

**IntechOpen**

# Advanced Geoscience Remote Sensing

*Edited by Maged Marghany*





---

# ADVANCED GEOSCIENCE REMOTE SENSING

---

Edited by **Maged Marghany**

## **Advanced Geoscience Remote Sensing**

<http://dx.doi.org/10.5772/57046>

Edited by Maged Marghany

### **Contributors**

Daniela Coltuc, María Victoria Soto, Maged Marghany, Vitalij Pavlenko, Viktor Speransky, Stuart Anderson, Edwige Vannier, Odile Taconet, Richard Dusseaux, Olivier Chimi-Chiadjeu, Yuanzhi Zhang, Yu Li, Othniel Likkason, Vladislav Olkhovsky

### **© The Editor(s) and the Author(s) 2014**

The moral rights of the and the author(s) have been asserted.

All rights to the book as a whole are reserved by INTECH. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECH's written permission.

Enquiries concerning the use of the book should be directed to INTECH rights and permissions department ([permissions@intechopen.com](mailto:permissions@intechopen.com)).

Violations are liable to prosecution under the governing Copyright Law.



Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 3.0 Unported License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be found at <http://www.intechopen.com/copyright-policy.html>.

### **Notice**

Statements and opinions expressed in the chapters are those 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, 2014 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](mailto:orders@intechopen.com)

Advanced Geoscience Remote Sensing

Edited by Maged Marghany

p. cm.

ISBN 978-953-51-1581-6

eBook (PDF) ISBN 978-953-51-5049-7

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,200+

Open access books available

116,000+

International authors and editors

125M+

Downloads

151

Countries delivered to

Our authors are among the  
Top 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)





# Meet the editor



Dr. Maged Marghany received his PhD from University Putra Malaysia in 2000. He was also awarded the ESA Post-doctoral Fellowship by the International Institute of Aerospace and Earth Observation (ITC) in Enschede, the Netherlands, in 2001. Dr. Marghany is a microwave remote sensing expert. Currently he is an associate professor at the Institute of Geospatial Science and Technology (INSTeG), Universiti Teknologi Malaysia (UTM). He has authored more than 200 reviewed papers and book chapters and had served as a reviewer of several international research journals. He led several projects related to application of microwave remote sensing to Malaysian coastal waters, funded by University Teknologi Malaysia (UTM), Ministry of Science and Technology, Malaysia (MOSTE), and Ministry of High Education, Malaysia (MOHE). Dr. Marghany was granted several awards: (i) best two research papers from Malaysian Space Agency (ANKASA) in 2010 and 2011; (ii) one of the best papers "Volterra Algorithm for Modelling Sea Surface Current Circulation from Satellite Altimetry Data". In Gervasi, O. and Gavrilova, M. (Eds.): *Lecture Notes in Computer Science. Computational Science and Its Applications - ICCSA 2008*, Springer-Verlag Berlin Heidelberg; (iii) the Excellent Services Award Year 2007 from Universiti Teknologi Malaysia; and (iv) best speaker award from Asian Association on Remote Sensing, November 1998. His Scopus H-index is 10. He is ranked 1 and 10 in two Distinctive Competencies (DC) for Universiti Teknologi in the World for Remote sensing: Algorithms; Oil spills; Surfaces. Finally, his biography has appeared in *Who's Who in World*, 1st Edition, 2013, 2009, 2007, and in *Who's Who in The World*, 23rd Edition, 2006 as a recognition for contribution in the field of science and engineering.



---

# Contents

---

## **Preface XI**

### **Section 1 Microwave Remote Sensing Applications 1**

Chapter 1 **Mathematical Description of Bayesian Algorithm for Speckle Reduction in Synthetic Aperture Radar Data 3**

Daniela Colţuc

Chapter 2 **Oil-Spill Pollution Remote Sensing by Synthetic Aperture Radar 27**

Yuanzhi Zhang, Yu Li and Hui Lin

Chapter 3 **Oil Spill Pollution Automatic Detection from MultiSAR Satellite Data Using Genetic Algorithm 51**

Maged Marghany

Chapter 4 **HF Radar Network Design for Remote Sensing of the South China Sea 73**

S. J. Anderson

Chapter 5 **A Three-dimensional of Coastline Deformation using the Sorting Reliability Algorithm of ENVISAT Interferometric Synthetic Aperture Radar 105**

Maged Marghany

### **Section 2 Nuclear, Geophysics and Telecommunication 123**

Chapter 6 **On the Modification of Nuclear Chronometry in Astrophysics and in Geophysics 125**

V.S. Olkhovsky

Chapter 7 **Exploring and Using the Magnetic Methods 141**

Othniel K. Likkason

- Chapter 8 **Identification of Communication Channels for Remote Sensing Systems Using Volterra Model in Frequency Domain** 175  
Vitaliy Pavlenko and Viktor Speranskyy
- Section 3 Environment Remote Sensing Investigations** 205
- Chapter 9 **Statistical Characterization of Bare Soil Surface Microrelief** 207  
Edwige Vannier, Odile Taconet, Richard Dusséaux and Olivier Chimi-Chiadjeu
- Chapter 10 **Simulation of Tsunami Impact on Sea Surface Salinity along Banda Aceh Coastal Waters, Indonesia** 229  
Maged Marghany
- Chapter 11 **Morphodynamic Environment in a Semiarid Mouth River Complex Choapa River, Chile** 253  
Joselyn Arriagada, María-Victoria Soto and Pablo Sarricolea

---

## Preface

---

Nowadays, advanced remote sensing technology plays tremendous roles to build a quantitative and comprehensive understanding of how the Earth system operates. The advanced remote sensing technology is also used widely to monitor and survey the natural disasters and man-made pollution. Besides, telecommunication is considered as precise advanced remote sensing technology tool. Telecommunication transmits at a distance by technological processes, principally through electrical signals or electromagnetic waves. Both remote sensing and telecommunication technologies are based on (i) a transmitter that takes information and converts it to a signal, (ii) a transmission medium, also called the "physical channel" that carries the signal. An example of this is the "free space channel"; and (iii) a receiver that takes the signal from the channel and converts it back into usable information. The increment needs of rapid, and comprehensive information are required to integrate between remote sensing platforms and telecommunication methods. These advanced technologies are compiled with network design, advanced mathematical and statistical methods such as Volterra, Bayesian, and Genetic Algorithm, multi-objective optimization problem, and Pareto dominance.

These advanced mathematical algorithms acquire precise understanding of their usages to the environment and geoscience. Indeed, it cannot build precise usages of remote sensing and telecommunication without a comprehensive understanding of mathematics and physics. However, some of academic institutes have the narrow point of view regarding applications of remote sensing and telecommunication technologies. Those academic institutes are using both techniques as mapping and surveying tools without understanding the philosophy of mathematics and physics beyond both concepts.

The most advanced remote sensing technology is microwave sensors. The microwave remote sensing platforms involve satellite platform, airborne, and High Frequency ground radar (HF) radar. The applications of microwave remote sensing are covering the wide range of the Earth system. These applications involve oceanography, meteorology, automatic detection of oil spill, and ship detections. It is a hard task to cover all the aspects of microwave remote sensing applications in this work. Instead, this book collected a number of high quality and comprehensive work on microwave remote sensing usages. Therefore, the optical remote sensing technologies are presented to study the tsunami 2004 effects on changing the sea surface salinity, bare of soil surface micro relief, and the morphodynamic environment in semiarid in mouth river. Finally, the magnetic method is also presented as a tool for mining exploration.

I also believe that there is a great integration between nuclear chronometer with remote sensing and telecommunication technologies. However, not many studies have established

this integration in comprehensive frames. In this regard, the book specifies one chapter for nuclear chronometer. At present, all known methods of nuclear chronometer there were usually taken into account the lifetime of only fundamental states of radioactive nuclei. The modification of nuclear chronometer is extremely needed for perfect understanding of the influence of constant cosmic radiation on the surface of the Earth.

This book has three parts (i) microwave remote sensing applications, (ii) nuclear, geophysics and telecommunication; and (iii) environment remote sensing investigations. The microwave remote sensing applications section contains chapters on synthetic aperture radar (SAR) speckle reduction, automatic detection of oil spill, interferometry synthetic aperture radar application to coastline deformation, and designing an HF network for oceanography and meteorology investigations the using optimization method of genetic algorithm. The nuclear, geophysics and telecommunications section has chapters on mathematical theoretical modification of nuclear chronometry, the application of the magnetic method to real field measurements of total-field aeromagnetic intensity, and identification telecommunication channels for remote sensing using the frequency domain which is based on Volterra model. Finally, environment remote sensing investigations section contains chapters on investigation of soil micro relief characteristics, the simulation of tsunami 2004 effects on the sea surface salinity and the morphodynamic environment in semiarid in mouth river.

I wish to convey my appreciation to all authors who have contributed to this book. Without their intense commitment this book would not have become such a precious piece of knowledge. I am also grateful to the InTech editorial team who has afforded the opportunity to publish this book.

**Maged Marghany**

Institute of Geospatial Science and Technology (INSTeG)  
Universiti Teknologi Malaysia  
Malaysia

---

# Microwave Remote Sensing Applications

---



---

# Mathematical Description of Bayesian Algorithm for Speckle Reduction in Synthetic Aperture Radar Data

---

Daniela Colțuc

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/57529>

---

## 1. Introduction

In remote sensing applications, the Synthetic Aperture Radar (SAR) images constitute a distinct class. The main particularity of this class is the image radiometry, which is a product between two signals: the terrain radar reflectivity  $R$  and the speckle  $s$ . The reflectivity depends on the terrain nature while the speckle is the consequence of the surface roughness. The speckle phenomenon is intrinsic to the radar technique. The radar sensor sends a coherent fascicle of microwaves toward the ground, which is scattered at the contact with the surface. The SAR antenna receives a non-coherent radiation resulted by the superposition of the reflected microwaves. These microwaves have random phase because of the surface roughness. The result is an image with a noisy aspect even in the regions with constant radar reflectivity. An example is the image in Figure 1, which represents a town surrounded by cultivated fields in French Guyana.

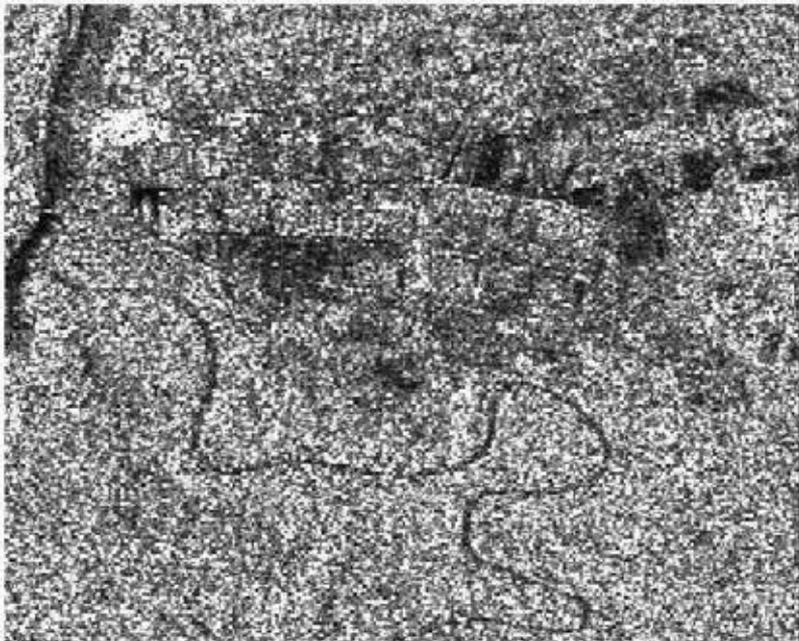
Although the speckle is carrying information that can be used in a series of applications, there are many cases where only  $R$  is of interest. In these situations, the speckle must be reduced in order to fully exploit the information in  $R$ .

The methods used to extract the reflectivity  $R$  from the SAR signal are known under the name of scene *reconstruction filters*. Among them, some of the most accurate in reconstruction are the Bayesian filters. These filters are based on the estimation theory, which has found a fertile terrain in SAR imagery due to the existing models for speckle. In the next section, we give a short overview of speckle distribution models in SAR imagery.

The Bayesian filters for speckle reduction were developed first for single SAR images and then extended to the multi-channel case. Here, the Bayesian filters have gained in efficiency due to the enhanced statistical diversity of the 3D data. This chapter presents several solutions for the

particular class of SAR multi-temporal series, which are sets of images of the same scene sensed at different dates. The rather long delays between acquisitions give the following special properties to these data: inter-channel speckle independency and radar reflectivity variation across the channels. The filters for multi-temporal SAR images presented in this chapter are either existing solutions, which were adapted to the above mentioned properties, or new approaches designed for the exclusive use on these data.

The multi-temporal SAR data have been used in a series of applications like change detection or land cover classification. While the majority of developed methods use a spatial speckle filtering as a pre-processing step [1-8], the recent publications have started reporting the use of multi-temporal filters [9-12].



**Figure 1.** Intensity SAR image.

The chapter begins with a brief description of speckle models, continues with the basics of estimation theory necessary to understand the Bayesian filters, presents some Bayesian filters for speckle reduction in single channel SAR images and ends with the solutions for speckle suppression in multi-temporal sets. At the end, we give some conclusions.

## 2. Speckle models in SAR images

The SAR images are of various types: complex, intensity, amplitude, logarithm, single-look and multi-looks. At the origin, it is the complex image that undergoes operations like pixels

modulus calculation (amplitude image), the square of modulus (intensity image), the logarithm of intensity (logarithm image) or the more complex multi-look processing consisting in pixels averaging and resampling.

In the complex SAR image, the pixels have complex values. The real part is the in-phase component of the measured radiation (same phase with the reference radiation of SAR sensor) and the imaginary part is the quadrature component.

In the case of fully developed speckle, both components have a Gaussian distribution:

$$P(z_1 | R) = \frac{1}{\sqrt{\pi R}} \exp\left(-\frac{z_1^2}{R}\right) \quad (1)$$

$$P(z_2 | R) = \frac{1}{\sqrt{\pi R}} \exp\left(-\frac{z_2^2}{R}\right) \quad (2)$$

where  $P$  is the probability density function (pdf),  $z_1$  is the real part and  $z_2$  the imaginary part. The eq. 1 and 2 show that, in a zone with constant reflectivity, the image pixels are varying because of speckle. By making  $R=1$ , one obtains the speckle pdf in complex SAR images.

The SAR intensity image  $I$  is obtained from the complex image by considering the square of the modulus:

$$I = z_1^2 + z_2^2 \quad (3)$$

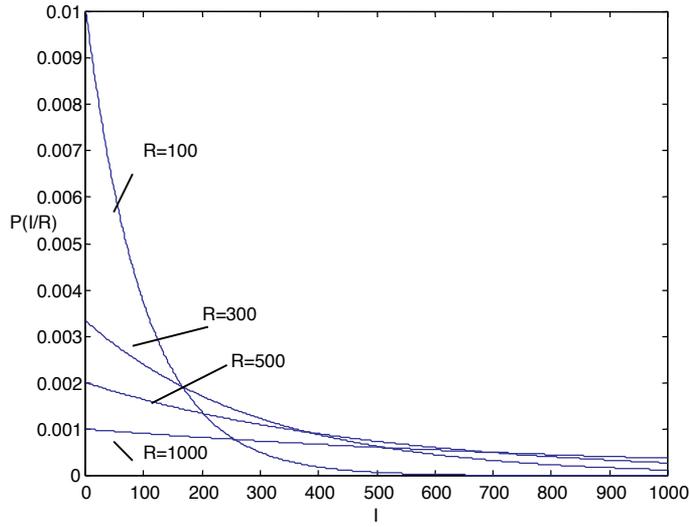
Generally, any mathematical operation with random variables (r.v.) changes the r.v. pdf. In the case of intensity images, the speckle pdf changes into an exponential distribution [13]:

$$P_I(I / R) = \frac{1}{R} \exp\left(-\frac{I}{R}\right) I \geq 0 \quad (4)$$

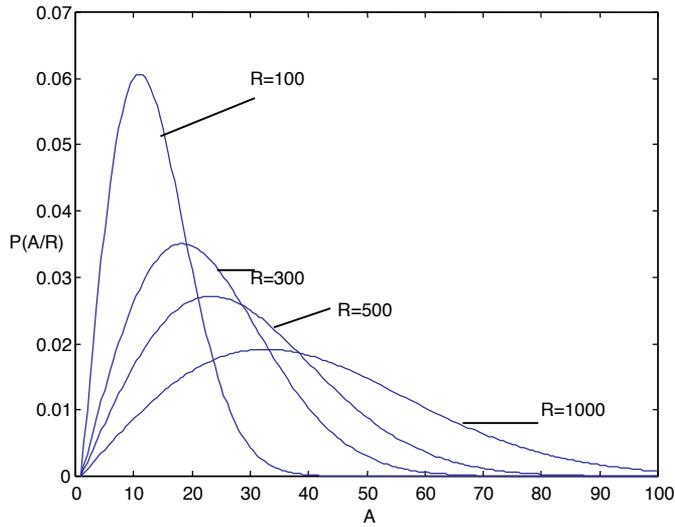
Fig. 2 shows the distribution in (4) for various values of  $R$ . The speckle distribution in SAR intensity images is obtained for  $R=1$ .

The amplitude SAR image is the square root of the intensity image,  $A = \sqrt{I}$ . In this case, the speckle has a Rayleigh distribution (Fig. 3):

$$P_A(A / R) = \frac{2A}{R} \exp\left(-\frac{A^2}{R}\right) A \geq 0 \quad (5)$$



**Figure 2.** Pixel pdf in an intensity image representing a zone with constant  $R$ .



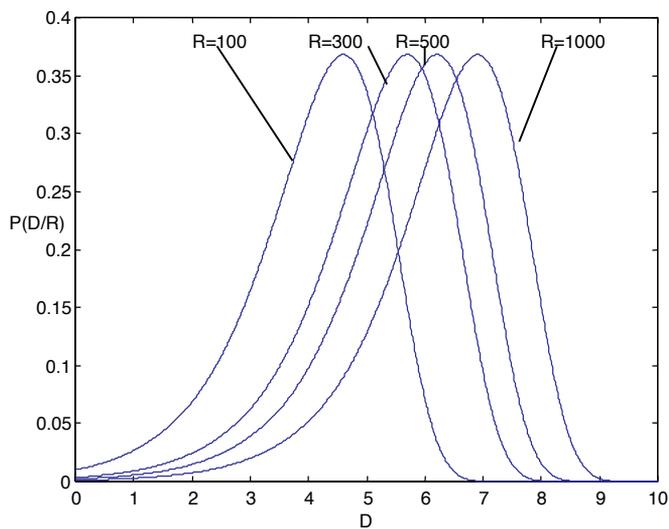
**Figure 3.** Pixel pdf in an amplitude image representing a zone with constant  $R$ .

The curves in Fig. 2 and 3 show that, for both intensity and amplitude images, the pixels variance increases with  $R$ . This explains why the speckle, although stationary, appears as stronger in the light areas. It is, in fact, a consequence of the multiplicative nature of the speckle.

The logarithm image is the logarithm of the intensity  $D=\ln(I)$ . The logarithm image has a Fisher-Tippet distribution [14]:

$$P_D(D/R)=\frac{e^D}{R}\exp\left(-\frac{e^D}{R}\right) \quad (6)$$

For various  $R$ , the distribution changes its mean but not its variance (Fig. 4). Indeed, since the logarithm transforms any product into a sum, the speckle in the logarithm SAR image is an additive stationary noise, which explains the variance constancy.



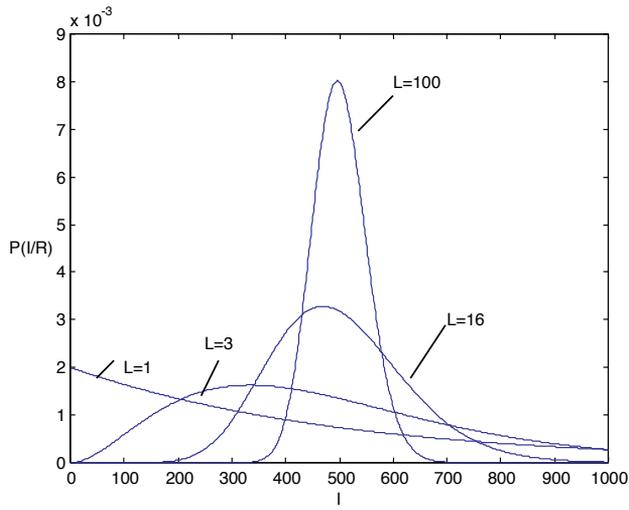
**Figure 4.** Pixel pdf in a logarithm image representing a zone with constant  $R$ .

The multi-look images are obtained by averaging, on columns, groups or 3 or 4 pixels. The resulting image must be interpolated and resampled in order to have the same resolution on rows and columns. Due to pixels averaging, the multi-look image is less noisy.

Depending on the image at the origin – intensity, amplitude or logarithm-the multi-look image can be of intensity, amplitude or logarithm type. The multi-look treatment changes the speckle distribution as follows:

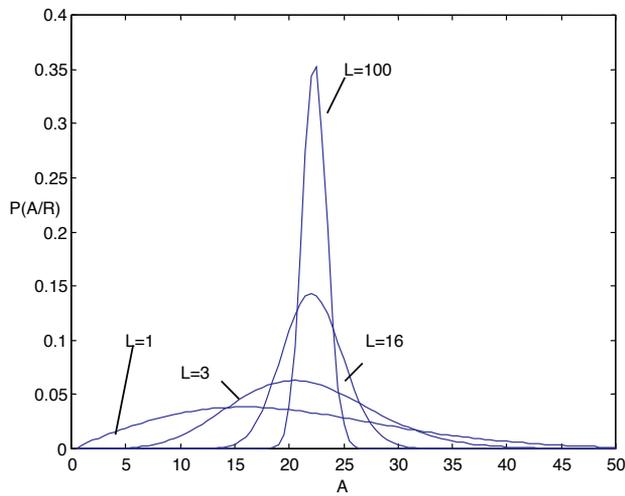
- Gamma distribution in multi-look intensity SAR image [13] (Fig. 5):

$$P_I(I/R)=\frac{1}{\Gamma(L)}\left(\frac{L}{R}\right)^L I^{L-1}\exp\left(-\frac{LI}{R}\right)I\geq 0 \quad (7)$$



**Figure 5.** Pixel pdf in a multi-look intensity image representing a zone with constant  $R$  where  $L$  is the number of looks (the number of averaged pixels).

- Generalized Gamma distribution in multi-look amplitude SAR image (Fig. 6):

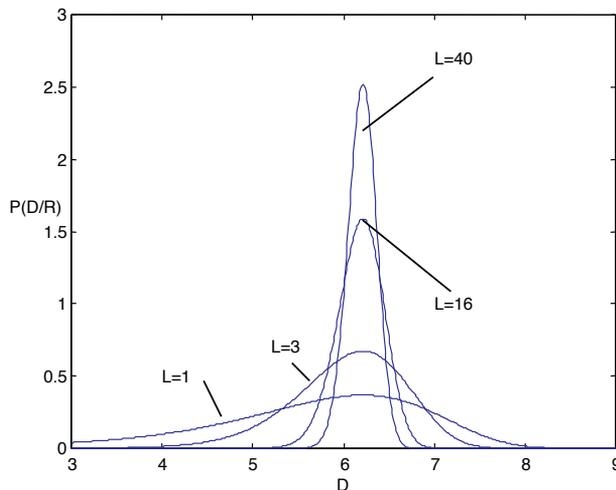


**Figure 6.** Pixel pdf in a multi-look amplitude image representing a zone with constant  $R$ .

$$P_A(A) = \frac{2}{\Gamma(L)} \left(\frac{L}{R}\right)^L A^{2L-1} \exp\left(\frac{-L A^2}{R}\right) A \geq 0 \tag{8}$$

- Fisher-Tippet distribution in multi-look logarithm SAR images [13] (Fig. 7):

$$P_D(D|R) = \frac{L^L}{\Gamma(L)} \exp(-L(D_R - D)) \exp(-L \exp(-(D_R - D))) \quad (9)$$



**Figure 7.** Pixel pdf in a multi-look logarithm image representing a zone with constant  $R$ .

where  $D_R = \ln(R)$ . As shown in Figures 5-7, in all the types of multi-look images, the speckle variance is reducing as the number of looks increases. The other side of the coin is the lower resolution of the resulted image.

Another aspect-important for Bayesian filters which are based on the calculus of estimates-is the speckle correlation. The speckle becomes a correlated signal at the sensor level that has a limited band. In estimation, the samples correlation worsens the estimate precision.

### 3. Basics of estimation theory

The multiplicative model  $I = R \cdot s$  shows that the SAR image radiometry  $I$  can be interpreted as a random signal having the reflectivity  $R$  as parameter. The Estimation Theory provides the means to estimate this parameter from the image pixels. The starting point in any estimation is a cost function  $C(\varepsilon)$  that has as argument the estimation error:

$$\varepsilon = \hat{R} - R \quad (10)$$

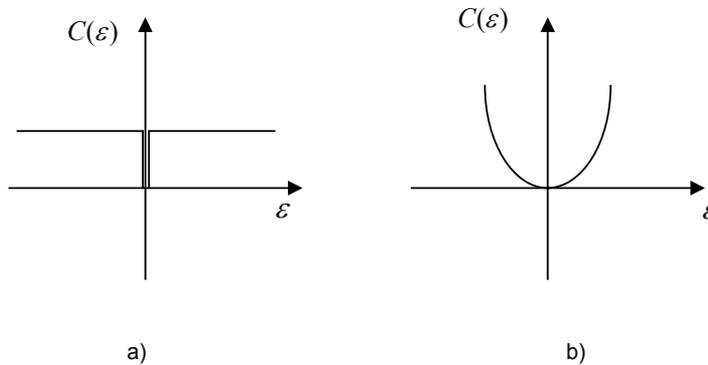
where  $\hat{R}$  is the estimated value of the reflectivity and  $R$  its actual value. The cost measures the penalty introduced by the estimation error. Therefore, it is natural to estimate the reflectivity under the constraint of minimum cost. Since the cost is itself a random variable that cannot be known by its particular realizations but by its statistical moments, the condition for the estimator is to minimize the average cost:

$$\bar{C} = \iint C(\hat{R} - R) \cdot P(R, I) dR dI \quad (11)$$

where  $P(R, I)$  is the joint pdf of  $R$  and  $I$ . The expression of  $\hat{R}$  is found by solving the equation  $\bar{C} = \min$ .

The cost function can be defined in several ways. The most common ones are the uniform and the quadratic functions. The uniform cost function is equal to one everywhere, excepting the origin where it is equal to zero (Fig. 8a) :

$$C(\varepsilon) = 1 - \delta(\varepsilon) \quad (12)$$



**Figure 8.** Two cost functions: a) the uniform function, b) The quadratic function.

With this function the penalty is the same disregarding the estimation error. The quadratic function has a penalty that increases with the estimation error (Fig. 8b):

$$C(\varepsilon) = \varepsilon^2 \quad (13)$$

### 3.1. Estimators based on the uniform cost function

By introducing (12) in the average cost (11), one obtains the following equation:

$$\begin{aligned} \bar{C} &= \iint [1 - \delta(\hat{R} - R)] P(R, I) dR dI = \\ &= \iint P(R, I) dR dI - P(\hat{R}, I) = 1 - P(\hat{R}, I) \end{aligned} \quad (14)$$

which means that, in the case of the uniform cost function, the estimate  $\hat{R}$  is the value of the reflectivity that maximizes  $P(\hat{R}, I)$ . Since in many cases,  $P(\hat{R}, I)$  has not an analytical expression or is too complex, it is replaced by  $P(R, I) = P(R/I)P(I)$  and the estimate  $\hat{R}$  is:

$$\max(P(R/I)) = P(\hat{R}/I) \quad (15)$$

This is the definition of the optimal estimator in the case of the uniform cost function. Since it is obtained by maximizing an a posteriori pdf, it is known as the *Maximum A Posteriori (MAP) estimator*. By using Bayes equation,  $P(R/I)$  can be expressed as a function of  $P(I/R)$ , a distribution that is known in SAR imagery:

$$P(R/I) = \frac{P(I/R)P(R)}{P(I)} \quad (16)$$

According to (8), maximizing  $P(R/I)$  is equivalent to maximizing  $P(I/R)P(R)$  (the reflectivity does not appear explicitly in  $P(I)$ ). When the reflectivity pdf  $P(R)$  is not known, it can be considered constant (uniform distribution) and the estimate  $\hat{R}$  is obtained by maximizing the a priori pdf:

$$\max(P(I/R)) = P(I/\hat{R}) \quad (17)$$

The equation (9) defines the *maximum likelihood estimator* of  $\hat{R}$ , which is suboptimal by respect to MAP because it ignores the scene reflectivity distribution.

### 3.2. Estimators based on the quadratic cost function

The quadratic cost function gives an average cost of:

$$\bar{C} = \iint (\hat{R} - R)^2 \cdot P(R, I) dR dI \quad (18)$$

The minimization of  $\bar{C}$  is equivalent to solving the equation:

$$\begin{aligned} \frac{d\bar{C}}{d\hat{R}} &= \iint 2(\hat{R} - R) \cdot P(R, I) dR dI = \\ &= \int P(I) \left[ \int 2(\hat{R} - R) P(R/I) dR \right] dI = 0 \end{aligned} \quad (19)$$

that gives the solution for the estimator  $\hat{R}$  :

$$\hat{R} = \int R \cdot P(R/I) dR = E[R/I] \quad (20)$$

The equation (20) shows that the optimal estimator, in the case of the quadratic cost function, is the conditional mean of the reflectivity. It can be derived from the SAR image radiometry if the a posteriori distribution  $P(R/I)$  is known. When  $P(R/I)$  is not known, a suboptimal estimate can be obtained by considering this distribution constant (uniform) and by using a linear estimate:

$$\hat{R} = ai + b \quad (21)$$

The coefficients  $a$  and  $b$  are obtained by minimizing the average cost, which in this particular case is:

$$\bar{C} = \iint (\hat{R} - R)^2 dR dI \quad (22)$$

## 4. Bayesian filters for single channel SAR images

### 4.1. Filters based on the quadratic cost function

#### a. Kuan Filter

The filter of Kuan is a linear estimator (21), based on the quadratic cost function. It is suboptimal because the a posteriori distribution  $P(R/I)$  is considered uniform as in (22). Other simplifying hypotheses used in deriving this filter are the speckle and the radar reflectivity whiteness, which is not realistic in none of the cases. These hypotheses worsen the estimator quality but simplify a lot its analytical expression.

The filter of Kuan estimates the radar reflectivity by [15]:

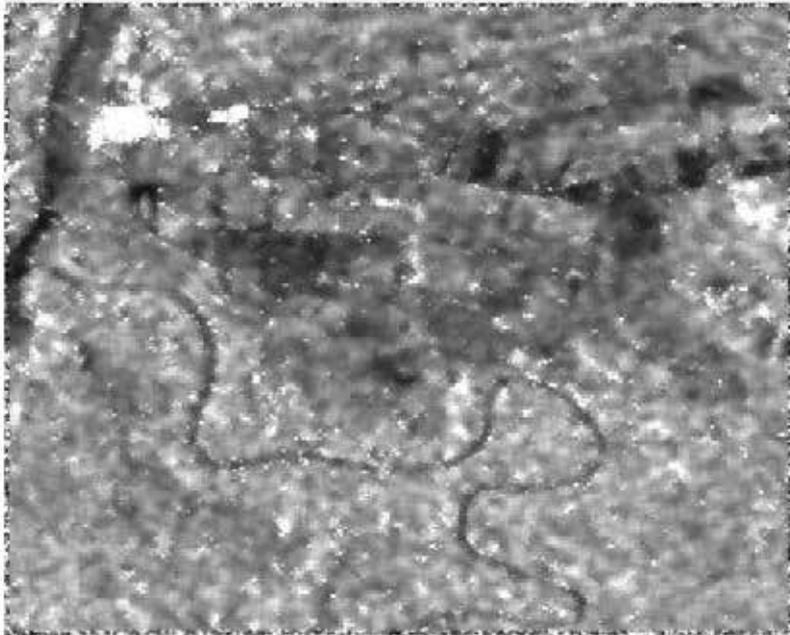
$$\hat{R}(x, y) = k \cdot I(x, y) + (1-k) \cdot \bar{I} \quad (23)$$

where  $I(x, y)$  is the current pixel,  $\bar{I}$  is the average radiometry in a window centered on  $I(x, y)$  and  $k$  is a coefficient estimated in the same window. It is given by:

$$k = \frac{1 - c_s^2 / c_I^2}{1 + c_s^2} \quad (24)$$

where  $c_s$  and  $c_I$  are the speckle and the image coefficients of variation. The coefficient of variation is the ratio between the signal standard deviation and mean. For the speckle, it is known (it depends on SAR image type) and for the image, it is estimated locally.

The filter smooths the image ( $\hat{R}(x, y) = \bar{I}$ ) in the areas with constant reflectivity and favors the current pixel  $I(x, y)$  on textured zones or edges. Figure 9 shows the result of filtering the SAR image in Figure 10, by Kuan filter. The coefficient of variation and the average intensity was estimated in a 9x9 neighborhood of the current pixel.

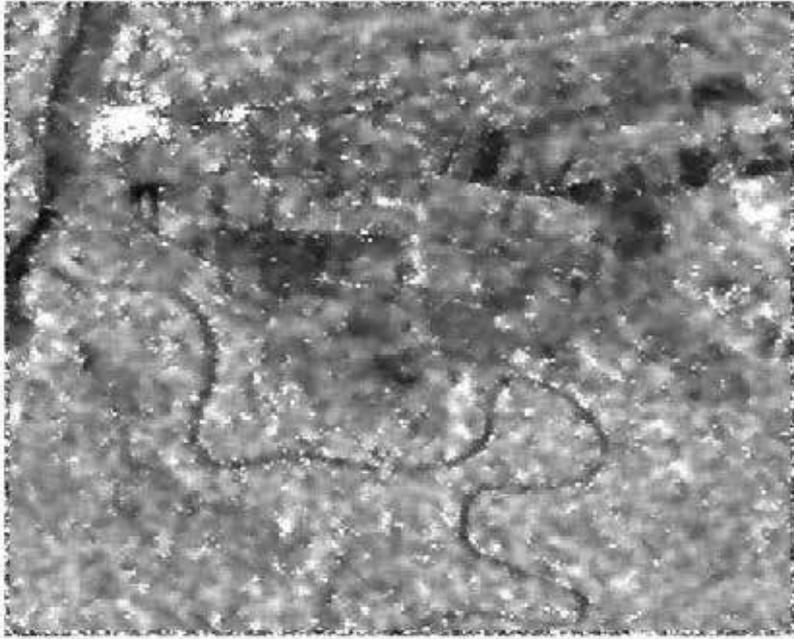


**Figure 9.** Scene reconstructed by Kuan filter with 9x9 neighborhoods.

#### **b. Lee filter**

The filter of Lee is of the same class as the filter of Kuan i.e., it is linear, based on the quadratic cost function and ignores the a posteriori distribution  $P(R/I)$ . The same hypotheses are considered: white speckle, white radar reflectivity and independence between speckle and reflectivity. The difference consists in the way it is derived. Being originally developed for additional noise, the speckle is first transformed into an additive noise by using the following approximation:

$$I(x, y) \approx A \cdot R(x, y) + B \cdot s(x, y) + C \quad (25)$$



**Figure 10.** Scene reconstructed by Lee filter with 9x9 neighborhoods.

where  $A=\bar{s}$ ,  $B=\bar{R}$ ,  $C=-\bar{R} \cdot \bar{s}$  (the coefficients are determined from the conditions to have an unbiased estimator and a minimum average cost). By applying the Lee filter for additive noise on this approximation, Lee obtained the following estimator [16]:

$$\hat{R}(x, y) = \bar{I} + k(I(x, y) - \bar{I}) \quad \text{where} \quad k = 1 - \frac{c_s^2}{c_I^2} \quad (26)$$

Figure 4 shows the SAR image filtered by Lee filter using a 9x9 neighborhood. The equations of Lee and Kuan filters are very similar. The difference is in the expression of  $k$ , which in the case of Lee filter is supplementary divided by  $1 + c_s^2$ . This number tends to 1 in multilook SAR images, showing that the two filters process similarly these data. It is not the case for single look images where  $1 + c_s^2 = 2$  and, consequently, Lee filter is less accurate.

### c. Frost filter

Although of the same class as Kuan filter – linear, minimizing the quadratic cost and ignoring the a priori distribution – the filter of Frost reduces better the speckle by modifying the hypothesis regarding the radar reflectivity. Differently from Lee and Kuan, Frost considers the reflectivity a correlated signal and uses for the correlation the exponential model:

$$\rho_R(d) = e^{-ad} \tag{27}$$

where  $\rho_R$  is the reflectivity coefficient of correlation,  $a$  is a parameter depending on the scene content and  $d$  is the pixel lag. As a consequence, the filter of Lee takes into account not only the current pixel  $I(x, y)$  but all the pixels in the neighborhood (usually a 3x3 neighborhood). The filter coefficients in the 1D case are [17]:

$$w(d) = K\alpha e^{-\alpha d} \quad \text{with} \quad \alpha^2 = K'c_I^2 \tag{28}$$

where  $K$  has the role to remove the estimator bias and  $K'$  is a constant that tunes the filtering effect. For images, the filter (28) it is applied on rows and columns.

Other filters based on the quadratic cost function are the EAP (Estimator a Posteriori) and the t-linear filter. The EAP minimizes the average cost in (18) not the simplified condition in (25), by considering a Gamma distribution for the radar reflectivity. It is more accurate than Kuan, Lee or Frost filters but it is time consuming.

#### 4.2. Filters based on the uniform cost function

The a posteriori distribution (15), which gives the MAP estimator, includes the a priori distribution (known for SAR images) and the reflectivity distribution. Each model used for the reflectivity distribution gives another estimator. The higher the model complexity, the better is the estimator. In this section, we give the estimators for the following three models of radar reflectivity pdf: uniform correlated distribution, Gamma uncorrelated distribution, Gamma correlated distribution.

A uniform correlated distribution for the reflectivity gives the following distribution for the intensity of the SAR image:

$$P_{loc}(I | R) = \left(\frac{\alpha}{R}\right)^L \frac{I^{\alpha-1}}{\Gamma(\alpha)} \exp\left[-\frac{\alpha I}{R}\right] \tag{29}$$

where  $R$  is the reflectivity of the current pixel,  $I$  is the intensity of any pixel in the 3x3 neighborhood of  $R$  and  $\alpha$  is a shape parameter:

$$\alpha = \frac{1}{c_I^2} \tag{30}$$

In the hypothesis of uncorrelated intensity, the maximization of the a posteriori probability gives the following expression for MAP estimator:

$$\hat{R}(x, y) = \bar{I} + k(I(x, y) - \bar{I}) \quad \text{with} \quad k = \frac{1}{1 + 8\alpha/L} \tag{31}$$

where  $L$  is the number of looks. Being derived by considering a uniform distribution for the reflectivity, (31) is in fact a maximum likelihood estimator. The parameter  $\alpha$  is estimated in the  $3 \times 3$  neighborhood of the current pixel and  $\bar{I}$  in a larger neighborhood.

A better estimator is obtained by introducing in MAP equation a non-uniform model for the radar reflectivity distribution. The non-uniform model currently used for the reflectivity is the Gamma distribution:

$$P(R) = \left( \frac{\nu}{E[R]} \right)^\nu \frac{R^{\nu-1}}{\Gamma(\nu)} \exp\left(-\frac{\nu R}{E[R]}\right) \quad (32)$$

where  $\nu$  is a shape parameter equal to  $1/c_R^2$ . Between  $\nu$  and  $\alpha$  it is the following connection:

$$\nu = \frac{1 + 1/L}{1/\alpha - 1/L} \quad (33)$$

The reflectivity estimator in this case is:

$$\hat{R}(x, y) = \frac{1}{2\nu} (E[R](\nu - L - 1) + \sqrt{E[R]^2(\nu - L - 1) + 4\nu LI(x, y)E[R]}) \quad (34)$$

where  $E[R]$  and  $\nu$  are estimated in a neighborhood of the current pixel. For the intensity SAR images,  $E[R] = E[I]$ .

The maximum likelihood estimator in (31) takes into account the radar reflectivity correlation but considers a uniform distribution while the estimator in (34) ignores the correlation and considers a reflectivity with Gamma distribution. A filter that uses both a correlated and a Gamma distribution model is the following estimator:

$$\hat{R}(x, y) = \frac{1}{2\nu} (E[R](\nu - L - 1 - 8\alpha) + \sqrt{E[R]^2(\nu - L - 1 - 8\alpha) + 4\nu(LI(x, y) + 8\alpha E[I])E[R]}) \quad (35)$$

where  $E[R]$ ,  $E[I]$ ,  $\nu$  and  $\alpha$  are estimated on the image, in a neighborhood of the current pixel. Like for the other two filters,  $\nu$  and  $\alpha$  are estimated in a  $3 \times 3$  window and  $E[R]$  and  $E[I]$  in larger neighborhood (the neighborhood size is chosen usually between  $7 \times 7$  and  $11 \times 11$  pixels).

The list of the Bayesian filters does not end here. This section has presented only several of them, which are the most used. Other filters from this class can be found in [18-21]. At a more attentive analysis, one can see that all these filters have in common the following behavior: in the areas with constant reflectivity, the reflectivity is estimated by the local mean of the intensity. The filters differ by the solution proposed in the textured areas and on the edges.

## 5. Bayesian filters for multi-temporal SAR images

The multichannel SAR images are 3D volumes of data constituted by assembling several SAR images representing the same scene. The images are registered i.e., the corresponding pixels in different channels (images) represent the same resolution cell at the ground. The Bayesian filters for multichannel images are in principal the same as for single SAR images: MAP, maximum likelihood and the conditional mean. The differences concern the neighborhood used by the estimators and the speckle statistics properties, more precisely the inter-channel correlation.

None of the Bayesian filters reconstructs perfectly a scene. Generally, the filtering result is a tradeoff between speckle reducing and texture and edges blurring. An essential role in making this tradeoff has the size of the neighborhood used for estimating the statistical moments included in filter expression. In a large neighborhood, the estimation is theoretically more accurate but many fine details can be lost. The main advantage of multichannel filters is in the neighborhood used for estimating the statistical moments. The existence of a third dimension allows the neighborhood extension by preserving a reasonable size in the image plane. For a given precision of the estimation, scene details are better preserved when 3D neighborhoods are used.

In the image plane, the speckle is a correlated signal. In estimation, the samples correlation is a withdraw because, at a given statistical population size, the estimation precision is worse than if the samples were independent. In multi-temporal sets, due to the significant time leg between acquisitions, the speckle components in different channels are not correlated. Therefore, the neighborhood extension on the 3<sup>rd</sup> dimension improves a lot the result of Bayesian filters.

This section presents three Bayesian filters for multi-temporal sets: the linear filter [22,23] (the equivalent of Lee filter), the Time-Space filter [24-26] and the Multi-temporal Non-Local Mean [30]. Other filters from the same class can be find in [28, 29].

### 5.1. The linear filter for multi-temporal SAR images

In Section 3.2, it was shown that when the a posteriori distribution of the reflectivity is not known, a linear estimator (21) can be obtained by minimizing the average quadratic error (22). It is the solution that gives the filters of Kuan, Lee or Frost. In the case of multichannel SAR images, the linear estimator is a weighted sum of the corresponding pixels in all the channels [22]:

$$\hat{R}_i(x, y) = \sum_{j=1}^N \alpha_j^{(i)} I_j(x, y) \frac{E[I_i]}{E[I_j]} \quad i=1, \dots, N \quad (36)$$

where  $N$  is the number of channels and  $\alpha_j^{(i)}$  is a set of coefficients that are determined from the following conditions:

$$\min(E[(\hat{R}_i - R_i)^2]) \quad (37)$$

$$\sum_{j=1}^N \alpha_j^{(i)} = 1 \quad \forall \quad i=1, \dots, N \quad (38)$$

The equations (37) are the condition for minimum cost and (38) the condition for unbiased estimators.

By using the linear estimator (36), Bruniquel derived a multi-temporal version of Lee filter (26). Additionally, he improved the filter hypotheses, by considering that the scene has also textures and not only homogenous zones like in the case of Lee's filter. For channels with identical textures, the conditions (37, 38) become [22]:

$$\begin{cases} \sum_{j=1}^N \alpha_j^{(n)} = 1 \\ \sum_{j=1}^N \alpha_j^{(n)} (\rho_{i,j} c_i c_j - \rho_{j,n} c_j c_n) = 0 \quad i=1, \dots, N \quad i \neq n \end{cases} \quad (39)$$

Where  $\rho_{i,j}$  is the coefficient of correlation between the channels  $i$  and  $j$  and  $c_i$  is the coefficient of variation in the channel  $i$ .

In multi-temporal sets, with high probability, the textures change from a channel to another because of the delay between the acquisitions (months, seasons or years). For this reason, in a second essay, Bruniquel introduced the hypothesis of different textures and obtained the following equations for the two conditions [23]:

$$\begin{cases} \sum_{j=1}^N \alpha_j^{(n)} = 1 \\ \sum_{j=1}^N \alpha_j^{(n)} (\rho_{i,j} c_i c_j - \rho_{n,j} c_n c_j) = \rho_{n,i} c_n c_i - c_n^2 \quad \begin{cases} i=1, \dots, N \\ i \neq n \end{cases} \end{cases} \quad (40)$$

In this case, the inter-channel correlation is not anymore the speckle correlation i.e., zero but the correlation between textures.

This filter uses the most realistic hypotheses for multi-temporal images and consequently it arrives to reconstruct rather well the scenes. The filter withdraw is the high complexity: for each pixel in the scene, one has to solve the system (40).

## 5.2. The time-space filter

A Bayesian filter for multi-temporal SAR images, both efficient and of low complexity but which passed almost unobserved, is the Time-Space filter. This filter takes advantage of the fact that, across the channels, the reflectivity is correlated while the speckle is independent, which means that they have different frequency bands. In order to separate in frequency these

two signals, the authors consider the logarithm of the images and then apply a 1D Discrete Cosine Transform (DCT) only on the 3<sup>rd</sup> dimension of the set. Due to the above mentioned properties, the reflectivity energy accumulates in zero frequency coefficients while that of speckle remains uniformly distributed over all frequencies. By filtering only the non-zero frequency coefficients, it is possible to obtain a good quality for the reconstructed scene. The non-zero frequency planes are filtered by using the filter of Lee for additive noise and the reconstructed channels are obtained by inverse DCT and exponential calculation.

The scene reconstruction is done in 6 steps [24, 26]:

1. Let be  $I_0, \dots, I_{N-1}$  the  $N$  channels of the multi-temporal set. The first step consists in calculating the logarithm of the images:

$$I_n \Rightarrow G_n = \ln(I_n) = \ln(R) + \ln(s) \quad \forall n \in [0, \dots, N-1] \quad (41)$$

For the pixels with value 0, a normalization is necessary: one can either consider that the logarithm is zero (the error is insignificant for an image on 16 bits) or add 1 to all the pixels before applying the logarithm.

2. A DCT is applied across the multi-temporal set. More precisely, the vector  $G_{xy}(n)$  constituted by the pixels with the same position  $(x, y)$  in the channels is transformed:

$$G_{xy}(n) = \begin{bmatrix} \ln I_0(x, y) \\ \dots \\ \ln I_{N-1}(x, y) \end{bmatrix} \Rightarrow T_{xy}(k) = DCT\{G_{xy}(n)\} = \begin{bmatrix} T_0(x, y) \\ \dots \\ T_{N-1}(x, y) \end{bmatrix} \quad (42)$$

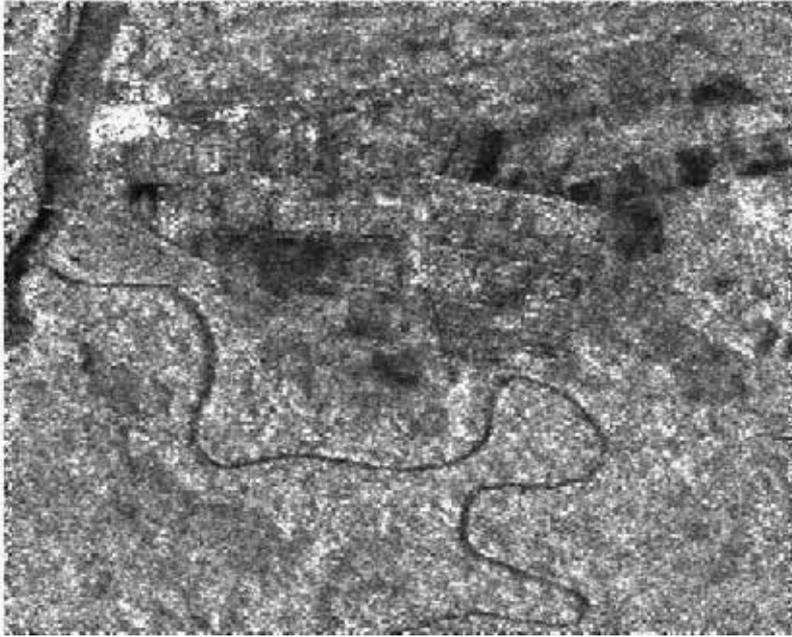
The result is a multispectral set with the same size as the multi-temporal set. The images are now constituted by DCT coefficients of the same frequency.

3. After DCT, the radar reflectivity, which is a correlated signal, is concentrated in  $T(0)$ . For this reason,  $T(0)$  is conserved and only the rest of the frequency planes are filtered by using Lee's filter for additive noise:

$$\hat{T}_{xy} = E[T_{xy}] + (T_{xy} - E[T_{xy}]) \frac{\sigma_T^2 - \sigma_{zgomot}^2}{\sigma_T^2} \quad (43)$$

where the mean  $E[T_{xy}]$  and the variance  $\sigma_T^2$  are estimated in a neighborhood of  $(x, y)$ . The variance  $\sigma_{zgomot}^2$  is a parameter that depends on speckle statistics. It was shown that it can be approximated by the speckle coefficient of variation.

4. Inverse DCT of the filtered set:



**Figure 11.** Scene reconstructed by Time-Space Filtering of 6 multi-temporal SAR images.

$$\hat{T}_{xy}(k) = \begin{bmatrix} T_0(x, y) \\ \hat{T}_1(x, y) \\ \dots \\ \hat{T}_{N-1}(x, y) \end{bmatrix} \Rightarrow \hat{G}_{xy}(n) = DCT^{-1}\{\hat{T}_{xy}(k)\} = \begin{bmatrix} \hat{G}_0(x, y) \\ \hat{G}_1(x, y) \\ \dots \\ \hat{G}_{N-1}(x, y) \end{bmatrix} \quad (44)$$

5. The genuine dynamic range of the set is obtained by calculating the exponential:

$$\hat{G}_n \Rightarrow \hat{I}_n = \exp(\hat{G}_n) \quad \forall n \in [0, \dots, N-1] \quad (45)$$

6.  $\hat{I}_n$  is a biased estimator because of the logarithm in the first stage. In [25], it was shown that bias is a multiplicative constant that depends on SAR image type and on the number of channels:

$$bias = \left( \int_0^{\infty} s^{\frac{1}{N}} P(s) ds \right)^N \quad (46)$$

where  $P(s)$  is the speckle pdf. The bias is corrected by:

$$\hat{I}_n \Rightarrow \hat{I}_n^{correct} = \frac{\hat{I}_n}{bias} \quad \forall n \in [0, \dots, N-1] \quad (47)$$

Figure 11 shows the scene in Figure 1 reconstructed by applying the Time-Space filter on a set of 6 multi-temporal SAR images. The filter of Lee uses a neighborhood of 11x11 pixels for moments estimation. A smaller neighborhood (9x9 pixels) is used for single SAR image filtering in Figures 3 and 4. Obviously, the image in Figures 6 has the details much better preserved than in the case of single channel filtering, where an effect of watercolor dominates. Even the finest details in the field at the right side of the image are preserved.

### 5.3. Two steps multi-temporal Non-Local Means

The Non-Local Means (NLM) is a denoising technique, which estimates each pixel by a weighted average of the similar pixels. NLM is a weighted maximum likelihood estimator. The NLM weights are defined according to the distance between the similar pixels neighborhoods.

The NLM derived for SAR images in [27] uses weights refined iteratively:

$$\hat{R}(i) = \sum_{j \in W} w(i, j) I(j) \quad (48)$$

where  $I(j)$  are the image pixels (indexed),  $w(i, j)$  the weights and  $W$  is the search neighborhood. The weights are given by:

$$w(i, j) = \frac{1}{Z} \exp \left[ -\frac{1}{h_0} S(i, j) - \frac{L}{h_1} R^{m-1}(i, j) \right] \quad (49)$$

with

$$S(i, j) = \sum \log \left[ \frac{I(i, k) + I(j, k)}{I(i, k)^{1/2} I(j, k)^{1/2}} \right] \quad (50)$$

$$R^{m-1}(i, j) = \sum_{k \in K} \frac{[\hat{R}_\Delta^{m-1}(i, k) - \hat{R}_\Delta^{m-1}(j, k)]^2}{R^{m-1}(i, k) R^{m-1}(j, k)} \quad (51)$$

where  $Z$  is a normalization parameter,  $h_0$  and  $h_1$  control the decay of the weights,  $K$  is a patch with pixel  $i$  (or  $j$ ) as center and  $I(i, k)$  is the  $k$ -th neighbor of  $I(i)$  in  $K$ .

The multi-temporal version of the iterated NLM for SAR images is a two steps algorithm [30]:

**Step 1.** Each channel is filtered by iterative NLM. Let be  $\hat{R}_1, \dots, \hat{R}_N$  the filtered channels. An improved image of a certain channel  $I_n$  is obtained by combining stable pixels in time while

keeping unchanged the pixels not in accordance with the other dates. This is done by deriving a binary mask for each couple  $(\hat{R}_n, \hat{R}_m)$ :

$$P_i(n, m) = \begin{cases} 1 & \text{if } \frac{[\hat{R}_n(i) - \hat{R}_m(i)]^2}{\hat{R}_n(i)\hat{R}_m(i)} > T \\ 0 & \text{otherwise} \end{cases} \quad (52)$$

where  $T$  is a threshold. The improved channel is the weighted average:

$$I_n^{improved} = I_n + \sum_m P(n, m) \hat{R}_m \quad (53)$$

**Step 2.** The improved channel is filtered by the iterated NLM. Denoising on the temporally weighted average image  $I_n^{improved}$  is comparable to (48). However, the pixels may have different (equivalent) number of looks depending on the number of averaged data. In this case the similarity  $S(i, j)$  between pixels has to be modified to take into account varying number of looks.

## 6. Conclusions

The Bayesian filters for SAR images are based on speckle statistics and, when possible, on scene distribution. The experiments have shown that none of the Bayesian filters arrives to reconstruct perfectly a scene. Generally, the result is a tradeoff between the speckle reduction and texture and edges fading out. A crucial factor is the quantity of a priori information included in the statistical models used by the Bayesian filters. A more complex model gives better chances in scene reconstruction but increases the computation time.

The single channel filters – Kuan, Lee and Frost – presented in Section 4.1 are all linear estimates optimizing the quadratic cost function. In the homogenous areas, they estimate the radar reflectivity in the same manner i.e., by calculating the mean of the pixels in the filtering window. The challenge is however represented by the edges and textures, where each filter gives another result. The best of them is Frost filter that includes in the statistical model the reflectivity correlation. The worst is Lee filter, which transforms the speckle into an additive noise by truncating a Taylor series.

The multi-temporal counterparts of the Bayesian filters improve the results but do not solve entirely the scene reconstruction problem. The reflectivity estimation is better due to the highest number of samples and mainly due to inter-channel speckle independence. Meantime, other specific problems arise, like the mixing of the features from different channels (in multi-temporal sets, the textures generally change from a channel to another because of the delays between acquisitions).

The filters for multi-temporal sets are using the same approaches as for single channel reconstruction. Additionally, they include in the statistical models the inter-channel characteristics i.e., speckle independency and variation of the radar reflectivity. This supplementary information contributes to the quality of the reconstruction result.

The filter for multi-temporal sets in Section 5.1 is a linear filter optimizing the quadratic cost function, like Kuan, Lee or Frost filters. Differently from these filters, the multi-temporal estimate is a linear combination of the corresponding pixels in all the channels. By introducing the hypothesis of variable textures, this filter arrives to reconstruct not only the homogenous zones but also the textured ones. The drawback is however its high complexity.

The Time-Space filter is basically the filter of Lee applied in the space of DCT. The Time-Space filter cleverness is the transformation solely on the temporal coordinate. As a consequence, the speckle, which is white, is uniformly spread in all the frequency planes while the radar reflectivity is concentrated in a single plane. By filtering all the planes excepting the one concentrating the reflectivity and by restoring the temporal channels by inverse DCT, the scene appears well reconstructed also in its fine details. For six 3-looks amplitude images, the Time-Space filter reconstructs the temporal channels with approximately 16 equivalent number of looks.

The multi-temporal NLM is a promising but not yet mature method. The NLM estimator, which gives good results in single image filtering, has definitely a higher potential in the context of multi-temporal sets. Some preliminary results on 6 six TerraSAR images show a good reconstruction of the fine details.

## Author details

Daniela Colțuc

University Politehnica of Bucharest, Romania

## References

- [1] Waske B. and Braun, M. Classifier ensembles for land cover mapping using multi-temporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2009; 64, 450–457.
- [2] Wang D. et al. Application of multi-temporal ENVISAT ASAR data to agricultural area mapping in the Pearl River Delta. *International Journal of Remote Sensing*. 2010; 31, 1555–1572.
- [3] Zhang Y. et al. Mapping paddy rice with multi-temporal ALOS PALSAR imagery in southeast China. *International Journal of Remote Sensing*. 2009; 30, 6301–6315.

- [4] Goodenough et al. Multitemporal evaluation with ASAR of Boreal forests. Proceedings of Geoscience and Remote Sensing Symposium IGARSS 2009, 12-17 July 2009, Seoul, Correa.
- [5] Maged M, Mazlan H. and Farideh M. Object recognitions in RADARSAT-1 SAR data using fuzzy classification. *International Journal of the Physical Sciences* 2011; 6(16) 3933 – 3938.
- [6] Maged M. and Mazlan H. Developing adaptive algorithm for automatic detection of
- [7] al linear features using RADARSAT-1 SAR data. *International Journal of the Physical Sciences* 2010; 5(14), 2223-2229.
- [8] Maged M. RADARSAT for Oil Spill Trajectory Model. *Environmental Modelling & Software* 2004; (19), 473-483.
- [9] Maged, M. Radar Automatic Detection Algorithms for Coastal Oil Spills Pollution. *International Journal of Applied Earth Observation and Geo-information* 2001; (3)(2), 191-196.
- [10] Yousif O. and Ban Y. Improving Urban Change Detection from Multitemporal SAR Images Using PCA-NLM. *IEEE Tr. on Geoscience and Remote Sensing.* 2013; 51(4) 2032 – 2041.
- [11] Thiel C. et al. Analysis of multi-temporal land observation at C-band. Proceedings of Geoscience and Remote Sensing Symposium IGARSS 2000, 24-28 July 2000, Honolulu, Hawaii, USA.
- [12] Satalino G. et al. Wheat crop mapping by using ASARAP data. *IEEE Tr. on Geoscience and Remote Sensing.* 2009; 47 527–530.
- [13] Martinez, J.M. and Toan, T.L. Mapping of flood dynamics and spatial distribution of vegetation in the Amazon flood plain using multitemporal SAR data. *Remote Sensing of Environment*, 2007; 108, 209–223.
- [14] Maître H. et al. Traitement des images de radar à synthèse d'ouverture. Hermes Science Europe Ltd. 2004.
- [15] Oliver C. and Quegan S. Understanding Synthetic Aperture Radar Images with CDROM. SciTech Publishing; 2004.
- [16] Kuan D et al. Adaptive restoration of images with speckle. *IEEE Transactions on Acoustics, Speech and Signal Processing* 1987; 35(3) 373-383.
- [17] Lee J-S. et al. Speckle filtering of synthetic aperture radar images: A review. *Remote Sensing Reviews* 1994; 8(4) 313-340.
- [18] Frost et al. A model for radar images and its implications to adaptive digital filtering of multiplicative noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2009; 4, 157–166.

- [19] Lopes. et al. Structure detection and statistical adaptive speckle filtering in SAR images. *Int. J. Remote Sens.*,1993; 14(9) 1735–1758, 1993.
- [20] Yu Y. and Acton S.T. Speckle reducing anisotropic diffusion. *IEEE Trans. Image Processing*, 2002; 11(11) 1260–1270.
- [21] Nicolas J.M., Tupin F. and Maitre H. Smoothing speckled SAR images by using maximum homogeneous region filters: An improved approach. *Proceedings of Geoscience and Remote Sensing Symposium IGARSS'01*, 9-13 July 2001, Sydney, Australia.
- [22] Feng H., Hou B. and Gong M. SAR image despeckling based on local homogeneous-region segmentation by using pixel-relativity measurement. *IEEE Trans. Geosci. Remote Sens.* 2011; 49(7) 2724–2737.
- [23] Bruniquel J. and Lopes A. Analysis and enhancement of multi-temporal SAR data. *Satellite Remote Sensing. International Society for Optics and Photonics* 1994; 342-353.
- [24] Bruniquel J. and Lopes A. Multi-variate optimal speckle reduction in SAR imagery. *International journal of remote sensing* 1997 18(3) 603-627.
- [25] Coltuc D. et al. Time-space filtering of multitemporal SAR images. *Proceedings of Geoscience and Remote Sensing Symposium IGARSS 2000*, 24-28 July 2000, Honolulu, Hawaii, USA.
- [26] Coltuc D., Trouvé E. and Bolon Ph. Bias correction and speckle reduction in time-space filtering of multi-temporal SAR images. *Proceedings of Geoscience and Remote Sensing Symposium IGARSS'01*, 9-13 July 2001, Sydney, Australia.
- [27] Coltuc D. and Radescu R. On the homomorphic filtering by channels' summation. *Proceedings of Geoscience and Remote Sensing Symposium IGARSS'02*, 24-28 July 2002, Toronto, Canada.
- [28] Deledalle C-A., Denis L. and Tupin F. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Transactions on Image Processing* 2009;18(12) 2661-2672.
- [29] Gineste P. A simple, efficient filter for multi-temporal SAR images. *International Journal of Remote Sensing*, 1999; 20, 2565–2576.
- [30] Quegan S. and Yu J.J. Filtering of multichannel SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 2001; 39, 2373–2379.
- [31] Su X. et al. Two steps multi-temporal non-local means for SAR images. *Proceedings of Geoscience and Remote Sensing Symposium IGARSS 2012*, 22-27 July 2012, Munich, Germany.



---

# Oil-Spill Pollution

## Remote Sensing by Synthetic Aperture Radar

---

Yuanzhi Zhang, Yu Li and Hui Lin

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/57477>

---

### 1. Introduction

Oil-spill, one of the major marine ecological disasters, can result in enormous damage to the marine environment and great difficulties to clean-up operations. In 2002, oil leaked from the sunken tanker Prestige off the coast of Galicia, polluted thousands of kilometers' coastline. In April 2010, the explosion of British Petrol's drilling rig Deepwater Horizon (DWH) in Mexico Gulf caused the largest accidental oil-spill throughout history. During 87 days of oil leaking from a sea-floor gusher, approximately 780,000 m<sup>3</sup> of oil, methane or other petro products was flowed into the Atlantic Ocean. In 2011, Bohai Sea, oil-spill was caused by the crack of submarine fault, triggered by the over pressure caused by oil extraction operation of Conocophillips Petro, China. Approximately 700 barrels of crude oil was leaked into the sea, and there were also about 2500 barrels of mineral oil-based mud deposited on the sea bed. Besides, a large proportion of oil-spills were caused by deliberate discharges from tankers or cargos, for the reason that there are still some vessels used to clean their tanks or engines before entering into the harbor. All these accidents and illegal acts caused huge damage to the coastal ecosystem and marine environment. As a result, early warning of oil-spill accidents by remote sensing is very important for coastal environmental protection and has become a vital task for maritime surveillance.

Optical sensors can be used for the application of oil-spill detection, however they are unavoidably suffered from weather and light conditions. For example during the Qingdao oil pipe accident that took place in Nov. 2013, crude oil flowed into the bay through drainage channel. Heavy smoke caused by the explosion covered the whole scene, which largely hampers the application of optical based sensors. Moreover there are also a lot of oil-spill accidents takes place during night or stormy weather. Due to its advantage of wide areal coverage under all-weather conditions the imaging capability during day-night times (Gade

and Alpers, 1999), satellite synthetic aperture radar (SAR) data from ERS-1/2, ENVISAT, ALOS, RADARSAT-1/2, and TerraSAR-X have been widely used to detect and monitor oil-spill (e.g., Alpers & Espedal, 2004a; Migliaccio *et al.*, 2007 & 2009; Topouzelis *et al.*, 2008 & 2009; Marghany & Hashim, 2011; Zhang *et al.*, 2012). Airborne SAR sensors like Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR, by JPL, L-band) and E-SAR, (by DLR, multi-land) have also proven their potential for scientific research of marine or land remote sensing. With a series of key breakthroughs of SAR remote sensing technology, especially the multi-polarimetric capability and increased resolution, oil-spill detection by SAR has become a very hot research area (Solberg, 2012).

Based on basic principles of marine SAR remote sensing, this chapter will discuss state-of-the-art technologies of SAR oil-spill detection, and introduce several ways of investigating the environmental impact of oil pollution to coastal environment. Several new trends within or beyond current applications, which will improve the performance of SAR oil-spill detection and provide more concrete information for studies in the near future, will also be introduced.

## 2. Basic principles of SAR oil-spill detection

SAR is an active radio equipment which transmit microwave and record the signal scattered by the target. The main mechanism of SAR marine remote sensing is observing the interaction between microwave and short-gravity and capillary waves of the sea surface. The backscattered signal from the sea surface can be described by backscattering coefficient  $\sigma^0$ , which stands for the uniformed radar cross-section.

At small incidence angle ( $<20^\circ$ ), the curvature radius are large enough compared with the wavelength of radar signal, specular scattering can be used to describe the scattering process and  $\sigma^0$  can be solved by Kirchoff approximation. At middle and large incidence angle ( $>30^\circ$ ), the backscattered signal is dominated by Bragg scattering:

$$\lambda_B = \frac{\lambda_r}{2\sin\theta} \quad (1)$$

where  $\lambda_B$  stands for the wavelength of Bragg scattering,  $\lambda_r$  stands for the wavelength of radar signal,  $\theta$  stands for the incidence angle. Currently small perturbation model (SPM) is always used for calculating the scatter coefficient caused by Bragg scattering.

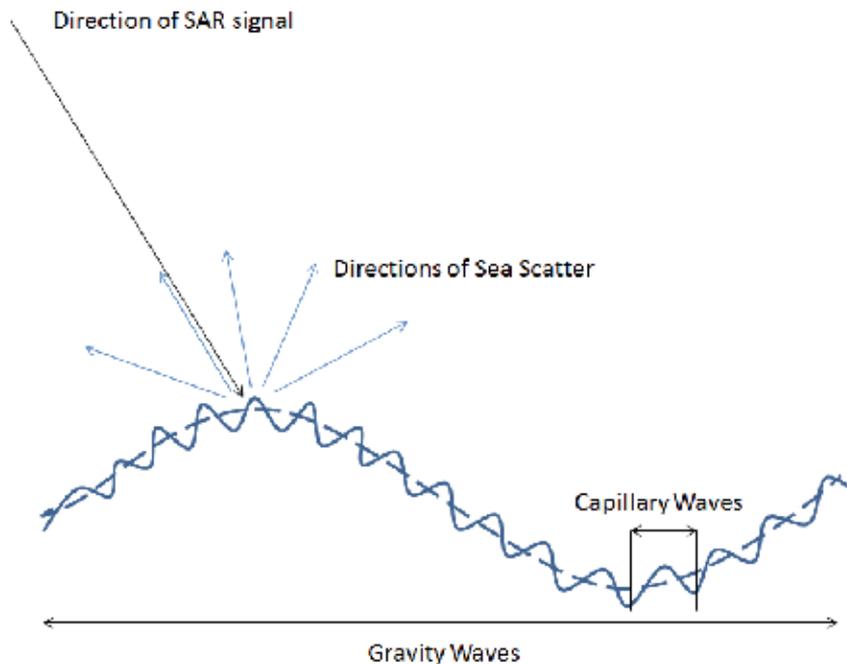
According to the composite sea surface model, the roughness of sea surface can be seen as small scale capillary waves superimposed on large scale grave waves (Fig. 1). So the backscattered radar signal can be treated as Bragg scattering modulated by the tilted scattering surface caused by large scale gravity waves. In mathematic expression, there are:

$$\sigma_0 = \sigma_0^S + \sigma_0^B \quad (2)$$

where  $\sigma_0^S$  and  $\sigma_0^B$  stand for backscatter coefficient calculated by model of specular and Bragg scattering respectively.

Since ancient Greek, physicists have noticed the damping effect of spilled oil films on the rough sea surface (Aristotle, *Problematica Physica*). Besides, some experienced sailors knew that pour oil to the sea surface could reduce the turbulence of sea surface and used this method to avoid ships from sinking in stormy seas. However systematic theory was not established to explain this phenomenon until Italian scientist Marangoni (Marangoni, 1872) pointed out that matter with different viscosity presents on the surface of fluid will produce elastic resistance to the movement of the surface, hence reduce the surface wave intensity.

The damping of oil films on the short-gravity and capillary waves can be measured by the ratio between backscatter coefficients of oil covered area and background sea surface. It is worth noting that the detectability of oil-spill by SAR relies closely on the wind speed above sea surface: if the sea surface wind speed is too slow, sea wave cannot be well developed and if it is too large, spills will break and be dispersed by mixing with seawater. So normally the ideal wind speed for oil spill detection is between 3m/s to 14m/s.



**Figure 1.** Demonstration of radar signal scattering from sea surface.

### 3. SAR Sensors and systems for oil-spill detection

Table 1 lists several main spaceborne SAR platforms. It can be found that multi-polarimetric capability has become a trend of advanced SAR satellites nowadays. And a series of SAR satellites constellations in plan will provide shorter revisit time for maritime monitoring in the near future. Discussion on the optimum frequency and polarization for SAR oil-spill detection can refer to Solberg's (2012) and other people's studies. Generally speaking, VV polarization performs better in highlighting oil slicks from sea background for stronger power return, however it may vary according to different oil types and sea conditions.

Sensors	Frequency	Polarization	Resolution* <sup>1</sup>	Swath width* <sup>1</sup>	Revisit/Repeat time* <sup>3</sup>
ERS 1/2 (Not operational)	5.3 GHz (C-band)	VV	4*20 m	80~100 km	35 d (Repeat)
Radarsat-1 (Not operational)	5.3 GHz (C-band)	HH	8~100m	50~500 km	2/24 d
Envisat-ASAR (Not operational)	5.331 GHz (C-band)	HH/HV; VV/VH; VV/HH	4*20, 150m	100, 400 km	2~3/35d
ALOS PALSAR (Not operational)	1.27 GHz (L-band)	HH, HV, VH, VV	7~100m	20~350 km	2 /46 d
TerraSAR-X	9.65 GHz (X-band)	HH, HV, VH, VV	1~16m	5~100 km	2 /11d
Radarsat-2	5.4 GHz (C-band)	HH, HV, VH, VV	3~100m	20~500 km	1/24d
Cosmo-SkyMed Constellation	9.6 GHz (X-band)	HH/HV; VV/VH; VV/HH	1~100m	10 km	1/16 d
HJ-1C SAR Constellation (1 /4)* <sup>2</sup>	3.2 GHz (S-band)	VV	5~20m	40,100 km	1/4d
Sentinel-1 Constellation (2 In plan)	5.4 GHz (C-band)	HH;VV; HH/HV; VV/VH;	5~20m	80,250,400 km	1~3/12d (6d with two satellites)
ALOS-2 (In plan)	1.3 GHz (L-band)	HH, HV, VH, VV	1~100m	25~490 km	14 d (Repeat)

\*<sup>1</sup>Resolution (approximate) and swath width may vary with operation modes;

\*<sup>2</sup>In operation/Total planed;

\*<sup>3</sup>Revisit time (general): revisit at different incidence angle; Repeat time: orbit revisit, applicable for InSAR.

**Table 1.** List of main spaceborne SAR sensors that can be used for oil pollution monitoring.

Airborne SAR sensors are also useful for quick respond of oil-spill accidents. The Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR), designed by Jet Propulsion Laboratory (JPL) at California Institute of Technology, US, is a reconfigurable, polarimetric L-band synthetic aperture radar with a 22-km-wide ground swath at 22° to 65° incidence angles. At present stage it is mounted beneath a Gulfstream III jet for experimental missions. The sensor has shown great performance in studies of land deformations, vegetation, ice & glaciers, and of course, oceanography. E-SAR is an airborne experimental SAR system operated by the German Space Agency (DLR). Since delivered its first images in 1988, the system has been continuously upgraded to become a multi-band and multi-polarimetric SAR sensor for a variety of Earth observation applications. In 2011, China successfully developed airborne multi-band, multi-polarimetric SAR interferometric mapping system, which includes radar systems and platform, data processing & mapping software and data distributing system.

There are several oil-spill monitoring systems in operation, among which CleanSeaNet satellite monitoring service is probably of the most famous one. It was set up and operated by the European Maritime Safety Agency (EMSA) since April 2007 and has developed to its 2<sup>nd</sup> generation. It takes advantage of images from a series of SAR and optical sensors. Together with auxiliary information such as meteorological and oceanographic data, the service could detect and report oil-spill accidents in near real time. Canadian Ice Service launched another real-time operational program called ISTOP (Integrated Satellite Tracking of Pollution), which uses the RADARSAT-1 data to monitor ocean and lakes for oil slicks and tracking the pollutants (Gauthier, 2007). Some software have also been developed such as semi-automatic SAR oil-spill detection system developed by Konsberg Satellite Services, Norway, and Ocean Monitoring Workstation developed by Satlantic Inc. Canada. State Oceanic Administration of China also developed SAR satellites based oil-spill monitoring system and used it to regularly monitor Bohai Sea since 2009.

#### **4. SAR automatic oil-spill detection algorithms**

Synthetic Aperture Radar is an active microwave remote sensing device, which takes advantage of relative motion between its antenna and the target to achieve higher spatial resolution. As mentioned before, the oil-covered region appears smoother than its surrounding sea surface. In other words, the Bragg scattering in these areas is weakened. And in SAR image this phenomenon is usually observed as dark patches. However, in SAR images, the backscattered signal from oil-spill are very similar to the values of the backscattering from calm sea areas and other ocean phenomena called “look-alikes”. In the following part, explanation and examples of some main look-alikes are provided:

Low wind area: area where the surface wind is very slow (<2m), such as areas sheltered by land. Low wind swayed sea surface acts just like a mirror, which reflect most radar signal to the opposite direction of the receiving antenna. So they are observed as low backscattered area.

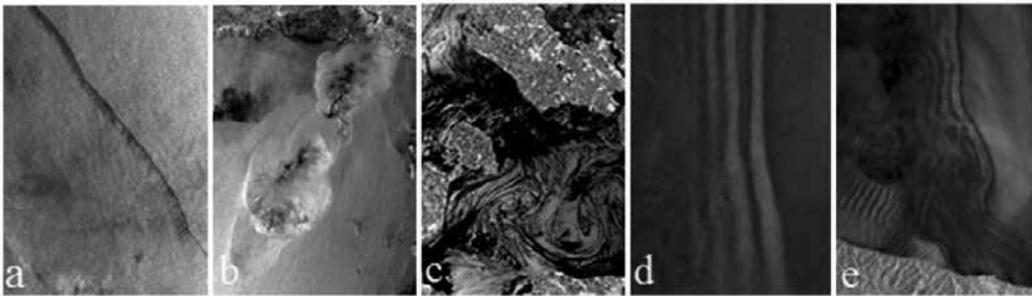
Biogenic films: natural films produced by phytoplankton and fishes are commonly found on the sea surface. Most of them are a very thin layer of surfactants with one to several molecules,

which reduces the surface tension dramatically. As a result, they are always observed as dark area.

**Rain cells:** also known as convective rain, which is the common form of rain in the tropics and subtropics (Alpers, 2004b). The typical structure of rain cell has downdraft as its core with circular gust front surrounded. The gust front will produce strong wind and always increase the sea roughness while the downdraft sometimes dampens the Bragg waves in form of precipitation. So some rain cells may have the signature of a darker core with much brighter surroundings compared with surrounded sea surface in SAR images (Alpers, 2004b).

**Oceanic Internal Waves (OIW) and Atmospheric Gravity Waves (AGW):** both of them have linear structures with bright and dark strips appear alternatively in SAR images. They can be distinguished based on solitary wave and radar imaging theories (Alpers et al. 2011).

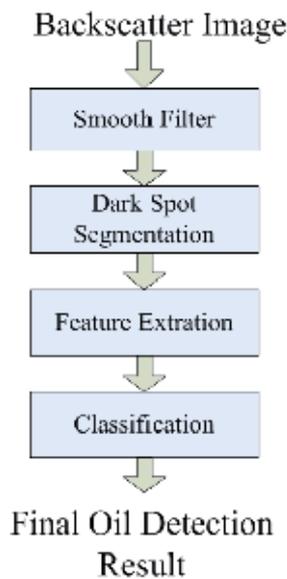
Besides, ocean physical phenomena such as grease ice and upwelling zones may also cause dark spots or zone in SAR images. Together with a stripe-like oil-slick, typical images of the above-mentioned look-alikes are provided in Fig. 2. However, in a lot of cases, their characteristics such as shape or texture may be very close to each other, which makes them very difficult to be distinguished using any simple criteria alone.



**Figure 2.** Examples of several main look-alikes in SAR images: a) typical oil (Photo: Alpers et al. 2004a); b) rain cell (Photo: ESA); c) biogenic slicks (Photo: Alpers et al. 2004a); d) OIW (Photo: Alpers et al. 2011); e) AGW (Photo: Alpers et al. 2011).

As the result, in operation it is very necessary to identify real oil slicks and look-alikes in SAR images based on comprehensive analysis of their different characteristics. Compared with manual inspection, automatic or semi-automatic techniques yield better efficiency and stability, hence they have been extensively studied in recent years. The standard framework of oil-spill detection algorithms is shown in figure 3, which mainly includes steps of pre-processing (calibration, speckle filtering), dark spot segmentation, feature extraction, as well as classification.

Data pre-processing is a very fundamental procedure for improving the accuracy of oil-spill detection. Many SAR image products (ground range projected products) are descriptions of radar brightness of targets in the scene, with elevation antenna patterns and range spreading losses corrected but no angle compensation. So sometimes it is also important to remove the



**Figure 3.** Flowchart of classic oil spill detection algorithms

effect of incidence angle variance before further processing (Mladenova, et al. 2013). Then images will be geo-coded to the geographical locations in the study area, for the convenience of further Geographic Information System (GIS) based study and applications.

SAR images are unavoidable suffered from speckle noise due to its coherent processing during image formation. A lot of studies have been done on segmenting dark patches that likely represent oil-slicks from the sea background. Both adapted and non-adapted threshold algorithms have been applied in dark patch detection. Solberg et al. (1999) set the threshold to  $k$  dB below the mean value of the moving window and used a multi-scale pyramid approach and clustering step in the calculation. Shu et al. (2010) made use of spatial density feature to separate dark spots and the background. Migliaccio et al. implemented oil-spill processing over single-look SAR images based on a physical model (Migliaccio et al. 2005) and used constant false alarm rate (CFAR) (Migliaccio et al., 2007a) filter to reduce the speckle within homogeneous areas without loss of information. Huang et al. (2005) used the level-set method to carry on oil-spill detection and accurately extracted oil-slicks from SAR images. However the traditional level-set methods are much more time-consuming. Zhang et al. (2012) jointly used level-set method and wavelets to provide a fast and accurate segmentation algorithm for SAR images.

In order to distinguish real oil-slicks between look-alikes, a variety features can be extracted from SAR images. Some mainly used features are listed in table 2. Polarimetric features have also been used to improve the accuracy of oil-spill classification in recent years and they will be introduced in details in the next section. It is also highly necessary to select the most suitable features for oil-slick classification. However the optimized feature sets will vary according to actual situations and also related to classification algorithm adopted.

Intensity	Morphology	Texture*	Context
Slick Backscattering ( $\mu_{obj}$ )			
Slick Std. Deviation ( $\sigma_{obj}$ )	Area (A)	Homogeneity	Distance to Coast
Surrounding Backscattering ( $\mu_{sce}$ )	Perimeter (P)	Contrast	Distance to Nearest Dark
Intensity Ratio ( $\mu_{obj}/\mu_{sce}$ )	Complexity (C)	Dissimilarity	Patch
Intensity St. Dev Ratio ( $\sigma_{obj}/\sigma_{sce}$ )	Asymmetry	Entropy	Number of Surrounding
ISRI ( $\mu_{obj}/\sigma_{obj}$ )	Euler number	Mean value	Dark Patches
ISRO ( $\mu_{obj}/\sigma_{sce}$ )	Shape Index	Std. Deviation	Number of Surrounding
Min Slick Value (MSV)	Axis Length	Correlation	Ships
Max Contrast ( $\sigma_{sce}$ -MSV)	Compactness		
Edge Gradient			

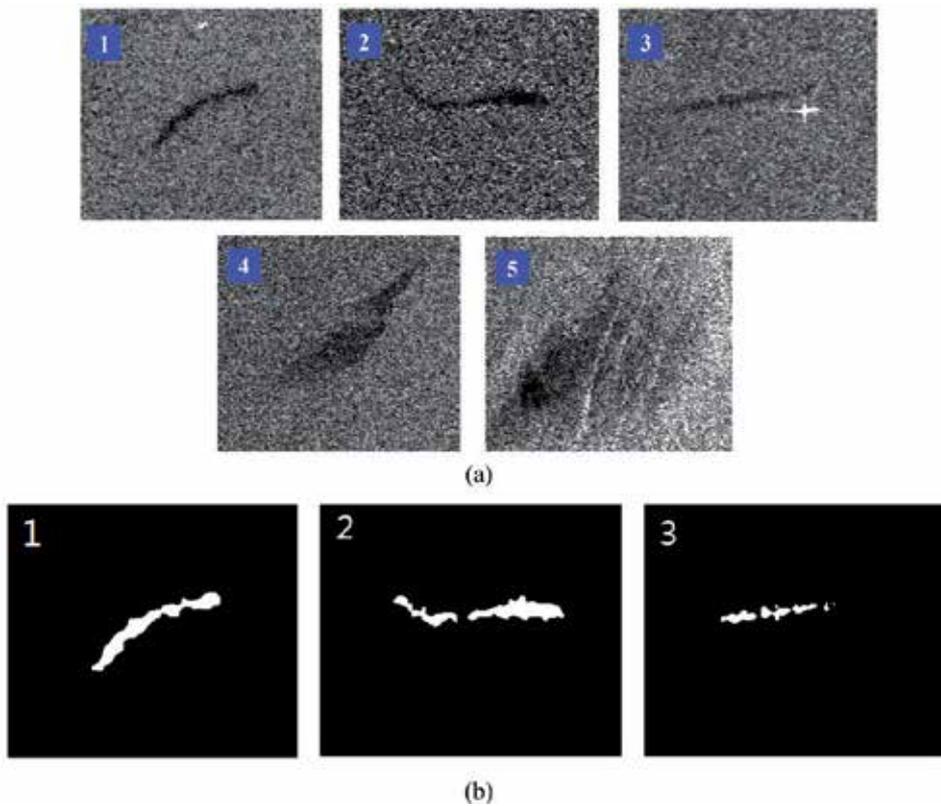
\* Gray-Level Co-occurrence Matrix (GLCM).

**Table 2.** Mainly used characteristics for SAR oil-spill detection.

An example of automatic oil-spill detection using dual-threshold segmentation and characteristic possibility function is provided as follow:

The ground satellite station of CUHK received a scene of Envisat ASAR image at Hong Kong coast region on June 5<sup>th</sup>, 2007, in which several significant dark spots (Fig. 4a) can be observed. Double-threshold segmentation was firstly implemented on the received SAR data to extract both high and low levels of grayscale information from the SAR backscatter image. By using high level segmentation result, look-alikes with large area are classified from oil-spills by basic morphological analysis. By taking advantage of low threshold grayscale information, other look-alikes were distinguished from oil-spill by means of probability likelihood function derived from morphological characteristics such as complexity, length to width ratio, Euler number, etc.. Finally, the detected spills were obtained by fusing classification result of different level and other auxiliary information. The detected oil-spills are shown in Fig. 4b. From which it can be seen that using basic morphological characteristics such as complexity and length to width ratio, oil-slicks and potential look-alikes can be distinguished. However in some other cases, it is not that easy to achieve high accuracy oil-spill classification. So the studies on selecting proper characteristic as well as developing highly efficient classification algorithms are extremely important for automatic oil spill classification. And sometimes auxiliary data such as wind speed, the location of drilling platforms, ships and ship tracks, will be helpful to verify possible look-alikes.

Marghany (2001) developed a model to discriminate textures between oil and water by using co-occurrence textures. Based on Bayesian or other statistical decision, some classification algorithms were developed and applied to oil-spill detection. But the complex process of these algorithms makes it difficult to define the classification rules, since there are a lot of nonlinear and poorly understood factors in the algorithms (Topouzelis et al., 2008a; Gambardella et al., 2010). One alternative is to train the classifier using only samples of oil-spill, instead of using oil-spill and look-alikes, which is called one-class classification. Gambardella et al. (2008) proposed and used this technique with optimized feature selection algorithm and obtained

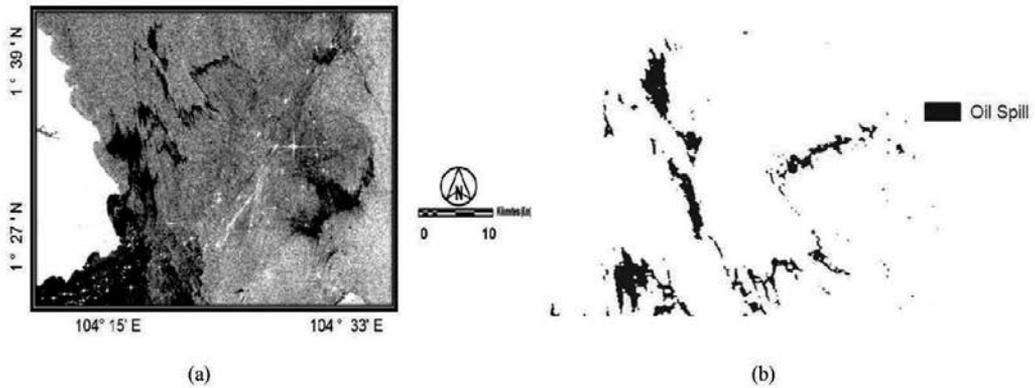


**Figure 4.** (a) SAR image of Hong Kong coastal region on June 5th 2007, (b) Result of automatic oil spill detection.

promising results with two case study datasets. By employing probability distribution formula, Marghany et al. (2009; 2011a) proposed fractal box counting based algorithms for oil-spill classification and further proved that this technique has promising performance in discriminating oil-spill from look-alikes on RADARSAT-1 and AIRSAR/POLSAR data.

