA) are provided by entire debt capital (DC). Own current assets, being financed at the expense of own capital, are called working capital (WC):

$$WC = CA - CL \tag{6}$$

An enterprise with negative working capital may lack the funds necessary for growth and is in a state of instability.

2. Determination of the indicator of financial and economic stability (IFS):

$$IFS = OC - NA \tag{7}$$

$$IFS = FA - DC \tag{8}$$

Creation of three-positional static scale (see table 4).

The given scale is simple, but it does not exclude the possibility of deepening (but simultaneously complicating) the analysis as well as determining FIS in dynamics.

[2] Non-mobile assets comprise long-term assets, all kinds of accounts receivable and quick deposits.

[3] Including fixed assets, intangible assets, incomplete construction.

3. After further division of assets into subgroups it is possible to single out five different variants of the state of enterprise stability (see table 5).

The author supplemented these characteristics of stability with risk scale. According to the research by Abrjutina & Grachev (1998) the number of complex characteristics accounts for 340 combinations of shifts. The research of all types of shifts from one level of the state of enterprise stability to another over the analysed periods is not expedient while using ETB as a method of express diagnosis.

4. Conducting positioning of the state of financial stability on the map of financial stability (see fig. 3).
5. Creation of dynamic scale of financial stability.

The risk zone is characterised by the lack of own funds including those in monetary terms. A larger part of enterprise property is borrowed. To leave the risk zone what is required is growth of own capital, introduction of the plan of economical usage of own monetary funds and costs reduction. The financial strain zone is a zone of relative financial stability and welfare.

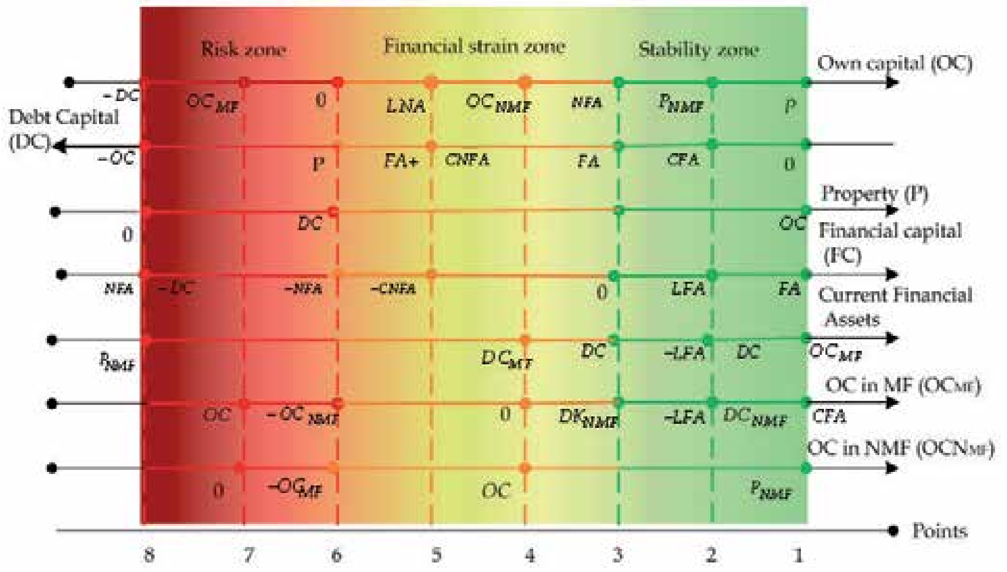
Variants of the state of enterprise stability	Value of indicator *FIS*	Evaluation of own capital	Evaluation of debt capital
State of stability	$FIS > 0$	$OC > NSFA$	$DC > FA$
State of equilibrium	$FIS = 0$	$OC = NSFA$	$DC = FA$
State of uncertainty	$FIS < 0$	$OC < NSFA$	$DC < FA$

Table 4. Static (main) scale of financial-economic stability of an enterprise

Characteristics of variants	Name		Characteristics of financial risk level of	
			risk	influence
Mobile financial assets exceed other liabilities	Super stability (absolute solvency)	Stability	Not substantial	Not significant
Mobile financial assets are less than all the other liabilities, but the amount of all financial assets is bigger than all the other liabilities	Sufficient stability (solvency guaranteed)		Little value	Not significant
Own capital is equal to non-financial assets, but financial assets are equal to all liabilities	Financial balance (lack of stability margin)		Moderate	Justified
Own capital is more than long-term financial assets, but less than all the amount of non-financial assets	Admissible financial strain (potential solvency)	Instability	Sensitive	Not accessible
Own capital is less than long-term non-financial assets	Risk zone (loss of solvency)		Critical	Not accessible

Table 5. Variants of financial state stability and risk of enterprise (Abrjutina & Grachev with the author's supplements)

Enterprise own capital is used to finance long-term non-financial assets but not sufficiently. Financial capital has a negative value and own monetary funds appear only by approaching the financial equilibrium point. It is necessary to undertake operative planning of cash flow usage, improve the work in the field of strategic planning and introduce changes to the enterprise marketing policy.



1 – super stability point; 2 - point of sufficient stability; 3 - point of financial equilibrium; 4 - point of stability loss; 5 - point of insufficient stability; 6- instability point; 7- risk point; 8 - bankruptcy point.

Fig. 3. Map of financial stability

4.3 A system of tests for determining financial risks for enterprises being at different stages of life cycle

Based on the studies of the origin and application of financial ratios for assessing financial risks described in chapter 3, we can come to the conclusion that the application of financial ratios to analyse financial risks in general brings good results.

However, the length of the life cycle of an enterprise for different countries in various branches differs. The theory of economic cycles has been known for over 2,500 years. The most recognised cycles are those proposed by N. Kondratjev and Kutchin with a length of three to five years concerned with a relative value of material resource reserves at enterprises, the Dzagler cycles, lasting seven to ten years occurring as a result of the interaction of different credit-monetary factors and the S. Kuznet cycles with a length of around 29 years, resulting from the terms of reproduction in building (Dagum,2010). The problems of risk and business cycles are investigated by Cower (2003), the issues of financing procedures in the framework of business cycle theory are covered by Mallick (2008) within six and 11-16 years. The link between credit risk and cyclical processes in the economy was researched by Koopman and Lucas (2003). However, the given research related to macroeconomic models of default assessment and does not allow for assessing the default of a certain borrower.

Most authors, for example Kislingerová (2004), mention that those phases of corporate life cycle can be identified according to the value of cash flow. According to the model by Reiners (2004) there are in total 16 combinations of phases of corporate and market life cycle. Different researchers (Shorokova, 2006), emphasise various sets of characteristics unique to each stage of their models.

Based on the research of entrepreneurial cycles (Dagum,2010; Korotayev & Trirel, 2010) we offer some practical recommendations – before starting to research a risk system it is recommended to position an enterprise according to the age scale of enterprise risk 5-6, 11, 16-17, 22, 45-67 … years[4]. An attentive reader may come to the practical conclusion, why planning with horizon of 5 years is used.

However, regardless of the number of stages (Shorokova, 2006), there are some commonalities in their conclusions. Firstly, the model stages are sequential. Secondly, each stage is a result of the previous one and is difficult to revert. Thirdly, all models consider a wide range of contextual organisations. To analyse risks the author uses three-stage models of life cycle of an enterprise. At the start-up stage an enterprise from the point of view of financial risk is more prone to the impact of external factors. Taking into account that an enterprise can control only internal factors at this stage it is required to control a share of loan capital, profitability of main activity and liquidity indicators. However, because of instability of stability indicators at a given stage, further factors are the indicators of the efficiency of the main activity of the enterprise: volume of sales, production, cost value. It is important to assess their dynamics. Depending on the change in dynamics of profit from sales and cost value, the level of financial risk will be determined (see Table 6).

Level of financial risk	Evaluation indexes		
	Fixed assets cove rage ratio $FACR$	Dynamics of indexes	
		Profit from sales	Cost value per unit
Minimum	$FACR \geq 1$	$P_p(t) = P_p(0) + at$	$I_p(t) = I_p(0) - at$
Moderate	$FACR \geq 1$	$P_p(t) = P_p(0) + at$	$I_p(t) = I_p(0) + at$
High	$FACR \leq 1$	$P_p(t) = P_p(0) - at$	$I_p(t) = I_p(0) - at$
Situation of occurrence of financial risk	$FACR \leq 1$	$P_p(t) = P_p(0) - at$	$I_p(t) = I_p(0) + at$

Table 6. Determination of the level of financial risk of an enterprise at the start-up stage

At the stage of the assessment of the risk of enterprises being at the start-up stage in addition to the above – it is necessary to take into consideration the condition of financial stability of an enterprise in operation. At the expansion stage, a rapid increase in profit and stabilisation of financial indicators in relation to own and loan capital occurs. The profit is considered as fast growing dynamics of indicators. It is possible to assess the level of financial risk of an enterprise at this stage on indicators of relationship between own and loan capital. The profit is viewed as fast growing, dynamics of volume of sales is positive (see table 7). At the maturity stage an enterprise operates in full swing, the indicators are stable but due to increasing competition and wearing out of capital it may shift to the

[4]For example, in autogenetics it is possible to determine critical points on the age scale of enterprise risk with more precision. We find it possible to use astrogenetics for this purpose (Budjashkina, 2003, 2003a).

decline stage. In this case it is necessary to control the volume of sales and the turnover of assets will testify to the reduction of competitiveness of production and increase in stock. At this stage it is not possible to judge the level of the financial risk of an enterprise according to the indicators of cost value and profits from sales. A state of renovation of assets may be introduced which can show the indicators, but it does not always mean the increase of the financial risk of an enterprise (see table 8).

Level of financial risk	Evaluation indexes			
	Fixed assets coverage ratio ($FACR$)	Borrowed and own capital ratio	Dynamics of indexes	
			Borrowed and own capital ratio	Financial leverage
Minimum	$FACR \geq 1$	$K_{boc} \leq 1$	$K_{boc}(t) = K_{boc}(0) - at$	$FL(t) = FL(0) + at$
Moderate	$FACR \geq 1$	$K_{boc} \geq 1$	$K_{boc}(t) = K_{boc}(0) - at$	$FL(t) = FL(0) - at$
	$FACR \geq 1$	$K_{boc} \leq 1$	$K_{boc}(t) = K_{boc}(0) + at$	$FL(t) = FL(0) - at$
High	$FACR \geq 1$	$K_{boc} \geq 1$	$K_{boc}(t) = K_{boc}(0) + at$	$FL(t) = FL(0) - at$
	$FACR < 1$	$K_{boc} \leq 1$	$K_{boc}(t) = K_{boc}(0) - at$	$FL(t) = FL(0) + at$
	$FACR \leq 1$	$K_{boc} \geq 1$	$K_{boc}(t) = K_{boc}(0) - at$	$FL(t) = FL(0) + at$
	$FACR \geq 1$	$K_{boc} \leq 1$	$K_{boc}(t) = K_{boc}(0) + at$	$FL(t) = FL(0) + at$
Situation of occurrence of financial risk	$FACR < 1$	$K_{boc} \geq 1$	$K_{boc}(t) = K_{boc}(0) + at$	$FL(t) = FL(0) + at$

Table 7. Determination of level of financial risk of an enterprise at the expansion stage

At the maturity stage the criteria of the level of financial risk assessment are the following dynamics of sales volume and working capital turnover. Methods of special coefficients are usually applied to discover financial risk by using publically available information. To investigate the causes of insolvency it is common practice to mention the following combinations of two reasons: small value of own capital and loss-making activities as well as changes in the market and loss-making activities.

Level of financial risk	Evaluation indexes		
	Fixed assets coverage ratio (FACR)	Dynamics of indexes	
		Net turnover(NT)	Turnover of circulating assets
Minimum	$FACR \geq 1$	$NT(t) > NT(0)$[5]	$K_a(t) > K_a(0)$
Moderate	$FACR \geq 1$	$NT(t) > NT(0)$	$K_a(t) < K_a(0)$
High	$FACR \leq 1$	$NT(t) < NT(0)$[6]	$K_a(t) > K_a(0)$
Situation of occurrence of financial risk	$FACR < 1$	$NT(t) < NT(0)$	$K_a(t) < K_a(0)$

Table 8. Determination of the level of financial risk of an enterprise at the maturity stage

[5] $NT(t) = NT(0) + at$

[6] $NT(t) = NT(0) - at$

4.4 Case study of the risk occurrence reasons

To manage financial risks it is not enough to assess them by using the instruments of financial analysis. It is necessary to probe into the reasons for risk occurrence and employ opportunities for removing/decreasing the impact of these reasons within the financial state of an enterprise. BPEST, SWOT and E-SWOT analyses are recommended in order to identify the role of financial problems in the system of strengths and weaknesses of an enterprise. Expert methods of financial risks assessment and the reasons for their occurrence are especially useful in cases of the shortage of necessary information and the involvement of enterprise personnel into the process of risk assessment. Incorporation of large numbers of employees at the stage of identification, description and risk assessment is an important step in the creation of the system of risk management. To apply an expert method we recommend using scales of the assessment of the value of the reasons of risk development (or value of risk types) and the occurrence of risk possibility assessment in scores or in terms of probability.

Assessments of the value of reasons/types of risks can be conducted by using a score scale (for example ten score). If the reasons/risks can be sorted out in decreasing scale then in order to find the coefficient of significance it is possible to use Fishburn's formula (9) (Fishburn 1970; Baron & Barrett 1996; Potapov & Evstafjeva 2008):

$$w_i = \frac{2 \cdot (m - i + 1)}{(m + 1) \cdot m} \tag{9}$$

where
m -a number of reasons/risks in a group;
i – an ordinal number of reason/risk in a group.

If the reasons/risks have equal value, then the significance coefficient is identified according to the formula:

$$w_i = {}^{1}\!/_{m} \tag{10}$$

To define the probability of the occurrence of the certain risk types/reasons we can assume as a basis the following distribution of probabilities (p_i): low risk 0 – 0.25; moderate risks 0.26 – 0.4, high risk 0.41 - 0.7 and critical risk 0.71 - 1.0.

Table 9 gives an example (conventional) of the assessment of the reasons of the occurrence of enterprise liquidity risks. The given example illustrates the possibility of using an expert method for the analysis and quantitative assessment of the reasons of financial risks occurrence.

Identification of the possibility of risk occurrence is carried out according to the formula (11)

$$R_i = \sum_{i}^{n} w_i \cdot p_i \tag{11}$$

By engaging several experts assessment is conducted either on the basis of all experts' collective opinion or on the basis of individual assessment.

The results of the individual experts' assessments are summarised taking into account coefficients of expert competence, then is determined coefficient of concordance of experts' opinions and is carried out the concordances test analysis.

Reasons of the origin of liquidity risk (R_{li})	Range	w_i	p_i	$w_i \cdot p_i$
R_{l1} Irrational usage of fixed assts				0.144
R_{l11} Presence of large amounts of non-used equipment	8	0.04	0.6	0.024
R_{l12} Presence of spare areas	4	0.13	0.5	0.065
R_{l13} Possible losses at the process of realisation of the investment object due to changes in its quality assessment	5	0.11	0.5	0.055
R_{l2} Irrational management of working capital				0.172
R_{l21} Presence of super normative reserves of finished goods	1	0.2	0.4	0.08
R_{l22} Presence of super normative material reserves	2	0.18	0.5	0.09
R_{l23} Errors in the system of operational management of product shipment to customers	9	0.02	0.1	0.002
R_{l24} Inappropriate control over terms of payment for production	6	0.09	0.5	0.045
R_{l3} Operational risks. Problems with liquidity management				0.076
R_{l31} Absence of the system of internal audit of working capital accounts	3	0.16	0.3	0,048
Negligence of employees engaged in loading/unloading of production/materials	7	0.07	0.4	0.028
Assessment of the possibility of the occurrence of liquidity risk taking into account the reasons for risk origin - moderate risk				0.392

Table 9. Example of the assessment of the reasons of the occurrence of enterprise liquidity risk

If the number of reasons/types of risks is over ten then for the convenience of implementing an expert assessment they should be united in subgroups and a coefficient of groups' values should be determined according to the formula (9) with further convolution.

Financial risk management is an integral part of a complex system of enterprise risk management and financial risks should be positioned in the map of enterprise risk. As a result, by carrying out an assessment of financial risks it is necessary to use expert methods (assessment with the usage of the scale of the occurrence of an event and the severity of consequences). The introduced approach of risk assessment using the range of reasons/types of risk is supplemented by determining the priority of reasons/types of risk according to Fishburn's formula.

5. Discussion

A pessimistically oriented reader could come to the conclusion that if methods and models do not allow reaching an acceptable result and risk assessment, then they will not be evaluated if the supervisory bodies do not require assessment. Such a position is not in compliance with the idea of entrepreneurship i.e. making reasonable decisions in the circumstances of risk and uncertainty.

What about the way out? To use, for assessment, not a single model but a combination of models. A.Miller (2002), S. Kealhofer (2003), A. Boykova (2010) and many other researchers (whose position we also share) conducted the monitoring of developed models holding the same view.

Enterprise financial risks have different risk-forming factors which change in relation to the stage of enterprise life cycle. It is necessary to position an enterprise on a temporary risk scale and use different indicators for the assessment of financial risk level according to the stage of life cycle. Our test system of enterprise risks related to its location at the life cycle stage and can be developed for four to five stage cycles in the future.

As for small and medium-sized enterprises we recommend using express analysis methods such as Quick Tests and Duran's technique at the stage of the creation of a system of risk management, including financial risk management. The usage of these types of models allows assessing which risk zone an enterprise is located or will be located according to the results of plan indicators. The application of these methods does not require calculation of a great number of indicators (from four to eight). However, in order to use Kralicek's Quick Test we need information about cash flow which is not always accessible for public analysis. Application of Kralicek's Quick Test in express assessment of financial risks needs experimental checks with a view to specific economic conditions of the countries.

Research in selecting models of discriminant type, particularly in the Baltic countries, Ukraine and Russia, suitable for assessing financial risks of bankruptcy should be continued. It is necessary to monitor MDA type models more efficiently (not less than once in five years) to select the most viable models with a view to branch and country specifics.