Another alternative is taking advantage of Artificial Neural Network (ANN). Neural networks have been largely investigated and recognized as a robust tool for classification. Frate et al. (2000) proposed a semi-automatic detection of oil-spill by neural network of which the input is a vector describing features of an oil-spill candidate. It was proved that the neural network could correctly discriminate oil-spill and look-alikes over a set of independent samples. Topouzelis et al. (2007) used neural networks in both dark formation detection and oil-spill classification. During their experiment 94% of the dark formations segmentation and 89% accuracy of classification were obtained respectively. Topouzelis et al. (2009) also carried out a detailed robustness examination of the combinations derived from 25 common used features. They found that a combination of 10 features yields the most accurate results. Marghany et al. (2011b) compared three algorithms for oil-spill classification including co-occurrence textures, post supervised classification and neural network. Quantitative study on standard deviation of the estimated error shows that the artificial neural network has the largest accuracy among

all these methods. Garcia-Pineda et al. (2013) developed the Textural Classifier Neural Network Algorithm (TCNNA) to map oil-spill in Gulf of Mexico Deepwater Horizon accident by combining Envisat ASAR data and wind model outputs (CMOD5) using a combination of two neural networks. Recently, Marghany (2013) used genetic algorithm (GA) for automatic detection of oil-spill from ENVISAT ASAR data. Experimental results proved that the cross-over process and fitness function used in the study could generate accurate pattern of oil slick, and the results are shown in Fig. 5.

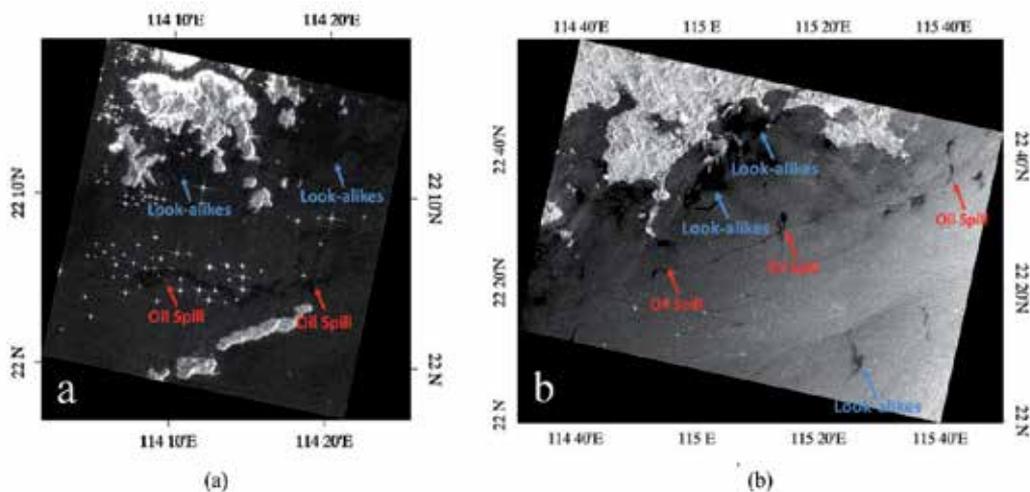


**Figure 5.** (a) ENVISAT ASAR data acquired on June 3rd 2010; (b) Oil spill automatic detection result by Genetic Algorithm (GA) (Marghany, 2013)

Although neural network has proven its advantage in oil-spill monitoring for SAR images, it is time consuming. Besides, the probability of misclassification does not always decrease as the number of features increases, especially when sample data is insufficient. This phenomenon is known as “the curse of dimensionality” (Jain et al., 2000). In a study (Li, et al., 2013), Support Vector Machine (SVM) was developed to automatically detect oil-spill from SAR images. A dual-threshold segmentation was implemented on SAR backscatter images to extract information in different grayscale levels. In the classification phase, a linear SVM was firstly built according to training data set. Then the constructed SVM was used to distinguish probable oil slicks and look-alikes based on morphological characteristics extracted from the dilated low threshold segmentation result. Final detection result was obtained by fusing the classification result of both low and high threshold segmentations. Experiment was carried out on two SAR images received in three days consecutively by the ground station of remote sensing satellite, CUHK. The first image received on May 19<sup>th</sup> was treated as training data sample (figure 6a). Oil-spills and look-alikes were artifically classified and morphological features were extracted to built a SVM. The second image (figure 6b) obtained three days later was used to examine the performance of the proposed algorithm. Experimental results illustrates that this new procedure is able to effecively detect oil-slicks, including small-size ones in coastal region based on prior morphological characteristics. In the comparison between characteristic possibility function (CPF) based method, the SVM method has better

performance in reducing both type I (incorrect rejection) and type II (incorrect acceptance) error. And in the comparison with ANN based methods, lower false alarm rate was obtained.

With years of study, the accuracy and efficiency of automatic oil-spill detection and classification has been largely improved. However at present stage, they are not very much likely to operate fully alone without the assistance and guidance of human experts. That is because there are always some unpredictable phenomenon observed in SAR images, which is new to the sample case based automatic recognition system, calls for comprehensive analysis based on related knowledge and auxiliary data support.



**Figure 6.** (a) ENVISAT ASAR image received on 10:13 a.m., May 19<sup>th</sup> 2010, Hong Kong coastal region, in which oil slicks lie in the lower middle part, along the main ship lane, so they are probably caused by deliberate discharges of ships. Look-likes are also found in the upper part of this image, where the wind are blocked by islands; (b) ENVISAT ASAR SAR image near Shanwei's coast on 10:18 a.m. 22nd May, 2010. It could be calculated that during three days the main part of the slick has drifted for almost 80 km, and it is largely stretched by the combined effect of wind and sea current. There are also several wind shelter look-alikes (upper) and look-alikes probably caused by biogenic film (lower right).

## 5. Trends of the studies on SAR oil-spill detection

### 5.1. Polarimetric SAR for oil-spill detection

Sometimes oil slicks and look-alikes can be well distinguished using above mentioned characteristics, however these characteristics of oceanic phenomenon in SAR images may vary according to specific conditions, which may results in misclassification. For example in some cases biogenic films holds very similar shape and texture as mineral oil. So the largest challenge that automatic oil-spills detection facing is to reduce the false alarm rate. Polarimetric SAR capabilities of new SAR sensors will help reduce false targets in SAR imagery (Brown and

Fingas, 2003). The feasibility of polarimetric SAR oil-spill classification relies on the fact that the polarimetric mechanisms for oil-free and oil-covered sea surface are largely different (Migliaccio et al., 2010). Before the availability of polarimetric observation, mineral oil and look-alikes such as biogenic slicks are difficult to be distinguished because they damp short gravity-capillary waves with almost the same strength (Alpers, 2002). Based on different polarimetric scattering behaviors, mineral oil and biogenic slicks can be better distinguished: for oil-covered area, Bragg scattering mechanism is largely suppressed and high polarimetric entropy can be witnessed, while in case of a biogenic slick, Bragg scattering is still dominant, but with a lower intensity. So similar polarimetric behavior as oil-free area is expected for biogenic films (Migliaccio et al., 2009). In the following part, polarimetric characteristics that may boost the performance of oil-spill detection will be listed and discussed:

- i. **Cross & Co-Polarization Ratio:** the power ratio between HH and VV or HV and HH/VV channels. They are easy to be derived and can be obtained through dual-pol systems. In tilted Bragg scattering model adopted by Minchew et al. (2012a), they are only functions of dielectric constant and incidence angle. However actual situations are always more complicated and further study is needed. Co-polarization ratio was also proved to be possible to discriminate slicks from look-alike features associated with low-wind conditions and surface current effects (Kudryavtsev et al., 2013).
- ii. **Pedestal Height:** describes the polarization signature. It was tested on L-band ALOS-PALSAR, C-band RADARSAT-2 and C-band SIR-C/X-SAR full-polarimetric SAR data to distinguish oil slicks from weak-damping look-alikes (Nunziata et al. 2011).
- iii. **Co-Polarized Phase Difference (CPD):** describes the complex correlation between HH and VV channel signal. Migliaccio et al. (2009) firstly used the standard deviation of CPD to distinguish oil slicks and biogenic slicks and also pointed out that theoretically it performs better in C-band rather than L-band or lower wavelength.
- iv. **Degree of Polarization (DOP):** a fundamental characteristic of partially polarized EM field (Shirvany, et al. 2012) derived from stokes parameters. It can take the advantage of full-pol, compact or linear dual polarization SAR data and has proven its effectiveness on oil-spill detection by Radarsat-2 and UAVSAR L-band data.
- v. **Pauli Decomposition Parameters:** Entropy  $H$ , Anisotropy  $A$ , Alpha angle  $\alpha$ , Alternative Entropy  $A_{12}$  and  $\lambda_i$  ( $i=1, 2, 3$ ), they are very important parameters in almost all polarimetric analysis. Based on SIR-C/X SAR data, the performance of  $H$  and  $A$  was validated (Migliaccio et al., 2005). Then a polarimetric constant false alarm rate filter was developed to detect oil slicks over SAR images (Migliaccio et al., 2007b).
- vi. **Conformity Coefficient:** an indicator of Bragg scattering (Zhang et al., 2011). It was tested using RADARSAT-2 quad-polarization SAR image of oil slicks in the Gulf of Mexico.
- vii. **Bragg Likelihood Ratio:** Based on the same principle as CPD, Salberg et al. (2012) developed a generalized likelihood ratio test to see whether the Bragg scattering model is followed.

Some of the above mentioned polarimetric characteristics are tested by L-band UAVSAR data obtained during the Gulf of Mexico Deep Water Horizon oil-spill accident. It is noted that for different bands, environmental conditions and types, status of leaked oil, their polarimetric characteristic may be different. Nevertheless with the help of polarimetric information, scattering mechanism of sea surface can be better identified and more accurate classification result could be obtained.

The different polarimetric characteristics obtained by the observation of L-band UAVSAR, in case of DWH oil-spill accident can be perceived from figure 7. Within these parameters: co- and cross-polarization ratio, Degree of Polarization, Entropy  $H$ , Alpha angle  $\alpha$ , Anisotropy  $A$  (within certain incidence angle range), works relatively better in distinguishing oil and sea water; while the Co-polarized phase difference, Conformity coefficient, Co-pol correlation coefficient, less work. Actually according to the discussion in Migliaccio et al (2009), CPD in L-band SAR does not perform well in the classification between mineral and biogenic oil films. Besides, Pol-SAR characteristics are observed varies with incidence angle. It has to be noted that the DWH oil-spill is a very rare case for its extreme amount of oil leaked, so actually the sea is "full" of oil and as first approximation it is possible to consider the scattering surface as a homogeneous and thick oil layer.

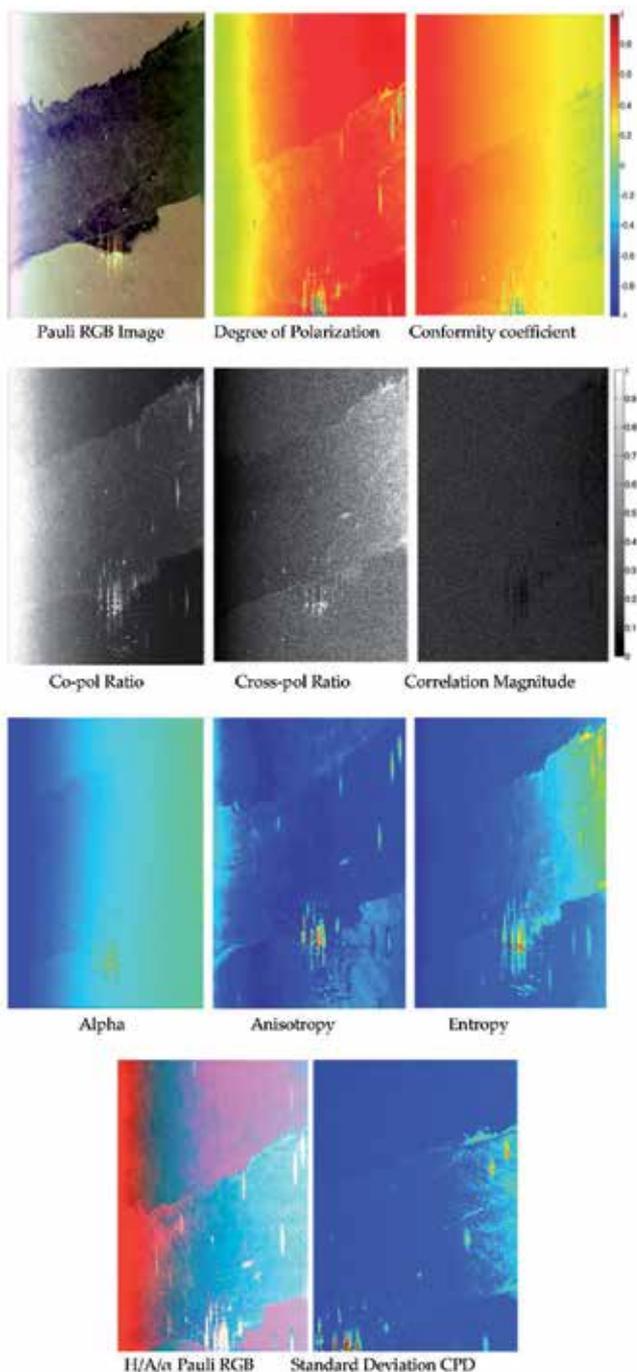
Analysis was conducted by Minchew et al. on the same dataset (Minchew, 2012a), in which it is also found that the major eigenvalue  $\lambda_1$  of oil covered area is constantly lower than that of clean water, making it a very good indicator for oil-spill. The study also highlighted the effect of the noise floor of the radar system to the polarimetric characteristics, especially those closely related to cross-channel signal. Systematic errors such as cross-talk, channel imbalance and thermal noise largely affects this dataset, which is also one of the reason that some polarimetric characteristics varies so much from near to far incidence angle. More precise calibration is required if higher classification accuracy is expected by using this UAVSAR dataset.

Similar study was conducted by Skrunes et. al (2012) during the Norwegian oil-on-water experiment, in which Radarsat-2 C-band and TerraSAR-X X-band data was used. The largest difference in findings between Skrunes' experiment and the above analysis is that for C and X band, CPD and correlation magnitude works well in the classification between mineral and biogenic oil.

Moreover, because of the complexity of sea surface polarimetric scattering mechanisms, it is unrealistic to consider using one single characteristic to distinguish variety kinds of oil-spill under different conditions. As the result, a synthetic and proper use of the aforementioned or other polarimetric characteristics is the key to accurate detection and successful interpretation of oil slicks.

## 5.2. Compact polarimetric SAR for oil-spill detection

Although being proved helpful in oil-spill classification, full polarimetric SAR is facing the challenges of system complexity and reduced swath width caused by doubled Pulse Repetition Frequency. Compact polarimetric (CP) modes have the advantage of un-halved swath width and comparable polarimetric capabilities compared with full polarimetric mode, which has



**Figure 7.** Polarimetric characteristics of DWH oil-spill: the horizontal direction in SAR images is accordance with range direction from the sight of the sensor, and vertical side stands for azimuth direction.

become a new study trend (Souyris et al., 2005; Chen et al., 2011; Nord et al., 2009). In the consecutive years after this technique was proposed, most of the studies were on its applications of land monitoring, e.g., biomass and soil moisture estimation (Dubois-Fernandez et al., 2008). Until recent years, it has been considered in maritime surveillance applications (Yin et al., 2011; Collins et al., 2013) but there are very few studies on oil-spill detection using CP SAR. Actually for sea surface the scattering mechanisms are relatively simpler than that of ground surface, so it is easier to extract and analyze useful information from CP SAR data of sea surface.

In case of fully-polarized SAR data, the backscattering properties of an object under certain condition can be described through a backscattering matrix as:

$$S = \begin{pmatrix} S_{HH} & S_{HV} \\ S_{HV} & S_{VV} \end{pmatrix} \quad (3)$$

where  $S_{xy}$  is the complex backscattering coefficient, with  $x$  denoting the received wave polarization and  $y$  indicating the transmitted wave polarization.

In the compact polarimetric SAR modes, the radar transmits only linear combination of horizontal and vertical ( $\pi/4$ ) or circularly (CTLR, DCP) polarized signal and linearly (CTLR,  $\pi/4$ ) or circularly (DCP) receives both horizontal and vertical polarizations. The 2-D measurement vector  $\vec{K}$  is the projection of the full backscattering matrix on the transmit polarization state. The measurement vector  $\vec{K}$  of three main compact polarimetric SAR modes can be defined as:

$$\vec{k}_{\pi/4} = [S_{HH} + S_{HV} \quad S_{VV} + S_{HV}]^T / \sqrt{2} \quad (4)$$

$$\vec{k}_{CTLR} = [S_{HH} - iS_{HV} \quad -iS_{VV} + S_{HV}]^T / \sqrt{2} \quad (5)$$

$$\vec{k}_{DCP} = [S_{HH} - S_{VV} + i2S_{HV} \quad i(S_{HH} + S_{VV})]^T / 2 \quad (6)$$

In the data received via CP SAR modes, only part of the polarimetric information is preserved. There are basically two ways to take advantage of CP SAR data. The first one is directly extract and analysis some properties from CP SAR data. Stokes parameters can be calculated in CTLR mode and further polarimetric analysis can be employed (Ranny, 2007). Then some important polarimetric parameters such as DOP, Relative Phase, Entropy, Anisotropy and  $\alpha$  can be derived (Shirvany, et al., 2012; Cloude, et al., 2012). It is noted that the processing method and definitions of some parameters for CP SAR data, in the process of calibration, decomposition and classification, will be different.

The other way is to reconstruct quad-polarimetric matrix from CP SAR data. In order to reconstruct the quad-pol data from the compact polarimetric SAR data using iteration based algorithm, Souyris et al. linked the magnitude of linear coherence and the cross polarization ratio with parameter  $N$  (Souyris et al., 2005). Nord modified Souyris' algorithm by replace  $N$

with  $(|S_{HH} - S_{VV}|^2 / |S_{HV}|^2)$ , and update it during the calculation process iteratively (Nord et al. 2009). Yin et al. proposed a reconstruction algorithm based on polarimetric decomposition and proved its soundness in ships detection (Yin et al. 2011). Collins proposed an empirical model to estimate  $N$  in variance of incidence angles by fitting the observed data with negative exponent function (Collins et al. 2013).

Experiments on JPL-UAVSAR data can be taken to illustrate the feasibility of quad-pol reconstruction with compact polarimetric SAR modes (Li et al., 2013). Compact polarimetric SAR signals are derived from full polarimetric SAR matrix of UAVSAR data. Then Souyris' quad-pol reconstruction algorithms were employed with a slightly modified hypothetical function, to better fit the character of sea surface scattering:

$$\frac{\langle |S_{HV}|^2 \rangle}{\langle |S_{HH}|^2 \rangle + \langle |S_{VV}|^2 \rangle} = \frac{1 - |\rho_{HHVV}|}{N} \quad (7)$$

Then by using iterative algorithms, quad-pol SAR image can be reconstructed. We are not going to repeat the iterative quad-pol reconstruction algorithms here, readers who are interested may refer to papers of Souyris et al., (2005), Nord et al., (2009) and Li et al. (2013). From visual inspection, it can be found that the reconstructed quad-pol image from CP data is highly similar with the pseudo color image of the original FP data. Based on quantitative analysis, it is found that for HH and VV channel, the reconstruction accuracy are much higher in both statistical and information theoretical analysis than HV channel. The reason is that due to the dominant Bragg scattering on the sea surface, the cross-pol return is much lower than the co-pol channels. Then some polarimetric characteristics such as co-and cross-polarization ratio, standard deviation of CPD, can be estimated from the reconstructed quad-pol image, to improve the accuracy of oil-spill classification.

### 5.3. Oil properties retrieve

For higher level applications of SAR oil-spill remote sensing, obtaining the location and coverage of oil-spill is definitely far from enough. It is also crucial to identify oil types for estimating its potential harm to ecosystem and arranging proper cleanup methods.

Although the terminology used to describe oil-spill is not always consistent, NOAA has developed a general glossary of terms to describe the thickness and mixture status of oil floating on the sea water (NOAA, 1996), by which oil slicks can be roughly classified into Light Sheen, Silver Sheen, Rainbow Sheen, Brown oil, Mousse, Black oil, Streamers, Tarballs, Tarmats and Pancakes.

Simply by SAR remote sensing, detailed parameters of oil-spill used to be hard to obtain due to complicated sea surface status and low interaction of oil film with microwave. Among these parameters, dielectric constant (permittivity) is a key one. It describes the material's capability of holding electro-magnetic energy or polarization, which is helpful in identifying the types and conditions of the oil. During BP Deepwater Horizon oil-spill accident, oil-spill were observed by UAVSAR (Minchew, et. al, 2012a; 2012b). Because of the extreme amount of oil

leakage (approximately 700,000 m<sup>3</sup>), the thickness of oil film/emulsion is sufficient to significantly affect the permittivity of sea surface at frequency of L-band (with wave length of approximately 24cm). Based on tilted Bragg scattering theory, the effect of oil on the dielectric permittivity and roughness spectrum was separated. It is worth noting that the DWH accident is very special for its large amount and deep source of leakage. On the way to sea surface crude oil was naturally refined and mixed with sea water, formed very complicated mixture, e.g., emulsion. So in that case, L-band SAR could observe the change of permittivity. Unfortunately in most other cases the situation is largely different.

Two-scale Boundary Perturbation Method (BPM) scattering model was used to analyze sea surface with and without biogenic slicks (Nunziata, et al., 2009). The model has the advantage of considering both large-scale sea roughness and the small-scale wave superposed on it. Compared with untitled SPM approach, the Two-scale BPM contrast model is much more accordance with the actual behavior of microwave sea surface scattering, and will be a very powerful tool for the future study. More studies are hence needed on the modeling of oil weakened sea scattering and dielectric constant of mixed oil and sea water.

The thickness of the oil slick provides vital information for estimating the amount of the spill and forecast its dispersing (Kasilingam, 1995). Near-infrared and color sensors such as MERIS and MODIS are considered to be very suitable for retrieving the thickness of oil films (Carolus et al., 2012). Hyperspectral signature is also proved to be adequate (Cong et al., 2012). However although theoretically possible, there aren't many soundness methods for retrieving oil thickness by SAR. In 1990s, polarimetric radar signature was firstly considered to estimate the thickness of thin oil films (Kasilingam, 1995). Recently Ivanova is trying to use Radar Imaging Model to obtain the thickness of the oil film from 0.37 to 0.45mm on the sea surface using satellite data.

## **6. The environmental impact study of oil-spill pollution**

Oil-spill is considered as one of the most devastating marine pollutions, which have serious influence on coastal environment (Ventikos & Psaraffis, 2004). It damages costal ecosystem in several aspects, including: biomass production, biotope landscape, greenhouse gas regulation and nutrient cycling (Mei & Yin, 2009). The environmental impact of oil pollutions depends on the type and amount of oil as well as the sensitivity of living organisms in the polluted area (Gauthier et al., 2007). As the result, the coastal data should be analyzed in order to determine to what degree oil-spill in coastal waters and their probable trajectories will influence coastal environment.

One clue of current researches is investigating how oil-spill contamination affects marine phytoplankton. A study proved that oil increase phytoplankton boom by providing surplus organic nutrient or reducing the amount of predators (Pan & Tang et al., 2011; 2012). The study concentrated on temporal variations of key phytoplankton biomass indicator: chlorophyll *a* before and after different oil-spill occurrence period. Although not much, SAR has been reported possible to observe phytoplankton blooms. For example ERS-1 observed low

backscatter areas caused by the accumulation of biological surfactants released by the plankton blooms (Svejkovsky et al., 2001). Since observation of phytoplankton blooms by SAR requires relatively strict conditions such as surface wind and incidence angle, at present stage it has better to be considered as complementary method of ocean color or radiometer observations.

Another example is the ecosystem impact study of Deepwater Horizon oil-spill based on JPL-UAVSAR data (Jones et al., 2011), in which the intrusion of oil pollution to wetlands in Barataria Bay is investigated using the polarimetric SAR data and its effect on wetland vegetation and costal algae boom is analyzed. This study proves that with fine-resolution polarimetric SAR, the damage caused by oil pollution to the coastal environment can be evaluated and the recovery of ecosystem can be tracked.

## 7. Conclusions

According to previous introduction and analysis, the following conclusions can be drawled: first, extraction of dark spots is a very fundamental step that calls for the physical understanding of imaging mechanisms of SAR. Second, effective and robust pattern recognition algorithms should be further developed to distinguish oil slicks from look-alikes. With the development of SAR technology, polarimetric properties have been used for target classification more frequently, and can be one of the key techniques to boost the accuracy of oil-spill classification. Moreover, compact polarimetric SAR modes have great potential in future applications, especially marine oil-spill surveillance that require large coverage area and short revisit time. Detailed properties of oil slicks at present is hard to be retrieved by SAR sensors alone, which required sophisticated damping model, refined scattering functions and high SNR SAR data. Last but not least, oil-spill has huge affection on marine environment, especially coastal zone. As a result, the study on its impact to the marine ecology is in urgent demand and calls for the common attention of scholars all over the world.

Due to the complexity of ocean physical processes, oil-spill detection by SAR is not an easy task, which calls for close cooperation between remote sensing scientists, oceanographers and electronic engineers. It is definitely not just a matter of image processing or pattern recognition. The understanding electro-magnetic behavior of oil presented sea surface is the key to understand the whole process, without which reliable and satisfactory results will never be obtained.

At last, it is worth noting that in operational oil pollution monitoring, the most cost-effective way is to take advantage of satellite-based SAR and aircraft surveillance jointly (Solberg, 2012). The former technique has unique advantage of large coverage area and all whether all day capabilities while the latter is helpful in further verification of oil-spill and collect evidence to prosecute the polluter.

## Acknowledgements

This research was jointly supported by the National Science Foundation of China (41271434), GRF (CUHK457212), ITF (GHP/002/11GD), the National Key Technologies R&D Program in the 12th Five Year Plan of China (Applied Remote Sensing Monitoring System for Water Quality and Quantity in Guangdong, Hong Kong and Macau, 2012BAH32B01 & 2012BAH32B03), and the funding of Shenzhen Municipal Science and Technology Innovation Council (JCYJ20120619151239947). And the author would like to thank JPL, NASA for providing the UAVSAR data.

## Author details

Yuanzhi Zhang<sup>1,2</sup>, Yu Li<sup>1</sup> and Hui Lin<sup>1,3</sup>

1 Institute of Space and Earth Information Science & Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong & Shenzhen, China

2 National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China

3 Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong

## References

- [1] Alpers, W. (2002), Remote sensing of oil spills, Proceedings of the symposium "Maritime Disaster Management", King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, pp. 19-23.
- [2] Alpers W., & H. Espedal (2004a). Oils and surfactants. In *"Synthetic Aperture Radar Marine User's Manual"*, US Department of Commerce: Washington, DC, pp. 263-275.
- [3] Alpers, W., and Ch. Melsheimer: Rainfall (2004b), In *"Synthetic Aperture Radar Marine User's Manual, National Oceanic and Atmospheric Administration, Center for Satellite Application and Research, NOAA/NESDIS, Washington, D.C., USA, ISBN 0-16-073214-X, pp. 355-371.*
- [4] Alpers, W., and W. Huang (2011), On the discrimination of radar signatures of atmospheric gravity waves and oceanic internal waves on synthetic aperture radar images of the sea surface, *IEEE Trans. Geosci. Rem. Sens.*, vol. 49, no. 3, pp. 1114-1126.

- [5] Brown, C. E., & M. F. Fingas (2003), Synthetic Aperture Radar Sensors: Viable for Marine Oil Spill Response. *Proc. of 26th Arctic and Marine Oil spill Program (AMOP)*, Canada.
- [6] Carolis G., Adamo M., Pasquariello G. (2012). Thickness estimation of marine oil slicks with near-infrared MERIS and MODIS imagery: The Lebanon oil spill case study, *Geoscience and Remote Sensing Symposium (IGARSS), IEEE International*, pp. 3002-3005.
- [7] Chen J., Quegan S. (2011). Calibration of Spaceborne CTLR Compact Polarimetric Low-Frequency SAR Using Mixed Radar Calibrators, *Geoscience and Remote Sensing, IEEE Transactions on*, vol.49, no.7, pp.2712,2723.
- [8] Cloude, S. R., Goodenough, D. G., Chen, H. (2012). Compact Decomposition Theory, *Geoscience and Remote Sensing Letters, IEEE*, vol.9, no.1, pp.28-32.
- [9] Collins, M. J., Denbina, M., Atteia, G. (2013). On the Reconstruction of Quad-Pol SAR Data From Compact Polarimetry Data For Ocean Target Detection, *Geoscience and Remote Sensing, IEEE Transactions on*, vol.51, no.1, pp.591-600.
- [10] Dubois-Fernandez, P. C., Souyris, J-C, Angelliaume, S., Garestier, F. (2008). The Compact Polarimetry Alternative for Spaceborne SAR at Low Frequency, *Geoscience and Remote Sensing, IEEE Transactions on*, vol.46, no.10, pp.3208-3222.
- [11] Frate F., Petrocchi A., Lichtenegger J.& Calabresi G. (2000). Neural Networks for Oil Spill Detection Using ERS-SAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2282-2287.
- [12] Gade, M. & W. Alpers (1999). Using ERS-2 SAR for routine observation of marine pollution in European coastal waters, *Sci. Total Environ.*, vol. 237, pp. 38441-38448.
- [13] Gambardella, A., Giacinto, G., Migliaccio, M. (2008). On the Mathematical Formulation of the SAR Oil-Spill Observation Problem, *Geoscience and Remote Sensing Symposium, IGARSS IEEE International*, vol.3, pp.III-1382,III – 1385.
- [14] Garcia-Pineda O., MacDonald I. R., Li X., Jackson C. R., Pichel W.G. (2013). Oil Spill Mapping and Measurement in the Gulf of Mexico With Textural Classifier Neural Network Algorithm (TCNNA), *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, no.99, pp.1-9.
- [15] Gauthier M., Weir L., Ou Z., Arkett M., and Abreu R. (2007). Integrated satellite tracking of pollution: A new operational program, *in Proc. Int. Geosci. Remote Sens. Symp.*, pp. 967–970.
- [16] Huang B., Li H., & Huang X. (2005). A Level Set method for Oil Slick Segmentation in SAR Images. *Int. J. Remote Sens.*, vol. 26, pp. 1145-1156.

- [17] Jain A. K., R. P. W. Duin, J. Mao. (2000). Statistical Pattern Recognition: A Review. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4-37, Jan. 2000.
- [18] Jones, C. E., B. Minchew, B. Holt, and S. Hensley (2011). Studies of the Deepwater Horizon oil spill with the UAVSAR radar, in Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise, *Geophys. Monogr. Ser.*, vol. 195, edited by Y. Liu et al., pp. 33–50, AGU, Washington, D. C.
- [19] Kasilingam D. (1995). Polarimetric radar signatures of oil slicks for measuring slick thickness, Combined Optical-Microwave Earth and Atmosphere Sensing. *Conference Proceedings., Second Topical Symposium on*, pp. 3-6.
- [20] Kudryavtsev, V. N., Chapron, B., Myasoedov, A. G., Collard, F., Johannessen, J. A. (2013), On Dual Co-Polarized SAR Measurements of the Ocean Surface, *Geoscience and Remote Sensing Letters, IEEE, IEEE*, vol.10, no.4, pp.761-765.
- [21] Li Y., Zhang Y. (2013), Synthetic aperture radar oil spill detection based on morphological characteristics, *Geo-spatial Information Science*, to be published.
- [22] Li, Y., Zhang, Y., Chen, J., Zhang, H. (2014), Improved Compact Polarimetric SAR Quad-Pol Reconstruction Algorithm for Oil Spill Detection, *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, no.6, pp. 1139-1142.
- [23] Lin C., Nutter, B., Liang D. (2012). Estimation of oil thickness and aging from hyperspectral signature, Image Analysis and Interpretation (SSIAI), *IEEE Southwest Symposium on*, pp.213-216.
- [24] Marangoni, C. (1872). Sul principio della viscosith superficiale dei liquidi stabili, *Nuovo Cimento*, Ser. 2, 5/6, 239-273.
- [25] Marghany, M. (2001). RADARSAT automatic algorithms for detecting coastal oil spill pollution. *International Journal of Applied Earth Observation and Geo-information*. vol. 3, iss. 2, pp. 191-196.
- [26] Marghany M. and Hashim M. (2011). Comparison between Mahalanobis classification and neural network for oil spill detection using RADARSAT-1 SAR data. *International Journal of the Physical Sciences*. vol. 6 (3), pp. 566 – 576.
- [27] Marghany M., Cracknell A. and Hashim M. (2009). Modification of Fractal Algorithm for Oil Spill detection from RADARSAT-1 SAR data. *International Journal of Applied Earth Observation and Geoinformation*. vol. 11, pp.96-102.
- [28] Marghany M.and Hashim M. (2011a). Discrimination between oil spill and look-alike using fractal dimension algorithm from RADARSAT-1 SAR and AIRSAR/POLSAR data. *International Journal of the Physical Sciences*.

- [29] Marghany, M., & M. Hashim (2011b). Comparative Algorithms for Oil Spill Automatic Detection Using Multimode RADARSAT-1 SAR Data. *IEEE International Geoscience and Remote Sensing Symposium*, pp. 2173-2176.
- [30] Marghany M. (2013). Genetic Algorithm for Oil Spill Automatic Detection from Envisat Satellite Data. Edited by Misra S., Carlini M., Carmelo M. Torre, H. Nguyen, D. Taniar, Bernady O. Apduhan, and Osvaldo Gervasi, In Beniamino Murgante, *Computational Science and Its Applications – ICCSA 2013*, vol. 7972, pp 587-598.
- [31] Migliaccio M., Tranfaglia M., Ermakov S. A. (2005). A Physical Approach for the Observation of Oil Spills in SAR Images. *IEEE Journal of Oceanic Engineering*, vol. 30, no. 3., pp. 496-507.
- [32] Migliaccio, M. (2010). Pedestal height for sea oil slick observation, *Radar, Sonar & Navigation, IET*, vol 5, no. 2, 103-110.
- [33] Migliaccio, M., F. Nunziata, & A. Gambardella (2009). On the co-polarized phase difference for oil spill observation. *Interl. J. Remote Sens.*, vol. 30, 1587-1602.
- [34] Migliaccio, M., G. Ferrara, A. Gambardella, & F. Nunziata (2007a). A new stochastic model for oil spill observation by means of single-look SAR data. *Environ. Res. Eng. Manag.*, vol. 39, pp. 24-29.
- [35] Migliaccio, M., Gambardella, A.; Tranfaglia, M. (2007b). SAR Polarimetry to Observe Oil Spills, *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 2, pp. 506-511.
- [36] Migliaccio, M., Tranfaglia, M. (2005). A study on the use of SAR polarimetric data to observe oil spills, *Oceans 2005-Europe*, vol. 1, pp.196-200.
- [37] Minchew, B., Jones, C.E., Holt, B. (2012a). Polarimetric Analysis of Backscatter From the Deepwater Horizon Oil Spill Using L-Band Synthetic Aperture Radar, *Geoscience and Remote Sensing, IEEE Transactions on*, vol.50, no.10, pp.3812-3830.
- [38] Minchew, B. (2012b). Determining the Mixing of Oil and Sea Water Using Polarimetric Synthetic Aperture Radar, *Geophys. Res. Lett.*, in press.
- [39] Mladenova, I. E., Jackson, T. J.; Bindlish, R., Hensley, S. (2013). Incidence Angle Normalization of Radar Backscatter Data, *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 3, pp.1791-1804.
- [40] NOAA (1996). Aerial observations of oil at sea, Office of Ocean Resources Conservation and Assessment. ([http://docs.lib.noaa.gov/rescue/NOAA\\_E\\_DOCS/E\\_Library/ORR/oilspills/OilatSea.pdf](http://docs.lib.noaa.gov/rescue/NOAA_E_DOCS/E_Library/ORR/oilspills/OilatSea.pdf)).
- [41] Nord, M. E., Ainsworth, T.L., Jong-Sen Lee, Stacy, N. J. S. (2009). Comparison of Compact Polarimetric Synthetic Aperture Radar Modes, *Geoscience and Remote Sensing, IEEE Transactions on*, vol.47, no.1, pp.174-188.

- [42] Nunziata, F., Gambardella, A., Migliaccio, M. (2008). On the Use of Dual-Polarized SAR Data for Oil Spill Observation, *Geoscience and Remote Sensing Symposium, IEEE International*, vol.2, pp.II-225,II-228, pp. 7-11.
- [43] Nunziata, F., Migliaccio, M. (2011). Gambardella, A.. Pedestal height for sea oil slick observation, *Radar, Sonar & Navigation, IET*, vol. 5, no. 2, pp.103-110.
- [44] Nunziata, F.; Sobieski, P.; Migliaccio, M. (2009), The Two-Scale BPM Scattering Model for Sea Biogenic Slicks Contrast, *Geoscience and Remote Sensing, IEEE Transactions on*, vol.47, no.7, pp.1949-1956.
- [45] Pan, G., Tang, D., Zhang, Y. (2012). Satellite monitoring of phytoplankton in the East Mediterranean Sea after the 2006 Lebanon oil spill. *International journal of remote sensing*, 33(23), pp. 7482-7490.
- [46] Raney, R. K. (2007). Hybrid-Polarity SAR Architecture, *Geoscience and Remote Sensing, IEEE Transactions on*, vol.45, no.11, pp.3397-3404.
- [47] Salberg, A.-B.; Rudjord, O.; Solberg, A.H.S. (2012). "Model based oil spill detection using polarimetric SAR," *Geoscience and Remote Sensing Symposium (IGARSS), IEEE International*, pp.5884-5887.
- [48] Shirvany, R., Chabert, M., Tournet, J.-Y. (2012). Ship and Oil-Spill Detection Using the Degree of Polarization in Linear and Hybrid/Compact Dual-Pol SAR, *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 3, pp.885-892.
- [49] Shu Y., Li J., Yousif H. (2010). Dark-spot detection from SAR intensity imagery with spatial density thresholding for oil-spill monitoring, *Remote Sensing of Environment*, vol. 114, issue 9, pp. 2026–2035.
- [50] Skrunes S., Brekke C., Eltoft T. (2012), An experimental study on oil spill characterization by multi-polarization SAR, *Synthetic Aperture Radar, 2012. EUSAR. 9th European Conference on*, pp.139,142.
- [51] Solberg, A. H. S. (2012). Remote Sensing of Ocean Oil-Spill Pollution, *Proceedings of the IEEE*, vol.100, no.10, pp.2931,2945.
- [52] Solberg, A., Storvik, G., Solberg, R., Volden, E (1999). Automatic Detection of Oil Spills in ERS SAR Images. *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 4, pp. 1916-1924.
- [53] Souyris J. C., Imbo P., Fjortoft R., Mingot, S. and Lee J.-S. (2005). Compact polarimetry based on symmetry properties of geophysical media: The  $\pi/4$  mode, *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 634-646.
- [54] Svejkovsky J., Shandley J. (2001). Detection of offshore plankton blooms with AVHRR and SAR imagery, *International Journal of Remote Sensing*, vol. 22, iss. 2-3, pp. 4625-4633.

- [55] Topouzelis K., Karathanassi V., Pavlakis P., Rokos D. (2007). Detection and discrimination between oil spills and look-alike phenomena through neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*. vol. 62, Iss. 4, pp. 264-270.
- [56] Topouzelis K., Stathakis D. and Karathanassi V.. (2008a). Investigation of Genetic Algorithms Contribution to Feature Selection for Oil Spill Detection. *International Journal of Remote Sensing*, vol. 30, no. 3, pp. 611–625.
- [57] Topouzelis, K., D. Stathakis, & V. Karathanassi (2009). Investigation of genetic algorithms contribution to feature selection for oil spill detection. *Intl. J..Remote Sens.*, vol. 30 no. 3, pp. 611-625.
- [58] Topouzelis, K., V. Karathanassi, P. Pavlakis, & D. Rokos (2008). Dark formation detection using neural network. *Intl. J. Remote Sens.*, vol. 29 no. 16, pp. 4705-4720.
- [59] Yin J., Yang J., Zhang X. (2011). On the ship detection performance with compact polarimetry, *Radar Conference (RADAR), IEEE*, pp.675-680, 23-27.
- [60] Zhang B., Perrie W., Li X.& Pichel W., (2011). Mapping sea surface oil slicks using RADARSAT-2 quad-polarization SAR image, *Geophys. Res. Lett.*, vol. 38, L10602.
- [61] Zhang Y., Lin H., Liu Q., Hu J., Li X. and Yeung K. (2012). Oil-spill monitoring in the coastal waters of Hong Kong and vicinity, *Marine Geodesy*, vol, 35, pp. 93-106.

---

# Oil Spill Pollution Automatic Detection from MultiSAR Satellite Data Using Genetic Algorithm

---

Maged Marghany

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/58572>

---

## 1. Introduction

Oil spill pollution has a substantial role in damaging marine ecosystem. Oil spill that floats on top of water, as well as decreasing the fauna populations, affects the food chain in the ecosystem [1-5]. In fact, oil spill is reducing the sunlight penetrates the water, limiting the photosynthesis of marine plants and phytoplankton. Moreover, marine mammals for instance, disclosed to oil spills their insulating capacities are reducing, and so making them more vulnerable to temperature variations and much less buoyant in the seawater. Therefore, oil coats the fur of sea otters and seals, reducing its insulation abilities and leading to body temperature fluctuations and hypothermia. Ingestion of the oil causes dehydration and impaired digestions [5].

Over recent years, there has been explosive increments in marine oil spill pollutions. Deepwater Horizon oil spill in 2010, for instance, is the most serious marine pollution disaster has occurred in the history of the petroleum industry (Figure 1).

This disaster has dominated by three months of oil flows in coastal waters of the Gulf of Mexico. Incidentally, the Deepwater Hoizon oil spill has serious effects on feeble maritime, wildlife habitats, Gulf's fishing, coastal ecologies and tourism industries [30, 31]. Therefore, the resulting oil slicks are difficult to control, as their evolution be influenced by weather, currents, tides, and many chemical and physical factors [1][10][31]. Further, oil sources is challenging to verify and be subject to the type of oil, its volume and location, duration of the seepage, and surrounding environmental conditions [1].

Deepwater Horizon oil spill in 2010 is the most serious marine pollution disaster has occurred in the history of the petroleum industry. This disaster has dominated by three months of oil flows in coastal waters of the Gulf of Mexico. In this regard, the Deepwater Horizon oil spill has serious effects on feeble maritime, wildlife habitats, Gulf's fishing, coastal ecologies and



**Figure 1.** Oil spill disaster in Gulf of Mexico

tourism industries. As a result, human health problems are caused because of the spill and its clean-up [25]. Consistent with Marghany and Hashim [7], Synthetic aperture radar (SAR) is a precious foundation of oil spill detection, surveying and monitoring that improves oil spill detection by various approaches. The different SAR tools to detect and observe oil spills are vessels, airplanes, and satellites [5]. Vessels can detect oil spills at sea, covering restricted areas, say for example, (2500 m x 2500 m), when they are equipped with the navigation radars [20]. On the other hand, airplanes and satellites are the main tools that are used to record sea-based oil pollution [7] [22, 31].

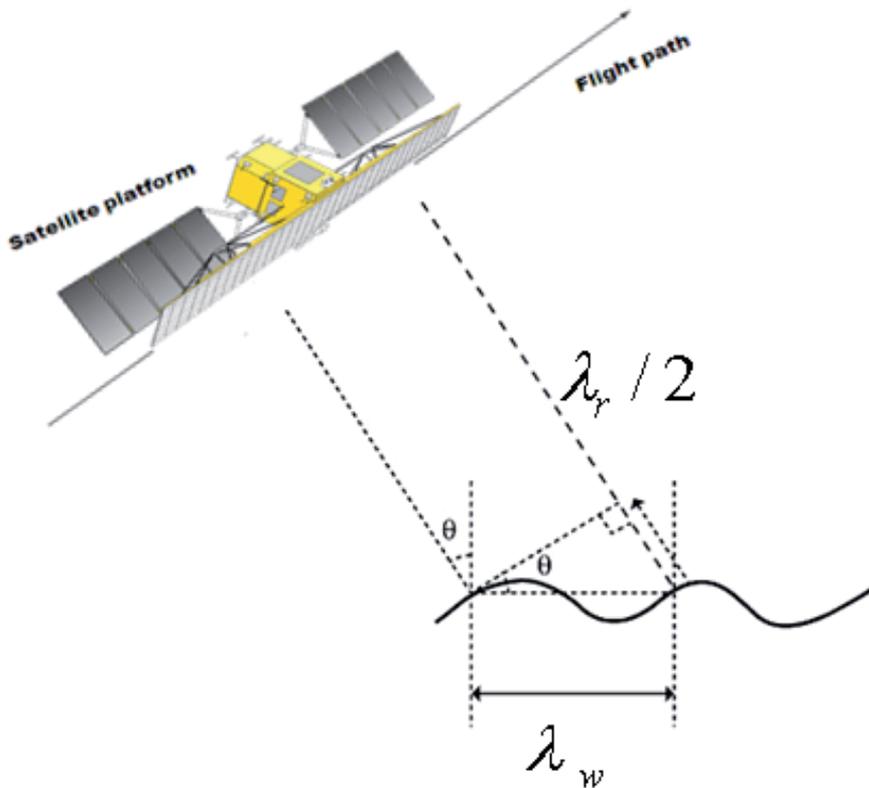
### 1.1. Principle of oil spill detection in SAR data

The main theory of oil spill imaging using SAR images is based on the concept of resonance fundamental theory. Very short ocean waves travel on the larger ocean waves, or swells. As a superposition, radar backscatter at incident angles of  $20^\circ$  to  $70^\circ$  is principally produced by Bragg resonance [23].

#### 1.1.1. General concept of Bragg scattering

According to According to Topouzelis [22] and Trivero [24], backscatter can model in two ways: specular reflection and Bragg scattering. Specular reflection occurs when the water surface is tilted which creates a small mirror pointing to the radar. For the perfect specular reflector, radar returns (backscatter) exist only near vertical incidence. This is because of a  $90^\circ$ -depression angle or the slope of the surface. In addition, the reflected energy is localized to the small angular region around the angle of reflection. Even for non-vertical incidence, however, backscatter can exist for a rough subsurface. This can occur if radar is penetrating deep enough[25]. Nevertheless, because of highest dielectric of water ocean, radar 's signal cannot penetrate sea surface.

Bragg or resonant scattering is scattering from a regular surface pattern. Resonant backscattering occurs when phase differences between rays backscattered from subsurface pattern interfere constructively. The resonance condition is  $2\lambda_w \sin\theta = \lambda_r$ , where  $\lambda_w$  and  $\lambda_r$  are water wavelength and the radar wavelength, respectively.  $\theta$  is the local angle of incidence (Figure 2). According to Topouzelis et al., [20], the short Bragg-scale waves form in response to wind stress, if the sea surface is rippled by a light breeze and no long waves are present. The radar backscatter is due to the component of the wave spectrum. This resonates with a radar wavelength.



**Figure 2.** Bragg Scatter concept.

Swell waves being imaged have much longer wavelengths than the short gravity waves, which cause the Bragg resonance. Further, Bragg-resonance from the ocean might be considered as coming from facets. The term facet refers to a relatively flat portion of the long wave structure with cover of ripples containing Bragg-resonant facets. This behaves like specular points. The beam-width and scattering gain of each facet are determined by its length in the appropriate direction [4, 10, 22].

### 1.1.2. Mathematical expression of Bragg scattering

As the incidence angle of the SAR is oblique to the local mean angle of the ocean surface, there is almost no direct specular reflection except at very high sea states. It is therefore assumed that at first approximation Bragg resonance is the primary mechanism for backscattering radar pulses. The Bragg equation defines the ocean wavelengths  $\lambda_w$  for Bragg scattering as a function of radar wavelength  $\lambda_r$  and incidence angle  $\theta$  :

$$\lambda_w = \frac{\lambda_r}{2 \sin \theta} \quad (1)$$

The short Bragg-scale waves are formed in response to wind stress. If the sea surface is rippled by a light breeze with no long waves present, the radar backscatter is due to the component of the wave spectrum which resonates with the radar wavelength. The threshold wind speed value for the C-band waves is estimated to be at about 3.25 m/s at 10 meters above the surface. The Bragg resonant wave has its crest nominally at right angles to the range direction [3].

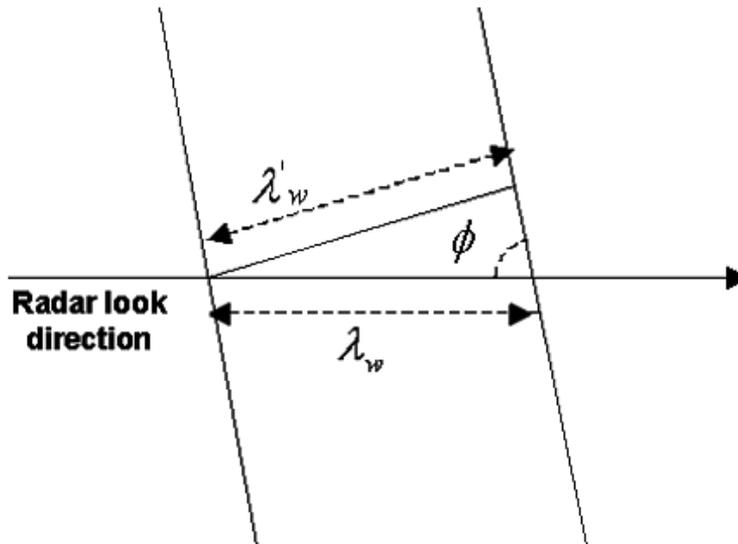
Under circumstance of Bragg scattering, largest incidence angle reduces the SAR backscatter. Indeed, the SAR backscattered power is proportional to the spectral energy density of the Bragg and the spectral distribution decays at shorter wavelength. SAR images tend to become darker with increasing range. Backscatter is related to the local incident angle (i.e. as the local incident angle increases, backscatter decreases), which is in turn related to the distance in the range direction. Backscatter is also related to wind speed [3].

According to Brekke and Solberg [3], For RADARSAT-1 and ENVISAT ASAR with C-band frequency, a radar wavelength of 5.7 cm and incidence angles in the range of 20°-50° will this model give Bragg resonant sea wavelengths  $\lambda_w$  in the range of 8.3-3.7 cm. For surface waves with crests at an angle  $\phi$  to the radar line-of-sight (Figure 3) the Bragg scattering criterion is

$$\lambda'_w = \frac{\lambda_r \cdot \sin \phi}{2 \sin \theta} = \lambda_w \cdot \sin \phi \quad (2)$$

where  $\lambda'_w$  is the wavelength of the surface waves propagating at angle  $\phi$  to the radar line-of-sight. The resonant surface wavelengths will increase when frequency increases.

The SAR directly images the spatial distribution of the Bragg-scale waves. The spatial distribution may be effected by longer gravity waves, through tilt modulation, hydrodynamic modulation and velocity bunching. Moreover, variable wind speed, changes in stratification in the atmospheric boundary layer, and variable currents associated with upper ocean circulation features such as fronts, eddies, internal waves and bottom topography effect the Bragg waves [3, 22].



**Figure 3.** Crests at an angle  $\phi$  to the look direction of the SAR [3].

### 1.1.3. Oil spill and surface films impact on Bragg scattering

Bragg scattering is a significant concept to understand the radar signal interaction with ocean surface. In this regard, the presence of capillary wave will produce backscatter that assists radar in imaging sea surface. Short gravity waves and capillary waves are damped by dynamic elasticity of the water surface, that is by changes in surface tension which occur when the surface is stretched or compressed [3]. This has the effect of the extracting of this energy from those waves which depends wholly or partly on surface tension to provide the restoring force necessary for wave propagation. If a surface film is present, the surface tension is lower than it would be in the absence of the film, and stretching and compression of the film because of the presence of waves provides the dynamic elasticity which enhances the wave damping. Thus capillary waves and short gravity waves are always damped in the presence of surface films [4]. If the surface film is spatially patchy, varying in thickness, or lined up in slicks because of surface convergence, it is to expect that the capillary/gravity wave energy will reflect that patchiness, being greatest where is no surface film. Oil films on the sea surface damp the capillary waves of the surface height spectrum. This hydrodynamic damping influences the normalized radar cross section (NRCS) of contaminated seas, comparatively to clean seas [22].

Therefore, oil slicks dampen the Bragg waves (wavelength of a few cm) on the ocean surface and reduce the radar backscatter coefficient. This results in dark regions or spots in satellite SAR images. Topouzelis et al., [21], emphasizes the importance of weathering processes, as they influence the oil spills physicochemical properties and detect-ability in SAR images. The processes that play the most important role for oil spill detection are evaporation, emulsifica-

tion and dispersion. Lighter components of the oil will evaporate into the atmosphere. The rate of evaporation is dependent on oil type, thickness of the spill, wind speed and sea temperature. Emulsification is estimated based on water uptake as a function of the wind exposure of the actual oil type. Dispersion is an important factor in deciding the lifetime of an oil spill and it is strongly dependent on the sea state [3].

## 1.2. Oil spill detection in SAR data

Scientists which have agreed with that oil spill detection with SAR data is based on: (i) all dark patches present in the SAR images are isolated [3, 16]; (ii) features for each dark patch are then extracted [1, 12]; (iii) these dark patches are tested against predefined values [22]; and (v) the probability for every dark patch is calculated to determine whether it is an oil spill or a look-alike phenomenon [1, 14, 20, 22, 25]. Therefore, Topouzelis et al., [21] and Topouzelis [22] reported that most studies use the low resolution of SAR data such as quick-looks, with the nominal spatial resolution of 100 m x 100 m, to detect oil spills. In this regard, quick looks' data are sufficient for monitoring large scale area of 300 km x 300 km. On the contrary, they cannot efficiently detect small and fresh spills [22].

Further, SAR data have distinctive features as equated to optical satellite sensors which makes SAR extremely valuable for spill watching and detection [5] [7] [9] [13]. These features are involved with several parameters: operating frequency, band resolution, incidence angle and polarization [10] [12]. Marghany and Hashim [7] develop comparative automatic detection procedures for oil spill pixels in Multimode (Standard beam S2, Wide beam W1 and fine beam F1) RADARSAT-1 SAR satellite data post the supervised classification (Mahalanobis), and neural network (NN) for oil spill detection. They found that NN shows a higher performance in automatic detection of oil spill in RADARSAT-1 SAR data as compared to the Mahalanobis classification with a standard deviation of 0.12. In addition, they W1 beam mode is appropriate for oil spill and look-alikes discrimination and detection [10] [11] [12]. Recently, Skrunes et al., [9], nevertheless, reported that there are several disadvantages are associated with Current SAR based oil spill detection and monitoring. They stated that, SAR sensors are not able to detect thickness distribution, volume, the oil-water emulsion ratio and chemical properties of the SAR data. In this regard, they recommended to utilize multi-polarization acquisition data such as RADARSAT-2 and TerraSAR-X satellites. They concluded that the multi-polarization data show a prospective for prejudice between mineral oil slicks and biogenic slicks.

Finally Marghany, [30] used Genetic algorithm for oil spill detection in ENVISAT ASAR data along Singapore Straits. He found that crossover process, and the fitness function generated accurate pattern of oil slick in SAR data. He used the receiver –operational characteristics (ROC) curve to verify oil spill detection in SAR data. He stated that ROC verified 85% for oil spill, 5% look-alikes and 10% for sea roughness.

## 1.3. Problem statement

The main challenge with SAR satellite data is the difficulty to drawbacks, which makes it difficult to develop a fully automated detection of oil spills. Due to the inherent difficulty of

discriminating between oil spills and look-likes, an automatic algorithm with a reliable confidence estimator of oil spill would be highly desirable. The needs for automatic algorithms rely on the number of images to be analyzed, but for monitoring large ocean areas, it is a cost-effective alternative to manual inspection. Automatic detection algorithms of oil spill are normally divided into three steps: (i) dark spot detection; (ii) dark spot feature extraction; and (iii) dark spot classification [3,10, 22,31].

One of the main problems in oil slick combat and management are forecasting the behavior (movement and spreading) of oil slicks. Commonly, the objective of predicting the behavior of oil slicks is to determine the time-evolving shape of the slick under different weather patterns in waters where currents exist [11-14]. Wind direction and speed are the most important climate parameters can impact the oil spill imagine in SAR data. Although great progress made in detecting and surveying oil slicks, a general model for oil slick movement and spreading has not yet been devised [1]. Models for oil slick behavior are important in environmental engineering and used as a decision support tool in environmental emergency responses. These models also use to help ships avoid oil slicks [1].

Modeling the progress and extent of oil in the ocean are not always tested against authentic spills, and models are not regularly developed with real databases, but rely instead on theoretical scenarios. In fact, some spill-threat scenarios have not been based on real oil movement data at all [15]. Yet there are frequent demands to provide just such models with credible precision. Consequently, it is important to study the behavior and movement of spilled oil in the sea in order to describe a suitable management plan for mitigating adverse impacts arising from such accidents. Simulation of oil spills using mathematical models form an important basis for subsequent study, according to Marghany [9]. Collectively, with the information of position of weak resources in time and space, the simulation outcome may develop a basis for evaluating the damage potential from an eventual oil spills. This may help the regulatory authorities to take direct preventive measures [9].

In addition, most of the studies that have been conducted in the coastal waters of Malaysia used a single radar image, which is inadequate to ensure a high degree of accurate detection of oil spills using SAR data. Some of the work involved the implementation of non-appropriate techniques for oil spill automatic detection. For instance, Mohamed et al., [16] have used data fusion techniques in a single RADARSAT-1 SAR image, with different co-occurrence texture algorithm results. However, the data fusion technique must apply with two or more different sensors for example, ERS-1, LANDSAT, and SPOT. In addition, according to Brekke and Solberg [3], using such PCA analysis does not consider an appropriate method for data fusion. Data fusion technique involves several methods such as high pass filtering technique, IHS transformation method, Brovey method, and à Trous wavelet method [3].

#### **1.4. Hypotheses and objective**

Concern with above prospective, we address the question of Genetic algorithm ability for oil spill detection in multiSAR satellite. This work has hypothesized that the dark spot areas (oil slick or look-alike pixels) and its surrounding backscattered environmental signal complex looks in the ENVISAT ASAR data can detect using Genetic Algorithm. However, previous

work has implemented post classification techniques [4, 9, 16, 18] or artificial neural network [19],[21],[24] which are considered as semi-automatic techniques. The objective of this work can divide into two sub-objectives: (i) To examine GA [27, 31] for oil spill automatic detection in multiSAR satellite data; and (ii) To design the multi-objective optimization algorithm for oil spill automatic detection in multiSAR satellite data that are based on algorithm's accuracy.

## 2. Data acquisition

The SAR data acquired in this study are from the ENVISAT ASARR and RADARSAT-2 SAR data that are involved Standard beam mode (S2); W1 beam mode (F1) image. SAR data are C-band and have a lower signal-to noise ratio due to their HH polarization with wavelength of 5.6 cm and a frequency of 5.3 GHz [7]. Further, RADARSAT-SAR data have 3.1 looks and cover an incidence angle of  $23.7^\circ$  and  $31.0^\circ$  [10]. In addition, RADARSAT-2 SAR data cover a swath width of 100 km.

SAR operation modes	
Frequency range	C-band (5.405 GHz)
Channel bandwidth	11.6, 17.3, 30, 50 and 100 MHz
Channel polarization	HH, HV, VH, VV
SAR antenna dimensions	15m x 1.5m

**Table 1.** RADARSAT-2 SAR different mode.

ASAR (Advanced Synthetic Aperture Radar) operates in the C band in a wide variety of modes. It can detect changes in surface heights with sub-millimeter precision. It served as a data link for ERS-1 and ERS-2, providing numerous functions such as observations of different polarities of light or combining different polarities, angles of incidence and spatial resolutions (Table 2).

Mode	Id	Polarisation	Incidence	Resolution	Swath
Alternating polarisation	AP	HH/VV, HH/HV, VV/VH	15 – 45°	30 – 150 m	58 – 110 km
Image	IM	HH, VV	15 – 45°	30 – 150 m	58 – 110 km
Wave	WV	HH, VV		400 m	5 × 5 km
Suivi global (ScanSAR)	GM	HH, VV		1 km	405 km
Wide Swath (ScanSAR)	WS	HH, VV		150 m	405 km

**Table 2.** ENVISAT different Modes.

### 3. Study area

The study conducted in two different coastal zones. The first area is located along the Gulf of Mexico and the second is located along Singapore Straits. It is surprising that the two oil spill disaster events occurred in April 20, 2010, and May 25, 2010, respectively.

#### 3.1. Gulf of Mexico

The Deepwater Horizon, an offshore oil-drilling rig, exploded on the night of April 20, 2010 while working on a well on the sea floor in the Gulf of Mexico. The blast occurred 41 miles from the Louisiana coast (Figure 5). For nearly three months, oil leaked from the Macondo well at a rate estimated between 35,000 and 60,000 barrels per day. (There are 42 gallons in a barrel, so that's equal to 1.4 to 2.5 million gallons per day.) Repeated attempts to stop the flow failed until mid-July, when a tighter-fitting cap sealed the well head. In all, the well spilled 4.9 million gallons: the biggest offshore oil spill in history. Further, the oil slick spread quickly over the ocean surface, covering 1,500 square kilometers (580 sq miles) by April 25 and over 6,500 square kilometers (2,500 sq miles) by the beginning of May.



**Figure 4.** Location of oil spill event in Gulf of Mexico.

#### 3.2. Singapore straits

The second study area is located along the Singapore Straits. The SAR data acquired in this study are from ENVISAT ASAR data on June 3, 2010, in single look complex format. On May 25, 2010 a merchant ship (Figure 6) collided with a Malaysian oil tanker on Tuesday morning, puncturing the tanker's hull and spilling 2,500 tons of crude oil into the Singapore Strait (Figure 7), maritime officials reported. The damage appeared to be limited to one compartment in the double-hulled tanker, the Bunga Kelana 3, with the spill amounting to about 18,000 barrels.



Figure 5. The merchant ship collided with a Malaysian oil tanker.



Figure 6. Location of oil spill is shown by red triangular.

## 4. Genetic algorithm

On the word of Kahlouche et al., [27], Genetic algorithms (GA) differ from classification algorithm. In classification algorithm, a single point at every iteration is generated. Moreover, classification algorithms correspondingly choose the next point in the classification by a deterministic computation. In contrast, the genetic algorithm (GA) generates a population of cells at every iteration, where the superlative cell in the population approaches an optimal solution. Moreover, GA, implements probabilistic transition rules not deterministic rules as compared to classification algorithms [22] [28, 31]. Consequently, Cellular Automata (CA) are mathematical algorithms that involve a large number of relatively simple individual units, or "cells," which is connected only locally, without the existence of a central control in the system. Each cell is a simple finite automaton that repeatedly updates its own state, where the new cell state depends on the cell's current state and those of its immediate (local) neighbors. The limited functionality of each cell, and the interactions, however, being restricted to local neighbors. Thus the system as a whole is capable of producing intricate patterns, and performing complicated computations [29].

A constrained multi-objective problem for oil spill discrimination in SAR data deals with more than one objective and constraint namely look-alikes, for instance, currents, eddies, upwelling or downwelling zones, fronts and rain cells). The general form of the problem is adapted from Sivanandam and Deepa [29] and described as

Minimize

$$f(\beta) = [f_1(\beta), f_2(\beta), \dots, f_k(\beta)]^T \quad (3)$$

Subject to the constraints:

$$g_i(\beta) \leq 0, i = 1, 2, 3, \dots, I \quad (4)$$

$$h_j(\beta) \leq 0, j = 1, 2, 3, \dots, J \quad (5)$$

$$\beta_s \leq \beta \leq \beta_U \quad (6)$$

where,  $f_i(\beta)$  is the  $i$ -th pixel backscatters  $\beta$  in SAR data,  $g_i(\beta)$  and  $h_j(\beta)$  represents the  $i$ -th and  $j$ -th constraints of backscatter in raw direction and column direction, respectively.  $\beta_L$  and  $\beta_U$  are the lower and upper limit of values of the backscatter. The transition rules for the cellular automata oil spill detection is designed using the input of different backscatter values  $\beta$  to identify the slick conditions required in the neighbourhood pixels of kernel window size of 7x7 pixels and lines for a  $\beta$  pixel to become oil slick. These rules can be summarized as follows:

1. IF test pixel is sea surface, OR current boundary features THEN  $\beta \geq 0$  not oil spill.
2. IF test pixel is dark patches (low wind zone, OR biogenic slicks OR shear zones)  $\beta \leq 0$  THEN It becomes oil slick if its.

#### 4.1. Data organization

Let the entire backscatter of dark patches in multiSAR data are  $\beta_1, \beta_2, \beta_3, \dots, \beta_k$  where  $K$  is the total number backscatter of dark patches in the multiSAR data. Therefore,  $K$  is made up from genes which is representing the backscatter  $\beta$  of dark patches and its surrounding environment and genetic algorithms is started with the population initializing step.

#### 4.2. Population initializing

Let  $P_i^j$  is a gene which corresponds to backscatter of dark pixels and its surrounding pixels. Consequently,  $P_i^j$  is randomly selected and representing both of backscatter variations of dark patches and its surrounding environmental pixels. Moreover,  $i$  varies from 1 to  $K$  and  $j$  varies from 1 to  $N$  where  $N$  is the population size.

#### 4.3. The fitness function

Following Kahlouche et al., [27], a fitness function is selected to determine the similarity of each individual backscatter of dark patches in RADARSAT-2 SAR and ENVISAT ASAR data. Then the backscatter of dark patches in RADARSAT-2 SAR and ENVISAT data be symbolized by  $\beta_i$  where  $i=1,2,3, \dots, K$  and the initial population  $P_i^j$  where  $j=1,2,3, \dots, N$  and  $i=1,2,3, \dots, K$ . Formally, the fitness value  $f(P^j)$  of each individual of the population is computed as follows [27]:

$$f(P^j) = \left[ \sum_{i=1}^K |P_i^j - \beta_i| \right]^{-1} \quad j = 1, \dots, N. \quad (7)$$

where,  $N$  and  $K$  are the number of individuals of the population used in fitness process. Generally, Equation 7 used to determine the level of similarities of dark patches that are belong to oil spill in RADARSAT-2 SAR and ENVISAT ASAR data.

#### 4.4. Selection step

The key parameter in the selection step of genetic algorithm which is chosen the fittest individuals  $f(P^j)$  from the population  $P_i^j$ . The threshold value  $\tau$  is determined by the maximum values of fitness of the population  $Max f(P^j)$  and the minimum values of fitness of the population of  $Min f(P^j)$  Indeed, in the next generations, this step serves the populations  $P$ . Therefore, the values of the fittest individuals dark patches in RADARSAT-2 SAR and ENVISAT ASAR data are greater identifies threshold  $\tau$  which is given by

$$\tau = 0.5 [Max f(P^j) + Min f(P^j)] \quad (8)$$

Equation 8 used as selection step to determine the maximum and minimum values of fitness of the population, respectively. This is considered as a dark patches' population generation step in GA algorithm.

#### 4.5. The reproduction step

According to Sivanandam and Deepa [29], Genetic algorithm is mainly a function of the reproducing step which involves the crossover and mutation processes on the backscatter population  $P_i^j$  in SAR data. In this regard, the crossover operator constructs the  $P_i^j$  to converge around solutions with high fitness. Thus, the closer the crossover probability is to 1 and the faster is the convergence [27]. In crossover step the chromosomes interchange genes. A local fitness value effects each gene as

$$f(P_i^j) = |\beta_i - P_i^j| \quad (9)$$

Then the crossfire between two individuals consists to keep all individual populations of the first parent which have a local fitness greater than the average local fitness  $f(P_{av}^j)$  and substitutes the remained genes by the corresponding ones from the second parent. Hence, the average local fitness is defined by:

$$f(P_{av}^j) = \frac{1}{K} \sum_{i=1}^K f(P_i^j) \quad (10)$$

Therefore, the mutation operator denotes the phenomena of extraordinary chance in the evolution process. Truly, some useful genetic information regarding the selected population could be lost during reproducing step. As a result, mutation operator introduces a new genetic information to the gene pool [27].

##### 4.5.1. Morphological operations

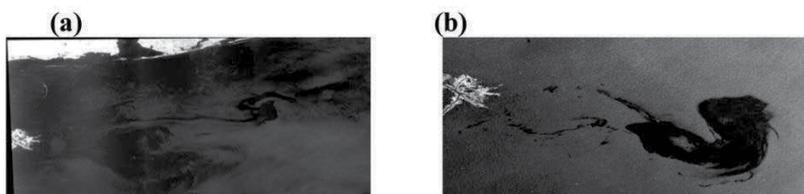
Morphological operation on the selected individuals is performed prior to the cross-over and the mutation process. This is to exploit connectivity property of the RADARSAT-2 SAR and ENVISAT ASAR data [28]. The morphological operators are implemented through reproduction step: (i) closing followed by (ii) opening. In this regard, the accuracy of dark patch segmentations are function of the size and the shape of the structuring element. Therefore, kernel window size of a square of structuring of  $7 \times 7$  is chosen to preserve the fine details of oil spill in RADARSAT-2 SAR and ENVISAT ASAR data [29, 31].

## 5. Results and discussion

### 5.1. Gulf of Mexico

In this study, RADARSAT-2 SAR data with RADARSAT-2 in ScanSAR Narrow B Beam on April 28, 2010 at 11:51:29 UTC is implemented for oil spill detection in the Gulf of Mexico. The Satellite has a Synthetic Aperture Radar (SAR) with multiple polarization modes, including a fully polarimetric mode in which HH, HV, VV and VH polarized data are acquired. Its highest resolution is 1 m in Spotlight mode (3 m in Ultra Fine mode) with 100 m positional accuracy requirement. In ScanSAR Wide Beam mode the SAR has a nominal swath width of 500 km and an imaging resolution of 100 m. An oil platform located 70 km from the coast of Louisiana sank on Thursday April 22, 2010 in the Gulf of Mexico spilling oil into the sea [9].

On these RADARSAT image we can clearly see the evolution of the spill, which has a darker tone than the surrounding water, as well as some boats in the area (Figure 8a). Figure 8 b shows the ENVISAT ASAR data was acquired on May 9, 2010. These data are C-band and had the lower signal-to-noise ratio owing to their HH polarization with a wavelength range of 3.7 to 7.5 cm and the frequency of 5.331 GHz. ASAR can achieve a spatial resolution generally around 30 m. The ASAR is intended for applications which require the spatial resolution of spatial resolution of 150 m. This means that it is not effective at imaging areas in depth, unlike strip map SAR. The azimuth resolution is 4 m, and range resolution is 8 m. This confirms the study of Marghany [31] and [32].



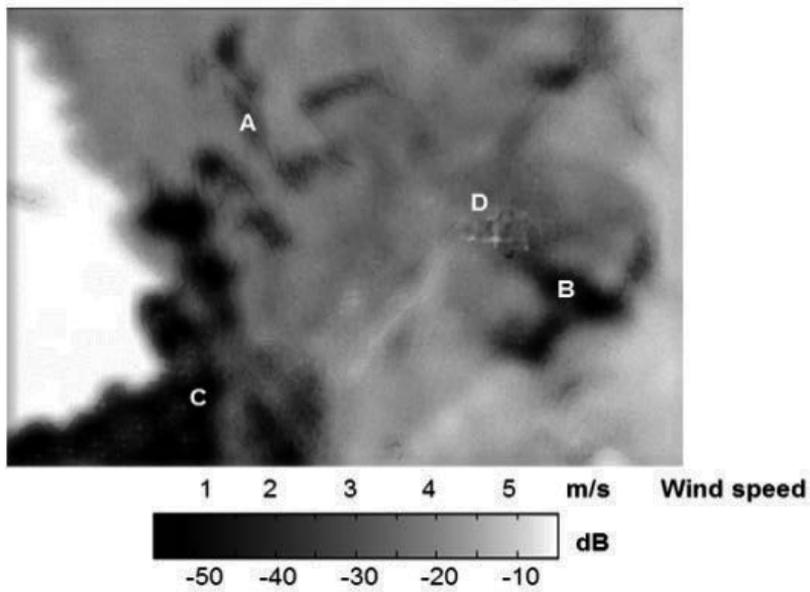
**Figure 7.** MultiSAR data over Gulf of Mexico (a) RADARSAT-2 SAR and (b) ENVISAT ASAR data

### 5.2. Singapore straits

Figure 9 shows the ENVISAT ASAR was acquired on June 3<sup>rd</sup> 2010 after the Merchant ship collided with a Malaysian oil tanker near the Singapore and Malaysian coastal waters. Clearly, there are various of dark patches which are scattered over a large area of coastal waters. The lowest backscatter of -40 dB, -45dB, -50dB are noticed in areas A, B and C, respectively. The highest backscatter of -10 dB is represented ships in area D (Figure 10). In fact, oil spills change the roughness of the ocean surface to smoothness surface which appear as dark pixels as compared to the surrounding ocean [1-22]. Therefore, the speckle caused difficulties in dark patch identifications in SAR data [14, 16]. Further, wind speed was recorded during 3<sup>rd</sup> June 2010 was ranged between 1 to 6 m/s.



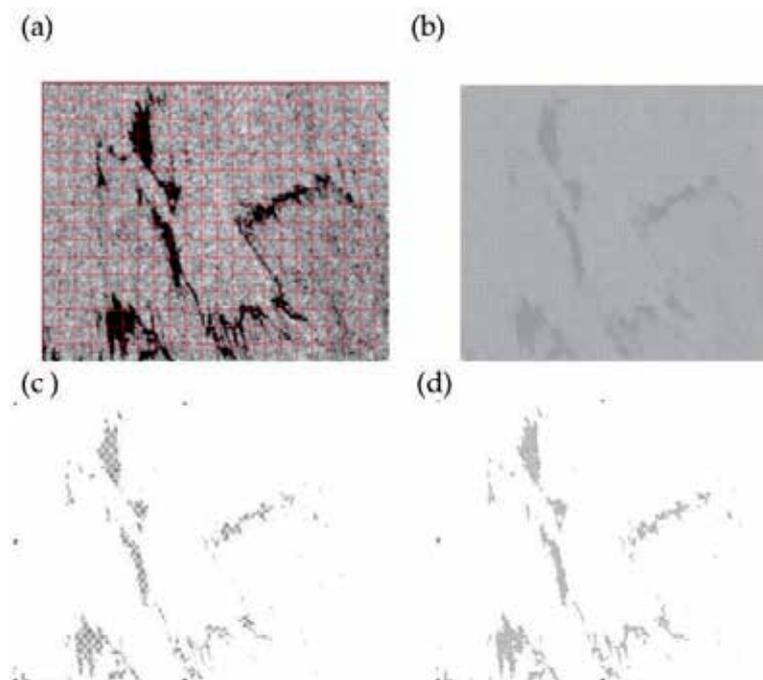
**Figure 8.** ENVISAT ASAR data during 3<sup>rd</sup> June 2010 along Straits of Singapore and South China Sea.



**Figure 9.** Backscatter and wind speed variations in ENVISAT ASAR data.

### 5.3. Genetic algorithm output

However, the result of backscatter values is different as compared to previous studies of Marghany et al., [10] [11] and Marghany and Hashim [12]. This is because of previous studies used different radar sensor of the RADARSAT-1 SAR and these studies have done under different weather and ocean conditions compared to recent work. Figure 11 shows example of the crossover process with 10 individuals. In this 10 individuals, the positive dark patches are represented oil spill pixels while negative dark patches represent the surrounding pixels. Accordingly, every cellular is compared with the corresponding cell in the others to determine either to be positive or negative.



**Figure 10.** Crossover procedures (a) original data, (b) first individual, (c) resulting from an individual prior cancellation, and (d) after cancellation.

In these procedures, cell has a positive value and must be strengthened when cell in the intermediate prototype has a value larger than zero and greater than threshold 's value. In this regard, these cells are represented an oil spill event in ENVISAT ASAR data. On the contrary, the cell represents look-alikes when it has a negative value. As a result, the cell in the intermediate prototype is less than zero and threshold 's value. In this regard, this cell must be diminished. The variation cell value (positive or negative) is a function of dissimilarity of the comparable cells. This study confirms and extends the capabilities of GA introduced by Kahlouche et al., [27].

Clearly the genetic algorithm is able to isolate oil spill dark pixels from the surrounding environment. In other words, look-alike, low wind zone, sea surface roughness, and land are marked by white colour while oil spill pixels are marked all black (Figures 12 and 13). Further, Figures 12 and 13 show the results of the GA, where 100% of the oil spills in the test set were correctly classified. This study is not similar to previous work done by Marghany and Hashim [12]. The dissimilarity is because this work provides the automatic classifier based on GA but Marghany and Hashim work is considered as a semi-automatic tool for oil spill detection.

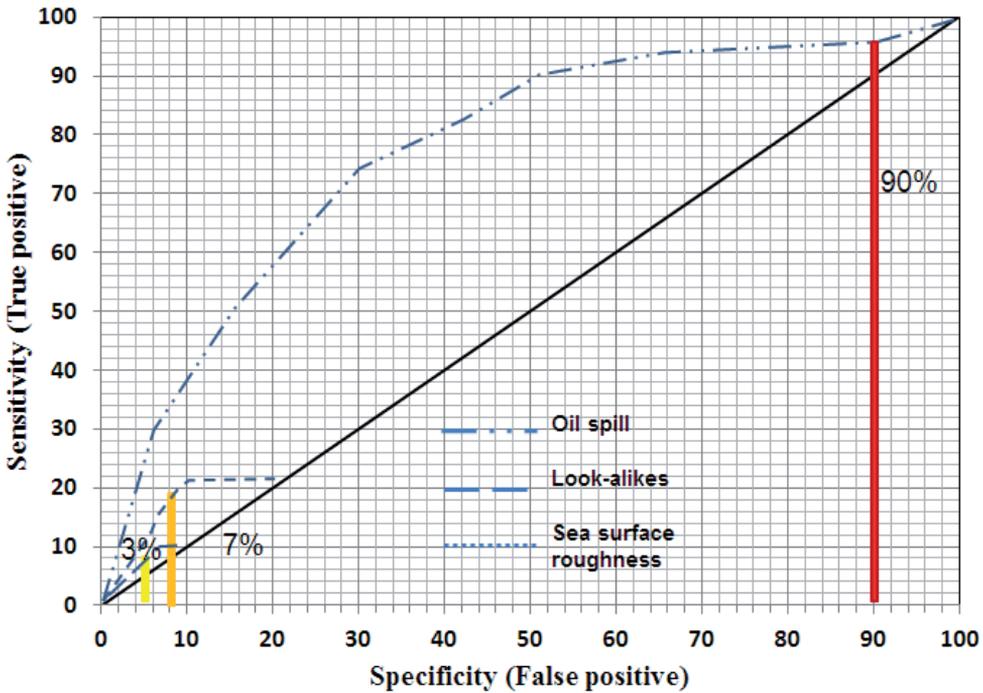


**Figure 11.** Oil spill automatic detection by Genetic Algorithm (GA) using (a) RADARSAT-2 SAR and (b) ENVISAT ASAR in Gulf of Mexico



**Figure 12.** Oil spill automatic detection by Genetic Algorithm (GA) in Singapore Straits

The receiver-operator characteristics (ROC) curve in Figure 14 indicates a significant difference in the discriminated between oil spill, look-alikes and sea surface roughness pixels. In terms of ROC area, the oil spill has an area difference of 85% and 5% for look – alike and 10% for sea roughness and a  $q$  value less than 0.0005 which confirms the study of Marghany et al., [10] [11]. This suggests that genetic algorithm is an excellent classifier to discriminate region of oil slicks from surrounding water features. This conforms the study of Marghany [30] and [31].



**Figure 13.** ROC for oil spill discrimination using Genetic Algorithm (GA).

In general, genetic algorithm contains the crossover procedure. In this regard, a new population is generated in each crossover process. As a result, individual populations are examined by the fitness function and added to the population. Thus, new populations are continuously generated based on the dissimilarities between the two successive fitness values. In addition, the crossover procedure produced a more refined oil spill pattern by despeckle and maintenance the morphology of oil spill pattern features. This is because of the fitness function is used to sustenance the oil spill pixel classification. Indeed, fitness function selected oil spills morphological pattern which is close to the requested spill prototype.

In contrast to previous studies of Fiscella et al., [4] and Marghany and Mazlan [13], the Mahalanobis classifier provides a classification pattern of oil spill where the slight oil spill can distinguish from medium and heavy oil spill pixels. Nevertheless, this study is consistent with Topouzelis et al., [20-22]. In consequence, the genetic algorithm extracted oil spill pixels automatically from surrounding pixels without using different segmentation algorithm as stated at Solberg et al., [19]; Samad and Mansor [18]; Marghany and Mazlan [12].

## 6. Conclusions

This study has demonstrated new approach for oil spill detection in multiSAR data using genetic algorithms. The RADARSAT-2 SAR and ENVISAT ASAR along the Gulf of Mexico and Singapore Straits are involved in this study. The study shows that crossover process, and the fitness function generated accurate pattern of oil slick in multiSAR data. Further, the study also shows that genetic algorithm provides accurate pattern of oil slick in SAR data. This shown by 90% for oil spill, 3% look-alike and 7% for sea roughness using the receiver –operational characteristics (ROC) curve. The Genetic algorithm also shows excellent performance in both RADARSTA-2 and ENVISAT ASAR data. In conclusion, Genetic Algorithm can be used as automatic detection tool for oil spill in multiSAR satellite data such as RADARSAT-2 SAR and ENVISAT ASAR.

## Author details

Maged Marghany\*

Institute of Geospatial and Science Technology (INSTEG), University of Technology, Malaysia

## References

- [1] Adam, J.A.: Specialties: Solar Wings, Oil Spill Avoidance, *On-Line Patterns*. IEEE Spect. 32: (1995) 87--95
- [2] Aggoune M.E., Atlas L.E., Cohn D.A., El-Sharkawi, M.A. and Marks, R. J.: Artificial Neural Networks For Power System Static Security Assessment. IEEE Int. Sym. on Cir. and Syst.. Portland, Oregon. (1989) 490--494
- [3] Brekke, C. and Solberg, A.: Oil Spill Detection by Satellite Remote Sensing. Rem. Sens. of Env. 95(2005)1--13
- [4] Fiscella B., Giancaspro A., Nirchio, F., Pavese, P. and Trivero, P.: Oil Spill Detection Using Marine SAR Images. Int. J. of Rem. Sens. 21: (2000) 3561--3566

- [5] Frate, F. D., Petrocchi, A., Lichtenegger, J. and Calabresi, G.: Neural Networks for Oil Spill Detection Using ERS-SAR Data. *IEEE Tran. on Geos. and Rem. Sens.* 38: (2000) 2282--2287
- [6] Hect-Nielsen, R.: Theory of the Back Propagation Neural Network. *Proc. of the Int. Joint Conf. on Neu. Net.* IEEE Press, I: (1989) 593--611
- [7] Marghany, M. and Hashim, M.: Comparative algorithms for oil spill detection from multi mode RADARSAT-1 SAR satellite data. *Lecture Notes in Computer Science*. D. Taniar et al. (Eds.). *Computational Science and Its Applications – ICCSA 2011*, 6783: (2011) 318--329
- [8] Marghany, M.: RADARSAT Automatic Algorithms for Detecting Coastal Oil Spill Pollution. *Int. J. of App. Ear. Obs. and Geo.* 3: (2001) 191--196
- [9] Marghany, M.: RADARSAT for Oil spill Trajectory Model. *Env. Mod. and Sof.* 19: (2004) 473--483
- [10] Marghany, M., Cracknell, A. P. and Hashim, M.: Modification of Fractal Algorithm for Oil Spill Detection from RADARSAT-1 SAR Data. *Int. J. of App. Ear. Obs. and Geo.* 11: (2009) 96--102
- [11] Marghany, M., Cracknell, A. P. and Hashim, M.: Comparison between Radarsat-1 SAR Different Data Modes for Oil Spill Detection by a Fractal Box Counting Algorithm. *Int. J. of Dig. Ear.* 2: (2009) 237--256
- [12] Marghany, M., Hashim, M. and Cracknell, A.P.: Fractal Dimension Algorithm for Detecting Oil Spills Using RADARSAT-1 SAR. In Gervasi O. and Gavrilova M. (Eds.): *LNCS, Springer-Verlag Berlin Heidelberg Part I*: (2007) 1054--1062
- [13] Marghany, M. and Hashim, M.: Texture Entropy Algorithm for Automatic Detection of Oil Spill from RADARSAT-1 SAR data. *Int. J. of the Phy. Sci.* 5: (2010) 1475--1480
- [14] Michael, N. :*Artificial Intelligence: A guide to Intelligent Systems*. 2nd editon, Harlow, England: Addison Wesley (2005)
- [15] Migliaccio, M., Gambardella, A. and Tranfaglia, M.: SAR Polarimetry to Observe Oil Spills. *IEEE Tran. on Geos. and Rem. Sen.* 45: (2007) 506--511
- [16] Mohamed, I. S, Salleh, A.M. and Tze, L.C.: Detection of Oil Spills in Malaysian Waters from RADARSAT Synthetic Aperture Radar Data and Prediction of Oil Spill Movement. *Proc. of 19th Asi. Conf. on Rem. Sen.* Hong Kong, China, 23--27 November. *Asian Remote Sensing Society, Japan*, 2: (1999) 980--987
- [17] Provost, F. and Fawcett, T.: Robust classification for imprecise environments. *Mach. Lear.*42: (2001)203--231
- [18] Samad, R., and Mansor, S.B.: Detection of Oil Spill Pollution Using RADARSAT SAR Imagery. *CD Proc. of 23rd Asi. Conf. on Rem. Sens.* Birendra International Convention Centre in Kathmandu, Nepal, November 25-29, 2002, *Asian Remote Sensing* (2002)

- [19] Skrunes, S., C. Brekke, and T. Eltoft, "An Experimental Study on Oil Spill Characterization by Multi-Polarization SAR," in Proc. European Conference on Synthetic Aperture Radar, Nuremberg, Germany (2012) 139--142
- [20] Topouzelis, K., Karathanassi, V., Pavlakis, P. and Rokos, D.: Potentiality of Feed-Forward Neural Networks for Classifying Dark Formations to Oil Spills and Look-alikes. *Geo. Int.* 24, (2009)179--19
- [21] Topouzelis, K., Karathanassi, V., Pavlakis, P. and Rokos, D.: Detection and Discrimination between Oil Spills and Look-alike Phenomena through Neural Networks. *ISPRS J. Photo. Rem. Sens.* 62: (2007) 264--270
- [22] Topouzelis, K.N.: Oil Spill Detection by SAR Images: Dark Formation detection, Feature Extraction and Classification Algorithms. *Sens.* 8: (2008) 6642--6659
- [23] Trivero, P., Fiscella, B. and Pavese, P.: Sea Surface Slicks Measured by SAR, *Nuo. Cim.* 24C, (2001) 99--111
- [24] Trivero, P., Fiscella, B., Gomez, F., and Pavese, P.: SAR Detection and Characterization of Sea Surface Slicks. *Int. J. Rem. Sen.* 19: (1998) 543--548
- [25] Velotto, D., M. Migliaccio, F. Nunziata, and S. Lehner, "Dual-Polarized TerraSAR-X Data for Oil-Spill Observation," *IEEE Trans. Geosci. Remote Sens.*, 49: (2011) 4751--4762
- [26] Chaiyaratana, N.; Zalzala, A.M.S: Recent developments in evolutionary and genetic algorithms: theory and applications. *Genetic Algorithms in Engineering Systems: Innovations and Applications,GALESIA 97. Second International Conference On 2-4 Sept. 1997, Glasgow, (1997) 270--277*
- [27] Kahlouche,S., K.Achour, and M. Benkhelif: Proceedings of the 2002 WSEAS International Conferences, Cadiz, Spain, June 12-16, 2002. [www.wseas.us/e-library/conferences/spain2002/papers/443-164.pdf](http://www.wseas.us/e-library/conferences/spain2002/papers/443-164.pdf): (2002)1--5
- [28] Gautam, G., and B.B. Chaudhuri: A distributed hierarchical genetic algorithm for efficient optimization and pattern matching: *Pattern Recognition Journal*, 40: (2007) 212--228
- [29] Sivanandam, S.N, and Deepa S.N: *Introduction to Genetic Algorithms*, by Springer Berlin Heidelberg New York (2008)
- [30] Marghany M. Genetic Algorithm for Oil Spill Automatic Detection from Envisat Satellite Data. In Beniamino Murgante, Sanjay Misra, Maurizio Carlini, Carmelo M. Torre, Hong-Quang Nguyen, David Taniar, Bernady O. Apduhan, and Osvaldo Gervasi. *Computational Science and Its Applications – ICCSA 2013*, 7972, pp 587-598.
- [31] Marghany M. Genetic Algorithm for Oil Spill Automatic Detection from Multisar Satellite Data. *Proceedings of the 34th Asian Conference on Remote Sensing 2013*. Bali – Indonesia, October 20-24, 2013. pp. SC03-671-SC0-3677.