6. Conclusions

The development of the system of managerial accounts and usage of a simplified conception enterprise economic turnover balance sheet (ETB) (as the first step to its implementation) is an instrument for both assessment and creation of a system of financial risk monitoring.

Based on the analysis of over 100 studies of the authors from nine countries the author chose those Z-score methods (see Appendices) which are more interesting. They were developed by means of representative selection of enterprises and are frequently mentioned/used for the purposes of the risk assessment of their own enterprises and business partners. We are not going to set the task of teaching how to build discriminant models, on the contrary – our goal is to provide information about sufficiently simple instruments of the second and third types of risk assessment (risk of financial stability loss, insolvency and bankruptcy) using official accountancy data and established financial ratios which have a long-standing history of practical utilisation.

The main principle of financial risk management of small and medium-sized enterprises is to plan and control their activities, using simple proven instruments but not rejecting new ones; in this way you will be able to manage the development of your business.

7. Appendices

Symbols adopted for tables 2, 3, 4 and 5: $A = TA$ - assets (total assets); AA - average assets; AAR - average account receivable; ATA - average level of total assets; BSP - balance sheet profit; C -cash; CA - current assets; C_b - cash and bank account; CF - cash flow; CL - current liabilities; CS - cost of sales; CR - rate loan repaid; C_u - customers; COA - costs of operation activity; $COGM$ - cost of goods sold; D - debt; D - debt capital; D_a - depreciation; D_u - duties; E - equity; $EBIT$ - earnings before interest and taxes; EMV -equity market value; FA -fixed assets; FU - funds provided by operations;

$FOWC$ - financial or onerous working capital; $FSTA$ - financial short term assets; $FINLEV$ - financial leverage; GP - gross profit; I - inventories; I_n - interest; IE - interest expense; IN – interest paid; L - liability (short- and long-term); TL - total liabilities; LTL - long- term liability; $\ln(E)$ - natural logarithm of an own capital of the enterprise; NCS - net credit sales; ND - net debt; NP - net profit; $NRS = NR = NS = NSR$ - net revenue from sales (net revenue, net sales, net sales revenue); NI - net income; NE - net earnings; $NPATI$ - net profit after tax and interest; NWC -net working capital; NP – net profit; NI - net income; NS - net sales; NT - net turnover; NI_t - net income for the most recent period; PBT - profit before tax; $NWKSA$ - needs of working capital; OC - own capital; OCA - own current assets; OCF - cash flow from operating cash flows; OE - operating expenses; $OSTL$ - onerous short term liabilities; OP - operating profit; OOC - other operating costs; OWC - owner's working capital; P - profit; P_o - overdue payable; PS - profit from sales; PBT - profit before tax; $PBIT$ - profit before interest and tax; QR - Quick ratio; T - turnover; TA - total assets; TC - total capital; TCA -total current assets; TL -total liabilities; TCL - total current liabilities; TD - total debt; TI -total income; TS - total sales R - revenue; Rr -refinancing rate; RS - return on sales(RS - revenues from sales); ROA - return of assets; ROE - return on equity; RTO -revenue from total operations; RP - retained profit; S - sales (net) (E-expected; B- breakeven); S_t - stocks; STD - short-term debtors; STL - short-term liabilities; T_A - growth rate of enterprise assets; T_E - growth rate of own capital of the enterprise; WC - working capital.

Model	Type of model
Logit-model Ohlson J (1980)	$P = \frac{1}{1+e^y}$ $\qquad CHIN = \frac{NI_t - NI_{t-1}}{\lvert NI_t\rvert + \lvert NI_{t-1}\rvert}$ $y = -1{,}32 - 0.407 \cdot SIZE - 6.03 \cdot TL / TA - 1.43 \cdot WC / TA + 0.076 \cdot CL / CA -$ $-1.72 \cdot OENEG - 2.37 \cdot NI / TA - 1.83 \cdot FU / TL - 0.285 \cdot INTVO - 0.521 \cdot CHIN$ $ZIZE = \log(\text{TA/GNP price-level index})$ $OENEG = \begin{cases} 1, if \ TL > TA \\ 0, if \ TL < TA \end{cases} \quad INTVO = \begin{cases} 1, if \ \text{NI} < 0 \ \text{for the last two years} \\ 0, if \ \text{NI} > 0 \end{cases}$
Logit-model Begley J., et al. (1996)	$P = \frac{1}{1+e^y}$, $y = -1.249 - 0.211 \cdot SIZE - 2.262 \cdot TL / TA - 3.451 \cdot WC / TA - 0.293 \cdot CL / CA -$ $-0.907 \cdot OENEG + 1.080 \cdot NI / TA - 0.838 \cdot FU / TL - 1.266 \cdot INTVO -$ $-0.960 \cdot CHIN$
Logit-model Grigaravičiaus (2003)	$\text{Pr}(1) = \frac{e^{-z}}{1+e^z}$; $\text{Pr}(0) = \frac{1}{1+e^z}$; $Z = -0.762 + 0.003 \cdot CA / TL + 0.424 \cdot WC / TA - 0.06 \cdot TA / E + 0.22 \cdot E / CL -$ $-0.774 \cdot PBIT / TA + 6.842 \cdot ROA - 12.262 \cdot CS / NWC - 5.257 \cdot R / TA$
Logit-model Minussi J. et al. (2007)	$P = \frac{1}{1+e^y}, WOWKSA \ = \ FOWC / S; \ FOWC = FSTA - OSTL;$ $y = -5.76 - 2.53 \cdot FOWKSA + 0.48 \cdot FINLEV - 0.17 \cdot EBIT / IE - 1.02 \cdot OWKSA -$ $-0.63 \cdot (CA - CL) / S; \ \ OWKSA = OWC / S.$

<table>
<tr>
<td>Logit – model Lukason (2006) trading companies</td>
<td>

$$P = \frac{1}{1+e^{K}},$$

$$K = 0.123\ln(A) + 37.188\cdot NP / AA + 0.006\cdot TN / AAR\cdot +22.816\cdot CF / S$$

</td>
</tr>
<tr>
<td>Logit-model Haydarshina G.A. (2008)</td>
<td>

$$C^{BR} = \frac{e^{y}}{1+e^{y}} \quad y = \alpha_0 + \alpha_1\cdot Corp\_age + \alpha_2\cdot Cred + \alpha_3\cdot CA / CL + +\alpha_4\cdot EBIT /$$

$$/IE + \alpha_5\cdot\ln(E) + \alpha_6\cdot Rr + \alpha_7\cdot \mathrm{Re}\,g + \alpha_8\cdot ROA + \alpha_9\cdot ROE + \alpha_{10}\cdot T\_E + \alpha_{11}T\_A$$

$$Corp\_age = \begin{cases} 0 \text{ - if an enterprise was established more than 10 years ago,} \\ 1 - \text{if an enterprise was established less than 10 years ago.} \end{cases}$$

$$Cred = \begin{cases} 0 - \textit{if credit history is positive,} \\ 1 - \text{ othewise.} \end{cases}$$

<table>
<tr><th rowspan="2">Ratio</th><th colspan="3">Economic sector</th><th rowspan="2">Ratio</th><th colspan="3">Economic sector</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>1</th><th>2</th><th>3</th></tr>
<tr><td>α_0</td><td>10.2137</td><td>30.7371</td><td>35.0326</td><td>α_7</td><td>-1.3698</td><td>-0.6913</td><td>-0.8023</td></tr>
<tr><td>α_1</td><td>0.0303</td><td>3.7033</td><td>4.1834</td><td>α_8</td><td>-6.3609</td><td>-5.0894</td><td>-8.4776</td></tr>
<tr><td>α_2</td><td>6.7543</td><td>8.9734</td><td>9.0817</td><td>α_9</td><td>-0.2833</td><td>-15.3882</td><td>-10.8005</td></tr>
<tr><td>α_3</td><td>-3.7093</td><td>-8.6711</td><td>-8.7792</td><td>α_{10}</td><td>2.5966</td><td>7.3667</td><td>7.1862</td></tr>
<tr><td>α_4</td><td>-1.5985</td><td>-7.0110</td><td>-8.5601</td><td>α_{11}</td><td>-7.3087</td><td>-20.0294</td><td>-22.7614</td></tr>
<tr><td>α_5</td><td>-0.5640</td><td>-1.6427</td><td>-1.6834</td><td colspan="4" rowspan="2">1 – industry; 2 – full and energy industry; 3 – trade.</td></tr>
<tr><td>α_6</td><td>-0.1254</td><td>-0.1399</td><td>-0.4923</td></tr>
</table>

$$\mathrm{Re}\,g = \begin{cases} 0 - \text{ if an enterprise is located in Moscow,} \\ 1 - \text{ if an enterprise is located in other regions of Russia.} \end{cases}$$

Characteristics of enterprise bankruptcy: If $0.8 < C^{BR} \le 1$ is maximum risk; if $0.6 < C^{BR} \le 0.8$ high risk; if $0.4 < C^{BR} \le 0.6$ average risk; $0.2 < C^{BR} \le 0.4$ low risk; if $0 < C^{BR} \le 0.2$ minimum bankruptcy risk.

</td>
</tr>
<tr>
<td>Logit-model Genriha, Pettere& Voronova (2011)</td>
<td>

$$PD = \frac{1}{1+e^{-z}}\ ; \quad L(x_i;a,b)_i = \frac{1}{1+e\{-(a+b\cdot x_i)\}};$$

$$Z = 25{,}998\cdot L_1 + 33{,}358\cdot L_2 + 16.208\cdot L_3 - 5.662$$

<table>
<tr><th>x_i</th><th>Rate</th><th>a</th><th>b</th></tr>
<tr><td>x_1</td><td>PBT / E</td><td>-3,03080</td><td>-2,06326</td></tr>
<tr><td>x_2</td><td>NT / TA</td><td>-1,07324</td><td>-3,59669</td></tr>
<tr><td>x_3</td><td>LTL / TA</td><td>-6,80139</td><td>4,35983</td></tr>
</table>

Characteristics of enterprise bankruptcy risk: PD states default probability when during one year an enterprise will not be able to pay back its credit obligations for a period longer than 90 days.

</td>
</tr>
</table>

Logit-model (Static and dynamic) Alekseeva (2011)	$P = \frac{1}{1+e^{y}}$; $y = (-(32.633 - 1.082 \cdot S/TA - 6.932 \cdot NP/TA + 3.697 \cdot L/A - 5.712 \cdot LTL/A -$ $-1.573 \cdot \ln(E))$; $P_{t+1} = \frac{1}{1+e^{-Y}}$; $\mathrm{Y} = 9.91 \cdot P_t + 0.213 \cdot \frac{\mathrm{P_t}}{\mathrm{P_0}} - 3.58$ P_{t+1} -probability of enterprise bankruptcy per year *t+1*, per year *t* and in a year *t-1*.

Table A1. Brief description of the logit models for assessing the risk of enterprises bankruptcy

Country and name of model		Model type and interpretation of results. Assessment of bankruptcy threat level
Czech Republic	Classic Altman's model for the Cz	$Z = 1.2 \cdot NWC/A + 1.4 \cdot E/A + 3.3 \cdot EBIT/A + 0.6 \cdot EMV/L +$ $+1 \cdot S/A + 1 \cdot P_0/S$ If $Z > 2.9$ - financially stable enterprise and if $Z > 2.9$ - threat of bankruptcy.
	Bonity index IB	$IB = 1.53 \cdot CF/D + 0.08 \cdot A/D + 10 \cdot EBT/A + 5 \cdot EBT/R +$ $+0.3 \cdot I/R + 0.1 \cdot R/A$ If $IB > 3$ can be marked as extremely, $IB > 1$ is good and $IB < -2$ are directly jeopardised by bankruptcy.
	Czech index IN05 Neumaierová &Neumaier (2005)	$IN05 = 0.13 \cdot A/D + 0.04 \cdot EBIT/I_n + 3.97 \cdot EBIT/A + 0.21 \cdot R/A +$ $+0.09 \cdot CA/STL$ If $IN05 < 0.9$, the enterprise with a high probability (86%) moves towards bankruptcy and if $IN05 > 1.6$ the enterprise creates EVA with probability of 67%.
Poland	FD model I	$FD = 9.498 \cdot OP/AA + 3.566 \cdot E/A + 2.903 \cdot (NE + D_a)/TL + 0.452 \cdot CA/$ $/CL - 1.498$ If $FD \leq 1.1$ the probability of bankruptcy is high; $1.1 < FD < 2.6$ the probability of bankruptcy is indefinite (grey area); $FD > 2.6$ the probability of bankruptcy is low.
	Holda (ZH) (2006)	$Z_H = 0.605 \cdot +0.681 \cdot CA/CL - 0.0196 \cdot TL/TA \cdot 100 + 0.00969 \cdot (NP/AA)$ $\cdot 100 + 0.000672 \cdot (ACL/(COA - OOC) \cdot 360 + 0.157 \cdot R/AA$ If $Z_H = -0.3$ the probability of bankruptcy is high; $-0.3 < Z_H < 0.1$ the probability of bankruptcy is indefinite (grey area); $Z_H > 0.1$ the probability of bankruptcy is low.
	Model B. Prusak (Z_{BP1})	$Z = 6.5245 \cdot OP/AA + 0.148 \cdot OE/(ASL - SF - SFL) +$ $+0.4061 \cdot (CA - DC)/STL + 2.1754 \cdot OP/NS - 1.5685$ If $Z < -0.13$, the enterprise with a high bankruptcy risk; if $-0.13 \leq Z \leq 0.65$ grey area, if $Z > 0.65$ the probability of bankruptcy is low.

	J. Gajdka D. Stosa (1996)	$Z2 = 0.7732059 - 0.0856425 \cdot S / AA + 0.0007747 \cdot L / COGS \cdot 365 +$ $+0.9220985 \cdot NP / AA + 0.6535995 \cdot GP / NS - 0.594687 \cdot TL / AA$ If $Z > 0.45$ no threat to the enterprise; if $Z < 0.45$ threat of bankruptcy.
	Poznan Model *Hamrol, Czajka,* (2004),	$Z_{HCP} = 3.562 \cdot NR / TA + 1.588 \cdot (CA - I) / CL + 4.288 \cdot FA / TA +$ $+6.719 \cdot PS / NSR - 2.368$ If $Z > 0$ very low chance of bankruptcy; $Z = 0$ critical point; if $Z < 0$ - very high probability of bankruptcy.
Romania	Model B – Băileştea nu (1998)	$B = 0.444 \cdot CA / CL + 0.909 \cdot (NP + D_a) / (CR + I_n) + 0.0526 \cdot CA /$ If $/C_u + 0.0333 \cdot P / CS \cdot 100 + 1.414$ $B < 0.5$ the enterprise with a high bankruptcy risk; if $0.5 < B < 1.1$ - high financial risk zone; $1.1 < B < 2.0$ - moderate financial risk zone; $B > 2.0$ - financially appropriate zone.
	Model I – Ivonciu (1999)	$I = 0.333 \cdot S / FA + 5.555 \cdot GP / TI + 0.0333 \cdot NCS / AAR + 0.714229 \cdot$ $\cdot ND / EBIT + 1.333 \cdot (CA - I) / CL + 4.0 \cdot (ES - BS) / BS - 1.66032$ $EBIT = R - Expenses$. If $I < 0$ bankruptcy is imminent; $0 < I < 1.5$ - high bankruptcy risk with a 64-81% probability; $1.5 < I < 3$ - uncertainty bankruptcy risk with a 46-64% probability; $3 < I < 4.5$ - there is a moderate risk of bankruptcy with a 29-46% probability; $4.5 < I < 6.0$ - low bankruptcy level with a 12-29% probability; $I > 6$ - shows a good financial state and the bankruptcy probability is very low (0-12%).
	Angel Model (2002)	$A = 5.676 + 6.3718 \cdot NP / R + 5.3932 \cdot CF / A - 5.1427 \cdot D / A -$ $-0.0105 \cdot L / T \cdot 360$ If $A < 0.0$ bankruptcy/failure situation; $0 \leq A \leq 2.0$ uncertainty situation demanding prudence; $A > 2.05$ non bankruptcy situation, a good financial situation.

Table A2. MDA models for assessing the risk of insolvency/bankruptcy Czech Republic, Poland and Romania

Country and name of model		Model type and interpretation of results. Assessment of bankruptcy threat level
Belarus	Byelorussion Model	$Z_B = 0.111 \cdot OCA / CA + 13.239 \cdot CA / FA + 1.676 \cdot S / TA +$ $+0.515 \cdot NP / A + 3.80 \cdot E / TC$ If $Z_B > 8$ then the enterprise bankruptcy does not threaten; if $5 < Z_B < 8$, risk of bankruptcy is small; if $3 < Z_B < 5$, financial condition is average, there is the risk of bankruptcy under certain circumstances; if $Z_B < 1$ the enterprise is bankrupt.
Rus-sia	Fedotova (1995)	$R = -0.3877 - 1.0736 \cdot CA / CL + 0.0579 \cdot D / TA$ If $Z < 0$ there is probability that an enterprise remains solvent; if $Z > 0$ bankruptcy is probable.

	Sayfullin& Kadykov (1996)	$R = 2 \cdot OCA / CA + 0.1 \cdot CA / STL + 0.08 \cdot P / RS + 0.45 \cdot R / AA + NI / O$ If $Z < 0$ bankruptcy is imminent.
	Davidova &Beljakov (1999)	$Z = 8.38 \cdot WC / A + EBIT / AA + 0.054 \cdot$ EBT/$AA + 0.63 \cdot EBIT / OC$ If $Z < 0$ bankruptcy is imminent; If $0 < Z < 0.18$ - high bankruptcy risk with a 60-80% probability; $0.18 < Z < 0.32$ - average risk of bankruptcy with a 35-50% probability; $0.32 < Z < 0.42$ - there is low bankruptcy level with a 15-20% probability; $Z > 0.42$ - bankruptcy probability is very low (10%).
	SMEs Model Lugovskaya (2010)	$Z = -0.05 - 0.61 \cdot C / CL + 0.07 \cdot$ CA/CL $+ 0.34 \cdot (C + STD) / CL -$ $-1.13 \cdot (C + STD) / TA + 1.35 \cdot ROA + 8.42 \cdot C / TA$
Ukraine	Martinenko (2000)	$LV = 1.0 \cdot CA / CL + 2.5 \cdot OC / L + 2.86 \cdot WC / E + 2.0 \cdot (FA + S_t) /$ $/ A + 3.3 \cdot NI / S$ Level of viability: $LV > 5.01$ - high; $4.16 < LV \leq 5.0$ - average; $4.15 < LV \leq 2.26$ low; $LV \leq 2.25$ very low.
	Small enterprise model (2009)	$Z = 0.0820 \cdot QR + 0.0209 \cdot TD / E + 0.0987 \cdot CA / CL + 0.9915 \cdot S /$ $/TA - 1.253$ If $Z < 0$ financial state is not satisfactory and an enterprise is not in crisis state or has the threat of crisis development. If $Z > 0$ financial state is satisfactory and crisis state is less probable.
	Tereshhenko (2003)	$Z = 1.04 \cdot CA / CL + 0.75 \cdot OC / TC + 0.15 \cdot NR / AA + 0.42 \cdot$ $\cdot OCF / S + 1.8 \cdot OCF / TA - 0.63 \cdot NS / BC - 2.16$ If $Z < -0.55$, then the financial state of an enterprise is not satisfactory; if $-0.55 < Z < 0.55$, then it is impossible to make plausible conclusions about the financial state of the enterprise; if $Z > 0.55$, then the financial state of an enterprise is considered to be satisfactory.

Table A3. MDA models for assessing the risk of insolvency/bankruptcy Belarus, Russia and Ukraine

Country and name of model		Model type and interpretation of results. Assessment of bankruptcy threat level
Estonia	T-model industrial bankruptcy model	$T = 2.44E \cdot /TA + 0.348 \cdot T / TA + 0.306 \cdot TA / CL \cdot$ If $T > 0.37$, the enterprise with a high bankruptcy; if $0.37 < T < 1.22$ the probability of bankruptcy is small; $T > 1.22$ the enterprise's condition is good.
	P-model (Trade)	$P = 0.603 \cdot NS / CA - 0.71 \cdot \log(NS \cdot 100 / WTE) + 0.88 \cdot TL / TA$ P turning point model offers 0.616

	E-model (energetic)	$E = 0.370 \cdot OP / TA - 0.843\log(T) + 0.587 \cdot C_b / NWC$
Latvia	Shorin/ Voronova Z_{2L}	$Z_{2L} = 2.5 \cdot WC / TA + 3.5 \cdot RP / TA + 4.4 \cdot PBT / TA + 0.45 \cdot OC / L +$ $+0.7 \cdot S / TA - 2.4$ If $Z_{2L} > 0$ very low chance of bankruptcy; $Z_{2L} = 0$ critical point; If $Z_{2L} < 0$ - very large probability of bankruptcy
Lithuania	Garškaite´ (2003)	$Z = -0.3877 - 1.0736 \cdot CA / CL + 0.0579 \cdot L / E$ If $Z = 0$ value is 0, then probability of company's bankruptcy equals 50%. If $Z < 0$, then probability of company's bankruptcy is very low. The lower Z value the lower probability for company to go into bankruptcy. If $Z > 0$, then probability of company's bankruptcy is higher than 50% and higher ratio indicates higher probability for company to go into bankruptcy.
	Stoškus et al., 2007	Classification functions: for successful enterprises $Y_0 = 4.77 \cdot NP / NS + 5.88 \cdot TCA / TCL + 9.51 \cdot TD / TL - 5.80 \cdot (CA - I)$ $/CL - 6.42$ For failed enterprises $Y_1 = 2.82 \cdot NP / NS + 2.90 \cdot TCA / TCL +$ $+6.43 \cdot TD / TL - 2.92 \cdot (CA - I) / CL - 2.94$ The enterprise is classified to group, the function of which has a greater value

Table A4. MDA models for assessing the risk of insolvency/bankruptcy Estonia, Latvia and Lithuania

8. References

Abrjutina, M.S.; Grachev, A. V. (1998). *Analiz finansovo-ekonomicheskoy deyatelьnosti predpriyatiya,* Delo i Servis, ISBN 5-8018-0037-9, Moscow, Russia

Alekseeva, J.A. (2011). Oczenka finansofogo sostoyaniya i prognozirovaniya bankrotstva predpriyatiya. Moskva, 31p. Available from: <www.hse.ru/data/2011/03/18/12 11254568/Автореферат_Алексеева.pdf>

Anjum, S. (2010) An overview of financial ration from 1900's till present day. International Journal of Research in Commerce & Management, Vol. No. 1 , Iss. No 8, pp.126-130.

Altman, E. (1968). Financial ratio, discriminant analysis and prediction of corporate bankruptcy. *Journal of Finance*, 23, pp. 589-609.

Altman, E.; Haldeman, R.; Narayanan, P. (1977). ZETA Analysis. A new model to identify bankruptcy risk corporations. *Journal of Banking and Finance. Vol1.* No 1. pp. 29-54.

Altman E.; Resti A.; Sironi A. (2004). Default and Recovery Rates in Credit Risk Modelling: A Review of the Literature and Empirical Evidence, *Economic Notes*, Vol. 33, Iss.3, pp. 183-208.

Altman, E.; Hotchkiss, E. (2006). *Corporate Financial Distress and Bankruptcy.* 3 end. John Wiley and Sons, New York.

Altman, E.; Lavellee, M.Y.(1981). Business failure classification in Canada. *Journal of Business Administration* Summer, pp.147-164.

Altman, E. (2010). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, Vol. 23, No. 4 (Sep., 1968), pp. 589-609.

Andruchak, E.M. (2004). *Diagnostika bankrotstva ukrainckih predprijatiy*, Kiev: Ukraine

Antonowicz, P. (2007). *Metody oceny i prognoza kondycji ekonomiczno-finansowej przedsiębiorstw* Wyd. Oddk, Gdańsk , pp. 1-207, ISBN: 978-83-7426-411-2

Antonowicz, P. (2011). Procesy upadłościowe przedsiębiorstw w Polsce w latach 2009-2010 - 2-letni raport z badań, Wyd. KPF w Polsce, luty, 2011, Gdańsk – Warszawa

Aziz, A.; Lawson, G.H. (1989) Cash Flow Reporting and Financial Distress Models: Testing of Hypothesis. *Financial Management,* 1989, 18(1), pp. 55-63.

Baron, H. F.; Barrett, B. E. (1996). Decision Quality Using Ranked Attribute Weights, *Management Science* 42(11): 1515–1523, DOI: 10.1287/mnsc.42.11.1515

Basel Committee on Banking Supervision (2003), *Overview of the New Basel Capital Accord: Consultative Document,* Basel: Bank for International Settlements.

Bellovary, J.; Giacomino, D.; Akers, M. (2007). A review of Bankruptcy Prediction Studies: 1930 to Present. Marquette. *Journal of Financial Education*, 1-1-2007. Pp.1-47.

Begley, J.; Ming, J.; Watts, S. (1996). Bankruptcy Classification Errors in the 1980s: An Empirical Analysis of Altman's and Ohlson's Models, *Review of Accounting Studies*, 1, pp. 267—284.

Beaver, W. (1967). Financial ratios as predictors of failure. *Journal of Accounting Research, Supplement*, Vol. 5, pp. 71-111.

Boyкova A.V. (2010). Prognozirovanie vozmozhnosti bankrotstva predpriyatiya: podhody I modeli. *Ekonomicheskie i gumanitarnye issledovaniya regionov*, N4, 51, pp. 347-364, Available from: <http://cegr.ru/downloads/journal_4_2010.pdf>

Brancia, A. SMEs risk management: an analysis of existing literature considering the different risk streams. The 8th AGSE International Entrepreneurship Research Exchange, Swinburne University of Technology, Melbourne, Australia. pp. 225-239. ISBN 978-0-9803328-7-2

Budjashkina, S. (2003). *Astrogenetika dlya predprinimateley*. Riga. Centrs "Astrogenetika", Latvia.

Budjashkina, S. (2003). Riska vadīšanas modelēšanas komercdarbībā. RTU 44. Starptautiskā zinātniskā konference. 9.-11.10. Riga, pp. 88.

Dagum, E. B.(2010) Business Cycles and Current Economic Analysis. *Estudios Economa Aplicada,*Vol. 28-3, pp. 577-594.

Davydova, G.V.; Belikov, A. Ju. (1999). Metodika kolichestvennoy oczenki riska bankrotstva. *Upravlenie riskom*. N 3, pp. 13-20.

Doncova, L.V.; Nikiforova, N.A. (2009). *Analiz finansovoy ochetnosti*. Delo i Servis,ISBN 978-5-8018-0340-1, Moscow, Russia.

Dulova, I.N.; Dubrovskij, V.Ž.; Kuzьmin, E.A.(2011). Ochenka finansivogo riska v prognozah denezhnyh potokov mnogoproduktovogo predpriyatiya. Vypusk Chelyabinskogo gosudarstvennogo universiteta. No.6 (221), Ekonomika, Vyp. 31, pp.100-107.

Dumičić, M. Dumičić and R. Curkov (2005). Financial risk management instrument usage in large and medium-sized enterprises – business survey in a transition country. Irving Fisher Committee Central Bank Statistics Bulletin 22, pp. 96-99.

Fishburn, P. (1970). *Utility theory for decision making.* Wiley. ISBN 0471260606, New York. *Finansovoe polozhenie predpriyatiya (oczenka, analiz, planirovanie). (*1998). Pod. red. A.V. Chupisa. Universalnaya kniga. Sumy.

Foster,B.P.; Ward,T.J. (1997) Using Cash Flow Trend to Identify Risk of Bankruptcy. *The CPA Journal*, September 1997. pp. 60-61.

Frazer, J.; Simkins B.J. (2010). *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives.* Kolb Series in Finance, ed. Wiley, Hoboken, New Jersey: John Wiley & Sons. ISBN 9780-470499085

Genriha, I.; Pettere, G.; Voronova, I. (2011). Entrepreneurship Insolvency Risk Management: case Latvia. *International Journal of Banking, Accounting and Finance* (IJ BAAF).

Genriha, I.; Voronova, I. (2011). Methods of evaluating the creditworthiness of borrowers. SCEE'2010. Conference Abstracts proceedings. RTU Publishing House, Riga. pp. 41-42.

Genriha, I.; Voronova, I. (2010). Maksātnespējas noteikšanas modeli Latvijas uzņēmumiem *RTU IEVF Zinātniskie raksti. Ekonomiskie pētījumi uzņēmējdarbībā*. 8. *Sējums*, pp. 38-55, ISBN 1691-0737, Riga, Latvia.

Grachev, A.V. (2002). *Analiz i upravlenie finansovoj ustojchevostьju predpriyatija. Ot buhgalterskogo ucheta k ekonomicheskomu.* Fipress, 2, ISBN 5 -8001-0028-4, Moscow.

Gritcenko, L. A.; Bojarko, I.N.; Gubar, A.A. (2009). Dickriminantna model diagnostiki bankrotstva malikh pidpriiemstv. *Aktualbni problem ekonomiki*, N5 (95), pp.256-262

Grigaravičius, S. (2003). *Įmonių nemokumo diagnostika ir jų pertvarkymo sprendimai,* VDU leidykla, Kaunas

Hamrol, M.; Czajka, B.; Piechocki, M. (2004), Upadłość przedsiębiorstwa - model analizy dyskry - minacyjnej, Przegląd Organizacji, nr 6.

Hawkins, G. (2003). Company Financial Risk (And Why Attorneys Must Understand It). *Banister Financial*, INC. pp. 1-4.

Henschel, T. (2006). Risk management practices in German SMEs: an empirical investigation. *International Journal of Entrepreneurship and Small Business.* Vol. 3, No 5. pp. 584-571.

Henschel, T. (2010). SME's Appetite of Risk Management. Available from: Error! Hyperlink reference not valid..ac.uk/business/rmgic/2010/T_Henschel.pdf

Henschel, T. (2010a). Typology of Risk Management Practices: An Empirical Investigation into German SMEs. *Journal of International Business and Economics Affairs,* pp. 1 -29

Hołda,A. (2006). *Zasada kontynuacji działalności i prognozowanie upadłości w polskich realiach gospodarczych*, Wydawnictwo Akademii Ekonomicznej w Krakowie.

Holton, G. (2004) *Defining risk. Financial Analysts Journal*. Vol. 60(6), pp.19-25. Available from: <http//www.riskexpertise.com/papers/risk.pdf>

Horrigan, J. O. (1968) A short history of financial ratio analysis. *The Accounting Review*, 43(2), pp. 284-294

ISO(2009). ISO 31000:2009 Risk Management – Principles and guidelines, International Organization. Available from: <http//www.iso.org>.

Jackendoff, N. (1962). A Study of Published Industry Financial and Operating Ratios. *Economics and Business Bulletin (Temple University)*, p. 34.

Jones, F. (1987). Current techniques in bankruptcy prediction. Journal of Accounting Literature 6, pp. 131-167.

Kealhofer, S. (2003). Quantifying credit risk I: Default prediction', *Financial Analyst. Journal*, Jan/Feb *2003*, pp. 30-44.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. London, Macmillan Press. Available from: <http://cepa.newschool.edu/het/texts/keynes /gtcont.htm>

Korotayev, A.V.; Tsirel, S.V. (2010). A Spectral Analysis of World GDP Dynamics: Kondratieff Waves, Kuznets Swings, Juglar and Kitchin Cycles in Global Economic Development, and the 2008-2009 Economic Crisis, *Structure and Dynamics*. Vol. 4, Nr.1. pp. 3-57.

Kovalev V.V. (2000). *Finansovyj analiz: Upravlenie kapitalom. Vybor invecticij. Analiz otchetnosti.* Moskva, Finansy i statistika. ISBN 5-279-02043-5

Kralicek, P. (1993). *Základy finančního hospodaření*. Přel. J. Spal., Praha: Linde ISBN 80-85647- 11-7.

Kralicek, P. Quick-Test zur Beurteilung von Unternehmen. Available from: <http://members.aon.at/ai-management/qt_qt.htm>aon.at/ai-management/qt_qt.htm>

Kramin, T.V. (2003). Oczenka finansovoj ustojchivocti firmy (predpriyatiya), *Problemy sovremennoj ekonomiki*, N 3/4 (7/8) Available from:< http://www.m-economy.ru /author.php?nAuthorId=250>

Krioni, A.E. (2009). Risk bankrotstva rossijskih malyh predpriyatij i metody ego predotvrazheniya .*Menedzment v Rossii i za rubezom*, N1, pp. 94-100.

Künnapas, I.(1998). Eesti tööstusettevõte pankrotiohu mudeli koostamine finantssuhtarvude ja diskriminantanalüüsi abili. Bakalaureusetöö, TÜ, 55 p.

Koopman, S.J.; Lucas, A. (2003). Business and Default Cycles for Credit Risk. Tinbergen Institute Discussion Paper TI 2003-062/2, pp. 1-22.

Lace, N.; Sundukova, Z. (2010). Company's Standards for Financial Soundness Indicators. 6th International Scientific Conference, *Business and Management* 2010, Selected papers. Vilnius, doi:10.3846/bm.2010.016

Lennox, C. (1999). Identifying Failing Companies: A Re-evaluation of Logit, Probit and DA Approaches. *Journals of Economic and Business,* 51, pp. 347-364.

Lugovskaya, L. (2010). Predicting default of Russian SMEs on the basis of financial and non-financial variables. *Journal of Financial Services Marketing,* 14, pp. 301-313. Doi:10.1057/fsm.2009.28.

Lukason, O. (2004). Esti energeetinaettevõtete pankrotimudel. TÜ ärirahan duse ja investeeringute õpperool, 62 p.

Lukason, O. (2006). Pankrotistumiste modelleerimine eesti kaubandusettevõtete näitel. Disserdatsiooni autoreferaat magister artiumi, TÜ, 79 p.

Luo, D.; Sun, B.(2010). Study on Financial Issues of Merger and Acquisition. *Orient Academic Forum*. Available from: <htt://www.<htt://www.seiofbluemountain.com/upload/product2010 019/2010cwjrhy06a2.pdf>

Mackavičus, J., Rakšteliené A. (2005). Altmanų taikymas Lietuvos įmonių bankrotui prognozuoti. *Ekonomikas teorija ir praktika.* Nr.1, I.

Martynenko, V.P. (2000). Ekonomiko-statistiskaja modelь opredeleniya veroyatnosti bankrotstva dlya predpriyatij obshhestvennogo pitaniya. *Mehanizm regulirovaniya ekonomiki, економіки prirodokorisnuvannia, економіка nidpriiemstva ma organizatsiia virobnitstva,* Vipusk 2, Sumi, pp.151-155.

Meijerng, E. (2002). A Chronology of interpolation: From Ancient Astronomy to Modern Signal and Image Processing. IEEE, vol. 3, pp.319-342, Available from: <http:// http://www.imagescience.org/meijering/publications/download/pieee2002.pdf>

Merwin, C. L. (1942). *Financing Small Corporation in Fife Manufacturing Industries 1926-36*. NBER. ISBN: 0-870-14130-9

Miller, A.J. (2002). *Subset Selection in Regression*. Chapman &Hall/CRC, ISBN 1-58488-171-2, Boca Raton, London

Minussi, J.; Soopramanien, D.; Worthigton D. (2007). Statistical modelling to predict corporate default for Brazilian companies in the context of Basel II using a new set of financial ratios. *Management Science Working Paper Series,* pp. 1-35.

Moeller, R. R. (2007). *COSO Enterprise Risk Management: understanding the new integrated ERM framework.* Wiley, ISBN: 978-0-471-74115-2

Moses, D.; Liao, S. S. (1987). On developing models for failure prediction. *Journal of Commercial Bank Lending*, Vol. 69, pp. 27-38.

Neumaierová, I.; Neumaier, I. (2005). Index IN05: In *Evropské finanční systémy: Sborník příspěvků z mezinárodní vědecké konference*, Brno, Masarykova univerzita v Brně, pp. 143-146.