---

# HF Radar Network Design for Remote Sensing of the South China Sea

---

S. J. Anderson

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/57599>

---

## 1. Introduction

HF surface wave radar (HFSWR) is a highly cost-effective technology for remote sensing of ocean surface conditions and monitoring of ship traffic; several hundred radars of this type are in operation around the world. While an individual radar, operating alone, is able to provide a great deal of useful information, the integration of multiple radars into a network results in a system capability which is far more than the sum of its parts. For example, an estimate of a ship's velocity vector can be obtained in seconds, not tens of minutes or hours as is the case with a single radar. As another example, ocean currents can be estimated unambiguously, even in the presence of eddies and upwelling. Apart from these well-known considerations, there is a class of benefits which has special significance for long range HFSWR systems, namely, the potential for bistatic operations. As shown later in this paper, the fusion of monostatic and bistatic measurements enhances radar performance in a number of ways, a gain which is especially important for very long range operations.

While some HFSWR systems have been designed and deployed with a single mission in mind, it is increasingly recognised that the versatility of this technology supports a variety of applications. For instance, one might wish to detect and track shipping but also to measure surface currents so that risks of collision or grounding can be minimised and any transport of pollution predicted. In addition, information on sea state is of considerable economic value for ship routing, planning for offshore wave energy extraction, coastal development, port operation scheduling, search and rescue, fishing, tourism and recreational activities, so extraction and dissemination of environmental data would be welcomed by a wide range of user communities. Of course, these various applications will have relative priorities which vary with location, time of day and season, as will the radar's ability to accomplish them.

The physics of HFSWR dictates that many site-dependent factors contribute to the accuracy, reliability and availability of the various radar products. Moreover, the sensitivity to network configuration varies according to the type of measurement (or 'mission') being undertaken. Thus the choice of geographical sites which together comprise the network must reflect not only the family of radar outputs required but also their relative priorities. The resulting optimisation problem is extremely challenging.

An effective methodology for optimising HFSWR network design for the case where multiple missions must be addressed has been developed recently and demonstrated in the context of a hypothetical two-radar system deployed in the Strait of Malacca (Anderson, 2013). The results of that study demonstrated that quite disparate criteria can be accommodated within a genetic algorithm framework and confirmed that the method yielded the true optimum site configurations. Yet that study left a key question unanswered. In practice we are unlikely to be satisfied knowing that our choice of sites is the best for a given budget; we want to know that the network will meet prescribed levels of performance. This could well mean that, in a particular situation, a mix of quite different radar types would be required, adding another dimension to the network design problem.

In this chapter we review the genetic algorithm methodology for multi-objective optimisation in the HFSWR context, and show how it can be extended to handle the inverse problem of designing networks to meet specified performance levels. In order to illustrate the steps involved in formulating and applying the methodology, the discussion is framed in the context of a specific scenario: the design of an HFSWR network for providing surveillance and remote sensing of the South China Sea.

The spatial resolution and ultimate sensitivity of HFSWR is primarily a function of radar design, but performance in its various candidate roles is also dependent on a wide variety of geophysical factors, lithospheric, oceanic, atmospheric and ionospheric. Further, the relative priority of different missions reflects economic, geopolitical and strategic considerations. As all these aspects would (or should) be taken into account by network designers, it is appropriate to examine ways in which they can be incorporated in the objective functions employed for optimisation. In the following section we set the scene for the subsequent analysis by reviewing the physical environment and the associated human activities which an HFSWR radar network might be expected to monitor. Next we outline the capabilities and limitations of HFSWR in this context, based on the nominal performance of four existing radar systems. Once the radar capabilities have been established, we turn to the central issue, namely, that of formulating the network design problem in mathematical terms, which leads us to focus on evolutionary algorithms of nonlinear optimisation. Here the genetic algorithm approach of Anderson (2013) is emphasised, as it lends itself naturally to multi-objective optimisation, though in order to handle the enormous computational burden in the present case, a recently-reported convergence acceleration technique (Anderson et al, 2013) is introduced. We proceed to describe practical methods for constructing chromosomes and objective functions for a number of missions, illustrating these by relating them to the South China Sea context.

## 2. The South China Sea

### 2.1. Physical geography

Formally the South China Sea extends from Bangka Belitung, between Sumatera and Borneo, to the northern extremity of Taiwan, and from the Gulf of Thailand to the Philippines, as shown in Figure 1.

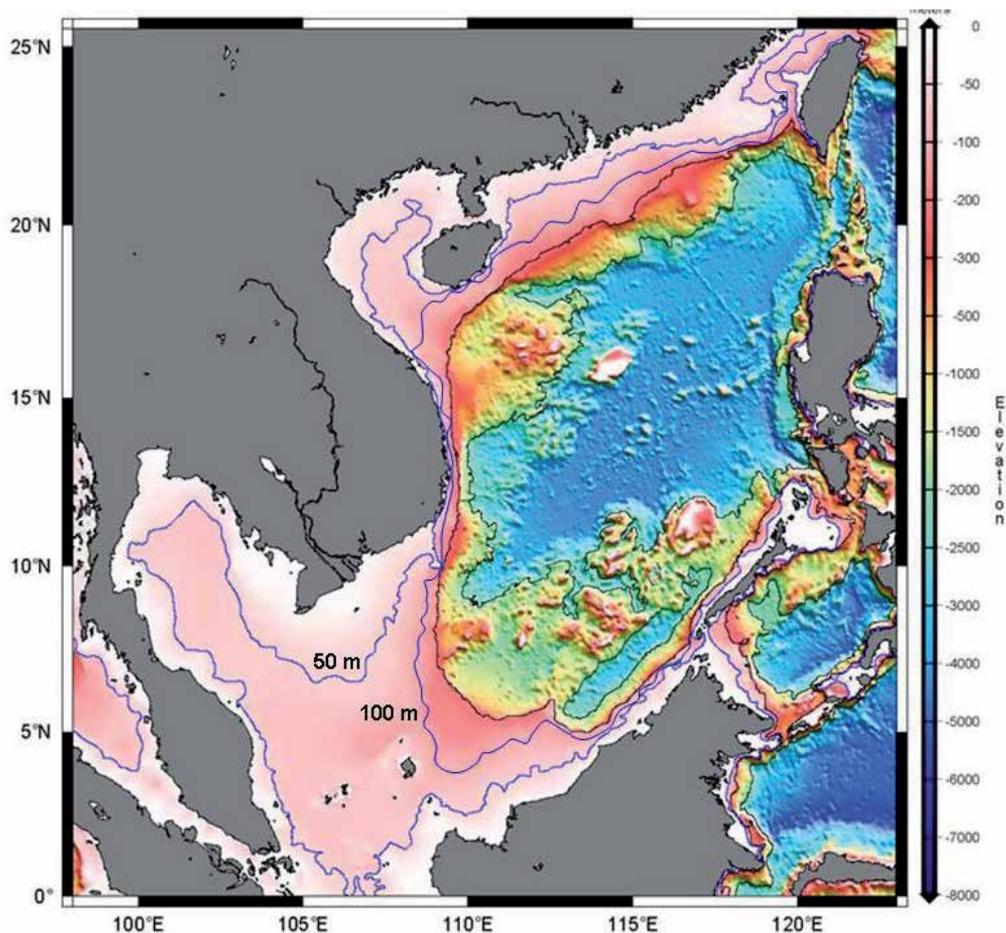
Within its area of some 3,500,000 square kilometres lie several hundred islands, of which most are grouped into two clusters, the Paracel and Spratly Island chains. A great many of the islands are little more than exposed reefs and even the important Spratly Island group has a total land area of less than 5 square kilometres and a maximum elevation above sea level of only 4 metres. Some important features are entirely submerged, as is the case with Macclesfield Bank – actually an atoll-which is on average about 10 m below sea level, yet has an area of some 6500 square kilometres. Scarborough Shoal (*aka* Panatag Shoal), has reefs and small islets above water amounting to only a few hectares, surmounting an area of some 150 square kilometres of about 15 metres depth. beyond which the sea floor drops away rapidly to a depth of several kilometres. Only a handful of islands are large enough to be home to an airstrip and some of these facilities may not survive even a modest rise in sea level.

The large-scale bathymetry of the South China Sea is particularly striking. South of a line joining Brunei to the southern tip of Vietnam, the depth is less than 100 metres, but north of that line the sea floor descends rapidly to 1000 – 5000 metres, except around the island chains and atolls. With the exception of a few narrow but deep channels between Luzon and Taiwan, connecting to the East China Sea, the South China Sea is essentially a basin.

Ocean surface conditions are influenced by the orography of adjacent land masses which helps steer the prevailing winds. In the case of the South China Sea the principal land feature that is relevant to HF radar system performance is the mountain range along almost the entire coast of Vietnam.

### 2.2. Meteorology and oceanography

The wind regime over the South China Sea is dominated by the monsoon winds, punctuated by mesoscale systems such as tropical cyclones. During the boreal winter, the northeasterly winter monsoon winds impose a fairly uniform stress over most of the South China Sea, whereas in summer, June, July and August, the southwesterly monsoon winds show somewhat greater spatial variability, especially south of about 6°N. Average wind speeds in winter tend to fall in the range 8 – 12 m/s whereas the southwesterly summer monsoon winds are typically approximate 6 – 8 m/s in the Southern SCS and somewhat less in the northern SCS. Highly variable winds and surface currents are observed during the transitional periods. Moreover, synoptic systems often pass by the SCS and causes temporally and spatially varying wind fields. Severe weather most frequently takes the form of an increase in the strength of the prevailing monsoon winds or as meso-scale disturbances concentrated in either of two regions: a localised area east of the southern part of Vietnam, centred on 10°N, 110°N, and the band between Luzon and southern China. The mean wind regimes for summer and winter are shown in Figure 2.



**Figure 1.** The bathymetry of the South China Sea (adapted from the World Data System for Marine Environmental Sciences, <http://www.wdc-mare.org/>)

Tropical cyclones form in the waters between 12°N and 24°N, usually making landfall over Hong Kong and southern China, the north and central coasts of Vietnam or the northern Philippines. The most severe cyclones occur to the east of the Philippines and Taiwan, as shown in Figure 3 but, even so, the South China Sea north of 15°N is occasionally subjected to category 3 and 4 events.

Wave and current distributions due to the wind forcing are less uniform than the wind fields. Significant waveheight (SWH) distributions are higher in the northern and central SCS (north of 10°N) than in the southern SCS (south of 10°N) with upper quartile values exceeding 2.25 m. (The Wavewatch III model has been found to yield fairly accurate results (Chu et al, 2004), so serves as a useful adjunct in modelling radar performance.) As shown in Figure 4, the orientation of the high SWH region coincides with the orientation of the monsoon winds.

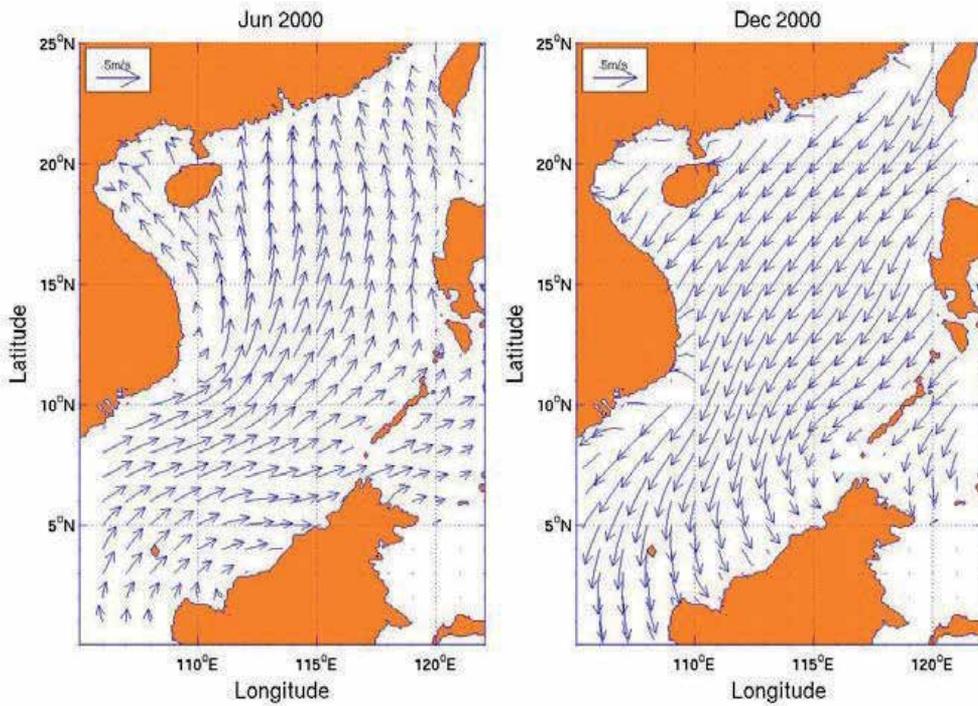


Figure 2. Synoptic-scale wind patterns during the summer and winter monsoon seasons (from Chu et al, 2003)

### Tropical Cyclones, 1945–2006

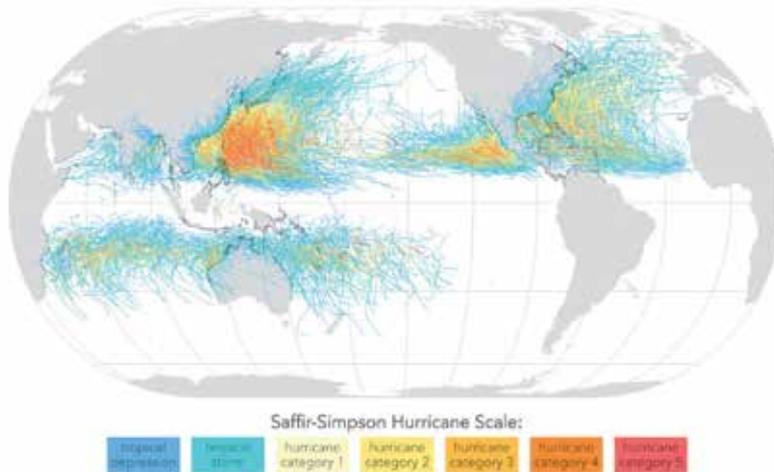
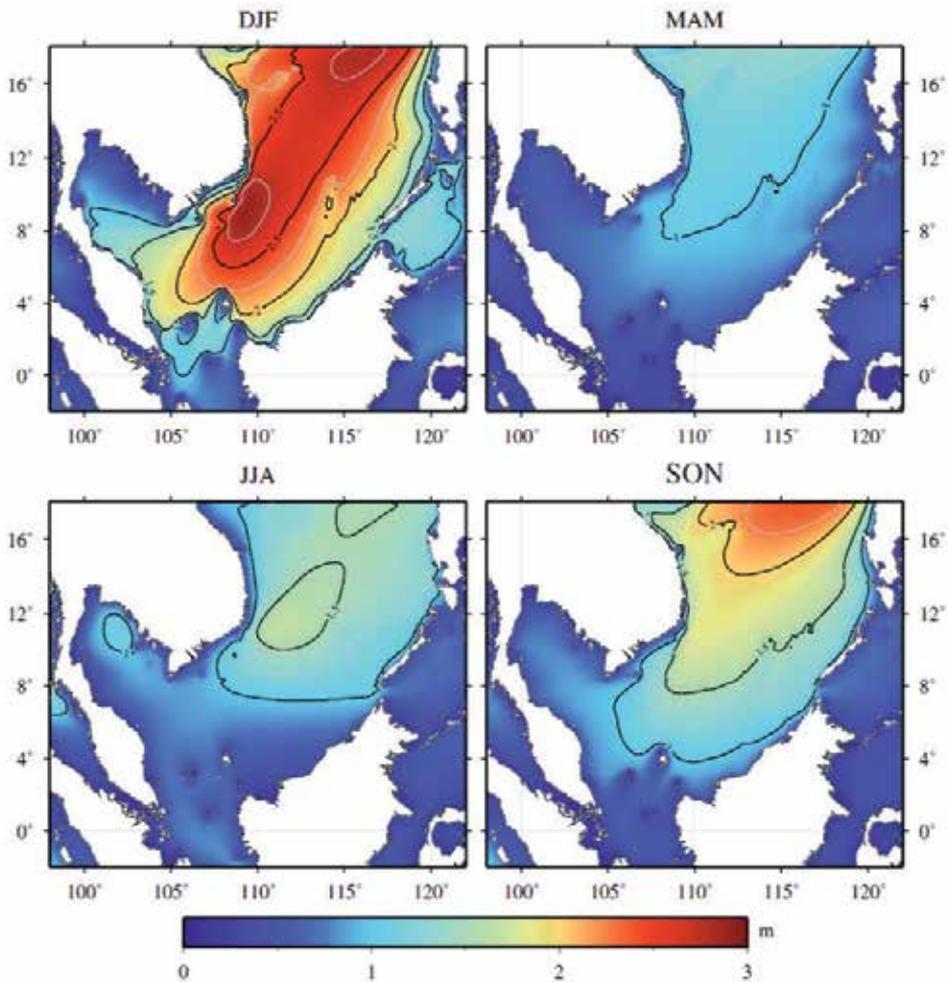


Figure 3. The distribution of tropical cyclones over the period 1945 – 2006.



**Figure 4.** Seasonal variation of mean wave height for the period 1979-2009, WaveWatch III hindcast (Mirzaei et al, 2013). Note that the dominant wave direction is aligned with the monsoon winds, which is to say southward-propagating in DJF and northward-propagating in JJA.

The prevailing winds have a direct effect on the surface water currents of the shelf region. The current speeds are about 0.6 knots to the SW during the winter monsoon. They change to 0.2 to 0.4 knots to the NE during the summer monsoon. Stronger currents flow adjacent to the Vietnamese coast in particular, attaining speeds in excess of 1 m/s, while the islands of the Spratly archipelago can induce fairly complex local current variations. Primary or climatological current patterns in summer and winter are shown in Figure 5, but these convey an incomplete picture of the flow field. To gain a better understanding of the complexity of the current distribution, consider Figure 6, which shows the outputs of a detailed hydrodynamic model of the current field in the northern and central South China Sea. A key feature of this model is the inclusion of the wind-induced current, which was found to dominate the

geostrophic current in many places. Moreover, note the appearance of the mesoscale eddies. These have been validated by observation. In a similar vein, Marghany (2009, 2011, 2012) has shown how local current patterns can be extracted from spaceborne SAR using observations off the east coast of peninsular Malaysia. The lesson to be drawn from this kind of modelling is that the flow field has significant structure on length scales of 50 km or less; given that the cross-range dimension of a typical HF radar resolution cell at long range may approach this magnitude, it is evident that HFSWR could provide unique validation data, though conventionally measured Doppler spectra will not always have discrete Doppler shifts and hence current velocity estimation will be compromised on those occasions.

One particular form of current perturbation which has received a lot of attention by the HF radar community is that associated with a tsunami (Lipa et al, 2012). It has been demonstrated that HFSWR is an effective tool for early warning of tsunamis provided that the bathymetry is favourable, which is to say reasonably shallow so that the speed of the tsunami is much reduced from its high deep water value. As Figure 1 shows, the southern part of the sea occupied by the Sunda Shelf and the north-western margins certainly satisfy this requirement.

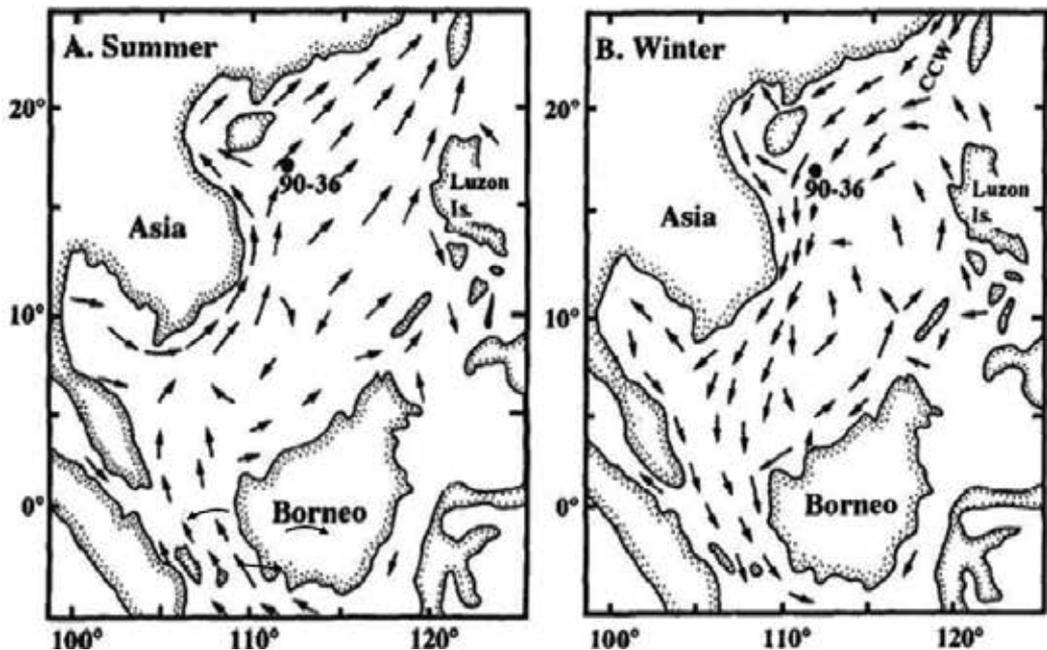
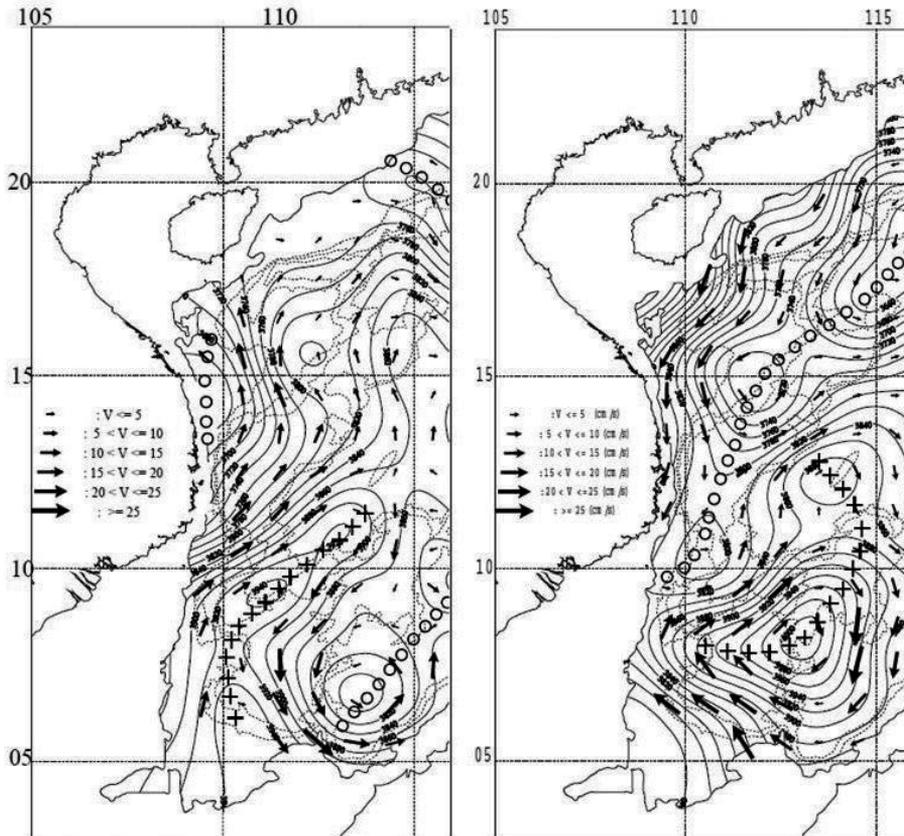


Figure 5. Primary currents during the summer and winter monsoon periods (Chen et al, 1985)

The SW winds blowing along the SW-to-NE part of continental shelf may induce upwelling during the summer, bring nutrients to the eutrophic zone on the outer portion of the shelf and, enhance primary production of the waters (Wang and Kester, 1988). The seasonal stratification stimulates the seasonal changes in primary production and nutrient cycling, with a strong signature evident in the high chlorophyll distributions in two coastal upwelling regions: the northwestern Luzon in winter and the eastern coast of Vietnam in summer. Mesoscale eddies provide another mechanism responsible for seasonal and interannual variability of the surface chlorophyll distribution.



**Figure 6.** Modelled surface current fields for the central and northern South China Sea as computed with a 2-D numerical code (Ninh et al, 2000); (a) summer, and (b) winter.

The South China Sea surface layer is 50-100m thick. Its distributions are different in winter and summer. The monsoon-driven reversal of surface currents affects the temperature and salinity of the water masses, and hence the conductivity which impacts on HF radar performance. In winter, due to the influence of the northeast monsoon, the temperature increases progressively from the coast to the outer sea and the salinity decreases progressively from north to south. The temperature ranges from 22°C in the north to 26°C in the south, while the salinity varies

from 33.2 to 34.5 PSU. In summer, due to the influence of southwest monsoon, the surface temperature is generally 28-29°C and the salinity is low – near 32 PSU – in the north and south, and high – about 33.6 PSU – in the central region. In summer, the SW monsoon results in the large increase in rainfall and river discharges. This results in the reduction of salinity in the coastal waters and the production of seasonal pycnoclines. Particularly low salinity occurs off the east coast of peninsular Malaysia.

The diurnal and semi-diurnal tides are of about equal magnitude in the South China Sea, though the latter is more effective at generating internal waves. These are exceptionally strong in the northern region, where they are generated in the Luzon Strait before propagating westward. Amplitudes reach 200 m with horizontal scales upwards of 200 km. These internal waves take about 4 days to cross the South China Sea, modulating the surface gravity wave field as they progress and hence influencing the radar scattering properties of the sea surface.

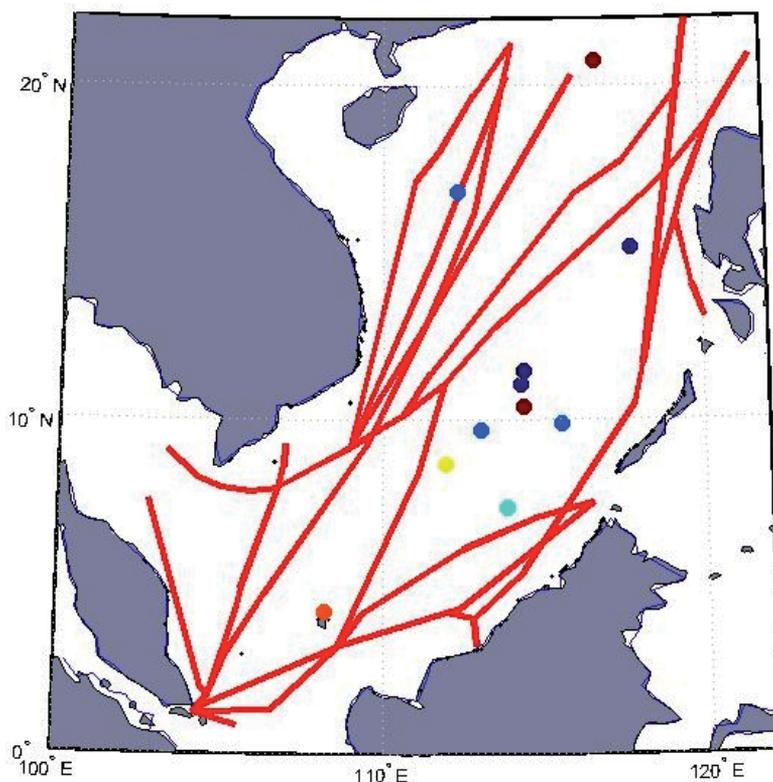
### 2.3. Shipping

The volume of shipping activity in the South China Sea can be illustrated by a few key statistics:

- i. nearly half the world's annual merchant fleet tonnage moves through its waters, carrying commodities valued at over \$5 trillion
- ii. one third of global oil tanker traffic and over half of global LNG traffic crosses the South China Sea, most from the Strait of Malacca but the very largest supertankers via the Sulu Sea
- iii. ore carriers, predominantly from Australia, transport roughly half a billion tonnes of iron ore and a similar amount of coal through the South China Sea annually
- iv. six of the world's ten largest ports lie on the coastlines of the South China Sea
- v. the annual growth rate for liquid petroleum fuels consumption in recipient countries – mainly China and Japan and South Korea – is presently 2.6 %, while that for natural gas is 3.9 %. For Australian minerals the figure is 4.6%
- vi. over half a billion people live within 100 miles of its margins
- vii. perhaps as many as 18,000 small fishing boats ply its waters

The major shipping routes are shown on Figure 7, using data derived from Wang et al, 2013).

Given the density of traffic, it is perhaps not surprising that shipping hazards in the South China Sea continue to take a toll on vessels in transit, as exemplified by several recent incidents: the sinking of the Bright Ruby (severe storm, November 2011), Royal Prime (hit reef, and sank, December 2012), Harita Bauxite (sank after engine failure, February 2013), Jung Soon (sank after hull failure, September 2013). Another form of hazard is piracy, for which the South China Sea was once notorious. While less frequent than a few years ago, hijacking and armed robbery remain a significant threat in some waters. Mimicking the 'mothership' refuelling station tactic used by pirates off the coast of Somalia, pirates in Indonesia and Malaysia tend to camp on a small island near to narrow shipping lanes and launch their strikes from there. Pirates in South



**Figure 7.** Principal shipping lanes through the South China Sea, and islands selected for use in radar mission definitions.

East Asia also tend to launch their attacks at night, which makes it much harder for ship captains to spot them coming. Between 2008 and 2010, 57 incidents of ‘cluster piracy’ took place around the Abambas / Natuna / Tambalan corridor. In the first six months of 2013, attacks involving pirates boarding vessels and assaulting the crew were recorded in the Singapore Straits, in Malaysian waters, in the Straits of Malacca and in the Philippines. Within the main body of the South China Sea, 2013 escaped serious incident.

Another consequence of heavy ship traffic is oil pollution, both accidental and deliberate, such as that caused by tankers flushing their tanks on the voyage back to the Middle East. Offshore facilities and undersea pipelines are other man-made sources.

#### 2.4. Economic activity

It is evident from the preceding discussion that ‘through traffic’ is critically important to the destination countries of China, Japan and South Korea, but, for the littoral states around the South China Sea, fishing is the most vital maritime activity, as it has been for centuries. Fish protein constitutes nearly a quarter of the average Asian diet and demand continues to grow

strongly. Thus, whereas the extent of oil and gas reserves beneath the South China Sea remains questionable, the value and importance of its fisheries and aquaculture is not in doubt. It is therefore of great concern that the relative stability of traditional fishing practices is now threatened by over-fishing, together with rising water temperatures which appear to be resulting in migration of fish populations, primarily further north. These developments are stoking tensions between the countries whose populations depend on accessible and reliable stocks.

It is widely reported that the South China Sea holds immense untapped natural reserves of oil and gas, and that the contested ownership of the Spratly Islands and other parts of the sea is primarily a fight for these resources. It is certainly the case that confrontation and armed skirmishes have taken place where exploration has been pursued in disputed waters. Yet a considered analysis does not support the more extreme assertions regarding the magnitude of the reserves in the contested regions. The most recent assessment by the US Energy Information Administration estimates that the total of proven and probable reserves in South China Sea amounts to approximately only 11 billion barrels of oil and 190 trillion cubic feet of natural gas. Another US expert source places the figure for oil at 2.5 billion barrels. Allowing for additional reservoirs in under-explored areas, the EIA says, could add between 5 and 22 billion barrels of oil and 70 to 290 trillion cubic feet of gas. These figures contrast with those of the Chinese National Offshore Oil Company which estimates undiscovered reserves amount to 125 billion barrels of oil and 500 trillion cubic feet of natural gas. In the absence of detailed prospecting, the actual quantities cannot be known with any certainty. What is undeniable is that the preponderance of known resources resides in the uncontested areas close to the coasts of the surrounding countries, especially Vietnam, Malaysia and Brunei. Thus the fierce competition for control, if not ownership, of the islands, reefs and shoals of the South China Sea, is probably driven by a combination of factors, economic, political and strategic.

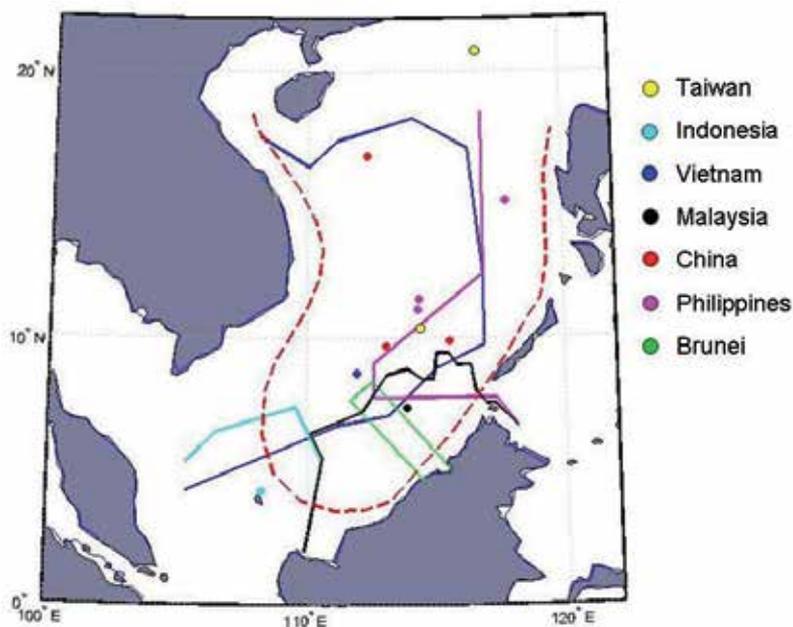
We note too that Malaysian researchers have identified potentially valuable elements in seabed sediment, including manganese, zinc, chromium, lead, copper and aluminium.

## **2.5. Strategic and geopolitical issues**

The South China Sea has a long history of tension and conflict. Much of this derives from overlapping territorial claims and disputed ownership of maritime features, as indicated in Figures 7 and 8, exacerbated by a race to exploit the maritime zone's natural resources. In addition, there is also a growing element of overt strategic rivalry and nationalism, which poses a substantial risk to regional security and prosperity. Specific areas of dispute include:

- the Spratly Islands, disputed between the People's Republic of China, the Republic of China, and Vietnam, with Malaysia, Brunei, and the Philippines claiming part of the archipelago
- the Paracel Islands, disputed between the People's Republic of China, the Republic of China, and Vietnam
- the Pratas Islands, disputed between the People's Republic of China and the Republic of China

- the Macclesfield Bank, disputed between the People's Republic of China, the Republic of China, the Philippines, and Vietnam
- the Scarborough Shoal, disputed between the People's Republic of China, the Philippines, and the Republic of China.



**Figure 8.** Territorial claims over the South China Sea, together with occupied islands.

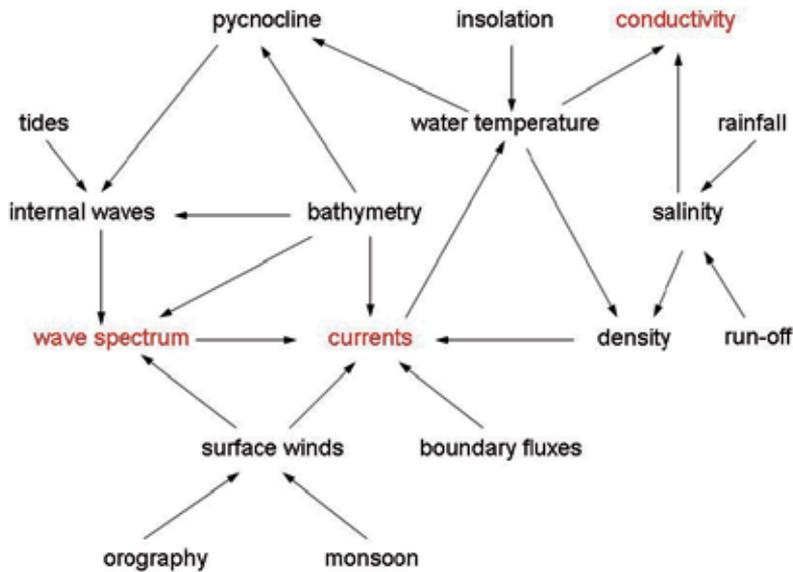
Many detailed discussions of these issues can be found in the open literature (International Crisis Group 2012a, 2012b); here it suffices to make the point that timely, comprehensive, robust and persistent surveillance can be a useful means of establishing trust and defusing incidents which could spiral out of control.

### 3. Capabilities and limitations of HF surface wave radar

#### 3.1. The suitability of HFSWR for maritime remote sensing and surveillance

The physical quantities which impact directly on HF radar capability in both its remote sensing and surveillance roles are (i) water electrical conductivity, (ii) surface currents, and (iii) the geometry and dynamics of the sea surface, usually represented as a spectrum of surface gravity waves. It is no exaggeration to state that radar performance in any of its missions is highly dependent on these primary quantities. Moreover, the primary quantities are coupled with other geophysical variables and processes, as illustrated in Figure 10. Therefore, as part of the

network design procedure, it is absolutely essential to take into account the kind of information presented in Section 2. Figure 10 shows the linkages between the primary quantities and other geophysical variables and processes. From this figure it is apparent that, by appropriate analysis and interpretation of the radar echoes, it may be possible in some circumstances to use the primary measurements of currents and wave spectra to infer secondary phenomena, surface winds being the prime example.



**Figure 9.** The relationships between the geophysical parameters and phenomena which impact on HF radar performance

It is evident from the discussion in Section 2 that real-time monitoring of these environmental conditions and ship traffic over the South China Sea could have substantial value for a wide range of users. Yet the practical application of any remote sensing technology requires that we first establish whether the coverage, resolution and accuracy of the measurements are commensurate with the needs of the users. To illustrate this step, we shall consider the nominal performance of four well-known HFSWR products.

The commercial marketplace for so-called ‘oceanographic’ HFSWR systems is dominated by two manufacturers: CODAR, with its Seasonde radars (Barrick, 1998), and Helzel Messtechnik, with its WERA systems (Helzel et al, 2010). These radars each cost in the vicinity of 0.5M\$ per system consisting of one transmitting station and one receiving station, and have excellent track records for delivering ocean current information at ranges out to 200 km, with sea state measurements available for significantly shorter ranges. While some extravagant claims are made about the ability of these low power radars to detect ship targets at ranges of several hundreds of kilometres, experience has tended to show that reliable detection is confined to 50 – 150 km, depending on target type, time of day, and other factors. These radars have quite

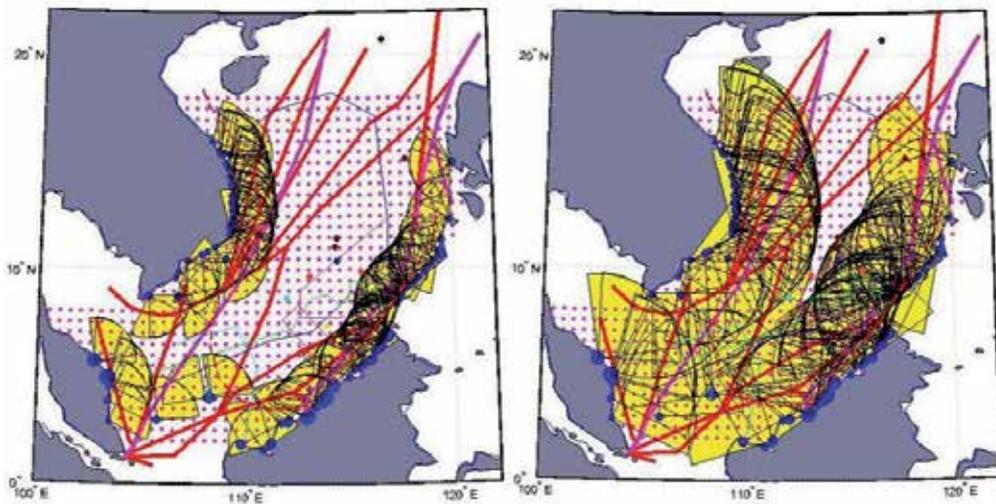
different characteristics so it is appropriate to include them both in the list of options for a heterogeneous network.

The South China Sea is roughly 1000 km east-to-west at 10° N so it is clear that oceanographic radars based on undisputed territories are not able to provide comprehensive surveillance. The possibility of deploying radar systems on small islands may change this assessment somewhat, as discussed later in this paper, but a priori it would seem that radars with far superior long range performance are required if comprehensive surveillance is to be achieved. For the study reported here, two commercial HFSSWR radars were chosen to represent such 'military-class' systems: Raytheon's SWR-503 (Ponsford, 2012) and Daronmont Technology's SECAR radar (Anderson et al, 2003). Each of these systems has demonstrated ship detection at ranges well in excess of 400 km. It is important to point out that the diverse observations and opinions from which these figures were inferred correspond to a variety of environmental conditions, so the estimates are really just indicative. Still, the table serves to provide numerical values for the purpose of exercising the network optimisation suite.

Observable	Typical performance			
	low-cost civilian radar		military radar	
	max. range (km)	accuracy	max. range (km)	accuracy
surface current	60 - 200	± 0.02 – 0.20 m/s	350 - 450	± 0.02 – 0.10 m/s
wave height	30 - 100	± 10 – 25 %	150 - 350	± 10 – 20 %
wind direction	50 - 180	± 30° - 60°	320 - 400	± 20° - 30°
wind speed	30 - 150	± 20 %	150 - 350	± 20 %
large ship	50 - 180	± 0.5 - 3 km	300 - 450	± 0.5 – 3 km
fishing boat	20 - 65	± 0.5 – 2 km	120 - 280	± 0.5 – 2 km
small boat	10 - 45	± 0.5 – 1 km	70 - 150	± 0.5 – 1 km

**Table 1.** Capabilities of selected HFSSWR systems, expressed in terms of typical maximum ranges at which measurements can be made reliably

A natural first test is to see whether even the most potent (and expensive) network could deliver the desired coverage. In order to obtain a large and realistic set of possible sites in the present case, we visually searched the coastlines around the South China Sea as presented in Google Earth™, selecting as candidate locations all those places characterised by reasonably flat, low-lying ground with linear sea frontage in excess of 300 m. These criteria were applied to ensure a choice of radar type at every location; only the CODAR Seasonde is able to be deployed on almost any topography. Some 141 sites emerged from this procedure; they are marked on Figure 10.



**Figure 10.** Maps of the South China Sea showing the major shipping lanes, candidate radar sites (blue dots), the associated potential radar coverage (yellow sectors), the discrete points in the sea area at which objective functions can be evaluated (magenta dots), and selected islands of interest (dots, various colours). Figure 10a shows radar coverage to 200 km, Figure 10b to 400 km.

Figure 10a shows the nominal current measurement coverage for oceanographic-class radars deployed at each of these sites, while Figure 10b shows the corresponding information for a military-class radar. These figures reveal that no fully-compliant solution exists, even with the maximal deployment of radars, but they suggest that a solution employing a combination of radar types, at a suitable subset of sites, might achieve an acceptable outcome leaving relatively few areas unsurveyed for this particular mission.

Regarding spatial resolution, while the cross-range dimension of a cell at nominal maximum range exceeds the along-range dimension by up to an order of magnitude, the broad features of oceanographic fields remain distinguishable, and discrete ship echoes can be finely resolved in the Doppler domain, so HFSWR is certainly able to provide the required detail for most objectives. The accuracy of measurements is limited not so much by radar design as by the intrinsic spatial and temporal variability of natural phenomena; the widespread acceptance of HFSWR remote sensing products confirms that the information is of adequate fidelity.

### 3.2. Performance limitations and constraints

It is helpful to be aware of the factors which limit HFSWR performance, as radar and network design can be adapted to minimise the deleterious effects of some of them. First there is the nature of surface wave propagation, which results in increasingly rapid signal decay as one moves beyond about 50 km range, with higher frequencies decaying much more quickly. Second, there is the frequency dependence of the radar signatures of both ship targets and the ocean surface, which have complicated forms that jointly have a strong influence on radar performance. Third, there is the external HF noise from lightning and man-made emissions,

which almost always defines the noise floor against which radar echoes of interest must compete. External noise is highly dependent on time of day. Fourth, there are Doppler-spread echoes from the ionosphere, which have a complex spatial and temporal pattern of occurrence and can mask echoes of interest.

Measures which can be taken to mitigate these factors include antenna design, advanced signal processing, frequency agility and, of special relevance to the present study, siting relative to the locations and velocities of the phenomena under observation. As a simple example, Figure 11 shows the obscuration of ship echoes due to sea clutter, plotted in Doppler space.

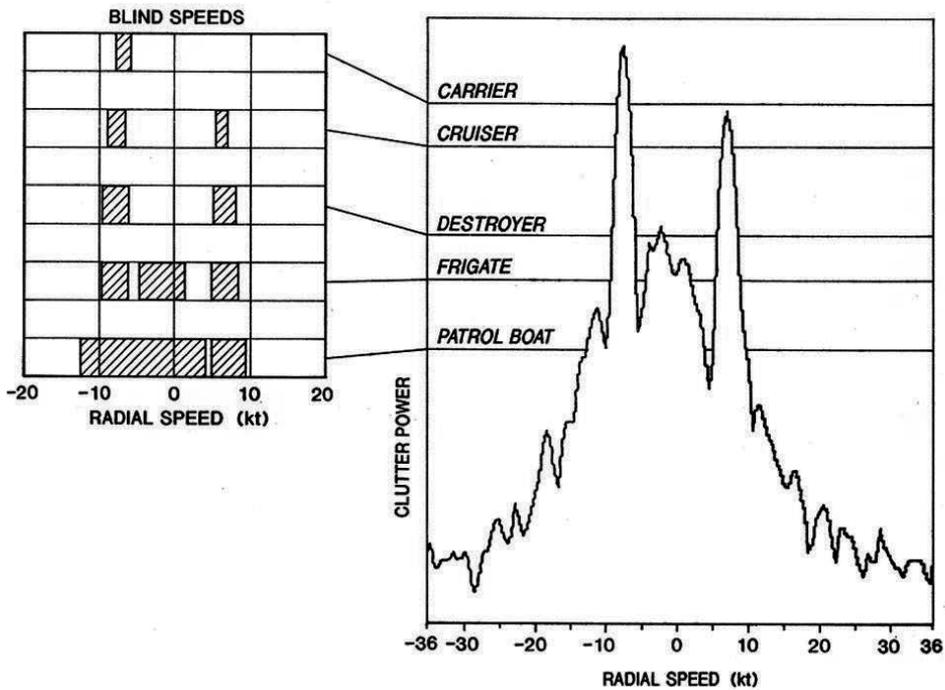


Figure 11. Blind speeds for various ship types, against a specific sea clutter spectrum, for a 32 second integration time

## 4. Radar siting and configuration design as a multi-objective optimisation problem

### 4.1. Elements of the formulation

At the outset we need to identify the data structures, procedures and supporting information that need to be integrated into the problem formulation. These are:

- i. the parameter space  $P$  in which the solutions must lie. A particular solution  $x^{mn} \in X^{mn}$  will have a fixed number  $m$  of transmitting systems, each specified by location, orientation and design, together with a fixed number  $n$  of receiving systems, each similarly described by its location, orientation and design. The ability of receiving system  $p$  to acquire and process signals from transmitting system  $q$  is represented by a coupling matrix  $C^{mn}$ :

$$C_{pq}^{mn} = \begin{cases} 1 & \text{if receiver } p \text{ can process signals from transmitter } q \\ 0 & \text{otherwise} \end{cases}$$

Usually compatibility demands that the systems belong to the same product family. Thus, for example, a Seasonde receiver can process signals from a Seasonde transmitter, but not signals from a WERA transmitter. In the problem under consideration, we do not know a priori what the network membership numbers  $m$  and  $n$  should be; accordingly we define  $P$  as the disjoint union of the  $X^{mn}$ ,

$$P = \bigcup_{n=1}^N \bigcup_{m=1}^M X^{mn}$$

where  $M$  and  $N$  are upper bounds on the numbers of radar transmit and receive sites.

A solution  $x$  thus can be written in the form of a pair of two-dimensional arrays,

$$x \triangleq [\{R_j^{lat}, R_j^{lon}, \Psi_j, \varphi_j\}_{j=1,N}; \{T_k^{lat}, T_k^{lon}, \Psi_k, \varphi_k\}_{k=1,M}]$$

where the dimensions correspond to parameter type (latitude, longitude, radar class, orientation) and parameter index (labelling the set of Tx/Rx sites which make up the configuration)

- ii. the specific coordinates of feasible sites. These constitute a subset of the set of points  $C$  which comprise the coastlines which border the South China Sea or, in the case of radar sites on very small islands, the nominal location at which the installation is most feasible
- iii. the amenity of each location to the installation of a radar, taking into account factors such as accessibility, power supply, field of view and environmental impact
- iv. the range and azimuthal coverage of the individual radars to be used
- v. the wind, wave and current climatology of the waters of the South China Sea
- vi. the recognised shipping lanes
- vii. the types of vessels of interest and their typical speeds and radar cross sections
- viii. the surveillance and remote sensing missions assigned to the radar system and the associated performance thresholds which must be exceeded (at least in a statistical sense)
- ix. algorithms which compute the radar network response for any given combination of ship type, course, speed and environmental conditions

- x. the objective function space  $Y$ , that is, the  $k$ -dimensional space whose coordinates measure the radar performance against the  $k$  tasks assigned to the radar
- xi. a search algorithm which finds the extrema of a scalar function over a specified domain
- xii. a criterion for ranking solutions which achieve extrema in one or more coordinates of  $Y$

It is common practice to formulate optimisation problems in terms of minimising the objective functions rather than maximising them, which is trivially achieved by redefining the coordinates of  $Y$ ; we shall follow this practice.

#### 4.2. Criteria for network optimality

With this palette of ingredients, various radar siting problems with cost and performance constraints can be formulated. Three of the most important are:

- find the solution  $x \equiv x^{opt} \in X$  which maximises performance against a specific task for a given cost
- find the minimum cost solution  $x \equiv x^{opt} \in X$  which exceeds a specified threshold of performance
- given an existing deployment  $x_{old}$  of a number of radars illuminating parts of an area of interest, find the solution  $x_{new}$  such that the augmented network  $x_{aug} \equiv x_{old} \cup x_{new}$  meets some specified performance/cost criterion

but there are many other possibilities. We observe that some of these can be expressed as inverse problems with the threshold vector taking the role of the data vector.

#### 4.3. Multi-objective optimisation via Pareto dominance

The definition of the problem given above is in one sense incomplete – it does not specify the choice of norm for the space  $Y$ . In a single objective optimisation problem, the objective space is usually a subset of the real numbers and a solution  $x_1 \in P$  is better than another solution  $x_2 \in P$  if  $y_1 < y_2$  where  $y_1 = \mu(x_1)$  and  $y_2 = \mu(x_2)$ . In the case of a vector-valued objective function mapping, comparing solutions is more complex and one must endeavour to capture the essential priorities of the problem in the choice of norm. Herein lies the crucial distinction between single objective and multi-objective problems-whereas the former afford simple scalar measures of fitness that can be used to rank individual members of the design space, the latter are characterised by conflicts of interest among the competing objectives as measured by  $\mu_i, i=1,m$ .

There are several ways to deal with this complication. Perhaps the simplest is to create a scalar figure of merit as a weighted sum of the separate objective measures,

- i. minimize  $\mu^{(1)} = \sum_{i=1}^m \alpha_i \mu_i$

Another approach is to convert all but one of the objectives into constraints,

- ii. minimise  $\mu_j$  subject to  $\mu_i \leq z_i \forall i=1, m; i \neq j$

While convenient, these methods shed little light on the nature of the trade-offs made. As there may be subtle, non-quantifiable considerations involved in site selection, such as risks to personnel or to equipment, a better approach is to map the trade-off surface so that the decision maker can execute judgment in making a final selection. To perform this mapping, it is not necessary to run (i) or (ii) above for a large number of parameter selections  $\alpha_i, z_i$  and to inspect the outcomes. Instead, we can use an evolutionary stochastic optimisation algorithm to reveal the Pareto front, as described below.

Pareto optimality is based on the binary relation of dominance. A solution  $x_1 \in X$  is said to be dominated by another solution  $x_2 \in X$ , written  $x_2 < x_1$ , if  $x_2$  is at least as good on all counts (objectives) and better on at least one, that is,

$$\mu_i(x_2) \leq \mu_i(x_1) \forall i=1, m \text{ and } \mu_j(x_2) < \mu_j(x_1) \text{ for some } j.$$

With this relation, the Pareto set of optimal (non-dominated) solutions  $P^*$  will usually have multiple entries, associated with different trade-offs between the objectives. The image  $Y^* \subset Y$  of the Pareto set  $P^* \subset P$  is referred to as the Pareto front and knowledge of its shape greatly assists in choosing the best compromise solution.

#### 4.4. Implementation via genetic algorithms

Classical techniques for finding extrema of functions defined on prescribed domains rely, in most cases, on gradient search methodologies. Such techniques are vulnerable to being trapped on local extrema, rather than the global extremum of main interest. In addition, the convergence may be slow, especially near the extrema, necessitating the invocation of higher-order derivatives. While there are ways to alleviate these weaknesses, they come at considerable cost. An alternative approach, now in widespread use, is to emulate evolutionary mechanisms which we observe in action in the natural world. The best known of these evolutionary optimisation techniques are genetic algorithms.

Genetic algorithms encode the parameter values associated with each candidate solution as a string, usually in binary format. For each parameter, the number of bits provided must be sufficient to encode the full range of possible values associated with that parameter. The string representing a solution is simply the concatenation of the sub-strings corresponding to the individual parameters; by analogy with biology, this string is referred to as a chromosome. Starting with an initial population of candidate solutions (ie, chromosomes) constructed by means of a random number generator, a genetic algorithm iteratively applies three basic steps: (i) rank the members of the current population according to fitness, (ii) select superior members which will be used to breed the next generation, and (iii) apply operators on randomly-selected pairs of these members to mimic the transfer of genetic material to offspring that occurs during biological reproduction, thereby producing a new generation with statistically superior characteristics.

A common mechanism for the transfer of information from one generation to the next is variable length cross-over. For each pair of chromosomes selected to breed together, the start and end indices of a sub-string are selected by a random number generator and the corresponding sub-strings are exchanged. The excisions are not forced to align with the parameter sub-string boundaries. The offspring of this coupling have parts in common with each parent, and in general will represent new solutions. A small fraction of this new set of chromosomes is then subjected to mutation, that is, one or two bits may be flipped to produce a different string, which of course maps onto a different candidate solution. This completes the process of constructing a new generation.

With single objective optimisation, it is a simple matter to rank the members of the resulting population so that selection of candidates for constructing the next generation can proceed. Chromosomes representing the best solutions are carried over unchanged to the next generation, as well as participating in the breeding cycle, while the least fit are discarded. The resulting population is then allowed to breed in its turn, via cross-over and mutation. After passing through a large number of generations, the population tends to converge towards a uniform composition whose members share the most desirable parameter values. Importantly, by virtue of the randomness of the cross-over and mutation operations, candidate solutions from all over the solution domain are potentially represented, and mutation ensures that this property is maintained, so that the population is unlikely to be trapped on a local extremum if a superior solution exists.

With multi-objective optimisation, the key objective is to find the Pareto front, but experience has shown that coverage and convergence can be improved by relying on more than just Pareto dominance for selection. In our approach, each chromosome was tested against its contemporaries and those which were Pareto dominant were automatically selected, while those which had only one or two dominators were also short-listed. In addition, members that performed particularly well against just one objective function were retained. Supplementing these criteria, a scalar figure of merit was defined by taking the product of the individual objective functions; this provided another metric for selection. The total size of the population was maintained at the initial value by allowing each of these different selection mechanisms to contribute a fraction of the membership, with the relative proportions changing with the generation index. We modified the single objective genetic algorithm developed by Anderson (2013) to embody these ideas and hence to compute an estimate of the Pareto front.

#### **4.5. Acceleration techniques**

Genetic algorithms tend to be computationally expensive, so special techniques continue to be developed to accelerate convergence. Some methods which have proven efficacious are:

- eugenics – a recent hybrid scheme which combines the virtues of GA with a very efficient gradient search
- class identifiers – partitioning chromosome space into dissimilar clusters and constraining cross-over to avoid in-breeding, thereby increasing and maintaining diversity

- smart seeds – using intuition, experience and common sense to insert some chromosomes with high potential

#### 4.6. Methods for handling variable solution space dimensionality

One of the challenges of the general network optimisation problem is that, unlike the case in Anderson (2013), the number of radars is itself a variable. Given that the gene length for representing an individual radar is fixed, it follows that the minimum chromosome length will change. This introduces some very fundamental modifications to the elements of the GA algebra, so a number of approaches have been explored:

- loop through dimension index; a straight-forward extension of conventional GA structure
- set the chromosome length to the maximum number of radar sites considered feasible and work within this space; likely to be computationally expensive
- adopt a hierarchical scheme, with fixed length chromosomes containing genes serving as pointers to subspaces of different dimensionality; potentially effective but complicated
- employ variable length chromosomes; this requires a whole new class of genetic operators able to work with strings of different lengths

In the implementation we have used for the South China Sea example, the first of these options has been adopted.

#### 4.7. Constructing the chromosomes

The chromosomes do not need to encode all the detailed information about site properties, radar characteristics, and so on. It is more efficient to use the genes as pointers to data files in which the numerical specifications are stored. In our illustrative example, we allow for four different radar types, so 2 bits are required for that purpose. Our survey of the coastlines of the South China Sea identified 141 candidate sites, so it might seem logical to allocate  $\log_2 141$  bits to represent them. This causes a problem, as not all 8-bit strings correspond to radar sites, and the extra algorithmic structure that would be required to deal with this issue would arise in a section of the code which is run intensively. In the present case it is better to prune the set back to 128 sites, with minimal impact on the outcome, though conceivably another problem could justify increasing the state space to 256 sites. Thus our basic gene has  $2 + \log_2 128 \approx 9$  bits. As it necessary to extract the separate radar-type and radar-site parameters, an efficient ‘gene scissors’ is required, easily implemented in Matlab.

The specific context imposes other constraints that need to be carefully considered. For instance, in the present network design study, we found candidate sites located on the mainlands of Vietnam, Malaysia, Brunei and the Philippines, as well as on several islands some of which are of disputed sovereignty. It may be that network operations embracing radars in all ASEAN nations could be negotiated, but such arrangements are never simple. The situation becomes even more complicated when we contemplate radars on those islands which presently are home to airstrips, ideal for basing array-type HFSWR systems, since islands

meeting that description are owned or occupied by China, Indonesia, Malaysia, the Philippines, Taiwan and Vietnam. In addition, a number of mostly submerged reefs and seamounts in the Spratly Islands bear constructions on which CODAR Seasonde radars could easily be fitted. All these possibilities need to be taken into account when proposing the extent of the solution space in which the set of optimal solutions is to be sought.

## 5. Constructing the objective function space

### 5.1. Objective functions for priority missions

While it is certainly possible to conceive of many useful missions which could be addressed by a network of HFSWR systems, it is generally the case that one focusses on those which have a high level of economic or geo-strategic relevance. As an example, the palette of tasks which one might wish to address could take the form:

- i. maintain surveillance around most, preferably all, of the important islands with the targets of interest being vessels of at least patrol boat size, typically 50 – 80 m in length and long endurance research ships of around 100-120 m length
- ii. provide full ocean current vector information over those parts of the South China Sea which are traversed by large vessels such as tankers and container ships; with emphasis on the major shipping routes
- iii. provide sea state information for the areas in which fishing fleets operate

It is readily seen that the spatial domains over which the performance of these three tasks is of interest are of different dimensionality. As the objective function used to define fitness for a given task involves integration over the corresponding domain, there is a strong link between task domain and computational load. The cases of most concern to the network optimisation problem under consideration are as follows:

Domain dimensionality	Examples
0	islands, shoals, offshore oil platforms,
1	shipping lanes, transects, sovereignty and EEZ boundaries
2	fishing grounds, oil exploration leases, wave energy surveys

For each of the designated tasks, it is necessary to define some criterion that quantifies performance and which can thus be used to govern the search for the Pareto optimum configurations. To illustrate, we shall outline the construction of objective functions – also known as fitness functions or figures of merit – for the first two tasks mentioned above.

## 5.2. Ship detection

Suppose an HF surface wave radar operating at a fixed centre frequency  $f$  is deployed with the goal of detecting ships whose radar cross section (RCS) exceeds some specified threshold. For detection we require that a ship echo exceed the clutter and noise power in the same Doppler bin by some margin  $\varepsilon$ , that is, there exists  $\omega \in [-\Omega, \Omega]$  such that  $s(\omega) > c(\omega) + n(\omega) + \varepsilon$  where  $s(\omega)$ ,  $c(\omega)$  and  $n(\omega)$  are the target, clutter and noise power spectral densities respectively and  $[-\Omega, \Omega]$  is the extent of the Doppler domain. At modest ranges, dependent on the radar type, the clutter power spectral density exceeds that of external noise, but at longer ranges external noise dominates and sets the detection limit. This we need to have a database which provides these distributions. From the description in Section 2 we know that the wind stress and hence the sea state is relatively constant over the South China Sea during each of the two monsoon periods, comprising some 80% of the year, so a reasonable approach is to compute clutter Doppler spectra for just these two sea states. In the context of large ships on the major lanes in the South China Sea, proceeding along known shipping lanes at fairly uniform speeds,  $v \in [v_{min}, v_{max}]$ , the Doppler perceived by a radar from a given ship is a function of a single coordinate, representing the ship's position along its chosen lane, since that determines the viewing geometry. Accordingly, for these targets it makes good sense to define a figure of merit which measures the fraction of the time (equivalently distance along the route under surveillance) for which such ships are detectable. In the case of small or medium size ships near particular islands or facilities, the direction of travel and the speed cannot be assumed, so the figure of merit should reflect the need to maximise detectability against all eventualities. Assuming a maximum speed  $v_{max}$

$$OF_1 = \frac{1}{2\omega_m} \sum_{k=1}^n \int_{-\omega_m}^{\omega_m} H(s(\omega; r_k) - c(\omega; r_k) - n(\omega) - \varepsilon) d\omega \quad (1)$$

where  $H(x)$  is the Heaviside function,

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

and

$$\omega_m = \frac{2v_{max}f}{c}$$

where  $f$  is the radar frequency and  $c$  the speed of light. The  $r_k$  are the coordinates of the discrete islands, offshore oil platforms or other discrete features of interest.

It is a computationally trivial but operationally useful generalisation to apply a priority weighting to the individual islands,

$$OF_2 = \frac{1}{2\omega_m} \left( \frac{1}{\sum_k w_k} \right) \sum_{k=1}^n w_k \int_{-\omega_m}^{\omega_m} H(s(\omega; r_k) - c(\omega; r_k) - n(\omega) - \varepsilon) d\omega \quad (2)$$

which could reflect the distribution of navigation hazards, risk of piracy, cross-Strait traffic density and so on. To evaluate these integrals, we need expressions for  $s(\omega; r)$  and  $c(\omega; r)$ , as well as noise data. The first of these can be written

$$s(\omega; r) = R(\psi_{Rx}) \left( \frac{c^2}{4\pi f^2} \right) G(r_{Rx}, r) \sigma(\omega; \varphi_{scat}, \varphi_{inc}, r) \delta(\omega - \omega_D) G(r, r_{Tx}) T(\psi_{Tx}) P_{Tx} \quad (3)$$

with  $P_{Tx}$  the transmitted power,  $T(\psi_{Tx})$  and  $R(\psi_{Rx})$  denoting the azimuthal gain patterns of the transmit and receive antennas,  $G(r_2, r_1)$  representing the propagation loss factor between positions  $r_1$  and  $r_2$ , and  $\sigma(\varphi_{scat}, \varphi_{inc})$  the bistatic radar cross section for an incident angle  $\varphi_{inc}$  and scattered angle  $\varphi_{scat}$  as defined at  $r$ , and  $\omega_D$  the Doppler shift associated with the target echo,

$$\omega_D = \frac{-f}{c} \times \frac{d}{dt} (|r - r_{Tx}| + |r - r_{Rx}|) \quad (4)$$

For target-specific criteria, the RCS must be calculated using a computational electromagnetics code such as NEC4 or FEKO™.

The corresponding expression for  $c(\omega; r)$  takes the form

$$c(\omega; r) = R(\psi_{Rx}) \left( \frac{c^2}{4\pi f^2} \right) G(r_{Rx}, r) \sigma(\omega; \varphi_{scat}, \varphi_{inc}, r) G(r, r_{Tx}) T(\psi_{Tx}) P_{Tx} A \quad (5)$$

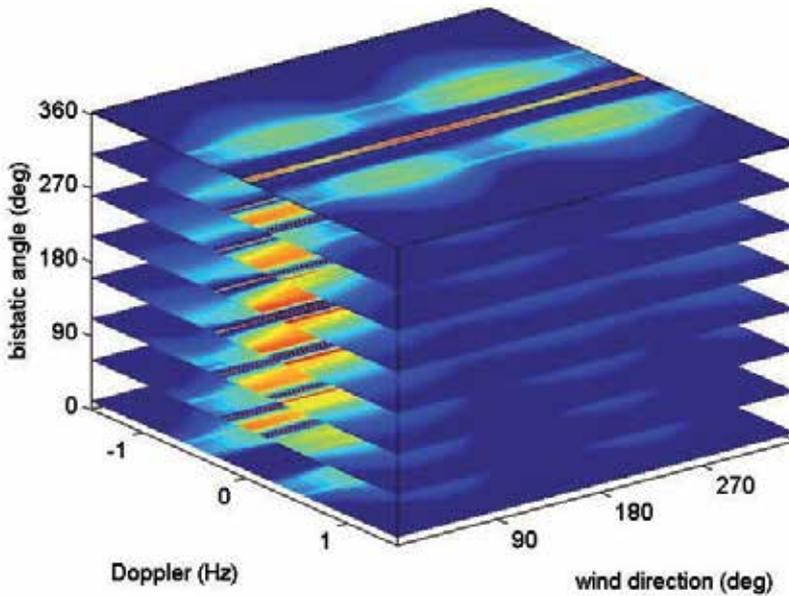
Here  $A$  denotes the area of the resolution cell, whose cross-range dimension increases with range from the receiver. The cell's range extent is determined, in general, by the bandwidth  $B$  of the transmitted waveform and, for a phased array system of aperture  $L_{Rx}$  we can write

$$A \approx \frac{c^2 |r - r_{Rx}|}{2BL_{Rx} \cos \psi_{Rx}} \quad (6)$$

The sea surface scattering coefficient  $\sigma(\omega; \varphi_{scat}, \varphi_{inc}, r)$  has a continuum of spectral content and, being dependent on sea state, will normally vary with position.

$$\begin{aligned} \sigma(\omega; \varphi_{scat}, \varphi_{inc}, r) = & 2^6 \pi k_0^4 \left[ \sum_{m=\pm 1} \int S(m\kappa; r) \delta(\omega - m\sqrt{g\kappa}) \delta(\kappa + k_{inc} - k_{scat}) d\kappa \right. \\ & + \sum_{m_1=\pm 1} \sum_{m_2=\pm 1} \iint \Gamma^2(m_1\kappa_1, m_2\kappa_2) S(m_1\kappa_1; r) S(m_2\kappa_2; r) \\ & \left. \times \delta(m_1\kappa_1 + m_2\kappa_2 + k_{inc} - k_{scat}) \delta(\omega - m_1\sqrt{g\kappa_1} - m_2\sqrt{g\kappa_2}) d\kappa_1 d\kappa_2 \right] \end{aligned} \quad (7)$$

where  $S(\kappa; r)$  is the directional wave spectrum at location  $r$  and  $\Gamma^2(m_1\kappa_1, m_2\kappa_2)$  is a kernel which contains, inter alia, the polarisation dependence, though that does not play a role here. An segment of the resulting database of Doppler spectra, for a given frequency, evaluated for all bistatic angles and wind directions, is shown in Figure 12. The sea parameters were those from Section 2.



**Figure 12.** A small subset of the database of Doppler spectra evaluated for a particular wind speed and radar frequency, but with all combinations of wind direction and bistatic scattering angle

For an operational deployment, one would compute figures of merit averaged over time of day and the seasons, for which we would need wind, wave and current climatologies. If appropriate, a weighting factor could be applied to effect diurnal or seasonal priorities.

For each of these figures of merit, the value lies in the interval  $[0, 1]$ , increasing with the merit of the solution. Two simple options for the function to be minimised are  $(1 - OF)$  and  $OF^{-1}$ .

The figures of merit developed above apply to individual radars but the essence of the problem under consideration is optimisation of a network. The extension to the network case begins

from the observation that, at any given moment, the target will be detected if at least one radar is able to achieve detection. For a set of radars operating in monostatic mode – what has been termed ‘stereoscopic radar’ (Anderson, 1990) – this can be encapsulated in the following expression:

$$OF_3 = \frac{1}{n} \sum_{k=1}^n \left[ 1 - \prod_{j=1}^n \left[ 1 - \max_{\omega \in Z} H(s_j(\omega; r) - c_j(\omega; r) - \varepsilon_j) \right] \right] \quad (8)$$

However, we need also to allow for bistatic detection, which has been shown (Anderson, 1990) to increase the probability of detection by circumventing the possibility of double blind speeds in stereoscopic configurations. This leads to

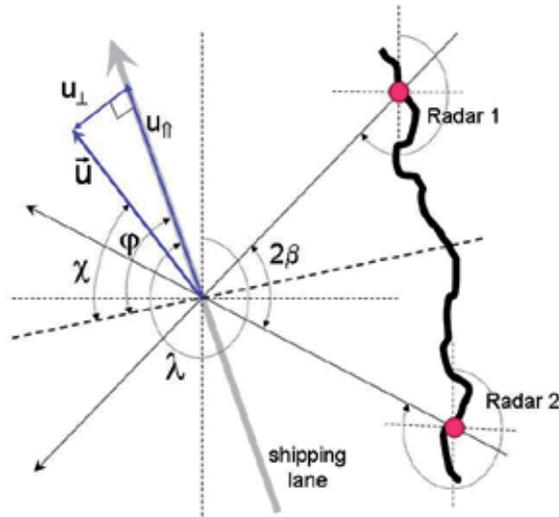
$$OF_4 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left[ 1 - \prod_{i=1}^n \prod_{j=1}^n \left[ 1 - \max_{\omega \in Z} H(s_{ij}(\omega; r) - c_{ij}(\omega; r) - \varepsilon_i) \right] \right] \quad (9)$$

While this formulation seems reasonable as far as detection is concerned, it does not take into account the advantage of detecting a target with two radars simultaneously, from different directions. Not only is the probability of detection increased but detection-to-track association is improved; this is an important consideration in the dense traffic environment of the South China Sea where ships are on average only ~ 10 km apart, not much more than the radar range resolution and less than the azimuthal resolution of the smaller radars. Accordingly, when two radars can view a region, we could take dual detectability into account via a performance enhancement factor which is a function of the angle subtended at the target by the two radars.

### 5.3. Current mapping

In principle, surface current mapping is a relatively simple operation, relying as it does on two very strong peaks in the Doppler spectrum. A fairly rudimentary objective function is the predicted clutter-to-noise ratio, which can be defined for both monostatic and bistatic measurements. Given the length scales of fine structure in the current field in the South China Sea, as shown in Figure 6, some refinements are needed if the function is to serve its purpose effectively.

The most important is consideration of the phenomenon of geometric dilution of precision. The parameters which govern the GDOP for current measurement are (i) the bistatic angle  $2\beta$  subtended by the two radar axes, and (ii) the crossing angle  $\chi$ , that is the angle between the nominal current direction and the bisector of the two radar axes. These are indicated on Fig. 4. The theory of GDOP is widely reported (see for example Chapman et al, 1997, Emery et al, 2004) and will not be repeated here. If, instead of the total current vector, one is interested in the components along and perpendicular to a given direction, a slightly different function emerges; we are then interested in the component shown as  $u_{\perp}$ , so the relevant crossing angle is  $\varphi$  and the error associated with GDOP must be computed using this angle.



**Figure 13.** Geometry of bistatic illumination, with current vector and shipping lane

Let the radial component of current velocity derived from a measurement by radar 1 be designated  $u_1 + \varepsilon_1$ , where  $\varepsilon_1$  represents random measurement error, and similarly, that of radar 2 by  $u_2 + \varepsilon_2$ . The estimates of velocity parallel to and normal to the bisector axis are then given by

$$u_p = \frac{(u_1 + u_2 + \varepsilon_1 + \varepsilon_2)}{2 \cos \beta} \quad (10)$$

$$u_n = \frac{(u_1 - u_2 + \varepsilon_1 - \varepsilon_2)}{2 \sin \beta} \quad (11)$$

We transform this vector measurement into the coordinate system defined by the lane axis and its normal by means of a rotation:

$$\begin{bmatrix} u_{\parallel} \\ u_{\perp} \end{bmatrix} = \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix} \begin{bmatrix} u_p \\ u_n \end{bmatrix} \quad (12)$$

Solving for  $u_{\perp}$ ,

$$u_{\perp} = \left( \frac{\cos \varphi}{2 \sin \beta} - \frac{\sin \varphi}{2 \cos \beta} \right) u_1 - \left( \frac{\cos \varphi}{2 \sin \beta} + \frac{\sin \varphi}{2 \cos \beta} \right) u_2 \pm \left[ \left( \frac{-\sin \varphi}{2 \cos \beta} + \frac{\cos \varphi}{2 \sin \beta} \right) \varepsilon_1 - \left( \frac{\cos \varphi}{2 \sin \beta} + \frac{\sin \varphi}{2 \cos \beta} \right) \varepsilon_2 \right] \quad (13)$$

Assuming  $\varepsilon_1$  and  $\varepsilon_2$  are independent and identically distributed, the rms error  $\varepsilon$  is found by squaring and averaging the error term,

$$\varepsilon = \frac{\left[ \cos^2(\varphi + \beta) + \cos^2(\varphi - \beta) \right]^{1/2}}{\sin \beta} \varepsilon_1 \quad (14)$$

The GDOP is defined as the ratio of the rms error  $\varepsilon$  to the error  $\varepsilon_1$  associated with an individual radar. Numerical evaluation (Anderson, 2013) shows that the geometry has a major bearing on the accuracy of HFSWR current estimates, more than doubling the errors once the radars depart from orthogonal viewing geometry by more than  $50^\circ$ .

#### 5.4. Visibility and topographic constraints

The figures of merit and associated objective functions developed in the preceding sections have made one assumption which demands explicit representation – the spatial integrations have made no allowance for blocking of the signal path from radar to patch of interest by an intervening land mass, either an island or part of the mainland. As it happens, HF surface waves can propagate across land, though with much greater attenuation than across sea, and there is an unusual effect (the Millington effect) through which a considerable fraction of signal strength is restored once the signal reaches the sea beyond the intervening land mass. Nevertheless, unless it cannot be avoided, it is better not to entertain the possibility of exploiting signals which have propagated across one or more islands. We can formalise this constraint on single site acceptability as follows.

Suppose there are  $K$  landmasses  $\{D_j\}_{j=1,K}$  with coastlines  $\{\partial D_j\}_{j=1,K}$  adjoining a sea or ocean of which a region  $W$  is to be monitored. From the  $k$ -th coastline,  $\partial D_k$ , construct  $\partial D_k^+$  as follows:

$$\partial \tilde{D}_k^+ = \left\{ r \in \partial D_k \left| \begin{array}{l} \{\alpha r + (1-\alpha)r'\} \cap D_m = \{0\} \\ \forall r' \in W, \\ \forall m = 1, K; \alpha \in [0, 1] \end{array} \right. \right\} \quad (15)$$

Then  $\partial D_k^+ \subset \partial D_k$  is the subset of the coastline of the  $k$ -th landmass which has an unobstructed view of the region  $W$ . If a radar is to be placed on  $D_k$ , then it must lie on  $\partial D_k^+$ .

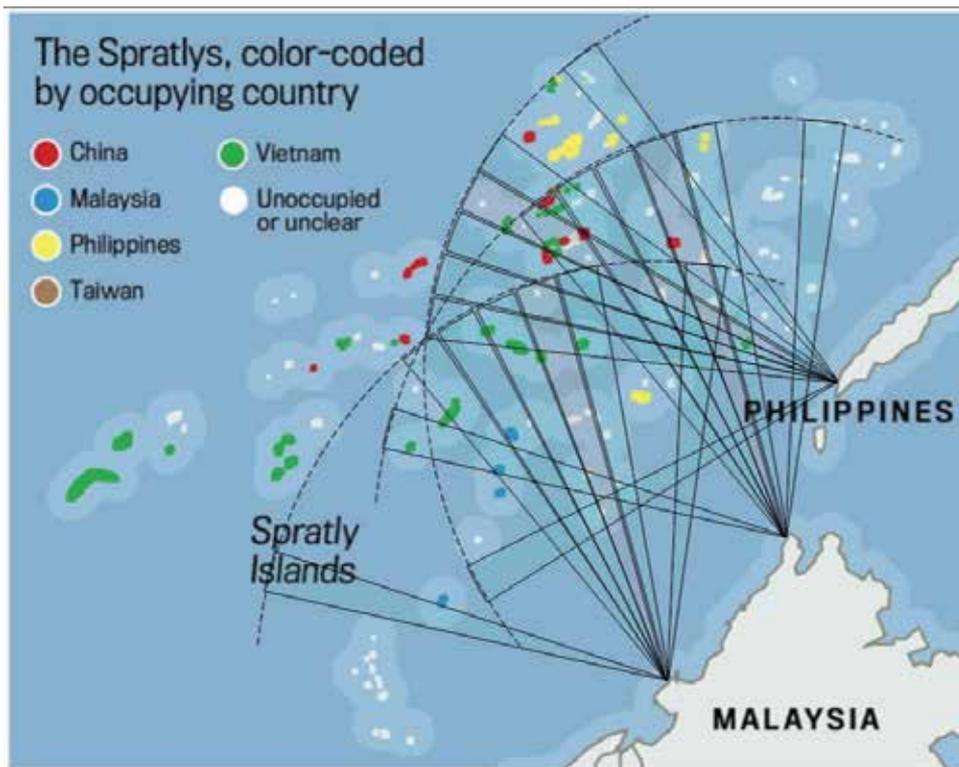
In the present study, where each radar can hope to survey at most a part of the area of concern, it is more appropriate to assign the coverage arc at each candidate radar site and measure the effectiveness of that coverage according to the metrics defined earlier.

Accordingly we have chosen to define the coverage arcs by the requirement that they exclude any directions which meet regions  $W$  for which the site does not belong to  $\partial D_k^+$ .

For the present illustrative purposes we shall not impose other site-specific constraints such as conditions on local topography or coastline orientation and curvature, though these too could be added if desired.

## 6. Radar networks for the South China Sea

The tools and procedures described in the preceding sections are of general applicability, but the success of the network optimisation relies on making best possible use of site-specific environmental information, not only oceanographic and meteorological but also the levels and patterns of HF noise. Often this information is unavailable or incomplete, but in most cases one can find climatological data which will serve adequately. We have used the South China Sea mainly as a context to illustrate the ways in which the geophysical information can be exploited, as well as the general issues that could drive network deployment. Needless to say, the South China Sea is of particular interest, so a number of network design experiments have been undertaken. They confirm the applicability of the genetic algorithm methodology to this context, reinforcing the conclusions of Anderson (2013) for the Strait of Malacca.



**Figure 14.** Part of an optimum solution, focussing on missions over the Spratly islands, showing the receive beam structure for a military-class phased array system.

An example of a candidate solution to a particular network optimisation problem is presented in Figure 14. It shows the coverage of the Spratly Islands afforded by a network of three military-class radars, addressing a combination of mission types. In this example, the optimality criterion of maximising performance for a fixed cost, discussed in Section 4.2, was used to constrain the solution space. A more comprehensive set of solutions, invoking a variety of optimality constraints, is in preparation (Anderson, 2014).

Note again that performance is a statistical quantity, with somewhat greater or perhaps far lesser coverage achieved on any given occasion.

## 7. Conclusion

The optimum deployment of a network of HFSWR systems is a highly complex task with many factors to be considered, especially when the radars are expected to perform multiple roles. Failure to treat the design problem with appropriate care could seriously degrade performance in one or more radar missions.

In this paper we have described a practical technique for HFSWR network design, based on a genetic algorithm adapted to multi-objective optimisation, and illustrated the method by placing it in the context of designing a multi-radar configuration system for remote sensing and surveillance of the South China Sea. The treatment pays particular attention to the construction of chromosomes and objective functions and extends previous measures of performance to allow for bistatic radar operations. In addition, we emphasise the importance of exploiting a priori knowledge about the regional geography, meteorology and oceanography in the design procedures.

A key advantage of the Pareto dominance formulation developed by Anderson (2013) and incorporated here is that it efficiently identifies those solutions which are superior to their fellows according to every criterion tested, and hence greatly reduces the range of design options which need to be considered. The designer is presented with a range of candidate optimal solutions which can be assessed according to additional considerations which may not be meaningfully quantifiable. When we consider domains such as the South China Sea, where complex geopolitical issues may arise, the virtues of this approach are self-evident.

## Author details

S. J. Anderson

Address all correspondence to: [stuart.anderson@dsto.defence.gov.au](mailto:stuart.anderson@dsto.defence.gov.au)

Defence Science and Technology Organisation, Edinburgh SA, Australia

## References

- [1] Anderson, S.J. (1990). Stereoscopic and Bistatic Skywave Radars : Assessment of Capabilities and Limitations, Proceedings of Radarcon-90, Adelaide, Australia, 305-313.
- [2] Anderson, S.J., Edwards, P.J., Marrone, P., and Abramovich, Y.I. (2003). Investigations with SECAR - a bistatic HF surface wave radar, Proceedings of IEEE International Conference on Radar, RADAR 2003, Adelaide.
- [3] Anderson, S.J. (2013). Optimizing HF Radar Siting for Surveillance and Remote Sensing in the Strait of Malacca, IEEE Transactions on Geoscience and Remote Sensing, Vol.51, No.3, 1805-1816.
- [4] Anderson, S.J., Darces, M., Helier, M., and Payet, N. (2013 ). Accelerated convergence of genetic algorithms for application to real-time inverse problems, Proceedings of the 4<sup>th</sup> Inverse Problems, Design and Optimization Symposium, IPDO-2013, Albi, France, 149-152.
- [5] Anderson, S.J. (2014). HF radar network optimisation : Case studies for the South China Sea. In preparation.
- [6] Barrick D E. (1998). EEZ Surveillance-The Compact HF Radar Alternative, EEZ Technology, Edition 3. London: ICG Publishing Ltd., 125-129. See also the CODAR website <http://www.codar.com/> and references therein for detailed information about the SeaSonde family of HFSWR systems.
- [7] Chapman, R.D., Shay, L.K., Graber, H.C., Edson, J.B., Karachintsev, A., Trump, C.L., and Ross, D.B. (1997). On the accuracy of HF radar surface current measurements : Inter-comparisons with ship-base sensors, Journal of Geophysical Research, vol. 102, C8, 118737-18748.
- [8] Chu, P.C., Qi, Y., Chen, Y., Shi, P., and Mao, Q. (2003). Validation of Wavewatch-III Using the TOPEX/POSEIDON Data, Proceedings of SPIE Conference on Remote Sensing of the Ocean and Sea Ice, Barcelona, Spain.
- [9] Chu, P.C., Qi, Y., Chen, Y., and Mao, Q. (2004). South China Sea Wind-Wave Characteristics. Part I: Validation of Wavewatch-III Using the TOPEX/POSEIDON Data, Journal of Oceanic Technology, Vol.21, No.11, 1718-1733.
- [10] Emery, B.M., Washburn, L., and Harlan, J.A. (2004). Evaluating radial current measurements from CODAR high-frequency radars with moored current meters, Journal of Atmospheric and Oceanic Technology, vol. 21, no. 8, pp. 1259-1271.
- [11] Helzel, T., Kniephoff, M., and Pettersen, L. (2010). Oceanography radar system WERA : features, accuracy, reliability and limitations, Turkish Journal of Electrical Engineering and Computer Science, vol. 18, no. 3, 389-397. See also the Helzel Messtechnik website <http://www.helzel.com/> and references therein for detailed information about the WERA family of HFSWR systems.