Netsymailo K.V. (2009). Metody upravleniya riskami v deyatelьnosti cybjektov malogo predprinimatelьstva. *Vestnik OG,* N9 (103), pp. 46-52. Available from: <http://vestnik. osu.ru/2009_9/9.pdf>

Ohlson, J.A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research,* 8(1), pp. 109-131.

Picard, R.G. (2004). Typology of risk in family media enterprises. *Journal of Media Business Studies*, 2, pp. 71-83.

Potapov, D.K.; Evstafjeva, V.V. (2008). O metodah opredelenija vesovyh koefficientov v zadache ocenki nadeznosti kommerchiskih bankov. *Socialno-ekonomicheskoe polozenie Rossii v novyh geopoliticheskih i finansovo-ekonomicheskih uslovijah:realii i perspektivy razvitija,* Sankt-Petersburg, Institut biznesa i prava, pp. 191–195. Available from: <http://www.ibl.ru/konf 041208 /60.pdf>

Savickaja, G.V. (2009). *Analiz hozaistvennoy dejatelnosti predprijatij*a. Infra-M. ISBN 978-5-160 03428-7, Moscow, Russia.

Scholey, C. (2006). Risk and the Balanced Scorecard: organizations around the globe need to step up their risk management initiatives. The Balanced Scorecard can help. Available from: <http://findarticles.com/p/articles/mi_hb6419/is_4_80/ai_n292 86251/?tag=content;col1>

Sharma, D. S. (2001). The Role of Cash Flow Information in Predicting Corporate Failure; The State of Literature. *Managerial Finance*. Volume 27, No 4, pp.3-26.

Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model, *Journal of Business*, 74, pp.101-124, Available from: < http://ihome.ust.hk/accl /Shumway%20%282001%29.pdf>

Solvency II framework. *Official Journal of the European Union.* Available from: < http://eur-lex.europa.eu/Lex UriServ/LexUriServ.do?uri=OJ:L:2009:335:0001: 155:en:PDF>

Smetanyuk O.A. (2007). Algoritm vizacheniia antikrizovikh zakhodiv na osnovi rezulbtiv diagnostiki finansovogo ctanu shdpriiemstva. *Vesnik Sumskogo derzavnogo universieta*. Seriia Ekonomika, N1, pp. 163-168.

Stoškus, S., Beržinskienė, D.; Virbickaitė, R. (2007). Theoretical and Practical Decisions of Bankruptcy as one of Dynamic Alternatives in Company's Performance, Engineering Economics. No 2 (52). ISSN 1392-2785.

Stunžiené, A.; Boguslauskas, V. (2006). Valuation of Bankruptcy Risk for Lithuanian Companies. *Engineering Economics*, 49(4), pp.29-36., ISSN 1392-2785

Shirokova, G. V. (2006). Strategies of Russian Companies at the Different Stages of Organizational Life Cycle: the Attempt of Empirical Analysis. Discussion Paper

Nr.5(R)–2006. St. Petersburg State University: SPb., Available from: <http: //dspace.gsom.pu.ru/jspui/bitstream/123456789/54/1/1128E%29_2006.pdf>

Šneidere, R. (2009). *Finanšu analīzes metodes uzņēmuma maksātnespējas prognozēšanai*, SIA "Lietišķās informācijas centrs", ISBN 978-9984-826-26-4, Riga, Latvia

Tamari M. (1966). Financial ratios as a means of forecasting bankruptcy. *Management International Review*. Vol. 4, pp. 15-21.

Tereshhenko, O. (2003*).* Diskriminantnya modelь integralьnoj oczenki finansovogo položeniypredpriyatiya. *Ekonomika Ukraini*. №*8* (*501*), pp. *38-44.*

Vaino, M. (1999). Esti jae- ja hulgikaubandusettevõtete pankrotimudeli koostamine finantssuhtarvude ja diskriminantanaüüsi abil. Bakalaureusetöö, TÜ, 73 lk

Vichnjakov, Ja. D.; Kolesov, A.V.; Shemjakin, V. L. (2000). Oczenka i analiz riskov predpriyatij v usloviyah vrazhdebnoj okruzhajushhej sredy. *Menedzment v Rossis i za rubezom,* N3

Voronina, V. M. (2007). Upravlencheskie resheniya v antikrizisnom upravlenii predpriyatiem na osnove diagnostiki chistyh aktivov. *Izvestija Sankt – Peterburgskogo universiteta ekonomiki i finansov,* pp. 9-16.

Voronova, I.; Romancēviča, J. (2005). Uzņēmumu ekonomiskās stabilitātes novērtēšana. 6 *International Scientific Conference. Public Relations: Quality, Benefits and Risks*, Selected papers, Riga, Biznesa augstskola TURĪBA, pp.134-142.

Watson, I. (1996). Financial Distress - The State of the Art in 1996. *International Journal of Business Studies*, Vol. 4, No. 2, pp. 39-65.

Xajdarshina, G.A. (2011). Effektivnaja ocenka riska bankrotstva d sovremennoj praktike finansovogo menedzhmenta na predprijatii. Available from:<http://viperson.ru/ Wind.php?ID=601029>

Zavgren, C.V. (1983). The Prediction of Corporate Failure: The State of the Art, *Journal of Accounting Literature*, Vol. 2, pp.1-37.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, pp. 59-86.

Supply Chain Risk Management in the Electronics Industry

Frank Zwißler and Marco Hermann
Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Stuttgart, Germany

1. Introduction

Current trends and events show that it is more urgent than ever for companies to deal with supply chain management and, in consequence, with the associated supply chain risk management. In recent years, manufacturing processes are characterized by a decreasing vertical range of manufacture. While in 1980 about 60% of products were manufactured in-house, this Figure fell to 43% in 2004, so that the proportion of purchased parts, according to Brossardt (2005), outstrips internal production.

In a globalized world, the situation gets even more complicated as companies are geographically spread all over the world. Following Ziegenbein (2007), the success of a company increasingly depends on its suppliers and its ability to integrate into supply networks.

In addition, companies are faced with a fast moving and unpredictable economic environment. Intensified competition due to globalization, rising customer expectations along with declining customer loyalty, shorter product life cycles, and a supply shortage of raw materials and energy bring companies under growing competitive pressure (Ziegenbein, 2007).

Heightened competition leads to growing cost pressure in the supply chain. Also, the number of supply chain risks has grown in recent years and their effects have become more devastating.

Ziegenbein (2007) takes one of the most prominent and widespread examples in the literature, i.e. the failure of a supplier to deliver, to demonstrate the serious effects that supply chain risks can have on companies. Fire in the production cell of a manufacturer of microchips brought production to stand still for three weeks. At that time, Ericsson had taken measures to optimize its supply chain and cut back on alternative microchip suppliers, resulting in a fall in output for several months and an estimated loss of US $ 400 million. Natural disasters, terrorism, and political as well as economic crises in individual countries are also raising the demands on supply chain risk management (Ziegenbein, 2007).

In view of the above, it becomes evident that supply chain risk management has gained in significance. In recent years, a number of studies have been conducted to analyze its importance and practical application. Overall, it can be said that more and more companies recognize the need for risk management and the need for appropriate measures, but fail to put them into practice in a systematic manner.

A study of BearingPoint Consultants claims that potential disruptions of the supply chain pose a major risk for companies today. More than 80% of the 300 companies surveyed said they were facing significantly more supply risks than five years ago. For almost half of the companies, supply is associated with significant risks. In spite of this, only every second of the surveyed companies had established a risk management system to support supply chain management.

- A survey of 162 purchasing managers in large enterprises published in 2006 by Marco Moder (Moder, 2008) also confirms the growing importance of supply risk management. At the time of the survey, 37% of the participating companies admitted the great significance of supply risk management, compared to 82 % in 2011.
- A 2009 IBM study showed that companies deal with their risks in a proactive and systematic manner. Only 38% said they would manage their processes and services in the supply chain as well as the related risks in a targeted way. The study also revealed that supply chain managers are well aware of the risks associated with their lack of information but are not drawing the necessary conclusions. Only 16% of the interviewees think that they are able to fully use and interpret the information and data they get and to smoothly integrate their external partners into the supply chain. 70% of the study participants say that the main problem in supply chain management is to handle, structure, and interpret large and distributed data sets (Kümmerlein, 2009).
- Although a 2005 study identified supply chain risks as most critical threats to the whole company and despite the fact that the importance of supply chain risk management is recognized by both science and industry, it is still rarely used in practice.

This chapter sets out to define basic terms before presenting the results of a study of 2010 and finally introducing a methodical toolkit for identifying, assessing, and managing risks in a supply chain. The methodical toolkit has been developed to help SMEs with risk management. We proceed by reviewing what the literature has to offer. Then, we introduce the study and the research methodology of the study. After the introduction of the basic conditions the actual research findings are presented and discussed.

2. Literature review

The topic risk management and with special attention on supply chain has been widely discussed in the literature and scientific community. It can be said that the risk management process is a developed methodology for the industry. However, there is a gap between the literature and the actual use of the risk management in companies. Therefore the next section discusses the literature on supply chain management and risk management. Supply Chain Management

The origins of Supply Chain Management (SCM) in its simple form can be traced back to the USA. Here, supply chain management was first implemented in the early 1980s. In 1982, the term 'Supply Chain Management' was introduced by the consultants Oliver and Webber at Booz & Company in London (Christopher, 1992). Even today, a uniform official DIN standard for SCM has not been established, and in the literature many different definitions of SCM can be found. Among the most accurate is that of (Scholz-Reiter and Jakobza,1999):

„Supply Chain Management is the cross-company coordination of material and information flows throughout the value-chain from raw material production via various value-adding stages to the end user, aimed at optimizing the overall process in terms of time and costs." (Scholz-Reiter and Jakobza, 1999).

The Supply Chain Operations Reference Model (SCOR) is most frequently mentioned in the literature as providing the basis of SCM design. The idea behind the SCOR model is to design the communication and production processes based on cross-company standardization so that companies can speak the same „tongue". In Arne Ziegenbein's dissertation on „ Supply Chain Risks", a detailed description of the SCOR model can be found (Ziegenbein, 2007).

2.1 Risk management

Risk management in general has become a widely used and well-established term in recent years. However, it was not before autumn 2009 that a basic standard for risk management processes was published as ISO 31000:2009 „Risk management – Guideline on principles and implementation of risk management". This standard is a fairly general description of a process that identifies, assesses, controls, and monitors risks. A definition of the generic term of risk management could be: Risk management is the totality of measures put in place to identify and assess potential risks and control known risks using appropriate strategies and methods.

2.2 Risk management: Process and methods

In the professional literature, the descriptions of the risk management process differ only slightly and more with regard to the names of the phases than to their content. Figure 1 shows the steps of the risk management process. In the following, the individual phases are described in greater detail, with a particular focus on supply risks.

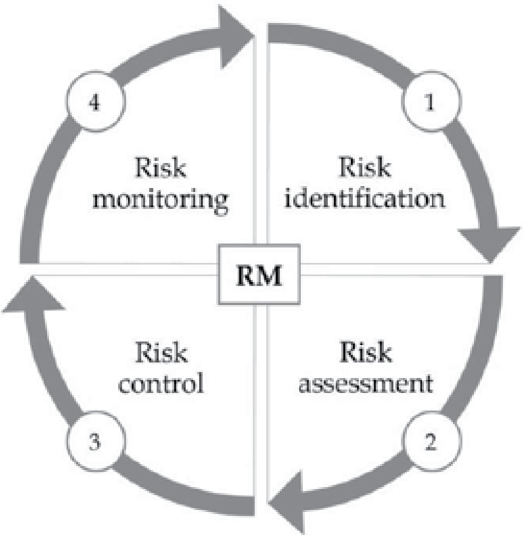


Fig. 1. The risk management process following Ziegenbein (Ziegenbein, 2007)

2.2.1 Risk identification

The starting point of the process is the identification of risks. It is often described as the most important phase, as only an identified risk can become part of the risk management process and be controlled by appropriate measures. Every unidentified risk carries an increased potential for disruptions and, in the worst case, may lead to losses jeopardizing the survival of the business. The first step therefore is to identify potential risks early on and to record them in a systematic way (Wildemann, 2006).

In practice, a number of different methods have been established to serve this purpose, especially risk checking, employee interviews, and the Failure Mode and Effects Analysis (FMEA).

Supply risk identification primarily means the continuous monitoring of suppliers. To this end, supplier performance measures can be used, such as delivery performance in terms of time and quantity, quality of the delivered products, payment performance, but also the monitoring of the financial situation of the supplier. If a major deviation from preset values occurs, this indicates a potential risk situation and calls for a more in-depth analysis. An analysis of the economic, political and geographical environment of the supplier is also recommended to check for related risks such as natural disasters, currency instability, or possible trade embargos.

2.2.2 Risk assessment

Risk identification is followed by risk assessment. In this step, risks are assessed as to their probability of occurrence and the potential severity of impact. The two dimensions can be defined both in qualitative and in quantitative terms.

Qualitative techniques for assessing the probability of occurrence include expert estimates or fault tree analysis, while quantitative techniques are based on the statistical analysis of past data or on simulation models. Qualitative techniques for assessing the severity of impact cover expert estimates or event tree analysis, while quantitative techniques use past data to calculate adverse variances in sales, profit margin or operational costs (Ziegenbein, 2007).

Usually, it is not worth the effort for medium-sized companies to quantitatively determine risks. Quantified values for the severity of impact are virtually never calculated and, besides, not needed for raising risk awareness and drawing attention to specific risk sources.

In practice, most companies confine themselves to a qualitative determination of risks. They use reference values for the severity of impact such as the purchasing volume per supplier or ABC classifications as indicators of the strategic significance of each supplier.

The probability of occurrence as well as the severity of impact is often assessed on a scale from 1 to 5 by the responsible staff or on the basis of past data (Blome and Henke, 2009). Figure 2 shows a classification of supply risks considering the two dimensions of probability of occurrence and severity of impact.

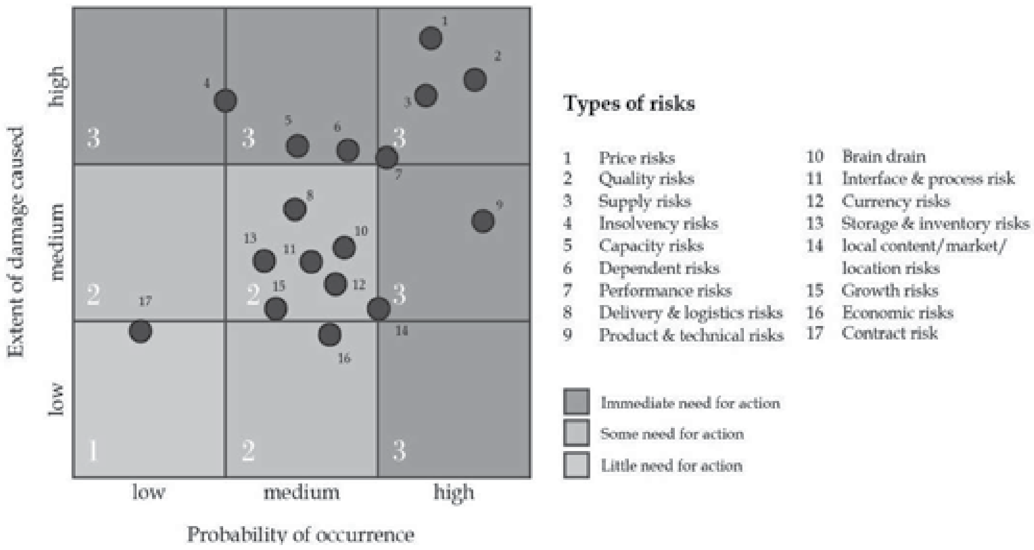


Fig. 2. The risk portfolio (Wildemann, 2005)

For price and quality risks, both probability of occurrence and severity of impact are rated highly. Accordingly, there is immediate need for action to develop measures for counteracting these risks.

2.2.3 Risk control

The next step after risk assessment is risk control. At this stage, specific measures are put in place to manage risks.

Basically, there are two types of measures: Proactive measures that are used to avoid or prevent risks and reactive measures that are concerned with what needs to be done if losses occur (Grundmann, 2008).

- Avoiding an identified risk by taking appropriate measures (Mikus, 2001). This is surely the final goal, even though the cost-benefit relationship should be taken into account. If this is perfectly balanced, measures should and can be implemented as part of supply chain risk management. If a healthy balance cannot be maintained, the management must decide how to further proceed.
- Mitigating an identified risk by taking appropriate measures. If a risk cannot be avoided, at least it should be mitigated (Mikus, 1998).
- Limiting or transferring an identified risk by taking appropriate measures (Lück, 2000). This strategy helps to limit risks and their consequences. With a view to trust and cooperation in a supply chain, this is the worst possible solution for stabilizing a risk. If a risk occurs, and even if it takes place at the supplier's site, your own company could still be indirectly affected. Therefore, this strategy should be avoided, if possible.
- Sharing an identified risk by taking appropriate measures. The effect of the measure should be that the entrepreneurial activities are shared between the supply chain parties (Mikus, 1998). This ensures that the consequences of a risk, if it occurs, affect the individual parties to a much lesser extent.

- Accepting an identified risk by taking appropriate measures. This is the case, if the risk cannot be fully avoided or if one of the above strategies cannot be applied. A company accepts the residual risk deliberately, knowing very well that the residual risk may occur, but assuming that the consequences are justifiable compared to the cost or the expense arising from reducing or even avoiding the risk.

2.2.4 Risk monitoring and documentation

The phase of risk documentation and/or monitoring is regarded as a continuous part of risk management, overlapping the above-mentioned phases. A system for documenting risks enables, above all, an overview of identified risks, lists all measures already taken to counteract risks, and keeps track of the implementation progress. An IT-based tool is recommended for use in risk documentation. Moreover, it is an important aspect of documentation to let a centralized risk management unit monitor and control all risk management steps (Greitemeyer, 2004).