- [12] International Crisis Group. (2012a). Stirring up the South China Sea (I), International Crisis Group Asia Report No. 223.
- [13] International Crisis Group. (2012b). Stirring up the South China Sea (II): Regional Responses, International Crisis Group Asia Report No. 229.
- [14] Lipa, B., Isaacson, J., Nyden, B., and Barrick, D. (2012), Tsunami arrival detection with high frequency radar, *Remote Sensing*, Vol.4, No.11, 1448-1461.
- [15] Marghany, M. (2009). Volterra - Lax-wendroff algorithm for modelling sea surface flow pattern from Jason-1 satellite altimeter data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) Volume 5730 LNCS*, 1-18 .
- [16] Marghany, M. (2011). Developing robust model for retrieving sea surface current from RADARSAT-1 SAR satellite data, *International Journal of the Physical Sciences*. Vol. 6 (29), 6630-6637.
- [17] Marghany, M. (2012). Three-Dimensional Coastal Front Visualization from RADARSAT-1 SAR Satellite Data. In Murgante B. et al. (eds.): *Lecture Notes in Computer Science (ICCSA 2012)*, Part III, LNCS 7335, 447-456.
- [18] Mirzaei, A., Tangang, F., Liew, J., Mustapha, M.A., Husain, M.L., and Akhir, F.A. (2013). Wave climate simulation for southern region of the South China Sea, *Ocean Dynamics*, Vol.63, 961-977.
- [19] Ninh, P.V., Quynh, D.N., Lanh, V.V., and Lien, T.V. (2000). Geostrophic and drift current in the South China Sea, Area IV: Vietnamese waters, *Proceedings of the SEA-DEC Seminar on Fishery Resources in the South China Sea, Area IV: Vietnamese Waters*, 365-373.
- [20] O'Rourke, R. (2013). Maritime Territorial and Exclusive Economic Zone (EEZ) Disputes Involving China: Issues for Congress, *Congressional Research Service Report for Congress 7-5700 R42784*.
- [21] Ponsford, A.M. (2012). Persistent surveillance of the 200 nautical mile Exclusive Economic Zone using Raytheon's land-based high frequency surface wave radar, *Raytheon Technology Today*, 2012, issue 2, 25-27.
- [22] Wang, J., Li, M., Liu, Y., Zhang, H., Zou, W., and Cheng, L. (2014). Safety assessment of shipping routes in the South China Sea based on the fuzzy analytic hierarchy process, *Safety Science*, Vol.62, 46-57.

---

# **A Three-dimensional of Coastline Deformation using the Sorting Reliability Algorithm of ENVISAT Interferometric Synthetic Aperture Radar**

---

Maged Marghany

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/58571>

---

## **1. Introduction**

At present, because of sea-level rise, and climate change (such as an increase in storm surges, hurricanes, etc.) there is a worldwide increase in coastal erosion, usually apparent in the progressive retreat of backshore cliffs, dunes, spits, and the concomitant landward displacement of the shoreline [10, 13, 21]. In this regard, coastal erosion requires standard procedures for monitoring, modelling, and mapping [12, 22].

In the last two decades, scientists have developed a powerful technique to measure the millimeter-scale of the Earth 's surface deformation by comparing complex synthetic aperture radar (SAR) data that were acquired a few days or a few years apart. This technique is known as interferometric synthetic aperture radar (InSAR). Accurate Earth 's surface deformation or digital elevation maps can be produced by implementing the single look complex synthetic aperture radar (SAR) images that are received by two or more separate antennas. The phase image is produced by multiplying the complex SAR image by the coregistered complex conjugate pixels of the other SAR data. Image coregistration, InSAR interferometric phase estimation (or noise filtering) and interferometric phase unwrapping [3, 4, 5, 6, 32, 37, 41] are three key processing procedures of InSAR.

Therefore, advance application of remote sensing technology such as interferometric synthetic aperture radar (InSAR) for coastal geomorphology study is in a preliminary stage. InSAR has shown that it can provide DEMs with 1-10 cm accuracy, which can be improved to millimeter level by Differential synthetic aperture radar Interferometry (DInSAR) [1, 3, 5, 8, 22].

Therefore, InSAR techniques are proven to provide precisely digital elevation model map (DEMs). In this context, mapping purposes, geomorphological studies based on aspect maps and slopes are acquired highly accurate DEMs products. In this regard, sub-centimeter target displacements can be detected using DInSAR along the sensor-target direction [5, 10, 17]. In comparison with other conventional techniques such as levelling, and GPS, DInSAR provides valuable information about target displacement, over large areas, at a relatively low cost [2, 6]. In fact, conventional methods are time consuming and they are required some control points [1, 9]. DInSAR has many other applications as well, such as monitoring geophysical natural hazards, for instance, earthquakes, volcanoes and landslides, also in engineering, in particular recording of subsidence and structural stability. Overtime-spans of days to years, InSAR can detect the centimeter-scale of deformation changes [4]. Furthermore, the precision of DEMs from the InSAR technique is quite high, compared to conventional remote sensing methods. In many countries, the 90 m SRTM data, or the 30 m. DEM data from ASTER are the main sources of DEMs [15, 17].

### 1.1. Problems in interferometric techniques

It is well known that the performance of the interferometric phase estimation suffers seriously from poor image coregistration. Interferogram filtering algorithms such as adaptive contoured window, pivoting mean filtering, pivoting median filtering, and adaptive phase noise filtering [15] are the main methods for the conventional InSAR interferometric phase estimation. These filtering algorithms, nevertheless cannot retrieve the accurate terrain interferometric phases from a poor interferogram because of coregistration errors and high decorrelation. Indeed, the interferometric phases are random in nature with their variances being inversely proportional to the correlation coefficients between the corresponding pixel pairs of the two coregistered SAR data. Therefore, the terrain interferometric phases should be estimated statistically [34-41].

It is well known that the accuracy of the image coregistration is a serious issue for accuracy interferometric phase unwrapping. Therefore, the conventional phases unwrapping algorithms, such as the branch-cut method, region-growing method and the least-square one, which are required accurate image coregistration to be about 1/10 to 1/100 resolution cell size (i.e., SAR pixel). Nevertheless, the obtained InSAR phase unwrapping will be extremely noises if InSAR data coregistration is decorrelated. In this regard, the accurate InSAR data coregistration is a hard task with low coherence data which is extremely change dynamically [41]. In this context, Hai L and Renbiao [41] stated that the interferometric phase estimation method based on subspace projection can provide an accurate estimation of the terrain interferometric Phase (interferogram) even if the coregistration error reaches one pixel.

Incidentally, the phase difference of the two SAR data is processed to acquire height and deformation of the Earth's surface. Therefore, scientists have agreed that the accurate results of InSAR are required standard criteria for data acquisitions. According to Hanssen [3], short temporal baseline, appropriate spatial baseline, good weather conditions and ascending and descending SAR data are regular criteria to reduce decorrelation and noise and produce a reliable DEM.

Nonetheless, alternative SAR datasets must obtain at high latitudes or in zones of run-down coverage [6, 13]. The baseline decorrelation and temporal decorrelation, nevertheless, make InSAR measurements unrealistic [8, 9, 10, 11]. Incidentally, Gens [12] stated the length of the baseline designates the sensitivity to height changes and sum of baseline decorrelation. Additionally, Gens [12] reported the time difference for two data acquisitions is a second source of decoration. Indeed, the time differences while compare data sets with a similar baseline length acquired one and 35 days apart suggests only the temporal component of the decorrelation. Therefore, the loss of coherence in the same repeat cycle in data acquisition is most likely because of baseline decorrelation. According to Roa et al. [7], uncertainties could arise in DEM because of limitation InSAR repeat passes. In addition, the interaction of the radar signal with troposphere can also induce decorrelation. This is explained in several studies [3, 8, 15].

Commonly, the propagation of the waves through the atmosphere can be a source of error exist in most interferogram productions. When the SAR signal propagated through a vacuum it should theoretically be subjected to some decent accuracy of timing and cause phase delay [3, 21]. A constant phase difference between the two images caused by the horizontally homogeneous atmosphere was over the length scale of an interferogram and vertically over that of the topography. The atmosphere, however, is laterally heterogeneous on length scales both larger and smaller than typical deformation signals [9, 23]. In other cases the atmospheric phase delay, however, is caused by vertical inhomogeneity at low altitudes and this may result in fringes appearing to correspond with the topography [24]. Under this circumstance, this spurious signal can appear entirely isolated from the surface features of the image, since the phase difference is measured other points in the interferogram, would not contribute to the signal [3, 17]. This can reduce seriously the low signal-to-noise ratio (SNR) which restricted to perform phase unwrapping. Accordingly, the phases of weak signals are not reliable. According to Yang et al., [11], the correlation map can be used to measure the intensity of the noise in some sense. It may be overrated because of an inadequate number of samples allied with a small window [9]. Weights are initiated to the correlation coefficients according to the amplitudes of the complex signals to estimate accurate reliability [11, 18].

According to Pepe [40] DinSAR has recently been applied with success to investigate the temporal evolution of the detected deformation phenomena through the generation of displacement time-series. In this case, the analysis is based on the computation of deformation time-series via the inversion of a properly chosen set of interferograms, produced from a sequence of temporally-separated SAR acquisitions relevant to the investigated area. In this context, two main categories of advanced DInSAR techniques for deformation time-series generation have been proposed in literature, often referred to as Persistent Scatterers (PS); and Small Baseline (SB) techniques, respectively. The PS algorithms select all the interferometric data pairs with reference to a single common master image, without any constraint on the temporal and spatial separation (baseline) among the orbits. In this case, the analysis is carried out on the full resolution spatial scale, and is focused on the pixels

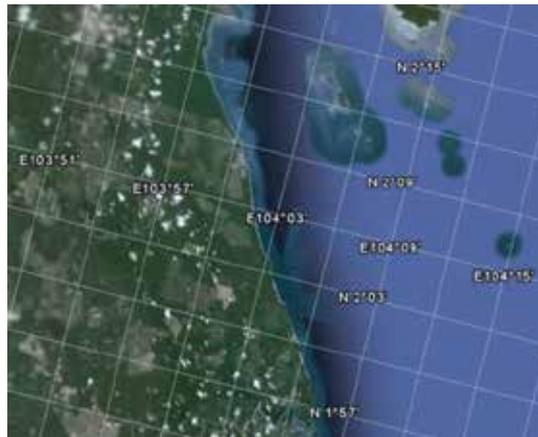
containing a single dominant scatterer thus ensuring very limited temporal and spatial decorrelation phenomena.

## 1.2. Hypotheses and objectives

In this paper, we address the question of utilization three-dimensional phase unwrapping algorithm to estimate rate changes of shoreline deformation. In fact, there are several factors could impact the accuracy of DEMs was derived from phase unwrapping [21]. These factors are involved radar shadow, layover, multi-path effects and image misregistration, and finally the signal-to-noise ratio (SNR) [11]. This demonstrated with ENVISAT ASAR data. The main contribution of this study is to implement three three-dimensional phase unwrapping algorithm with InSAR technique. Three hypotheses examined are: (i) three-dimensional phase unwrapping algorithm can be used as filtering technique to reduce noise in phase unwrapping; (ii) 3-D shoreline reconstruction can be produced using satisfactory phase unwrapping by involving the three-dimensional phase unwrapping algorithm; and (iii) high accuracy of deformation rate can be estimated by using the new technique.

## 2. Study area

The study area is located along the coast of Johor in the southern eastern part of Peninsular Malaysia. The area is approximately 20 km of Johor (Figure 1), located in the South China Sea, between  $1^{\circ} 57' N$  to  $2^{\circ} 15' N$  and  $103^{\circ} 51' E$  to  $104^{\circ} 15' E$ . This coastline is exhibited a variety of geomorphologies that includes sandy beaches, rocky headlands which is broken by small river mouths. In addition, the coastline has hilly terrain with steep slopes and deep narrow valleys. Further, the coastline is bordered with varying widths of alluvial plains. Further, sand materials make up the entire of the eastern Johor shoreline. Consistent with Marghany [42], this area lies in an equatorial region dominated by two seasonal monsoons and two inter monsoon periods. The southwest monsoon lasts from May to September while the northeast monsoon lasts from November to March. The inter monsoon periods involve April which is between end of northeast monsoon and beginning of the southwest monsoon periods. Further, the second inter monsoon period is October which is between the end of the southwest monsoon and the beginning of northeast monsoon periods. The monsoon winds affect the direction and magnitude of the waves. Further, Marghany [14, 42] stated that strong waves are prevalent during the northeast monsoon when the prevailing wave direction is from the north (November to March), while during the southwest monsoon (May to September), the wave directions are propagated from the south. According to Marghany [14] the maximum wave height during the northeast monsoon season is 4 m. The minimum wave height is found during the southwest monsoon which is less than 1 m [18].



**Figure 1.** Location of study area

### 3. InSAR data processing

Two methods are involved to perform InSAR from ENVISAT ASAR SAR data (i) conventional InSAR procedures; and (ii) three-dimensional phase unwrapping algorithm i.e. three-dimensional sorting reliabilities algorithm (3D-SRA) [25, 42]

#### 3.1. Conventional InSAR method

According, Zebker et al., [4], two complex SAR data are required to achieve InSAR procedures. These complex SAR data have a real (cosine) and an imaginary (sine) components. Then real and an imaginary components are combined as vectors to acquire information of phase and intensity of the SAR signal [10]. In Marghany [26], the surface displacement can estimate using the acquisition times of two SAR data  $S_1$  and  $S_2$ . The component of surface displacement, thus, in the radar-look direction (Figure 2), contributes to the further interferometric phase ( $\phi$ ) as [9]

$$\phi = \frac{4\pi(\Delta R)}{\lambda} = \frac{4\pi(B_h \sin \theta - B_v \cos \theta)}{\lambda} \quad (1)$$

where  $\Delta R$  is the slant range difference from satellite to target respectively at different time,  $\theta$  is the look angle (19.2-26.7°), (Table 1) and  $\lambda$  is the ENVISAT ASAR wavelength Single Look Complex (SLC) which is about 5.6 cm for C-band. Therefore,  $B_h$ ,  $B_v$  are horizontal and vertical baseline components[19].

According to Lee [9], for the surface displacement measurement, the zero-baseline InSAR configuration is the ideal as  $\Delta R=0$ , so that

$$\phi = \phi_d = \frac{4\pi}{\lambda} \zeta \tag{2}$$

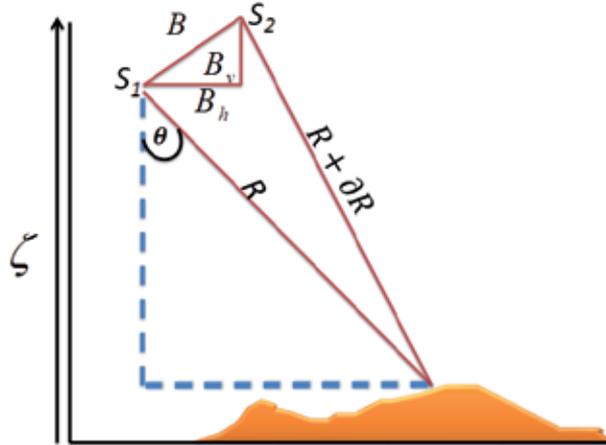


Figure 2. InSAR Geometry.

Parameters	Values
Radar Wavelength (cm)	5.6
Ground Resolution	30
Swath Width (km)	105
Incident Angle (°)	19.2-26.7
Polarization	HH/VV

Table 1. ENVISAT ASAR characteristics were used in this study [19, 42]

In actual fact, zero-baseline, repeat-pass InSAR configuration is hardly achievable for either spaceborne or airborne SAR. Therefore, a method to remove the topographic phase as well as the system geometric phase in a non-zero baseline interferogram is needed. If the interferometric phase from the InSAR geometry and topography can strip off from the interferogram, the remnant phase would be the phase from block surface movement, providing the surface maintains high coherence [5]. Then, the phase difference  $\Delta\phi$  between the two ENVISAT ASAR data positions and the pixel of target of terrain point is given by

$$\Delta\zeta = \frac{\lambda R \sin\theta}{4\pi B} \Delta\phi \tag{3}$$

Equation 3 is a function of normal base line  $B$ , and the range  $R$ . In addition, equation 3 can provide information about the heights and phase differences estimations. In fact, the estimated height of each pixel of ENVISAT ASAR data is an important task to generate a raster form of the DEM.

### 3.2. DEM reconstruction using a three-dimensional Sorting Reliabilities Algorithm (3D-SRA)

Marghany [42] has adapted the algorithm introduced by Hussein et al., [25] for three-dimensional phase unwrapping the algorithm that is called a three-dimensional sorting reliability algorithm (3D-SRA). The quality of each edge of phase unwrapping is a function of the connection of two voxels in 3-D Cartesian axis e.g.,  $x, y, z$ . Starting, to carry out the unwrapping path from high quality voxels to bad quality voxels [25]. In addition, following a discreet path, the 3D-SRA algorithm unwraps the phase volume, which is significant to determine the 3-D volume change rate of shoreline. In this regard, the voxels connects the highest reliable edges that are unwrapped first with border surfaces. Consistent with Hussien et al., [26], the reliability value of an edge that connects a border voxel with another voxel in the phase volume is set to zero.

Let  $E_x, E_y,$  and  $N$  are the horizontal, vertical, and normal second differences, respectively which are given by

$$E_x(i, j, k) = \gamma[\phi(i-1, j, k) - \phi(i, j, k)] - \gamma[\phi(i, j, k) - \phi(i+1, j, k)], \quad (4)$$

$$E_y(i, j, k) = \gamma[\phi(i, j-1, k) - \phi(i, j, k)] - \gamma[\phi(i, j, k) - \phi(i, j+1, k)], \quad (5)$$

$$N(i, j, k) = \gamma[\phi(i, j, k-1) - \phi(i, j, k)] - \gamma[\phi(i, j, k) - \phi(i, j, k+1)], \quad (6)$$

where  $i, j, k$  are the neighbors' indices of the voxel in  $3 \times 3 \times 3$  cube, and  $\gamma$  defines a wrapping operator that wraps all values of its argument in the range  $[-\pi, \pi]$ . This can be done by adding or subtracting an integer number of  $2\pi$  rad to its argument [26].  $\gamma$  can calculate from the wrapped-phase gradients in the  $x, y,$  and  $z$  directions as follows [25,26]

$$\gamma(i, j, k) = \frac{\partial \phi_{i,j,k}^x}{[\phi_{i+1,j,k} - \phi_{i,j,k}]}, \quad (7)$$

$$\gamma(i, j, k) = \frac{\partial \phi_{i,j,k}^y}{[\phi_{i,j+1,k} - \phi_{i,j,k}]}, \quad (8)$$

$$\gamma(i, j, k) = \frac{\partial \phi_{i,j,k}^z}{[\phi_{i,j,k+1} - \phi_{i,j,k}]}, \quad (9)$$

Equations 7 to 9 are represented 3-D array of the wrapped-phase gradients  $\partial \phi^x$ ,  $\partial \phi^y$ ,  $\partial \phi^z$  and each has the same dimensions as the wrapped-phase volume. In addition, the maximum phase gradient measures the magnitude of the largest phase gradient that is, partial derivative or wrapped the phase difference in a  $v^*v^*v$  volumes [25]. Using the sum of equations 4 to 6, the second difference quality  $Q$  can be obtained [26]

$$Q_{i,j,k} = \sqrt{E_x^2(i, j, k) + E_y^2(i, j, k) + N^2(i, j, k)} \quad (10)$$

The unwrapping path is performed based on equation 10 where entirely the edges are stored in a 3D array and sorted with their edge quality values [26]. Further, unwrapping a voxel or a group of voxels concerning another group may require the addition or subtraction of multiples of  $2\pi$  [25]. In addition, the badness of each voxel in a  $v^*v^*v$  volume is determined by

$$b = \max \left\{ \begin{array}{l} \max \left\{ \left| \partial \phi_{i,j,k}^x \right| \right\} \\ \max \left\{ \left| \partial \phi_{i,j,k}^y \right| \right\} \\ \max \left\{ \left| \partial \phi_{i,j,k}^z \right| \right\} \end{array} \right\} \quad (11)$$

$b$  is the badness of voxel  $(m, n, l)$ , which is given by

$$b = v^{-3} * [\sqrt{\sum (\overline{\partial \phi_{i,j,k}^x} + \overline{\partial \phi_{i,j,k}^x})^2} + \sqrt{\sum (\overline{\partial \phi_{i,j,k}^y} + \overline{\partial \phi_{i,j,k}^y})^2} + \sqrt{\sum (\overline{\partial \phi_{i,j,k}^z} + \overline{\partial \phi_{i,j,k}^z})^2}] \quad (12)$$

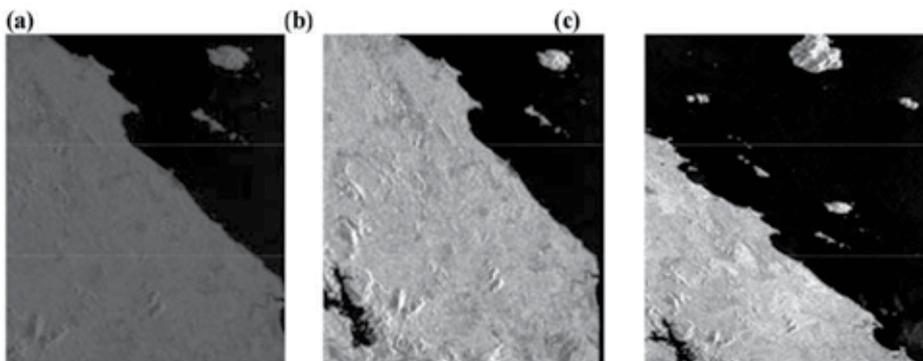
where  $\overline{\partial \phi_{i,j,k}^{x,y,z}}$  are the mean value of wrapped phase gradients in  $x, y, z$  directions, respectively through a  $v^*v^*v$  volume that are centered at that voxel [25].

### 3.3. Ground survey

At the rear of Marghany [14], the GPS survey used to: (i) to record the exact geographical position of the shoreline; (ii) to determine the cross-sections of shoreline slopes; (iii) to corroborate the reliability of InSAR data co-registered; and finally, (iv) to create a reference network for future surveys. The geometric location of the GPS survey was obtained by using the new satellite geodetic network, IGM95. After a careful analysis of the places and to identify the reference vertexes, we thickened the network around such vertexes to perform the measurements for the cross sections (transect perpendicular to the coastline). The GPS data collected within 50 sample points scattered along 10000 m coastline. The interval distance of 2000 m between each sample location is considered. In every sample location, Rec-Alta (Recording Electronic Tachometer) was used to acquire the coastline elevation profile. The ground truth data were acquired on January 25 2011 during satellite passes.

## 4. Results and discussion

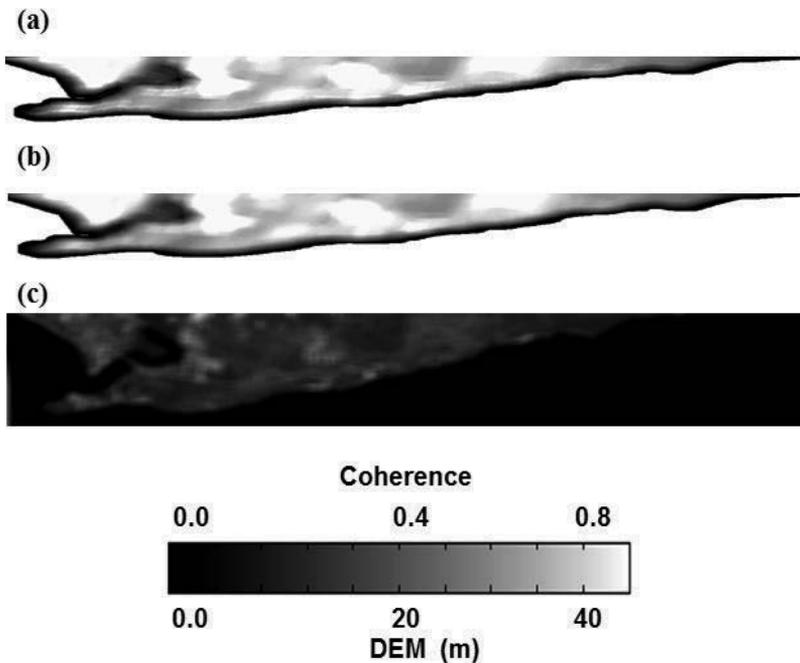
In this study, ENVISAT ASAR satellite data were used to investigate shoreline change using three-dimensional sorting reliability algorithm. InSAR methods are implemented on ENVISAT ASAR data sets of 5 March 2003 (SLC-1), 28 December 2010 (SLC-2) and January 25, 2011, (SLC-3) of Wide Swath Mode (WSM) (Figure 3). They are acquired from ascending (Track: 226, orbit: 5290), descending (Track:490, Orbit: 6055) and descending (Track 420,Orbit 4655), respectively. These data are in C band with VV polarization.



**Figure 3.** ENVISAT ASAR data used in this study (a) 5 March 2003; (b) 28 April 2003; and (c) January 25 2011

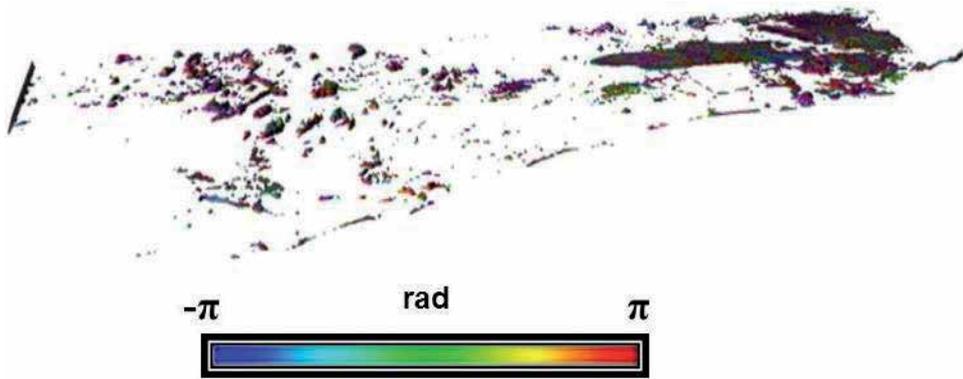
These data are C-band and had the lower signal-to-noise ratio owing to their VV polarization with a wavelength range of 3.7 to 7.5 cm and a frequency of 5.331 GHz. ASAR can achieve a spatial resolution generally around 30 m. The ASAR is intended for applications which require the spatial resolution of spatial resolution of 150 m. This means that it is not effective at imaging areas in depth, unlike stripmap SAR. The azimuth resolution is 4 m, and range resolution ranges between 8 m.

Figure 4 presents the reference DEM which has been generated from topographical 1:50000, and in situ measurements, respectively while Figure 4c shows the ENVISAT coherence data. It is clear that the maximum elevation of 50 m found inland. The maximum elevation of 10 m is shown along the coastline. The high coherence rates are existed in urban zone and along infrastructures with 0.9 while low coherence of 0.2 is found in a vegetation zone along the coastline. Since three ASAR data acquired in the wet northeast monsoon period, there has been an impact of wet sand on a radar signal penetration which causing weak penetration of radar signal because of the dielectric. Indeed, the total topographic decorrelation effects along the radar-facing slopes are dominant and highlighted as lowest coherence value of 0.2. According to Marghany [27] the micro-scale movement of the sand particles driven by the coastal hydrodynamic, and wind speed of 12 m/s during the northeast monsoon period [30] could change the distribution of scatters resulting in rapid temporal decorrelation which has contributed to lowest coherence along coastline. This result agreed with Marghany [27, 31].



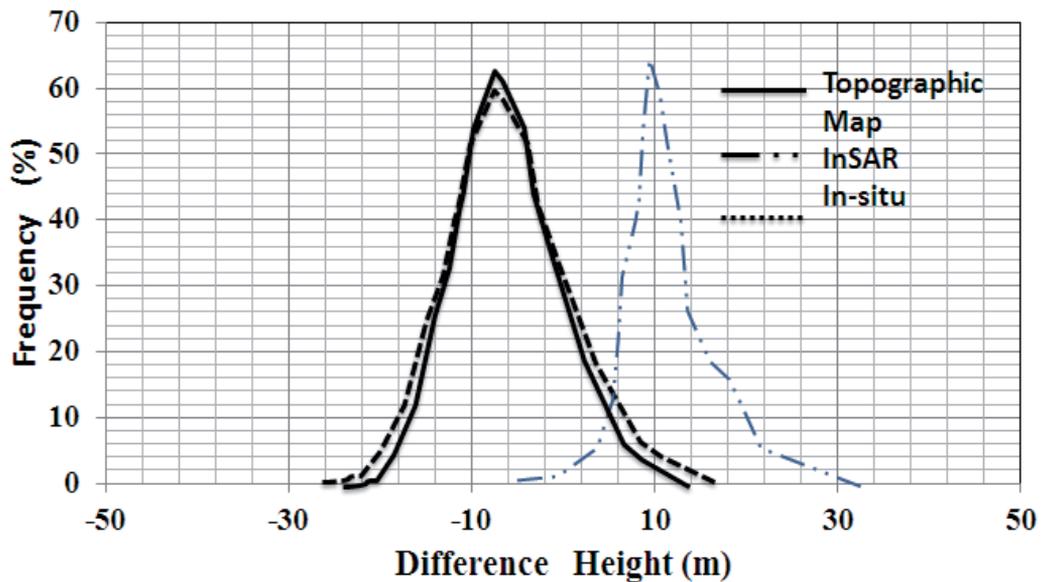
**Figure 4.** DEM generated from (a) topographic map; (b) in situ measurements; and (c) coherence ENVISAT ASAR

Clearly, there is huge differences between InSAR DEM and DEM generated from in situ measurements with 9 m differences while in situ measurements are concurred with DEM produced from topographic map within 1.3 m differences (Figure 5). This is because the impact of decorrelation. This result confirms the work done by Marghany [27, 29]. The overall scene is highly incoherent (Figure 4) which extremely effected the accuracy of InSAR DEM. This decorrelation caused a poor detection of InSAR DEM which induce large ambiguities because of poor coherence and scattering phenomenology. Indeed, the signatures of the interferometric



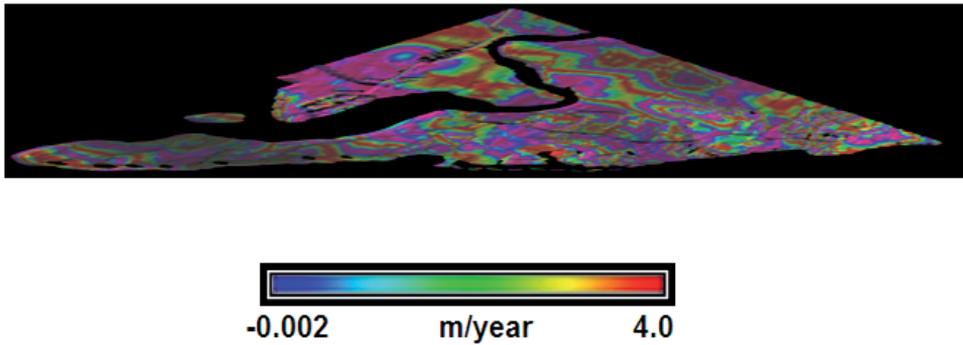
**Figure 6.** Interferogram generated from InSAR data

coherence and phase of vegetation are extremely impacted by temporal decorrelation in ASAR c-band data.



**Figure 5.** Height differences between DEM generated from InSAR, topographic map and in situ measurements.

According to Marghany [27], the ground ambiguity and ideal assumption that volume-only coherence can be acquired in at least one polarization. This assumption may fail when vegetation is thick, dense, or the penetration of an electromagnetic wave is weak. This agrees with the studies of Lee [9]; Marghany [14]; and Marghany [28, 29]. This can be seen clearly in InSAR interferogram has been made from ASAR data (Figure 6).



**Figure 7.** Fringe interferometry generated by a the three-dimensional sorting reliabilities algorithm

Figure 7 shows the interferogram created using three-dimensional sorting reliabilities algorithm 3-DSR. The full color cycle represents a phase cycle, covering the range between  $-\pi$  to  $\pi$ . In this context, the phase difference given module  $2\pi$ ; is color encoded in the fringes. Seemingly, the color bands change in the reverse order, indicating that the center has a critical coastline erosion of -3.5 m/year. This shift corresponds to 2 of coastal deformation over the distance of 10000 m.

Table 2 shows the statistical comparison between the simulated DEM from the InSAR, real ground measurements and with using three-dimensional sorting reliabilities algorithm. This table represents the bias (averages mean the standard error, 90 and 95% confidence intervals, respectively). Evidently, the InSAR using three-dimensional sorting reliabilities algorithm has a bias of -0.08 m, lower than ground measurements and the InSAR method. Therefore, a three-dimensional sorting reliabilities algorithm has a standard error of mean of  $\pm 0.05$  m, lower than ground measurements and the InSAR method. Overall performances of InSAR method using a three-dimensional sorting reliabilities algorithm is better than the conventional InSAR technique which is validated by a lower range of error ( $0.06 \pm 0.32$  m) with 90% confidence intervals.

Statistical parameters	InSAR techniques			
	InSAR		Three-dimensional sorting reliabilities algorithm	
	Lower	Upper	Lower	Upper
Bias	3.3		-0.08	
Standard error of the mean	2.2		0.05	
90% (90% confidence interval)	2.3	3.8	0.06	0.24
	1.3	3.2	0.04	0.32

**Table 2.** Statistical Comparison between InSAR and InSAR-three-dimensional sorting reliabilities algorithm

This study confirms the work have been done by Hussein et al., [25, 26]. The three-dimensional sorting reliabilities algorithm provides an excellent 3-D phase unwrapping, which leads to high quality of 3-D coastline reconstruction. This could be contributed to quality map. Indeed, the 3-DSR algorithm is guided by including the maximum gradient quality maps. Therefore, quality maps guide the unwrapping path through a noisy region so that the interferogram patterns are in completing cycle as compared to InSAR interferogram. Moreover, as stated by Hussien et al., [26], changing the cube size has a great effect to reduce the effect of noise and improve the calculated quality. Consistent with Marghany [42], the 3-DSR algorithm follows discrete unwrapping paths to ensure the processing of the highest quality regions even if they are separated from each other. In other words, within the 3-DSR algorithm, the edges are stored in an array which is based on the terms of the terms of their edge quality values. This means it relies on edge quality to guide the unwrapping path that produces accurate 3-D coastline reconstruction as compared to InSAR techniques. Generally, the 3-DSR algorithm can be an excellent solution for decorrelation problems in such tropical area as Malaysia.

## 5. Conclusions

The paper has demonstrated InSAR phase unwrapping using the three-dimensional sorting reliabilities algorithm (3-DSR). In addition, three-dimensional (3-D) coastline deformation from interferometry synthetic aperture radar (InSAR) is estimated. Further, three-dimensional sorting reliabilities algorithm (3D-SRA) is implemented with phase unwrapping technique. Consequently, the 3D-SRA is used to eliminate the phase decorrelation impact from the interferograms. The study shows that InSAR produces discontinues interferogram pattern because of the high decorrelation. On the contrary, the three-dimensional sorting reliabilities algorithm generated 3-D coastline deformation with a bias of 0.08 m, lower than ground measurements and the InSAR method. Therefore, the three-dimensional sorting reliabilities algorithm has a standard error of mean of  $\pm 0.05$  m, lower than ground measurements and the InSAR method. Consequently, the 3D-SRA is used to eliminate the phase decorrelation impact from the interferograms. The study also shows the performance of InSAR method using the 3D-SRA is better than InSAR procedure which is validated by a lower range of error ( $0.06 \pm 0.32$  m) with 90% confidence intervals. In conclusion, the 3-DSR algorithm can be used to solve the problem of decorrelation and produced accurate 3-D coastline deformation using ENVISAT ASAR data.

## Author details

Maged Marghany\*

Address all correspondence to: [maged@utm.my](mailto:maged@utm.my), [magedupm@hotmail.com](mailto:magedupm@hotmail.com)

Institute of Geospatial Science and Technology (INSTeG), University of Technology, Skudai, Johor Bahru, Malaysia

## References

- [1] Massonnet, D. and Feigl K. L.: Radar interferometry and its application to changes in the earth's surface. *Rev. Geophys.* 36 (1998) 441--500
- [2] Burgmann R. Rosen P.A. and Fielding E.J.: Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation. *Ann. Rev. of Earth and Plan. Sci.* 28: (2000)169--209
- [3] Hanssen R.F.: Radar Interferometry: Data Interpretation and Error Analysis, Kluwer Academic, Dordrecht, Boston, (2001)
- [4] Zebker H.A., P.A. Rosen, and Hensley S.: Atmospheric effects in interferometric synthetic aperture radar surface deformation and topographic maps. *J. Geophys. Res.* 102 (1997) 7547--7563
- [5] Askne J., M. Santoro, G. Smith, and Fransson J. E. S.: Multitemporal repeat-pass SAR interferometry of boreal forests," *IEEE Trans. Geosci. Remote Sens.* 41 (2003) 1540--1550
- [6] Nizalapur V., R. Madugundu, and Shekhar Jha C.: Coherence-based land cover classification in forested areas of Chattisgarh, Central India, using environmental satellite-advanced synthetic aperture radar data. *J. Appl. Rem. Sens.* 5 (2011) 059501-1--059501-6
- [7] Rao K. S., H. K. Al Jassar, S. Phalke, Y. S. Rao, J. P. Muller, and. Li Z.: A study on the applicability of repeat pass SAR interferometry for generating DEMs over several Indian test sites," *Int. J. Remote Sens.* 27 (2006) 595--616
- [8] Rao K. S. and Al Jassar H.K.: Error analysis in the digital elevation model of Kuwait desert derived from repeat pass synthetic aperture radar interferometry. *J. Appl. Remote Sens.* 4(2010)1--24
- [9] Lee H.: Interferometric Synthetic Aperture Radar Coherence Imagery for Land Surface Change Detection. Ph.D theses, University of London (2001)
- [10] Luo X., F.Huang, and Liu G.: Extraction co-seismic Deformation of Bam earthquake with Differential SAR Interferometry. *J. New Zea. Inst. of Surv.* 296 (2006) 20--23
- [11] Yang J. T.Xiong and Peng Y.: A fuzzy Approach to Filtering Interferometric SAR Data. *Int. J. of Remote Sens.* 28 (2007) 1375--1382
- [12] Gens R. The influence of input parameters on SAR interferometric processing and its implication on the calibration of SAR interferometric data", *Int. J. Remote Sens.* 2 (2000) 11767--1771
- [13] Anile A.M., B. Falcidieno, G. Gallo, M. Spagnuolo, Spinello S.: Modeling uncertain data with fuzzy B-splines", *Fuzzy Sets and Syst.* 113 (2000) 397--410

- [14] Marghany M.: Simulation of 3-D Coastal Spit Geomorphology Using Differential Synthetic Aperture Interferometry (DInSAR). In I. Padron.,(ed.) Recent Interferometry Applications in Topography and Astronomy. Croatia: InTech-Open Access Publisher, (2012) 83–94
- [15] Spangnolini U.: 2-D phase unwrapping and instantaneous frequency estimation. *IEEE Trans.Geo.Remot. Sensing* 33:(1995) 579--589
- [16] Davidson G. W. and Bamler R. Multiresolution phase unwrapping for SAR interferometry *IEEE Trans. Geosci. Remote Sensing*, 37 (1999) 163--174
- [17] Marghany M., Z. Sabu and Hashim M.: Mapping coastal geomorphology changes using synthetic aperture radar data". *Int. J. Phys. Sci.* 5,(2010)1890–1896
- [18] Marghany M.:Three-dimensional visualisation of coastal geomorphology using fuzzy B-spline of dinsar technique. *Int. J. of the Phys. Sci.* 6 (2011) 6967--6971
- [19] ENVISAT, "ENVISAT application [online] Available from <http://www.esa.int> [Accessed 2 February 2013].
- [20] Zebker H.A.,Werner C.L., P.A. Rosen and Hensley S.: Accuracy of Topographic Maps Derived from ERS-1 Interferometric Radar. *IEEE Geosci. Remote Sens.*2. (1994) 823--836
- [21] Baselice F., G. Ferraioli, and Pascazio V.: DEM reconstruction in layover areas from SAR and Auxiliary Input Data *IEEE Geosci. Rem. Sensing Letters*,6: (2009) 253--257
- [22] Ferraiuolo G., V. Pascazio, Schirinzi G. Maximum a Posteriori Estimation of Height Profiles in InSAR Imaging. *IEEE Geosci. Rem. Sensing Letters* (2004) 66--70
- [23] Ferraiuolo G., F. Meglio, Pascazio V., and Schirinzi G.: DEM reconstruction accuracy in multichannel SAR interferometry. *IEEE Trans. On Geosci. and Rem.* 47 (2009) 191--201
- [24] Ferretti A.,C. Prati, Rocca F.: Multibaseline phase unwrapping for InSAR topography estimation", *Il Nuov Cimento*, 24 (2001) 159--176
- [25] Hussein S.A., Gdeist M., Burton D., Lalor M.: Fast three-dimensional phase unwrapping algorithm based on sorting by reliability following a non-continuous path, Optical Measurements System for Industrial Inspection IV conference, Eds Osten, Gorecki and Novak, Proc. SPIE, Vol. 5856-Part 1, (2005) 32--40
- [26] Hussein S.A., Munther A., Gdeisat, David R.,Burton, Michael J., Lalor, F.Lilley, and Moore C.: Fast and robust three-dimensional best bath phase unwrapping algorithm. *Appl. Opt.*, 46: (2007) 6623--6635.
- [27] Marghany M.: DInSAR technique for three-dimensional coastal spit simulation from radarsat-1 fine mode data. *Acta Geophy.* 61:(2013) 478--493

- [28] Marghany M.: Three dimensional Coastal geomorphology deformation modelling using differential synthetic aperture interferometry. *Verlag Der Zeitschrift fur Naturforschung*. 67a: (2012) 419--420.
- [29] Marghany M.: DEM reconstruction of coastal geomorphology from DInSAR. In Murgante B. et al. (eds): *Lecture Notes in Computer Science (ICCSA 2012), Part III, LNCS 7335*: (2012) 435--446
- [30] Marghany M.: Intermonsoon water mass characteristics along coastal waters off Kuala Terengganu, Malaysia. *Int.J.of Phys. Sci* 7:(2012) 1294--1299
- [31] Marghany M.: Modelling shoreline rate of changes using holographic interferometry. *Int.J.of Phys. Sci.* 6: (2011) 7694--7698
- [32] Wei X and Cumming I. A Region-Growing Algorithm for InSAR Phase Unwrapping. *IEEE Trans. On GRS*, 1999, 37(1): 124-134
- [33] M. Costantini. A Novel Phase Unwrapping Method based on Network Programming. *IEEE Trans. On GRS*, 1998, 36(3): 813-831.
- [34] Goldstein R M, Zebker H A, and Werner C L. Satellite radar interferometry : two-dimensional phase unwrapping. *Radio Sci.*, 1988,23(4):713-720.
- [35] Ireneusz B, Stewart M P, Kampes P M, Zbigniew P, and Peter L. A Modification to the Goldstein Radar Interferogram Filter. *IEEE Trans. On GRS*, 2003, 41(9):2114-2118.
- [36] Nan W, Da-Zheng F, and Junxia L. A locally adaptive filter of interferometric phase images. *IEEE GRS Letters*, 2006, 3(1):73-77.
- [37] Qifeng Yu, Xia Yang, Sihua Fu, Xiaolin Liu, and Xiangyi Sun. An Adaptive Contoured Window Filter for Interferometric Synthetic Aperture Radar. *IEEE GRS Letters*, 2007, 4(1):23-26.
- [38] Marghany M. Simulation of 3-D Coastal Spit Geomorphology Using Differential Synthetic Aperture Interferometry (DInSAR). In Padron I. (ed.) "Recent Interferometry Applications in Topography and Astronomy". InTech-Open Access Publisher, University Campus STeP Ri, Croatia. (2012),83-94.
- [39] Marghany M. Simulation of three dimensional of coastal erosion using differential interferometry synthetic aperture radar. *Global nest Journal* (2014), 16(1): 80-86.
- [40] Pepe A. Advanced Multitemporal Phase Unwrapping Techniques for DInSAR Analyses. In Padron I. (ed.) "Recent Interferometry Applications in Topography and Astronomy". InTech-Open Access Publisher, University Campus STeP Ri, Croatia. (2012),57-82.
- [41] Hai L and Renbiao W. Robust Interferometric Phase Estimation in InSAR via Joint Subspace Projection. Hai Li and Renbiao Wu (2012) Robust Interferometric Phase Estimation in InSAR via Joint Subspace Projection. In Padron I. (ed.) "Recent Interfer-

ometry Applications in Topography and Astronomy". InTech-Open Access Publisher, University Campus STeP Ri, Croatia. (2012), 111-132.

- [42] Marghany. Three Dimensional Coastline Deformation from Insar Envisat Satellite Data. In Beniamino Murgante, Sanjay Misra, Maurizio Carlini, Carmelo M. Torre, Hong-Quang Nguyen, David Taniar, Bernady O. Apduhan, and Osvaldo Gervasi Computational Science and Its Applications – ICCSA 2013, 2013, (7972) 599-610.



---

# Nuclear, Geophysics and Telecommunication

---



---

# On the Modification of Nuclear Chronometry in Astrophysics and in Geophysics

---

V.S. Olkhovsky

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/57436>

---

## 1. Introduction

The  $\alpha$ -decays of nuclei-chronometers from the only *ground* states are usually taken into account by the practical applications of the standard methods of nuclear chronometry for the stellar and terrestrial processes. The nuclear cosmic nucleo-synthesis includes also the analysis of formation of the initial long-living isotopes during *s*- and *r*-processes. In these processes, inside stars and supernova not only the ground states but also all possible excited states of synthesized nuclei are being formed as a result of nucleon radiation captures ( $n, \gamma$ ) and ( $p, \gamma$ ) (see, for instance, [1-3]).

And from the Geiger and Nutall  $\alpha$ -decay law it follows directly that the lifetime  $\tau_{\text{exc}}$  of the  $\alpha$ -decaying nucleus does very strongly depend on the  $\alpha$ -particle kinetic energy. In many cases the lifetime  $\tau_{\text{exc}}$  is diminished by several orders with the increasing of the  $\alpha$ -particle energy by 1-2 MeV! But up to now no systematic experimental study of the excited radioactive nuclei relative to the  $\alpha$ -decays on their excitation energy had been undertaken in the analysis of lifetimes because of the much more rapid (within  $10^{-13}$ - $10^{-9}$  sec) and so strong  $\gamma$ -decays of the excited nuclei. Previously it had usually assumed that there is no practical reason to take into account the much more slow and so weak processes of the  $\alpha$ -decays from the excited states. But if there are chains of subsequent emissions and quasi-resonance absorptions of  $\gamma$ -quanta by the  $\alpha$ -radioactive nuclei inside stellar masses, then the influence of the excited  $\alpha$ -decaying nuclei can be much stronger.

Of course, there are the emissions-absorptions chains of excited nuclei  $\gamma$ -decays and also  $\beta$ -decays of  $\beta$ -radioactive chronometers in stars. It had been firstly supposed in [4,5] that inside large masses of stellar substance a part of radioactive nuclei could be supported in the excited states during a long time due to the chains of subsequent emissions and quasi-resonance absorptions of  $\gamma$ -quanta by nuclei-chronometers if the energy losses, caused by recoils during

emitting and absorbing, are compensated by the nucleus thermal kinetic energy. Further in [6, 7] this supposition it had been continued for modification of the nuclear chronometry in stars and it had been expanded for planets. The formation of the excited states of the  $\alpha$ -radioactive nuclei in the terrestrial surface layers and in the meteorites is also present under the influence of the weak but constant cosmic radiation. And it has been supposed that the cosmic radiation can also accelerate the  $\alpha$ -decays.

As to the  $\beta$ -radioactive chronometers, we shall say later (at the end of the section **Nuclear chronometry in astrophysics**) that sometimes the lifetimes of them become near  $10^9$  lesser in a bare (without electrons) atom than in the usual atom with filled electronic shell. And inside the stars these nuclei partially or totally are being deprived their electronic shells.

## 2. To the standard nuclear chronometry

Here we shall deal with  $\alpha$ -radioactive nuclei-chronometers. In the standard nucleo-chronometry techniques (see, for instance, [1]) one uses the abundances  $P$  and  $D$  of parent and daughter nuclei. They are connected with the decay function  $L(t-t_0)=\exp(-\Gamma_0(t-t_0)/\hbar)$  ( $\Gamma_0$  being the  $\alpha$ -decay width of the ground state,  $t_0$  being an initial time) and the surviving function  $W(t)=1-L(t)$ , defined and utilized in [4,5], by the following evident relations:

$$L(t-t_0)=P(t)/P(t_0), W(t-t_0)=[P(t)-P(t_0)]/P(t_0) \quad (1)$$

and

$$P(t) + D(t) = P(t_0) + D(t_0), \quad (2)$$

or

$$D(t) - D(t_0) = P(t_0)W(t-t_0) = P(t)[\exp(\Gamma_0(t-t_0)/\hbar) - 1], \quad (3)$$

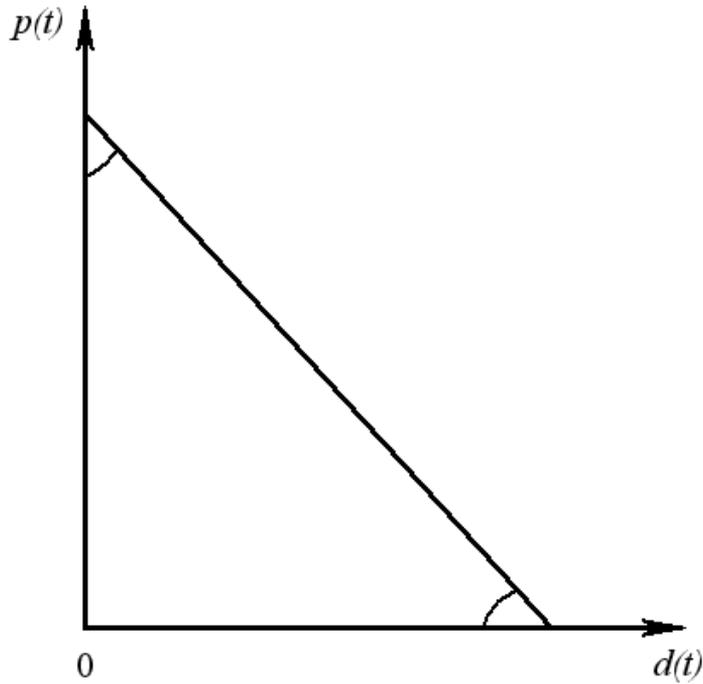
or

$$D(t) - D(t_0) = P(t_0) - P(t) = P(t_0)[1 - \exp(\Gamma_0(t-t_0)/\hbar) - 1]. \quad (4)$$

Usually equality (3) is divided by the abundance of *another* stable isotope  $D_x$  which does not obtain any contribution from the decay of the parent nuclei (i.e.  $D_x$  does not depend from time). As a result, (3) acquires the following form

$$p(t)[\exp(\Gamma_0(t-t_0)/\hbar) - 1] - d(t) + d(t_0) = 0 \quad (5)$$

where  $p=P/D_x$  and  $d=D/D_x$ . Measuring  $p=P/D_x$  and  $d=D/D_x$  in different samples (or in different separate parts of the same sample), we obtain the plot (2b) at the plane  $p=P/D_x$ ,  $d=D/D_x$  which has the form of a straight line, the slope of which with respect to axis  $p$  does simply permit to define age  $t-t_0$  (see, for instance, [1] and Fig. 1 here).



**Figure 1.** The typical plot (4).

For the description of the  $\alpha$ -decay evolution inside stars and under the cosmic radiation on the earth surface one usually uses the Krylov-Fock theorem [8] (see also its application during the generalization in [4,5-7,9]):

$$L(t-t_0) = \langle |\Psi(t-t_0)| \Psi(t_0) \rangle^2 = |f(t-t_0)|^2 / |f(t_0)|^2 \tag{6}$$

where

$$f(t-t_0) = \langle \Psi(t-t_0) | \Psi(t_0) \rangle = \int_0^\infty |G(E)|^2 \exp[-iE(t-t_0)/\hbar] dE \tag{7}$$

is the characteristic function for the energy ( $E$ ) distribution in a decaying state with the weight amplitude  $G(E)$ ,  $t_0$  is the chosen initial time moment and  $\Psi(t_0)$  is normalized by the condition  $|\Psi(t_0)| = 1$ .

### 3. A short previous history of my investigations on the nuclear chronometry

We use in time analysis of nuclear processes elaborated on time as a quantum observable, canonically conjugate to energy, the general expression for the duration  $\langle \tau_{fi} \rangle$  like [10,11]

$$\langle \tau_{fi} \rangle = \langle t_f \rangle_f - \langle t_i \rangle_i \equiv \int_{-\infty}^{\infty} t w_f(z_f, t) dt - t_{in} \tag{8}$$

where  $w_f(z_f, t) = j(z_f, t) / \int_{-\infty}^{\infty} j(z_f, t) dt$ ,  $j_f(z_f, t) = \text{Re}(\psi_{\beta}(-i\hbar/2\mu_f \partial/\partial z_{\beta})\psi_{\rho})_f$  is the  $z_f$  the component of the

probability flux; the brackets  $(\dots)_f$  define the integration over all coordinates, entered in the wave functions  $\psi_f$ , besides  $z_f$ ;  $z_f$  is the axis, directed along the mean velocity  $\langle \vec{v} \rangle_f$  of the relative motion of the decay pair (for instance,  $\alpha$ -particle and the daughter nucleus) with the reduced mass  $\mu_v$  and their wave function of the internal motion  $|v\rangle$  with energy  $e_v$ ;  $\vec{v}_v = \hbar \vec{k}_v / \mu_v$ ,  $\epsilon_v = \hbar^2 k_v^2 / 2\mu_v \equiv E - e_v$ ;  $t_{in}$  is defined as  $\langle t_i \rangle_i$  where  $i$  is the initial channel of forming the decaying nucleus or simply as an initial time moment  $t_{in}$

The wave packet of the relative motion of the decay pair in one-dimensional radial asymptotic limit is described as

$$\Psi_f(r_f, t) = r_f^{-1} \int_0^{\infty} dE g(E) F_{if}(E) \exp[ikr_f - iEt / \hbar]. \tag{9}$$

where

$$F_{if}(E) = \frac{C_{if}}{E - E_r + i\Gamma/2}, \tag{10}$$

with  $C_{if}$  being a constant or smooth function of final-particle kinetic energy  $E$  in the region  $(E_r - \Gamma/2, E_r + \Gamma/2)$ ,  $E_r$  and  $\Gamma$  are the resonance energy and width, respectively. For  $z_f \geq R_{\beta}$ , where  $R_{\beta}$  is the radius of interaction in the final channel, and under condition  $\Gamma \ll \Delta E \ll E_r$  it is possible to re-write (9) in the following simplified form

$$\Psi_f(R_f, t) = A \int_0^{\infty} dE \frac{\exp[-iEt / \hbar]}{E - E_r + i\Gamma/2}, \tag{11}$$

where  $A$  is a constant. For  $\Gamma = \text{constant}$  we obtain

$$\Psi_{\beta}(R_{\beta}, t) = \begin{cases} B \exp[-iE_r t / \hbar - (\Gamma / 2\hbar)t], & \text{for } t > 0 \\ 0, & \text{for } t < 0 \end{cases} \quad (12)$$

(shifting, for very small  $\Gamma$ , the lower limit of the integration in (11) from 0 till  $-\infty$  and utilizing the residue theorem). Here  $B$  is constant and more precisely here will be  $t-t_{in}$  instead of  $t$ .

Of course,  $w_f$  for wave function (12) is equal

$$L(t) = (\Gamma / \hbar) \exp(-\Gamma t / \hbar). \quad (13)$$

The function  $L(t)$ , generally speaking, defined by (6) and concretely by (13), sometimes are named by the *decay function* (see, for instance, [4-10]). The more detailed study (see, for instance, [12-14]) shows that even for an isolated resonance the exponential form of the decay law is approximately performed in the time interval, limited by below by the quantity  $t_1 \sim t_0 \Gamma / E_r$  and from above by the quantity  $t_2 \sim t_0 \ln(\Gamma / E_r)$ , where  $t_0 \sim \hbar / \Gamma$ , and the accuracy of describing by the exponential function is the better, the lesser is the relation  $\Gamma / E_r$ , [14]. The functions (6) and (13) describe the essence of the Krylov-Fock theorem which was derived by another way (see the section before) and states that the decay law of the meta-stable state is totally defined by the energy spectrum of the initial state [8].

The author (V.S.O.) had been extended the contents of the Krylov-Fock theorem, considering the formation of the  $\alpha$ -radioactive chronometers in stars and also the emissions, the successive absorptions etc of  $\gamma$ -quanta by the nuclei-chronometers in [4] on the base of his earlier article [9].

#### 4. Nuclear chronometry in astrophysics

This section is based on [4-7] and in principle on the preceding author articles, later generalized in [10,11] and described here in the preceding section.

The decay rate per a time unit is

$$\rho(t-t_0) = d[1 - L(t-t_0)] / d(t-t_0). \quad (14)$$

When the decay is going on by several channels (for example, by  $\alpha$ - and  $\gamma$ -decays from the first excited state of the parent nucleus) we have

$$\rho_{1i}(t-t_0) = (\Gamma_i^1 / \Gamma_1) d[1 - L_1(t-t_0)] / d(t-t_0) = (\Gamma_i^1 / \hbar) \exp(-\Gamma_1(t-t_0) / \hbar) \quad (15)$$

instead of (14).

The decay of an ensemble of radioactive nuclei can go on simultaneously with its preparation (in particular, with the nucleo-synthesis or with decays from the previous state). One can use the following expression for the decay rate per a time unit:

$$I(t-t_0) = \int_{t_0}^t \rho_1^{(0)}(t') \rho_0^{(1)}(t-t') dt', \quad (16)$$

where  $\rho_m^{(n)}(t)$  are defined by the decay functions  $L_m^{(n)}(t)$  ( $m \neq n=0,1$ ) with relevant spectral distributions  $G_m^{(n)}$  for a "preparative" decay from the *initial* (first excited) state and the *subsequent* (ground) state, formed after the  $\gamma$ -decay of that initial one, respectively;  $\Gamma_1 = \Gamma_\alpha^{-1} + \Gamma_\gamma^{-1}$ ,  $\Gamma_\alpha^{-1}$  and  $\Gamma_\gamma^{-1}$  being the  $\alpha$ -decay and  $\gamma$ -decay widths of the excited parent  $\alpha$ -radioactive nucleus (with the first excited state of the internal  $\alpha$ -particle).

In the approximation of the single (one-step) absorptions of the emitted  $\gamma$ -quanta after the  $\gamma$ -decay one can use the expression

$$M(t) = q M_0 \exp(-ct/m) \quad (17)$$

(with  $c, \mu$  and  $q$  being the light velocity, the inversed  $\gamma$ -quanta absorption coefficient inside the matter and the dimensionless quantity, which depends on small loss of the  $\gamma$ -quantum energy due to the  $\gamma$ -scattering by nuclei and by electrons, respectively) for the  $\gamma$ -quanta propagation inside the matter sample [4,5]. After the simplifications at the approximations of small *recoil nucleus kinetic energy*  $\varepsilon_{\text{recoil}} \sim (E_1 - E_0)^2 / 2Am_n c^2$  ( $A$  and  $m_n$  being the atomic number of parent nucleus and the mean nucleon mass, respectively) after  $\gamma$ -quantum emission or absorption and the Doppler width for the resonant  $\gamma$ -emission and absorption  $\mathbf{D} \sim 2[\varepsilon_{\text{recoil}} kT]^{1/2}$  ( $k$  and  $T$  being the Boltzman constant and the sample temperature, respectively), then  $\Gamma_0 \ll \Gamma_1$ ,  $c \hbar / \mu$  and  $t - t_0 \gg \hbar / \Gamma_1$ ,  $\mu/c$  and also considering the diminution of *directly* decaying parent nuclei due to the  $\gamma$ -absorption, one obtains the following expression:

$$D(t) = D(t_0) + [P_0(t_0) + P_1(t_0) [1 - \exp(-\Gamma_0(t - t_0) / \hbar)] + P_1(t_0) Q_l [1 - \exp(-\Gamma_0(t - t_0) / \hbar)] \quad (18)$$

with

$$Q_l = (\Gamma_\alpha^{-1} / \Gamma_1) q M_0 [P_0(t_0) + P_1(t_0) (2\Gamma_1 + c\hbar/l) / 2(\Gamma_1 + c\hbar/l)],$$

or

$$D(t) = D(t_0) + [P_0(t_0) + (1 - Q_l) P_1(t_0)] \times [1 - \exp(-\Gamma_0(t - t_0) / \hbar)] + Q_l P_1(t_0), \quad (19)$$

$P_0$  and  $P_1$  being the abundances of the ground and the excited states of the parent  $\alpha$ -radioactive nuclei, respectively. The tentative estimations of  $q$  give values which are approximately within (1/2, 1).

Taking into account not only single (one-step)  $\gamma$ -absorptions but all possible *multiple*  $\gamma$ -absorptions, one can evaluate an every step of  $\gamma$ -absorptions by an equal contribution with  $M \cong qM_0 (2\Gamma_1 + c \hbar / \lambda) / 2(\Gamma_1 + c \hbar / \lambda)$  and easily sum up them in (18) or (19). Then, if  $\Gamma_\alpha^1 / \Gamma_1$  is not very small (let us say,  $\geq 0.1$ ), one can also roughly take  $N\Gamma_\alpha^1 / \Gamma_1 \cong 1$  in (19) ( $N$  being an effective number of the considered  $\gamma$ -absorptions steps). We can confirm such valuations with the help of the simple evident reasoning. For the lifetimes  $\tau_\gamma = \hbar / \Gamma_\gamma^1$  which are larger than  $10^{-13}$  sec and  $\Gamma_\gamma^1 \ll D$  and, moreover, for  $A \cong 250$ ,  $\varepsilon_\gamma \cong 50\text{keV}$  and  $N \cong 10$ , the quantity  $N\varepsilon_{\text{recoil}}$  satisfies the following expressions :

$$N\varepsilon_{\text{recoil}} < D \text{ when } T > 300^\circ\text{K (for the terrestrial lumps)}$$

and

$$N\varepsilon_{\text{recoil}} \ll D \text{ when } T \cong 10^9\text{K (inside the stars).}$$

Although up to now the partial  $\alpha$ -decay widths for excited states experimentally are not studied, we can expect that the condition  $N < 10$  is rather realistic on account of the Geiger and Nuttall law, at least for the high-energy excited states. If the values of  $\Gamma_\gamma$  and  $N\varepsilon_{\text{recoil}}$  get into such spreads  $D$  we can generalize (19) for the multiple  $\gamma$ -absorptions and write (with  $Q_\lambda \rightarrow q$ )

$$D(t) = D(t_0) + [P_0(t_0) + P_1(t_0)(1-q)][1 - \exp(-\Gamma_0(t - t_0) / \hbar)] + qP_1(t_0) \quad (20)$$

and hence

$$D(t) = D(t_0) + P_0(t_0)[1 - \exp(-\Gamma_0(t - t_0) / \hbar)] + P_1(t_0) \quad (21)$$

at the approximation of  $q \rightarrow 1$ .

The results (20) and (21) are valid at the approximations of infinitely large medium volumes, and sufficiently large times  $t - t_0$  (which are much larger than mean lifetimes of excited states and mean times of the free flight of  $\gamma$ -quanta inside the medium and also than times of quantum oscillations caused by different interference processes described by applying the Krylov-Fock theorem [6], generalized in [4,5](see also [7]) and at the condition  $(\hbar\Gamma_0 / \Gamma_\alpha^1)N \geq 1$ .

Let us analyze the new relation (21) in comparison with the previously known equation (4). They are basic for the determination of the age of a pattern but under the different conditions: with and without taking into account the intermediate  $\gamma$ -absorptions inside the sample matter. From the simple comparison of (4) and (21) one can see that

- i. for the same  $D(t_0)$ ,  $P_0(t_0)$ ,  $P_1(t_0)$ ,  $t_0$  and  $\Gamma_0$  at any moment  $t$  the value of  $D(t)$  in (21) is larger

than in (4) by the quantity  $P_1(t_0)$ ;

- ii. the same value  $D(t)$  is obtained in (21) at an earlier moment  $t$  than in (4);
- iii. the larger is the contribution of  $P_1$  into the sum  $P_0+P_1$ , the earlier is the moment  $t$  in (21) in comparison with (4) at which the same value of  $D(t)$  is obtained.

Now we illustrate the inference (ii) by the following instance:

if  $P_0(t_0)=P_1(t_0)=(1/2)P(t_0)$ , then the same value of  $D(t)$  is obtained for the values of  $t=t_{standard}$  in (5) and  $t=t_{real}$  in (20) which are connected by the following striking relation

$$t_{real} = t_{standard} - (\Gamma_0 / \hbar) \ln 2 \tag{22}$$

(of course, we imply that  $t_{standard} - t_0 > (\Gamma_0 / \hbar) \ln 2$ ).

The table 1 represents some impressive calculation results for the  $\alpha$ -decay of the nucleus-chronometer  $^{238}\text{U}$  with the half-life  $T_{1/2} (= \ln 2 \hbar / \Gamma_0) = 4.5 \cdot 10^9$  years on the base of (22).

$t_{real}$ years	$2 \cdot 10^3$	$4 \cdot 10^3$	$6 \cdot 10^3$	$8 \cdot 10^3$	$1 \cdot 10^4$	$1 \cdot 10^6$	$0.8 \cdot 10^8$	$1.8 \cdot 10^9$
$t_{standard}$ years	$3.32002 \cdot 10^9$	$3.32004 \cdot 10^9$	$3.32006 \cdot 10^9$	$3.32008 \cdot 10^9$	$3.3201 \cdot 10^9$	$3.321 \cdot 10^9$	$3.4 \cdot 10^9$	$3.5 \cdot 10^9$

**Table 1.** Comparison of the calculation results for  $t_{real}$  and  $t_{standard}$  the base of formula (22).

In Fig. 2 there is represented qualitative behavior of the corrected chronometry in astrophysics in comparison with the standard chronometry.

So, sometimes billions years obtained by the usual nuclear-chronometry method can correspond to several thousands years and always there is no exponential decay law of the alpha-radioactive chronometer at all.

Of course, the results (i) and (ii) will be even stronger, if one will consider more than one excited state of the parent nuclei with the excited internal  $\alpha$ -particles. It is possible to state that the usual (non-corrected) "nuclear clocks" do really indicate to the *upper limits* of the durations of real  $\alpha$ -decay processes. So, sometimes billions years obtained by the usual nuclear-chronometry method can correspond to several thousands years. Of course, the results (i) and (ii) will be even stronger if one will consider more than one excited state of the parent nuclei with the excited internal  $\alpha$ -particles. It is possible to state that the usual (non-corrected) "nuclear clocks" do really indicate to the *upper limits* of the durations of real  $\alpha$ -decay processes.

Further, also it is experimentally shown in [15] that the lifetime of the  $\beta$ -radioactive isotope  $^{187}\text{Re}$  in a bare (without electrons) atom is less  $10^9$  times than the lifetime of the usual atom with totally filled electronic shell (it becomes  $\sim 30$  years instead of  $\sim 3 \cdot 10^{10}$  years!). And it is known that the stellar matter is in the plasma state with the free nuclei and electrons. So, the nuclei partially or totally are being deprived their electronic shells inside the stars.

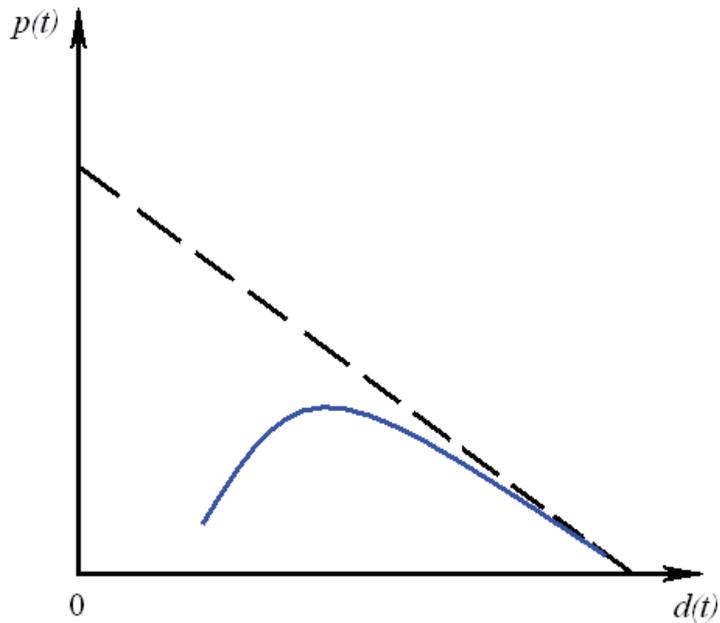


Figure 2. The typical plot (21)21 in comparison with (4).

## 5. Nuclear chronometry in geophysics

Now we consider the revision of the methods of the terrestrial nuclear chronometry taking into account the relatively weak but constant cosmic radiation on the upper earth layers up to the depth of  $\sim 5$  meters. Under the influence of the cosmic radiation, the following processes are going on: (1) the constant formation of the excited nuclei-chronometers with the smaller life-times than for the nuclei-chronometers in the ground state, (2) the effective acceleration of the  $\alpha$ -decay through knocking-out the  $\alpha$ -particles by cosmic protons and (3) the constant nonzero removal of nuclei-chronometers through the channels of inevitable rearrangement nuclear reactions. *These processes lead to the real diminishment of the results of measurements of the decay times.*

Let us analyze the diminishment caused by the first and the second kinds of processes.

In our case, taking into account the cosmic radiation (supposing its flux to be constant in time) and using the same method of the generalization of the Krylov-Fock theorem as in [2-4], we, instead of  $L(t-t_0)$  (with the consequences (1)-(2)), obtain:

$$L_p(t-t_0) = [1 - a \times (t - t_0)] L(t-t_0) \quad (23)$$

for a unit ( $1\text{cm}^3$ ) chronometer volume, where  $a = j_{\text{cosm}} \sigma \nu n$ ,  $j_{\text{cosm}}$  is the cosmic (mainly proton) radiation flux (in  $\text{cm}^{-2} \text{sec}^{-1}$ ),  $\sigma$  is the total cross section of all reaction proton-chronometer processes with the removal of the nuclei-chronometers,  $\nu$  is the number of multiple collisions in

the medium after the first proton-chronometer collision and  $n$  is the mean nucleus-chronometer number along the 1 cm-depth and  $L_p(t - t_0)$  includes all parent-nucleus diminutions through collisions of chronometers with the cosmic protons. Here for the simplicity we neglect the elastic and inelastic scattering. We see that for  $t - t_0 \geq 1/a$   $L_{p-D}(t - t_0) = 0$  and  $P(t - t_0) = 0$ .

If we select only the reactions ( $p, p' \alpha$ ) with the general cross section  $\sigma_D$  which strongly accelerate the emission of the  $\alpha$ -particles with the formation of daughter nuclei, neglecting all other processes, then we obtain the particular relation

$$L_{p-D}(t - t_0) = [1 - a_D \times (t - t_0)] L(t - t_0) \quad (24)$$

where  $a_D = j_{\text{cosm}} \sigma_D \nu n$ , and  $L_{p-D}(t - t_0)$  denotes the decay function in this case. And in this case relation (2b)4 passes into

$$p(t) [[1 - a_D \times (t - t_0)] \exp(\Gamma_0(t - t_0) / \hbar) - 1] - d(t) + d(t_0) = 0. \quad (25)$$

From the comparison of (25) with (4) one can see that when  $t - t_0 \rightarrow 1/a$  the time duration defined by the old method (without taking the cosmic radiation into account) becomes very large – much more than  $\Gamma_0 / \hbar$ .

So, one can see that that qualitatively in the terrestrial layers of geophysics the main equation (25) is rather similar to the standard chronometry but with one modification – the insertion of a term  $a_D(t - t_0)$  which can noticeably diminish the age of the earth and also distort the exponential law of  $\alpha$ -decay of the  $\alpha$ -radioactive chronometers.

For qualitative evaluations we have taken  $\sigma_D = 3 \cdot 10^{-25} \text{ cm}^2$  and  $\nu$  between  $10^2 \cdot 10^3$  for the mean proton energy  $\sim 10^9 \text{ eV}$ , the flux  $j_{\text{cosm}} = 0.85 \text{ (cm}^2\text{sec)}^{-1}$  in the top atmosphere layer or  $1.75 \cdot 10^2 \text{ (cm}^2\text{sec)}^{-1}$  on the sea level [16] and  $n = 1 \text{ cm} / 3 \cdot 10^{-8} \text{ cm} = 0.33 \cdot 10^8$ . Of course, practically it is impossible now to calculate  $\nu$ , because the effective value of  $\nu$  is defined not only by mean proton energy and by nucleon, cluster and fragment binding energies but also by usually unknown cross sections of all possible reactions for wide energy region – and hence we have used very simplified evaluations. Then we obtain values of  $a$  between  $(1/1.5) \cdot 10^8$  years and  $(1/2.7) \cdot 10^5$  years. From the result we can see that when  $P(t) \rightarrow 0$ , for both values of  $a$  and also in both cases (with valid and invalid relation (25)) the real time duration is essentially less than without taking the cosmic radiation into account.

As to the more deep terrestrial layers and the core of the earth, we have to take into account the history of the earth formation. Now there is no unique generally accepted theory or even model of the planet origin. There are two known groups of such models: (1) models where one considers the planet origin from the final cooled pieces of the exploded star or super-nova, (2) models where one considers planets as results of the cosmic dust condensation.

Relative to any model from the first group, if one takes any piece of the terrestrial mass, it is impossible to distinguish the genetically real parent and daughter nuclei from the admixture of the same kinds of nuclei formed in other parts of the cooled and transformed parent stellar

piece. Therefore one can approximately suppose that the earth age is the sum of the parent star (or the preceding super-nova) age before exploding and of the consequent age of the formed earth. The last age can be determined also by the methods of the nuclear chronometry but in different ways inside the earth and on the surface of the earth. Deeply inside the earth one has to consider consequences of the

formation and decay of the excited nuclei-chronometers during the preceding stellar (and super-nova) nucleo-synthesis and during subsequent planet cooling in the melted magma (inside the earth). And in the surface layers of the earth (up to the depth of ~ 5 meters) we can consider the influence of the cosmic radiation which was presented above. For both cases it is also necessary to take into account the unknown now initial nonzero quantity of the daughter nuclei in the earth (in the examined earth pieces) from the previous stellar (nucleo-synthesis and chronometer-decay-chain) processes.

Relative to any model from the second group, from the very beginning it is necessary to take into account the cosmic dust origin. Hypothetically the cosmic dust was born partially simultaneously with first stars after Big Bang and partially during the star evolution – from the cooled micro-rejections out of stars and super-nova during their perturbations and explosions. And now there is neither a systematic theory of the dust origin, independent from the star origin, nor a systematic theory of the dust condensation → clotting into a planet. We can approximately evaluate the mean existence time of the earth beginning from the hypothetical mean instant of the conventional dense clotting of the condensed dust into the planet by the methods of nuclear chronometry if we know the real initial quantities of the parent and the daughter nuclei just in this mean instant. *And a nonzero initial quantity of the daughter nuclei always leads to a real diminishment of the evaluation of the decay time – a larger diminishment for a larger initial quantity of the daughter nuclei.* Moreover, we have to take into account the constant excitation of radioactive nuclei-chronometers by the cosmic radiation and then, in the case of large masses, also the formation of  $\gamma$ -emission-absorption chains with accompanied multiple excitations of nuclei-chronometers.

## 6. A short previous history of my investigations on the nuclear chronometry

We use in time analysis of nuclear processes elaborated on time as a quantum observable, canonically conjugate to energy, the general expression for the duration  $\langle \tau \rangle$  like [10,11]

$$\langle \tau_{f_i} \rangle = \langle t_f \rangle_f \dots - \langle t_i \rangle_i \equiv \int_{-\infty}^{\infty} t w_f(z_f, t) dt - t_{in} \tag{26}$$

where  $w_f = j(z_f, t) / \int_{-\infty}^{\infty} j(z_f, t) dt$ ,  $j_f(z_f, t) = \text{Re}(\psi_f(-i\hbar/2\mu_f \partial/\partial z_f) \psi_f)$  is the  $z_f$  the component of the

probability flux; the brackets (...) define the integration over all coordinates, entered in the wave functions  $\psi_f$ , besides  $z_f$ ;  $z_f$  is the axis, directed along the mean velocity  $\langle \vec{v} \rangle_f$  of the relative motion of the decay pair (for instance,  $\alpha$ -particle and the daughter nucleus) with the reduced mass  $\mu_v$  and their wave function of the internal motion  $|v\rangle$  with energy  $e_v$ ;  $\vec{v}_v = \hbar \vec{k}_v / \mu_v$ ,  $\epsilon_v = \hbar^2 k_v^2 / 2\mu_v \equiv E - e_v$ ;  $t_{in}$  is defined as  $\langle t_i \rangle$  where  $i$  is the initial channel of forming the decaying nucleus or simply as an initial time moment  $t_{in}$

The wave packet of the relative motion of the decay pair in one-dimensional radial asymptotic limit is described as

$$\Psi_f(r_f, t) = r_f^{-1} \int_0^\infty dE g(E) F_{if}(E) \exp[ikr_f - iEt / \hbar] \tag{27}$$

where

$$F_{if}(E) = \frac{C_{if}}{E - E_r + i\Gamma/2}, \tag{28}$$

with  $C_{if}$  being a constant or smooth function of final-particle kinetic energy  $E$  in the region  $(E_r - \Gamma/2, E_r + \Gamma/2)$ ,  $E_r$  and  $\Gamma$  are the resonance energy and width, respectively. For  $z_f \geq R_\beta$ , where  $R_\beta$  is the radius of interaction in the final channel, and under condition  $\Gamma \ll \Delta E \ll E_r$ , it is possible to re-write (27) in the following simplified form

$$\Psi_f(R_f, t) = A \int_0^\infty dE \frac{\exp[-iEt / \hbar]}{E - E_r + i\Gamma/2}, \tag{29}$$

where  $A$  is a constant. For  $\Gamma = \text{constant}$  we obtain

$$\Psi_\beta(R_\beta, t) = \begin{cases} B \exp[-iE_r t / \hbar - (\Gamma/2\hbar)t], & \text{for } t > 0 \\ 0, & \text{for } t < 0 \end{cases} \tag{30}$$

(shifting, for very small  $\Gamma$ , the lower limit of the integration in (29) from 0 till  $-\infty$  and utilizing the residue theorem). Here  $B$  is constant and more precisely here will be  $t - t_{in}$  instead of  $t$ .

Of course,  $w_f$  for wave function (30) is equal

$$L(t) = (\Gamma / \hbar) \exp(-\Gamma t / \hbar). \tag{31}$$

The function  $L(t)$ , generally speaking, defined by (6) and concretely by (31), sometimes are named by the *decay function* (see, for instance, [4-10]). The more detailed study (see, for instance, [12-14]) shows that even for an isolated resonance the exponential form of the decay law is

approximately performed in the time interval, limited by below by the quantity  $t_1 \sim t_0 \Gamma/E_r$  and from above by the quantity  $t_2 \sim t_0 \ln(\Gamma/E_r)$ , where  $t_0 \sim \hbar/\Gamma$ , and the accuracy of describing by the exponential function is the better, the lesser is the relation  $\Gamma/E_r$  [14]. The functions (6) and (31) describe the essence of the Krylov-Fock theorem which was derived by another way (see the section before) and states that the decay law of the meta-stable state is totally defined by the energy spectrum of the initial state [8].

The author (V.S.O.) had been extended the contents of the Krylov-Fock theorem, considering the formation of the  $\alpha$ -radioactive chronometers in stars and also the emissions, the successive absorptions etc of  $\gamma$ -quanta by the nuclei-chronometers in [4] on the base of his earlier article [9].

## 7. Several words on measurements of $P$ and $D$ for astrophysical and geophysical processes

Of course, it should be noted that is impossible for us to measure *directly* with usual radioactive methods  $P$  and  $D$  in the sun or any astrophysical  $\alpha$ - and  $\gamma$ -radioactive decay processes. And the only way for such measurements is to adapt the photon ( $\gamma$ -ray) spectroscopic method for defining  $P$  and  $D$  for the  $\alpha$ -radioactive chronometer with correspondent  $\gamma$ -transitions, analyzing the correspondent energy peaks of  $\alpha$ - and  $\gamma$ -lines with the application of  $\alpha$ - and  $\gamma$ -sensors of the sun or star photon radiation.

As to the geophysical processes, it is possible to use the usual for earth measurements of  $P$  and  $D$ , which are usual for earth. As to moon or some other planet or satellite, it is possible to use the remote satellite sensors with the broadcast of their measurements for the earth through the earth atmosphere and in cosmos.

## 8. Conclusions and perspectives

The presented simplified estimations brings to the conclusion that the usual (non-corrected) "nuclear clocks" do really indicate not to the realistic values but to the *upper limits* of the durations of the  $\alpha$ -decay stellar and planet processes and also that the realistic durations of these processes have to be noticeably (at least several orders) smaller. And the results of this paper represent *what physical processes* can strongly influence on estimations of the sun and earth ages, which has to be much smaller than usual estimations.

As a continuation and expansion of the exposed results, it is necessary to propose the following program for future investigations, taking into account:

1. *the chains of the successive decays of every star and terrestrial chronometer,*
2. *all the initial excited states of every star and terrestrial chronometer,*
3. *joint considering the both  $\alpha$ - and  $\beta$ -radioactive decays and finally,*

4. the possible modifications of processes of the stellar and cosmic nucleo-synthesis of all nuclei-chronometers.

## Author details

V.S. Olkhovsky

Address all correspondence to: olkhovsky@mail.ru

Department of Theory of Nuclear Processes in Institute of Nuclear Research of NASU, Kiev, Ukraine

## References

- [1] Audouze J. and Vauclair S., An Introduction to Nuclear Astrophysics. The Formation and the Evolution of Matter in the Universe (Reidel Publish.Comp.)1980, Ch.VIII.
- [2] Fowler W.A., Rev.Mod.Phys.1984; 56 149-184.
- [3] Wasserburg G.J., Rev.Mod.Phys.1995; 67 805-905.
- [4] Olkhovsky V.S., Atti dell'Accademia Peloritana dei Pericolanti, Sci.Fis., Mat.e Nat. (Messina) 1998; LXXVI 59-65.
- [5] Olkhovsky V.S., Influence of excited radioactive nuclei for results in large-scale nuclear chronometry, International Conference of Nuclear Physics at Border Lines 2001: conference proceeding, May 21-24,2001, Messina, Italy; 2001, pp.241-247.
- [6] Olkhovsky V.S., Dolinska M.E. and Doroshko N.L., Problems of Atomic Science and Technology (Ukraine, Khar'kov) 2009; N3 9-14;
- [7] Olkhovsky V.S. and Dolinska M.E., Central Eur. J. Phys.2010;.8(1) 95-100.
- [8] Krylov N.S. and Fock V.A., Zh.Eksp.Teor.Fiz.(USSR)1947; 17 93-99.
- [9] Olkhovsky V.S., Izvestia AN SSSR, serie phys.(USSR)1985 49(5) 938-47.
- [10] Olkhovsky V.S.(Ol'khovskii), Sov. J. Particles Nucl.(Engl.Transl.),(United States), 1984;15(2) 130-148.
- [11] Olkhovsky V.S., Advances in Math.Phys.2009 vol. 2009, article ID 859710, 83 pages, doi:10.1155/2009/859710.
- [12] Halfin L.A. [in Russian: Халфин Л.А., ЖЭТФ 1957;33, 1371-1378].
- [13] Rosenfeld L., Nucl.Phys.1965; 70 1-11.

- [14] Baz' A.I., Perelomov A.M., Zel'dovich Ya.B., Scattering, reactions and decays in non-relativistic quantum mechanics, Jerusalem:Israel Program for Scientific Translations; 1969.
- [15] Wefers E., Bosch F., Faestermann T. et al, Nuclear Phys.1997; A626 215-224.
- [16] Friedlander M.W., Cosmic Rays,Canbridge,MA: Harward University Press);1989, 160p.; see also references in: Cosmic Rays, from: The Columbia Encyclopedia, sixth edition; 2007.



---

# Exploring and Using the Magnetic Methods

---

Othniel K. Likkason

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/57163>

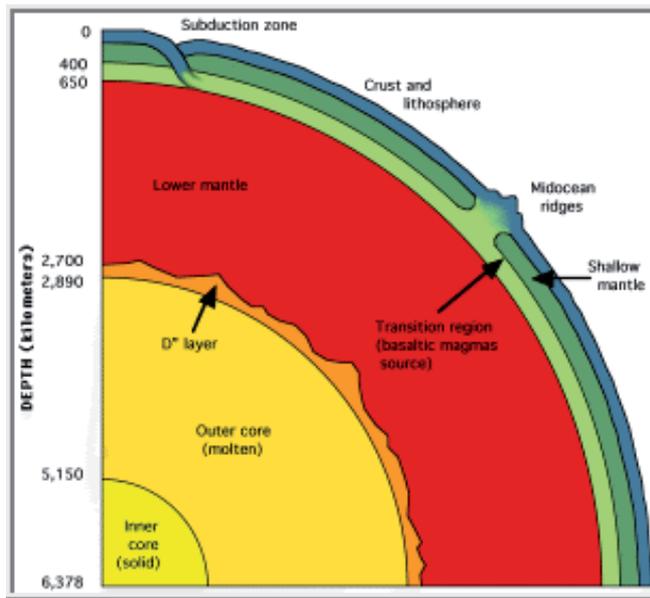
---

## 1. Introduction

The Earth is principally made up of three parts: core, mantle and crust (Fig. 1). As understood today, right at the heart of the Earth is a solid inner core composed primarily of iron. At 5,700°C, this iron is as hot as the Sun's surface, but the crushing pressure caused by gravity prevents it from becoming liquid. Surrounding this is the outer core, a nearly 2,000 km thick layer of iron, nickel, and small quantities of other metals. Lower pressure than the inner core means the metal here is fluid. Differences in temperature, pressure and composition within the outer core cause convection currents in the molten metal as cool, dense matter sinks while warm, less dense matter rises. This flow of liquid iron generates electric currents, which in turn produce magnetic fields (Earth's field). These convection processes in the liquid part of core (outer core) give rise to a dipolar geomagnetic field that resembles that of a large bar magnet aligned approximately along the Earth's rotational axis. The mantle plays little part in the Earth's magnetism, while interaction of the past and present geomagnetic field with the rocks of the crust produces magnetic anomalies recorded in detail when surveys are carried out on or above the Earth's surface.