2.3 Supply risk management as part of supply chain management

The great variety of companies involved in a supply chain makes it necessary to implement a specific risk management approach. There are different types of business culture, visions, and objectives clash that along the supply chain. Opportunism or a lack of trust might result in unwillingness to share information, creating an imbalance of information regarding potential risks. Even communication problems caused by a lack of proper channels and interfaces may pose a threat to supply chain processes (Beckmann, 2004).

Companies may also differ in their risk strategy and risk-bearing capacity. And lastly, different national risk management requirements can impede successful collaboration in international supply chains (Siepermann and Vahrenkamp, 2007). Supply chain risk management is defined by Norrman and Lindroth as follows:

„Supply Chain Risk Management is to collaboratively with partners in a supply chain apply risk management process tools to deal with risks and uncertainties caused by, or impacting on, logistics related activities or resources" (Norrman and Lindroth, 2002).

That means supply chain risk management is handled by a chain of companies taking coordinated risk management activities including three elements: 1) Internal risk management, 2) marketing risk management and 3) supply risk management. Internal risk management deals with production-related risks occurring exclusively within one's own company. Marketing risk management focuses on the identification, analysis, control, and monitoring of consumer-related risks. Finally, supply risk management centers on supply-related risks. They can be classified according to internal and external supply risks, with external risks referring to the supplier while internal risks refer to the supply management of a company. Accordingly, (Moder, 2008) defines supply risk as „a possible negative deviation from target on the part of the supplier, on the supply markets or in the internal supply management, adversely affecting the business function of supply management and impacting the supply processes, other internal functions or the customers of the buying company ". For instance, if a supplier is not able to deliver a sufficient quantity of products, this may lead to a shutdown or delay in production for the companies concerned. Within the framework of this study, the term supply chain risk has the same meaning as the term supply risk.

3. Research methodology

Different methods of data collection can be used for an empirical study. First, a distinction must be made between primary and secondary research. While secondary research uses and re-analyzes existing data collected for other research purposes, primary research collects new data directly from the relevant target group. Surveys on supply chain risk management and their secondary data already exist but differ so much in their focus and the questions asked that only a primary data collection was possible. The primary data collection combines several methods, as Figure 3 shows:

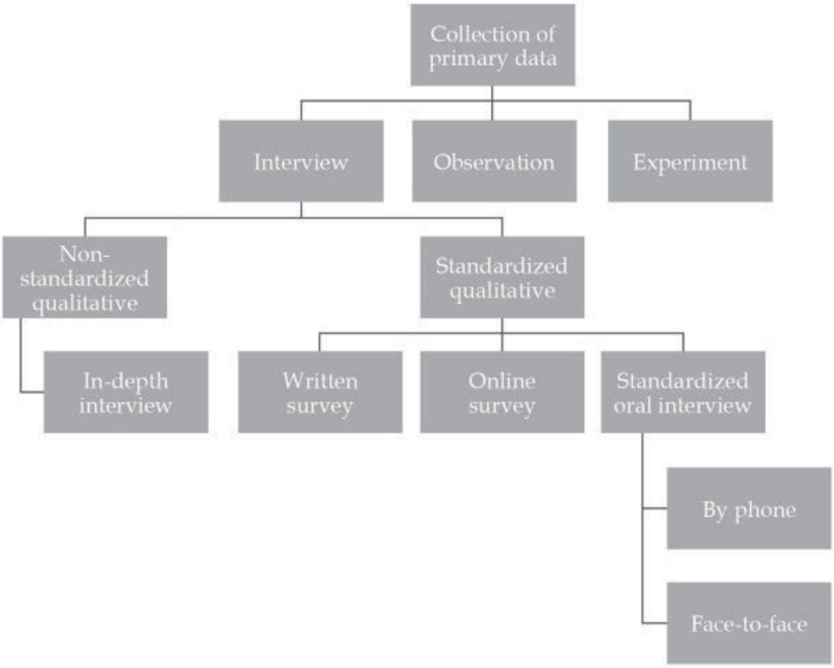


Fig. 3. Different ways to collect primary data (own drawing based on Albers et al., 2007)

A survey can be performed in a standardized or in a non-standardized manner. The non-standardized method aims at qualitative responses. Questions are derived from the context of expected responses or added to existing questions. When the standardized method is applied, the number, formulation and order of questions is precisely defined. A pre-designed questionnaire forms the basis of the interview and is filled in by the participants either in writing or online, or serves as a guideline for an oral interview (Raithel, 2008).

The questions are classified according to the type of suggested responses. Open questions require the respondent to formulate an answer himself, whereas closed questions („multiple-choice questions“) provide possible answers from which the participant must choose either one or several answers (multiple vs. single response). Basically, the advantage of open questions is that it gives the participant greater leeway for his responses. So he/she can include aspects that would not have been touched upon in the predefined categories of the closed survey. However, open questions can also prove problematic, as participants may feel annoyed when compelled to formulate responses of their own, prompting them to give short responses and so possibly withholding important information. Also, the analysis is very difficult and time-consuming, as processing the data in a structured way requires to painstakingly elicit those details that the given answers have in common. This may, if nothing else, distort the evaluation results. By contrast, the advantage of closed questions is that the answers can be more easily compared and that the evaluation is more objective (Raithel, 2008).

With a view to the scope of the planned study and the expected return of a large amount of data, only the standardized survey with closed questions using multiple-choice answers suited best. In certain sections, however, open questions were also acceptable to increase the level of detail and the information content.

When asking the respondents to assess certain statements, a rating scale was used based on verbalized categories.

Example: Please rate the current importance of the following operational risks to your company:				
1= minimal	2= rather low	3= rather high	4= very high	5= don't know

Table 1. Example of a rating scale

Special attention was given to meeting the recommendations of Raithel (2008, p.68) to establish precise, disjunctive (not overlapping), exhaustive and roughly equidistant categories (Raithel, 2008). Scales ranging from 1 to 4 are used throughout the present study, eliminating the option to select a middle category. So, the frequently observed "trend towards the middle" (cf. Berekoven 2009, p.70) is avoided to which participants often resort out of convenience or ignorance, forcing the participants to take a position.

The online survey was chosen as data collection method, enabling the quickest and most uncomplicated survey for the chosen type of questions and ensuring a high response rate.

In the next section we present our study and discuss the outcomes of it.

4. Study on supply chain risk management

Fifty-two companies took part in the study. Since a quarter of the participants belonged to the mechanical engineering industry, this turned out to be the dominant sector. The automotive (supply) industry, the electrical & electronics industry, and the metal (working) industry were also strongly represented in the study at about 13% each. The remaining industries together accounted for roughly a third of the study participants (see Figure 4).

At 48%, medium-sized companies with a workforce of 50 to 500 people are most strongly represented in the study. Even large companies with more than 1000 staff members account for a significant proportion at 33%. The study mainly covers module and system suppliers at 44% and OEMs at 39%, while component suppliers are represented at 15% and raw material suppliers at about 2%.

The study also considered the position in the supply chain, which determines the roles of the companies in the supply chain.

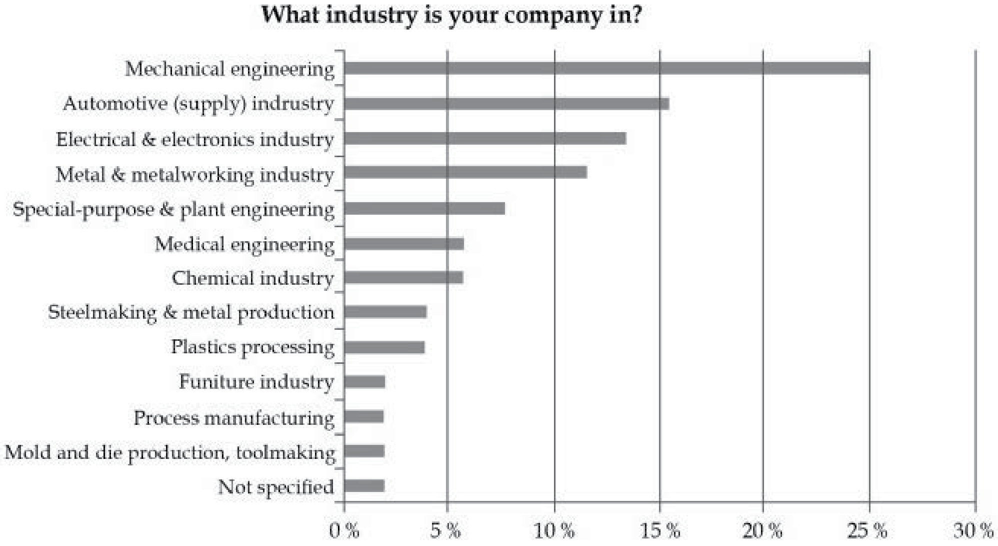


Fig. 4. Quantitative data on participants in the study according to industry

4.1 The importance of supply risk management

About 86% of interviewees indicated that risk management has become more important to their company in recent years. Only one participant said that risk management has not become more important, while 12% found the issue to have hardly gained in importance (cf. Figure 5).

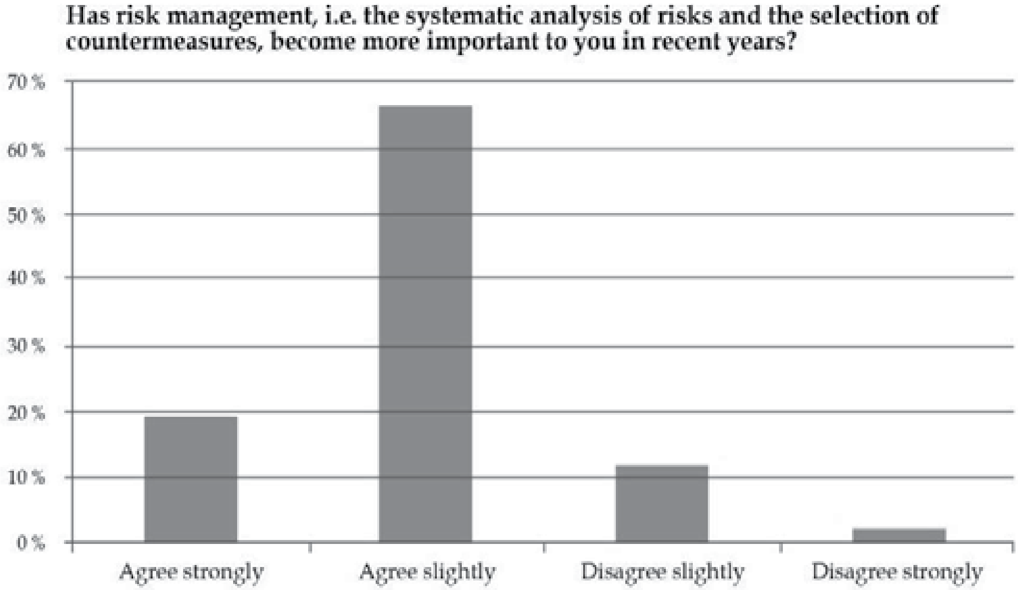


Fig. 5. Importance of risk management (overall)

Especially prominent German sectors, such as the automotive (supply) industry, mechanical engineering and, above all, the electrical and electronics industry, feel the threat of unforeseen events affecting their supply management. News on recurring production losses, e.g. due to a lack of material, do their best to confirm those fears. As this is hardly a new problem for those industries, the values indicating the gain in importance of risk management are beyond average.

Unlike raw material suppliers, component suppliers, and module/system suppliers, OEMs do not see a great increase in importance. The reason for this relatively low rating is that OEMS have to deal with risk management for quite some time, as the number of their suppliers tends to be rather high. In sum, almost all study participants admitted a significant increase in the importance of supply risk management.

The next section will deal with the risks that companies regard as relevant and critical.

4.2 An overview of risks – Their significance today and in the future

For a better analysis, the study divided the risks into operational and strategic risks and asked to provide an assessment of risks now and in three years' time.

Among the operational risks, the risk rated highest both today and in the future is that of late delivery by the suppliers. (cf. Figure 6)

The strong growth since the economic crisis has, for instance, repeatedly brought the production lines at automotive OEMs to a halt because supply bottlenecks occurred at semiconductor manufacturers who had reduced capacity during the crisis. The development of capacity, however, that now lies ahead and the associated investment take time to take effect. And anyway, synchronizing the planning and production strategies of the two supply chain partners is difficult enough.

The lean principles of the automotive industry require a high level of flexibility, which, due to technological restrictions and the resulting batch production, is not easily provided by the semiconductor industry. This risk is especially feared by OEMS due to their high proportion of purchased parts. In general, the vertical range of manufacture and the risk of late deliveries is directly connected, due to insufficient production capacity on the part of the suppliers.

Even in 2010, quality problems still range among the top risks. Unstable manufacturing processes, often in combination with poor quality control, present an overall high risk. A way to successfully avoid risks in globalized supply is to check and select suppliers very carefully to ensure process stability. A functioning quality assurance system including clearly defined responsibilities and standards in the supply chain can also be a lever to reduce cost risks in the supply chain. The sooner a quality problem in the chain is discovered, the less cost are incurred for its remedy. So, it becomes obvious how important an integrated supply risk management approach is. The automotive sector, which sells consumer goods, is at the forefront of public attention. Almost every year an automotive OEM hits the headlines because of the poor quality of some of its components. Toyota's recent recall campaigns, for instance, will cost the company 13 million Euros in Germany alone, not to mention the loss of image. The risk of "unstable manufacturing processes at the supplier" is often the cause of quality problems and therefore rated high.

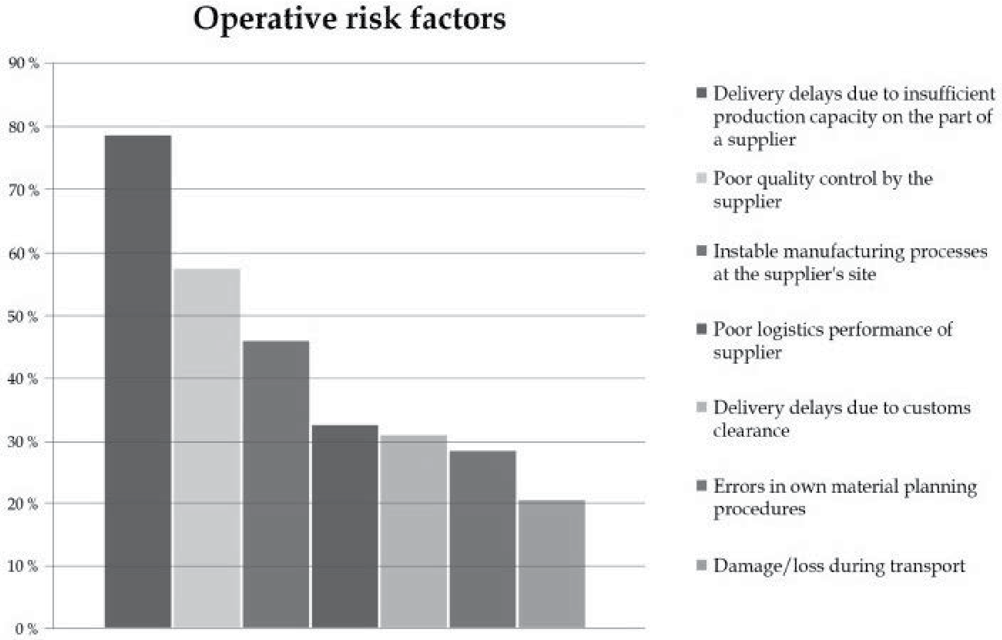


Fig. 6. Operational risk factors (overall)

The dependence brought about by single sourcing was also rated highly, as well as an increase in prices by the suppliers. There are various reasons for single sourcing: A common strategic reason is that companies fear the disclosure of knowledge, inducing them to cooperate with a single supplier only. It is, of course, possible that there were no alternatives available on the market. Other frequent reasons are financial, such as quantity discounts or long-established business relations, for which no alternatives were developed. So, the real question is when and where single sourcing makes sense and how to weigh the associated risks against other sourcing strategies? This is what makes risk management a key part of the supply strategy.

The assessment of strategic risk factors is influenced by the economic crisis and the related currency risks as well as by the rising raw material and energy prices. Accordingly, risks in the business environment received a high rating by 81% of the companies questioned (cf. Figure 7)

Similar to the 2008 Moder study, in this study the risk of force majeure events such as catastrophes and war/terrorism received a low rating. Even events such as the volcanic eruption in Iceland in May 2010 did not affect the assessment of risks. The evaluation of risks in the economic environment, however, is different from past investigations. While currency or competitiveness are regarded by the Moder study as relevant risks, these are ranked top risks in this study. This certainly owes to the economic and financial crisis, and to currency fluctuations, business insolvencies and cutbacks in capacity associated with it. In spite of this, companies do not think they are the most important risks in the future. This demonstrates the companies' positive outlook on the future (Schatz and Hermann, 2010).

Next, we proceed by discussing the methodological expertise and the maturity level of risk management process in the companies.

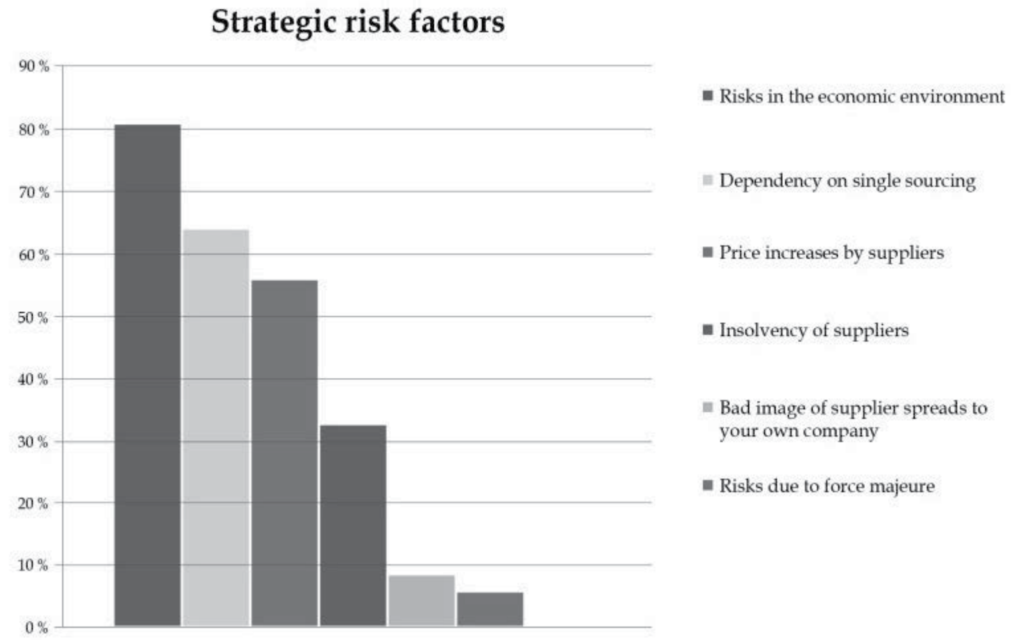


Fig. 7. Strategic risk factors (overall)

4.3 Methodological expertise, process maturity and organization of risk management

Looking at the methods in use, the companies apparently are searching for techniques and solutions to systematically identify, assess and control risks, but hardly any company has implemented suitable systems and processes. It is often claimed that risks are known, but they are not methodically tracked down or even ignored, for companies do not know what action strategies to apply. Another issue is the conflict between accuracy and the invested time and effort, as well as the desire to employ more systematic methods and not to rely on "gut-based" decisions (Ziegenbein, 2007).