The magnitude of the Earth's magnetic field averages to about  $5 \times 10^{-5}$  T (50,000 nT). Magnetic anomalies as small as 0.1 nT can be measured in continental magnetic surveys and may be of geological significance.

The magnetic methods, perhaps the oldest of geophysical exploration techniques bloomed after the World War II. Today, with improvements in instrumentation, navigation and platform compensation, it is possible to map the entire crustal section at a variety of scales from strongly magnetic basement at a very large scale to weakly magnetic sedimentary contacts at small scale. Methods of magnetic data treatment, filtering, display and interpretation have also advanced especially with the advent of high performance computers and colour raster graphics.



**Figure 1.** Internal structure of the Earth (from <http://zebu.uoregon.edu>)

As is well known today, magnetic methods are used to solve various problems such as:

1. Mapping the basement surface and sediments in oil/gas exploration
2. Detecting different types of ore bodies in mining prospecting
3. Detecting metal objects in engineering geophysics
4. Mapping basement faults and fractures
5. Determining zones with different mineralization in logging as well as inspecting casing parameters
6. Studying the magnetic field of the Earth and its generators and
7. A variety of other purposes such as natural hazards assessment, mapping impact structures and environmental studies.

Magnetic observations are obtained relatively easily and cheaply and a few corrections are applied to them. This explains why the magnetic methods are one of the most commonly used geophysical tools. Despite these obvious advantages, interpretations of magnetic observations suffer from a lack of uniqueness due to dipolar nature of the field and other various polarization effects. Geologic constraints, however, can considerably reduce the level of ambiguity. Information from magnetic surveys comes from rock units at depth as well as from those at or near the surface. This is the strength of the magnetic method (or any surface geophysical method), making it more powerful than any other remote sensing method which relies on the information from reflections of electromagnetic (EM) waves by materials on the Earth surface.

Thus while the natural magnetic field of the Earth is measured in magnetic method, EM radiation normally is used as the information carrier in remote sensing. Electromagnetic radiation is a form of energy with the properties of a wave, and its major source is the sun. Solar energy traveling in the form of waves at the speed of light is known as the electromagnetic spectrum. Passive remote sensing systems record the reflected energy of electromagnetic radiation or the emitted energy from the Earth, such as cameras and thermal infrared detectors. Active remote sensing systems send out their own energy and record the reflected portion of that energy from the Earth's surface, such as radar imaging systems.

In this chapter, we explore the magnetic methods of geophysical exploration. The first part of the chapter covers the fundamental concepts of magnetic force field, the Earth's magnetic field and its relationship with gravity field. The second part deals with the measurement procedures and treatment of the magnetic field data, while the third part covers the magnetic effects of simple geometric bodies, processing and interpretation of magnetic data and ending it with treatment, analysis and interpretation of real field data.

## 2. Fundamental magnetic theories

Any magnetic grain is a dipole. That is, it has two poles,  $P_1$  and  $P_2$  of opposite signs diametrically linked. Charles Augustin de Coulomb in 1785 showed that the force of attraction or repulsion between electrically charged bodies and between magnetic poles obeys an inverse square law similar to that derived for gravity by Newton.

The mathematical expression for the magnetic force,  $F_m$  experienced between two magnetic monopoles is given by:

$$F_m = \frac{1}{\mu} \frac{P_1 P_2}{r^2} \quad (1)$$

Where  $\mu$  is a constant of proportionality known as the magnetic permeability,  $P_1$  and  $P_2$  are the 'strengths' of the magnetic monopoles and  $r$  is the distance between the poles.

We note that the expression in equation (1) is identical to the gravitational force,  $F_g = G \frac{m_1 m_2}{r^2}$  and electrical force,  $F_e = k \frac{q_1 q_2}{r^2}$  expressions. Here  $m_1$ ,  $m_2$  and  $q_1$ ,  $q_2$  are respectively masses and electrical charges separated by distance  $r$ ,  $G$  is the universal gravitational constant while  $k$  is the Coulomb's law constant for the medium. However, unlike the gravitational constant,  $G$ , the magnetic permeability,  $\mu$  is a property of the material medium in which the two monopoles,  $P_1$  and  $P_2$  are situated. If they are placed in a vacuum, then  $\mu$  is for the free space. Also, unlike  $m_1$  and  $m_2$ ,  $P_1$  and  $P_2$  can be either positive or negative in sign. If  $P_1$  and  $P_2$  have the same sign, the force,  $F_m$  between the two monopoles is repulsive. If  $P_1$  and  $P_2$  have opposite signs,  $F_m$  is attractive.

We may seem to easily compare the gravitational force between masses  $m_1$  and  $m_2$  separated by  $r$  to that of either the attractive or repulsive magnetic force between two monopoles.

However, the magnetic monopoles have never existed! Rather the fundamental magnetic element appears to consist of two magnetic monopoles: one positive and the other negative, separated by a distance. Thus the fundamental magnetic element consisting of two monopoles is called a magnetic dipole. Every magnetic grain is therefore a dipole.

We can therefore determine the force produced by a dipole by considering a force produced by two monopoles. Since the dipole is simply two monopoles, each of strength  $P_1$  and  $P_2$ , we expect that the force generated by a dipole is simply the force generated by one monopole added vectorially to the force generated by the second monopole. Consequently, the force distribution for a dipole is nothing more than the magnetic force distribution observed around a simple bar magnet. Thus a bar magnet can be thought as two magnetic monopoles separated by a length of the magnet. The magnetic force appears to originate out of the North Pole (N) of the magnet and to terminate at the South Pole (S) of the magnet. Some of the field lines pass through the material of the magnet (high concentration because of high  $\mu$ ), some pass through air (low concentration because of low  $\mu$ ). Notice that even in air; the poles have high density of field lines. Also, the lines radiate out from N (vertically outward) and radiate into S (vertically inward). Between the length of the bar in air, the magnetic field directions are variable, but with the middle of the bar having a near horizontal field direction. Again, the field strength and direction at any point around the bar magnet is a vector sum of the force field contributed by each of the monopole (N or S).

When we examine equation (1) in terms of unit of measurement, we see that the magnetic force,  $F_m$  retains its fundamental unit of newton (N) and  $r^2$  would be in square metre ( $m^2$ ). Permeability,  $\mu$  by the S. I. unit definition, is a unitless constant. The units of the pole strength,  $P$  are defined such that if a force of 1 N is produced by two unit poles separated by a distance of 1 m, then each unit pole has a strength of one ampere-metre (1 Am). Thus a unit pole has an S.I. unit of ampere-metre.

We can also define, from equation (1), the force per unit pole strength exerted by a magnetic monopole,  $P_1$  or  $P_2$ . This is called magnetic field strength or magnetizing force,  $H$ . Thus

$$H = \frac{F_m}{P_2} = \frac{P_1}{\mu r^2} \quad (2)$$

Here again, given the units associated with force (N) and magnetic monopoles (Am), the unit associated with magnetic field strength,  $H$  are N/A-m and by definition, 1 N/A-m is referred to as a tesla (T): named after a Croatian inventor, Nikola Tesla. Thus 1 T = 1 N/Am. Indeed from equation (2), the unit of  $H$  can be expressed as Am/ $m^2$  or  $Am^{-1}$  (ampere per metre). Thus 1 N/Am = 1  $Am^{-1}$  = 1 T. Similarly, the unit of magnetic flux is weber (Wb) and magnetic flux per unit area is the magnetic strength we have been talking about. Thus the unit of magnetic strength can also be expressed in weber per square metre (Wb/ $m^2$ ). Hence 1 Wb/ $m^2$  = 1 T.

When describing the magnetic field of the Earth, it is common to use units of nanotesla (nT), where 1 nT =  $10^{-9}$  T. The average strength of the Earth's magnetic field,  $H$  is about 50,000 nT (ranges from 20,000 to 70,000 nT). A nanotesla has the value as the old unit of gamma (1 nT = 1 gamma).

When magnetic materials or rocks are placed within a field,  $T$  (a magnetizing force such as  $H$  given in equation (2)), the magnetic materials or rocks will produce their own magnetizations or polarizations. This phenomenon is called induced magnetization, magnetic polarization or magnetic induction. The strength of the magnetic field induced on the magnetic material due to the inducing field,  $T$  is called the intensity of magnetization or magnetic polarization,  $J_i$ ; where

$$J_i = kT \quad (3)$$

The constant of proportionality,  $k$  is the magnetic susceptibility and is a unitless constant determined by the physical properties of the magnetic material. The susceptibility,  $k$  can either be positive or negative in values. Positive values imply that the field,  $J_i$  is in the same direction as the inducing field  $T$ . Negative  $k$  implies that the induced magnetic field is in the opposite direction as the inducing field. Details of the mechanisms of induced magnetization can be further obtained from [1].

In magnetic exploration method, the susceptibility is the fundamental material property whose spatial distribution, we attempt to determine. We see that magnetic susceptibility is analogous to density in gravity surveying. Unlike density, there is a large range of susceptibilities even within materials and rocks of the same type. This definitely will put limit to knowledge of rock type through susceptibility mapping of an area.

Magnetic susceptibility in SI unit is a dimensionless ratio having a magnitude much less than 1 for most rocks. Hence a typical susceptibility value may be expressed (as for example)  $k = 0.0064$  SI. In the old c.g.s. system of electromagnetic units (emu), the numerical value of magnetic susceptibility for a given specimen is smaller by a factor of  $4\pi$  than the SI value. Thus  $k(\text{SI}) = k(\text{emu}) \times 4\pi$ . Hence for  $k = 0.0064$  SI,  $k(\text{emu}) = k(\text{SI})/4\pi = 0.00051$  emu.

### 3. The Earth's magnetic field

Nearly 90% of the Earth's magnetic field (geomagnetic field) looks like a magnetic field that would be generated from a dipolar magnetic source located at the centre of the Earth and nearly aligned with the Earth's rotational axis. This field is believed to originate from convection of liquid iron in the Earth's outer core [2] and is monitored and studied using global network of magnetic observatories and various satellite magnetic surveys. If this dipolar description of the Earth's field were complete, then the magnetic equator would nearly correspond to the Earth's geographic equator and the magnetic poles would also nearly correspond to the geographic poles. The strength of the Earth's field at the poles is about 60,000 nT. This is called the Main Field of the Earth. This field changes slowly with time and is believed to go through a decay and collapse, followed by polar reversal on a time scale of the order of 100,000 years [3], [4]. The construction of a global magnetic reversal timescale is of fundamental importance in deciphering Earth's history. For details on such discussion, [5] can be consulted.

The remaining 10% of the Earth's magnetic field cannot be explained in terms of simple dipolar sources. The larger component of this 10% of the Earth's field originates in iron-bearing rocks near the Earth's surface where temperatures are sufficiently low (i.e. less than the Curie temperature of the rocks). This region is confined to the upper 30 – 40 km of the crust and is the source of the crustal field which is made up of induced field on magnetically susceptible rocks and remanent magnetism of the rocks. The smaller portion of the 10% comes from the upper atmosphere (external source).

The external source field is believed to be produced by interactions of the Earth's ionosphere with the solar wind. Hence some temporal variations (usually variable over hours at tens of nT or occasionally variable over a few hours at hundreds of nT: the magnetic storm) are correlated to solar activity. The external component (except for magnetic storm phenomenon) is usually regular and are corrected/removed appropriately from field measurements in a process similar to drift correction in gravity surveys. Where magnetic storm is detected, survey is most often discontinued until after the phenomenon has passed.

The crustal field, its relation to the distribution of magnetic minerals within the crust, and the information this relation provides about exploration targets are the primary subjects of the magnetic method in exploration. In a magnetic survey, the magnetic induction,  $B$  whose magnitude is measured at a point is the vector sum of four field components:

1. The Earth's **main field** which originates from dynamo action of conductive fluids in the Earth's deep interior [6];
2. An **induced field** caused by magnetic induction in magnetically susceptible earth materials polarized by the main field [7];
3. A field caused by **remanent magnetism** of earth materials [7]; and
4. Other (usually) less significant fields caused by solar, atmospheric [8] and cultural influences

While we can handle the external features (source 4) component (like drift correction in gravity survey: for the solar/atmospheric sources) and divesting from such features or recognizing their transient effects and removing them (for cultural features), the main field is examined from complex models that have been developed and are available. Our intent here is to characterize the global magnetic field (main field) in order to isolate the magnetic field caused by crustal sources (sources 2 and 3).

Spherical harmonic analysis provides the means with which to determine from measurements of a potential field and its gradient on a sphere whether the sources of the field lie within the sphere or outside the sphere. Carl Friederich Gauss in 1838 was the first to describe the geomagnetic field in this way and concluded that the observed field at the Earth's surface originates entirely from within the Earth. However, we know today from satellite observations, space probes and vast accumulation of information from field measurements that a small part of the geomagnetic field originates from outside the Earth.

We consider a magnetic induction vector,  $\vec{B}$  at a point on or above the Earth's surface and its potential,  $V$ , such that  $\vec{B} = -\nabla V$ . If we assume a source-free space,  $V$  is harmonic and satisfies Laplace's equation:

$$\nabla^2 V = 0 \tag{4}$$

Following [9], if no sources exist outside the sphere, then both  $V$  and  $\frac{\partial V}{\partial r}$  must vanish for  $r \rightarrow \infty$  and hence:

$$V^i = a \sum_{n=0}^{\infty} \left(\frac{r}{a}\right)^{n+1} \sum_{m=0}^n (A_n^{mi} \cos m\varphi + B_n^{mi} \sin m\varphi) P_n^m(\theta) : r \geq a \tag{5}$$

On the other hand, if all sources lie outside the sphere, then  $V$  and  $\frac{\partial V}{\partial r}$  must be finite within the sphere and appropriately,

$$V^e = a \sum_{n=0}^{\infty} \left(\frac{r}{a}\right)^n \sum_{m=0}^n (A_n^{me} \cos m\varphi + B_n^{me} \sin m\varphi) P_n^m(\theta) : r \leq a \tag{6}$$

Where in both equations (5) and (6), the superscripts,  $i$  and  $e$  denote internal and external sources respectively,  $\theta$  is the co-latitude (latitude =  $90^\circ - \theta$ ),  $\phi$  is longitude,  $r$  is the radial distance from the centre of the sphere,  $a$  is the radius of the sphere, and  $P_n^m(\theta)$  is an associated Legendre polynomial of degree  $n$  and order  $m$  normalized according to the convention of Schmidt. The magnitude of the normalized Schmidt surface harmonics when squared and averaged over the sphere can be expressed as

$$\frac{1}{4\pi r^2} \int_0^{2\pi} \int_0^\pi P_n^m(\theta) \left\{ \frac{\cos m\varphi}{\sin m\varphi} \right\} P_{n'}^{m'}(\theta) \left\{ \frac{\cos m'\theta}{\sin m'\theta} \right\} r^2 \sin \theta d\theta d\varphi = \begin{cases} 0 & n \neq n' \text{ or } m \neq m' \\ \frac{1}{2n+1} & n = n' \text{ and } m = m' \end{cases} \tag{7}$$

For example, the normalized surface harmonics for  $n = 0, m = 0$  is 1, for  $n = 1, m = 0$  is  $\cos\theta$ , for  $n = 1, m = 1$  is  $\sin\theta \left\{ \frac{\cos}{\sin} \right\} \varphi$ , etc.

Different types of surface harmonics can be deduced from the nature and forms of the normalized term:  $P_n^m(\theta) \left\{ \frac{\cos m\varphi}{\sin m\varphi} \right\}$ . If  $m = 0$ , the surface harmonic depends on colatitude,  $\theta$  or latitude ( $90^\circ - \theta$ ) as the longitude component vanishes. This surface harmonic is called the zonal harmonic. If  $n-m = 0$ , the surface harmonics depends on longitude and is called the sectoral harmonic (resembles the sectors of an orange). If  $m > 0$  and  $n-m > 0$ , the harmonic is termed as tesseral harmonic. These harmonics are useful in characterizing the relative importance of coefficients  $A_n^m$  and  $B_n^m$  in equations (5) and (6).

If sources exist both inside and outside the sphere, then the potential,  $V$  in source-free regions near the surface of the sphere is given by the sum of the equations (5) and (6). Thus it is further

convenient to express the combination of equations (5) and (6) in terms of Gauss' coefficients  $g_n^m$  and  $h_n^m$  for free-space (where the permeability,  $\mu_0 = 1$  c.g.s.). Thus

$$V = a \sum_{n=1}^{\infty} \left(\frac{a}{r}\right)^{n+1} \sum_{m=0}^n [g_n^m \cos m\varphi + h_n^m \sin m\varphi] P_n^m(\theta) \quad (8)$$

It is generally known that  $n = 1$  harmonic from equation (8) gives the first three coefficients ( $g_1^0$ ,  $g_1^1$  and  $h_1^1$ ) which have overwhelming dominance. There is no  $g_0^0$  term as this corresponds to the potential of a monopole which must therefore be zero. The first degree harmonic describes the potential of a dipole at the centre of the sphere and therefore the large amplitudes of these coefficients reflect the generally geocentric dipolar character of the main geomagnetic field.

Excluding the  $n = 1$  harmonic from equation (8) eliminates the dipole term from the geomagnetic field and leaving a remainder of the form called the non-dipole part.

At the point of observation,  $P$ ,  $T$  is the magnitude of the total field intensity, and  $X$ ,  $Y$ ,  $Z$  and  $H$  are the north, east, vertical and horizontal components respectively. The quantity,  $I$  is the angle  $T$  makes with the horizontal (along which  $H$  is directed) and is called the dip or inclination, while  $D$ , the declination is the angle the horizontal field,  $H$  makes with the true or geographic north. Note that  $H$  is not the same here as the one expressed in equation (2).

We note that a simple dipole theory predicts that the magnetic inclination,  $I$  is related to the geographic latitude,  $\varphi$  as  $\tan I = 2 \tan \varphi$ .

The vector elements of the Earth's magnetic field at a point are  $\vec{T}$ ,  $\vec{X}$ ,  $\vec{Y}$ ,  $\vec{Z}$ ,  $D$ ,  $I$ . Like the reference ellipsoid and theoretical gravity, the mathematical representation of the low-degree parts of the geomagnetic field is determined by international agreement. This mathematical description is called the International Geomagnetic Reference Field (IGRF) and is attributed to the International Association of Geomagnetism and Aeronomy (IAGA) and its umbrella organization, the International Union of Geodesy and Geophysics (IUGG).

The IGRF is essentially a set of Gaussian coefficients,  $g_n^m$  and  $h_n^m$  that are put forth every 5 years by IAGA for use in a spherical harmonic model. At each of these epoch years, the group considers several proposals and typically adopts a compromise that best fits the data available. The coefficients for a given epoch year are referred to by IGRF and then the year, as in IGRF2000. The model includes both the coefficients for the epoch year and secular variation variables, which track the change of these coefficients in nanotesla per year. These secular variation coefficients are used to extrapolate the Gaussian coefficients to the date in question. Once data become available about the actual magnetic field for a given epoch year, the model is adjusted and becomes the Definitive Geomagnetic Reference Field, or DGRF.

Practically the IGRF consists of Gauss' coefficients through degree and order 10 or slightly above as these terms are believed to represent the larger part of the field of the Earth's core. Subtracting these low-order terms from the measured magnetic fields provides in principle the magnetic field of the crust.

## 4. Similarities and differences with gravity methods

The gravity and magnetic survey methods exploit the fact that variations in the physical properties of rocks in-situ give rise to variations in some physical quantity which may be measured remotely (on are above the ground). In the case of gravity method, the physical rock property is density and so density variations at all depths within the Earth contribute to the broad spectrum of gravity anomalies. For the magnetic method, the rock property is magnetic susceptibility and/or remanent magnetization; both of which can only exist at temperatures cooler than the Curie point and thus restricting the sources of magnetic anomalies to the uppermost 30 – 40 km of the Earth’s interior. In practice, almost all magnetic properties of rocks in bulk reflect the properties and concentrations of oxides of iron and titanium (Fe and Ti): the Fe-Ti-O system, plus one sulphide mineral, pyrrhotite [1]. We also note that the highest density used typically in gravity surveys are about  $3.0 \text{ g cm}^{-3}$ , and the lowest densities are about  $1.0 \text{ g cm}^{-3}$ . Thus densities of rocks and soils vary very little from place to place. On the other hand, magnetic susceptibility can vary as much as four to five orders of magnitude from place to place, even within a given rock type.

Magnetic method	Gravity method
Passive and is a potential field bearing all the consequences	Passive and is a potential field bearing all the consequences
Mathematical expression for the force field is that of the inverse square law relation	Mathematical expression for the force field is that of the inverse square law relation
Force between monopoles can either be attractive or repulsive	Force between masses is always attractive
A monopole cannot be isolated. Monopoles always exist in pairs (dipole)	A single point mass can be isolated
A properly reduced field has variation due to variation in induced magnetization of susceptible rocks and remanent magnetization	A properly reduced field has variation due to density variation in rocks
Field changes significantly over time (secular variation).	Field does not change significantly over time.

**Table 1.** Other similarities and differences between gravity and magnetic methods

## 5. The magnetic properties of rocks

Geologic interpretation of magnetic data requires the knowledge of the magnetic properties of rocks in terms of magnetic susceptibility and remanent magnetization. Factors that influence rock magnetic properties for various rock types have been summarized appropriately [10], [1], [11], [12]. The rocks of the Earth’s crust are in general only weakly magnetic but can exhibit both induced and remanent magnetizations. Magnetic properties of rocks can only exist at

temperatures below the Curie point. The Curie temperature is found to vary within rocks but is often in the range 550°C to 600°C [13]. Modern research indicates that this temperature is probably reached by the normal geothermal gradient at depths between 30 and 40 km in the Earth and this so-called 'Curie point isotherm' may occur much closer to the Earth's surface in areas of high heat flow.

Indeed rock magnetism is a subject of considerable complexity. Clearly, all crustal rocks find themselves situated within the geomagnetic field described in section 3. These crustal rocks are therefore likely to display induced magnetization given by equation (3), where the magnitude of magnetization,  $J_i$  is proportional to the strength of the Earth's field,  $T$ . The magnetic susceptibility,  $k$  is actually the magnetic volume susceptibility that is encountered in exploration rather than mass or molar susceptibilities.

Apart from the induced magnetization, many rocks also show a natural remanent magnetization (NRM) that would remain even if the present-day geomagnetic field ceases to exist. The simplest way in which NRM can be acquired is through the process of cooling of rocks in molten state. As the rocks cool past the Curie point (or blocking temperature) a remanent magnetization in the direction of the prevailing geomagnetic field will be acquired. The magnitude and direction of the remanent magnetization can remain unchanged regardless of any subsequent changes in the ambient field.

## 6. Measurement procedures of magnetic field

Measurements can be made of the Earth's total magnetic field or of components of the field in various directions. The oldest magnetic prospecting instrument is the magnetic compass, which measures the field direction. Other instruments include magnetic balances and fluxgate magnetometers.

The most used instruments in modern magnetic surveys are the proton-precession or optical-pumping magnetometers and these are appreciably more accurate and all of these instruments give absolute values of field. The proton magnetometer measures a radio-frequency voltage induced in a coil by the reorientation (precession) of magnetically polarized protons in a container of ordinary water or paraffin. Its measurement sensitivity is about 1 nT. The optical-pumping magnetometer makes use of the principles of nuclear resonance and cesium or rubidium vapour. It can detect minute magnetic fluctuations by measuring the effects of light-induced (optically pumped) transitions between atomic energy levels that are dependent on the strength of the prevailing magnetic field. The sensitivity of the optical absorption magnetometer is about 0.01 nT and on this premise may be preferred to proton precession magnetometer in air-borne surveys.

Airborne magnetic surveys or aeromagnetic surveys are usually made with magnetometers carried by aircraft flying in parallel lines spaced 2 - 4 km apart at an elevation of about 500 m when exploring for petroleum deposits and in lines 0.5 - 1.0 km apart roughly 200 m or less above the ground when searching for mineral concentrations. Ship-borne magnetic surveys

or marine magnetic surveys can also be completed over water by towing a magnetometer behind a ship.

Ground surveys are conducted to follow up magnetic anomaly discoveries made from the air. Such surveys may involve stations spaced from 50 m apart. Survey may be along profiles or gridded network or may be in random pattern. Magnetometers also are towed by research vessels or mounted on the researcher on foot. In some cases, two or more magnetometers displaced a few metres from each other are used in a gradiometer arrangement; differences between their readings indicate the magnetic field gradient. A ground monitor is usually used to measure the natural fluctuations of the Earth's field over time so that corrections similar to drift correction in gravity can be made. Alternatively, like gravity observations where the temporal variation in field values were accounted for by reoccupying a base station and using the variation in this reading to account for instrument drift and temporal variations of the field, we could also use the same strategy in acquiring magnetic observations. The alternative is not the best as field variation in magnetic may be highly erratic and magnetometers which are electronic instruments do not drift. With these points in mind most investigators conduct magnetic surveys using two magnetometers. One is used to monitor temporal variations of the magnetic field continuously at a chosen "base station", and the other is used to collect observations related to the survey proper.

Surveying is generally suspended during periods of large magnetic fluctuation (magnetic storms).

## 7. Magnetic anomalies

Magnetic effects result primarily from the magnetization induced in susceptible rocks by the Earth's magnetic field. Most sedimentary rocks have very low susceptibility and thus are nearly transparent to magnetism. Accordingly, in petroleum exploration magnetics are used negatively: magnetic anomalies indicate the absence of explorable sedimentary rocks. Magnetics are used for mapping features in igneous and metamorphic rocks, possibly faults, dikes, or other features that are associated with mineral concentrations. Data are usually displayed in the form of a contour map of the magnetic field, but interpretation is often made on profiles.

The first stage in any ground magnetic survey is to check the magnetometers and the operators. Operators can be potent sources of magnetic noise. Errors can also occur when the sensor of the magnetometer is carried on a short pole or on a back rack. Compasses, pocket knives, metal keys, geological hammers, and cultural articles with metal blend (belt, shoes, bangles, etc) are all detectable at distances below about a metre and therefore the use of high-sensitivity magnetometers requires that operators divest themselves of all metallic objects. Attempts must be made to follow the operation manual provided along with a magnetometer!

Diurnal corrections are essential in most field work, unless only gradient data are to be used. If only a single magnetometer is available, diurnal corrections have to rely on repeated visits to a base, ideally at intervals of less than an hour. A more robust diurnal curve can be con-

structured if a second fixed magnetometer is used to obtain readings at 3 to 5 minute intervals. The second magnetometer need not be the same type as that being used in the field. Thus a proton magnetometer can provide adequate diurnal control for surveys conducted with cesium vapour magnetometer and vice versa. Note that base should be remote from magnetic interferences and must be describable for future use.

In aeromagnetic surveys, great pains must be taken to eliminate spurious magnetic signals that may be expected to arise from the aircraft itself. Airframes of modern aircraft are primarily constructed from aluminum alloys which are non-magnetic and so the potential magnetic sources are the aircraft engines. Thus magnetometer sensors must be mounted far away from these engines.

Aeromagnetic data usually obtained have gone through on-board processing such as magnetic compensation, checking/editing, diurnal removal, tie line and micro leveling. For example, the basis of magnetic compensation is the reduction of motion-induced noise on the selected magnetic elements. These can be from individual sensors or various gradient configurations. The motion noise comes from the complex three-dimensional magnetic signature of the airframe as it changes attitude with respect to the magnetic field vector. The noise comes from permanent, induced and eddy effects of the airframe plus additional heading effects of the individual sensors. Thus the magnetic interference in a geophysical aircraft environment comes from several sources which must be noted and compensated for.

On-board data checking and editing involves the removal of spurious noise and spikes from the data. Such noise can be caused by cultural influences such as power lines, metallic structures, radio transmissions, fences and various other factors. Diurnal removal corrects for temporal variation of the earth's main field. This is achieved by subtracting the time-synchronized signal, recorded at a stationary base magnetometer, from survey data. Alternatively, points of intersection of tie lines with traverse/profile lines can form loop networks which can be used to correct for the diurnal variation similar to drift correction in gravity survey.

Tie leveling utilizes the additional data recorded on tie lines to further adjust the data by consideration of the observation that, after the above corrections are made, data recorded at intersections (crossover points) of traverse and tie lines should be equal. Several techniques exist for making these adjustments and [14] gives a detailed of the commonly used techniques. The most significant cause of these errors is usually inadequate diurnal removal. Micro-leveling, on the other hand, is used to remove any errors remaining after the above adjustments are applied. These are usually very subtle errors caused by variations in terrain clearance or elevated diurnal activity. Such errors manifest themselves in the data as anomalies elongate in the traverse line direction. Accordingly they can be successfully removed with directional spatial filtering techniques [15].

When all the above considerations to raw magnetic data have been recognized and attended to, the IGRF correction (main field effect) is now carried out to give the 'magnetic anomaly' defined as the departure of the observed field from the global model.

## 8. Potential fields and models

### 8.1. Potential field in source free space

The potential field,  $\varphi(x, y, z)$  in free space, i.e. without any sources satisfies the Laplace equation  $\nabla^2\varphi = 0$  or when sources are present the potential satisfies the so-called Poisson equation,  $\nabla^2\varphi = -\rho(x, y, z)$ , where  $\rho(x, y, z)$  stands for density or magnetization depending upon whether  $\varphi$  stands for gravity or magnetic potential. Laplace's equation has certain very useful properties particularly in source-free space such as the atmosphere where most measurements are made. Some of these properties are (1) given the potential field over any plane, we can compute the field at almost all points in the space by analytical continuation; (2) the points where the field cannot be computed are the so-called singular points. A closed surface enclosing all such singular points also encloses the sources which give rise to the potential field. These properties of the potential field are best brought out in the Fourier domain.

The Fourier transform in one-dimension can be found in most text books of applied mathematics. The Fourier transform in two or three dimensions possess additional properties worth noting [16]. The two-dimensional Fourier transform is given by:  $\Phi(u, v) = \iint_{-\infty}^{+\infty} \varphi(x, y)e^{-i(ux+vy)} dx dy$  and its inverse is given by  $\varphi(x, y) = \frac{1}{4\pi^2} \iint_{-\infty}^{+\infty} \Phi(u, v)e^{i(ux+vy)} du dv$

Where  $u$  and  $v$  are spatial frequency numbers in the  $x$ - and  $y$ -directions respectively ( $u = 2\pi/L_x$  and  $v = 2\pi/L_y$ ), with  $L_x$  and  $L_y$  as length dimensions in the  $x$ - and  $y$ -directions respectively. It is important to note that  $\varphi(x, y)$  and  $\Phi(u, v)$  are simply different ways of looking at the same phenomenon. The Fourier transform maps a function from one domain (space or time) into another domain (wave number or frequency). For details, [17] can be consulted.

Magnetic potential field is caused by the variation in magnetization in the Earth's crust. This potential field is observed over a plane close to the surface of the Earth. If the magnetization variations are properly modeled consistent with other geological information, it is possible to fit the model to the observed potential field. Note that the magnetic field induction usually observed is the derivative of this potential. The model parameters (body shape factors, susceptibility values, burial depth, and magnetization direction) are then observable. These models may be (1) excess magnetization confined to a well-defined geometrical object, (2) geological entity such as basins (sedimentary or metamorphic with intrusive bodies). Sedimentary basins are of great interest on account of their hydrocarbon potential and since these rocks are generally non-magnetic, the observed magnetic field is probably entirely due to the basement on which sediments are resting and (3) with available resources and technology (as in airborne magnetic surveys) large areas can be covered in a survey and so permitting maps that cover several geological provinces or basins and therefore allow inter basin studies such as delineating of extensive shallow and deep features as faults, basin boundaries, etc. to be extracted.

We shall briefly outline a few examples of the rigors that an interpreter goes through to synthesize information from these potential fields.

## 8.2. Dipole field

We consider two monopoles of opposite sign: one at the origin of the 3-coordinate system and the other positioned below such that their common axis is along the z-axis, with  $-\Delta z$  as the separation between the monopoles (Fig. 2)

The potential at P,  $V(P)$  due to both monopoles is the sum of the potentials caused by each monopole. This is given generally for monopoles that are not aligned along any particular axis as [18]:

$$V(P) = -C_m \vec{m} \cdot \vec{\nabla}_P \left( \frac{1}{r} \right) \quad (9)$$

Where  $\vec{m}$  is the dipole moment ( $\vec{m} = q \vec{ds}$  with  $q$  as the pole strength and  $\vec{ds}$  is the displacement from monopole 1 to monopole 2.),  $C_m$  is the Coulomb's law constant ( $C_m = \mu_0/4\pi$ :  $\mu_0$  is the permeability of free space) and  $\vec{\nabla}_P$  is the derivative vector operator towards point P.

According to Helmholtz theorem, the magnetic field,  $\vec{B}$  can be derived from this potential,  $V(P)$  such that  $\vec{B} = -\vec{\nabla}_P V(P)$ . Using this on equation (9) yields

$$\vec{B} = C_m \frac{m}{r^3} [3(\hat{m} \cdot \hat{r})\hat{r} - \hat{m}], \quad r \neq 0 \quad (10)$$

Where  $m$  is the magnitude of the dipole moment while  $\hat{r}$  and  $\hat{m}$  are unit vectors in the increasing  $r$  and  $m$  directions respectively. Thus equation (10) shows that the magnitude of  $\vec{B}$  is proportional to the dipole moment and inversely proportional to the cube of the distance to the pole.

Equation (10) can also be expressed in cylindrical coordinates as [8]:

$$\vec{B} = C_m \frac{m}{r^3} (3\cos\theta \hat{r} - \hat{m}) \quad (11)$$

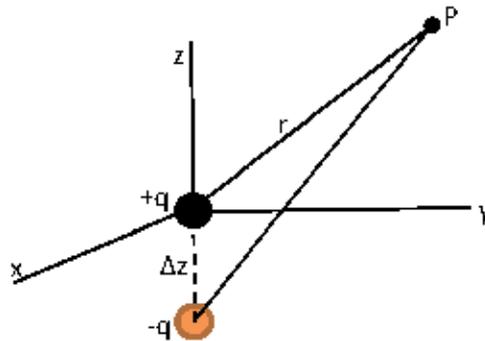
Where  $\theta$  is now the angle between  $\hat{m}$  and  $\hat{r}$  so that with  $\vec{B} = -\vec{\nabla}_P V$ , we can compute the components of the field in the directions of  $r$  and  $\theta$ , respectively as  $\vec{B}_r$  and  $\vec{B}_\theta$  expressed as

$$\vec{B}_r = -\hat{r} \frac{\partial V}{\partial r} = 2C_m \frac{m \cos\theta}{r^3}, \quad \vec{B}_\theta = -\hat{\theta} \frac{1}{r} \frac{\partial V}{\partial \theta} = C_m \frac{m \sin\theta}{r^3} \text{ and so the magnitude of } \vec{B} \text{ is expressed as}$$

$$|\vec{B}| = \sqrt{|\vec{B}_r|^2 + |\vec{B}_\theta|^2} = C_m \frac{m}{r^3} (3\cos^2\theta + 1)^{1/2} \quad (12)$$

Equation (12) shows that the magnitude  $|\vec{B}|$  of the dipole field along any direction extending from the dipole decreases at a rate inversely proportional to the cube of the distance to the dipole. The magnitude of  $\vec{B}$  also depends on  $\theta$ .

Many magnetic bodies exist that are dipolar in nature to a first approximation. For example, the entire field of the Earth appears nearly dipolar from the perspective of other planets. It is also known that in aeromagnetic survey, the inhomogeneity of a massive pluton at the survey height appears to be a dipole source.



**Figure 2.** Two monopoles of opposite sign with the monopole of positive sign at the origin and other situated at  $z = \Delta z$

### 8.3. Three-dimensional distribution of magnetization

We can consider a small element of magnetic material of volume,  $dv$  and magnetization,  $\vec{M}$  (Fig. 3) to act like a single dipole such that  $\vec{M} dv = \vec{m}$ . Then the potential at some point, P outside the body is given as in equation (9) to be



**Figure 3.** A 3-D magnetic body of volume  $dv$  and uniform magnetization  $\vec{M}$

$$V(P) = -C_m \vec{M} \cdot \vec{\nabla}_P \left( \frac{1}{r} \right) dv \quad (13)$$

Where  $r$  again is the distance from  $P$  to the dipole. In general, magnetization,  $\vec{M}$  is a function of position where both direction and magnitude can vary from point to point, i.e.  $\vec{M} = \vec{M}(Q)$ , where  $Q$  is the position of the volume element,  $dv$ . Integrating equation (13) over all elemental volumes provides the potential of a distribution of magnetization as

$$V(P) = C_m \int \vec{M}(Q) \cdot \vec{\nabla}_Q \left( \frac{1}{r} \right) dv \quad (14)$$

The magnetic induction,  $\vec{B}$  at P is then given by

$$\vec{B}(P) = - \vec{\nabla}_P V(P) = - C_m \vec{\nabla}_P \int \vec{M}(Q) \cdot \vec{\nabla}_Q \left( \frac{1}{r} \right) dv \quad (15)$$

The subscripts in the gradient operator from P to Q are now  $\vec{\nabla}_P$  and  $\vec{\nabla}_Q$  when the operator is inside the volume integral showing that the gradient is taken with respect to the source coordinates rather than with respect to the observation point.

For a two-dimensional source, we may start with a body of finite length  $2a$  and so the volume integral in equation (15) is replaced with surface integral over the cross sectional area,  $dS$  of the body and a line integral along its length (the  $z$ -axis) as:

$$V(P) = C_m \int \vec{M}(Q) \cdot \vec{\nabla}_Q \left( \frac{1}{r} \right) dv = C_m \int \vec{M}(Q) \cdot \left( \int_{-a}^a \vec{\nabla}_Q \left( \frac{1}{r} \right) dz \right) dS \quad (a) \quad (16)$$

$$V(P) = 2C_m \int \frac{\vec{M}(Q) \cdot \hat{r}}{r} dS \quad (b)$$

Where  $S$  is the cross sectional area of the body. As  $a \rightarrow \infty$  (a 2-D case), the integral approaches a potential of infinite line of dipoles of unit magnitude. Hence (16b).

The magnetic field induction,  $\vec{B}(P)$  can be obtained from equation (16b) as:

$$\vec{B}(P) = - \vec{\nabla} V(P) = 2C_m \int \frac{\vec{M}(Q)}{r^2} [2(\hat{M} \cdot \hat{r})\hat{r} - \hat{M}] dS \quad (17)$$

Note also that  $\hat{r}$  is the normal to the long axis of the cylinder and  $r$  is a perpendicular distance.

#### 8.4. Poisson relation

We had noted in section one, and by implication, that the mutual force,  $F$  between a particle of mass  $m_1$  centred at point  $Q(x', y', z')$  and a particle of mass,  $m_2$  at  $P(x, y, z)$  is given by

$$F = G \frac{m_1 m_2}{r^2} \quad (18)$$

Where  $r = [(x - x')^2 + (y - y')^2 + (z - z')^2]^{1/2}$ . If we let  $m_2$  be a test particle with unit magnitude, then the gravitational attraction,  $\vec{g}(P)$  produced by  $m_1$  at the location  $P$  of  $m_2$  (the test particle) is

$$\vec{g}(P) = - G \frac{m_1}{r^2} \hat{r} \quad (19)$$

Where  $\hat{r}$  is a unit vector directed from  $m_1$  to the observation point, P and expressed as  $\hat{r} = \frac{1}{r}[(x - x')\hat{i} + (y - y')\hat{j} + (z - z')\hat{k}]$  and the minus sign in equation (19) indicates that as  $r$  increases  $g(P)$  decreases in absolute value.

If a potential exists, then the gravitational attraction (also known as the gravitational acceleration),  $g(P)$  can be derived from this potential. Let this potential be  $U$ . Then at P,  $U = U(P)$ , such that

$$\vec{g}(P) = -\vec{\nabla} U(P) \tag{20}$$

Where here,  $U(P)$  can be expressed as

$$U(P) = G \frac{m_1}{r^2} \tag{21}$$

Where the gravitational potential is defined as the work done by the field on the test particle.

Equations (19) and (12) show that the magnetic scalar potential of an element of magnetic material and the gravitational attraction of mass are identical. Starting from equation (13), and considering a body with uniform magnetization,  $\vec{M}$  and uniform density,  $\rho$ , the magnetic scalar potential at a point, P is

$$V(P) = -C_m \int \vec{M} \cdot \vec{\nabla}_P \left(\frac{1}{r}\right) dv = -C_m \vec{M} \cdot \vec{\nabla}_P \int \frac{1}{r} dv \tag{22}$$

The gravitational potential (equation (22)) is written as

$U(P) = G \int \frac{\rho}{r} dv = G \rho \int \frac{1}{r} dv$ , so that  $\int \frac{1}{r} dv = \frac{U(P)}{G\rho}$  and substituting this in equation (21) gives

$$V(P) = -\frac{C_m}{G\rho} \vec{M} \cdot \vec{\nabla}_P U(P) = -\frac{C_m M}{G\rho} g_m \tag{23}$$

Where  $g_m$  is the component of gravity in the direction of magnetization and  $M$  is the magnitude of the magnetization.

Equation (23) is the so-called Poisson's relation and can be stated as follows: if the boundaries of a gravitational and magnetic source are the same and its magnetization and density distribution being uniform, then the magnetic potential is proportional to the component of gravitational attraction in the direction of magnetization, i.e.  $V(P) \propto M \cdot g(P)$ .

Poisson relation can be used to transform a magnetic anomaly into pseudo-gravity, the gravity anomaly that would be observed if the magnetization were replaced by a density distribution of exact proportions [19]. Pseudo-gravity transformation is a good aid in interpretation of magnetic data. In addition, Poisson's relation can be used to derive expressions for the magnetic induction of simple bodies when the expression for gravitation attraction is known.

### 8.5. Magnetic field over simple geometrical bodies

The form of magnetic anomaly from a given body is complex and generally depends on the following factors:

- i. The geometry of the body
- ii. The direction of the Earth’s field at a location of the body
- iii. The direction of polarization of the rocks forming the body
- iv. The orientation of the body with respect to the direction of the Earth’s field
- v. The orientation of the line of observation with respect to the axis of the body.

Thus the computations of models to account for magnetic anomalies are much more complex than those for gravity anomalies. As earlier stated, when the gravity expressions for simple geometrical bodies are given, we can use the Poisson’s relation to find the magnetic expressions over these models [18]. Table 2 is a summary of few of such computations.

Shape of Body	Gravity potential, U	Magnetic potential, V	Magnetic field, $\vec{B} = \nabla V$
Sphere of radius, a	$\frac{4}{3}\pi a^3 \frac{G\rho}{r}$	$C_m \frac{\vec{m} \cdot \hat{r}}{r^2}$	$C_m \frac{m}{r^3} [3(\hat{m} \cdot \hat{r})\hat{r} - \hat{m}]$
Horizontal cylinder of infinite length of cross sectional radius, r	$2\pi a^2 \rho G \log\left(\frac{1}{r}\right)$	$2C_m \pi a^2 \frac{\vec{M} \cdot \hat{r}}{r}$	$\frac{2C_m m'}{r^2} [2(\hat{m}' \cdot \hat{r})\hat{r} - \hat{m}']$
Horizontal slab of thickness, t	$2\pi \rho G t z$	$-2\pi C_m M t$	Zero

**Table 2.** Gravity and magnetic potentials caused by simple sources, along with magnetic induction for bodies of uniform density,  $\rho$  and magnetization,  $\vec{M}$  observed at some point, P (x, y, z) away from the centre of the body (other symbols are defined as in section 8.2)

## 9. Magnetic data processing and interpretation

The total-field magnetic anomaly of section 7 which was obtained by subtracting the magnitude of a suitable regional field (the IGRF or DGRF model for the survey date) from the total-field measurement is referred to as the crustal field. As earlier stated, this field is a vector sum of the remanent and the induced fields of the magnetically susceptible rocks of the crust down to the bottom of the Curie depth. The induced field component is usually in the same direction as the ambient field during the survey period.

Magnetic data processing includes everything done to the data between acquisition and the creation of an interpretable profile map or digital data set. Apart from the effect earlier discussed which are ignored or avoided, rather than corrected for, the correction required for ground magnetic data are insignificant especially when compared to gravity. The influence of

topography (terrain) on ground magnetics on the other hand can be significant. Magnetic terrain effects can severely mask the signatures of underlying sources as demonstrated by [20]. Many workers have attempted to remove or minimize magnetic terrain effects by using some form of filtering as summarized in [21].

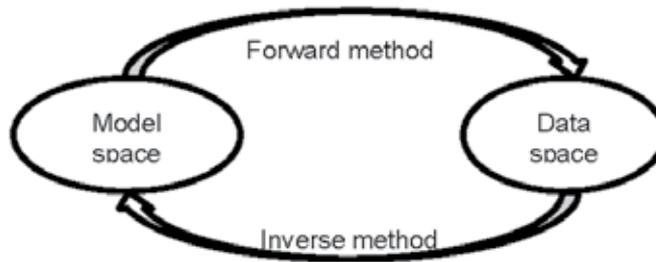
Interpretation of magnetic anomalies has to do with (a) studying the given magnetic map, profile or digital data to have a picture of the probable subsurface causes (qualitative interpretation), (b) separating the effect of individual features on the basis of available geophysical and geological data (further separation of broad-based or long-wavelength anomalies) and (c) estimating the likely parameters of the bodies of interest from the corresponding 'residual' or anomalies (quantitative interpretation).

The last two categories of interpretation procedures can be further broken into three parts. Each part has the goal of illuminating the spatial distribution of magnetic sources, but they approach the goal with quite different logical processes.

1. **Forward method:** an initial model for the source body is constructed based on geologic and geophysical intuition. The initial model anomaly is calculated and compared with the observed anomaly; and model parameters are adjusted in order to improve the fit between the two anomalies until the calculated and observed anomalies are deemed sufficiently alike. Forward method is source modeling in which magnetic anomalies were interpreted using characteristics curves [22] calculated from simple models (before the use of electronic computers) or using computer algorithms. Such schemes include 2-D magnetic models [23] and many workers that followed), 3-D magnetic models [24] and many other improvements that followed), Fourier-based models [25] and other improvements that followed) and voxel-based models [26] and others that followed).
2. **Inverse method:** one or more body parameters are calculated automatically and directly from the observed anomaly through some plausible assumptions of the form of the source body. Under this category, we have depth-to-source estimation techniques such as Werner deconvolution [27] and other workers that followed), frequency-domain techniques [28] and others that followed), Naudy matched filter based method [29] and others that followed), analytic signal method [30], [31] and others that followed), Euler deconvolution [32] and others that followed), source parameter imaging [33] and others that followed) and statistical methods [34] and others that followed). Physical property mapping under the inverse method include terracing [35] and susceptibility mapping [36] and others that followed). Other inversion techniques have to do with automated numerical procedures which construct models of subsurface geology from magnetic data and any prior information [37], [38], [39], [40], [41] among many others).
3. **Data enhancement and display:** model parameters are not calculated, but the anomaly is processed in some way in order to enhance certain characteristics of the source bodies, thereby facilitating the overall interpretation. This category involves all filter-based analyses such as reduction to pole (RTP) and pseudogravity [19] and others that followed), upward/downward continuation [42] and others that followed), derivative filters [43] and many others), matched filtering [44] and others that followed) and wavelet transforms

[45] and others that followed). Some of the enhancement techniques are artificial illumination [46], automatic gain control [47] and textural filtering [48]. Data displays can be in form of stacked profiles, contour maps, images and bipole plots.

In summary, any geophysical survey has two domains or spaces of interest: data space (data collected from the field) and model space (earth models) that account for the data space. The task is to establish a link between a data space and a model space (Fig.4).



**Figure 4.** Connecting link between data space and model space in forward and inverse method

The task of retrieving complete information about model parameters from a complete and precise set of data is inversion. Thus if we have a set of data collected from the field, we try to say about the earth model with those finite data set. How many different ways can one try to travel within a data space and a model space? What difficulties are encountered? How many different ways can one try to overcome those difficulties? How much information can one really gather and what are their limitations? What precautions should be taken as one moves from one space to another? Inverse theory seems to address these questions. Inverse method is a direct method in which source parameters are determined in a direct way from field (e.g. magnetic field) measurements.

Forward method on the other hand entails starting from model space (Fig. 4) by guessing initial model parameters and then calculating the model anomaly (in data space). The model anomaly is compared with observed (data) anomaly. If the match between the two is acceptable, the process stops, otherwise model parameters are adjusted and the process repeated. Forward method is an indirect method.

Various formulae exist for computing magnetic field of regular shapes such as the ones given in Table 2.

Analysis of magnetic data and their various enhancements via a suite of qualitative and quantitative methods as outlined above results in an interpretation of the subsurface geology. Qualitative interpretation relies on the spatial patterns that an interpreter can recognize in the data. Faults, dykes, lineaments and folds are usually identified. Intrusive bodies are often recognized by virtue of the shape and amplitude of their anomalies and so on. For example, detection of a fault in a magnetic map is an important exercise in mineral prospecting. Usually faults and related fractures serve as major channel-ways for the upward migration of ore-bearing fluids. Fault zones containing altered magnetic minerals can be detected from series

of closed lows on contour maps, by inflections or by lows on magnetic profiles or by displaced magnetic marker horizons.

Quantitative interpretation on the other hand is meant to compliment the qualitative method and seeks to provide useful estimates of the geometry, depth and magnetization of the magnetic sources. Broadly categorized as curve matching, forward modeling or inversion, quantitative techniques rely on the notion that simple geometric bodies, whose magnetic anomaly can be theoretically calculated, can adequately approximate magnetically more complex bodies. Geometric bodies such as ellipsoids, plates, rectangular prisms, polygonal prisms and thin sheets can all be calculated. Complex bodies can be built by superposing the effects of several simple bodies. Faults are often modeled using thin sheet model.

Like most other geophysical methods, magnetics is ambiguous to the extent that there are an infinite number of models that have the same magnetic anomaly. Acceptable models should be tested for geological plausibility.

## **10. Case study: Aeromagnetic field over the upper Benue trough, Nigeria**

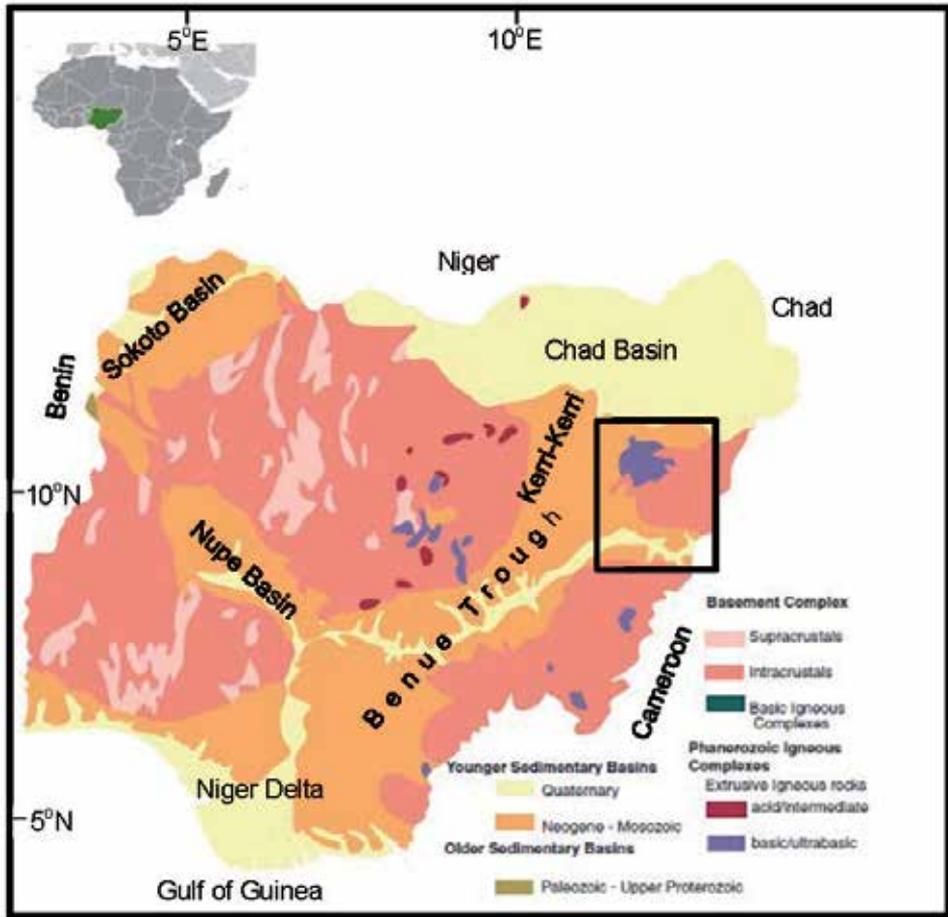
### **10.1. Geological framework**

The Upper Benue Trough, Nigeria (UBT) is the northern end of the nearly 1000 – km NE-SW trending sediment-filled Benue Trough, Nigeria (Fig. 5). Apart from the adjacent Niger Delta area and the offshore region where oil/gas exploration and production are taking place, the Benue Trough as an inland basin lacks the same full attention of the oil/gas companies. However, with the upbeat in the exploration efforts of the national government towards the search for hydrocarbon prospects of this inland basin, particularly in the light of new oil discoveries in the nearby genetically related basins, attention is directed at the structural setting of this basin.

The Benue Trough as a NE-SW trending sedimentary basin has a Y – shaped northern end (a near E – W trending branch of the Yola – Garoua and north-trending branch of Gongola – Muri) (Fig. 5). The Benue Trough is filled with sediments that range from Late Aptian to Palaeocene in age and whose thickness could reach up to 6000 m at places. The environments of deposition also varied over time such that the sediments vary from continental lacustrine/fluviatile sediments at the bottom through various marine transgressive and regressive beds, to immature reddish continental sands at the top.

For the past 50 years, the published works on the geology of the Benue Trough have substantially increased. The most important regional geological work on the Benue Trough by [49] was a basis for subsequent geological interpretations. [49] interpreted the Benue Trough origin in terms of rift faulting and the folding of the Cretaceous age associated with the basement flexuring. The first geophysical contribution of [50] on the Benue Trough remains to date the unique published reference. These authors have proposed the same rift origin considering that the main boundary rift faults are concealed by the Cretaceous sediments. They observed the existence below the Benue Trough axis, of a central positive gravity anomaly interpreted as a

basement high. Field evidence indicates that a set of deep-seated faults is superimposed on the axial high and controlled the tectonic evolution of the trough.



**Figure 5.** Top is map of Africa showing Nigeria. Below is a simplified geological map of Nigeria (modified from <http://nigeriaminers.org>). The inset rectangle is the Upper Benue Trough, Nigeria

The rift origin of the Benue Trough supported by numerous authors was interpreted in the plate tectonics concept and from the 1970s several models were proposed to explain the origin of the Benue Trough. For example, seen as a direct consequence of the Atlantic Ocean opening, the Benue Trough was considered to be the third arm of a triple junction located beneath the centre of the present Niger Delta [51] and proposed a Ridge-Ridge-Ridge (RRR) triple junction. This hypothesis has been widely discussed and replaced in the general framework of African Rift System.

The Benue Trough is subdivided into 3 units on the basis of stratigraphic and tectonic considerations. The southern ensemble is called the Lower Benue Trough (LBT) and includes two

main units: the Abakaliki Uplift or Anticlinorium and the Anambra Basin. The Abakaliki Anticlinorium (AA) is formed by tightly folded Cretaceous sediments intruded by numerous magmatic rocks. From the Niger Delta, AA extends for about 250 km to the Gboko – Ogoja area in a N50E direction. To the north of AA is a vast synclinal structure called the Anambra Basin and trends in a N30E direction. This basin comprises a thick undeformed Cretaceous series. On the northern margin of the Anambra Basin, is the Nupe or Middle Niger Basin which stretches along a NW-SE direction. To the south, AA is flanked by the Afikpo Syncline and by a narrow strip of thin, undeformed sediments resting on the Basement Complex (the Mamfe Basin) and to the northwestern border is the Oban Massif. South of the Oban Massif is the Calabar Flank, which belongs to the coastal basins of the Gulf of Guinea.

The Upper Benue Trough (UBT) which is the northern ensemble is the most complex part (Fig. 5). It is characterized by cover tectonics and can be further subdivided into several smaller units. The Gongola – Muri and Yola – Garoua branches are digitations of the Benue Trough and present a similar tectonic style. The Gongola - Muri rift disappears beneath the Tertiary sediments of the Chad Basin and so the margins of this rift are geologically the most difficult to established. The Yola – Garoua rift to some extent is the least known of the West African Rift and strikes E-W into Cameroon. On the western margin of the UBT is the flat-lying Paleocene Kerri Kerri basin resting unconformably upon the folded Cretaceous. The development of the Kerri Kerri basin is said to be controlled by a set of faults between it and the Basement Complex of the Jos Plateau [52]. The basin formation and its tectonic activity seem to be a response of the general uplift of the UBT due to late Cretaceous folding.

The UBT is contiguous with the Nigerian sector of the Chad Basin which extends northwards into the Termit Basin of Chad and Niger and southwestwards into Cameroon and southern Chad as Bongor, Doba, Doseo and Salamat basins. This rift system is closely linked with oil-rich Muglad Basin of Sudan.

## 10.2. Aeromagnetic field

Magnetic data over Nigeria have been largely collected above the ground surface in form of systematic surveys on behalf of the national government. These airborne surveys were carried out principally by a consultant, namely: **Fugro Airborne Surveys**, on behalf of the Nigerian Geological Survey Agency (NGSA) between 2003 and 2009. The main aim of these surveys has been to assist in mineral and groundwater development through improved geological mapping. Flight line direction is nearly NW-SE and tie line direction is NE – SW. The flight height average is 100 m; profile line spacing is 500 m with tie line spacing of 2 km. Figure 6 shows one of the aircrafts of Fugro Airborne Surveys in flight.

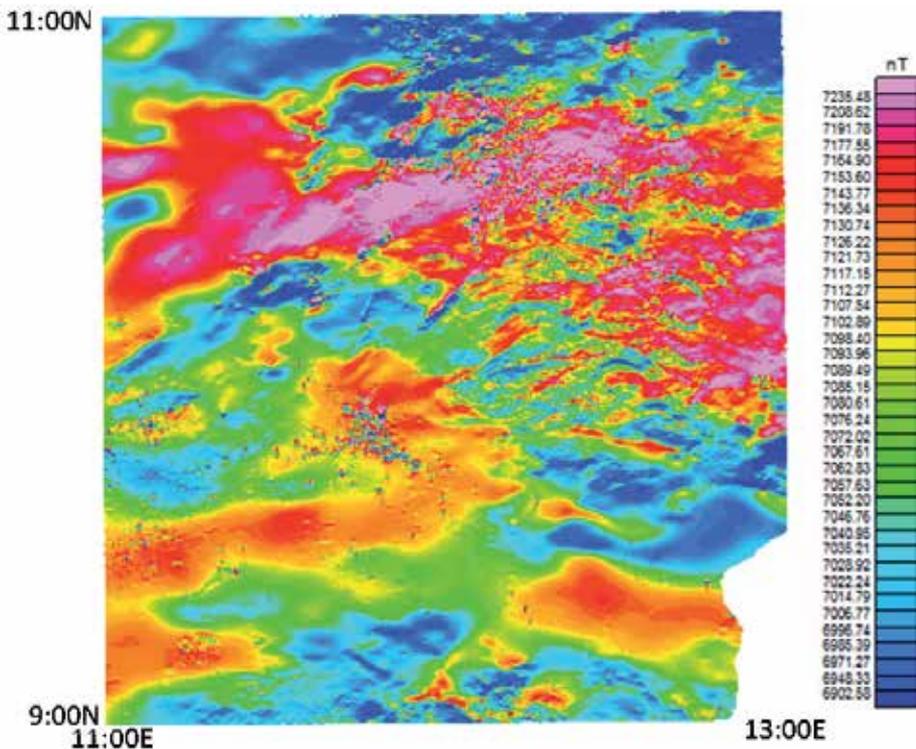
The total field aeromagnetic field intensity for the UBT comprises 16 half-degree grids acquired from NGSA and is used for the purpose of the present study. These are 131\_BAJOGA, 132\_GULANI, 133\_BIU, 134\_CHIBUK; 152\_GOMBE, 153\_WUYO, 154\_SHANI, 155\_GARKIDA; 173\_KALTUNGO, 174\_GUYOK, 175\_SHELLEN, 176\_ZUMO; 194\_LAU, 195\_DONG, 196\_NUMAN and 197\_GIREI total magnetic intensity (TMI) grids.



**Figure 6.** Fugro Airborne Surveys photo showing a magnetometer in a ‘stinger’ behind the fixed-wing aircraft.

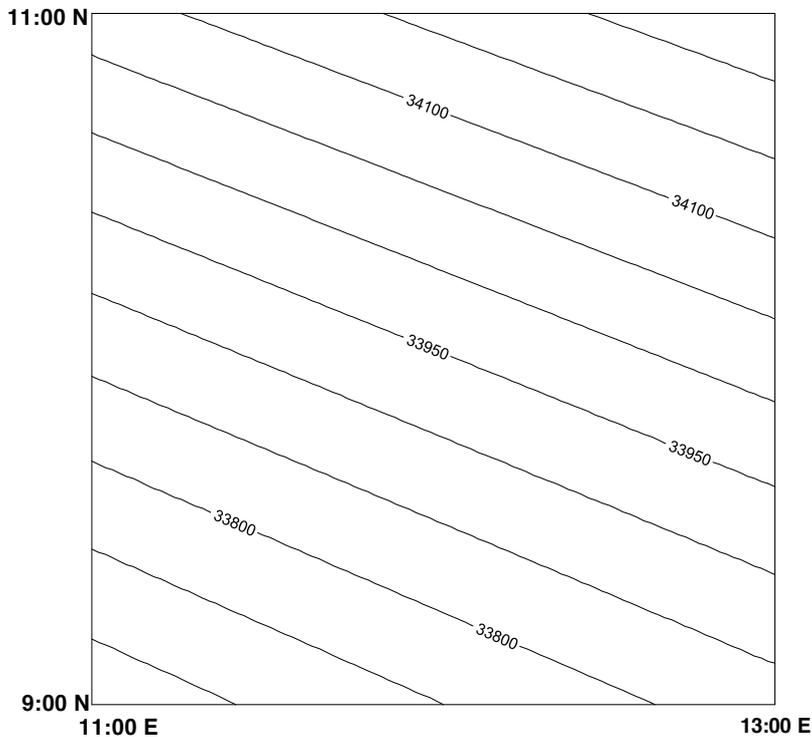
The aeromagnetic data obtained have gone through on-board processing such as magnetic compensation, checking/editing, diurnal removal, tie line and micro leveling.

The composite total field aeromagnetic data for the UBT are displayed in image (Fig. 7). The advantage of images is that they are capable of showing extremely subtle features not apparent in other forms of presentations (such as contour maps). They can also be quickly manipulated in digital form, thereby providing an ideal basis for on-screen GIS-based applications.



**Figure 7.** The total-field aeromagnetic intensity over UBT. A base value of 26, 000 nT should be added to map value for the total-field

We have further treated the composite total field aeromagnetic data (Fig. 7) for the UBT for the main field effect through the removal of the Definitive Geomagnetic Reference Field (DGRF – 2005) (Fig. 8) resulting in total magnetic field intensity anomaly (Fig. 9). This anomaly field has polarity signs that shows the impact of low geomagnetic inclination values for the study area and also reflects (1) the **induced field** caused by magnetic induction in magnetically susceptible earth materials polarized by the main field and (2) the field caused by **remanent magnetism** of earth materials. We call these two fields, the crustal field and have used the appropriate software (Geosoft Oasis Montaj version 8.0) for image processing and/or display of both the raw data (Fig. 7) and the anomaly data (Fig. 9). We have noted that a NW-SE mega feature dominated the middle of the study area. This linear feature, which is interrupted somehow towards the NW section of the map area, is believed to be central in the structural configuration and set-up of this Benue Trough subarea.



**Figure 8.** Definitive Geomagnetic Reference Model (DGRF2005) over UBT at 100 m above ground level. The Earth’s model used is geodetic and CI is 50 nT.

### 10.3. Analysis of aeromagnetic field

We have computed for the present study area (UBT) the field,  $F$  for the epoch year 2005 and is displayed in contour form (Fig. 8) and because of the relatively small size of the study area, the values of  $D$  and  $I$  cannot be contoured and imaged. However between these values of

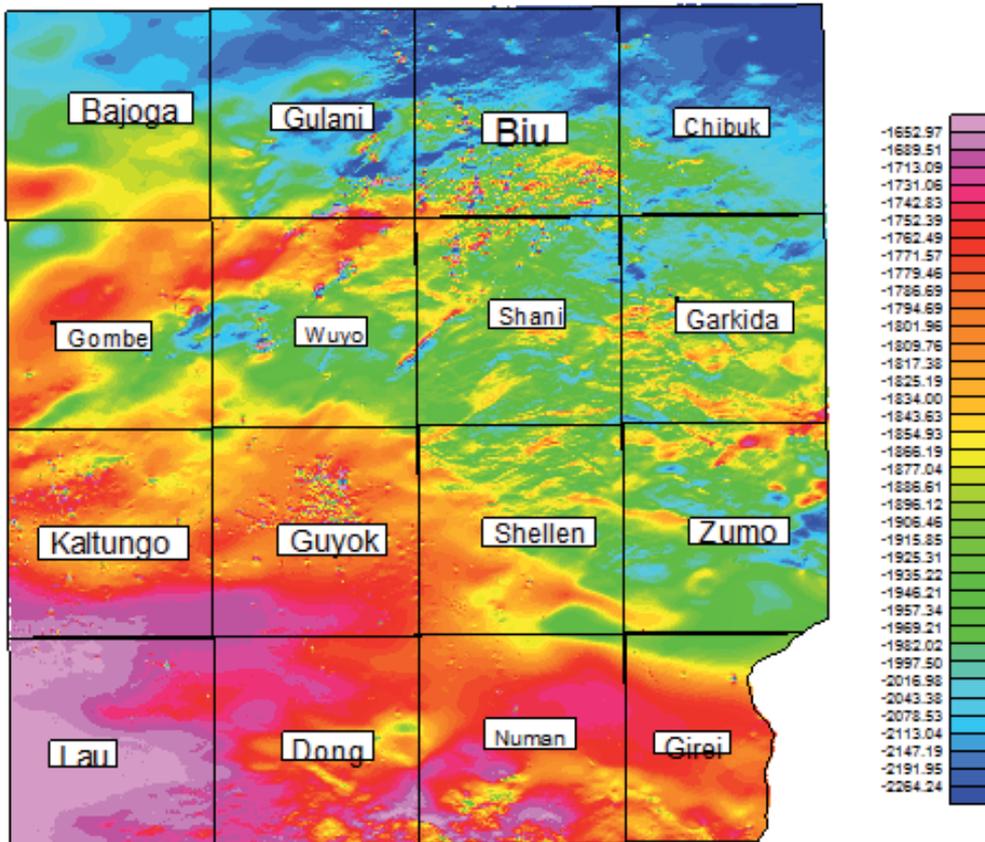
Longitudes and Latitudes ( $11^{\circ}\text{E} - 13^{\circ}\text{E}$ ,  $9^{\circ}\text{N} - 11^{\circ}\text{N}$ ),  $D$  ranges from  $-1.4^{\circ}$  ( $11^{\circ}\text{E}$ ,  $9^{\circ}\text{N}$ ) to  $-0.7^{\circ}$  ( $13^{\circ}\text{E}$ ,  $11^{\circ}\text{N}$ ) and  $I$  ranges from  $-5.7^{\circ}$  ( $11^{\circ}\text{E}$ ,  $9^{\circ}\text{N}$ ) to  $-0.2^{\circ}$  ( $13^{\circ}\text{E}$ ,  $11^{\circ}\text{N}$ ).

The Definitive Geomagnetic Reference Field (DGRF) or the main field map of the study area (Fig. 8) shows a NW-SE trending lines that have increasing values from the SW portion (minimum value of 33688 nT) to the NE portion (maximum value of 34271 nT) having an average value of 33974 nT and standard deviation of 129 nT. The inclination of the field for this epoch period (2005) decreases correspondingly from  $-5.7^{\circ}$  to  $-0.2^{\circ}$  indicating that slightly north of  $11^{\circ}$  latitude, the inclination of  $0^{\circ}$  or magnetic equator passes. We are therefore dealing with a low magnetic latitude area. Similarly the geomagnetic declination varies correspondingly from  $-1.4^{\circ}$  to  $-0.7^{\circ}$  which also shows that further north of  $11^{\circ}$  latitude, the declination would be  $0^{\circ}$ , indicating that geomagnetic and geographic meridians coincide. Computations of the rate of change of declination,  $D$  (in minutes per year) shows a constant value of 6 minutes/year, rate of change of inclination,  $I$  shows a northward increase from  $-4$  to  $-3$  minutes/year and also a northward increase of secular variation in the total field,  $F$  of between 21 – 24 nT/year. This shows that beginning 2010, Nigeria will be completely in the southern magnetic hemisphere in the next 40 years, where then the  $0^{\circ}$  latitude or magnetic equator will be passing through Niger Republic.

The image display of the aeromagnetic total field anomaly map (Fig. 9) has negative anomaly values. This is not surprising and in fact it is expected. The study area and generally Nigeria is situated in a magnetically low-latitude area. The polarizing field of the Earth in such areas is the horizontal component,  $H$ . Note that the structure of the Earth's magnetic field resembles that of a bar magnet. At the magnetic poles, the field is essentially vertical ( $Z$ ), at the centre of the magnetic bar the field is horizontal ( $H$ ). The horizontal component,  $H$  of the total field,  $F$  around or at the magnetic equator is therefore the polarizing field. Any magnetically susceptible (non-zero susceptibility) earth materials within this area will be magnetized or polarized by  $H$ . When  $H$  field induces a polarization field in a susceptible material, the orientation of the field lines describing the magnetic field is rotated  $90^{\circ}$ . Above this susceptible earth material, the polarization field now points in the opposite direction as the Earth's main field. Thus the total field measured will be less than the Earth's main field, and so upon removal of the main field, the resulting anomalous field will be negative. This is not the case in high-latitude areas, for the same susceptible earth material, where the anomalous field over such would be largely positive and/or negative where also the rotation of the polarizing field depends on the value of inclination,  $I$ .

The anomaly map in Figure 9 can be broadly characterized into at least 4 colour zones with the following grid values:  $-2264$  to  $-982$  nT,  $-982$  to  $-877$  nT,  $-877$  to  $-731$  nT and  $-731$  to  $-653$  nT running from the NE edge and terminating to the SW side. There appears to be a shear zone running NW-SE nearly bisecting the area and passing through Girei, Shellen, Wuyo and disappearing or being interrupted by Gombe grid probably by reason of an offset NE-SW feature occupying the middle of Gombe grid and pinching out on Biu grid (Fig. 9). The Biu basalts and other high-susceptibility rocks around must have been very influential in the recorded low magnetic anomaly values at NE portion of the map area, particularly towards the northern edges of Bajoga, Gulani, Biu and Chibuk grids.

To demonstrate the usefulness of digital tools in the analysis of magnetic data, we shall apply only one digital processing tool to the analysis of the aeromagnetic total-field anomaly over UBT. We shall use the analytic signal technique.



**Figure 9.** The total-field aeromagnetic anomaly over UBT. A value of 1000 nT should be added to map values.

The analytic signal for magnetic anomalies was initially defined as a ‘complex field deriving from a complex potential’ [30]. A 3-D analytic signal  $\vec{A}$  [43], [53] of a potential field anomaly,  $M$  (magnetic field or vertical gradient of gravity), can be defined as:

$$\vec{A}(x, y) = \frac{\partial M}{\partial x} \hat{x} + \frac{\partial M}{\partial y} \hat{y} + \frac{\partial M}{\partial z} \hat{z} \quad (24)$$

Where  $\hat{x}$ ,  $\hat{y}$ ,  $\hat{z}$  are unit vectors in the x-, y- and z-axis directions. The analytic signal amplitude or its absolute value can be expressed by a vector addition of the two real components in the x and y directions and of the imaginary component in the z-direction, i.e.

$$|\vec{A}(x, y)| = \sqrt{\left(\frac{\partial M}{\partial x}\right)^2 + \left(\frac{\partial M}{\partial y}\right)^2 + \left(\frac{\partial M}{\partial z}\right)^2} \quad (25)$$

The field and the analytic signal derivatives are more easily derived in the wave number domain. If  $F(M)$  is the Fourier transform of  $M$  in the 2-D wave number domain with wave number  $k = (k_x, k_y)$ , the horizontal and vertical derivatives of  $M$  correspond respectively to multiplication of  $F(M)$  by  $i(k_x, k_y) = ik$  and  $|k|$ . In 3-D the gradient operator in frequency domain is given by  $\vec{\nabla} = ik_x \hat{x} + ik_y \hat{y} + |k| \hat{z}$ . The Fourier transform of the analytic signal can then be expressed in terms of the gradient of the Fourier transform of the field  $M$  by the following equation equivalent to the space-domain equation above, i.e. [53]:

$$\hat{t} \cdot F(\vec{A}(x, y)) = \hat{h} \cdot \vec{\nabla} F(M) + i \hat{z} \nabla F(M) \quad (26)$$

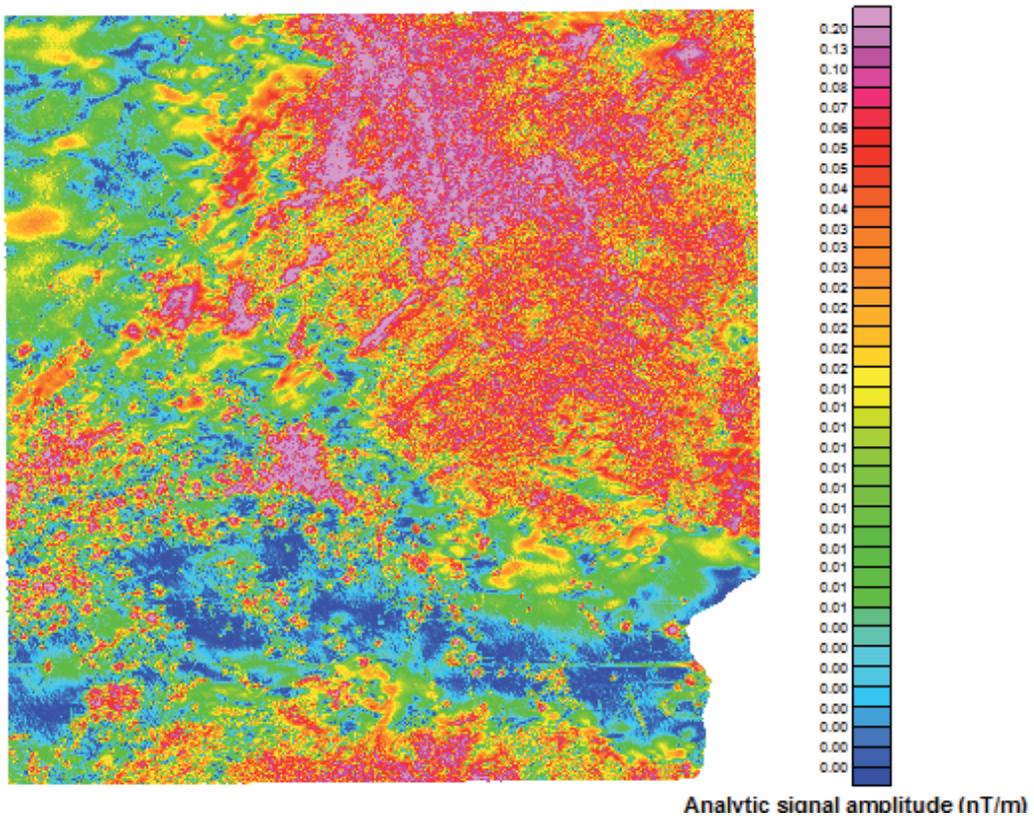
Where  $\hat{h} = \hat{x} + \hat{y}$  is the horizontal unit vector and  $\hat{t} = \hat{h} + \hat{z}$ .

The amplitude of the 3-D analytic signal of the total magnetic field anomaly produces maxima over magnetic contacts regardless of the direction of magnetization. The absence of magnetization direction in the shape of analytic signal anomalies is a particularly attractive characteristic for the interpretation of magnetic field near the magnetic latitude like the area under test. It is also known that the depths to sources can be determined from the distance between inflection points of analytic signal anomalies, but have not explored that option and interested readers can refer to [54].

In this method, we have applied the concept of analytic signal to the residual total magnetic field intensity of the UBT. These processes were accomplished by use of Geosoft Oasis Montaj (version 8.0).

Figure 10 shows the output of the analytical signal amplitude calculated from the original total-field magnetic anomaly (Fig. 9). Analytic signal of the total-field magnetic anomaly reduces magnetic data to anomalies whose maxima mark the edges of magnetized bodies and whose shape can be used to determine the depth of these edges (we have not done this second aspect).

The analytic signal amplitude over the UBT ranges from 0.00 nT/m to 7.93 nT/m: having a mean of 0.036 nT/m and standard deviation of 0.073 nT/m. Since amplitude of the analytic signal anomalies combines all vector components of the field into a simple constant, a good way to think of analytic signal is as a map of magnetization in the ground. With this in mind, we can picture strong anomalies to exist over where the magnetization vector intersects magnetic contrasts, even though one cannot know the source of the contrasts from the signal amplitude alone. Consequently, we can easily see the boundaries of the Biu basalts properly demarcated (Fig. 10) shown by the higher analytic signal values. Note also the few scattered imprints of the same basalts tailing to the SW direction from this major anomaly. The Biu area is composed of Tertiary and Quaternary periods (less than 65 Ma ago) basaltic lava flows containing abundant peridotite xenoliths.



**Figure 10.** Output of the analytic signal amplitude over UBT. The boundary of high amplitude anomaly over the Biu area (basaltic areas) are delineated

## 11. Conclusion

In this chapter we have explored the magnetic method for economic exploration of the Earth. The strength of the method lies in the adequate distribution of magnetization within the crustal materials of the Earth in the light of measurable magnetic field over them.

The Earth's magnetic field, that is central in the remanence and induced processes, is itself complex. Spherical harmonic analysis provides the means of characterizing the Earth's magnetic field and with such a representation; it is possible to predict the geomagnetic dipolar field and other non-dipolar components. The knowledge of the dipolar field of the Earth enables the magnetic anomaly to be determined over a survey area from measurements of the magnetic field induction.

We have applied the magnetic method to real field measurements of total-field aeromagnetic intensity data over the Upper Benue Trough, Nigeria. The working data were corrected for secular variation using the existing DGRF model. The anomaly field which is the summary of the crustal field was further processed to obtain the amplitude of analytic signal of the anomaly

field. The analytic signal transformations combine derivative calculations to produce an attribute that is independent of the main inclination and direction of magnetization as well as having peaks over the edges of wide bodies. Thus a simple relationship between the geometry of the causative bodies and the transformed data is observed such as seen in Figure 10. We note that the borders of the Biu basalt as exposed by the applied analytic signal technique of the magnetic anomaly data goes beyond the outcropping boundary that may be offered by remotely sensed data. We recognize that even though the magnetic data were remotely sensed, the result from such measurements goes beyond what the traditional remote sensing information can offer.

The greatest limitation of the magnetic method is the fact that it only responds to variations in the magnetic properties of the earth materials and so many other characteristics of the subsurface (e.g. regolith characteristics) are not resolvable. The inherent ambiguity in magnetic interpretation (for quantitative techniques) is problematic where several geologically plausible models can be attained from the data. Interpreters of magnetic data must therefore be aware of such limitations and be prepared to obtain confirmatory facts from other databases to decrease the level of ambiguity.

## Acknowledgements

I acknowledge the National Centre for Petroleum Research and Development (NCPRD), of the Energy Commission of Nigeria (ECN), Abubakar Tafawa Balewa University, Bauchi, Nigeria for supporting this research.

## Author details

Othniel K. Likkason

Physics Programme, Abubakar Tafawa Balewa University, Bauchi, Nigeria

## References

- [1] Grant, F. S. Aeromagnetism, geology and environments I, magnetite in igneous, sedimentary and metamorphic rocks: an overview. *Geoexploration* 1985; 23 303 - 333
- [2] Campbell, W. C. Introduction to geomagnetic fields: Cambridge University Press; 1997
- [3] Cox, A. Plate tectonics and geomagnetic reversals. W. H. Freeman and Company; 1973

- [4] McElhinny, W. M. Paleomagnetism and plate tectonics: Cambridge University Press; 1973
- [5] McElhinny, W. M., McFadden, P. L. Paleomagnetism – continents and oceans: Academic Press; 2000
- [6] Merrill, R. T., McElhinny, M. W., McFadden, P. L. Magnetic field of the earth; paleomagnetism, the core and the deep mantle: Academic Press, San Diego;1996
- [7] Doell, R., Cox, A. Magnetization of rocks, In: Mining Geophysics Volume 2: Theory, Society of Exploration Geophysicists SEG Mining Geophysics Volume Editorial Committee (Eds.):1967 446-453, Tulsa
- [8] Telford, W. M., Geldart, L. P., Sheriff, R. E., Keys, D. A. Applied geophysics. Cambridge University Press, London:1990
- [9] Chapman, S., Bartels, J. Geomagnetism. Oxford University Press:1940
- [10] Haggerty, S. E. The aeromagnetic mineralogy of igneous rocks. Canadian Journal of Earth Sciences 1979: 16 1281-1293
- [11] Reynolds, R. L., Rosenbaum, J. G., Hudson, M. R., Fishman, N. S. Rock magnetism, the distribution of magnetic minerals in Earth's crust and aeromagnetic anomalies. U. S. Geological Survey Bulletin 1990; 1924.24-45.
- [12] Clark, D. A. Magnetic petrophysics and magnetic petrology: aids to geological interpretation of magnetic surveys. AGSO Journal of Australian Geology and Geophysics 1997;17.83-103
- [13] Clark, D. A., Emerson, D. W. Notes on rock magnetization characteristics in applied geophysical studies. Exploration Geophysics 1991; 22.547-555.
- [14] Luyendyk, A. P. J. Processing of airborne magnetic data. AGSO J. Australian Geol. Geophys. 1998;17 31-38
- [15] Minty, B. R. S. Simple micro-leveling for aeromagnetic data. Exploration Geophysics 1991;22.591-592.
- [16] Kay, S. M. Fundamentals of statistical signal processing. Prentice Hall, Englewood Cliffs, NJ.; 1993
- [17] Naidu, P. S., Mathew, M. P. Analysis of geophysical potential fields. Elsevier Science Publishers, Netherlands; 1998
- [18] Blakely, R. J. Potential theory in gravity and magnetic applications. Cambridge University Press; 1996
- [19] Baranov, V. A new method for interpretation of aeromagnetic maps: pseudo-gravimetric anomalies. Geophysics 1957;.22.359-383

- [20] Grauch, V. J. S., Cordell, L. Limitations on determining density or magnetic boundaries from horizontal gradient of gravity or pseudogravity data. *Geophysics* 1987: 52.118-121
- [21] Grauch, V. J. S. A new variable magnetization terrain correction method for aeromagnetic data. *Geophysics* 1987: 52 94-107.
- [22] Nettleton, L. L. Gravity and magnetic calculations. *Geophysics* 1942:7.293-310
- [23] Talwani, M., Heirtzler, J. M. Computation of magnetic anomalies caused by two-dimensional structures of arbitrary shape, In: *Computers in mineral industries*, G. A. Parks (Ed.) 464-480, Stanford Univ.: 1964
- [24] Talwani, M. Computation with help of a digital computer of magnetic anomalies caused by bodies of arbitrary shape. *Geophysics* 165: 30 797-817
- [25] Bhattacharyya, B. K. Continuous spectrum of the total magnetic field anomaly due to a rectangular prismatic body, *Geophysics* 1966: 31 97-121
- [26] Hjelt, S. Magnetostatic anomalies of dipping prisms. *Geoexploration* 1972:10, 239-254
- [27] Werner, S. Interpretation of magnetic anomalies at sheet-like bodies. *Sveriges Geol. Undersok. Ser. C, Arsbok*, 1953:43(6)
- [28] O'Brien, D. P. CompuDepth: a new method for depth-to-basement computation. Presented at the 42<sup>nd</sup> Annual International Meeting 1972, SEG.
- [29] Naudi, H. Automatic determination of depth on aeromagnetic profiles. *Geophysics* 1971: 36 717-722
- [30] Nabighian, M. N. The analytic signal of two-dimensional magnetic bodies with polygonal cross-section: its properties and use for automated anomaly interpretation. *Geophysics* 1972:37 507-517
- [31] Nabighian, M. N. Additional comments on the analytic signal of two-dimensional magnetic bodies with polygonal cross-section. *Geophysics* 1974: 39 85-92
- [32] Thompson, D. T. EULDPH – a new technique for making computer-assisted depth estimates from magnetic data. *Geophysics* 1982: 47 31-37.
- [33] Thurston, J. B., Smith, R. S. Automatic conversion of magnetic data to depth, dip and susceptibility contrast using the SPI™ method. *Geophysics* 1997: 62.807-813
- [34] Spector, A., Grant, F. S. Statistical models for interpreting aeromagnetic data. *Geophysics* 1970: 35 293-302
- [35] Cordell, L., McCafferty, A. E. A terracing operator for physical property mapping with potential field data. *Geophysics* 1989: 54 621-634.
- [36] Grant, F. S. The magnetic susceptibility mapping method for interpreting aeromagnetic survey:43<sup>rd</sup> Annual International Meeting, SEG 1973 expanded abstract 1201.