Therefore, more and more companies regard risk management as important but they lack the methodical support for the implementation. So, the need for practicable approaches is immense (Siepermann and Vahrenkamp, 2008). Compared to past surveys, the following sections will show how the methodological expertise has improved and what strategies are used by the companies. Another objective is to determine the maturity level of the risk management process in industry.

4.3.1 Maturity level of the risk management process

Compared to past surveys, process maturity has greatly improved. Of the companies surveyed, 76% agreed slightly or strongly, when asked if the sequence of process steps is defined and the responsibilities are clearly outlined (cf. Figure 8).

Particularly interesting here is the survey's suggestion that a decreasing range of vertical manufacture goes along with a quality decrease in the risk management process. At 70%, the rating of OEMs is much below that of parts suppliers at about 85% (see Figure 9).

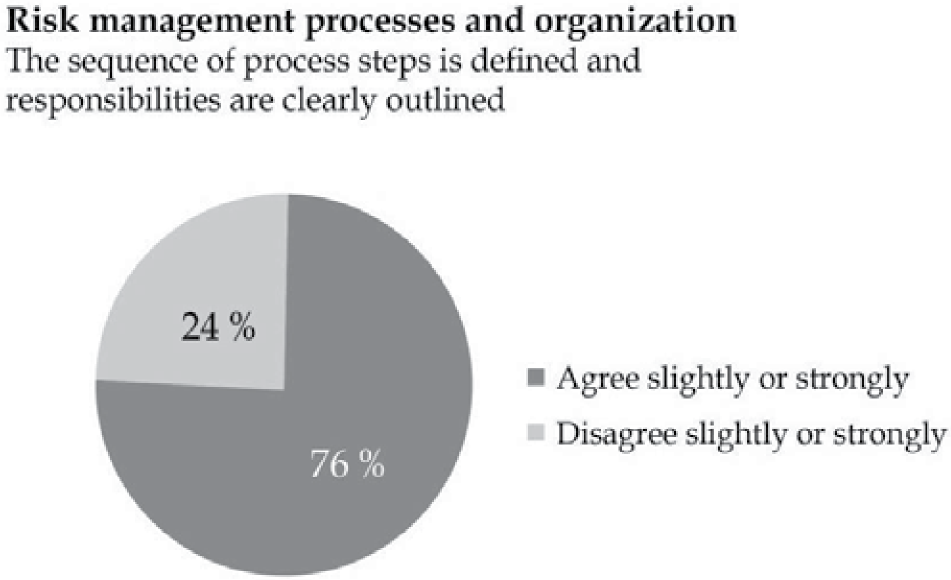


Fig. 8. Process steps of risk management (overall)

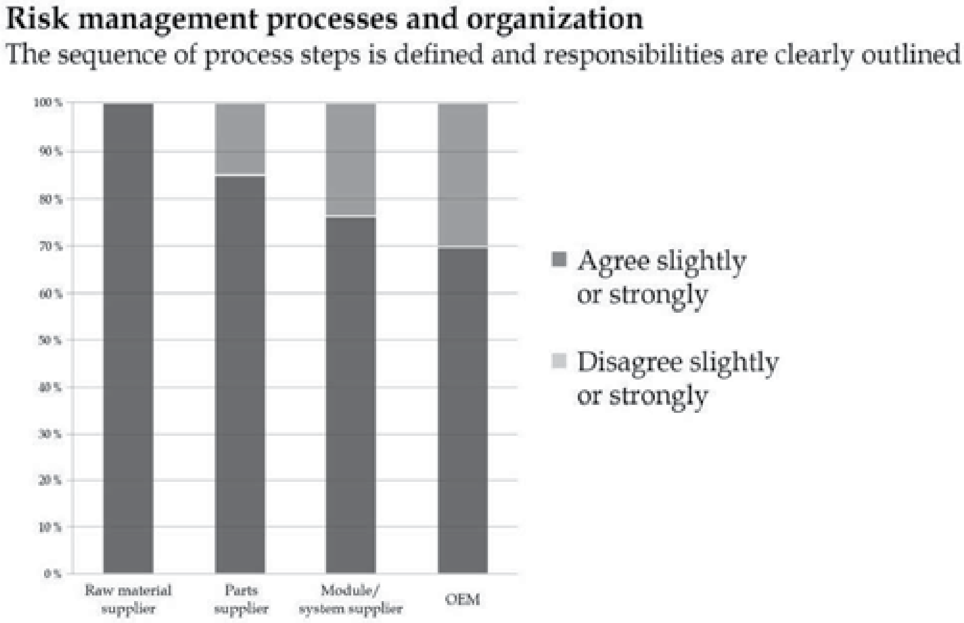


Fig. 9. Process steps of risk management – the role of the supply chain

The survey shows that the closer a company is to the end user, the worse the rating of the process quality gets. However, a defined risk management process is not enough if the necessary resources or methodical expertise are missing.

4.3.2 Methodological expertise

The methodological expertise of companies receives lower ratings than the maturity level of the process. Only 70% of the companies surveyed said that the methods for identifying and assessing supply risks are efficient, easy-to-understand and descriptive (cf. Figure 10).

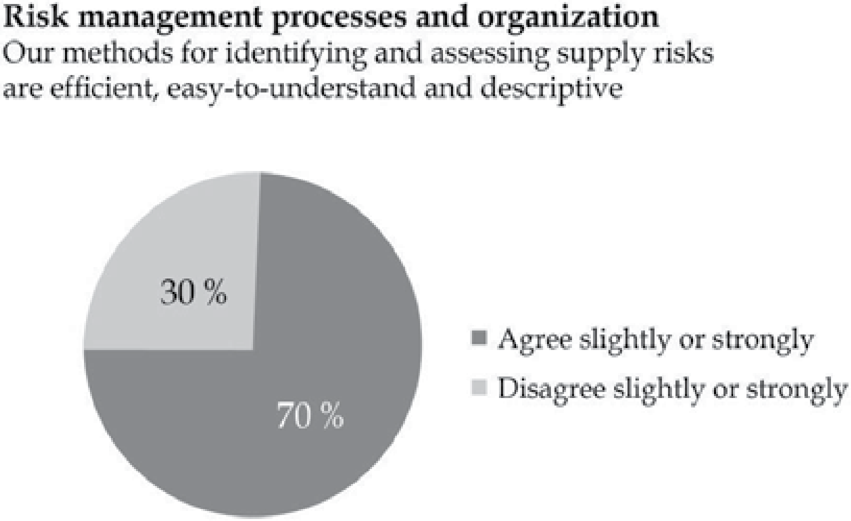


Fig. 10. Methodological expertise in risk management (overall)

While companies from the electrical and electronics industry found their process quality to be rather poor, they rated their methodological expertise as high. This contrasts with the situation in the metal and metal working industry. While here the process quality receives high ratings, weaknesses are recognized in the methodological expertise. In mechanical engineering, weaknesses are discovered in both areas. 55% of the interviewees indicate that the methods used are not very or not at all efficient, easy-to-understand and descriptive.

4.3.3 Quality of the tools

The weakest points in the risk management process are the tools available to the companies. Almost half of the companies surveyed indicate that the tools are not very or not at all mature (cf. Figure 11).

and the automotive (supply) industry rate their tools as being mature (cf. Figure 8).

In special-purpose and plant engineering, mechanical engineering and the electrical and electronics industry, the respondents again indicate that there is potential for improvement.

It is also surprising that in all areas the ratings of the OEMs are lower than those of the raw material suppliers, parts suppliers or module & system suppliers. This also applies to the assessment of the tools (cf. Figure 11).

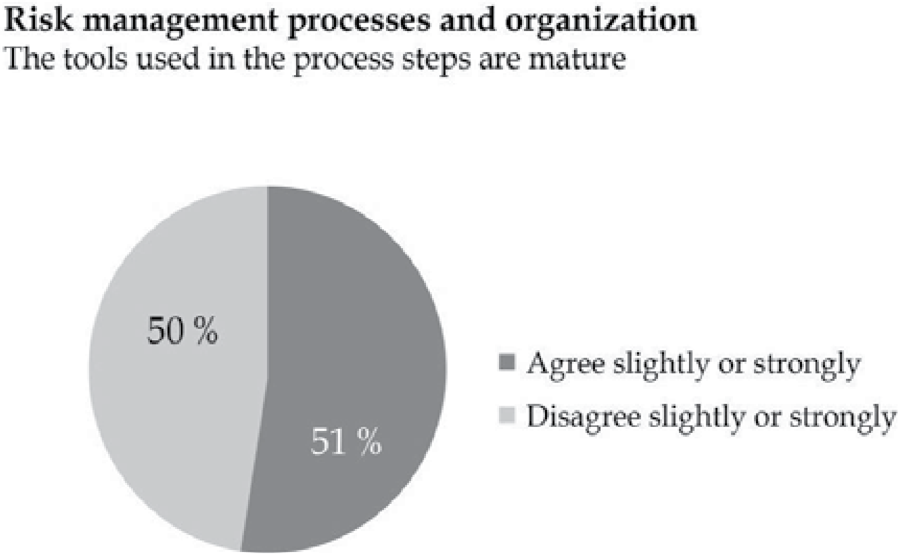


Fig. 11. The tools used in the risk management process (overall)

Only the steelmaking and the metal production sector, the process manufacturing sector,

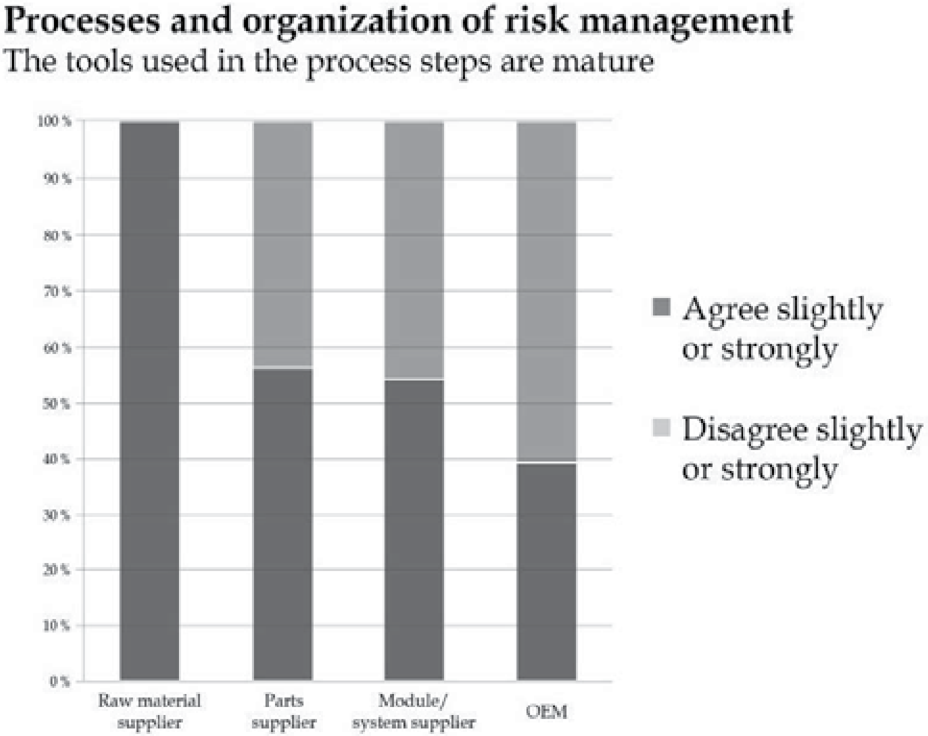


Fig. 12. The tools used in risk management – the role of the supply chain

4.3.4 System integration

Companies often use internally developed tools for IT support, which are not integrated with the system architecture. Accordingly, 48% of the tools used in the supply risk management process are internally developed. About a quarter uses ERP systems, though in most cases they can only be efficiently used after extensive customization. So, with the number of suppliers to be scheduled and controlled growing, the companies are increasingly turning away from using ERP systems for their operations. Commercial custom software or databases are rarely used, and a quarter of the companies surveyed do not use any software at all.

4.3.5 Summary

A rapid increase of the maturity level of the risk management process has taken place. In the 2006 study of Horst Wildemann, only 36% of companies indicated that they systematically capture supply risks, while in 2010, 76% of the participating companies said that the risk management process is mature.

Opportunities may exist in the methodological expertise, in the identification and assessment of risks, and especially regarding the available tools.

An analysis of the results of the separate industries shows that only the automotive (supply) sector receives a consistently high rating in all categories. All other sectors revealed weaknesses, suggesting considerable room for improvement. The outcome of the study should be taken as a warning by the special-purpose and plant engineering sector, the mechanical engineering and the electrical & electronics industry.

All in all, the findings of the study show that companies recognize the importance of supply chain risk management and that the corresponding responsibilities have been largely defined. Weaknesses can be found in the methodological expertise, in the tools used in the process, and in IT support.

These weaknesses were already at the focus of the research project „STABLE", which was initiated in 2008 by Fraunhofer IPA and BMWi, and funded by the AiF (German Federation of Industrial Research Associations). Together with Fraunhofer IML and three partners from the electronics industry, they developed an approach and a methodical toolkit for each phase of the risk management process. The objective of the project was to make methods and tools available to SME's to identify, assess and control supply chain risks. More information on this project can be found on the website of Fraunhofer IPA.[1] The results of the study show that there is still a lack of knowledge and use of risk management in the companies. Especially in SMEs, there is a lack of use of the mentioned process due to a shortage of human or financial resources or just due to a lack of interest. Therefore in the next section we introduce our developed methodical toolkit for SMEs that should reduce respectively overcome the mentioned barriers to implement a risk management process in companies.

[1] SMEs are small and medium sized businesses. According to the definition of the European Union SMEs are defined with an headcount ≤ 250, a turnover ≤ 50 million € or a balance sheet total ≤ 43 million € (European Commission, 2005)

5. A method for risk stabilization in SMEs of the electronics industry

The study highlights that there is an immense need for action in industry to implement supply chain risk management systems. It also shows how much the sectors differ in their management of supply chain risks. Not only that supply chain risk management is not evenly applied throughout the different sectors, there are also great differences in the use and the implementation of supply chain risk management. A particularly striking example is the electronics industry. It particularly reveals the gap between the necessity of supply chain risk management and the actual use of supply chain risk management systems. Moreover, the study demonstrates that, to a certain extent, companies doubt the comprehensibility and the descriptiveness of the processes and structures for supply chain risk management. The maturity level of the processes is seen as insufficient by 49%. For this reason, we have developed the following procedure to overcome the weaknesses in corporate supply chain risk management.

5.1 The particularities of the electronics industry

The following section will undertake a closer examination of the electronics industry and its supplier networks: This requires taking a closer look at its particular features to recognize the need for supply chain risk management in the electronics industry. Most supply chains in the electronics industry are global networks consisting of a single OEM, an "A-supplier", and several small and medium supply companies (SME). These networks are characterized, firstly, by the dominance of the OEM or the A-supplier and, secondly, by the volatile electronics market and its strong fluctuations in demand, short product life cycles, and tremendous potential for technical innovation (Kersten *et al.*, 2008).

It is the combination of environmental turbulence and the focus on ever higher efficiency & productivity of individual companies that makes things difficult for the SMEs in the supply chain. A vital issue is the ability to master existing or occurring risks within the supply chain.

This is why the stabilization of critical supply chains has gained in importance. Risks such as the loss of key suppliers, avoidable inaccuracies in the demand forecast, or unforeseen raw material shortages may have disastrous consequences for the supply chain (Zwißler and Hermann, 2010).

The above-mentioned environmental parameters and the results of the study reveal a pressing need for action in the electronics industry to implement supply chain risk management processes.

To satisfy this need, the Institute of Industrial Manufacturing and Management (IFF) and the Fraunhofer Institute for Manufacturing Engineering and Automation (IPA) together with the Fraunhofer IML from Dortmund carried out the research project STABLE. The project was funded by the BMWi and supported by the AiF.

5.2 A process model for stabilizing SME supply chains in the electronics industry

Since the process of risk management is not implemented in many companies as shown in the study, the next section introduces the process model for a risk stabilization in SMEs in the electronics industry.

5.2.1 Basics

The process model developed in the research project enables SMEs in the electronics industry to identify major risks in their supply chains and to produce measures to stabilize them. The separate phases are outlined in Figure 13. It shows that the process model follows the normal risk management cycle described in section 1.3.

1. The process model covers four phases. The first one is the risk identification phase. Here, the critical supply chain sections are identified as well as the exact location where the risk occurs. It includes an examination of all supply chain sections, both on the supply and on the distribution side.
2. In the second phase, the identified risks are assessed to enable a ranking of the risks. The overall goal of phase I is to identify all risks affecting the relevant supply chain sections mentioned by the employees. The given variety of risks must be prioritized to enable informed decisions on what measures to take. Defining measures for all identified risks would not make sense. Better is to manage critical risks first before, step-by-step, including less critical risks in the process.
3. The third Phase deals with the development of measures. The aim is to select and apply targeted measures for those risks ranked in phase II and so to enable risk avoidance, mitigation, limitation, sharing and transfer.
4. Phase four is very important for the lasting implementation of the first three phases. The entire supply chain risk management cycle should not be a one-off event. It is rather a continuous process, which must be put into practice in everyday work and thus needs to be firmly established in operations and organization.