- [37] Pedersen, L. B. Interpretation of potential field data – a generalized inverse approach. *Geophysical Prospecting*, 1977: 25 199-230.
- [38] Pilkington, M. & Crossley, D. J. Inversion of aeromagnetic data for multilayered crustal models. *Geophysics* 1986: 51 2250-2254.
- [39] Pustisek, A. M. Noniterative three-dimensional inversion of magnetic data. *Geophysics* 1990:55 782-785.
- [40] Li, Y. & Oldenburg, D. W. 3-D inversion of magnetic data. *Geophysics* 1996:61 394-408.
- [41] Shearer, S. & Li, Y. 3 D inversion of magnetic total gradient in the presence of remanent magnetization. 74<sup>th</sup> Annual International Meeting, SEG 2004: Abstracts 774-777
- [42] Kellogg, O. D. Foundations of potential field theory. Dover New York; 1953
- [43] Nabighian, M. N. Toward a three dimensional automatic interpretation of potential field data via generalized Hilbert transforms: fundamental relation. *Geophysics* 1984:49 957-966
- [44] Spector, A. Spectral analysis of aeromagnetic data. Ph.D thesis, University of Toronto; 1968
- [45] Moreau, F. D. G., Holschneider, M., Saracco, G. Wavelet analysis of potential fields. *Inverse Problems* 1997:13 165-178
- [46] Pelton, C. A computer program for hill-shading digital topographic data sets. *Computers and Geosciences* 1987:13 545-548.
- [47] Rajagopalan, S., Milligan, P. Image enhancement of aeromagnetic data using automatic gain control. *Exploration Geophysics* 1995: 25 173-178.
- [48] Dentith, M., Cowan, D. R., Tompkins, L. A. Enhancement of subtle features in aeromagnetic data. *Exploration Geophysics* 2000:31 104-108
- [49] Carter, J. D., Barber, W., Tait, E. A., Jones, G. P. The geology of parts of Adamawa, Bauchi and Bornu provinces in northeastern Nigeria. *Bull. Geol. Survey Nigeria* 1963:30 p108
- [50] Cratchley, C. R. & Jones, G. P. An interpretation of the geology and gravity anomalies of the Benue Valley, Nigeria, *Overseas Geological Surveys, Geophysical Paper* 1965 (1)
- [51] Burke, K., Dessauvagie, T. F. J., Whiteman, A. J. Geological history of the Benue Valley and adjacent areas, In: *African Geology*, T. F. J. Dessauvagie & A. J. Whiteman (Eds.) University of Ibadan Press, Nigeria; 1970
- [52] Benkhelil, M. J. The origin and evolution of the Cretaceous Benue Trough, Nigeria, *Journal of African Earth Sciences* 1989: 8 251-282

- [53] Roest, W. R., Verhoef, J., Pilkington, M. Magnetic interpretation using 3-D analytic signal. *Geophysics* 1992: 57 116-125
- [54] Debeglia, N. & Coperl, J. Automatic 3-D interpretation of potential field data using analytic signal derivatives. *Geophysics* 1997:62 87-96

---

# Identification of Communication Channels for Remote Sensing Systems Using Volterra Model in Frequency Domain

---

Vitaliy Pavlenko and Viktor Speranskyy

Additional information is available at the end of the chapter

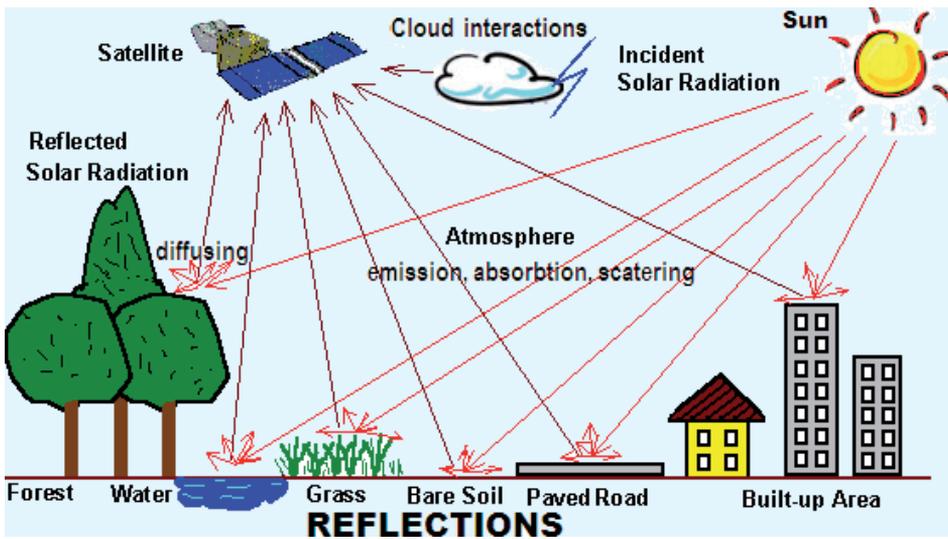
<http://dx.doi.org/10.5772/58354>

---

## 1. Introduction

The sensors in remote sensing systems are looking through a layer of atmosphere separating the sensors from the Earth's surface being observed. It is essential to understand the effects of atmosphere on the electromagnetic radiation travelling from the Earth to the sensor through the atmosphere. The atmospheric constituents cause wavelength dependent absorption and scattering of radiation due to environment interactions, emissions and so on (fig. 1) [6]. The atmosphere between radiating surface and sensor can be understood as communication channel (CC). The technical conditions of CC during operation should be considered for the effective communications. Changes during data transfer can decrease the rate of data transmission in digital CC up to stop of transmission. In analog CC it can be cause distortions and noise of the transmitted signals. Some of the atmospheric effects can be corrected before the sensing data is subjected to further analysis and interpretation. These effects degrade the adequateness of received data. The new methods and supporting tools are developed to automate the measurement and consideration of the characteristics of the CC. It helps to build the information and mathematical models of nonlinear dynamic object such as the CC [3, 19, 20], i.e. to solve the identification problem.

Building Volterra models and using them for visualization for such complex nature effects as waves of sea surface were well studied in [8-10]. This methodic allows building linear and nonlinear models for different systems. Modern continuous CCs are nonlinear stochastic inertial systems. The model in the form of integral Volterra series used to identify them [3, 4]. The nonlinear and dynamic properties of such system are completely characterized by a sequence of multidimensional weighting functions – Volterra kernels).



**Figure 1.** Environment effects in remote sensing systems

Building a model of nonlinear dynamic system in the form of a Volterra series lies in the choice of the test actions form. Also it uses the developed algorithm that allows determining the Volterra kernels and their Fourier-images for the measured responses (multidimensional amplitude–frequency characteristics (AFC) and phase-frequency characteristics (PFC)) to simulate the CC in the time or frequency domain, respectively [16-18, 23].

The additional research of new method of nonlinear dynamical systems identification, based on the Volterra model in the frequency domain is proposed. This method lies in  $n$ -fold differentiation of responses of the identifiable system by the amplitude of the test polyharmonic signals. The developed identification toolkit is used to build information model of the test nonlinear dynamic system in the form of the first, second and third order model. [12, 13, 15, 20, 21]

The aim of this work is to identify the continuous CC using Volterra model in the frequency domain, i.e. the determination of its multi-frequency characteristics on the basis of the data of the input-output experiment, using test polyharmonic signals and interpolation method to obtain model coefficients [14, 16-19, 21].

## 2. Volterra models and identification of dynamical systems in the frequency domain

Generally, “input–output” type ratio for nonlinear dynamical system can be presented by Volterra series [19].

$$\begin{aligned}
 y[x(t)] = & w_0(t) + \int_0^\infty w_1(\tau)x(t-\tau)d\tau + \int_0^\infty \int_0^\infty w_2(\tau_1, \tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \\
 & + \int_0^\infty \int_0^\infty \int_0^\infty w_3(\tau_1, \tau_2, \tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 + \dots = w_0(t) + \sum_{n=1}^\infty y_n[x(t)]
 \end{aligned}
 \tag{1}$$

where the  $n$ -th partial component of response of the system is

$$y_n[x(t)] = \int_0^\infty \dots \int_0^\infty w_n(\tau_1, \dots, \tau_n) \prod_{i=1}^n x(t-\tau_i) d\tau_i$$

$x(t)$  and  $y(t)$  are input and output signals of system respectively;  $w_n(\tau_1, \tau_2, \dots, \tau_n)$  – weight function or  $n$ -order Volterra kernel;  $y_n[x(t)]$  –  $n$ -th partial component of system’s response;  $w_0(t)$  – denotes free component of the series (for zero initial conditions  $w_0(t)=0$ );  $t$  – current time.

Commonly, the Volterra series are replaced by a polynomial, with only taking several first terms of series (1) into consideration. Nonlinear dynamical system identification in a form of Volterra series consists in  $n$ -dimensional weighting functions determination  $w_n(\tau_1, \dots, \tau_n)$  for time domain or it’s Fourier transforms  $W_n(j\omega_1, \dots, j\omega_n)$  –  $n$ -dimensional transfer functions for frequency domain.

Multidimensional Fourier transform for  $n$ -order Volterra kernel (1) is written in a form:

$$W_n(j\omega_1, \dots, j\omega_n) = F_n \langle w_n(\tau_1, \dots, \tau_n) \rangle = \int_0^\infty \dots \int_0^\infty w_n(\tau_1, \dots, \tau_n) \exp\left(-j \sum_{i=1}^n \omega_i \tau_i\right) \prod_{i=1}^n d\tau_i$$

where  $F_n \diamond$  –  $n$ -dimensional Fourier transform;  $j = \sqrt{-1}$ . Then the model of nonlinear system based on Volterra model in frequency domain can be represented as:

$$y[x(t)] = \sum_{n=1}^\infty F_n^{-1} \left\langle W_n(j\omega_1, \dots, j\omega_n) \prod_{i=1}^n X(j\omega_i) \right\rangle_{t_1 = \dots = t_n = t}$$

where  $F_n^{-1} \diamond$  – inverse  $n$ -dimensional Fourier transform;  $X(j\omega_i)$  – Fourier transform of input signal.

Identification of nonlinear system in frequency domain consists in determination of absolute value and phase of multidimensional transfer function at given frequencies – multidimensional AFC  $|W_n(j\omega_1, j\omega_2, \dots, j\omega_n)|$  and PFC  $\arg W_n(j\omega_1, j\omega_2, \dots, j\omega_n)$  which are defined by formulas:

$$|W_n(j\omega_1, \dots, j\omega_n)| = \sqrt{[\text{Re}(W_n(j\omega_1, \dots, j\omega_n))]^2 + [\text{Im}(W_n(j\omega_1, \dots, j\omega_n))]^2}
 \tag{2}$$

$$\arg W_n(j\omega_1, \dots, j\omega_n) = \arctg \frac{\text{Im}[W_n(j\omega_1, \dots, j\omega_n)]}{\text{Re}[W_n(j\omega_1, \dots, j\omega_n)]}
 \tag{3}$$

where Re and Im – accordingly real and imaginary parts of a complex function of  $n$  variables respectively.

**An approximation method** of identification of the nonlinear dynamical system based on Volterra series is offered [11, 14, 16-19]. During the identification of a Volterra kernel of  $m$ -th order significant effect on accuracy is rendered adjacent terms of a Volterra series. Therefore, it is necessary to apply the special methods, allowing minimizing this effect. The idea of such method lays in construction such expression of system responses to  $N$  ( $1 \leq m \leq N$ ) test input signals with the given amplitudes that with certain accuracy (accurate within to the thrown terms of order  $N+1$  and above) would be equal to  $m$ -th term Volterra series:

$$y_m[x(t)] = \sum_{j=1}^N c_j y[a_j x(t)] = \sum_{n=1}^{\infty} \left( \sum_{j=1}^N c_j a_j^n \right) \int_0^{\infty} \dots \int w_n(\tau_1, \dots, \tau_n) \prod_{i=1}^n x(t - \tau_i) d\tau_i \quad (4)$$

where  $a_j$  – amplitudes of test signals, random nonzero and pairwise different numbers;  $c_j$  – real coefficients which are chosen in such way that in a right part of (4) all first  $N$  terms are equal to 0, except  $m$ -th, and the multiplier at a  $m$ -multiple integral became equal to 1. This condition leads to a solution of the linear algebraic equations system concerning coefficients  $c_1, \dots, c_N$ :

$$\begin{aligned} c_1 a_1 + c_2 a_2 + \dots + c_N a_N &= 0 \\ &\dots \\ c_1 a_1^m + c_2 a_2^m + \dots + c_N a_N^m &= 1 \\ &\dots \\ c_1 a_1^N + c_2 a_2^N + \dots + c_N a_N^N &= 0. \end{aligned} \quad (5)$$

This system (5) always has a solution, and the unique one, as the system determinant differs from Vandermonde determinant with only a multiplier  $a_1 a_2 \dots a_N$ . Thus, with any real numbers  $a_j$ , that different from zero and pairwise different, it is possible to find such numbers  $c_j$  at which the linear combination (4) of system responses is equal to  $m$ -th term of a Volterra series accurate within to the thrown terms of series.

It is possible to build numberless assemblage of modes for expressions (4), by taking various numbers  $a_1, \dots, a_N$  and defining (5) coefficients  $c_1, \dots, c_N$  by them.

The choice of amplitudes  $a_j$  should provide the convergence of series (1) and an minimum error  $\Delta$  during extraction of a partial component  $y_m[x(t)]$  according to (4) defined by reminder of series (1) – terms of degree  $N+1$  and above

$$\sum_{j=1}^N c_j y[a_j x(t)] = \int_0^{\infty} \dots \int_0^{\infty} w_m(\tau_1, \dots, \tau_m) \prod_{i=1}^m x(t - \tau_i) d\tau_i + \sum_{j=1}^N c_j \sum_{n=N+1}^{\infty} y_n[a_j x(t)] = y_m[x(t)] + \Delta \quad (6)$$

If  $x(t)$  – is a test effect with maximum admissible amplitude at which a series (1) converges, amplitudes  $a_j$  should be no more than 1 by their absolute values:  $|a_j| \leq 1$  for  $\forall j=1, 2, \dots, N$ .

The amplitudes of the test signals  $a_{N_j}^{(m)}$  and the corresponding coefficients  $c_{N_j}^{(m)}$  for responses are shown in table 1, where  $m$  – order of the estimated Volterra kernel;  $j$  – number of the experiment;  $N$  – approximation order, i.e. quantity of identification experiments.

$N$	$m$	$a_{N_j}^{(m)}$						$c_{N_j}^{(m)}$						$\Delta$
2	1	-1		1		-0,5		0,5		-0,5		0,5		1
	2	-1		1		0,5		0,5		-0,5		0,5		1
4	1	-1	1	-0,5	0,5	0,167	0,167	1,333	1,333	3				
	2	-1	1	-0,644	0,644	-0,354	-0,354	2,061	2,061	4,8				
	3	-1	1	0,5	-0,5	-0,667	0,667	-1,333	1,333	4				
6	1	-0,8	1	-1	-0,3	0,3	0,8	0,28	0,09	-0,09	-2,13	2,13	0,28	5
	2	0,9	-1	0,9	0,4	-0,4	1	-0,8	0,41	-0,8	4,64	4,64	0,41	11,7
	3	0,3	1	0,8	-0,8	-1	-0,3	-5,46	-1,11	3,44	-3,44	1,11	5,46	20

**Table 1.** Numerical values of identification accuracy using approximation method

**An interpolation method** of identification of the nonlinear dynamical system based on Volterra series is offered [12, 13, 15, 20, 21]. It is used  $n$ -fold differentiation of a target signal on parameter-amplitude  $a$  of test actions to separate the response of the nonlinear dynamical system on partial components  $y_n[x(t)]$ .

*Affirmation 1.* Let at input of system test signal of  $ax(t)$  kind is given, where  $x(t)$  – is arbitrary function and  $a$  – is scale coefficient (amplitude of signal), where  $0 < |a| \leq 1$ , then for the selection of a partial component of the  $n$ -th order  $\hat{y}_n(t)$  from measurement of the response nonlinear system  $y[ax(t)]$  in the form of Volterra series, it is necessary to determine  $n$ -th partial derivative of the total response amplitude  $a$  where  $a=0$

$$\hat{y}_n(t) = \int_0^\infty \dots \int_0^\infty w_n(\tau_1, \dots, \tau_n) \prod_{i=1}^n x(t-\tau_i) d\tau_i = \frac{1}{n!} \left. \frac{\partial^n y[ax(t)]}{\partial a^n} \right|_{a=0} \quad (7)$$

We use the method of extracting the partial components with the help of  $n$ -fold differentiation of the response  $y[ax(t)]$  with respect to parameter–amplitude  $a$  and the use of the derivative value at  $a=0$ .

Injecting an input signal  $ax(t)$  where  $a$  is the scaling factor (signal amplitude), one has the following response of the nonlinear system:

$$\begin{aligned}
 y[a \cdot x(t)] = & a \int_0^t w(\tau) \cdot x(t-\tau) d\tau + a^2 \iint_{00}^{tt} w_2(\tau_1, \tau_2) x(t-\tau_1) x(t-\tau_2) d\tau_1 d\tau_2 + \\
 & + a^n \int_0^t \dots \int_0^t w_n(\tau_1, \dots, \tau_n) \prod_{r=1}^n x(t-\tau_r) d\tau_r + \dots
 \end{aligned}
 \tag{8}$$

To distinguish the partial component of the  $n$ -th order, differentiate the system response  $n$  times with respect to the amplitude:

$$\begin{aligned}
 \frac{\partial^n y[a \cdot x(t)]}{\partial a^n} = & n! \int_0^t \dots \int_0^t w_n(\tau_1, \dots, \tau_n) \prod_{r=1}^n x(t-\tau_r) d\tau_r + \\
 & (n+1)! \cdot a \int_0^t \dots \int_0^t w_{n+1}(\tau_1, \dots, \tau_{n+1}) \prod_{r=1}^{n+1} x(t-\tau_r) d\tau_r + \dots
 \end{aligned}
 \tag{9}$$

Taking the value of the derivative at  $a=0$ , we finally obtain the expression for the partial component (7).

*Formulas for numerical differentiation.* Partial derivative should be substituted by form of finite difference for calculation. Differentiation of function, which was set in discrete points, could be accomplished by means of numerical computing after preliminary smoothing of measured results. Various formulas for the numerical differentiation are known, which differ from each other by means of error.

Let's use universal reception which allows to substitute a derivative of any  $n$  order for differential ratio so that the error from such replacement for function  $y(a)$  was any beforehand set order of  $p$  approximation concerning a step of  $h=\Delta a$  of computational mesh on amplitude. Method of undetermined coefficient for equality

$$\frac{d^n y(a)}{da^n} = \frac{1}{h^n} \sum_{r=-r_1}^{r_2} c_r y(a+rh) + O(h^p),
 \tag{10}$$

where the coefficients  $c_r$  are taken not depending on  $h$ ,  $r = -r_1, -r_1+1, \dots, -1, 0, 1, \dots, r_2-1, r_2$ , so that equality (10) was fair. Limits of summation  $r_1 \geq 0$  and  $r_2 \geq 0$  could be arbitrary, but so that the differential relation  $h^{-n} \sum c_r y(a+rh)$  of  $r_1+r_2$  order have to satisfy to inequality  $r_1+r_2 \geq n+p-1$ .

To define the  $c_r$ , it is necessary to solve the following set of equations

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ -r_1 & -r_1 + 1 & \dots & r_2 \\ \dots & \dots & \dots & \dots \\ (-r_1)^{n-1} & (-r_1 + 1)^{n-1} & \dots & r_2^{n-1} \\ (-r_1)^n & (-r_1 + 1)^n & \dots & r_2^n \\ (-r_1)^{n+1} & (-r_1 + 1)^{n+1} & \dots & r_2^{n+1} \\ \dots & \dots & \dots & \dots \\ (-r_1)^{n+p-1} & (-r_1 + 1)^{n+p-1} & \dots & r_2^{n+p-1} \end{bmatrix} \begin{bmatrix} c_{-r_1} \\ c_{-r_1+1} \\ \dots \\ c_{-1} \\ c_0 \\ c_1 \\ \dots \\ c_{r_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \\ n! \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (11)$$

If  $r_1 + r_2 = n + p - 1$ , then inscribed in  $n + p$  equality forms linear system concerning the same number of  $c_r$  unknown. The determiner of this system is Vandermonde's determiner and differs from zero. Thus, there is only one set of  $n$  coefficients, satisfying the system.

If  $r_1 + r_2 \geq n + p$ , then there are many such sets of coefficients  $c_r$ .

On the basis of (10) in [20] the formulas of derivative calculation of the first, second and third orders are received at  $a=0$  with use of the central differences for equidistant nodes of the computational grid.

In work formulas for numerical differentiation with use of the central differences for equidistant assembly are used. Volterra kernel of the first order is determined by formulas as the first derivative at  $r_1=r_2=1$ ,  $r_1=r_2=2$  or  $r_1=r_2=3$  respectively

$$\begin{aligned} y'_0 &= \frac{1}{2h}(-y_{-1} + y_1), \\ y'_0 &= \frac{1}{12h}(y_{-2} - 8y_{-1} + 8y_1 - y_2), \\ y'_0 &= \frac{1}{60h}(-y_{-3} + 9y_{-2} - 45y_{-1} + 45y_1 - 9y_2 + y_3). \end{aligned} \quad (12)$$

Volterra kernel of the first order is determined by formulas as the first derivative at  $r_1=r_2=1$ ,  $r_1=r_2=2$  or  $r_1=r_2=3$  respectively

$$\begin{aligned} y''_0 &= \frac{1}{h^2}(y_{-1} - 2y_0 + y_1), \\ y''_0 &= \frac{1}{12h^2}(-y_{-2} + 16y_{-1} - 30y_0 + 16y_1 - y_2), \\ y''_0 &= \frac{1}{180h^2}(2y_{-3} - 27y_{-2} + 270y_{-1} - 490y_0 + 270y_1 - 27y_2 + 2y_3). \end{aligned} \quad (13)$$

Volterra kernel of the first order is determined by formulas as the first derivative at  $r_1=r_2=2$  or  $r_1=r_2=3$  respectively

$$\begin{aligned}
 y_0''' &= \frac{1}{2h^3}(-y_{-2} + 2y_{-1} - 2y_1 + y_2), \\
 y_0''' &= \frac{1}{8h^3}(y_{-3} - 8y_{-2} + 13y_{-1} - 13y_1 + 8y_2 - y_3).
 \end{aligned}
 \tag{14}$$

In the formulas written above, we use the following notations

$$y'_0 = y'(0), \quad y''_0 = y''(0), \quad y'''_0 = y'''(0); \quad y_r = y(rh), \quad r = 0, \pm 1, \pm 2; \pm 3$$

where we put  $y_0=0$ , since identification nonlinear systems is implemented with zero initial conditions. The test polyharmonic effects for identification in the frequency domain representing by signals of such type:

$$x(t) = \sum_{k=1}^n A_k \cos(\omega_k t + \varphi_k), \tag{15}$$

where  $n$  – the order of transfer function being estimated;  $A_k$ ,  $\omega_k$  and  $\varphi_k$  – accordingly amplitude, frequency and a phase of  $k$ -th harmonics. In research, it is supposed every amplitude of  $A_k$  to be equal, and phases  $\varphi_k$  equal to zero.

For identification in the frequency domain the test polyharmonic signals are used. We prove:

*Statement.* If test polyharmonic signal is used in form

$$x(t) = A \sum_{k=1}^n \cos \omega_k t = \frac{A}{2} \sum_{k=1}^n (e^{j\omega_k t} + e^{-j\omega_k t}) \tag{16}$$

then the  $n$ -th partial component of the response of test system can be written in form:

$$\begin{aligned}
 y_n(t) &= \frac{A^n}{2^{n-1}} \sum_{m=0}^{E(n/2)} C_n^m \sum_{k_1=1}^n \dots \sum_{k_m=1}^n |W_n(-j\omega_{k_1}, \dots, -j\omega_{k_m}, j\omega_{k_{m+1}}, \dots, j\omega_{k_n})| \times \\
 &\times \cos\left(\left(-\sum_{l=0}^m \omega_{k_l} + \sum_{l=m+1}^n \omega_{k_l}\right)t + \arg W_n(-j\omega_{k_1}, \dots, -j\omega_{k_m}, j\omega_{k_{m+1}}, \dots, j\omega_{k_n})\right),
 \end{aligned}
 \tag{17}$$

where  $E(n/2)$  – function used to obtain the of integer part of the value.

The component with frequency  $\omega_1 + \dots + \omega_n$  is extracted from the response to test signal (15):

$$A^n |W_n(j\omega_1, \dots, j\omega_n)| \cos[(\omega_1 + \dots + \omega_n)t + \arg W_n(j\omega_1, \dots, j\omega_n)]. \quad (18)$$

Certain limitations should be imposed while choosing of frequency polyharmonic test signals in a process determining multidimensional AFC and PFC. This is the reason why the values of AFC and PFC in this unallowable points of multidimensional frequency space can be calculated using interpolation only. In practical realization of nonlinear dynamical systems identification it is needed to minimize number of such undefined points at the range of multidimensional frequency characteristics determination. This was performed to provide a minimum of restrictions on choice of frequency of the test signal. It is shown that existed limitation can be weakened. New limitations on choice of frequency are reducing number of undefined points.

After analyzing the (18) it is defined: to obtain Volterra kernels for nonlinear dynamical system in frequency domain the limitations on choice of frequencies of test polyharmonic signals have to be restricted. These restrictions provide inequality of combination frequencies in the test signal harmonics. The theorem about choice of test signals frequencies is proven.

*The theorem about choice of test signals frequencies.* For the definite filtering of a response of the harmonics with combination frequencies  $\omega_1 + \omega_2 + \dots + \omega_n$  within the  $n$ -th partial component it is necessary and sufficient to keep the frequency from being equal to another combination frequencies of type  $k_1\omega_1 + \dots + k_n\omega_n$ , where the coefficients  $\{k_i | i=1, 2, \dots, n\}$  must satisfy the conditions:

- number  $K$  of negative value coefficients ( $k_i < 0$ ) is in  $0 \leq K \leq E(n/2)$  (where  $E$  – function used to obtain the of integer part of the value);
- $\sum_{i=1}^n |k_i| \leq n$ ;
- $\sum_{i=1}^n |k_i| \equiv n \pmod{2}, n - \sum_{i=1}^n |k_i| = 2l, l \in \mathbb{N}$ .

It was shown that during determination of multidimensional transfer functions of nonlinear systems it is necessary to consider the imposed constraints on choice of the test polyharmonic signal frequencies. This provides inequality of combination frequencies in output signal harmonics:  $\omega_1 \neq 0, \omega_2 \neq 0$  and  $\omega_1 \neq \omega_2$  for the second order identification procedure, and  $\omega_1 \neq 0, \omega_2 \neq 0, \omega_3 \neq 0, \omega_1 \neq \omega_2, \omega_1 \neq \omega_3, \omega_2 \neq \omega_3, 2\omega_1 \neq \omega_2 + \omega_3, 2\omega_2 \neq \omega_1 + \omega_3, 2\omega_3 \neq \omega_1 + \omega_2, 2\omega_1 \neq \omega_2 - \omega_3, 2\omega_2 \neq \omega_1 - \omega_3, 2\omega_3 \neq \omega_1 - \omega_2, 2\omega_1 \neq -\omega_2 + \omega_3, 2\omega_2 \neq -\omega_1 + \omega_3$  and  $2\omega_3 \neq -\omega_1 + \omega_2$  for the third order identification procedure.

The amplitudes of the test signals  $a_i^{(n)}$  and the corresponding coefficients  $c_i^{(n)}$  for responses are shown in Table 2, where  $n$  – order of the estimated Volterra kernel;  $i$  – number of the experiment ( $i=1, 2, \dots, N$ ), where  $N=r_1+r_2$ , i.e. number of interpolation knots (number of experiments).

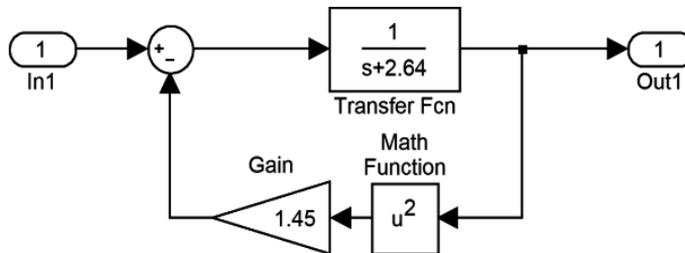
$n$	$N$	$a_1^{(n)}$	$a_2^{(n)}$	$a_3^{(n)}$	$a_4^{(n)}$	$a_5^{(n)}$	$a_6^{(n)}$	$c_1^{(n)}$	$c_2^{(n)}$	$c_3^{(n)}$	$c_4^{(n)}$	$c_5^{(n)}$	$c_6^{(n)}$
1	2	-1	1					-0,5	0,5				
	4	-1	-0,5	0,5	1			0,0833	-0,6667	0,6667	-0,0833		
	6	-1	-0,67	-0,33	0,33	0,67	1	-0,0167	0,15	-0,75	0,75	-0,15	0,0167
2	2	-1	1					1	1				
	4	-1	-0,5	0,5	1			-0,0833	1,3333	1,3333	-0,0833		
	6	-1	-0,67	-0,33	0,33	0,67	1	0,0111	-0,15	1,5	1,5	-0,15	0,0111
3	4	-1	-0,5	0,5	1			-0,5	1	-1	0,5		
	6	-1	-0,67	-0,33	0,33	0,67	1	0,125	-1	1,625	-1,625	1	-0,125

**Table 2.** Amplitudes and corresponding coefficients of the interpolation method

### 3. The techniques of test system identification

Described method was tested using nonlinear test system (fig. 2) represented by Riccati equation

$$\frac{dy(t)}{dt} + \alpha y(t) + \beta y^2(t) = u(t). \tag{19}$$



**Figure 2.** Simulink-model of the test system

Analytical expressions of AFC and PFC for the first, second and third order model were received:

$$|W_1(j\omega)| = \frac{1}{\sqrt{\alpha^2 + \omega^2}}, \arg W_1(j\omega) = -\arctg \frac{\omega}{\alpha};$$

$$|W_2(j\omega_1, j\omega_2)| = \frac{\beta}{\sqrt{(\alpha^2 + \omega_1^2)(\alpha^2 + \omega_2^2)[\alpha^2 + (\omega_1 + \omega_2)^2]}}$$

$$\arg W_2(j\omega_1, j\omega_2) = -\arctg \frac{(2\alpha^2 - \omega_1\omega_2)(\omega_1 + \omega_2)}{\alpha(\alpha^2 - \omega_1\omega_2) - \alpha(\omega_1 + \omega_2)^2};$$

$$|W_3(j\omega_1, j\omega_2, j\omega_3)| = \sqrt{[\operatorname{Re}(W_3(j\omega_1, j\omega_2, j\omega_3))]^2 + [\operatorname{Im}(W_3(j\omega_1, j\omega_2, j\omega_3))]^2} =$$

$$= \frac{2\beta^2}{3} \frac{1}{\sqrt{[\alpha^2 + (\omega_1 + \omega_2 + \omega_3)^2](\alpha^2 + \omega_1^2)(\alpha^2 + \omega_2^2)(\alpha^2 + \omega_3^2)}} \times$$

$$\times \frac{\sqrt{[3\alpha^2 - (\omega_1 + \omega_3)(\omega_1 + \omega_2) - (\omega_1 + \omega_2)](\omega_2 + \omega_3) - (\omega_1 + \omega_3)(\omega_2 + \omega_3)]^2 + 16\alpha^2(\omega_1 + \omega_2 + \omega_3)^2}}{\sqrt{[\alpha^2 + (\omega_1 + \omega_2)^2][\alpha^2 + (\omega_1 + \omega_3)^2][\alpha^2 + (\omega_2 + \omega_3)^2]}}$$

$$\arg W_3(j\omega_1, j\omega_2, j\omega_3) = \arctg \frac{\operatorname{Im}W_3(j\omega_1, j\omega_2, j\omega_3)}{\operatorname{Re}W_3(j\omega_1, j\omega_2, j\omega_3)} = -\arctg \frac{DA - CB}{AB + CD},$$

where

$$u = \alpha^3 - \alpha\omega_1\omega_3 - \alpha\omega_2\omega_3 - \alpha(\omega_1 + \omega_2 + \omega_3);$$

$$v = (\omega_1 + \omega_2 + \omega_3)(2\alpha^2 - \omega_1\omega_3 - \omega_2\omega_3);$$

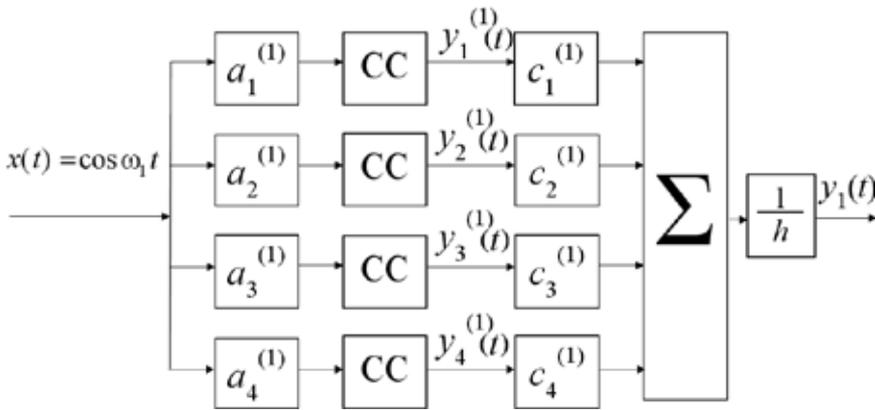
$$w = (\alpha^2 - \omega_1\omega_2 - \omega_2\omega_3)(\alpha^2 - \omega_1\omega_2 - \omega_1\omega_3) - \alpha^2(\omega_1 + \omega_2 + \omega_3)^2;$$

$$x = \alpha(\omega_1 + \omega_2 + \omega_3)(2\alpha^2 - 2\omega_1\omega_2 - \omega_1\omega_3 - \omega_2\omega_3).$$

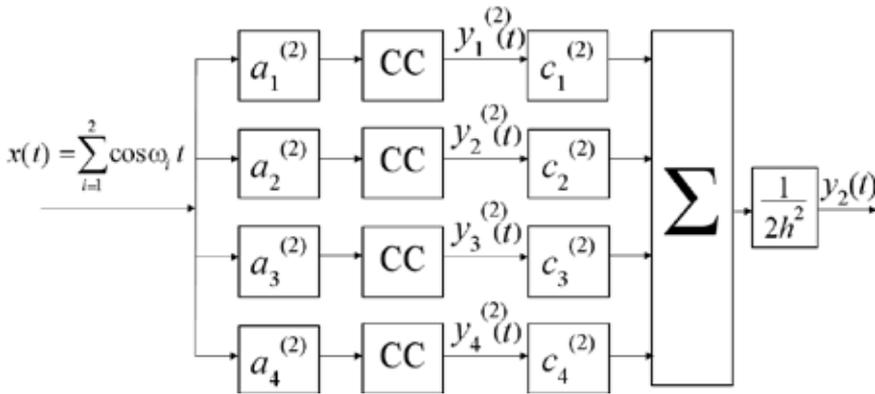
The main purpose was to identify the multi-frequency performances characterizing nonlinear and dynamical properties of nonlinear test system [11–21]. Volterra model in the form of the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> order polynomial is used. Thus, test system properties are characterized by transfer functions of  $W_1(j\omega)$ ,  $W_2(j\omega_1, j\omega_2)$ ,  $W_3(j\omega_1, j\omega_2, j\omega_3)$  – by Fourier-images of weight functions  $w_1(t)$ ,  $w_2(t_1, t_2)$  and  $w_3(t_1, t_2, t_3)$ .

Structure charts of identification procedure – determinations of the 1<sup>th</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> order AFC of CC are presented accordingly in figs. 3–5.

The weighted sum is formed from received signals – responses of each group from figs. 3–5. As a result the partial components of CC responses  $y_1(t)$ ,  $y_2(t)$  and  $y_3(t)$  are got. For each partial component of response the Fourier transform (the FFT is used) is calculated, and from received spectrum only an informative harmonics (which amplitudes represent values of required characteristics of the first, second and third orders AFC) are taken.



**Figure 3.** The structure chart of identification using 1<sup>st</sup> order Volterra model in frequency domain, number of experiments  $N=4$ :  $a_1=-2h, a_2=-h, a_3=h, a_4=2h; c_1=-1/12, c_2=-2/3, c_3=2/3, c_4=1/12$



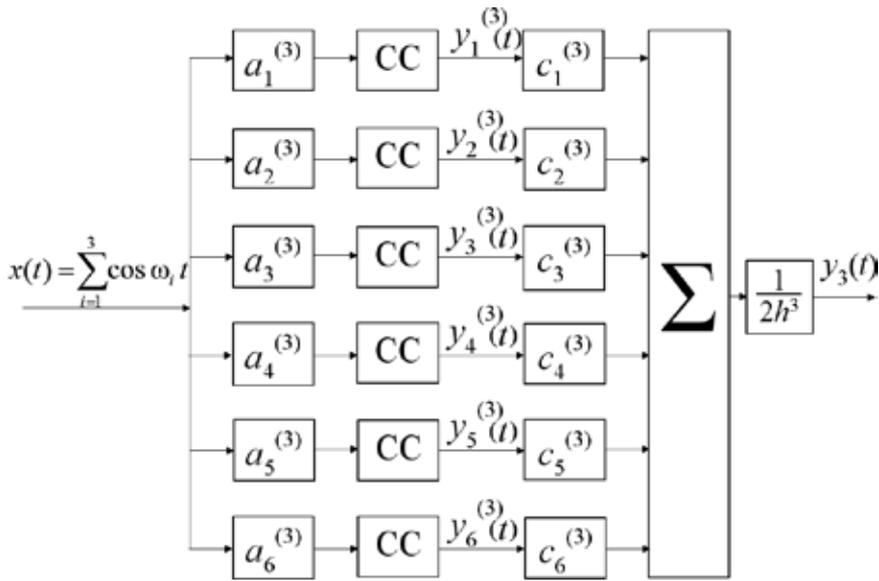
**Figure 4.** The structure chart of identification using 2<sup>nd</sup> order Volterra model in frequency domain, number of experiments  $N=4$ :  $a_1=-2h, a_2=-h, a_3=h, a_4=2h; c_1=-1/12, c_2=4/3, c_3=4/3, c_4=-1/12$

The first order AFC  $|W_1(j\omega)|$  and PFC  $\arg W_1(j\omega)$  is received by extracting the harmonics with frequency  $\omega$  from the spectrum of the CC partial response  $y_1(t)$  to the test signal  $x(t)=(A/2)\cos\omega t$ .

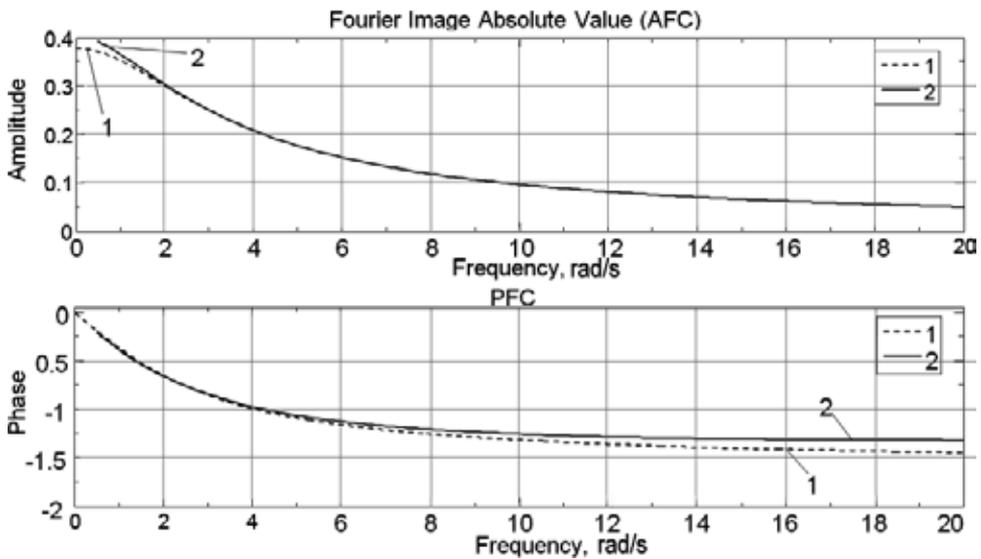
The second order AFC  $|W_2(j\omega, j(\omega+\Omega_1))|$  and PFC  $\arg W_2(j\omega, j(\omega+\Omega_1))$  having  $\omega_1=\omega$  and  $\omega_2=\omega+\Omega_1$  were received by extracting the harmonics with summary frequency  $\omega_1+\omega_2$  from the spectrum of the CC partial response  $y_2(t)$  to the test signal  $x(t)=(A/2)(\cos\omega_1 t+\cos\omega_2 t)$ .

The third order AFC  $|W_3(j\omega, j(\omega+\Omega_1), j(\omega+\Omega_2))|$  and PFC  $\arg W_3(j\omega, j(\omega+\Omega_1), j(\omega+\Omega_2))$  having  $\omega_1=\omega, \omega_2=\omega+\Omega_1$  and  $\omega_3=\omega+\Omega_2$ , were received by extracting the harmonics with summary frequency  $\omega_1+\omega_2+\omega_3$  from the spectrum of the CC partial response  $y_3(t)$  to the test signal  $x(t)=(A/2)(\cos\omega_1 t+\cos\omega_2 t+\cos\omega_3 t)$ .

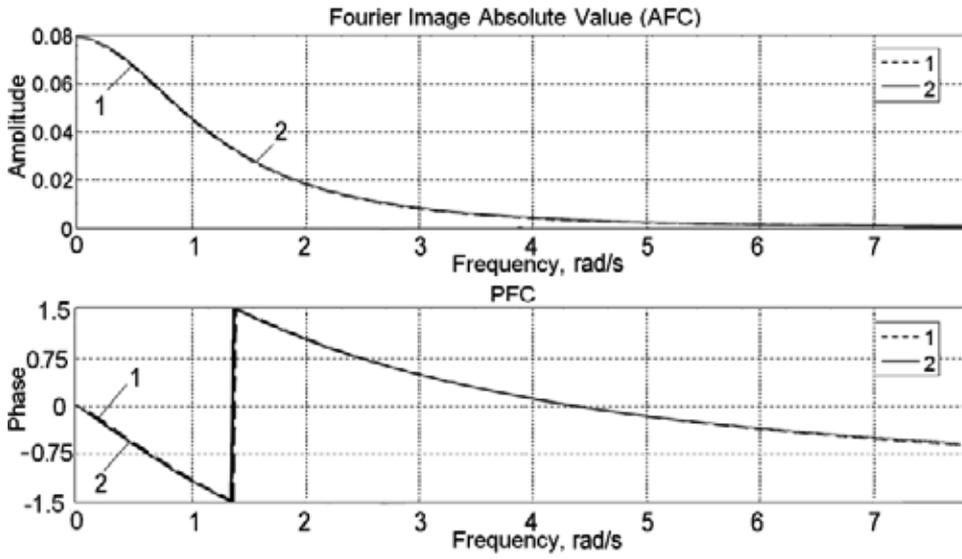
The results (first, second and third order AFC and PFC which had been received after procedure of identification) are represented in figs. 6–8.



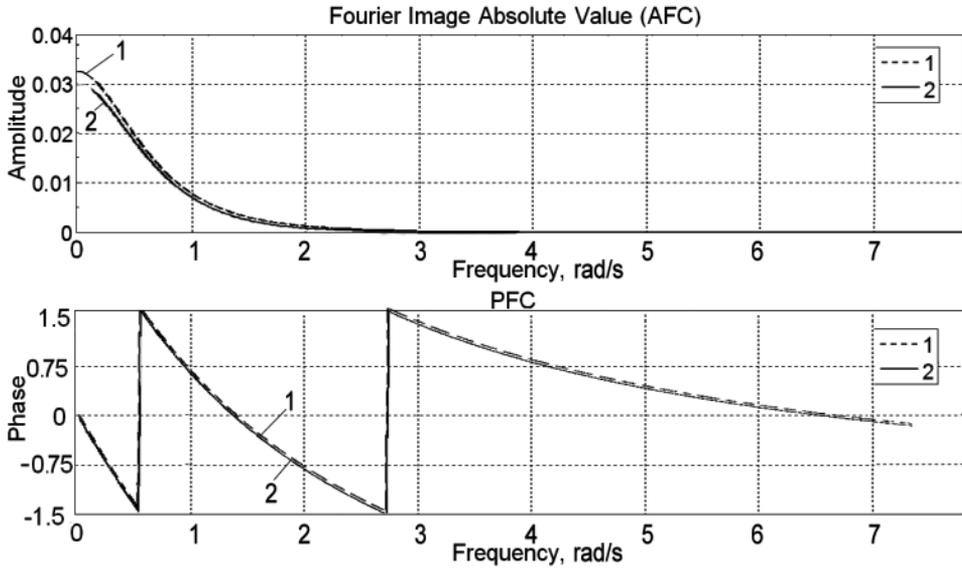
**Figure 5.** The structure chart of identification using 3<sup>rd</sup> order Volterra model in frequency domain, number of experiments  $N=6$ :  $a_1=-3h, a_2=-2h, a_3=-h, a_4=h, a_5=2h, a_6=3h$ ;  $c_1=1/8, c_2=1, c_3=13/8, c_4=-13/8, c_5=1, c_6=1/8$



**Figure 6.** First order AFC and PFC of the test system: analytically calculated values (1), section estimation values with number of experiments for the model  $N=4$  (2)

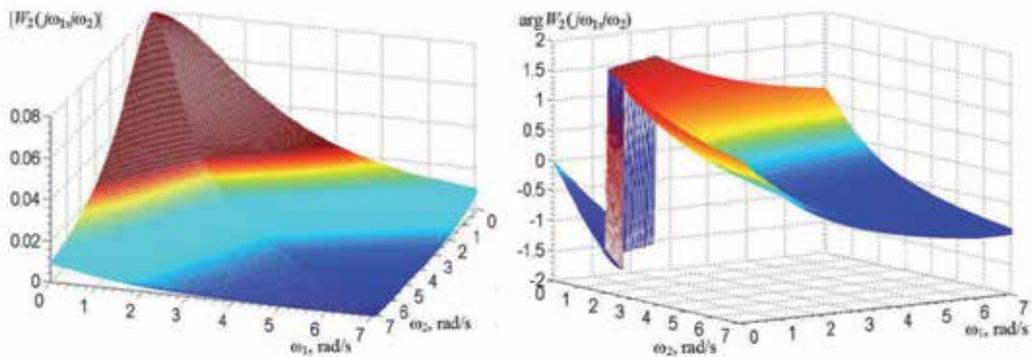


**Figure 7.** Second order AFC and PFC of the test system: analytically calculated values (1), sub-diagonal cross-section values with number of experiments for the model  $N=4$  (2),  $\Omega_1=0,01$  rad/s

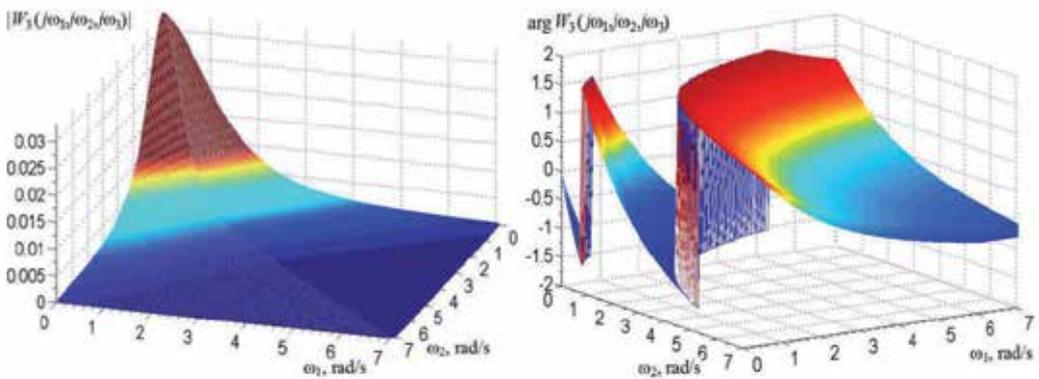


**Figure 8.** Third order AFC and PFC of the test system: analytically calculated values (1), sub-diagonal cross-section values with number of experiments for the model  $N=6$  (2),  $\Omega_1=0,01$  rad/s,  $\Omega_2=0,1$  rad/s

The second and third order surfaces for AFC and PFC received after procedure of the test system identification are shown in fig. 9 and fig. 10 respectively.



**Figure 9.** Surface of the test system AFC (left) and PFC (right) built of the second order subdiagonal cross-sections received for  $N=4$



**Figure 10.** Surface of the test system AFC (left) and PFC (right) built of the third order subdiagonal cross-sections received for  $N=6, \omega_3=0,1 \text{ rad/s}$

Presented surfaces are built from sub-diagonal cross-sections which were received separately.  $\Omega_1$  was used as growing parameter of identification with different value for each cross-section in second order characteristics. Fixed value of  $\Omega_2$  and growing value of  $\Omega_1$  were used as parameters of identification to obtain different value for each cross-section in third order characteristics.

Numerical values of identification accuracy using interpolation method for the test system are represented in table 3, where:  $n$  – order of the estimated Volterra kernel,  $N$  – approximation order/number of interpolation knots (number of experiments).

<i>n</i>	<i>N</i>	AFC relative error, %	PFC relative error, %
1	2	2.1359	2.5420
	4	0.3468	2.0618
	6	0.2957	1.9311
2	2	30.2842	76.8221
	4	2.0452	3.7603
	6	89.2099	5.9438
3	4	10.981	1.628
	6	10.7642	1.5522

**Table 3.** Numerical values of identification accuracy using interpolation method

Comparison of the numerical values for identification accuracy using interpolation method [17-18] and approximation one [13-16] for the test system is presented in table 4.

<i>n</i>	<i>N</i>	AFC relative error, %		PFC relative error, %	
		approximation	interpolation	approximation	interpolation
1	2	3.6429	2.1359	3.3451	2.5420
	4	1.1086	0.3468	3.1531	2.0618
	6	0.8679	0.2957	3.1032	1.9311
2	2	26.0092	30.2842	30.2842	76.8221
	4	3.4447	2.0452	2.0452	3.7603
	6	7.3030	89.2099	4.6408	5.9438
3	4	72.4950	10.981	10.9810	1.628
	6	74.4204	10.7642	10.7642	1.5522

**Table 4.** Identification accuracy using approximation and interpolation methods

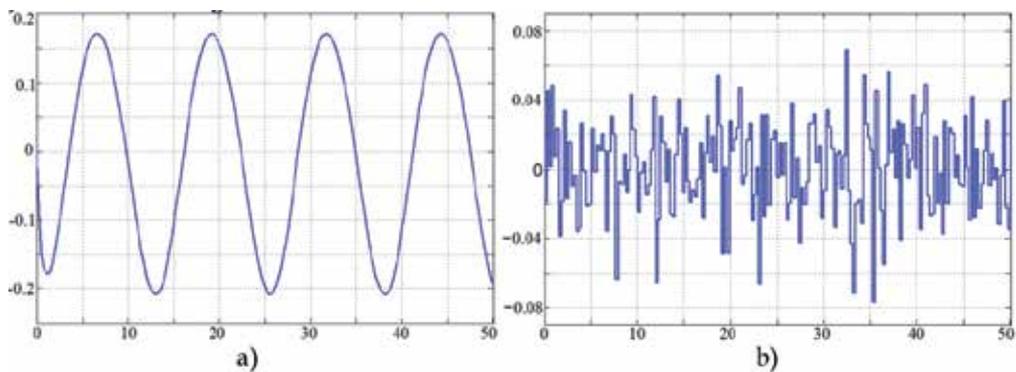
#### 4. The study of noise immunity of the identification method

Experimental researches of the noise immunity of the identification method were made. The main purpose was the studying of the noise impact (noise means the inexactness of the

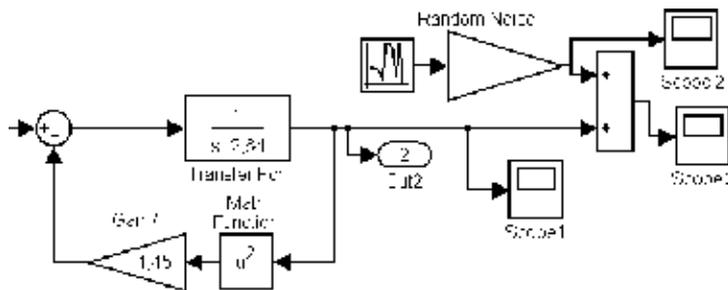
measurements) to the characteristics of the test system model using interpolation method in frequency domain.

The first step was the measurement of the level of useful signal (harmonic cosine test signal shown in fig. 11a) after test system (Out2 in fig. 12). The amplitude of this signal was defined as the 100% of the signal power.

After that procedure the Random Noise signal (with the form shown in fig. 11b) where added to the test system output signal. This steps where performed to simulate inexactness of the measurements in the model. The sum of these two signals for the linear test model signal is shown in fig. 13.



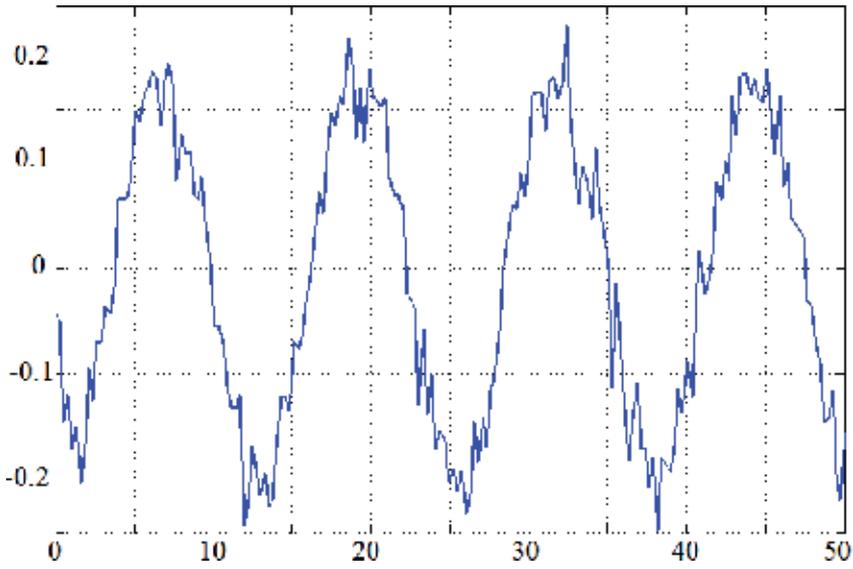
**Figure 11.** a) Test signal and b) random noise with 50% amplitude of test signal



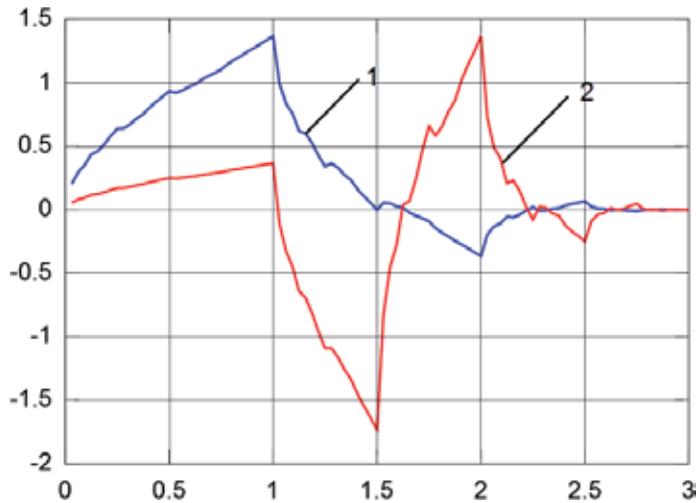
**Figure 12.** The Simulink model of the test system with noise generator and osillosopes

The simulations with the test model were performed. Different noise levels were defined for different order of the Volterra model.

The adaptive wavelet denoising was used to reduce the noise impact on final characteristics of the test system. The Daubechie wavelet of the 2 and 3 level was chosen (fig. 14) and used for the AFC and PFC denoising respectively [2, 5, 7].



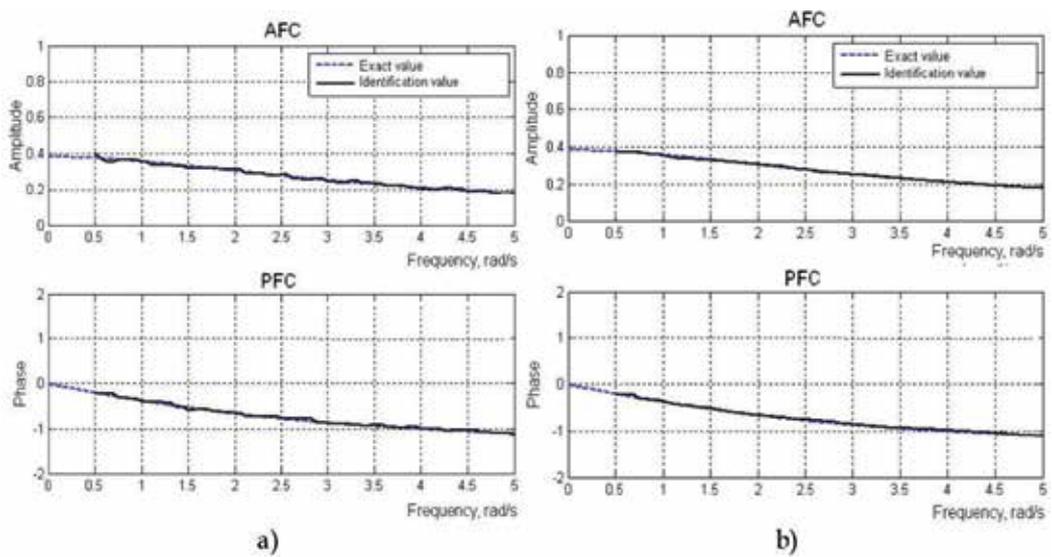
**Figure 13.** The “noised” signal of the test system, level of noise is 50% of source signal



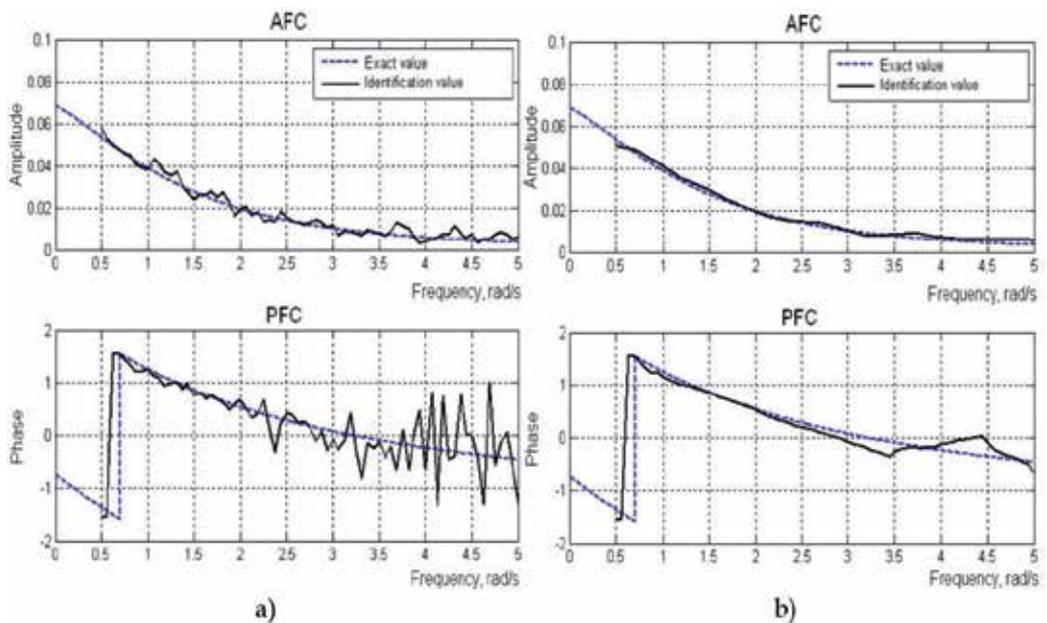
**Figure 14.** The Daubechies 2<sup>nd</sup> level scaling (1) and wavelet (2) functions

The first order (linear) model was tested with the level of noise 50% and 10% and showed excellent level of noise immunity. The noised (fig. 15a) and de-noised (filtered) (fig. 15b) characteristics (AFC and PFC) with level of noise 50% are presented.

The second order (nonlinear) model was tested with the level of noise 10% and 1% and showed good level of noise immunity. The noised (fig. 16a) and de-noised (filtered) (fig. 16b) characteristics (AFC and PFC) with level of noise 10% are presented.

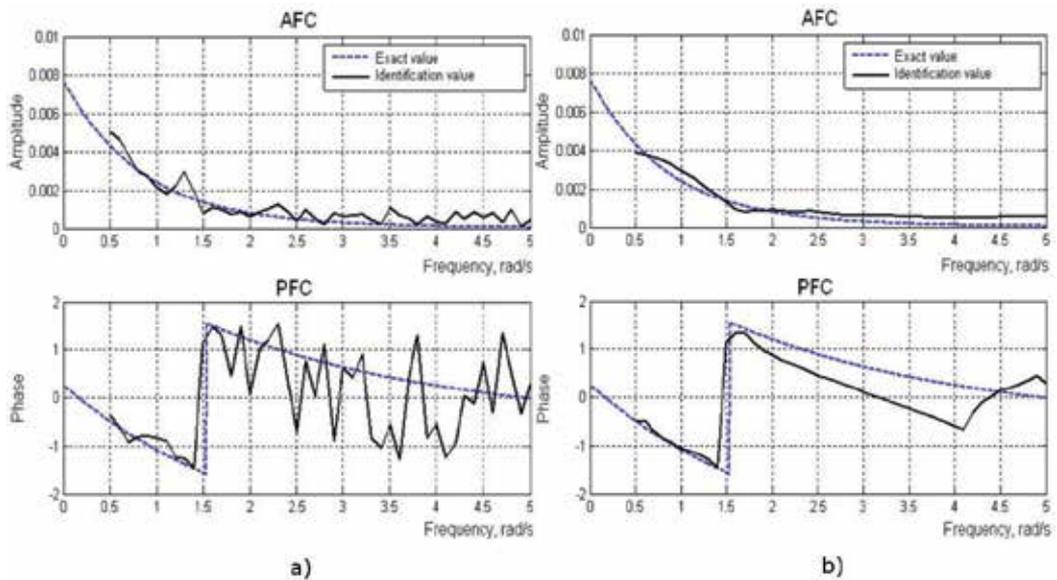


**Figure 15.** Noised (a) and denoised (b) characteristics (AFC – top, PFC – bottom) of the 1<sup>st</sup> order model of the test system with level of noise 50%



**Figure 16.** Noised (a) and denoised (b) characteristics (AFC – top, PFC – bottom) of the 2<sup>nd</sup> order model of the test system with level of noise 10%

The third order (nonlinear) model was tested with the level of noise 10% and 1% and showed good level of noise immunity. The noised (fig. 17a) and de-noised (filtered) (fig. 17b) characteristics (AFC and PFC) with level of noise 1% are presented.



**Figure 17.** Noised (a) and denoised (b) characteristics (AFC – top, PFC – bottom) of the 3<sup>rd</sup> order model of the test system with level of noise 1%

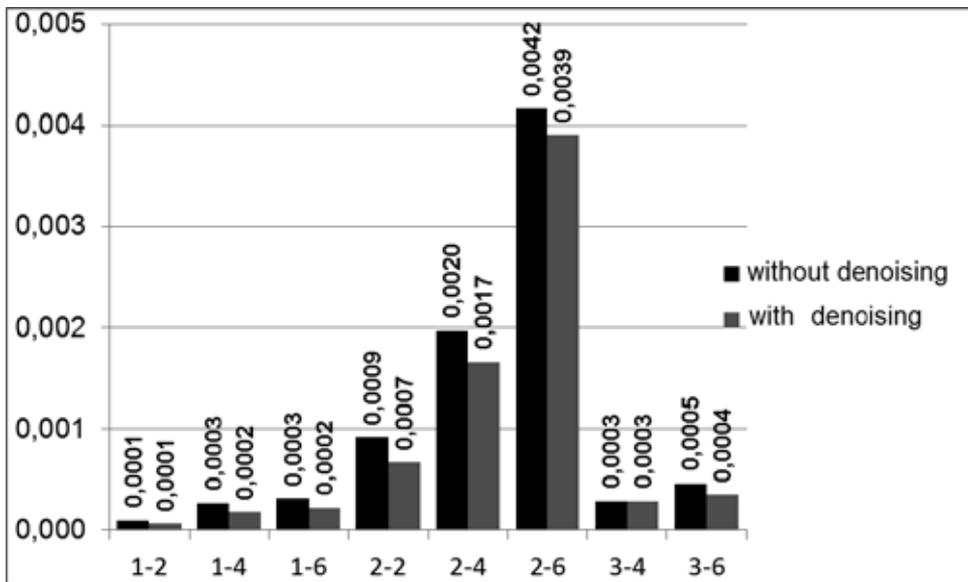
The numerical values of standard deviation (SD) of the identification accuracy before and after wavelet denoising procedure are presented in Table 5.

<i>n</i>	<i>N</i>	Noise level = 10%		Noise level = 1%		Improvement	
		SD for AFC	SD for PFC	SD for AFC	SD for PFC	for AFC, times	for PFC, times
(without / with denoising)							
	2	0.000097 / 0.000063	0.09031 / 0.07541	–	–	<b>1,540</b>	1,198
1	4	0.000271 / 0.000181	0.07804 / 0.06433	–	–	1,497	1,213
	6	0.000312 / 0.000223	0.12913 / 0.09889	–	–	1,399	1,306
2	2	0.000920 / 0.000670	0.52063 / 0.51465	–	–	1,373	1,012

n	N	Noise level = 10%		Noise level = 1%		Improvement	
		SD for AFC	SD for PFC	SD for AFC	SD for PFC	for AFC, times	for PFC, times
	4	0.001972 / 0.001663	0.28004 / 0.06877	–	–	1,186	<b>4,072</b>
	6	0.004165 / 0.003908	0.39260 / 0.19237	–	–	1,066	2,041
3	4	–	–	0.000288 / 0.000288	0.89857 / 0.61251	1,003	1,467
	6	–	–	0.000461 / 0.000352	0.84868 / 0.59319	1,310	1,431

**Table 5.** Standard deviation for interpolation method with noise impact (bold font shows the best values)

The diagrams showing the improvement of standard deviation for identification accuracy using the adaptive wavelet denoising of the received characteristics (AFC and PFC) are shown in fig. 18 and fig. 19 respectively.



**Figure 18.** Standard deviation changing for AFC using adaptive Wavelet-denoising

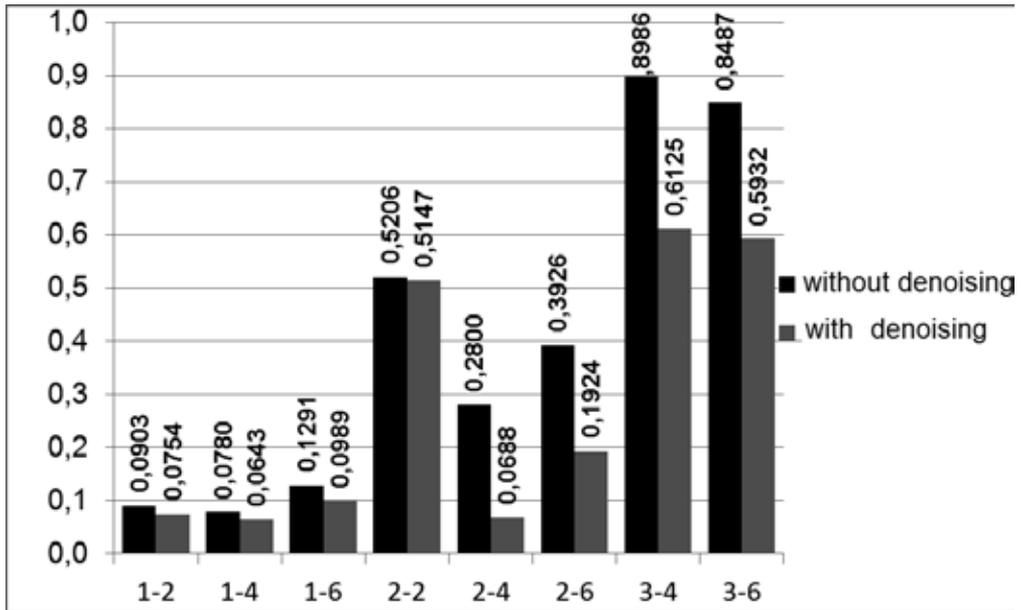


Figure 19. Standard deviation changing for PFC using adaptive Wavelet-denoising

## 5. The technique and hardware-software tools of radiofrequency CC identification

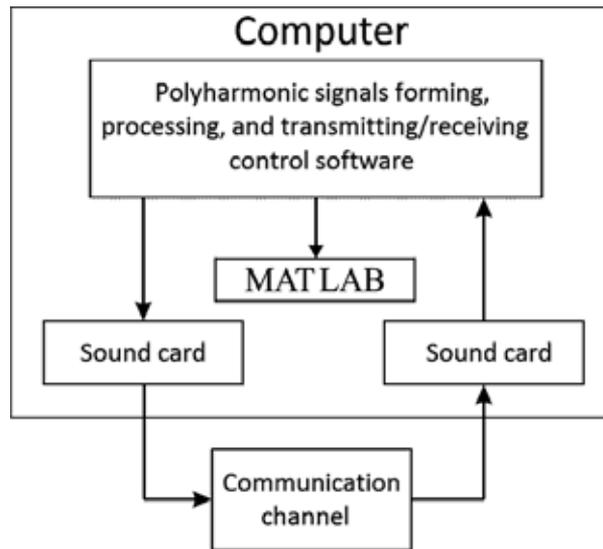
Experimental research of the Ultra High Frequency range CC were done. The main purpose was the identification of multi-frequency characteristics that characterize nonlinear and dynamical properties of the CC. Volterra model in the form of the second order polynomial is used. Thus physical CC properties are characterized by transfer functions of  $W_1(j2\pi f)$ ,  $W_2(j2\pi f_1, j2\pi f_2)$  and  $W_3(j2\pi f_1, j2\pi f_2, j2\pi f_3)$  – by the Fourier-images of weighting functions  $w_1(t)$ ,  $w_2(t_1, t_2)$  and  $w_3(t_1, t_2, t_3)$ .

Implementation of identification method on the IBM PC computer basis has been carried out using the developed software in Matlab software. The software allows automating the process of the test signals forming with the given parameters (amplitudes and frequencies). Also this software allows transmitting and receiving signals through an output and input section of PC soundcard, to produce segmentation of a file with the responses to the fragments, corresponding to the CC responses being researched on test polyharmonic effects with different amplitudes.

In experimental research two identical marine transceivers S.P.RADIO A/S SAILOR RT2048 VHF (the range of operational frequencies is 154,4–163,75 MHz) and IBM PC with Creative Audigy 4 soundcards were used. Sequentially AFC of the first and second orders were defined. The method of identification with number of experiments  $N=4$  was applied. Structure charts

of identification procedure – determinations of the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> order AFC of CC are presented accordingly in figs. 3–5. The general scheme of a hardware–software complex of the CC identification, based on the data of input–output type experiment is presented in fig. 20.

The CC received responses  $y[a_i x(t)]$  to the test signals  $a_i x(t)$ , compose a group of the signals, which amount is equal to the used number of experiments  $N$  ( $N=4$ ), shown in fig. 21. In each following group the signals frequency increases by magnitude of chosen step. A cross-correlation was used to define the beginning of each received response.

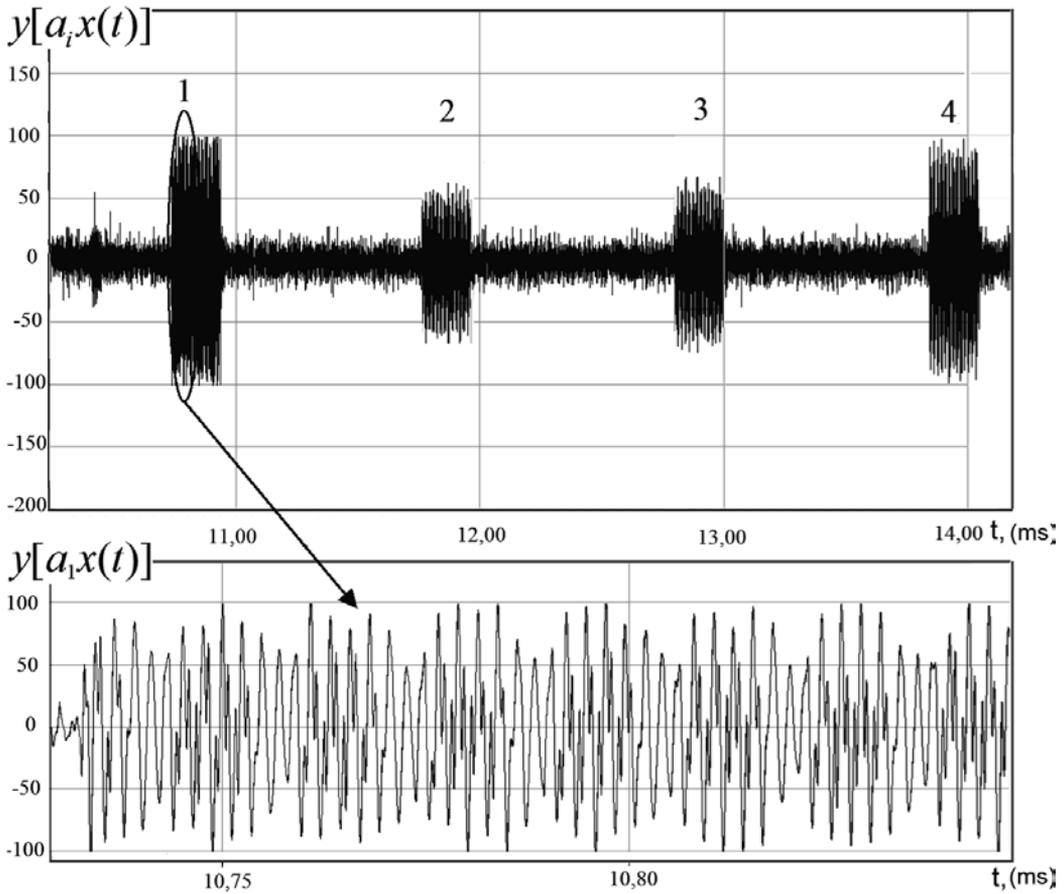


**Figure 20.** The general scheme of the experiment

Maximum allowed amplitude in described experiment with use of sound card was  $A=0,25V$  (defined experimentally). The range of frequencies was defined by the sound card pass band (20...20000 Hz), and frequencies of the test signals has been chosen from this range, taking into account restrictions specified above. Such parameters were chosen for the experiment: start frequency  $f_s=125$  Hz; final frequency  $f_e=3125$  Hz; a frequency change step  $F=125$  Hz; to define AFC of the second order determination, an offset on frequency  $F_1=f_2-f_1$  was increasingly growing from 201 to 3401 Hz with step 100 Hz.

The weighed sum is formed from received signals – responses of each group (figs. 3–5). As a result we get partial components of the response of the CC  $y_1(t)$  and  $y_2(t)$ . For each partial component of the response a Fourier transform (the Fast Fourier Transform is used) is calculated. Only informative harmonics (which amplitudes represents values of required characteristics of the first, second and third order AFC) are taken from received spectrum.

The first order amplitude-frequency characteristic  $|W_1(j2\pi f)|$  is received by extracting the harmonics with frequency  $f$  from the spectrum of the partial response of the CC  $y_1(t)$  to the test signal  $x(t)=(A/2)\cos 2\pi ft$ .



**Figure 21.** The group of signals received from CC with amplitudes: -1 (1); -1/2 (2); 1/2 (3); 1 (4);  $N=4$

The second order AFC  $|W_2(j2\pi f_1, j2\pi f_2)|$ , where  $f_1=f$  and  $f_2=f+F_1$  was received by extracting the harmonics with summary frequency  $f_1+f_2$  from the spectrum of the partial response of the CC  $y_2(t)$  to the test signal  $x(t)=(A/2)(\cos 2\pi f_1 t + \cos 2\pi f_2 t)$ .

The third order AFC  $|W_3(j2\pi f_1, j2\pi f_2, j2\pi f_3)|$ , where  $f_1=f$ ,  $f_2=f+F_1$  and  $f_3=127,5$  Hz were received by extracting the harmonics with summary frequency  $f_1+f_2+f_3$  from the spectrum of the partial response of the CC  $y_3(t)$  to the test signal  $x(t)=(A/2)(\cos 2\pi f_1 t + \cos 2\pi f_2 t + \cos 2\pi f_3 t)$ .

The wavelet noise-suppression was used to smooth the output data of the experiment [9]. The results received after digital data processing of the data of experiments (wavelet “Coiflet” denoising) for the first, second and third order AFC are presented in fig. 22–25.

The surfaces shown in fig. 24-25 are built from sub-diagonal sections that are received separately. We used  $F_1$  as growing parameter of identification with different value for each section.

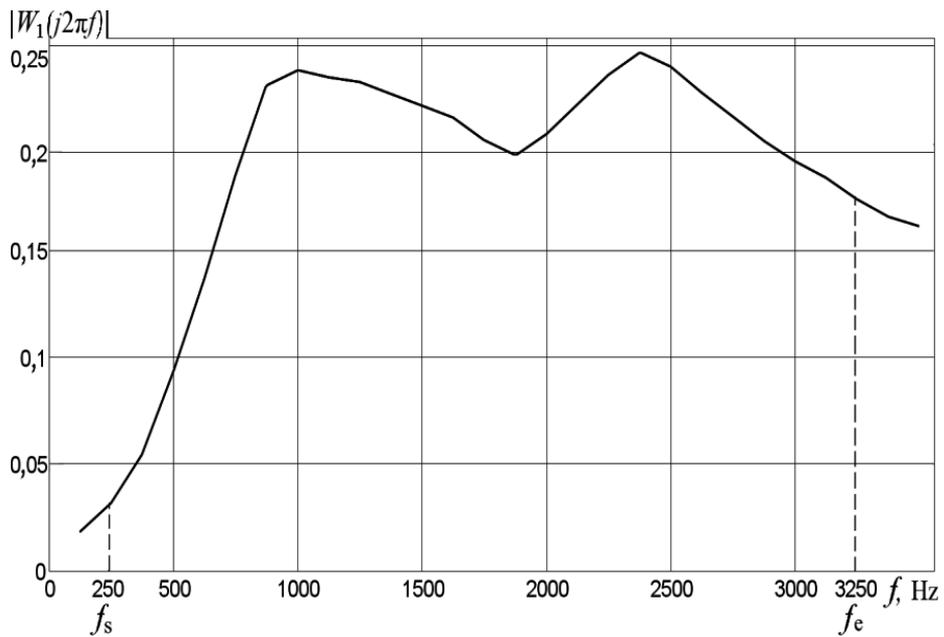


Figure 22. AFC of the first order after wavelet "Coiflet" 2<sup>nd</sup> level noise-suppression

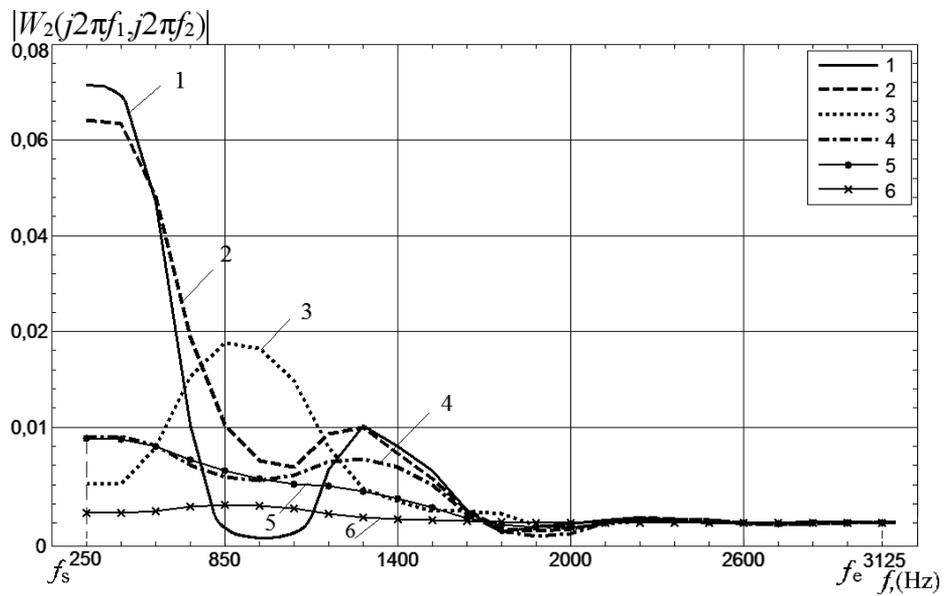


Figure 23. Subdiagonal sections of AFCs of the second order after wavelet "Coiflet" 2<sup>nd</sup> level noise-suppression at different frequencies  $F_1$ : 201 (1), 401 (2), 601 (3), 801 (4), 1001 (5), 1401 (6) Hz

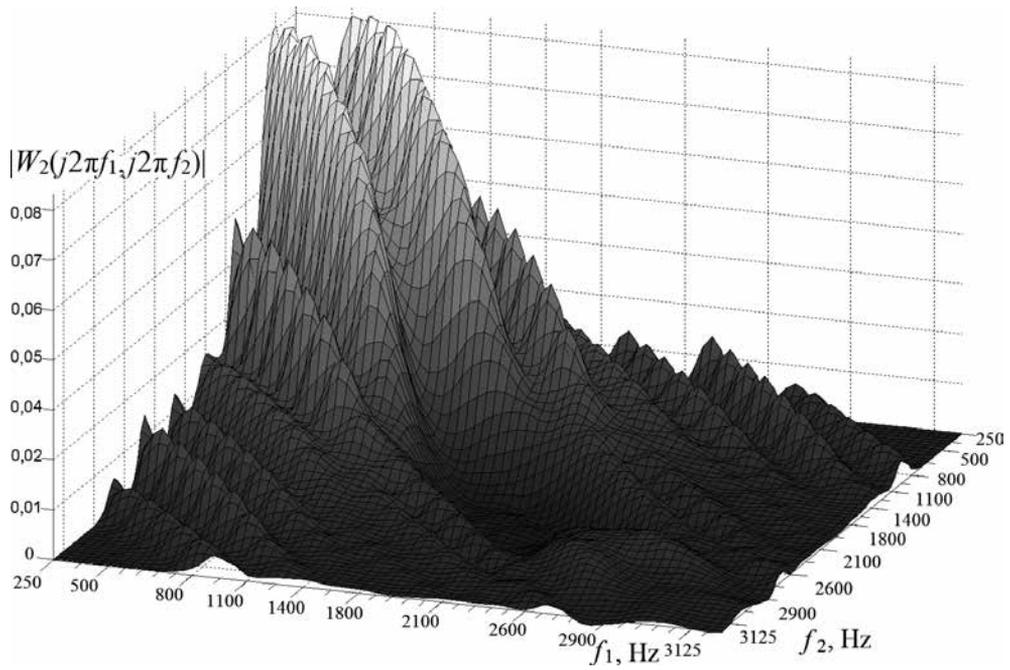


Figure 24. Surface built of AFCs of the second order after wavelet "Coiflet" 3<sup>rd</sup> level noise-suppression

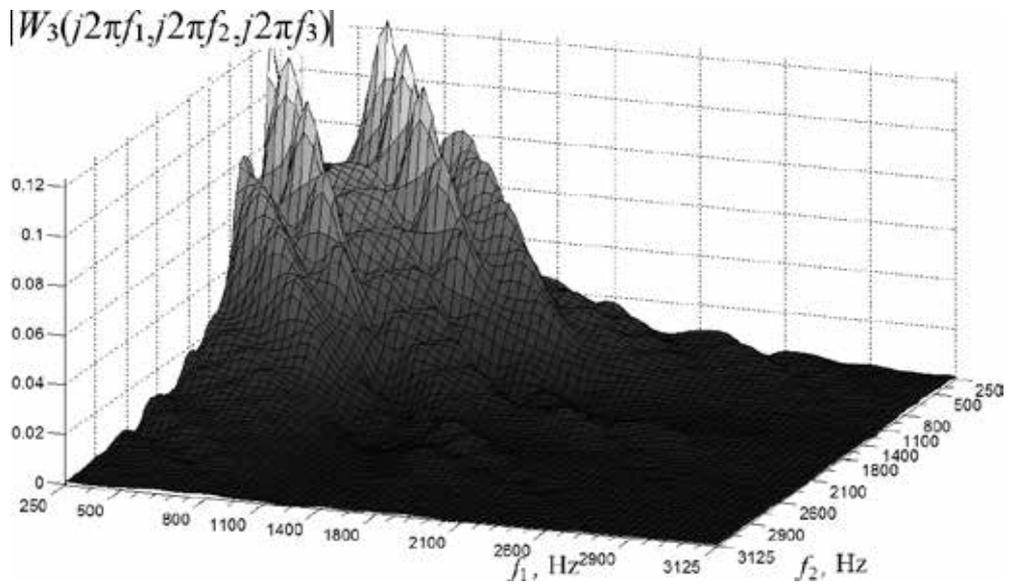


Figure 25. Surface built of AFCs of the third order after wavelet "Coiflet" 3<sup>rd</sup> level noise-suppression, where  $f_3=127,5$  Hz

## 6. Conclusion

Communication channel as a media for remote sensing systems functioning is analyzed. Nonlinear effects of the environments have great impact on result data received in experiments. The method based on Volterra model using polyharmonic test signals for identification nonlinear dynamical systems is analyzed. To differentiate the responses of system for partial components we use the method based on linear combination of responses on test signals with different amplitudes.

New values of test signals amplitudes were defined and they are greatly raising the accuracy of identification compared to amplitudes and coefficients written in [1]. The accuracy of identification of nonlinear part of the test system growth 2 times and the standard deviation in this case is about 5%.

The interpolation method of identification using the hardware methodology written in [5] is applied for construction of informational Volterra model as an APC of the first and second order for UHF band radio channel.

Received results reveal essential nonlinearity of the CC that leads to distortions of signals in radio broadcasting devices, reduces the important indicators of the TCS: accuracy of signals reproduction, throughput, noise immunity.

The noise immunity is very high for the linear model, high enough for the second order nonlinear model and has moderate noise immunity for the third order model. The wavelet denoising is very effective and gives the possibility to improve the quality of identification of the noisy measurements up to 1,54 and 4,07 times for the AFC and PFC respectively.

Final characteristics of the CC have to be used to maintain sensor systems to improve the adequateness of received data.

## Author details

Vitaliy Pavlenko and Viktor Speranskyy

Institute of Computer Systems, Odessa National Polytechnic University, Odessa, Ukraine

## References

- [1] Danilov LV, Mathanov PN, Philipov ES. The theory of nonlinear electrical circuits. Published Energoatomizdat, Leningrad; 1990.