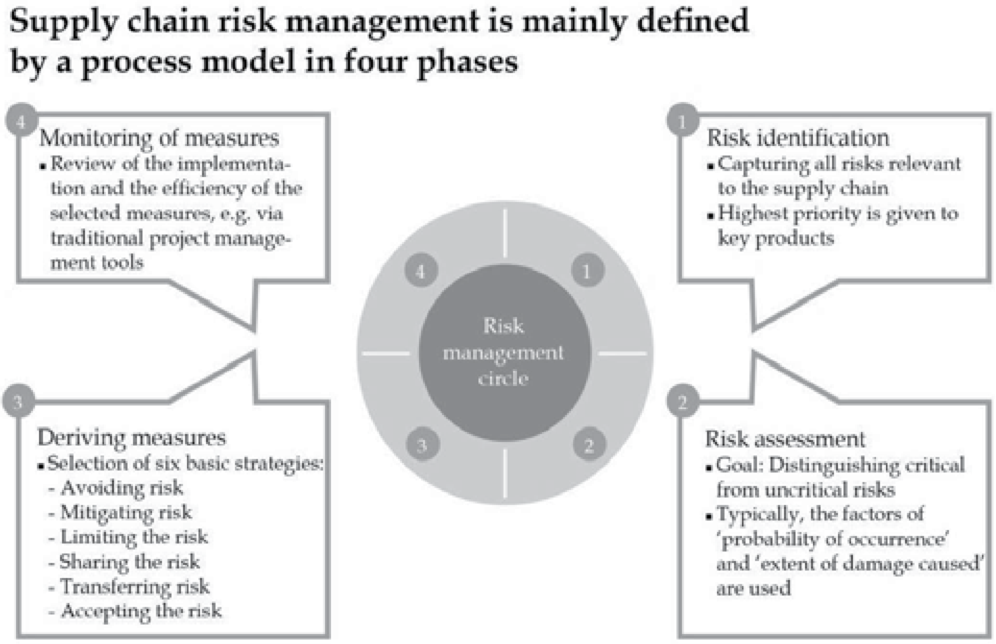

Supply chain risk management is mainly defined by a process model in four phases
4
Monitoring of measures
• Review of the implementation and the efficiency of the selected measures, e.g. via traditional project management tools
1
Risk identification
• Capturing all risks relevant to the supply chain
• Highest priority is given to key products
4
1
Risk management circle
3
2
3
Deriving measures
• Selection of six basic strategies:
- Avoiding risk
- Mitigating risk
- Limiting the risk
- Sharing the risk
- Transferring risk
- Accepting the risk
2
Risk assessment
• Goal: Distinguishing critical from uncritical risks
• Typically, the factors of 'probability of occurrence' and 'extent of damage caused' are used

Fig. 13. Supply chain risk management cycle

To find a suitable approach for the electronics industry, the literature was searched for methods and procedures applicable to the four phases of supply chain risk management. As Figure 14 shows, the 100 methods initially examined for their suitability for a future supply chain risk management approach were reduced to ten methods left for validation and finally to five methods that could be applied in the process model.

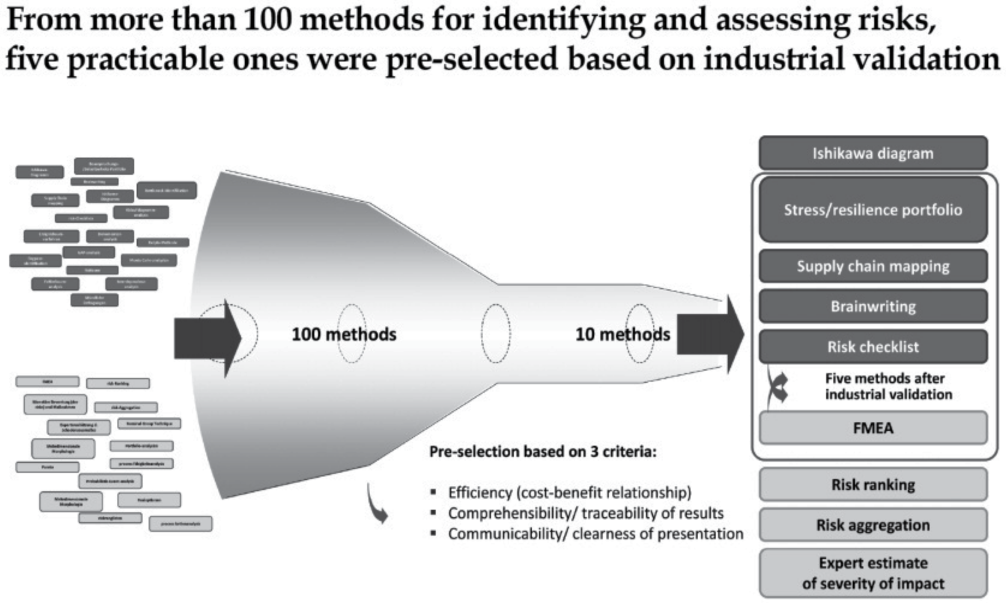

From more than 100 methods for identifying and assessing risks, five practicable ones were pre-selected based on industrial validation
100 methods
10 methods
Ishikawa diagram
Stress/resilience portfolio
Supply chain mapping
Brainwriting
Risk checklist
Five methods after industrial validation
FMEA
Risk ranking
Risk aggregation
Expert estimate of severity of impact
Pre-selection based on 3 criteria:
Efficiency (cost-benefit relationship)
Comprehensibility/ traceability of results
Communicability/ clearness of presentation

Fig. 14. Process of reducing the number of methods for supply chain risk management

5.2.2 Phase I: Risk identification

The first step presents methods for identifying critical supply chains that could support a proactive stabilization in subsequent steps. It should be noted that many methods found in the literature are not designed for use in the electronics industry. Accordingly, the following methods were selected and/or customized for use in the electronics industry. These methods were validated in industry and proved to be especially suitable for the identification of supply chain risks.

5.2.2.1 Supply Chain Mapping

The major advantage of Supply Chain Mapping is its particular time and resource efficiency. Supply chain maps give the user an overview of the key partners, processes and relations within the supply chain. The first step in drawing a supply chain map is to record all suppliers and customers on the map relevant to the supply chain and their locations (Kaufmann and Germer, 2001). Then, the material, information, and financial flows between the supply chain partners are illustrated on the map. Like a road map, the goal is to chart only the "motorways" and major "highways" (cf. Figure 15).

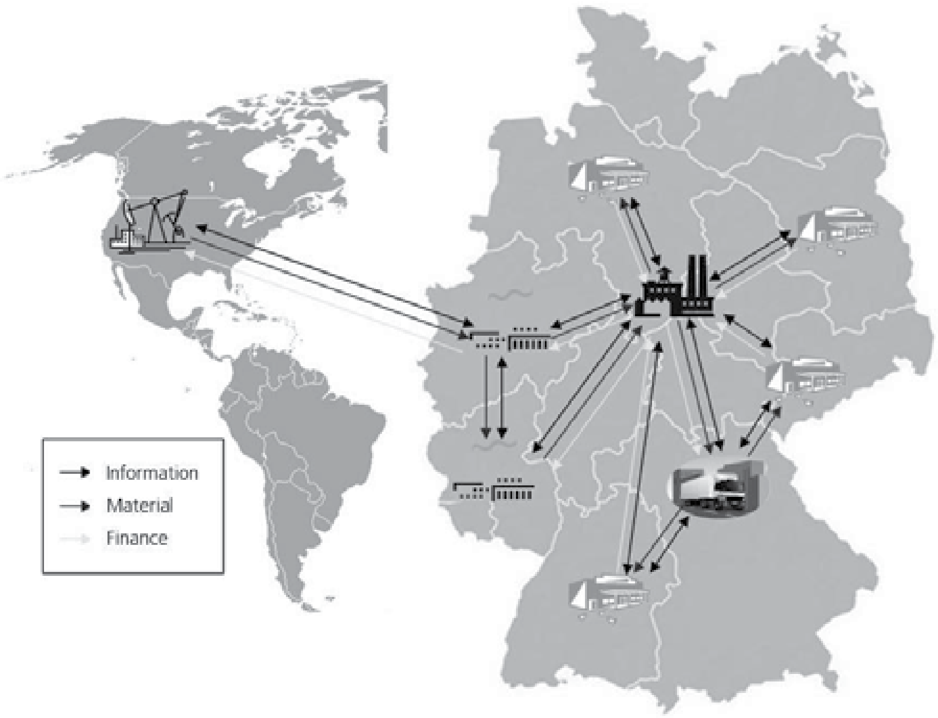


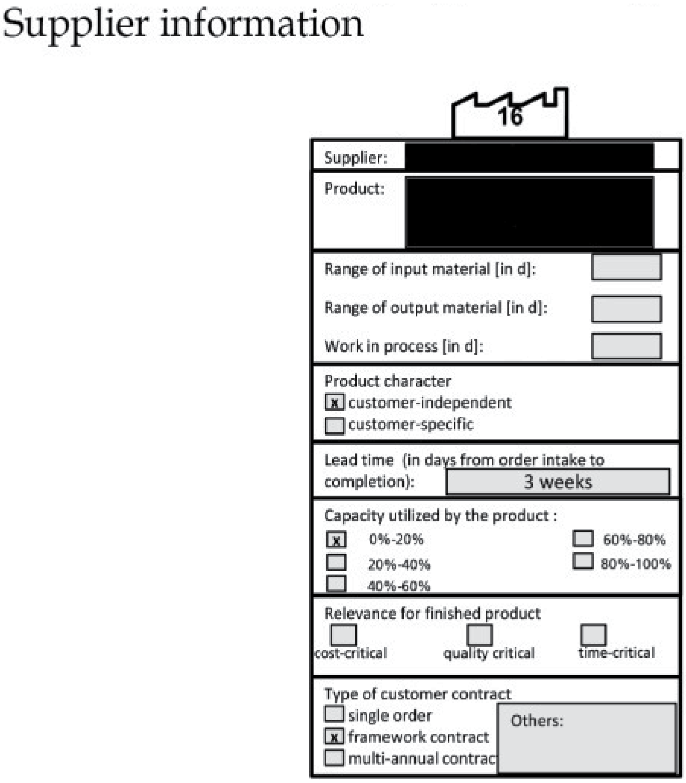
Fig. 15. Example of a supply chain map

Further details on individual locations and flows can also be included. For example, information on lead times, delivery performance, schedule performance, transportation times, capacity, stock levels, the logistics service provider involved, and the relevance of a supplied product for the end product. It should be defined in advance what key factors impacting the supply chain map are to be entered (cf. Figure 16). In view of the amount of information collated, it is necessary to take care that the level of detail is sufficient to make it a useful supply chain map; however, the map should not be too complex in presentation to ensure transparency and traceability.

5.2.2.2 Brainwriting method

Brainwriting is an intuitive group creativity technique for putting down ideas in writing or generating additional ideas based on the combination of ideas (risks) created by several participants.

This technique allows each participant, in a few minutes, to write down on index cards every single risk associated with the identified supply chain section. They are then discussed by the group (Rückle and Behn, 2007). As part of risk identification, this method uncovers by far the largest number of risks in the corresponding supply chain and the examined supply chain sections.



Supplier information

16

Supplier: [redacted]

Product: [redacted]

Range of input material [in d]: □

Range of output material [in d]: □

Work in process [in d]: □

Product character
☒ customer-independent
□ customer-specific

Lead time (in days from order intake to completion): 3 weeks

Capacity utilized by the product :
☒ 0%-20%
□ 20%-40%
□ 40%-60%
□ 60%-80%
□ 80%-100%

Relevance for finished product
□ cost-critical
□ quality critical
□ time-critical

Type of customer contract
□ single order
☒ framework contract
□ multi-annual contrac

Others:

Fig. 16. Information on the supply chain sections to be retrieved in the Supply Chain Map

5.2.2.3 Stress, resilience and expense portfolio

Based on the results of the supply chain mapping process, a stress and resilience portfolio was drawn up to identify the critical supply chain sections (Kaufmann and Germer, 2001). This method is aimed at defining the position of each supply chain section in the portfolio in order to find out what sections are particularly vulnerable, depending on stress and resilience. The analysis of the separate supply chain sections regarding the two dimensions of „stress“ and „resilience“ resorts to questions e.g. about the dynamics or robustness of the section. Then, each dimension is rated on a scale of 1 to 5 and so allows allocating the sections to the portfolio. This is the starting point for identifying and assessing the risks of the critical sections and for implementing measures to stabilize them.

A major drawback of the stress and resilience portfolio is that currently applied measures are only included in the assessment in terms of quantity but not with regard to their quality. As a consequence, measures with poor cost/benefit ratio could possibly be applied to increase the resilience of a supply chain. This makes the SRP a suitable tool to identify where action is needed but not to reveal potential cost savings in the supply chain. To overcome this drawback, the dimension of expense is added to the SRP procedure.

Expense is the general indicator for current risk control activities in each analyzed supply chain section. The higher the numerical value of the expense, the higher the investment in measures for risk control with a sub-optimal cost/benefit outcome.

In capturing the expense, it is particularly challenging to identify the hidden potential for improvement in currently used measures. Taking into account the expense invested, inefficient processes (e.g. three employees ordered to track the shipments of suppliers) are now identified, pointing out opportunities for cost savings (a single employee would be sufficient if the process were redesigned).

With the factor of expense added to the SREP , the user companies get a tool enabling them to analyze and assess the as-is situation (Zwißler and Hermann, 2010).

As a result of the SREP, the following three basic strategies can be presented:

1. In an isolated case of stress increase, no (substitution) investments need to be made, as the expense can be reduced or dropped if stress normalizes (e.g. cut down on temporary workers and floaters, etc.).
2. If stress increase is permanent, a lasting solution at a lower expense must be found. In terms of staff capacity, this might mean an increase in productivity (training schemes, staff qualification).
3. If stress increases periodically, the interval becomes relevant. For short intervals, a permanent solution must be found, while for long intervals a short-term increase of expense is usually more favorable. A frequent strategy for this scenario is changing the shift schedule to increase staff flexibility.

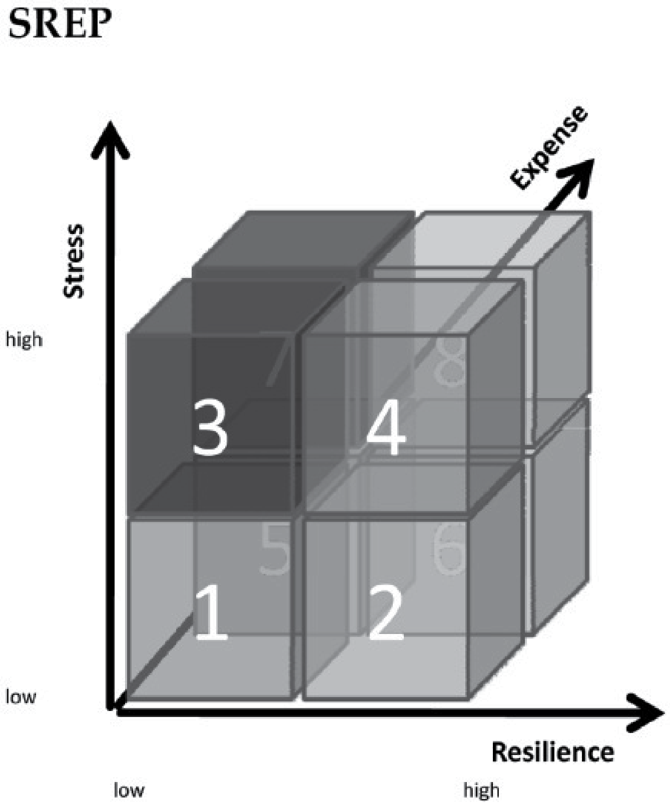


Fig. 17. SREP – The cube of dimensions

The SREP (cf. Figure 17) helps companies to uncover "hidden" opportunities faster and more efficiently, and quickly locate the highest risk supply chain sections, while enabling them to develop immediate strategies from standard strategies for a proactive stabilization.

5.2.2.4 Risk checklist

The risk checklist helps companies to identify existing risks. In sum, the checklist covers five risk categories: quality risk, risk of delay, risk of failure, cost risk, and planning risk. The checklist has been developed during the research project:

An example is given in the following for the category of „quality risks":

Q1) Damage to the material and/or to the end products

a. at your suppliers
b. during transport to your company
c. in the incoming goods department
d. during the manufacturing process in your company
e. during your subcontractors' manufacturing process
f. in the outgoing goods department
g. during transport to your customers

Q2) Insufficient quality control

Q3) Change of product specifications (without consultation)

Q4) Manufacturing fault (e.g. machine/tool failure, incorrect use, etc.)

Q5) Quality defects due to climate (e.g. dry, damp, cold, or warm climate)

The risk checklist facilitates the identification of certain risk factors and ensures that the process of risk identification covers all risks. The risk checklist brings a more objective method of identification to the complete sub-process of risk identification. The following sequence of steps represents the best and most efficient procedure for the first phase of risk identification with regard to the number and quality of the identified risks (cf. Figure 18).

At first, the supply chain should be mapped to get an overview of information, money and material flows in the analyzed supply chain. It is important not to adopt a too high level of detail in the process of supply chain mapping, as this would be of no avail to the next steps. The brainwriting method is used to uncover additional risks in the analyzed supply chain sections. During validation at the industry partners, this creativity technique turned out to be most effective and profitable, helping to identify by far the most risks. In the next step, the stress, resilience and expense portfolio is applied. It provides detailed information on critical sections in the supply chain, helping to identify those risks with crucial impact on the performance of the companies and the supply chain.

Finally, we recommend using the risk checklist to identify further risks not yet uncovered by the previous methods.