- [2] Donoho DL, Johnstone IM. Threshold selection for wavelet shrinkage of noisy data. Proc. 16th Annual Conf. of the IEEE Engineering in Medicine and Biology Society, 24a–25a, IEEE Press; 1994.
- [3] Doyle FJ, Pearson RK, Ogunnaike BA. Identification and Control Using Volterra Models. Published Springer Technology & Industrial Arts; 2001.
- [4] Giannakis GB, Serpedin EA. bibliography on nonlinear system identification and its applications in signal processing, communications and biomedical engineering. Signal Processing, EURASIP, Elsevier Science B.V. 81(3); 2001. p533-580.
- [5] Goswami JG, Chan AK. Fundamentals of Wavelets: Theory, Algorithms, and Applications. Publishing John Wiley&Sons Inc; 1999.
- [6] Liew SC. Principles of remote sensing. Space View of Asia, 2nd Edition. CRISP; 2001. <http://www.crisp.nus.edu.sg/~research/tutorial/rsmain.htm> (accessed 1 November 2013).
- [7] Misiti M, Misiti Y, Oppenheim G, Poggi J-M. Wavelets Toolbox Users Guide. The MathWorks. Wavelet Toolbox, for use with MATLAB; 2000.
- [8] Marghany M. Volterra-Lax-wendroff algorithm for modelling sea surface flow pattern from Jason-1 satellite altimeter data. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) Volume 5730 LNCS; 2009; p1-18.
- [9] Marghany M, Mazlan H, Cracknell AP. 3-D visualizations of coastal bathymetry by utilization of airborne TOPSAR polarized data. International Journal of Digital Earth; 3(2); p187–206.
- [10] Marghany M. Three-Dimensional Coastal Front Visualization from RADARSAT-1 SAR Satellite Data. In Murgante B. et al. (eds.): Lecture Notes in Computer Science (ICCSA 2012); Part III, LNCS 7335; p447–456.
- [11] Pavlenko V, Speranskyy V, Ilyin V, Lomovoy V. Modified Approximation Method for Identification of Nonlinear Systems Using Volterra Models in Frequency Domain. Applied Mathematics in Electrical and Computer Engineering. Proceedings of the AMERICAN–MATH'12 & CSST'12 & CEA'12, Harvard, Cambridge, USA, January 25–27, 2012. Published by WSEAS Press; 2012. p423–428.
- [12] Pavlenko VD, Pavlenko SV, Speranskyy VO. Interpolation Method of Nonlinear Dynamical Systems Identification Based on Volterra Model in Frequency Domain. Proceedings of the 7th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2013), 15–17 September 2013, Berlin, Germany; 2013. p173–178.
- [13] Pavlenko VD, Speranskyy VA. Analysis of identification accuracy of nonlinear system based on Volterra model in frequency domain. American Journal of Modeling and Optimization; Vol.1, No.2; 2013. p11–18. DOI: 10.12691/ajmo-1-2-2.

- [14] Pavlenko VD, Speranskyy VO. Communication Channel Identification in Frequency Domain Based on the Volterra Model. Recent Advances in Computers, Communications, Applied Social Science and Mathematics. Proceedings of the International Conference on Computers, Digital Communications and Computing (ICDCC'11), Barcelona, Spain, September 15-17, 2011. Published by WSEAS Press; 2011. p218–222.
- [15] Pavlenko VD, Speranskyy VO. Interpolation method modification for nonlinear objects identification using Volterra model in frequency domain. 23rd International Crimean Conference “Microwave & Telecommunication Technology” (CriMiCo'2013), Sevastopol, Ukraine; 2013. p257-260.
- [16] Pavlenko VD, Speranskyy VO, Lomovoy VI. Modelling of Radio–Frequency Communication Channels Using Volterra Model. Proc. of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2011), 15–17 September 2011, Prague, Czech Republic; 2011. p574–579.
- [17] Pavlenko VD, Speranskyy VO, Lomovoy VI. The Test Method for Identification of Radiofrequency Wireless Communication Channels Using Volterra Model. Proc. of the 9th IEEE East–West Design & Test Symposium (EWDTS'2011), Sevastopol, Ukraine, September 9–12, 2011; Kharkov: KNURE; 2011. p331–334.
- [18] Pavlenko V, Lomovoy V, Speranskyy V, Ilyin V. Radio frequency test method for wireless communications using Volterra model. Proc. of the 11th conference on dynamical systems theory and applications (DSTA'2011), December 5–8, 2011, Łódź, Poland; 2011. p446–452.
- [19] Pavlenko VD, Speranskyy VO. Analysis of nonlinear system identification accuracy based on Volterra model in frequency domain. *Electrotechnic and Computer Systems*. Kiev: «Technica»; 08(84); 2012. p66–71.
- [20] Pavlenko VD, Speranskyy VO. Identification of nonlinear dynamical systems using Volterra model with interpolation method in frequency domain. *Electrotechnic and Computer Systems*. Kiev: «Technica»; 2012. 05(81). p229–234.
- [21] Pavlenko VD, Speranskyy VO. Simulation of Telecommunication Channel Using Volterra Model in Frequency Domain. 10th IEEE East–West Design & Test Symposium (EWDTS'2012), Kharkov, Ukraine, September 14–17, 2012; 2012. p401–404.
- [22] Schetzen M. *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley&Sons, New York; 1980.
- [23] Westwick DT. *Methods for the Identification of Multiple–Input Nonlinear Systems*. Departments of Electrical Engineering and Biomedical Engineering, McGill University, Montreal, Quebec, Canada; 1995.



---

# Environment Remote Sensing Investigations

---



---

# Statistical Characterization of Bare Soil Surface Microrelief

---

Edwige Vannier, Odile Taconet,  
Richard Dusséaux and Olivier Chimi-Chiadjeu

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/57244>

---

## 1. Introduction

Because the soil surface occurs at the boundary between the atmosphere and the pedosphere, it plays an important role for geomorphologic processes. Roughness of soil surface is a key parameter to understand soil properties and physical processes related to substrate movement, water infiltration or runoff, and soil erosion. It has been noted by many authors that most of the soil surface and water interaction processes have characteristic lengths in millimeter scales. Soil irregularities at small scale, such as aggregates, clods and interrill depressions, influence water outflow and infiltration rate. They undergo rapid changes caused by farming implements, followed by a slow evolution due to rainfall events. Another objective of soil surface roughness study is investigating the effects of different tillage implements on soil physical properties (friability, compaction, fragmentation and water content) to obtain an optimal crop emergence. Seedbed preparation focuses on the creation of fine aggregates and the size distribution of aggregates and clods produced by tillage operations is frequently measured.

Active microwave remote sensing allows potential monitoring of soil surface roughness or moisture retrieving at field scale using space-based Synthetic Aperture Radars (SAR) with high spatial resolution (metric or decametric). The scattering of microwaves depends on several surface characteristics as well as on imagery configuration. The SAR signal is very sensitive to soil surface irregularities and structures (clod arrangement, furrows) and moisture content in the first few centimeters of soil (depending on the radar wavelength). In order to link the remote sensing observations to scattering physical models as well as for modelling purpose, key features of the soil microtopography should be characterized. However, this characterization is not fully understood and some dispersion of roughness parameters can be observed in the same field according to the methodology used. It seems also, that when describing

surface roughness as a whole, some information related to structured elements of the microtopography is lost.

## **2. Data SAR – Electromagnetic models and soil modelling: Position of problem**

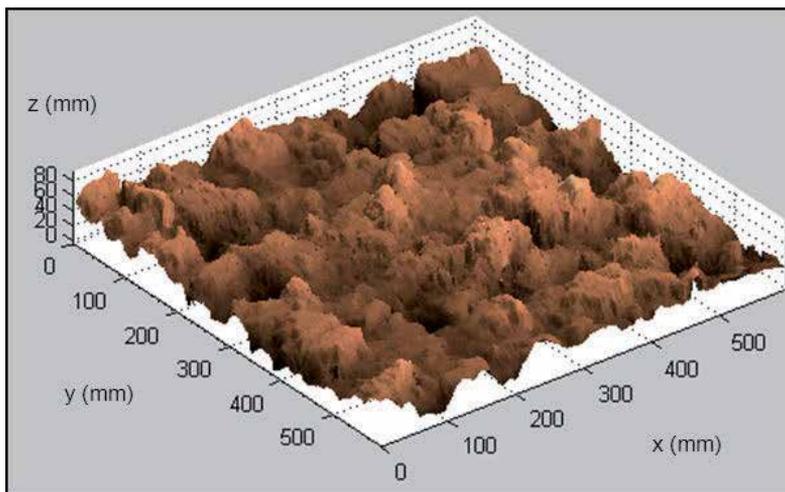
Synthetic Aperture Radars allow the study of agricultural bare soils by measuring the backscattered coefficient. The backscattered signal depends on the surface roughness, on the electrical permittivity which is closely linked to the soil moisture and texture. It varies depending on the frequency and polarization of the transmitted electromagnetic wave and on the incidence angle [1-4]. The electromagnetic modeling is a valuable tool for radar data inversion to intend to characterize the roughness and / or the humidity of agricultural soils. The electromagnetic models based on the Maxwell's equations and the boundary conditions can be classified into two categories: analytical models [5] and numerical models [6-9].

The analytical models are based on physical approximations which reduce the applicability domain. But, these models have an obvious interest: the electromagnetic field scattered by the surface and the coherent and incoherent intensities are given by analytical formulas arousing physical interpretations. The first-order small-perturbation method (SPM1) is only valid for surfaces with small roughness versus the wavelength [10] and the Kirchhoff approximation (KA) is applicable to surfaces with long correlation length [11-12]. The small slope approximation (SSA1) has an extended domain of applicability which includes the domains of the two previous approaches [13]. It is likewise for the integral equation model (IEM) which became the most quoted and implemented rough surface scattering model in the field of radar remote sensing for earth observation [14]. These four analytical models assume that the soil can be represented by a stationary random process. The stationarity is not clearly established for agricultural soils, especially in the presence of very marked furrows [15]. These models require the knowledge of the autocorrelation function. In addition, the models KA, SSA and IEM require the one-point and two-point height probability distributions. The calculations are typically conducted with the Gaussian probability density function. The Gaussian character is observed for seedbeds with furrows slightly marked but to our knowledge, it is not the case for ploughed soils [9, 16]. Furthermore, tilled agricultural soils are anisotropic surfaces and we have to take into account the quasi-periodicity of the surface in the estimation of the coherent and incoherent components of the backscattered signal [17]. Fractal and multi-fractal approaches for describing agricultural soils were also implemented. Some works allowed extending electromagnetic analytical models to the fractal character of soils [18-19]. So, the remote sensing studies using these analytical models consider a global description of the scattering surface and do not take into account their structuring objects such as soil clods, holes and aggregates [20]. Then, for the analysis of radar data, it is essential to extend the analytical models by taking into account these structuring objects which needs the understanding of their statistical properties and their influence on the autocorrelation function.

The second class of electromagnetic models relies on numerical methods for solving Maxwell's equations and boundary conditions [6-8]. These models are so called exact if no physical approximation is made. These models do not provide an analytical solution of the problem and require high computational times. This is a major drawback for the inversion of radar data. Advanced numerical methods have been developed over the past two decades to reduce computational time [21]. For these numerical methods, the scattered intensities and the backscatter coefficient are estimated over results of a finite set of surface realizations. The estimation error depends on the number and size of surfaces. These numerical models associated with the Monte Carlo method rely on surface generating algorithms. Agricultural bare soils have structuring objects contributing to the backscattered signal. Except for a few references [22], the surface generators are based on a global approach and do not take into account the objects characterizing agricultural soils. To improve the analysis of radar data, it makes sense to move forward in the statistical description of the structuring objects, to develop new surface generating algorithm in accordance with the statistical properties of these objects and to interface them with numerical electromagnetic models.

### 3. Statistical and spatial analysis of soil roughness

Several types of surface roughness may be recognized from micro relief variations to larger level variations representing the landscape. At millimetre or centimetre-scales, soil surface roughness mainly results from the breakdown of the superficial soil layer by tillage operations. Depending on the soil properties, in particular texture, organic matter content and moisture state of the soil material, and on the tillage tool, different aggregate sizes will be produced (figure 1).



**Figure 1.** Microtopography of a seedbed surface

### 3.1. Background

Two approaches are commonly investigated for statistical and spatial analysis of soil roughness at these scales. In the first one, the surface roughness is characterized as a whole by its autocorrelation function model or by its variogram and is summarized by statistical indices: root mean square of heights, correlation length, roughness exponent or Hurst coefficient, parameter  $Z_s$ , tortuosity or other specific indices [9, 15, 23-31]. Fractal and multifractal description are investigated too [32-34]. In the second one, local and morphological aspects of the surface are brought into focus such as aggregates, clods and mound-and-depression patterns [20, 28, 35-43].

Direct ground field measurement of surface roughness can be performed by either contact techniques [44-46] or non-contact techniques such as close range photogrammetry [15, 26, 47-50] or laser scanning [32, 51-53]. Some authors made comparative studies [30, 54-56] and a review of the measurement techniques can be found in [3]. Although close range photogrammetry and laser scanning enable retrieving a 3D digital elevation model (DEM) of the soil surface, many authors still carry on characterizing soil surface roughness by extracting 1D profiles on the 3D measurement [50, 57-58]. This can be explained by the fact that 1D statistical indices are widely spread and serve as reference, since measurement of 1D profiles was developed in the first place. Nevertheless, it traduces also a lack of microrelief modelling. In the field of soil moisture retrieval from SAR imagery, Lievens and Verhoest, [59], are led to propose methods to circumvent surface roughness parameterization problems.

Blaes and Defourny [30] and Dusséaux et al. [9] suggest using a bidimensional autocorrelation function to characterize bare soil agricultural surfaces. Bidimensional approach is more appropriate to the measurement of roughness anisotropy related to multiscale processes (row structured field, clod arrangement). In order to improve statistical characterization of bare soil surface microrelief, we highly focus on 3D DEM interpretation study. The proposed approach is detecting and localizing clods and big aggregates by segmentation methods, modelling them by a simplified geometric form, estimating the statistical distribution of the modelling structuring object parameters and then generate numerical surfaces by setting structuring objects onto a substrate according to statistical laws.

### 3.2. 3D-DEM interpretation

Millimetric DEMs of the soil surface were seldom used for detecting and localizing big aggregates and clods. In 3D, an individual clod (or big aggregate) is seen as a local bump standing out with slightly sharp boundaries (figure 2). It is a recognizable structure presenting rather medium or high level values, with a high gradient on the edges, but inside color level values are non-homogenous. Two newly built methods have been presented [41- 42, 60].

Addressing the problem of detecting and localizing clods requires also an evaluation methodology. We propose several qualification tools based on the knowledge of a ground truth serving as reference. The detection performance can be evaluated by estimating the sensitivity *sen* and specificity *spe*, defined as follows:

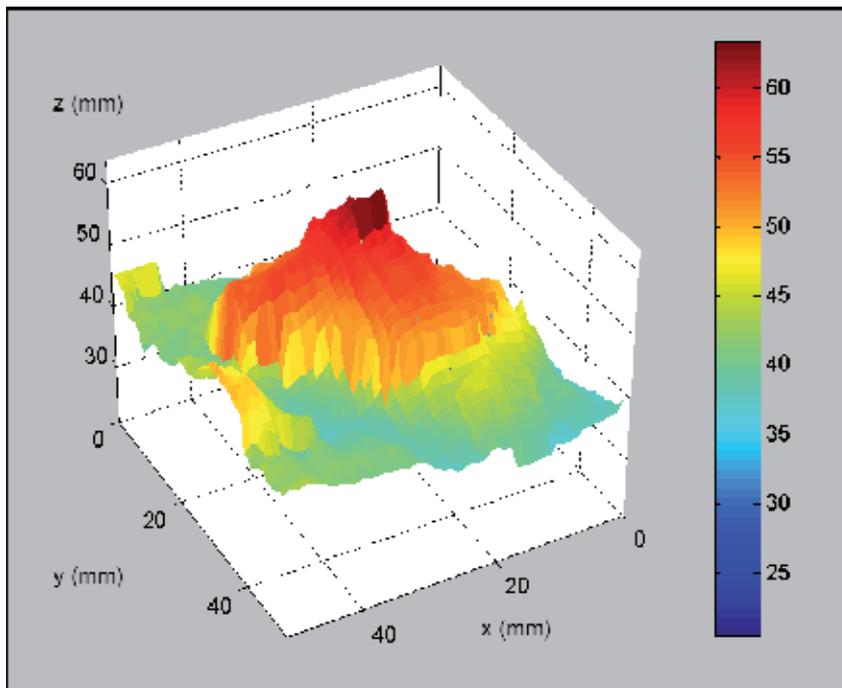
$$sen = 1 - f_n \tag{1}$$

where  $f_n = \frac{\text{number of not detected clods}}{\text{number of assessed clods}}$

$$spe = 1 - f_p \tag{2}$$

where  $f_p = \frac{\text{number of erroneously detected clods}}{\text{number of clods detected by the method}}$

$f_n$  and  $f_p$  represent the rates of respectively misdetections and false alarms.



**Figure 2.** Clod representation in 3D

There is usually a compromise between sensitivity and specificity for a detection method. Setting of the method parameters can be made by constructing the receiver operating curve (ROC), eventually introducing weight reflecting the cost of a bad decision [61].

The localization performance can be estimated by computing the overlap rates between each detected clods and matched reference clods:

$$O_v(i) = A(I_i \cap I_{ti}) / A(I_i \cup I_{ti}) \quad (3)$$

where  $I_i$  is the set of pixels located inside the identified boundaries of the clod  $i$ ,  $I_{ti}$  the set of pixels located inside the true boundaries of the clod  $i$ , and  $A(I_i)$  the cardinal of  $I_i$ .  $O_v$  is null either when a true clod is not identified or when a false positive has been detected. The ideal case corresponds to an overlap rate equal to 1 for all the clods. This ideal case leads to a cumulative distribution function (cdf) equal to 0 for an overlap rate strictly lower than 1 and equal to 1 for an overlap rate upper or equal to 1 [60].

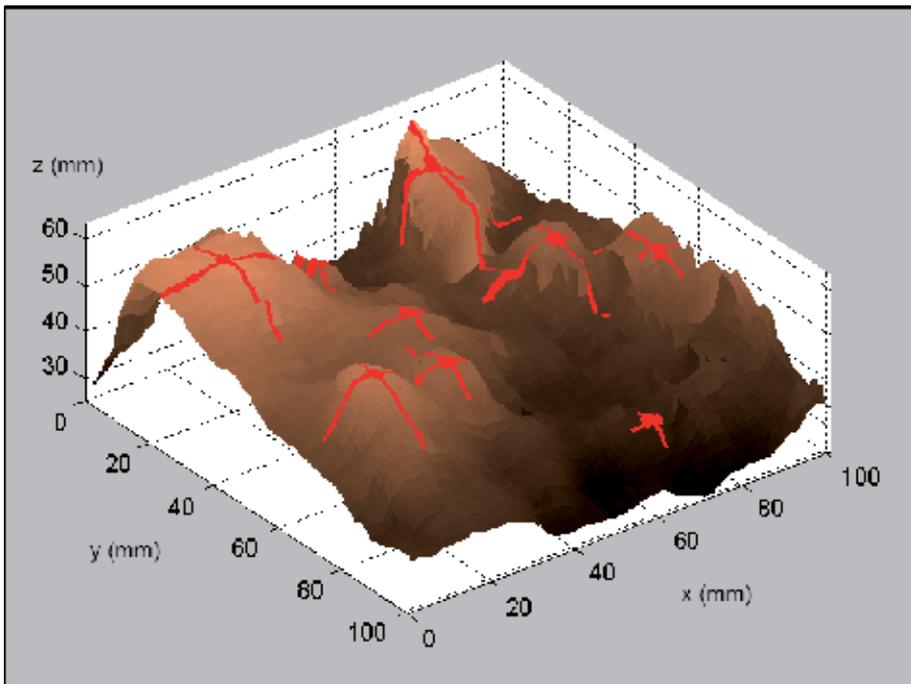
Now, in [41], clod identification is obtained by merging information on local maxima enhanced at different scales by wavelet transform. Clods are successfully identified by their summit and two diameters on several kinds of soil surfaces. Boundary points are estimated by thresholding the height slope. The detection performance was evaluated with the help of a soil scientist on a controlled surface made in the laboratory as well as on real seedbed and ploughed surfaces, made by tillage operations in an agricultural field. A minimum size, set here to 7 mm of diameter depending on the DEM horizontal resolution of 1mm, is required for aggregates, in order to prevent a high false positive rate. Indeed, one has to be able to count the aggregates. The identifications of the clods were in good agreement, with an overall sensitivity of 84% and a specificity of 94%. A limitation of this method is that it does not provide the clod boundaries. An example of clods identification on a part of seedbed surface is provided on figure 3. We can see that most of the mounds have been identified. A small one, with center coordinates located near (50, 15) has not been detected.

In order to get the clods boundaries, a contour-based method has been proposed by Taconet et al. [42]. The clod boundary is determined by hierarchy of closed elevation contours going through the pixels of highest gradient values. This second method gives satisfactory results with a clod identification sensitivity of 75% and specificity of 95%. Figure 4 displays the clods localization obtained in the same part of seedbed as on figure 3.

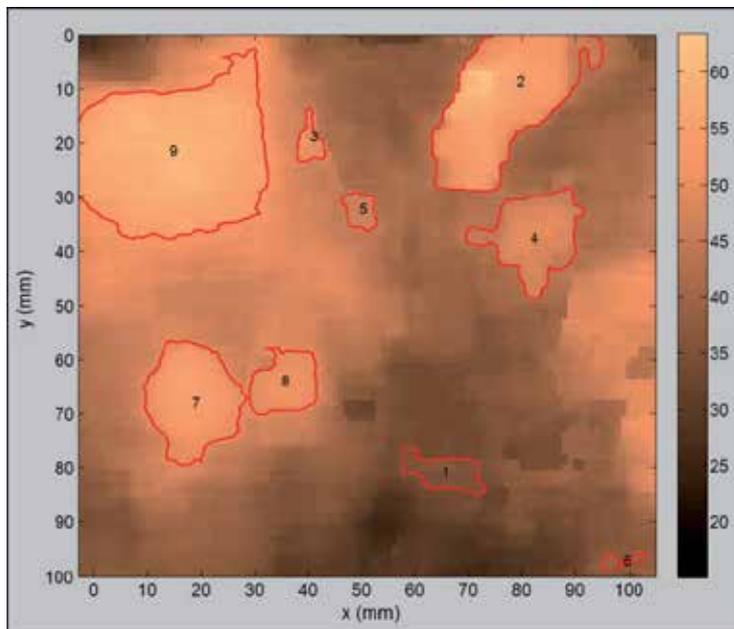
There is a good agreement with the identification of clods by wavelet-based method. The small clod that was not detected is now identified (clod number 5 in figure 4), but two clods are not detected anymore, illustrating the lower sensitivity of the contour-based method.

The contour-based method has been tested until now, on freshly tilled seedbed and artificial soils which both contain mostly distinct clods and aggregates. Furthermore, defining clod boundary as an elevation contour requires that the substrate of clods has little variation, which is not the case of ploughed surfaces. Indeed, the automatically retrieved clod boundaries are generally underestimated and lay at a constant height, which is not very realistic.

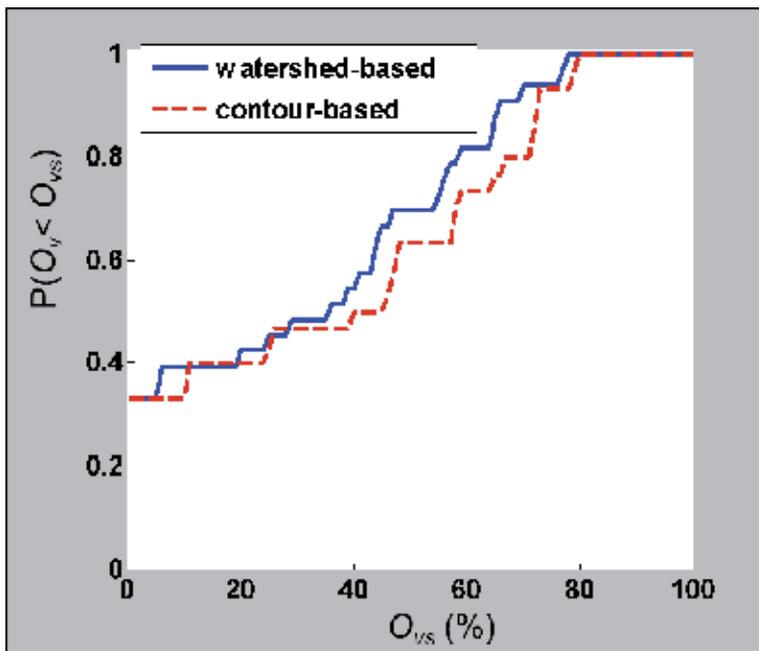
Another approach is developing a more classical method of image segmentation based on mathematical morphology, the watershed-based method [62]. Such a method has been introduced in [63]. It relies on a transformation of the DEM image in order to produce a pseudo



**Figure 3.** Automatic localization of clods by wavelet-based method



**Figure 4.** Automatic localization of clods by contour-based method



**Figure 5.** Cdf of overlap rates for watershed- and contour-based segmentation methods

elevation image of greater dynamic to enhance the clods and on mixing information on heights and gradients so that the local minima correspond to clods bases. Figure 5 shows the cdf of overlap rates estimated for the contour-based and the watershed-based methods on a seedbed DEM.

The cdfs have the same value at the origin, showing that both methods have the same rates of false detected and not detected clods. The clods localization is all the better than the cdf is low. The watershed-based method cdf being higher than the contour-based one, it shows that, for this surface, the classical image segmentation method does not perform better than the newly built method. However, it does not have the drawback of horizontal clods bases. Thus identifying clods and big aggregates on a millimeter DEM remains a complex problem. Also, a specific approach of contour moving based on the minimization of a cost function related to clods detection properties has been introduced [64]. Based on the characterization of clods, as recognizable structures presenting rather medium or high level values, with a high gradient on the edges and low height values on the boundary, we have defined four criteria, linked to heights and gradients again, that characterize the clod boundary. Contrary to [42], the hypothesis of a horizontal base does not exist anymore. The minimization is performed by a meta-heuristic inspired from the simulated annealing algorithm [65]. Let us notice that each identification method can be used to initialize the contour moving algorithm, however, until now, the results are better if the initial contours lay inside the reference boundaries. The overlap rates can be enhanced by up to 10%.

## 4. Modelling of soil characteristics

As clods and aggregates are irregular shaped objects, a way to define relevant indices in size and shape is representing them by a simple approached form which matches closely their horizontal and vertical extents. In numerical generation of soil surfaces, clods and aggregates are usually modelled by half circles [22] or half spheres [38] or part of spheroids [66] put onto a substrate which may be a plane surface, an exponentially correlated random surface, or a surface representing furrows. In [66], the spheroids are equal-sized and regularly placed on a square grid. In [38], an envelope-Boolean surface generation model was introduced, where half spheres were set onto a plane according to the desired statistic. Two artificial surfaces were manually created by distributing aggregates randomly on a plane. Then structuring elements were generated by a Poisson point process reproducing the occurrence of aggregates size classes. It was suggested that the use of other primary function than half-sphere, like ellipsoids, could improve the performance of the surface generation process. In [22], the soil surface is described as a substrate, modelled by a 1D profile of Gaussian distribution of heights, with an exponential autocorrelation, and clods, modelled by half circles, randomly and independently set onto the substrate. Size of clods and distance between them obey to statistical laws that are not estimated from observations.

The choice of a modelling shape is dependent on the type of soil considered in an experiment. In [67], soil compaction extent is described by a half-ellipse. In [20], it is proposed to fit an ellipse to clod horizontal boundary contour and a half cosine function to the height shape. A semi-ellipsoid can also be suitable to fit the height shape of clods and aggregates. Mathematical frame of these modelling is presented hereafter.

### 4.1. Equivalent ellipse modelling clod base contour

The equivalent ellipse denotes the best fitting ellipse of a plane closed contour. Let O be the barycentre of the considered clod, R = (Oxy) its barycentric reference and Re = (OXY) the inertia principal reference. Estimation of the three shape parameters (semi-major axis length  $a$ , semi-minor axis length  $b$  and orientation angle  $\theta$ , (i.e. the angle between the major axis supported by (OX) and the horizontal axis supported by (Ox)) is based on the method of geometric moments [68]. The central second order moments in the barycentric reference, where the summation extends over all the pixels located within the clod boundary contour are given by:

$$I_{Ox} = \frac{1}{N} \sum_{i=1}^N y_i^2, I_{Oy} = \frac{1}{N} \sum_{i=1}^N x_i^2 \text{ and } I_{xy} = \frac{1}{N} \sum_{i=1}^N y_i x_i \quad (4)$$

in the barycentric reference, where the summation extends over all the pixels located within the clod boundary contour. Using these geometric moments, the length of the semi-axes ( $a$ ,  $b$ ) and the orientation angle  $\theta$  are calculated as follows:

$$\theta = \frac{1}{2} \tan^{-1} \left( \frac{2I_{xy}}{I_{Oy} - I_{Ox}} \right) \pm k \frac{\pi}{2} \quad (k \in N) \theta \quad (5)$$

where only one orientation angle is kept so that  $\theta$  varies within  $]-\pi/2, \pi/2[$  interval, and:

$$m_1 = 2\sqrt{I_{Ox} \cos^2 \theta + I_{Oy} \sin^2 \theta - I_{xy} \sin 2\theta} \quad (6)$$

$$m_2 = 2\sqrt{I_{Ox} \sin^2 \theta + I_{Oy} \cos^2 \theta + I_{xy} \sin 2\theta} \quad (7)$$

$$a = \max(m_1, m_2) \quad (8)$$

$$b = \min(m_1, m_2) \quad (9)$$

#### 4.2. Semi-arch of cosine modelling clod height shape

The height of the semi-arch of cosine  $h_c$  is estimated by least squares minimization of the sum of squared residuals:

$$\varepsilon = \sum_{i=1}^{N_c} \left[ z(x_i, y_i) - h_b - h_c \cos \left( \frac{\pi}{2} \sqrt{\frac{x_i^2}{a^2} + \frac{y_i^2}{b^2}} \right) \right]^2 \quad (10)$$

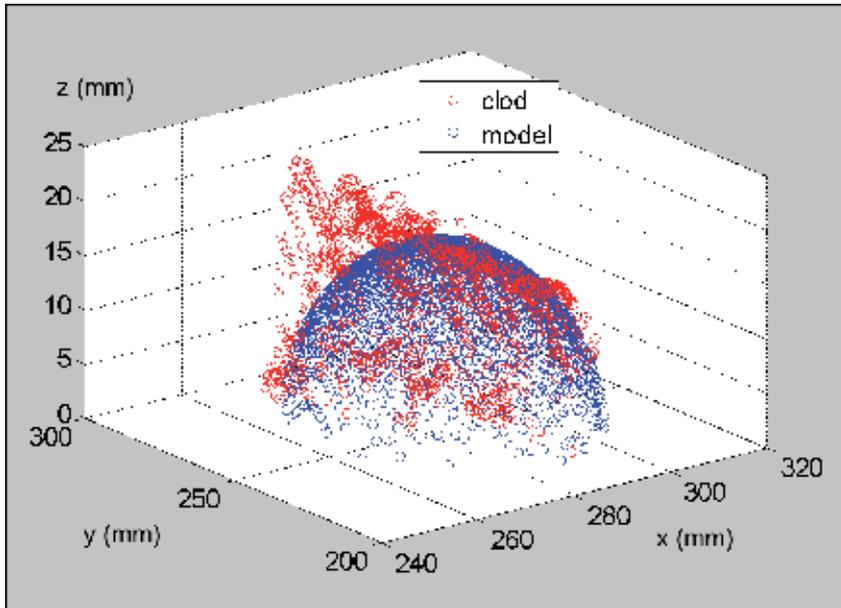
where  $z$ , is the elevation at point  $(x_i, y_i)$  in barycentric reference, and  $h_b$  the elevation of the clod base contour.

#### 4.3. Half ellipsoid modelling clod height shape

Parametric equation of half ellipsoid is:

$$\begin{cases} x = a \cos(v) \cos(u) \\ y = b \cos(v) \sin(u) \\ z = h \sin(v) \end{cases} \quad (11)$$

with  $u \in ]-\pi, \pi[$  and  $v \in [0, \pi/2[$ ,  $a$  and  $b$  semi axes of the equivalent ellipse and  $h$ , height of the half ellipsoid, that can be estimated by identifying the sum of squares of clod heights and the sum of squares of model heights, in order to keep the second order moment. An example of clod modelling is illustrated on figure 6.



**Figure 6.** Clod modelling by a half ellipsoid

Extracting the identified clods and setting them onto a plane gives a surface with mainly the same autocorrelation function as the same surface where clods are replaced by their half ellipsoid model (see figure 7). This illustrates the goodness of fit, when modelling clods and big aggregates by half ellipsoids. Let us notice that the finite length of surfaces causes oscillations in the autocorrelation functions.

#### 4.4. Generation process

Analysis of DEM plots of a soil surface allows deriving the statistical distributions characterizing structuring object parameters. As an example, when analyzing a seedbed surface containing 350 clods, we find that the centers of gravity  $G$  and the orientation angles  $\theta$  of big aggregates and clods can be assumed uniformly distributed. With the ellipsoid model, the variables  $a$ ,  $b$  and  $h$  are not independent. Therefore intermediate variables have to be introduced, such as horizontal and vertical compression factors. They can be defined as follows:

$$f_h = \frac{b}{a} \tag{12}$$

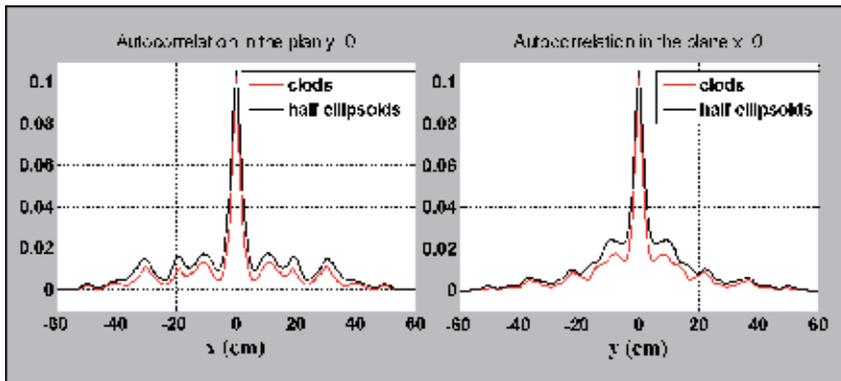


Figure 7. Autocorrelation functions of clods set onto a plane and fitted

$$f_v = \frac{h}{a} \tag{13}$$

Using Pearson’s chi-squared tests of independence, we found a plausible independence of  $a$  and  $f_h$  and of  $a$  and  $f_v$ . Then,  $b$  and  $h$  would be derived from  $f_h$  and  $f_v$  using equations (12) and (13).

Generating realistic numerical surfaces is expected for soil science studies as well as for numerical electromagnetic codes used in remote sensing studies. In order to generate a cloddy surface, several parameters have to be estimated:

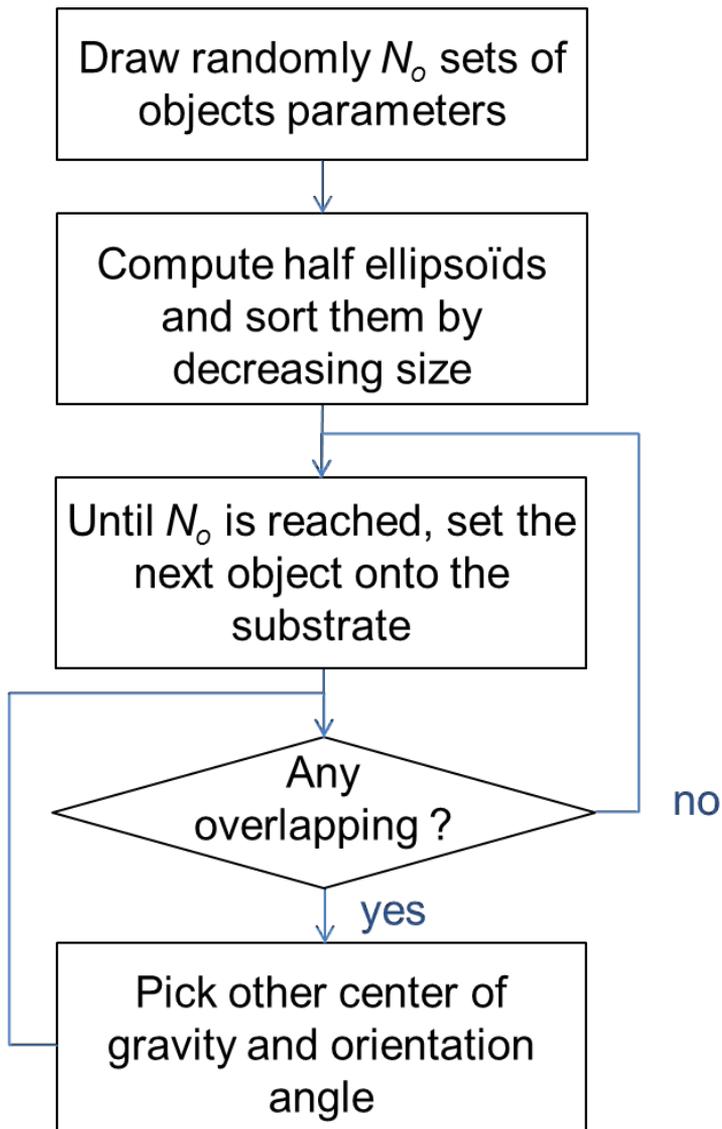
- Number of objects (according to the desired density)  $N_o$ ;
- Probability density functions (pdf) of  $a, f_h$  and  $f_v, \theta$  and  $G$ ;
- Shape of the substrate, where to put the structuring objects onto.

Then, the proposed generation process is composed of three main steps (see figure 8).

First step is drawing randomly  $N_o$  sets of objects parameters. Variables  $a, f_h$  and  $f_v$  as well as  $\theta$  and  $G$  are drawn according to their pdf. Variables  $b$  and  $h$  are deduced from  $f_h$  and  $f_v$  using equations (10) and (11). Second step is computing half ellipsoids and sorting them by decreasing size. Last step is setting each structuring object onto the chosen substrate only if it does not overlap with a former object; drawing another couple of center of gravity and orientation angle otherwise.

## 5. Results and discussion

This generation process has been applied, until now, only on a natural freshly tilled seedbed surface. We estimated Gamma pdf for  $a$  and Bêta pdf for  $f_h$  and  $f_v$ . A correction has to be made



**Figure 8.** Generation process flow chart

for  $a$ , since the minimum diameter of our aggregates is 7 mm and not zero. Therefore, the computing is performed with the translated variable  $a - a_{min}$ , where  $a_{min}$  designates the minimum value of  $a$ . Figures 9 to 11 show the cdf of observed and modelled major semi-axis  $a$  and compression factors  $f_h$  and  $f_v$ . The small difference between the model and the data in each case, less than 3%, confirms the theoretical distributions. The data modeling is also validated by Pearson's chi-squared tests. Figure 12 shows a generated surface obtained by placing half ellipsoids onto a plane.

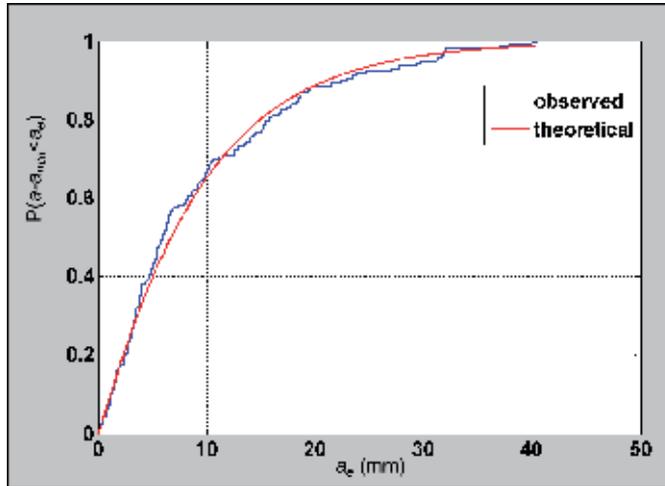


Figure 9. Cdf of half ellipsoid major semi-axis  $a$  fitted by Gamma function

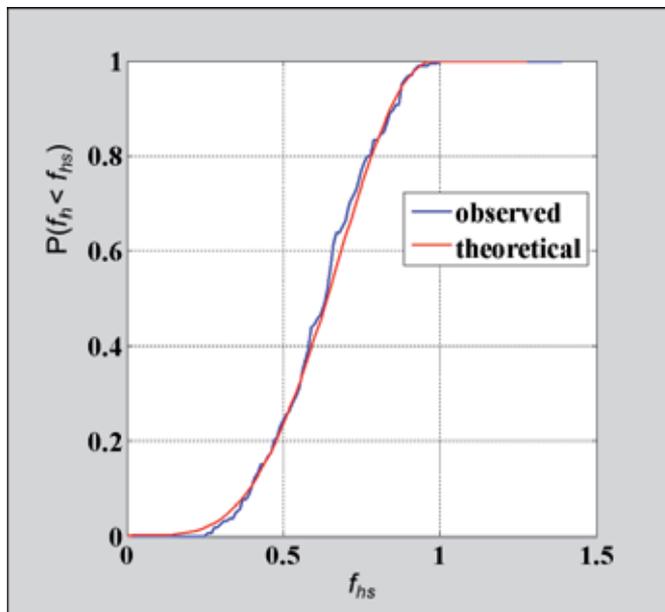
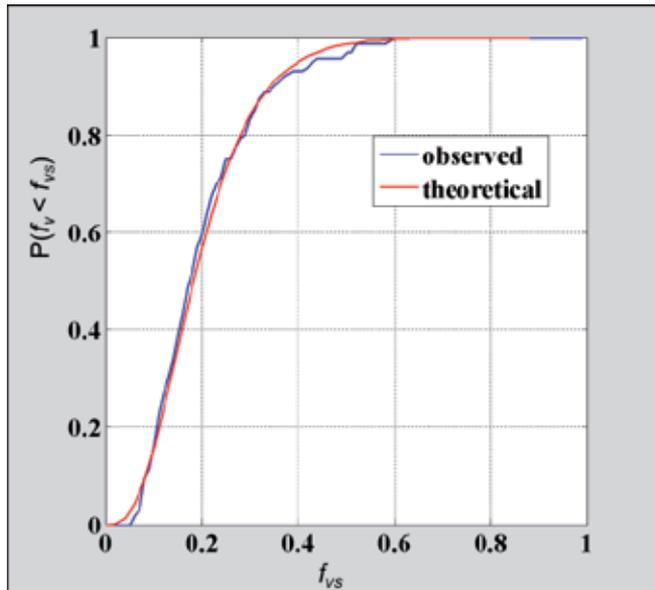
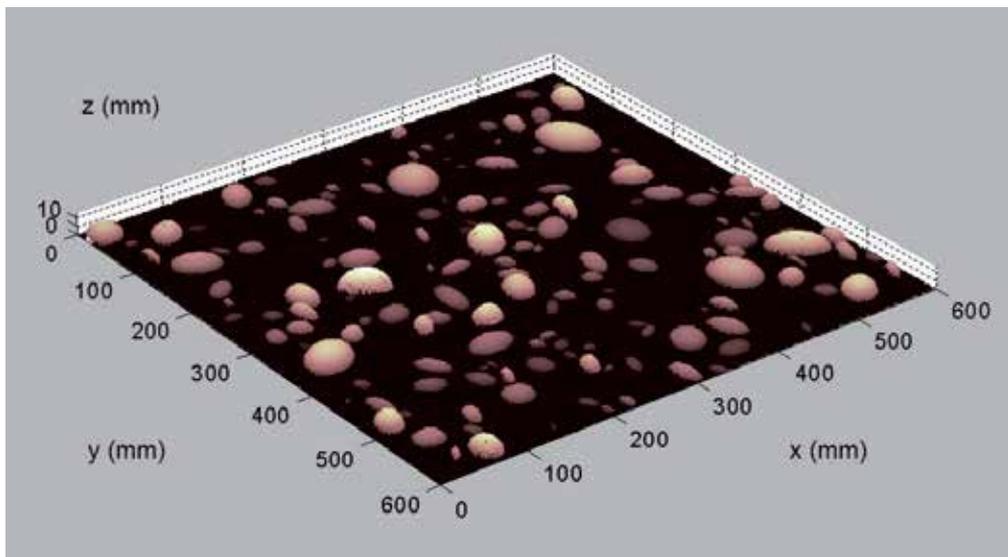


Figure 10. Cdf of horizontal compression factor  $f_h$  fitted by Beta function



**Figure 11.** Cdf of vertical compression factor  $f_v$  fitted by Beta function

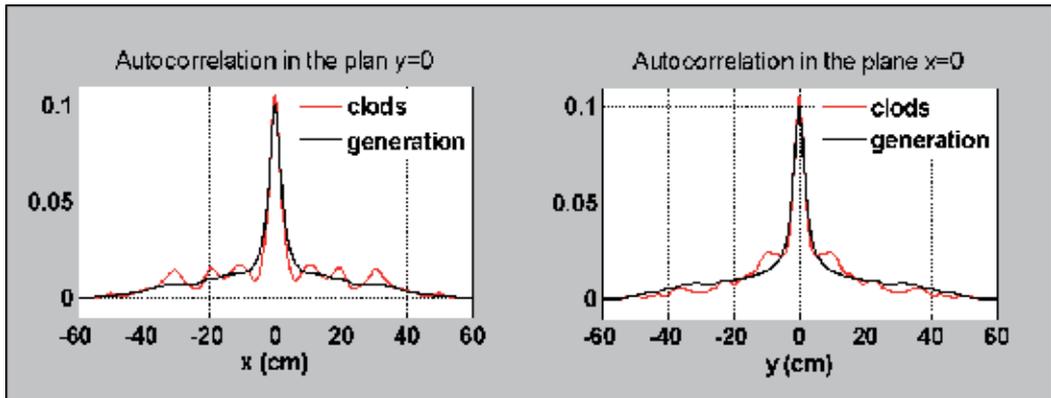


**Figure 12.** Generated numerical surface

The generated soil surface on figure 12 is a simplified surface representing only some irregularities of the soil. Aggregates and clods cover about 30% of a surface. A more complex model would include also the holes or depressions and an appropriate substrate. The segmentation

methods presented in this chapter have the potential to detect also the holes. Characterization of a more realistic substrate than a plane is a complex problem that has not been solved until now.

In order to validate the exposed generation process, the autocorrelation function of generated surface was compared to that of the surface obtained by setting the half ellipsoid modelling the automatically identified clods and aggregates of the seedbed surface under study onto a plane. Both autocorrelation functions can be seen on figure 13.



**Figure 13.** Autocorrelation functions of half ellipsoids set onto a plane by fitting detected clods or by generation process

One can see that the autocorrelation function of generated surface reproduces satisfactorily the shape of the autocorrelation function of the modeled clods set onto a plane, which is encouraging for future work. The black curve is a mean autocorrelation function, estimated on 15 realizations of generated surface. It has thus no more oscillations.

Recently, some authors have recalled the need to characterize the soil structure [69] or to define appropriate parameterization of the microtopography [70]. Addressing the arrangement of aggregates and clods contributes to these objectives. The proposed approach for modelling soil characteristics and generating soil surface is robust, as shown by the results. It relies on a geometric model of clods, which gives a physical basis to the characterization of surface microrelief. It represents also a progress in modelling compared to preceding works [22, 38, 66]. Future work should take into account soil depressions in relation to aggregates and clods.

## 6. Conclusion

Taking into account the structured elements when studying roughness effects is more and more brought into focus. The soil irregularities at millimeter or centimeter scales have an impact on hydrological processes, as they influence water outflow and infiltration rate. They have also an influence on remote sensing studies, by producing scattering and shadowing

effects. The proposed approaches in this chapter are detecting and characterizing some of the soil surface irregularities that are clods and big aggregates. Little studies were dedicated to this topic. For seedbed-type surfaces, a contour-based identification method enables to automatically retrieve the clods and big aggregates on a 3D DEM with a sensitivity of 75% and a specificity of 95%. Then modelling clods by semi-ellipsoids is suitable and allows reproducing the same autocorrelation function. The cdf of the semi-ellipsoids parameters were fitted with relative errors less than 3%. Numerical soil surfaces were successfully generated by placing semi-ellipsoids onto a plane, according to the statistical distributions. These recent works were applied to a limited number of soils and should be extended to more soils. Interest of a good surface roughness description is allowing generating realistic numerical surfaces. Such surfaces are then very useful for soil science studies as well as for electromagnetics codes used in remote sensing.

## Author details

Edwige Vannier, Odile Taconet, Richard Dusséaux and Olivier Chimi-Chiadjeu

Université de Versailles Saint-Quentin en Yvelines, LATMOS/IPSL/CNRS, Guyancourt, France

## References

- [1] Holah N., Baghdadi N., Zribi M., Bruand A., King C. Potential of ASAR/ENVISAT for the characterization of soil surface parameters over bare agricultural fields. *Remote Sens. Environ.* 2005, 96: 78-86.
- [2] Rahman M.M., Moran S.M., Thoma D.P., Bryant R., Sano E.E., Holifield Collins C.D, Skirvin S., Kershner C., B.J. Orr, A derivation of roughness correlation length for parameterizing radar backscatter models. *Int. J. Remote Sens.* 2007, 28: 1907-1920.
- [3] Verhoest N.E.C., Lievens H., Wagner W., Alvarez-Mozos J., Moran M.S., Mattia F. On the soil roughness parameterization problem in soil moisture retrieval of bare surfaces from Synthetic Aperture radar. *Sensors.* 2008, 8: 4213-4248.
- [4] Lievens H., Verhoest N.E.C., De Keyser E., Vernieuwe H., Matgen P., Alvarez-Mozos J., De Baets B. Effective roughness modelling as a tool for soil moisture retrieval from C- and L-band SAR. *Hydrology Earth System Sciences.* 2009, 15: 151-162.
- [5] Elfouhaily T.M., Guérin C.A. A critical survey of approximate scattering wave theories from random rough surfaces. *Waves Random Media.* 2004, 14: R1-10.
- [6] Warnick K.F., Chew, W.C. Numerical simulation methods for rough surface scattering. *Waves Random Media.* 2001, 11: R1-R30.

- [7] Saillard M., Sentennac, A. Rigorous solutions for electromagnetic scattering from rough surfaces. *Waves Random Media*. 2001, 11: R103-R137.
- [8] Dusséaux, R., Aït Braham K., Granet G. Implementation and validation of the curvilinear coordinate method for the scattering of electromagnetic waves from two-dimensional dielectric random rough surfaces. *Waves Random Complex Media*. 2008, 18: 551-570.
- [9] Dusséaux R., Vannier E., Taconet O., Granet G. Study of backscatter signature for seedbed surface evolution under rainfall — Influence of radar precision. *Progress in Electromagnetics Research* 2012 ; 125: 415-437.
- [10] Thorsos E.I., Jackson D.R., The validity of the perturbation approximation for rough surface scattering using a Gaussian roughness spectrum. *J. Acoust. Soc. Am.* 1989, 86: 261-277.
- [11] Beckmann P., Spizzichino A. *The Scattering of Electromagnetic Waves from Rough Surfaces*. 1963. Pergamon, New-York.
- [12] Thorsos, E.I. The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum, *J. Acoust. Soc. Am. A*. 1988, 82: 78-92.
- [13] Voronovich G. *Wave Scattering from Rough Surfaces*. 1994. Springer, Berlin.
- [14] Fung A. K. *Microwave scattering and emission models and their applications*. 1994. Artech House, Boston.
- [15] Taconet O., Ciarletti V. Estimating soil roughness indices on a ridge-and-furrow surface using stereo photogrammetric. *Soil and Tillage Research* 2007; 93: 64-76.
- [16] Chimi O., *Caractérisation probabiliste et synthèse de surfaces agricoles par objets structurants à partir d'images haute résolution*. PhD thesis. University of Versailles Saint-Quentin en Yvelines. 2012
- [17] Mattia F. Coherent and incoherent scattering from tilled soil surfaces. *Waves Random Complex Media*. 2011, 21: 278-300.
- [18] Franceschetti G., Migliaccio M., Riccio D. An electromagnetic fractal-based model for the study of fading. *Radio Sci.* 1996, 31: 1749–1759.
- [19] Guérin C.A., Holschneider M., Saillard M. Electromagnetic scattering from multi-scale rough surfaces. *Wave Random Media* 1997: 331-349.
- [20] Taconet O., Dusséaux R., Vannier E., Chimi-Chiadjeu O. Statistical description of seedbed cloddiness by structuring objects using digital elevation models. *Computers and Geosciences* 2013; 60: 117-125.
- [21] Tsang L., Kong J.A., Ding K.H., Ao C.O. *Scattering of electromagnetic waves – Numerical Simulations*. Wiley-Interscience, New York: 2001.

- [22] Zribi M., Le Morvan A., Dechambre M., Baghdadi N. Numerical backscattering analysis for rough surfaces including a cloddy structure. *IEEE Transaction Geosciences Remote Sensing* 2010; 48 (5): 2367-2374.
- [23] Currence H. D., Lovely W. G. The analysis of soil surface roughness. *Transaction ASAE* 1970; 13: 710-714.
- [24] Römken M.J.M., Wang, J.Y. Effect of tillage on surface roughness. *Transactions of the American Society of Agricultural Engineers* 1986; 29: 429-433.
- [25] Bertuzzi P., Rauws G., Courault D. Testing roughness indices to estimate soil roughness changes due to simulated rainfall. *Soil and Tillage Research* 1990; 17: 87-99.
- [26] Helming K., Roth C. H., Wolf R., Diestel H. Characterization of rainfall-microrelief interactions with runoff using parameters derived from digital elevation model (DEMs). *Soil Technology* 1993; 6: 273-286.
- [27] Takken I., Govers G. Hydraulics of interrill overland flow on rough, bare soil surfaces. *Earth Surface Processes and Landforms* 2000; 25: 1387-1402.
- [28] Darboux F., Davy P., Gascuel-Oudou C., Huang C. Evolution of soil surface roughness and flowpath connectivity in overland flow experiments. *Catena* 2001; 46: 125-139.
- [29] Zribi M., Dechambre M. A new empirical model to retrieve soil moisture and roughness from C-band radar data. *Remote sensing of environment* 2002; 84: 42-52.
- [30] Blaes X., Defourny P. Characterizing bidimensional roughness of agricultural soil surfaces for SAR modelling. *IEEE Transactions on geoscience and remote sensing* 2008; 46: 4050-4061.
- [31] Croft H., Anderson K., Kuhn N.J. Reflectance anisotropy for measuring soil surface roughness of multiple soil types. *Catena* 2012; 93: 87-96.
- [32] Huang C., Bradford J.M. Application of a laser scanner to quantify soil microtopography. *Soil Science Society of America Journal* 1992; 56: 14-21.
- [33] Vidal Vázquez E., Vivas Miranda J.G., Paz González A. Characterizing anisotropy and heterogeneity of soil surface microtopography using fractal models. *Ecological Modelling* 2005 ; 182: 337-353.
- [34] Vidal Vázquez E., García Moreno R., Miranda J. G. V., Díaz M.C., Saa Requejo A., Paz Ferreiro J., Tarquis A.M. Assessing soil surface roughness decay during simulated rainfall by multifractal analysis. *Nonlinear Processes in Geophysics* 2008; 15: 457-468.
- [35] Moldenhauer W.C. Influence of rainfall energy on soil loss and infiltration rates: II. Effect of clod size distribution. *Soil Science Society of America Journal* 1970; 34: 673-677.

- [36] Ambassa-Kiki R., Lal R. Surface clod size distribution as a factor influencing soil erosion under simulated rain. *Soil and Tillage Research* 1992; 22: 311-322.
- [37] Sandri R., Anken T., Hilfiker T., Sartori L., Bollhalder H. Comparison of methods for determining cloddiness in seedbed preparation. *Soil and Tillage Research* 1998; 45: 75-90.
- [38] Kamphorst E.C., Chadœuf J., Jetten V., Guérif J. Generating 3D soil surfaces from 2D height measurements to determine depression storage. *Catena* 2005; 62: 189-205.
- [39] Arvidsson J., Bölenius E. Effects of soil water content during primary tillage - laser measurements of soil surface changes. *Soil and Tillage Research* 2006; 90: 222-229.
- [40] Keller T., Arvidsson J., Dexter A.R. Soil structures produced by tillage as affected by soil water content and the physical quality of soil. *Soil Tillage Research* 2007; 92: 45-52.
- [41] Vannier E., Ciarletti V., Darboux F. Wavelet-based detection of clods on a soil surface. *Computers and Geosciences* 2009 ; 35: 2259-2267.
- [42] Taconet O., Vannier E., Le Hégarat-Masclé S. A boundary-based approach for clods identification and characterization on a soil surface. *Soil Tillage Research* 2010 ; 109: 123-132.
- [43] Borselli L., Torri D. Soil roughness, slope and surface storage relationship for impervious areas. *Journal of Hydrology* 2010; 393: 389-400.
- [44] Merrill D. Comments on the chain method for measuring soil surface roughness: use of the chain set. *Soil Science Society American Journal* 1998; 62: 1147-1149.
- [45] Planchon O., Esteves M., Silvera N., Lapetite J. Microrelief induced by tillage : measurement and modelling of surface storage capacity. *Catena* 2001; 46: 141-157.
- [46] Guzha A.C. Effects of tillage on soil microrelief, surface depression storage and soil water storage. *Soil and Tillage Research* 2004; 76: 105-114.
- [47] Brasington J., Smart R.M.A. Close range digital photogrammetric analysis of experimental drainage basin evolution. *Earth Surface Processes and Landforms* 2003; 28: 231-247.
- [48] Rieke-Zapp D., Nearing M.A. Digital close range photogrammetry for measurement of soil erosion. *Photogrammetric Record* 2005; 20 (109): 69-87.
- [49] Abd Elbasit M.A.M., Anyoji H., Yasuda H., Yamamoto S. Potential of low cost close-range photogrammetry system in soil microtopography quantification. *Hydrological Processes* 2009; 23: 1408-1417.
- [50] Bretar F., Arab-Sedze M., Champion J., Pierrot-Deseilligny M., Heggy E., Jacquemoud S. An advanced photogrammetric method to measure surface roughness : ap-

- plication to volcanic terrains in the Piton de la Fournaise, Reunion Island. *Remote Sensing of Environment* 2013 ; 135: 1-11.
- [51] Helming K., Römken M.J.M., Prasad S.N. Surface roughness related processes of runoff and soil loss: a flume study. *Soil Science Society of America Journal* 1998; 62: 243-250.
- [52] Darboux F., Huang C. An instantaneous-profile laser scanner to measure soil surface microtopography. *Soil Science Society America Journal* 2003; 67: 92-99.
- [53] Zhixiong L., Nan C., Perdok U.D., Hoogmoed W.B. Characterization of soil profile roughness. *Biosystem Engineering* 2005; 91: 369-377.
- [54] Frede H.-G., Gäth S. Soil surface roughness as the result of aggregate size distribution 1. Report: measuring and evaluation method. *Pflanzenernaehr. Bodenk.* 1995; 158: 31-55.
- [55] Jester W., Klik A. Soil surface roughness measurement- methods, applicability and surface representation. *Catena* 2005; 64: 174-192.
- [56] Aguilar M.A., Aguilar F.J., Negreiros J. Off-the-shelf laser scanning and close-range digital photogrammetry for measuring agricultural soils microrelief. *Biosystem Engineering* 2009; 103: 504-517.
- [57] Marzahn P., Ludwig R. On the derivation of soil surface roughness from multi parametric PolSAR data and its potential for hydrological modelling: *Hydrology Earth System Sciences* 2009; 13: 381-394.
- [58] Vulfson L., Genis A., Blumberg D.G., Sprintsin M., Kotlyar A., Freilikh V., Ben-Asher J. Retrieval of surface roughness parameters of bare soil from the radar satellite data. *Journal of Arid Environments* 2012; 87: 77-84.
- [59] Lievens H., Verhoest N.E.C. Spatial and temporal soil moisture estimation from RADARSAT-2 imagery over Flevoland, the Netherlands. *Journal of Hydrology* 2012; 456-457: 44-56.
- [60] Chimi-Chiadjeu O., Vannier E., Dusséaux R., Taconet O. Influence of gradient estimation on clod identification on seedbed digital elevation model. *Environmental & Engineering Geoscience* 2011; 17: 337-352.
- [61] Fawcett T. *ROC Graphs: Notes and Practical Considerations for Researchers*. Technical Report. HP Laboratories, Palo Alto 2004, USA.
- [62] Beucher S., Meyer F. The morphological approach to segmentation: the watershed transformation. In: E. R. Dougherty (ed) *Mathematical Morphology in Image Processing*. 1993. p433-481.
- [63] Chimi-Chiadjeu O., Vannier E., Dusséaux R., Le Hégarat-Masclé S., Taconet O. Segmentation of elevation images based on a mathematical morphology approach for

- agricultural clod detection: proceedings of 5th international congress on image and signal processing CISP 2012, Chongqing, China, 16-18 October 2012, 701-705.
- [64] Chimi-Chiadjeu O., Vannier E., Dusséaux R., Le Hégarat-Masclé S., Taconet O. Using simulated annealing algorithm to move clod boundaries on seedbed digital elevation model. *Computers & Geosciences* 2013 ; 57: 68-76.
- [65] Kim Y.H., Rana S., Wise S. Exploring multiple viewshed analysis using terrain features and optimisation techniques. *Computer & Geosciences* 2004; 30 : 1019–1032.
- [66] Cierniewski J., Verbrugghe M., Marlewski A. Effects of farming works on soil surface bidirectional reflectance measurements and modelling. *International Journal of Remote Sensing* 2002 ; 23 (6): 1075-1094.
- [67] Richard G., Boizard H., Roger-Estrade J., Boiffin J., Guérif J. Field study of soil compaction due to traffic in northern France: pore space and morphological analysis of the compacted zones. *Soil and Tillage Research* 1999; 51: 151-160.
- [68] Feral L., Mesnard F., Sauvageot H., Castanet L., Lemorton J. Rain cells shape and orientation distribution in south-west of France. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere* 2000; 25 (10): 1073-1078.
- [69] Roger-Estrade J., Richard G., Caneill J., Boizard H., Coquet Y., Defosse P., Manichon H. Morphological characterisation of soil structure in tilled fields: from a diagnosis method to the modeling of structural changes over time. *Soil and Tillage Research* 2004; 79: 33-49.
- [70] Martin Y., Valeo C., Tait M. Centimetre-scale digital representations of terrain and impacts on depression storage and runoff. *Catena* 2008; 75: 223-233.

---

# Simulation of Tsunami Impact on Sea Surface Salinity along Banda Aceh Coastal Waters, Indonesia

---

Maged Marghany

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/58570>

---

## 1. Introduction

Constantly later the catastrophe of the Indian Ocean tsunami of Boxing Day 2004, research in tsunami geoscience has augmented evidently [1]. Scientists have attempted to comprehend the mechanisms of the wide scale of the Indian Ocean tsunami of 2004. Nevertheless, with great efforts done by scientists since Boxing day 2004, the Japanese tsunami with great disaster occurred. However, on March 11, 2011, a magnitude of  $M_w$  9.0 earthquake struck off the coast of Honshu, Japan, sparking a tsunami that not only devastated the island nation, but also caused destruction and fatalities in other parts of the world, including Pacific islands and the United State (U.S.) West Coast [4].

Initial reports were eerily similar to those on December 26, 2004, when a massive underwater earthquake off the coast of Indonesia's Sumatra Island rattled the Earth in its orbit. The 2004 quake, with a magnitude of  $M_w$  9.1, was the largest one since 1964. But as in Japan, the most powerful and destructive aftermath of this massive earthquake was the tsunami that it caused. The death toll reached higher than 220,000 [1-3][10].

### 1.1. Definitions of tsunami

It is well known that the tsunami is the natural phenomena consisting of a series of waves generated when the waves are rapidly displaced on a massive scale. Tsunami (pronounced soo-NAH-mee) is a Japanese word which is meaning harbor ('tsu') and wave ('nami'). Tsunamis are fairly common in Japan and many thousands of Japanese have been killed by them in recent centuries. In this context, the term was coined by fishermen who returned to port to find the area surrounding the harbor devastated, in spite they had not been aware of any wave on high seas [2].

Haugen et al., [8] stated that tsunamis are long waves set in motion by an impulsive perturbation of the sea, intermediate between tides and swell waves in the spectrum of gravity water waves. Subsequently, Zahibo et al., [5] defined tsunami waves as surface gravity waves that occur in the ocean as the result of large-scale short-term perturbations (underwater earthquakes, eruptions of underwater volcanoes, landslides, rock falls, pyroclastic avalanche from land volcanoes entered in water, asteroid impact, and underwater explosions).

### 1.2. Comments on tsunami definition

In earlier times, seismic ocean waves were called “tidal” waves, incorrectly implying that they had some direct connection to the tides. In fact, when the tsunami approach coastal zone they began to characterize by a violent onrushing tidal rather than the sort of cresting waves that are generated by wind stress upon the sea surface. However, to eliminate this confusing the Japanese word “tsunami is used to describe the giant wave (Figure 1) in which is referring to a seismic wave and meaning harbor wave to replace the misleading term tidal wave. This tsunami is a synonym for seismic sea wave. In this regard, a tsunami is a seismic sea wave containing tremendous amounts of energy as a result of its mode of formation i.e. the factor that causing a seismic wave. Therefore, tsunamis are temporary oscillations of sea level with periods longer than wind, waves and shorter than tides for the tsunami, and shorter than a few days of storm surge [8].



**Figure 1.** Giant Tsunami Wave [3].

### 1.3. Tsunami characteristics

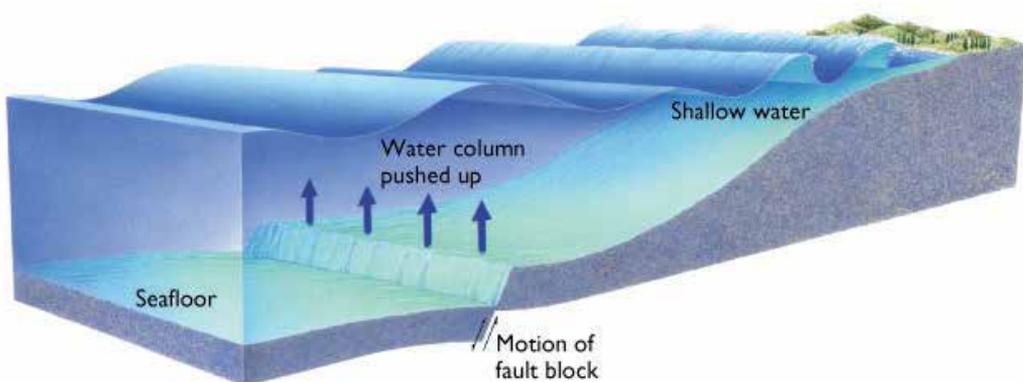
The physical parameters of duration, length, propagation speed, and heights are the keys description of tsunami. In this regard, tsunamis have duration is ranged between 5 to 100

minutes. They have long length which is ranged between 100 m to 1000 km. Further, tsunami propagation speed is between 1 to 200 m/s, and their heights can be up to tens of meters. Therefore, Zahibo et al., [5] stated that tsunami waves of the seismic origin are usually very long (50–1000 km). In this context, the source of the giant 2004 tsunami in the Indian Ocean (magnitude of  $M_w$  9.0–9.3) has approximated dimensions: (i) length; (ii) 670 km; (iii) width 150 km; and (iv) height 12 m [5].

This means that a tsunami can propagate long distances before it strikes a shoreline hundreds or thousands of kilometres from the earthquake source. To accurately model tsunami propagation over such large distances, the Earth's curvature should be taken into account. Other factors, such as Coriolis forces and dispersion, may also be important [7].

#### 1.4. Tsunami generation

It is interesting to understand the mechanisms of tsunami generation. Generation mechanisms of tsunamis are geological events like land- and rockslides in fjords and lakes, submarine gravity mass flows and earthquakes. The Storegga slide is one of the largest submarine slides in the world [8]. While the mechanism for generating the initial water waves by purely tectonic motions is reasonably well comprehended. Conversely, the modelling of tsunamis generated by submarine landslides is not yet explicated. Co-seismic deformation of the seafloor usually occurs rapidly relative to the propagation speeds of long water waves (Figure 2), allowing for simple specification of initial conditions by transferring the resultant permanent seafloor deformation to the free surface. However, sub-aerial and submarine landslides move less rapidly and the time-history of seafloor deformation (Figure 2) is important, necessitating the addition of source terms in the equations of motions. Compared with the understanding of earthquake-induced initial tsunami waves, the understanding of landslide-generated waves is marginal. Briefly, in terms of the semi-analytical empirical studies transferred the energy released by a moving block sliding from its initial position to its final position to solitary waves and calculated the height of the wave [9].



**Figure 2.** Tsunami generation due to motion of fault block.

### 1.5. Tsunami classifications

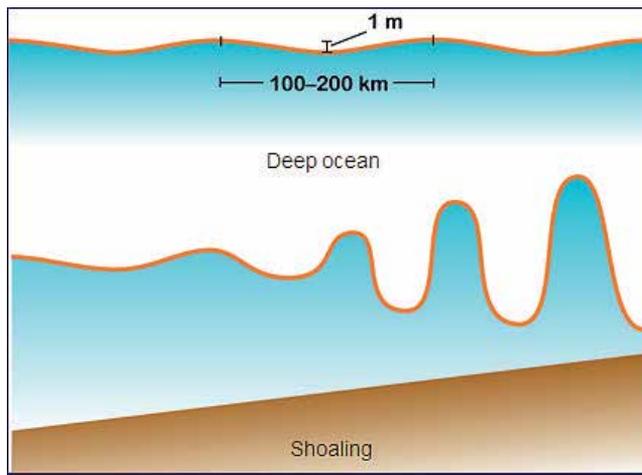
Tsunami can be classified by the distance from their source to the area of impact; i.e., local and remote tsunami. Locally generated tsunami have short warning times and relatively short wave periods; remote tsunami have longer warning times and relatively long periods. Typical periods for tsunami range from 15 minutes for locally generated tsunami to several hours for remote tsunami. Typical run-up height for tsunami range up to 15 m at the coast, although most are much smaller. Storm surges on the other hand are caused by variations in barometric pressure and wind stress over the ocean. Decreasing barometric pressure causes an inverse barometer effect where sea level rises. This is usually a slow and large-scale effect and thus does not usually generate waves in the frequency range typical of tsunami. However, there can be short-period meteorological events (such as meteorological tsunami – rissaga) with time-scales of a few hours that may be important. Wind stress on the other hand has a wide range of time-scales and causes coastal sea-level setup as well as wind waves, where the setup depends on the wind direction, strength, and wave height. Storm surge periods range from several hours to several days. Typical heights for storm surge alone range up to 1.0 m for instance, along the coast of New Zealand, although most are usually less than 0.5 m. Wind waves, on the other hand, can be quite large, producing wave setup and wave run-up of several meters in height.

In general, tsunamis can be categorized into: (i) microtsunami which has a small amplitude and would not even notice visually and (ii) local tsunami which has destructive impact due to its wide spread on coastal zone within hundred of kilometres and usually caused by plate tectonic movements. It might be internal wave named by internal tsunami due to its traveling along the a thermocline layer.

In a somewhat similar fashion, dropping a stone into a puddle of water creates a series of waves which radiate away from the impact point. In this context, the impact point of puddle of water is representing a sudden shifting of rocks or sediments on the ocean floor caused by cataclysmic event, such as a volcanic eruption, an earthquake, or a submarine landslide, can force the water level to drop  $\geq 1$  m generating a tsunami—a series of low waves with long periods, and long wavelengths (Figure 3). This indicates that the tsunami wavelength is bigger than its amplitude in the open ocean. Thus the tsunami height in the open ocean is approximately less than 1 m which it is not noticeable in open ocean. The tsunamis, however, cross the oceans at a rate of  $\sim 750$  km/hr which is equivalent to a speed of a modern jet aircraft! Despite these phenomenal speeds, tsunami pose no danger to vessels in the open ocean. Indeed, the regular ocean swell would probably mask the presence of these low sea waves. The tsunamis, however, grow to height of  $\geq 10$  m as it impinges on a shoreline and flood the coast, sometimes with catastrophic results, including widespread property damage and loss of life.

### 1.6. Geological descriptions of Sumatra earthquake

In previous sections, the fundamental mechanism of tsunami was explained. Therefore, this section is devoted for tsunami of 26 December 2004 which is known as Sumatra-Andaman earthquake. Consequently, this disaster is called as Asian Tsunami in Asia region and also known as a Boxing day in the Australia, Canada, New Zealand, and the United Kingdom,



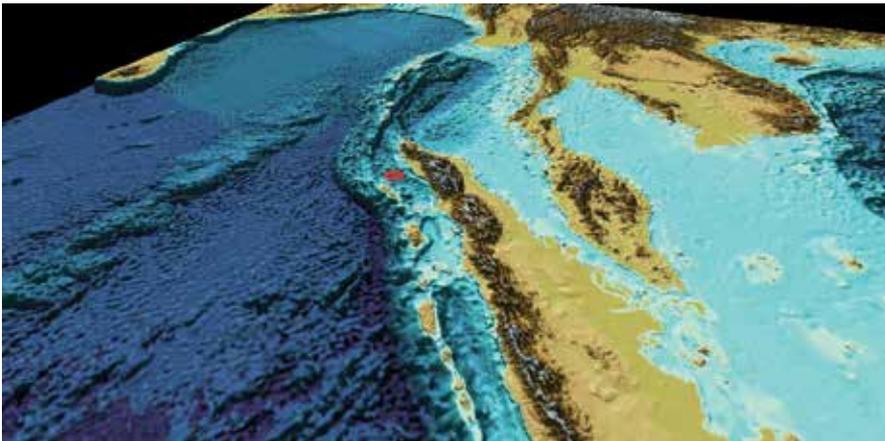
**Figure 3.** Tsunami Generation in Deep Water.

because it took place on Boxing Day. This earthquake was also reported to be the longest duration of faulting ever observed, lasting between 500 and 600 seconds (8.3 to 10 minutes), and it was large enough that it caused the entire planet to vibrate as much as half an inch, or over a centimeter.

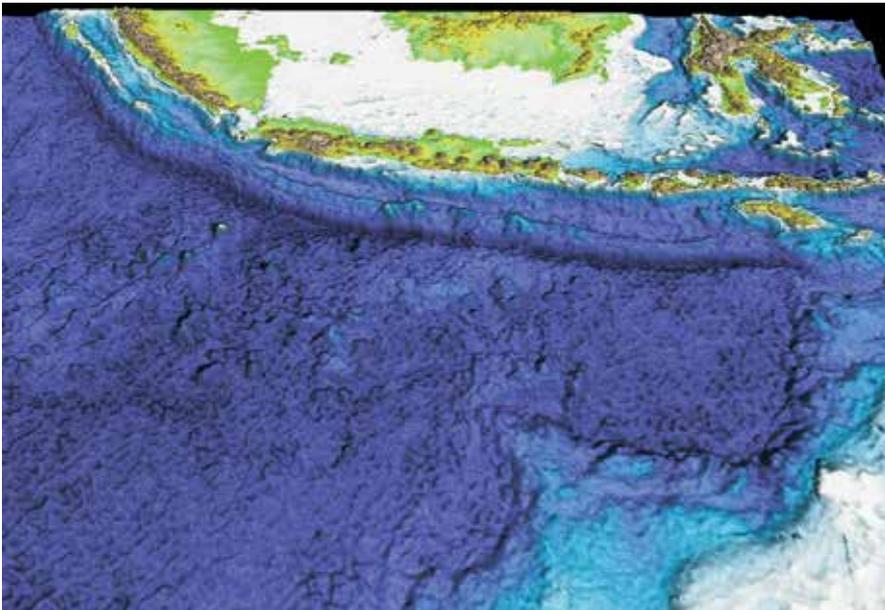
#### 1.6.1. Epicenter of the giant tsunami

The epicentre of the earthquake of the exceptionally high magnitude of 9.0 (Figure 4) is situated inside the trough as indicated between the northern edge of Sumatra and the small island of Simeulue, one member of the chain of islands next to the trench. Neither this trough relief reveals anything unusual nor does the comparatively moderate depth of the Sunda Trench of less than 5,000 m, the shallower sister of the conspicuous, adjacent Java-Trench southeast of it (Figure 5). The existing records of earthquakes of more than magnitude 5.5 in the Sumatra area since 1970 have clearly shown that their great majority, including the epicentre of the Giant Tsunami, occurred within the shallow strip between the coastline of Sumatra and the adjacent subduction trench rim; in other words: they occurred on the edge of the overriding plate!

According to Abbott [11], the Indian-Australian plate moves obliquely toward western Indonesia at 5.3 to 5.9 cm/yr (2 to 2.3 in/yr). The enormous, ongoing collision results in subduction-caused (Figure 6) earthquakes that are frequent and huge. Many of these earthquakes send off tsunamis. On 26 December 2004, 1,200 km (740 mi) long fault rupture began as 100 km (62 mi) long portion of the plate tectonic boundary ruptured and slipped during a minute (Figure 6). The rupture then moved northward at 3 km/sec (6,700 mph) for four minutes. Then slowed to 2.5 km/sec (5,600) during the next six minutes. At the northern end of the rupture, the fault movement slowed drastically and only traveled tens of meters during the next hour. It appears that a bend or scissors like tear in the subduction plate may have delayed the full rupture in December 2004.

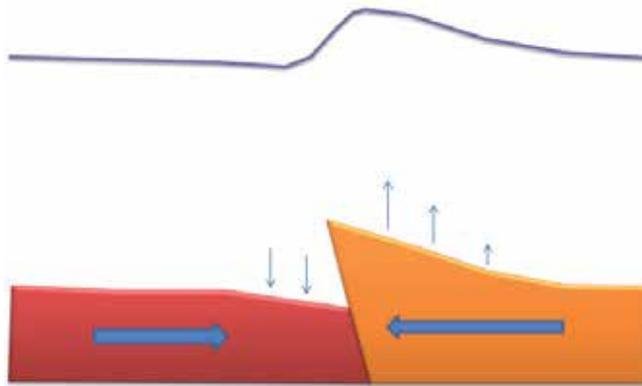


**Figure 4.** Epicenter of Giant Tsunami



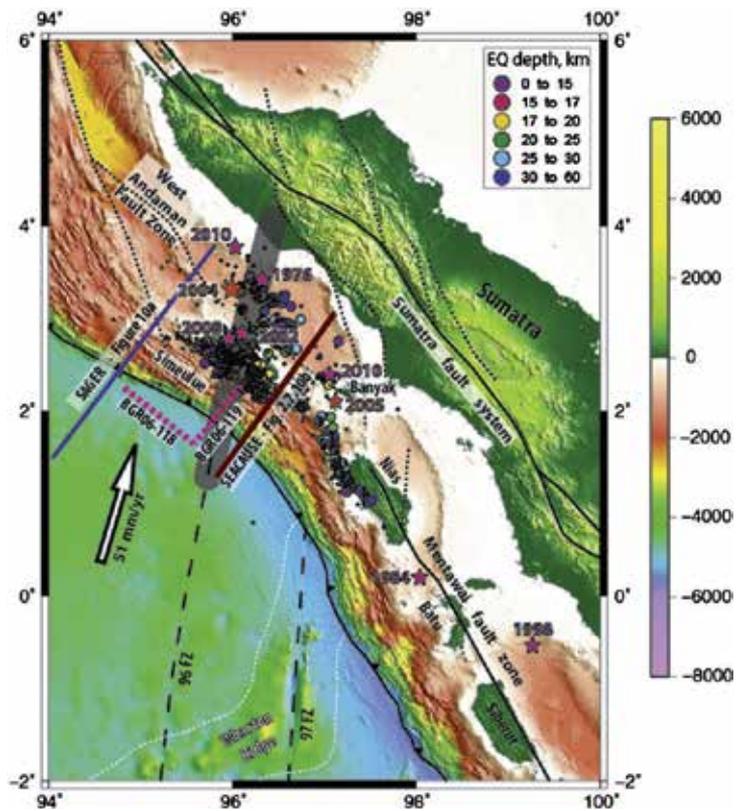
**Figure 5.** Java Trench.

Shulgin et al., [13] stated that the tectonics around Northern Sumatra are predominantly controlled by the subduction of the oceanic Indo-Australian plate underneath Eurasia. The current convergence rate offshore Northern Sumatra is estimated at 51 mm/yr. The increasing obliquity of the convergence northwards from the Sunda Strait results in the formation and development of a number of arc-parallel strike-slip fault systems (Figure 7). The most significant are the Sumatra and the West Andaman Fault systems, accommodating arc-parallel



**Figure 6.** Deformation of sea Floor and Sea Surface.

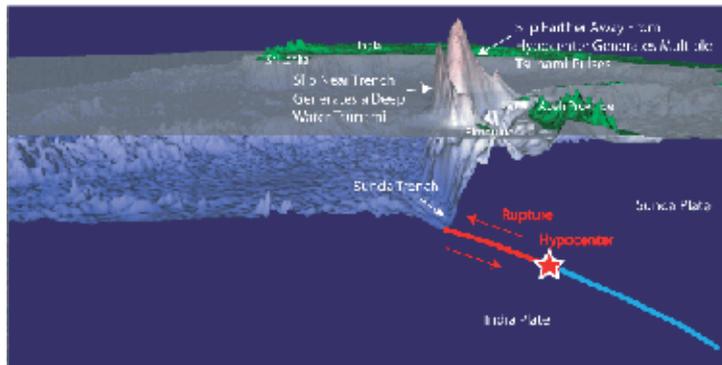
strain offshore central-southern Sumatra. For the Mentawai fault system, recent findings suggest deformation dominated by backthrusting.



**Figure 7.** Subduction of Indian-Australian plate beneath Indonesia Shulgin et al., [13]

Subsequently, Kelmelis et al., [10], the Indo-Australian plate is moving approximately 40 to 50 mm per year to the northeast, and the earthquake ruptured at the subduction zone boundary or interpolate thrust boundary (Figure 8). On the fault the earthquake had a maximum slip of approximately 15 to 20 meters with an average slip of >5 meters along the full length of the rupture. The sea floor overlying the thrust fault would have been raised by several meters. Further, velocities of displacement along 1200 to 1300 kilometres of the fault with at least three major energy bursts during the propagation of the rupture (50 to 150 seconds, 280 to 340 seconds, and 450 to 500 seconds).

### 2004 Sumatra-Andaman Earthquake (M=9.2)



### 2005 Northern Sumatra Earthquake (M=8.6)

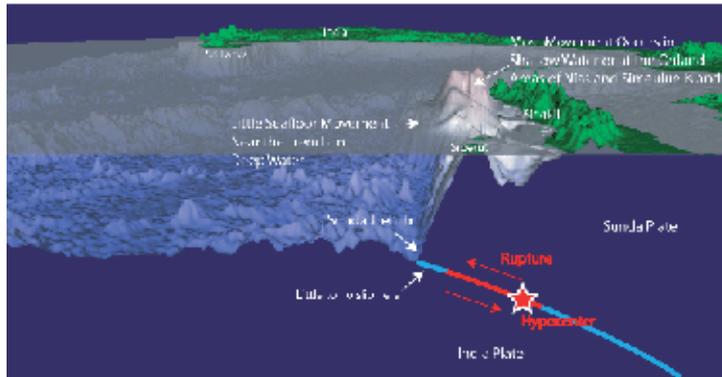
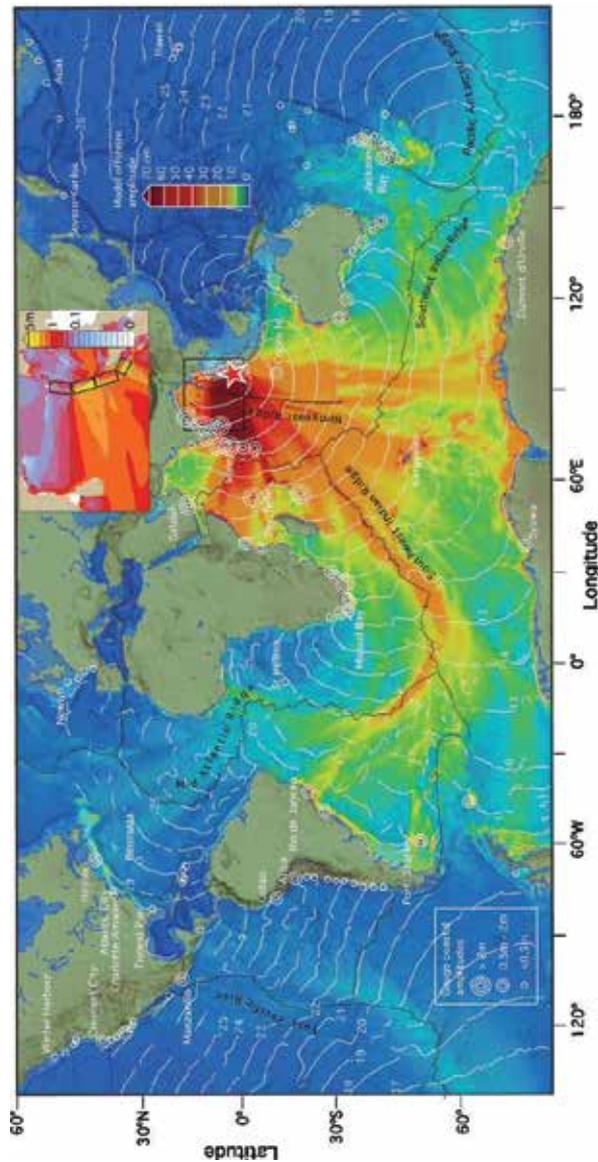


Figure 8. Location of Hypocenter of 2004.

#### 1.7. 2004 tsunami wave propagation

At 00:58:53 UTC (07:58:53 local time) December 26, 2004, an undersea earthquake was occurred along a tectonic subduction zone in which the India Plate. In general Figure 9 shows that the

tsunami reaches Phuket and Sri Lanka coasts in two hours after the earthquake, and African coast in 8-11 hours. The tsunami propagation is also animated (up to 5 hours) from a 1200 km fault. The red color means that the water surface is higher than normal, while the blue means lower. It indicates that initial tsunami to the east (e.g., Phuket) began with receding wave, while to the west (e.g., Sri Lanka) large wave suddenly reached. The darker the color, the larger the amplitude. The tsunamis were larger in the east and west directions.



**Figure 9.** Tsunami wave propagation during 2004 [14].

The tsunami was caused by movement along a fault line running through the seabed of the Indian Ocean. As the fault runs north-south, the waves travelled out across the ocean in mainly easterly and westerly direction with duration of 7 hours that shook the world. At 00.58 GMT, an undersea earthquake measuring 9.3 on the Richter Scale occurs off Indonesia. At +15 minutes later, the Indonesia Island of Sumatra, close to epicenter of the quake, is hit by the full force of tsunami. Many towns and villages in Aceh province on the western tip of the island are completely washed away, and the capital, Banda Aceh, is destroyed.