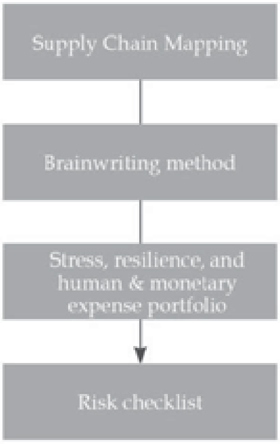


Fig. 18. The steps of phase I: Risk identification

5.2.3 Phase II: Risk assessment

In the second phase of the process model, the identified risks are assessed. The Failure Mode and Effects Analysis technique has proved to be an ideal risk assessment method for evaluating and ranking the identified risks of phase I.

Failure Mode and Effects Analysis was chosen as the basic risk assessment method. To serve this purpose, it was modified and fully adapted to the requirements of supply chain risk management for SMEs in the electronics industry (Figure 19). The identified risks were ranked for an FMEA assessment. The ranking helps to prioritize the risks for the future development of stabilization measures.

For a better understanding, an example is used to illustrate the FMEA process. The risk lies in insufficient quality control measures at a company supplying a manufacturer of construction equipment sensors.

In the first step, the risks are compared to the logistical goals for risks posing a threat to certain logistical goals. In our case, the poor quality control at the supplier threatens the logistical goal of „right quality“.

In the next step, the assessment refers to the parameters of "probability of occurrence" and "probability of detection". First, the risks are assessed for their probability of occurrence in a given period. The assessment is based on a numerical scale from 1 (very rare) to 5 (very frequent). In our example, the possible occurrence of insufficient quality control at the supplier' site is rare and accordingly is ranked a "2". Second, the probability of detection after the risk has occurred is assessed. In our case, where incoming goods are inspected at the sensor manufacturer, it is very high and therefore ranked a "1".

In the following step, the exact risk location, i.e. the place where the risk occurs, is determined. In our case, this place is easily identified at the supplier's site. However, some risks are not so easy to pin down, being composed of a number of consecutive errors, with the real cause of the risk lying somewhere else. But one thing always holds true: if you want a lasting stabilization of risks, treat the cause of the risk.

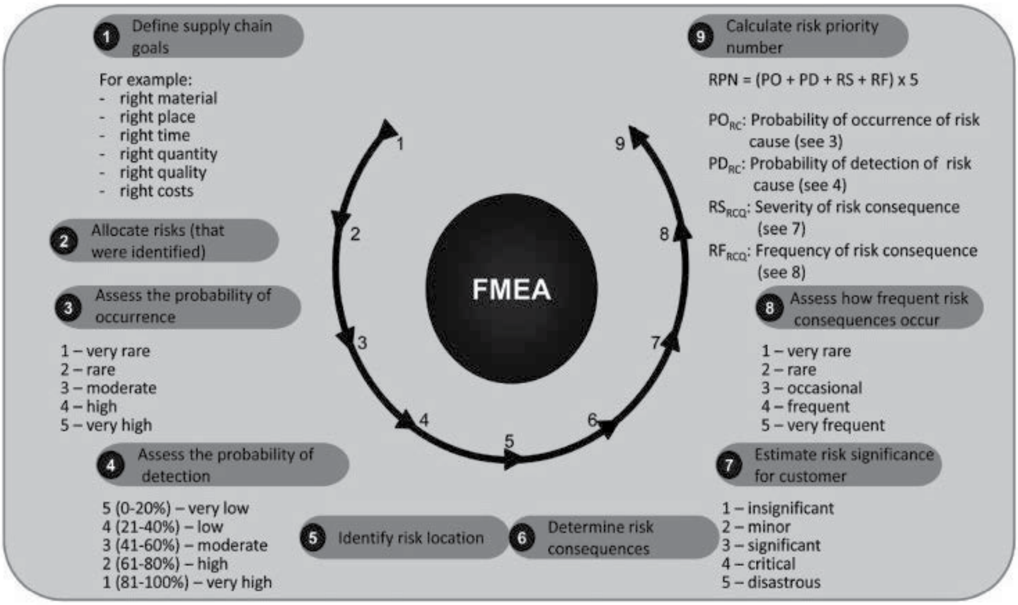

1 Define supply chain goals
For example:
- right material
- right place
- right time
- right quantity
- right quality
- right costs
2 Allocate risks (that were identified)
3 Assess the probability of occurrence
1 – very rare
2 – rare
3 – moderate
4 – high
5 – very high
4 Assess the probability of detection
5 (0-20%) – very low
4 (21-40%) – low
3 (41-60%) – moderate
2 (61-80%) – high
1 (81-100%) – very high
5 Identify risk location
6 Determine risk consequences
FMEA
1
2
3
4
5
6
7
8
9
9 Calculate risk priority number
RPN = (PO + PD + RS + RF) x 5
PO$_{RC}$: Probability of occurrence of risk cause (see 3)
PD$_{RC}$: Probability of detection of risk cause (see 4)
RS$_{RCQ}$: Severity of risk consequence (see 7)
RF$_{RCQ}$: Frequency of risk consequence (see 8)
8 Assess how frequent risk consequences occur
1 – very rare
2 – rare
3 – occasional
4 – frequent
5 – very frequent
7 Estimate risk significance for customer
1 – insignificant
2 – minor
3 – significant
4 – critical
5 – disastrous

Fig. 19. The steps of the FMEA

Then, the consequences of the risks are determined. Here, it is important to remember that a risk can give rise to more than a single risk consequence. In our case, the risk consequence could imply a loss of production, but just as well it could only mean internal rework or a return of goods without causing production losses.

Next follows an estimate of the severity of risk consequences for the customer in the supply chain. In our example, the consequences and their significance for the customers of the supply chain section, i.e. the sensor manufacturer, would be very high, depending on the severity of the quality defect. In this case, the assessment could range from a 2 (for rework) to a 4 (for production breakdown).

An additional distinction is made by assessing the probability of occurrence of the risk consequence. In our example, this could mean a loss of production, as worst-case scenario, or rework, as less serious risk consequence, which have to be assessed for their probability of occurrence. The probability of occurrences for each scenario is also ranked with values from 1 (very rare) to 5 (very frequent). The probability of occurrence for the production loss scenario is very low, resulting in a score of 1. By contrast, the probability of occurrence for the rework scenario is certainly much higher, and therefore scores a 4.

After assessing the different parameters, the so-called risk priority number (RPN) can be determined. The RPN is calculated by adding occurrence probability and detection probability of risk cause, severity of risk consequence and frequency of risk consequence, multiplied by the value of 5. The method of addition was chosen to avoid a too strong effect by a potentially unrealistic assessment, as would be the case with multiplication, leading to falsified results. Finally, the RPN of each risk can be used to establish a hierarchy of risks. The prioritization of risks affects particularly the important parameters of risk cause, risk consequence as well as significance and frequency of risk consequence.

5.2.4 Phase III: Developing measures and risk control

It is important to align the development of measures with the preceding prioritization of risks, enabling a useful cost-benefit estimate for the different risk-stabilizing measures. To stabilize the prioritized risks by measures, the following basic strategies can be applied. The basic strategies are described in detail in section 2.2.3:

- Avoiding an identified risk
- Mitigating an identified risk
- Limiting or transferring an identified risk
- Sharing an identified risk
- Accepting an identified risk

It has to be said that not all of the mentioned measures are transferable in business. In our point of view and concerning the electronics industry, the transferring of risks to supply chain partners may interrupt the whole supply chain in case of an occurring risk.

5.2.5 Phase IV: Monitoring of measures

Two fundamental questions must be answered when it comes to the monitoring of measures. One is the question how to integrate supply chain risk management into day-to-day operations. And the other question is who is responsible for supply chain risk management and what organizational unit is responsible for it? During validation in companies of the electronics industry, it became obvious that neither a pure top-down nor a bottom-up approach makes sense. Instead, an interdisciplinary team comprising experts from both operational and strategic (i.e. top management) areas of the business should be put together. This helps to take care of risks which only one of the two groups can identify, whether operational or strategic risks. Letting the top management take part also ensures the organizational integration of supply chain risk management. This leads on to the actual application of the process model. A common mistake in industry is to recognize the importance of the methods of supply chain risk management but to apply them only once. After that, the methods as well as the supply chain risk management process end up in a drawer and are no longer carried out. For the success and usefulness of supply chain risk management, however, it is essential that the process is repeated cyclically.

To ensure maximum benefit, the presented process model should be applied once every year. The identification and validation of risks is done by the core team based on the analysis data of the previous year. So, the data pool is always updated with existing and previously identified risks. Likewise, additional and new risks are integrated in this cycle. In the end, supply chain risk management undergoes a continuous cycle and so ensures that risks to which SMEs in the electronics industry can be exposed are identified early on to be counteracted by the development of appropriate measures.

6. Limitations and review of the research

This study has several limitations. First, the sample size of the study is sufficient but not large. Therefore, the results have to be evaluated in further research with a larger sample size. There are also limitations related to the measures that were used. The length of the questionnaire and the number of items were limited. Further on, the study was limited to mainly German companies. The results might be transferable for companies across Europe, since there is one European economic area. However, the outcomes of the study could be different in other economic areas in the world like the NAFTA, LAC or APEC due to different economic, political and geographical positions.

The developed methodology of a risk stabilization process is focused on the electronics industry. Therefore as there has been a special attention to the risks in this industry branch. Other industry branches might be slightly different in their initial position. But the developed steps and the methods are usable in all industry branches. Therefore, the focus of the risk management process might change but the methods used in the risk management process will be the same.

7. Future work and conclusion

The study on supply chain risk management has shown that the supply chain risk management is a particularly important and urgent issue in the industry. Especially in the electronics industry with its strong OEMs and weak suppliers the supply chain risk management should be a standardized method in a company and therefore needs to be taken into account.

Companies increasingly recognize the importance of supply chain risk management, due to globalization and the associated increasing distribution of value-added activities, as well as the expansion and the growing complexity of the supply chains. The study also shows that supply chain risk management must analyze and be based upon both strategic and operational risks. Another interesting finding of the study is that the quality of supply chain risk management decreases with a shrinking vertical range of manufacture, and also that the process quality along the supply chain is reduced on the way to the end consumer. This is an important discovery for the electronics industry, since the distribution of value added and the vertical range of manufacture among the supply chain partners is low in the electronics industry.

It has to be said that future research about the topic of supply chain risk management should focus on broadly based case studies conducted in the European Union and the other economic areas for significant, transnational equation. The developed process model has to be mitigated to other industries and branches. Therefore, future research will be concentrated on the development of methods for a process model that can be used and implemented in any industry and industrial sector.

The described process model was developed to help SMEs in the electronics industry to evaluate the criticality of their supply chains. The methodology and the process model provide a systematic procedure that supports companies to minimize risks, enables companies to enhance their performance thanks to more stable supply chains, increases the productivity of companies by reducing the supply-related waiting times, and supports companies to increase their competitiveness based on a stabilized supply chain.

Supply Chain risk management will be a key success factor for companies in a globalized world if they have implemented a risk management process in their organizational structure.

8. References

Albers, S., Klapper, D., Konrad, U., Walter, A. & Wolf, J. (2007). *Methodik der empirischen Sozialforschung*, Gabler I GWV Fachverlage GmbH, 978-3834904690, Wiesbaden

Beckmann, H. (2004). Konzept des Supply Chain Managaments – Risiken, In: *Supply Chain Management– Strategien und Entwicklungstendenzen in Spitzenunternehmen"*, pp. 17, Springer-Verlag, 978-3540443902, Berlin

Blome, C. & Henke, M. (2009). Risikomanagement-Standard im Einkauf – Brunnen graben, bevor der Durst kommt, In: *Beschaffung aktuell*, book 12, pp. 30-34, 0343-9704

Brossardt B. (2005). Wertschöpfung hat Wert!, In: *Verband der Bayerischen Metall –und Elektroindustrie & TCW Transfer-Centrum GmbH & Co. KG für Produktions-Logistik und Technologiemanagement*, 03/21/2010, Available from: <http://www.stmwivt.bayern.de/pdf/wirtschaft/Wert_der_Wertschoepfung.pdf>

Christopher, M. (1992). *Logistics: The Strategic Issues*, Chapman & Hall, 978-0412597701, London

European Commission (2005). *The new SME definition User guide and model declaration*. Office for Official Publications of the European Communities, 978-9289479097, Luxembourg

Grundmann, T. (2008). Der Risikomanagement-Prozess – Schritte, angewandte Methoden und Hilfsmittel, In: *Ein anwendungsorientiertes System für das Management von Produkt - und Prozessrisiken*, pp. 27, Apprimus Verlag, 978-3940565105, Aachen

Greitemeyer, J. (2004). Expertenbrief zum Thema: Integriertes Risiko-Management für den Mittelstand, In: *Wirtschaftsjunioren Deutschland*, 03/25/2010, Available from: <www.ratingweb.de/download.php?media_id=00000162>

Kaufmann, L., Germer, T. (2001). Controlling internationaler Supply Chains Positionierung – Instrumente – Perspektiven, In: *Supply Chain Management: Unternehmensübergreifende Prozesse, Kollaboration, IT-Standards*, Arnold, U., Mayer, R., Urban,G. (Ed.), pp. 177-192, 978-3932306396, Bonn

Kersten, W., Hohlrath, P., Winter, M. (2008). Risikomanagement in Wertschöpfungsnetzwerken – Status quo und aktuelle Herausforderungen, In:

Supply Chain Risk Management, Schlattau, E. (Ed.), pp. 7-21, Wien: Fachhochsch. des Bfi Wien, 978-3902624048, Wien

Kümmerlein, R. (2009). Betriebe gehen zu sorglos mit Lieferrisiken um, In: *Deutsche Verkehrs-Zeitung*, WISO-Datenbank der GBI-Genios Deutsche Wirtschaftsdatenbank GmbH, DVZ Nr. 45-406, Hamburg

Lück, W. (2000). Managementrisiken, In: P*raxis des Risikomanagements: Grundlagen, Kategorien, branchenspezifische und strukturelle Aspekte*, Dörner, D., Horváth, P., Kagermann, H. (Ed.), pp. 311-343, 978-3791014524, Stuttgart

Mikus, B. (2001). Zur Integration des Risikomanagements in den Führungsprozess, In: *Risikomanagement*, Götze, U., Henselmann, K., Mikus, B. (Ed.), pp. 67-94, 3-790814156, Heidelberg

Schlattau, E., Mikus, B. (2001). *Supply Chain Risk Management Make-or-buy-Entscheidungen in der Produktion: Führungsprozesse, Risikomanagement und Modellanalyse*, GUC Gesellschaft für Unternehmensrechnung und Controlling m.b.H., 978-3934235175, Wiesbaden

Moder, M. (2008). *Supply Frühwarnsysteme die Identifikation und Analyse von Risiken in Einkauf und Supply Management* (1st. Edition), Gabler Verlag, 978-3834912039, Wiesbaden

Otterbach, B. (2011). OEMs beziffern Produktionsausfälle. In*: automobil-industrie.vogel.de*, 03/31/2011, Available from:
<http://www.automobil-industrie.vogel.de/oems/articles/308934>

Raithel, J. (2008). Die Befragung als dominantes Datenerhebungsverfahren, In: *Quantitative Forschung - Ein Praxiskurs*, pp. 66, Gabler I GWV Fachverlage GmbH, 978-3531161815, Wiesbaden

Rückle, H., Behn, M., (2007). *Unternehmenserfolg mit Zielen: Mit Checklisten, Leitfäden und Übungen*, expert-Verlag, 978-3816926474, Renningen

Schatz, A., Hermann M. (2010*) Risikomanagement in der Beschaffung – Studie 2010*, Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA, Retrieved from <http://www.ipa.fraunhofer.de/fileadmin/www.ipa.fhg.de/pdf/Studien/Studie_Risikomanagement_in_der_Beschaffung_2010.pdf>

Schatz, A., Hermann M. (2011). Supply Chain Risk Management – Relevanz und Handlungsbedarf, In: *ZWF Produktionstechnik im Wandel*, book 5/2011

Scholz-Reiter, B., Jakobza J. (1999). Supply Chain Management: Überblick und Konzeption, In: *HMD: Praxis der Wirtschaftsinformatik*, Hildebrand, K., Meinhardt S., pp. 7-15, HMD(207), Heidelberg

Siepermann, C., Vahrenkamp, R. (2007). Grundlagen des Risikomanagements in Supply Chains. In: *Risikomanagement in Supply Chains*, pp. 16, Erich Schmidt Verlag, 978-3503100415, Berlin

Siepermann, C., Vahrenkamp, R. (2008). Empirische Untersuchung zu SC-Risiken und SC-Risikomanagement in Deutschland, In: *Risikomanagement in Supply Chains*, pp. 61-73, Erich Schmidt Verlag, 978-3503100415, Berlin

Wildemann, H. (2006). Gestaltung des Risikomanagements im Leistungserstellungsprozess – Risikomanagement in der Beschaffung. In: *Risikomanagement und Rating*, pp. 141-142, TCW Transfer-Centrum GmbH & Co. KG, 978-3937236261 , München

Ziegenbein, A. (2007). Identifikation, Bewertung und Steuerung von Supply Chain Risiken – eine Methodik, In *Supply Chain Risiken: Identifikation, Bewertung und Steuerung*, pp. 81, Vdf Hochschulverlag AG, 978-3728131669, Zürich

Zwißler, F. & Hermann, M. (2010). Methodik zur Identifikation und Bewertung von Risiken in Supply Chains. *Productivity Management*. 15, 3,(October 2010), pp. 31-33, 1868-8519

Edited by Jan Emblemsvåg

A large part of academic literature, business literature as well as practices in real life are resting on the assumption that uncertainty and risk does not exist. We all know that this is not true, yet, a whole variety of methods, tools and practices are not attuned to the fact that the future is uncertain and that risks are all around us. However, despite risk management entering the agenda some decades ago, it has introduced risks on its own as illustrated by the financial crisis. Here is a book that goes beyond risk management as it is today and tries to discuss what needs to be improved further. The book also offers some cases.

ISBN 978-953-51-5111-1

Photo by natasaadzic / iStock

IntechOpen