The remote Andaman which is lying only 100 km from the epicenter of the earthquake was struck within +30 minutes later. An hour after it hit Sumatra, the tsunami reached Thailand. It had lessened slightly in height and power but still struck the Thai coast with incredible force and the sea surged out for about 200 m. One of the worst-hit places was Sri-Lanka, which lay almost directly west of the earthquake's epicenter. The tsunami wave reached Sri Lanka within +2.00 hours later. The tsunami height recorded in Sri Lanka ranged between 5 to 10 m (Table 1).

Location	Heights
Sri Lanka, east coast	5-10 m
India, east coast	5-6 m
Andaman Islands	>5 m
Phuket, Thailand	3-5 m
Kenya	2-3 m

**Table 1.** Measured Tsunami Wave Heights on Boxing Day

With no continental shelf to lift the tsunami's waves as they near shore, the Maldives gets off relatively lightly within +3.30 hours. Finally tsunami waves strike Africa coast within +7 hours. The wave height along Kenya coast was between 2 and 3 m (Table 1).

Further, the Christmas tsunami was so powerful it actually sped up the rotation of the Earth reducing the length of its sidereal day. The earthquake that spawned it also caused the Earth to vibrate all over by as much as 1 cm. In this regard, the critical question may be raised is what the tsunami effects on the ocean physical properties such as temperature and salinity? In fact, there are no many studies have been conducted to answer this question. Temperature and salinity are the main parameters can be used to describe the ocean status. Both parameters can produce vertical current movement because of the their gradient changes. In addition, water density changes are function of gradual changes of temperature and salinity.

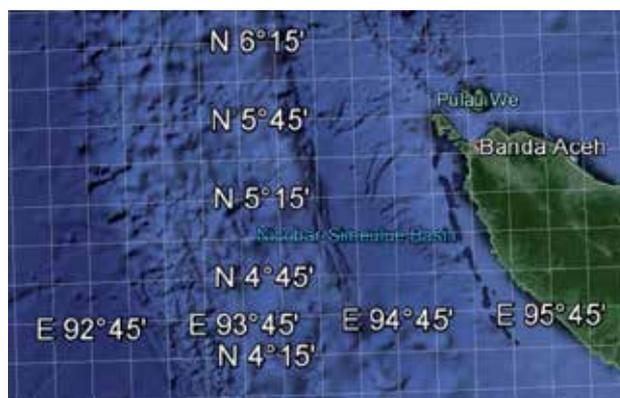
### 1.8. Hypotheses and objective

Concern with above prospective, we address the question of tsunami's impact on Sea Surface Salinity (SSS) pattern changes pro and post tsunami event of 2004. This is demonstrated with Moderate-resolution Imaging Spectrometer (MODIS) i.e. the Aqua/MODIS data level IB reflectance satellite data using least square algorithm. Three hypotheses are examined:

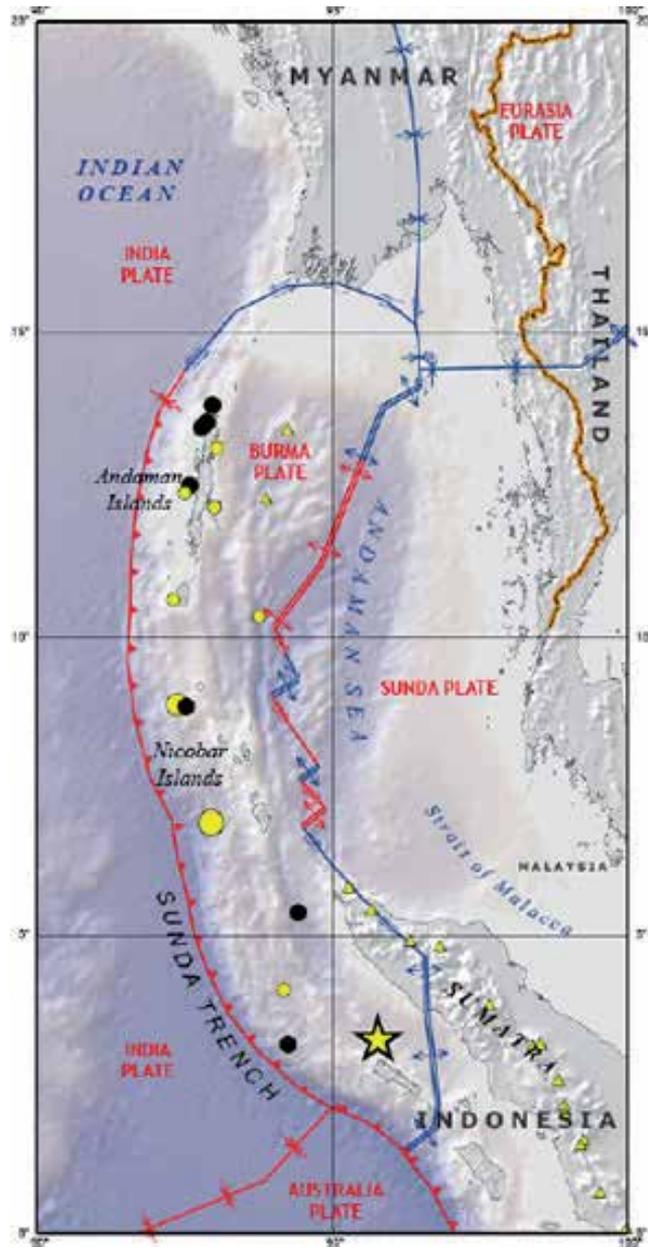
- i. Least square algorithm can be used to retrieve Sea Surface Salinity in MODIS satellite data during and post tsunami disaster;
- ii. Least square algorithm is automatic algorithm for retrieving SSS for short periods i.e. couple of days;
- iii. There is changes in SSS pattern post tsunami 2004 event along Banda Aceh coastline.

## 2. Study area

The study area is located along the western coastal zone of Aceh with boundaries of latitudes  $3^{\circ} 30' N$  to  $6^{\circ} 30' N$  and longitudes of  $93^{\circ} 30' E$  to  $99^{\circ} 30' E$  (Figure 10). The Sunda trench is running north-south along the coastal waters of Aceh towards the Andaman Sea. Running in a rough north-south line on the seabed of the Andaman Sea is the boundary between two tectonic plates, the Burma plate and the Sunda Plate. These plates (or microplates) are believed to have formerly been part of the larger Eurasian Plate, but were formed when transform fault activity intensified as the Indian Plate began its substantive collision with the Eurasian continent. As a result, a back-arc basin center was created, which began to form the marginal basin which would become the Andaman Sea, the current stages of which commenced approximately 3–4 million years ago (Figure 11). On December 26, 2004, a large portion of the boundary between the Burma Plate and the Indo-Australian Plate slipped, causing the 2004 Indian Ocean earthquake. This megathrust earthquake had a magnitude of 9.3. Between 1300 and 1600 kilometers of the boundary underwent thrust faulting and shifted by about 20 meters, with the sea floor being uplifted several meters. This rise in the sea floor generated a massive tsunami with an estimated height of 28 meters (30 ft) [14].



**Figure 10.** Location of study area.



**Figure 11.** Location of Sunda Trench

The average depth of the sea is about 1,000 meters (3,300 ft). The northern and eastern parts are shallower than 180 meters (600 ft) due to the silt deposited by the Irrawaddy River. This major river flows into the sea from the north through Burma. The western and central areas are 900–3,000 meters deep (3,000–10,000 ft). Less than 5% of the sea is deeper than 3,000 meters

(10,000 ft), and in a system of submarine valleys east of the Andaman-Nicobar Ridge, the depth exceeds 4,000 meters (13,200 ft). The sea floor is covered with pebbles, gravel and sand [14].

Further, the climate and water salinity of the Andaman Sea and Aceh are mostly determined by the monsoons of southeast Asia. Air temperature is stable over the year at 26 °C in February and 27 °C in August. Precipitation is as high as 3,000 mm/year and mostly occurs in summer. Sea currents are south-easterly and easterly in winter and south-westerly and westerly in summer. The average surface water temperature is 26–28 °C in February and 29 °C in May. The water temperature is constant at 4.8 °C at the depths of 1,600 m and below. Salinity is 31.5–32.5‰ (parts per thousand) in summer and 30.0–33.0‰ in winter in the southern part. In the northern part, it decreases to 20–25‰ due to the inflow of fresh water from the Irrawaddy River. Tides are semidiurnal (i.e. rising twice a day) with the amplitude of up to 7.2 m [23].

### 3. Least square model

In this section, we present the theoretical model of split window method that relates MODIS sea surface salinity with in situ salinity measured by thermal infrared sensors, these include multi-channel methods. We assume the MODIS image radiance  $I$  within multi-channels  $i$  have a linear relationship with measured sea surface salinity (SSS). A useful extension of linear function of  $k$  channels as in

$$SSS = b_0 + b_1I_1 + b_2I_2 + b_3I_3 + \dots + b_kI_k + \varepsilon \tag{1}$$

The model may be written in terms of the observations as

$$SSS = b_0 + \sum_{i=1}^k b_i I_i + \varepsilon \tag{2}$$

where  $SSS$  is the measured sea surface salinity,  $k$  is a number of MODIS radiance bands which equals 7 bands,  $b_0, b_i$  are constant coefficient of linear relationship between MODIS radiance data ( $I$ ) and  $\varepsilon$  is residual error of SSS estimated from selected MODIS bands. The unknown parameters in equation 2, that are  $b_0$  and  $b_i$  may be estimated by a general least square iterative algorithm. This procedure requires certain assumptions about the model error component  $\varepsilon$ . Stated simply, we assume that the errors are uncorrelated and their variance is  $\sigma_\varepsilon^2$ . In general, if  $\varepsilon_i$  and  $\varepsilon_j$  are two uncorrelated errors, then their covariance is zero, where we define the covariance as

$$Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \cdot \varepsilon_j) \tag{3}$$

The least-square estimator of the  $b_i$  minimizes the sum of squares of the errors, say

$$S_E = \sum_{j=1}^n (SSS_j - b_0 - \sum_{i=1}^k b_i I_i)^2 = \sum_{j=1}^n \varepsilon_j^2 \tag{4}$$

where  $SSS_j$  is the value of  $SSS$  measured at  $I_j$ ,  $n$  is the total number of data points and  $n \geq k$ . It is necessary that the least squares estimators satisfy the equations given by the  $k$  first partial derivatives  $\frac{\partial S_E}{\partial b_i} = 0, i=1,2,3,\dots,k$  and  $j=1,2,3,\dots,n$ . Therefore, differentiating equation 4 with respect to  $b_i$  and equating the result to zero we obtain

$$\begin{aligned} \hat{n}b_1 + (\sum_{j=1}^n I_{2j})\hat{b}_2 + (\sum_{j=1}^n I_{3j})\hat{b}_3 + \dots + (\sum_{j=1}^n I_{kj})\hat{b}_k &= \sum_{j=1}^n SSS_j \\ (\sum_{j=1}^n I_{2j})\hat{b}_1 + (\sum_{j=1}^n I_{2j}^2)\hat{b}_2 + \dots + (\sum_{j=1}^n I_{2j}I_{kj})\hat{b}_k &= \sum_{j=1}^n I_{2j}SSS_j \\ (\sum_{j=1}^n I_{kj})\hat{b}_1 + (\sum_{j=1}^n I_{kj}I_{2j})\hat{b}_2 + \dots + (\sum_{j=1}^n I_{kj}^2)\hat{b}_k &= \sum_{j=1}^n I_{kj}SSS_j \end{aligned} \tag{5}$$

The equations (5) are called the least-squares normal equations. The  $\hat{b}_k$  found by solving the normal equations (5) are the least-squares estimators of the parameters  $b_i$ . The only convenient way to express the solution to the normal equations is in matrix notation. Note that the normal equations (5) are just a  $k \times k$  set of simultaneous linear equations in  $k$  unknowns (the  $\{b_k\}$ ). They may be written in matrix notation as

$$H\hat{b} = h \tag{6}$$

where

$$H = \begin{bmatrix} n & \sum I_{2j} & \dots & \sum I_{kj} \\ \sum I_{2j} & \sum I_{2j}^2 & \dots & \sum I_{2j}I_{kj} \\ \sum I_{3j} & \sum I_{3j}I_{2j} & \dots & \sum I_{3j}I_{kj} \\ \dots & \dots & \dots & \dots \\ \sum I_{kj} & \sum I_{kj}I_{2j} & \dots & \sum I_{kj}^2 \end{bmatrix}$$

$$\hat{b} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_k \end{bmatrix} \text{ and } h = \begin{bmatrix} \sum SSS_j \\ \sum I_{2j}SSS_j \\ \vdots \\ \sum I_{kj}SSS_j \end{bmatrix}$$

Thus,  $H$  is a  $k \times k$  estimated matrix of MODIS radiance bands that used to estimate sea surface salinity,  $b$  and  $h$  are both  $k \times 1$  column vectors. The solution to the least-squares normal equation is

$$\hat{b} = H^{-1}h \tag{7}$$

where  $H^{-1}$  denotes the inverse of the matrix  $H$ . Given a solution to least squares normal equations, the retrieval  $SSS_{MODIS}$  is estimated using the fitted multiple regression model of equation 2 as given

$$SSS_{MODIS} = b_1 + \sum_{i=1}^k \hat{b}_i I_i \tag{8}$$

Following Sonia et al., (2007),  $\varepsilon$  errors that represents the difference between retrieved and in situ SSS are computed within 10 km grid point interval and then averaged over all grid points having the same range of distance to coast, where the bias  $\varepsilon$  on the retrieved  $SSS_{MODIS}$  is given by:

$$\varepsilon = \frac{\sum_{i=1}^N (SSS_{MODIS} - SSS_{situ})}{N} \tag{9}$$

where  $SSS_{MODIS}$  is the retrieved sea surface salinity from MODIS satellite data,  $SSS_{situ}$  is the reference sea surface salinity on grid point  $i$  and  $N$  is the number of grid points. Then, the empirical formula of  $SSS_{MODIS}$  (psu) which is based on equations 8 and 9 is

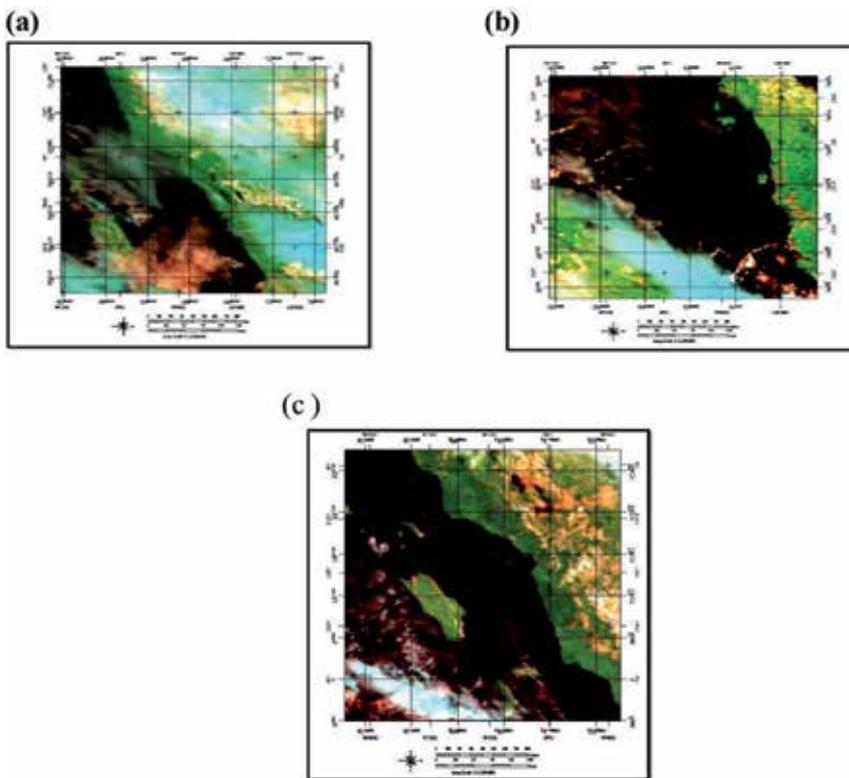
$$SSS_{MODIS}(\text{psu}) = 27.40 + 2.0I_1 - 3.4I_2 + 2.0I_3 + 2.2I_4 + 1.8I_5 + 0.3I_6 + - 1.8I_7 \pm 1.1 \tag{10}$$

This equation is different than the equation was obtained by Wong et al. [21] in terms of constant coefficient of linear model and involving the retrieved  $SSS_{MODIS}$  bias value. Finally, root mean square of bias (RMS) is used to determine the level of algorithm accuracy by comparing with in situ sea surface salinity. Further, linear regression model used to investigate the level of linearity of sea surface salinity estimation from MODIS data. The root mean square of bias equals [17]

$$RMS = [N^{-1} \sum_{i=1}^N (SSS_{MODIS} - SSS_{situ})^2]^{0.5} \tag{11}$$

#### 4. Results and discussion

The tsunami impact on sea surface salinity has been examined on three MODIS satellite data along Aceh coastal waters. These data are acquired on 23<sup>rd</sup>, 26<sup>th</sup> and 27<sup>th</sup> 2004 which represent pro, and post tsunami event, respectively (Figure 12). According to Marghany et al., [18], Moderate Resolution Imaging Spectroradiometer (MODIS) has 1 km spatial resolution and having 36 bands which are ranged from 0.405 to 14.285 $\mu\text{m}$  [24]. The MODIS satellite takes 1 to 2 days to capture all the scenes in the entire world, acquiring data in 36 spectral bands over a 2330 km swath.

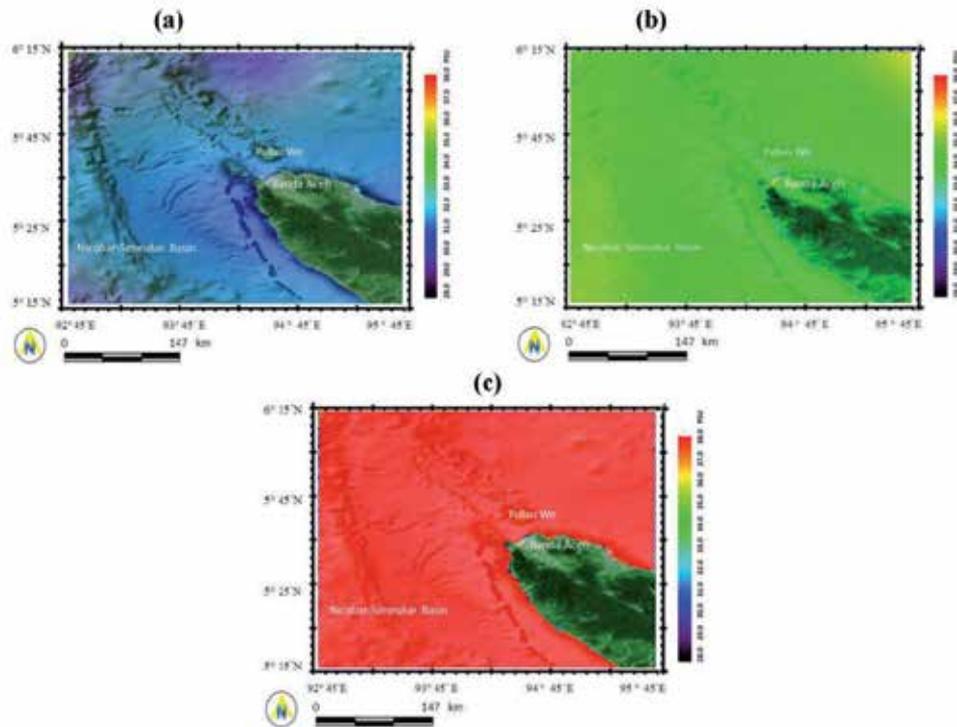


**Figure 12.** MODIS satellite data (a) pro tsunami event, (b) during Tsunami event and (c) post tsunami event.

It is interesting to find that heavy cloud covers have occurred on December 23<sup>th</sup> 2004 as compared to post tsunami event on December 27<sup>th</sup> 2004 (Figure 12). In fact, Aceh is located on tropical zone where the heavy cloud covers are one of the dominant features. Figure 9 shows the spatial variation of the salinity distribution along Aceh which are derived using the linear least square algorithm. On December 23<sup>th</sup> 2004, the sea surface salinity ranged between 28 psu to 31 psu. Nevertheless, during the tsunami event 25<sup>th</sup> 2004, the sea surface salinity dramatically

increased and ranged between 34 psu to 36 psu. The sea surface salinity was continued to increase post tsunami event of December 27<sup>th</sup> 2004 and ranged between 37 psu to 38 psu.

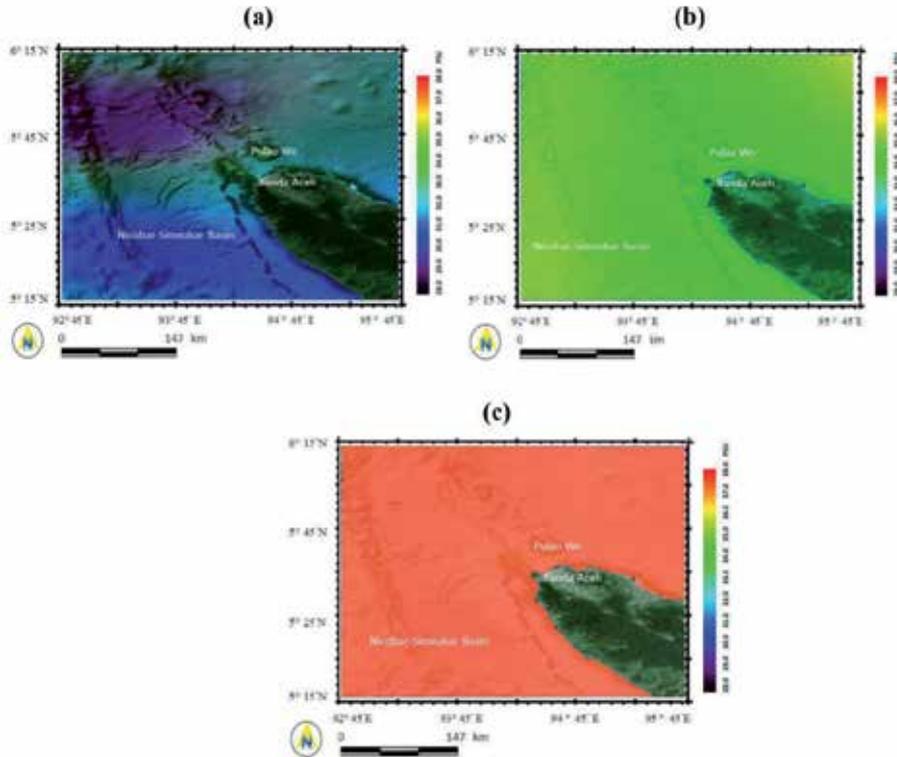
The isohaline contours of sea surface salinity which derived from in situ data. These data were acquired from <http://aquarius.nasa.gov/>. Figure 13 shows that the in situ sea surface salinity were increased dramatically. Pre tsunami event, the isohaline contours were ranged between 28.5 to 29.0 psu. However, the isohaline contours were increased to 36.7 psu during tsunami event and continued to increase to 37 psu (Figure 13).



**Figure 13.** Isohaline contours (a) pro tsunami event, (b) during tsunami event and (c) post tsunami event.

Both MODIS satellite data and in situ data show homogenous sea surface salinity pre tsunami and post tsunami with dramatically increment of sea surface salinity from 28.5 psu to 38.0 psu in coastal waters of Aceh. Figure 15 shows the comparison between in situ sea surface salinity measurements and SSS modeled from MODIS data. Regression model shows that SSS modeled by using linear least square methods are in good agreement with in situ data measurements. The degree of correlation is a function of  $r^2$ , probability (p) and root mean square of bias (RMSE). The relationship between estimated SSS from MODIS data and in situ data shows positive correlation as  $r^2$  value is 0.96 with  $p < 0.00007$  and RMS value of  $\pm 1.1$  psu. Further, accurate results of sea surface salinity in recent study can be explained as: using multiple MODIS bands i.e. 1 to 7 bands is a useful extension of linear regression model is the case where

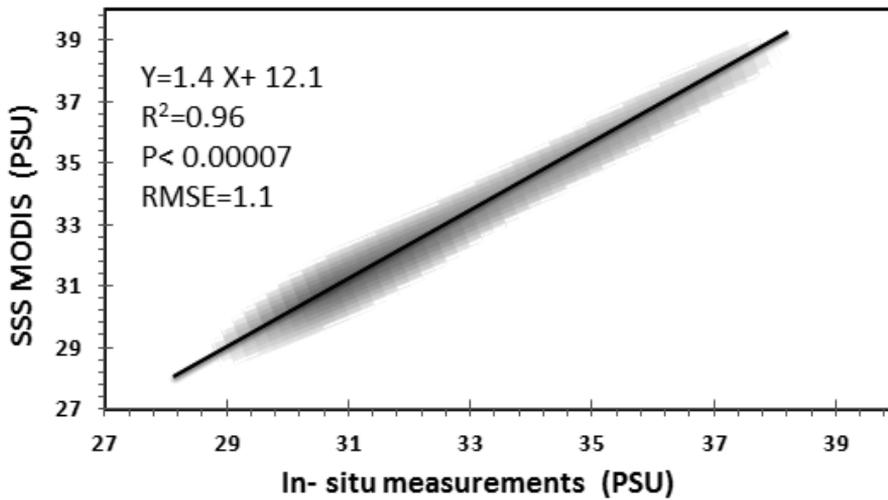
SSS is linear function of 7 independent bands. Such a practical is particularly useful when modeling SSS from MODIS data. This statement is agreed with Qing et al., [24].



**Figure 14.** Sea Surface Salinity (SSS) retrieved from MODIS data (a) pro tsunami event, (b) during Tsunami event and (c) post tsunami event.

Further, using least squares method derive a curve that minimizes the discrepancy between estimated SSS from MODIS data and in situ data[16]. This means that using a new approach based on least squares method would be to minimize the sum of the residual errors for the estimating SSS from MODIS data. Further, this study shows the possibilities of direct retrieving of the SSS from visual bands of MODIS satellite data without utilizing such parameter of colored dissolved organic matter,  $a_{CDOM}$  [15]. This study confirms the studies done by Marghany et al. [18]; Marghany and Mazlan [19].

Further dissolved salts, suspended substances have a major impacts on the electromagnetic radiation attenuation outside the visible spectra range [21]. In this context, the electromagnetic wavelength larger than 700 nm is increasingly absorbed whereas the wavelength less than 300 nm is scattered by non-absorbing particles such as zooplankton, suspended sediments and dissolved salts Marghany et al., [17].



**Figure 15.** Regression model between in-situ measurement and modeled SSS from MODIS data.

#### **4.1. Effect of changes the value of salinity towards the ocean**

The Sea Surface Salinity (SSS) along Banda Aceh has been changed dramatically pre and post tsunami event of 2004. During earlier run-up phases, the flows were extremely strong to erode sediments deposited especially when the coastal topography and bathymetry canalized the tsunami flow. In this regard, the backwash flows were potentially more erosive and powerful than run-up flows because of hyper-concentrated flow routing by coastal morphology. Therefore, tsunami resulted in massive deposits of sand, silt and fine gravel containing isolated boulders [25]. According to Font et al., [22], the area close to the shoreline was eroded by the passing wave front and deposition occurred when the wave front passed by. Just 6 min later, the direction of the flow close to the shoreline began to change to the seaward direction before the wave front reached the inundation limit. A portion of the sediment entrained by the wave front. In other words, tsunami of 2004 was caused brief coastal flooding with high overland flow velocities and strong abrasion and reworking of the nearshore materials (Figure 16). As most nearshore environments are composed by sand, mud and gravel. In this regard, the salinity values are extremely increased due to tremendous genetic sediment differences carried by wave from inland. According to Moore et al., [23] geochemical proxies have been provided evidence for saltwater inundation, associated coral and/or shell material, high-energy flows (heavy minerals, if present), and possible contamination associated with tsunamis. Finally, Pre and post event imagery can show the extent of erosion and areas most optimal for preservation of tsunami deposits (Figure 17). This information is necessary for comparing different events in the same locality. In addition, In the 2004 tsunami, sediments varied in their salinity levels, so sediments need to be assessed for salinity before any crops are planted in them.



**Figure 16.** High deposits flow due tsunami 2004 along Aceh coastal waters.



**Figure 17.** Satellite data along Aceh coastal waters (a) pro and (b) post tsunami event 2004 event.

In general, different grain size of sediments deposited in the coastal waters of Banda Aceh have increased the salinity in the sea surface. Indeed, these sediments contain different level of salts and mineral concentration. Post-tsunami, SSS has increased extremely because the sea coast was rough and turbid with suspended sediments (Figures 16 and 17). This is excellent evident of additional of extremely high amount of salts and minerals due high level of sediment deposit concentration in the coastal waters of Banda Aceh. This study confirms the work done by Saraf et al., [26].

## 5. Conclusions

Remote sensing technology has been recognized as powerful tool for environmental disaster studies. Ocean surface salinity is considered as a major element in the marine environment. In this study, we simulated the tsunami 2004 impacts on the Sea Surface Salinity along Banda Aceh using the least square algorithm from MODIS satellite data. This study shows significant variations in the values of SSS pro, during and post the tsunami event. The maximum salinity was observed post tsunami event was 38 psu as compared to before and during tsunami event. The results also show a good correlation between in situ SSS measurements and the SSS that is retrieved from MODIS satellite data with high  $r^2$  of 0.96 and RMSE of bias value of  $\pm 1.1$  psu. In conclusion, the least square algorithm is an appropriate method to retrieve SSS from MODIS satellite data. Clearly, the tsunami 2004 has significant impacts on the SSS because of high sediment deposit concentrations which added more salts and minerals to coastal waters of Banda Aceh.

## Author details

Maged Marghany\*

Institute of Geospatial and Science Technology (INSTEG), University of Technology, Malaysia

## References

- [1] Dawson A, Stewart I, Tsunami Geoscience Progress in Physical Geography. 2007, 31(6) 575– 590.
- [2] University of Washington. Earth & Space science: The Physics of Tsunamis. <http://earthweb.ess.washington.edu/tsunami/general/physics/physics.html>(accessed 11 November 2013).
- [3] Taru T, Japan Experts Warn of Future Risk of Giant Tsunami. <http://www.cosmostv.org/2012/04/japan-experts-warn-of-future-risk-of.html>. (accessed 11 November 2013).
- [4] Valdes R, Halabrin N, Lamb R How Tsunamis Work. <http://science.howstuffworks.com/nature/natural-disasters/tsunami.htm>. (accessed 11 November 2013).
- [5] Zahibo N, Pelinovsky E, Talipova T, Kozelkov A, Kurkin A. Analytical and numerical study of nonlinear effects at tsunami modelling. Applied Mathematics and Computation. 2006 (174) 795– 809.

- [6] Zaitsev A, Kurkin A, Levin B, Pelinovsky E, Yalciner A C, Troitskaya Y, Ermakov S. Modeling of propagation of the catastrophic tsunami (December 26, 2004) in the Indian Ocean, *Doklady Earth Sci.* 403 (3) (2005).
- [7] V V and Synolakis C E. Numerical modeling of 3-D long wave runup using VTCS-3. In *Long Wave Runup Models*, P. Liu, H. Yeh, and C. Synolakis (eds.), World Scientific Publishing Co. Pte. Ltd., Singapore, 1996 242–248.
- [8] Haugen K B, Løvholt F, Harbitz C B. Fundamental mechanisms for tsunami generation by submarine mass flows in idealised geometries. *Marine and Petroleum Geology.* 2005 (22) 209– 217.
- [9] Liu P L F, Wu T R, Raichlen F, Synolakis C. and Borrero J. Runup and rundown generated by three-dimensional sliding masses. *Journal Fluid Mech.* (2005) (536),107–144.
- [10] Kelmelis JA, Schwartz L, Christian C, Crawford M, and King D. Use of Geographic Information in Response to the Sumatra-Andaman Earthquake and Indian Ocean Tsunami of December 26, 2004. *Photogrammetric Engineering & Remote Sensing.* 2006 72 (8) 862-876.
- [11] Abbott P L. *Natural Disasters*. 6<sup>th</sup> edition, 2008, McGraw Hill Higher Education, New York.
- [12] Catherine J K, Gahalaut V K, Ambikapathy A, Kundu B, C. Subrahmanyam, S. Jade, Amit Bansal, R.K. Chadha, M. Narsaiah, L. Premkishore, D.C. Gupta. 2008 Little Andaman aftershock: Genetic linkages with the subducting 90°E ridge and 2004 Sumatra–Andaman earthquake. *Tectono physics* 2009 (479) 271–276.
- [13] Shulgin A, Kopp H, Klaeschen D, Papenberg C, Tilmann F, Flueh E R, Franke D, Barckhausen U, Krabbenhoeft A, Djajadihardja Y. Subduction system variability across the segment boundary of the 2004/2005 Sumatra megathrust earthquakes. *Earth and Planetary Science Letters* 365 (2013) 108–119.
- [14] Titov V V, Gonzalez F I, Bernard E N, et al. (2005): In: Real-time tsunami forecasting: challenges and solutions. *Nat. Hazards* 35(1), Special Issue, pp. 41–58. US National Tsunami Hazard Mitigation Program.
- [15] Ahn Y H, Shanmugam P, Moon J E, and Ryu J H. Satellite remote sensing of a low-salinity water plume in the East China Sea. *Annals of Geophys.* 2008 (26) 2019–2035.
- [16] Marghany M. Linear algorithm for salinity distribution modelling from MODIS data, *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009, 12-17 July 2009, Cape Town, South Africa*, 2009 (3) III-365-III-368.
- [17] Marghany M. Examining the Least Square Method to Retrieve Sea Surface Salinity from MODIS Satellite Data *European Journal of Science of Research.* 40 (2010) 377-386.

- [18] Marghany M Hashim M and Cracknell A P. Modelling Sea Surface Salinity from MODIS Satellite Data. Computational Science and Its Applications – ICCSA 2010, Lecture Notes in Computer Science. 2010, (6016) 545-556.
- [19] Marghany M and Hashim M. A numerical method for retrieving sea surface salinity from MODIS satellite data. International Journal of the Physical Sciences. 2011 6(13) 3116-3125.
- [20] Hu C, Chen T, Clayton P, Swarnzenski J, Brock I, and Muller-Karger F. Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL. Remote Sensing of Environment 2004 (93) 423-441
- [21] Wong M S, Kwan L, Young J K, Nichol J, Zhangging L, Emerson N Modelling of suspended solids and sea surface salinity in Hong Kong using Aqua/ MODIS satellite images. Korian Journal of Remote sensing. 2007 (23) 161-169.
- [22] Font E, Nascimento C, Omira R, Baptista MA, Silva PA. Identification of tsunami-induced deposits using numerical modelling and rock magnetism techniques: A study case of the 1755 Lisbon tsunami in Algarve, Portugal. Physics of the Earth and Planetary Interiors. 2010 (182) 187-198.
- [23] Moore A, Nishimura Y, Gelfenbaum G, Kamataki T, and Triyono R, Sedimentary deposits of the 26 December 2004 tsunami on the northwest coast of Aceh, Indonesia. Earth Planets Space, 2006 (58) 253-258.
- [24] Qing S Zhang J Cui T Bao Y. Retrieval of sea surface salinity with MERIS and MODIS data in the Bohai Sea. Remote Sensing of Environment. 2013 (136) 117-125.
- [25] Anthony E J. Developments in marine geology: Shore Processes and their palaeoenvironmental applications. Series Editor Chamley H, 2009 (4) 415-420.
- [26] Saraf A K, Choundhury and Dasgupta S. Satellite observation of the great mega thrust Sumatra earthquake activities. International journal of Geoinformatics. 2005, 1(4) 67-74.



---

# Morphodynamic Environment in a Semiarid Mouth River Complex Choapa River, Chile

---

Joselyn Arriagada, María-Victoria Soto and  
Pablo Sarricolea

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/57410>

---

## 1. Introduction

The knowledge about mouth systems, estuarine environments and associated processes, seen from the viewpoint of geomorphologic science in Chile, focusing on the contributions of Araya-Vergara [1, 2,3] for the central Chile and Paskoff [4], in the case of semi-arid.

The traditional European and American studies on deltas and estuaries can not address the reality of the most complex estuaries, through evolutionary and taxonomic schemes clear. Such is the case of the mouths on the coast of central Chile, which cannot be studied if it is enriched and modified some molds of the coastal geomorphology as a discipline [4].

According to work done by Pritchard and Caspers [1] indicates that the estuary is a term hydrological and not geomorphological, also provides a definition accepted by the estuarine specialists, defined as a coastal water body that has a free connection with the open sea and from which seawater is measurably diluted with freshwater derived from continental drainage. For its part, Caspers [1] emphasizes the importance of the tidal factor for the classification of the estuary.

Because estuarine delta expression, understood as the accumulation of banks that grow from the inside of the estuary to the sea, is applied to a delta developed into an estuary, it is necessary to seek ways of the land on which was developed estuary, and then pursued a concept of geomorphology. Thus, Araya-Vergara [3] concluded that this form in its most typical commonly is an estuary. Consequently, intends to call delta in *ria* to the internal deltas formed in valleys flooded by the sea.

Notably Davies [3] conceptualizes the estuarine delta as part of a continuum of types of mouth, forming the transitional phase between the estuary and the delta as such (understood as a typically delta marine and distal).

One of the conclusions reached by Araya-Vergara [3, 5] in central Chile, refers to the strong influence morphoclimatic that seems to define the domain of these environments mouth. The author acknowledges the following areas depending on the significance morphoclimatic:

Distal zone of deltas. Corresponds to the Morphoclimatic influence of the desert with summer contributions halogen derivatives of influences height.

Prograded *rias* zone. These coincide with the semi-arid conditions of the Norte Chico, from the Copiapo River (27 ° 20 'lat. S) to the Maipo river inclusive (33 ° 47 lat. S). These paleo – *rias* determined by Paskoff [5]. This reflects the higher solid load is attributed to the rivers of arid areas.

Estuarine delta zone. These seem to correspond to the transition area between the semi-arid and humid, from the mouth of the Rapel River (34 ° lat. S) and the Bio-Bio (37 ° lat. S) inclusive, where the river power is important to calibrate wide channel lower course and to provide plenty of substance delta, but also marine energy is sufficient to impede the progress of the progradation into the sea.

*Rias* Zone. This is connected with the moist and wet conditions in the Región de los Lagos (38° lat. S).

One way to characterize an estuarine system is through the characteristics of its key components. In this context, you can set the forms constituents of estuarine systems, which allows determining a zoning classification of these [3]:

A fluvial or proximal zone, with meandering estuarine.

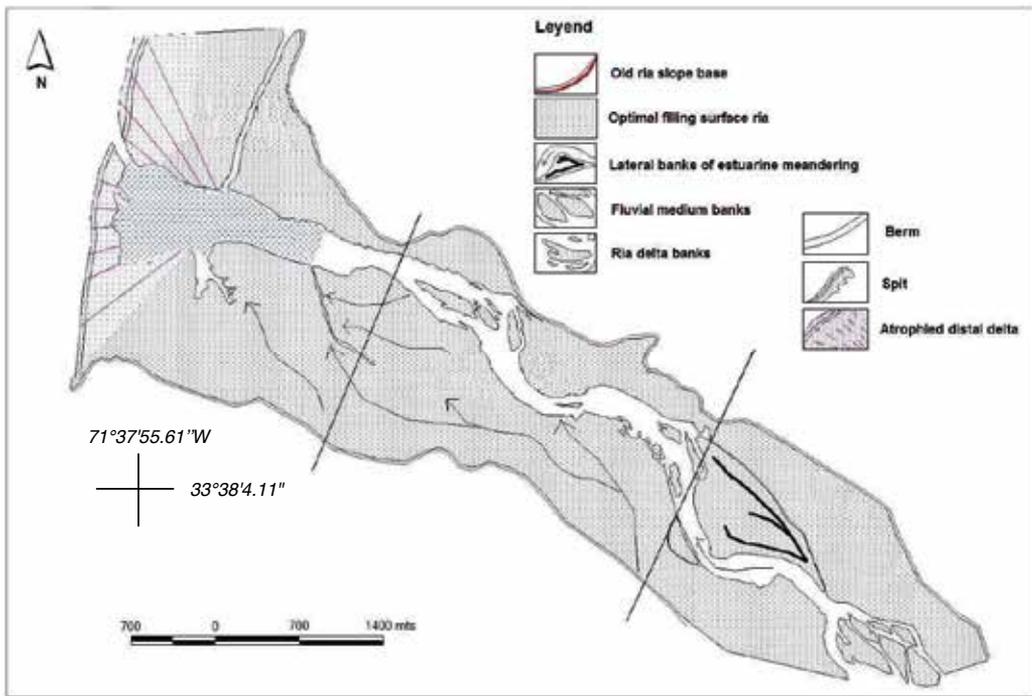
An interior or media delta zone, with media and distributary banks.

Lagoon or distal zone, with distal lake shore or barrier semiblock by spits or spit outer barrier (Fig. 1).

Proximal zone: indicating the presence of estuarine meanders, they owe their properties to the action of currents ebb and flow, which are related to the action of river and sea action. Estuarine meanders have an essential feature, which is reflected in the way you present: wide at the media and narrow at the ends, also the banks lateral has a shape of cusp.

Media zone: is determined by a change in the pattern of the channel downstream from the area of estuarine meanders. Clearly denoted start means estuarine banks, which have a form characterized by the pointed shape of its ends.

Distal: identify the existence of an estuarine lagoon, which is located between the distal part of the delta area and a spit in the mouth. The spit or barrier are built with the contributions of the littoral drift (in the case of the Chilean coast, it makes sense from south to north), which carries sediment from the existing extension ridge further south coast, the waves push these sediments towards the coast as an arrow, with its free end facing the waves and giving



**Figure 1.** Essential Components, Maipo estuarine system. From: Arriagada [6]

elongated in the same sense that the beach from which the materials. Its curvature, as a *hook*, is explained by the refraction of waves.

This classification has been applied to the case of estuarine systems of the Aconcagua, Maipo, Itata, among others. The case of the Aconcagua, considered by Cortez [7], Martinez and Cortez [8], has provided interesting results in relation to chemical, physical and mixing processes. In applying the model of Cooper [9] the system of Aconcagua river mouth, it was established its status as an estuary dominated by fluvial processes and the waves.

The Maipo River estuarine complex was analyzed by Arriagada [6], work from which it is concluded that estuarine dynamics, associated with changes in the essential forms has been greatly affected by the construction of the San Antonio port, constituting this element of anthropogenic influence in a major evolutionary factor. In this case, according to the analysis developed by applying the model of Cooper [9] concluded that the estuary has a tendency to be dominated by fluvial action. However, there is evidence to indicate the alternation of dominant effects, depending on the season.

Other estuarine systems have been analyzed Rapel, the Itata, the Maule and the Mataquito [3]. In each of them have been identified and characterized the essential forms of estuarine environments, where the presence of banks and spits cuspidate is ubiquitous in them.

Then, the most studied systems from the perspective geodynamics and where the model has been applied, corresponded to central Chile, with few information on systems related to semi-arid mouth.

### 1.1. Supply and mass transfer

The study of the evolution of coastal systems, must include not only the characterization of the estuaries, but must also incorporate the behavior and evolution of the mass associated dune system. This aspect is relevant because they are indicators of sediment supply and dynamics of coastal environments in general, and estuaries in particular systems [10, 11, 12].

In this aspect, Aqueveque [13] notes the development of three main generations of dunes, ancient and modern means, which are dominated by the degree of evolution and morphology, called *continuum dunaris*. Also, Araya-Vergara [14] adopts the term transmutation for the spatial variation of the dune system, and transmutation, to the change in the morphology of the dune system.

Thus, the mass balance of inner and outer beach is explained by the orientation of the beach, as well as the type of surf zone [14, 15]. This has been found in genetic- evolution-ary studies of barjan belts, categories of changes dunes and wave interaction system in large beach coves.

In this situation, one should expect that, in its evolution, the dunes adjacent to estuarine systems present a variation in the mass of dunes, backed by the dynamics and the associated contribution to the estuary.

In addition to the above, we can infer that the dynamic conditions of estuarine systems may be influenced by the processes associated with the basins. Such conditions are of global importance, as they are localized areas where most human settlements, whose impacts are significant morphological [16].

An example of these impacts was what happened at the mouth of the Maipo, where human influence changed the coastline less than 100 years [6]. Other cases have been studied by Federicci and Rodolfi [17, 18], who stated that the rapid retreat of the coastline in Ensenada de Atacama (Ecuador), is due to erosion caused by El Niño events, but also the indirect action of the destruction of wetlands (mangroves).

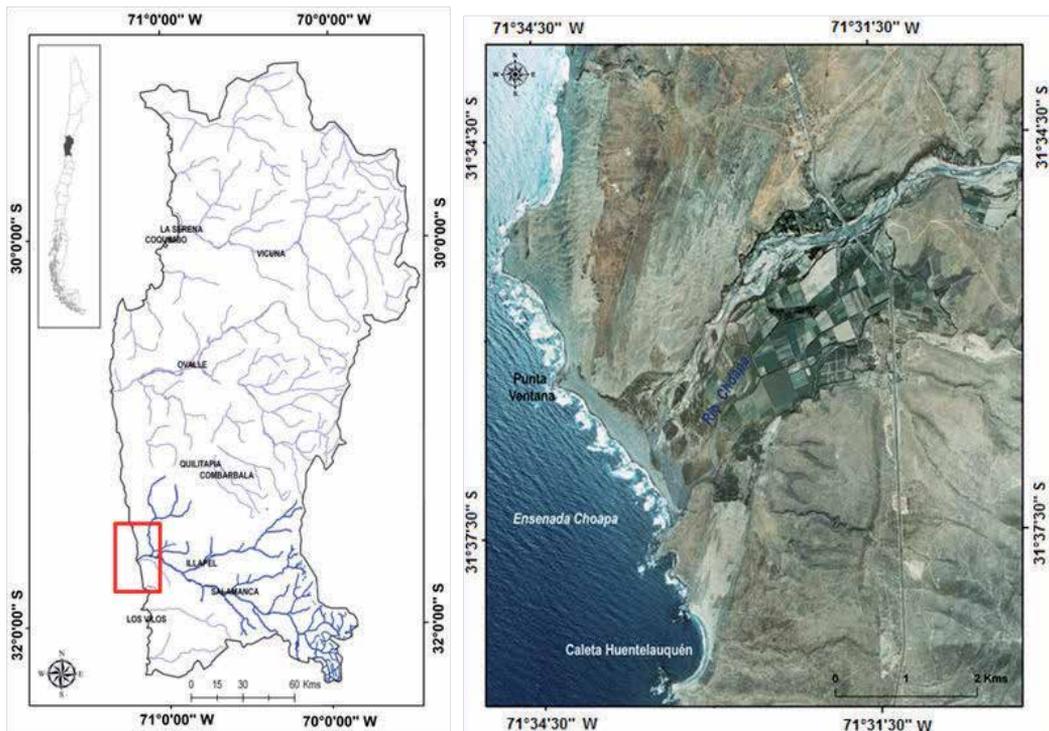
Araya-Vergara [5] taxonomy points to the classification of estuaries prograded from Copiapo to the Maipo River. These are associated with higher solid load of rivers is attributed to the semi-arid zones.

Consequently, the hypothesis behind this research, notes that in the mouth of Choapa system are recognized morphological traits that account for the evolution of these systems flows into the context of a domain morphoclimatic transitional semi-arid to temperate.

## 2. Materials and methods

### 2.1. Study area

In this research, we analyze the mouth of the river system Choapa (Region of Coquimbo) (Fig. 2). The rationale for the study of these systems is based on the differentiation of morphoclimatic domains that support these environments, which highlights the semiarid condition.



**Figure 2.** Study Area, Choapa river mouth in the national and regional context.

From the point of view of the climatic characteristics, the study area is located in the steppe climate with abundant cloudiness; geologically the area is inserted into the structural domain premesozoico correspond to the basement, where you can distinguish Choapa Metamorphic Complex [19].

There is also the presence of Huentelauquén formation, interpreted as a deposit of open marine platform [19]. It highlights the unconsolidated Quaternary deposits: aeolian, coastal, alluvial and colluvial. The sand dunes have a transverse component and indicate a contribution of wind SSW-NE direction. These correspond to barjanas and transverse coastal sand dunes.

The geographical context into which you insert the specific problems of the mouth of Choapa, is related to a watershed with land use changes and intensity, but less rapidly than in the case

of the Copiapo and Aconcagua river. For Choapa, recently begun to observe the process of conversion of agriculture and traditional farming to fruit farminggoats [20].

## 2.2. Methodology

The recognition of the essential systems linked to Choapa and Copiapo river mouths and sustained took place on the following concepts morphological from recognized estuarine systems theory for the domain morphoclimatic semiarid in Chile:

- a. The conceptual basis of the systematization of Araya-Vergara [3]. Among the essential forms to be identified, indicating the estuarine meanders and banks, cuspidate lateral banks, and spit estuarine lagoon, located in the proximal, media and distal, respectively.

The proximal part has meandering estuaries, characterized by fluvial influence that still operates in this area, in the media zone, it means banks are estuarine or cuspidate, due to the influence of river flows, as well as the marine influence. In the distal zone is possible to identify the presence of the spit, as well as estuarine lagoon.

Was realized the identification of the forms cited by photo interpretation of aerial photographs of a period of 30 years, as well as recognition in the field. The final information was mapped by using the software ArcGIS 9.2.

- b. The zoning of each estuarine system was carried out according to the scheme of Araya-Vergara [3], in this way, we use a general conceptual geomorphological model for all estuarine systems, classifying them in the following areas: proximal, media and distal. It included an analysis of the components in each zone.

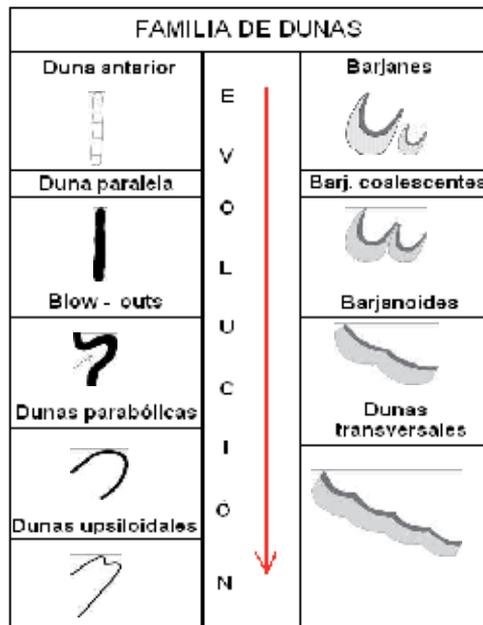
Analysis of morphodynamics temporal-spatial of estuarine complex through the identification of significant changes that occurred over a period covering the last 30 years, using the topographical maps and photo interpretation of aerial photographs (30 years).

Status of change and evolution of the systems analyzed.

The patterns of change and evolution trend of the morphology surrounding the estuary, was systematized by producing map geomorphology of each system, for the period of time as indicated. Thus, together with analysis of the essential forms of estuarine system, emphasis was placed on the dynamics of the dune systems. The main method of analysis refers to the mapping principle follow the design of dune ridges, which are classified according to the notions of sequence [21] and *continuum dunaris* [14, 22], combining families to determine the dunes (Fig. 3)

Patterns of change and development trends of these estuarine systems, related to the domain morphoclimatic in the context of headland bay beach systems in which they reside.

In the characterization of the near shore environment shows the importance of processes derived from swell. This results in temporal-spatial distribution of the energy of breaking waves. An extensive bibliography on the subject is reviewed by Martínez [23], who delivers new observations to Central Chile. Considered essential advances in active sensing and morphodynamic [24, 25, 26], however working with passive sensors.



**Figure 3.** Family of dunes and continuum dunaris. Source: Prepared by the author, from notions of Verstappen [21], Araya-Vergara [14, 22].

Analysis of wave-dominated beaches was performed according to the scheme of Wright & Short [27], as amended by Araya-Vergara [15] for use in aerial photographs. This taxonomy classifies the beach as dissipative, intermediate and reflective. The nomenclature used is summarized in Table 1.

Type of breaker zone	Nomenclature
Reflective	R
Low tide terrace	LTT
Transverse bars and rip	TBR
Rhythmic bars and rip	RBB
Long shore bar and trough	LBT
Dissipative	D

**Table 1.** Classification of type of wave-dominated beach [28, 29].

To detect change between 1997 and 2008 used three images Landsat 5 Thematic Mapper (TM) (table 2). Landsat TM operate in circular, sun synchronous, near polar orbits at a nominal altitude of 705 km. The inclination of the orbit is 98.2 degrees. The TM sensors view the entire earth every 16 days with a complete cycle of 233 orbits. Full scene products contain 5728 lines of 6120 pixels, corresponding to a ground area of 172 km by 184 km. The TM is a sensor, which

records a continuous strip image of the earth in 7 spectral bands, between 0.45 and 12.5 microns. The satellite images were corrected geometric and atmospheric correction in all the bands required. The spectral water index is a single number derived from an arithmetic operation of two or more spectral bands.

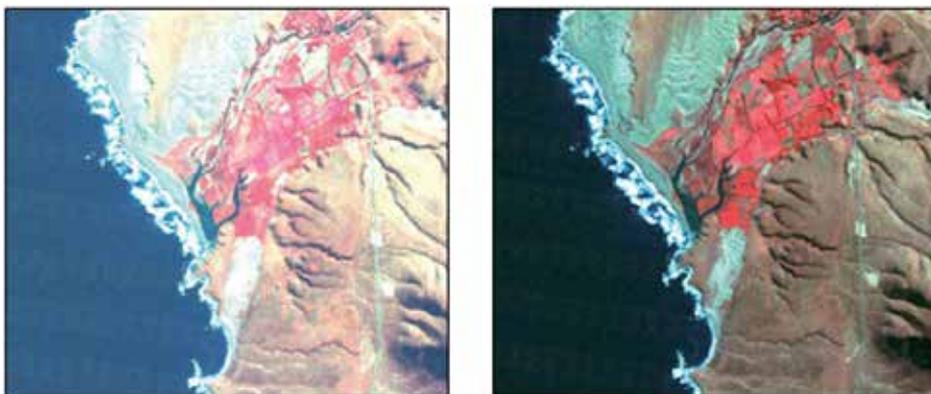
Acquisition Date	Satellite Images	Resolution
1997/06/26	Landsat 5 TM	30 m
2001/06/21	Landsat 5 TM	30 m
2008/09/12	Landsat 5 TM	30 m

**Table 2.** Used images.

An appropriate threshold of the index is then established to separate water bodies from other land-cover features based on the spectral characteristics. Adopting the format of the normalized difference vegetation index (NDVI), McFeeters [30] and Kuleli et al. [31] developed the normalized difference water index (NDWI), defined as:

$$NDWI = (\rho_{green} + \rho_{NIR}) / (\rho_{green} - \rho_{NIR}) \tag{1}$$

Where  $\rho_{green}$  and  $\rho_{NIR}$  are the reflectance of green and NIR bands, respectively. The NDWI value ranges from +1 to -1. McFeeters [30] set zero as the threshold. The NDWI data derived from Landsat TM (Thematic Mapper) of the years 1997, 2001 and 2008. They were processed in the software IDRISI Taiga for correcting some of the distortions mentioned by Chuvieco [32] needed in order to obtain a clear index. After the georeferencing, an atmospheric correction using the COST model was used. This model was chosen because his accuracy in semiarid regions, and specially for his accuracy on the Landsat TM blue band. (Fig.4). Moreover, the figure 5 illustrates the flow of this method.



**Figure 4.** Example of the COST model for atmospheric correction. The left image is a false color composition using the raw bands of the TM image for the mouth of the Choapa River for 2001. The right image is the same composition using the same bands after being corrected by the COST model on the software IDRISI Taiga.

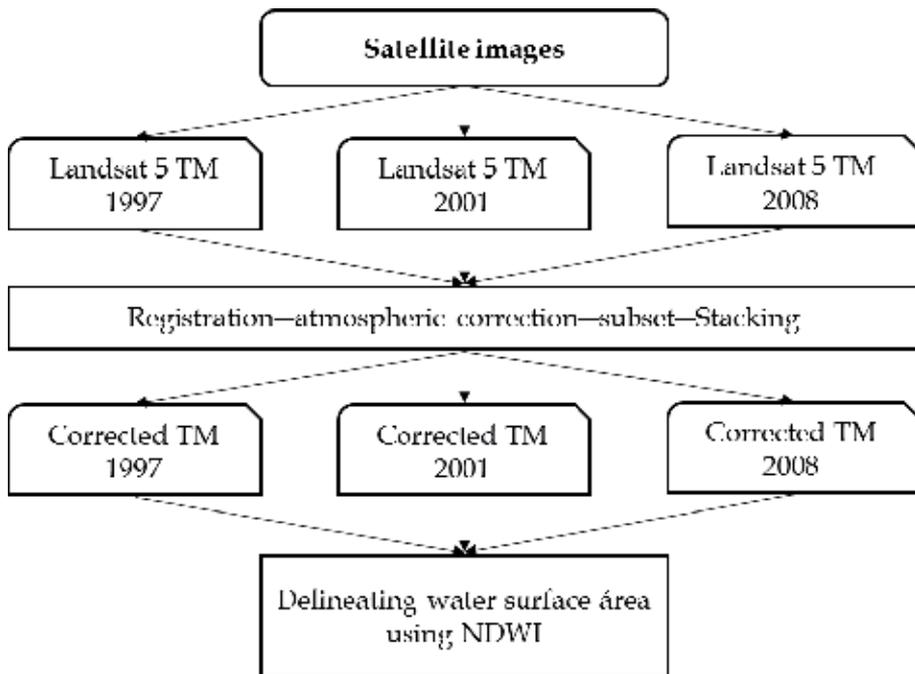


Figure 5. Flow diagram showing the methodology steps

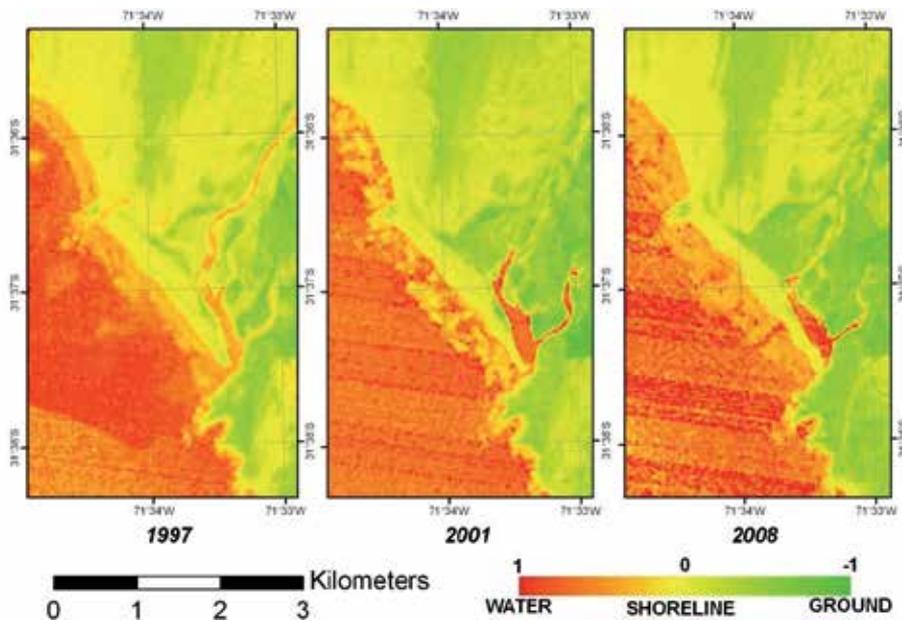
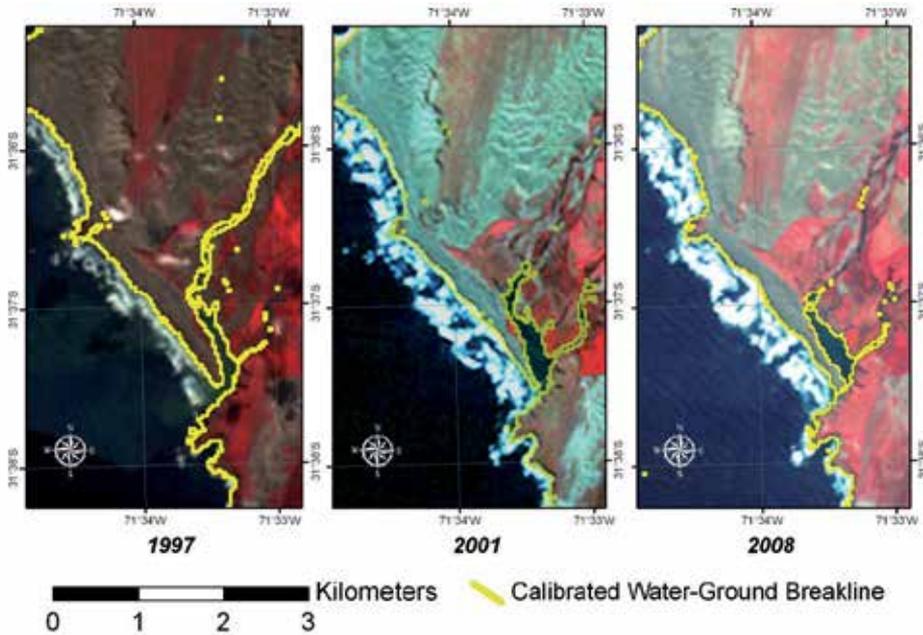


Figure 6. NDWI raw values on the corrected TM images for the years 1997, 2001 and 2008.

The result of the NDWI for each pixel, in which positive values represent water, and negative values represent ground (Fig. 6).

However, this index tend to show wet sand pixels as water due the reflectivity in the band 2, but these values, while positive, are close very close to zero, so a calibration of the water-ground break value was performed though photo interpretation, reaching the breaking value of  $-0.045$  for the 1997 image,  $0.053$  on the 2001 image, and  $0.0048$  for 2008. The resulting calibrated water-ground break line is shown on the Figure 7.



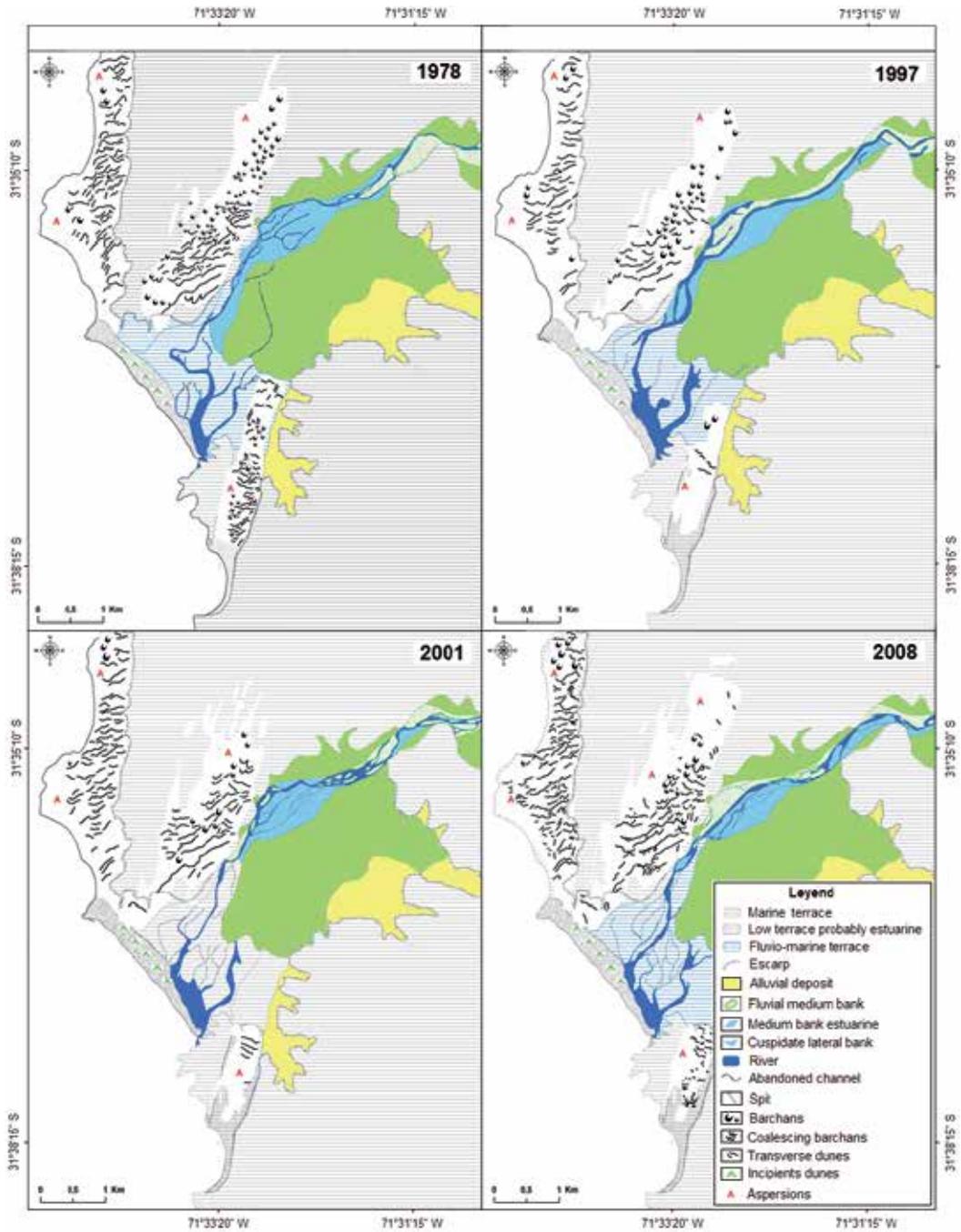
**Figure 7.** Resulting Water-Ground limit after calibrating the NDWI raw break value.

Finally, a reclassification of these values was made on the software ArcGis 9.2 creating polygons for water and ground categories, allowing a direct comparison for the years used on the images.

### 3. Results

#### 3.1. Geodynamics of the Choapa mouth system (1978 to 2008)

As shown in Figure 8, for 1978 highlights the presence of a channel in the proximal diffluent, where a glimpse of the river influence that dominates the environment in this sector, where there are large lateral river banks extension. Then in the media part of the system there are the rotation from fluvial media banks and lateral fluvial banks. Note the presence of abandoned



**Figure 8.** Evolution of the Choapa mouth river, since 1978-2008.

channels on the terrace river banks and in the estuarine side. In the distal part, called attention to the extension of the fluvial-marine terrace, which extends aprox. 2 km.

For 1997, in the proximal part, the estuary has the same type of channel had meandrante to 1978, this continues to the middle of the estuary, where the channel is meandering and winding. Towards the middle, it is possible to clearly distinguish estuarine character means banks, which are detected cuspidate, just before the river course becomes estuarine lagoon. Estuarine lagoon (distal) presents a greater extent compared to 1978. Relative to the sand dune system located south of the estuary include the presence of barjans and certain transverse sand dunes.

For the year 2001 in the distal zone is appreciated the meandering channel, as in previous years. In the middle part, is where the biggest differences you notice it tends to notice a greater number of banks means, indicating a greater sediment load, as a geodynamic behavior like braided channel. This extends to the edge of the middle, where the river channel again becomes meandering. In the distal part, the lagoon estuary is clearly distinguishable. We highlight the presence of abandoned channels, which could be indicators of lower river load.

The situation of 2008 is not much different to the situation of another year. In the proximal part, it shows the meandering course is appreciate where fluvial banks. On the average, banks still appear fluvial as well as the presence of banks estuarine media. The river course is still meandering, reaching its mouth.

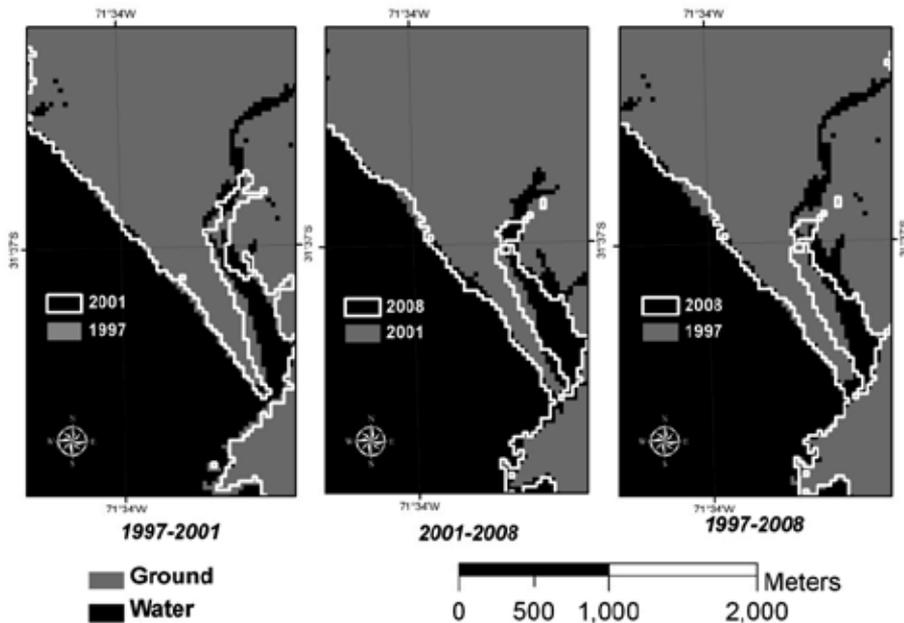
The results of the NDWI comparison (Fig. 9), they are completely consistent as they clearly show the variations and some changes on the cannels and banks. However, the most evident change shown by the comparison is the growth of the estuarine lagoon.

### **3.2. Characterization of Choapa Cove (1978 to 2008)**

For the year 1978, in the case of Choapa system, the most striking are the dunes. According to the figure 10, there is a predominance of transverse and barjanes sand dunes. In the coastal sector, there are some cross-sectional and longitudinal patterns less important, these dunes overlie the marine terrace, which is covered. Towards the north-inner estuary, dune field is quite distinguishable, repeating the pattern of tranverse dune, but more remarkable is the barjanic sand fiel, which is considerably large.

The system located to the west, corresponds to the evolution of barjanics sets, so it is assumed to be the oldest, on the contrary, the system located to the east (individual sand dunes barjanic) are recent and are moving in the direction NE.

For 1997, the sand dunes express changes in morphology and in the case of the northern section - coastal estuary, there appears to be an increase in the different dunes, where the cross pattern becomes much more noticeable (Fig 9). Barjan remain, but at lower densities. In the same figure, we see the same patterns and barjanics cross in the north-interior, although it refers to a decrease in the number of barjanic units. In turn, the spray area includes increasing magnitude, so one might infer that the barjan have evolved and in turn, transverse dunes have been dismantled.



**Figure 9.** NDWI comparison between 1997, 2001 and 2007. On the 1997-2008 comparison, the estuarine lagoon variation and growth is accurately detected by the NDWI.

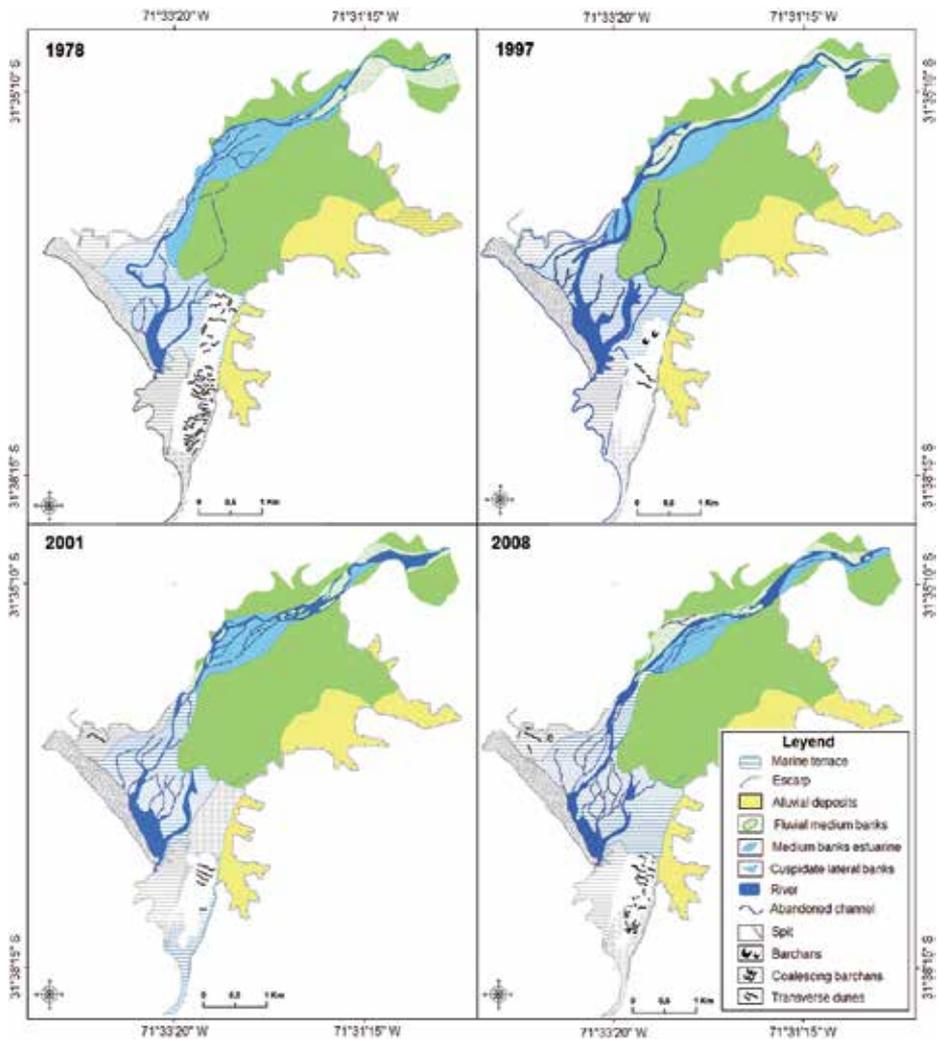
In 2001, major changes are in the dunes, as well as for 1978-1996. This year, they become barjanic and transverse barjanoides sand dunes, in the north, we can differentiate the extension of coverage for spraying (Fig. 9). There is a low representation of barjan. In the dune field located south of the estuary, it is noteworthy that the area near the estuary is devoid of mass contribution, and that is clearly seen that there is associated to vegetation, which has no clear traces of individual dunes.

For the year 2008 in the northern part there is a progressive decrease in barjans, reflected in the increased presence of barjanoides, as well as the largest area of spraying, these dune systems (barjanoides, barjans coalescing) cover large areas. Generate distinct forms. There is a continuous supply of material to the marine terrace, which is reflected not only in the designated spray, but also in the genesis and stabilization of dunes.

## 4. Discussion

### 4.1. Geodynamic system

In the Choapa mouth river system, the presence of the essential forms differ according to the area to be analyzed, so that in the proximal part, is a clear presence meandering course, while in the middle is not as noticeable presence estuarine banks. The depositional landforms in the



**Figure 10.** Dynamics Choapa Cove, since 1979 to 2008

middle (banks) have a great dynamic in the study period, varying in shape and location, a situation that hindered the categorization of these banks, in its identification as a river or estuarine. In the distal, zoning Araya-Vergara [3] is quite applicable, because the presence of estuarine lagoon and spit presence of essential forms to correspond to the characterization of this area. This condition is similar to the conditions analyzed in central Chile, as is the case of the Maipo [6], Aconcagua [7], Rapel [1, 3] and Itata [33].

From the point of view of the relative influence of the waves, tide and fluvial component [9, 34] in this estuary river influence seems to prevail, as evidenced by the presence of the spit, it demonstrate a continuous supply of sediment mass, which seems to be more evident in the last year analyzed. This condition is comparable to the situation in the Maipo [6] where the

spit remains almost intact over a long period of time. However, consider the case of the Maipo and Aconcagua morphoclimatic environments are restricted to the case of Choapa different, yet similar conditions exist particularly in the presence of the spit.

#### 4.2. The relationship with the zone surf dynamic

The Choapa Cove present the sand dune complexes identified by Araya-Vergara [13]. There is a predominance sand dunes systems from of genetic barjanic, which also presents transmutation conditions and transmutation [14]. However, Choapa Cove there is a high wave energy because the width of the surf zone is enough. This becomes important as the contribution of mass: in the case of Choapa Cove is possible to see the *continuum* of dunes, led to the sand dunes

These processes are quite notorious during the study period (30 years), indicating a strong dynamics related to sediment this is where the component of the wave and surf zone type is important (Fig. 11), main component is transversal.

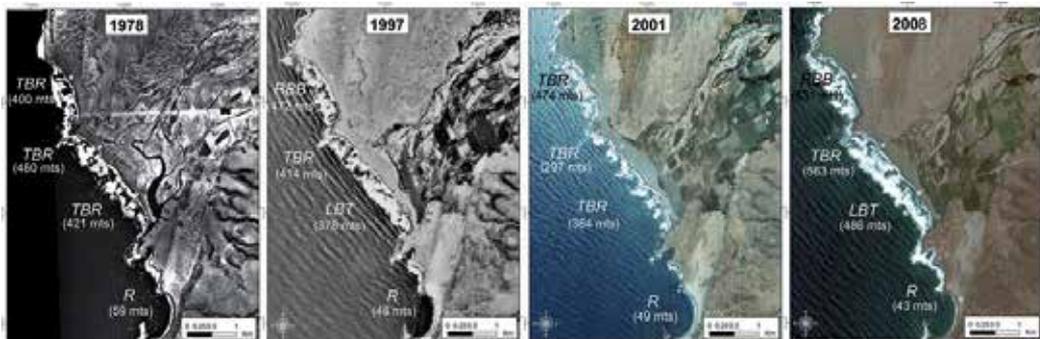


Figure 11. Space-time dynamics of wave-dominated beaches

## 5. Conclusions

In relation to the Choapa estuarine system, it presents conditions similar to those found in central Chile, mainly in the distal zone. Presents an estuarine lagoon flanked by a spit, which is quite stable in the middle section you can find river banks and estuaries, in alternancy. In the proximal part, the river is clearly a meandering stream.

You could say that the Choapa estuarine system realize greater sediment supply, because the river course change in the distal part in the review period: meandering change to a braided dynamic. So you can say that the system supplies material to the coast on a continuous basis over a period of 30 years.

In the case of Choapa mouth system, evolution has been much more pronounced: by 2007, the sand dunes barjanes change to coalescent barjan and barjanoides, and even become quite

transverse dunes. That is, in this system can be seen in 30 years the *continuum dunaris* from origin barjanic coastal sand dunes.

It is also important to note the change associated with the estuary Choapa, in 1978 is presented as meandering in its distal part, but in 2007, there are differential patterns that denotes a greater amount of sediment, resulting in an barided river in its distal zone. In addition, it adds to the change in the direction of the river. In the Choapa river system there are basic shapes, such as media cuspidate banks, lateral cuspidate banks and spit, which presents constant throughout the period under review.

For this reason, it opens the discussion about the limits proposed by Araya-Vergara [3] regarding the classification of estuaries and the possible conclusion that morphologically mouth system can be termed as estuarine delta and not as part of prograded rias system, by displaying dynamics similar characteristics with the mouths of central Chile, as the Aconcagua and Maipo rivers.

The calculation of the NDWI is a valid method for an automatic delimitation of a shoreline or boundary between ground and water corpses (like the estuarine lagoon) as is shown successful using the presented materials and being consistent with the overall data. However, the accuracy of this index is dependent of the pixel resolution, the preciseness of the geometric and atmospheric corrections, but even more dependent of the calibration phase of the break value between ground and water. This break value differs in every image but is always close to 0. A care photo interpretation can achieve a satisfactory boundary line between ground and water and obtain the data needed for an easy comparison between years.

## Acknowledgements

To FODECYT and the Faculty of Architecture and Urbanism of the University of Chile.

## Author details

Joselyn Arriagada<sup>1,2,3\*</sup>, María-Victoria Soto<sup>2</sup> and Pablo Sarricolea<sup>2</sup>

\*Address all correspondence to: joarriag@uchile.cl

1 Proyecto FONDECYT, Chile

2 Departamento de Geografía, Facultad de Arquitectura y Urbanismo, Universidad de Chile, Chile, Santiago, RM, Chile

3 Laboratoire Environnements et Paléoenvironnements Océaniques et Continentaux (EPOC), Université Bordeaux 1, Francia, 351 Cours de la Libération, Talence, France

## References

- [1] Araya-Vergara J. Contribución al estudio de los procesos estuariales en las desembocaduras de los ríos Rapel y Maipo. *Revista Informaciones Geográficas* 1970; 20:17-38.
- [2] Araya-Vergara J, Castro C, Andrade B. Análisis de la turbidez fluvial en el mar mediante teledetección y control terrestre. In: Instituto Geográfico Militar de Chile, editors. *Primer Symposium Internacional sobre Percepción Remota*. 1978, Apr. Santiago de Chile.
- [3] Araya-Vergara J. El concepto de delta en ría y su significado en la evolución litoral (ejemplo de Chile Central). *Informaciones Geográficas Chile*. 1981; 28:19-42.
- [4] Paskoff R. *Recherches geomorphologiques dans le Chili Semi-aride*. Vol 1. Bordeaux: Bordeaux Biscaye; 1970.
- [5] Araya-Vergara J. Análisis de la localización y de los procesos y formas predominantes de la línea litoral de Chile. *Informaciones Geográficas de Chile*. 1982; 29:35-55.
- [6] Arriagada J. Cambios en el sistema estuarial del Maipo y su relación con obras portuarias, Chile Central. Memoria de título de Geógrafo, Universidad de Chile; 2005.
- [7] Cortéz C. Observaciones dinámicas y geomorfológicas en el estuario del Aconcagua, Chile Central. Memoria de título de Geógrafo, Universidad de Chile; 2002.
- [8] Martínez C, Cortéz C. Características hidrográficas y sedimentológicas en el estuario del río Aconcagua, Chile Central. *Revista de Geografía Norte Grande*. 2007; 37:63-74.
- [9] Cooper JA. Sedimentary processes in the river-dominated Mvoti stuary, South Africa. *Geomorphology*. 1994; 9:271-300.
- [10] Arriagada J. Geomorfología comparada en la zona semiárida de Chile: casos Copiapó y Choapa. Tesis de Magister. Universidad de Chile: 2009.
- [11] Arriagada J, Soto MV, Castro CP. Dinámica del complejo estuarial del Choapa: istema estuario-playa-duna, región de Coquimbo. In: *Resumen XXIX Congreso Nacional y XIV Internacional de Geografía*. 2008 Oct. Temuco, Chile.
- [12] Soto MV, Arriagada J, Castro CP, Märker M, Rodolfi G. Aspectos geodinámicos de n paleo estuario del desierto marginal de Chile. Río Copiapó. *Revista de Geografía Norte Grande*. 2010; 46:123-135.
- [13] Aqueveque C. Análisis de suelos desarrollados en dunas litorales antiguas de Chile entral. Memoria de título de Geógrafo, Universidad de Chile; 2008.
- [14] Araya-Vergara J. The evolution of modern coastal dune systems in central Chile. n: *International Geomorphology 1986 Part II*. V.Gardiner. John Wiley & Sons Ltd.; 987. p. 1231-1244.

- [15] Araya-Vergara J. Sistema de interacción oleaje-playa frente a los ergs de Chanco y Arauco, Chile. *Gayana Oceanología*. 1996; 4(2):159-167.
- [16] Verhaar PM, Biron PM, Ferguson RI, Hoey TB. A modified morphodynamic model or investigating the response of Rivers to short-term climate change. *Geomorphology*. 2008; 101(4):674-682.
- [17] Federici P, Rodolfi G. Rapid shorelines retreat along the Esmeraldas Coast, Ecuador: natural and man-induced processes. *Journal of Coastal Conservation*. 001; 7:163-170.
- [18] Federici P, Rodolfi G. Geomorphological features and evolution of the Ensenada de Atacames (Provincia de Esmeraldas, Ecuador). *Journal of Coastal Research*. 2004; 20(3):700-708.
- [19] Rivano S, Sepulveda P. Hoja Illapel, Región de Coquimbo, 1: 250.000. In: Servicio Nacional de Geología y Minería, editors, *Serie Geología Básica (69)*. 1986. Santiago de Chile.
- [20] Castro CP, Meza M, Soto MV. Pérdida de calidad de suelos en el periurbano de ciudades intermedias. Salamanca y Quillota. In: *Anales XXIX Congreso Nacional y XIV Internacional de Geografía*. 2008 Oct. Temuco, Chile.
- [21] Verstappen H. On dune types, families and sequences in areas of unidirectional winds. *Gottiner Geogr. Abh.* 1972; 60:341-353.
- [22] Araya-Vergara J. Significance of barchans in Beach-dune Systems interactions in Central Chile. *Thalassas*. 1986; 4(1):23-27.
- [23] Martínez C. 2001. El efecto de ensenada en los procesos litorales de las ensenadas de Valparaíso, Algarrobo y Cartagena, Chile Central. Tesis de Magíster, Universidad de Chile; 2001.
- [24] Marghany M. Modelling shoreline rate of changes using holographic interferometry. *International Journal of the Physical Sciences*. 2011; vol. 6(34):7694-7698.
- [25] Marghany M. DEM Reconstruction of Coastal Geomorphology from DINSAR. In Murgante B. et al., editors *Lecture Notes in Computer Science (ICCSA 2012), Part II, LNCS 7335*; 2012. p. 435-446.
- [26] Marghany M. DInSAR technique for three-dimensional coastal spit simulation from Radarsat-1 fine mode data. *Acta Geophysica*. 2013 Apr.; vol. 61(2):478-493.
- [27] Short AD. A note on the control of beach type and change, with S. E. Australian Examples. *Journal of Coastal Research*. 1987; 3(3): 387-395.
- [28] Wright LD, Short AD. Morphodynamic variability of surf zones and beaches: A synthesis. *Marine Geology*. 1984; 56 (1): 93-118.
- [29] Short AD. Waves-dominated beaches. In Short A, editors. *Handbook of beach and shoreface morphodynamics*. Chichester; John Wiley & Sons; 1999. p. 173-191.

- [30] Mcfeeters SK. The use of Normalized Difference Water index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*. 1996; vol 17(7): 1425–1432.
- [31] Kuleli T, Guneroglu A, Karsli F, Dihkan M. Automatic detection of shoreline change on coastal Ramsar wetlands of Turkey. *Ocean Engineering*. 2011; 38:1141-1149.
- [32] Chuvieco E. *Teledetección Ambiental. La observación de la tierra desde el espacio*. Barcelona: Ariel; 2002.
- [33] Castro CP. (1987). *Influencia del río Itata en el desarrollo de líneas de costa arenosas en la parte sur de Chile Central*. Memoria de Título de Geógrafo. Universidad de Chile; 1987.
- [34] Dalrymple R, Zaimln B, Boyd R. Estuarine facies models. Conceptual basis and stratigraphic implications. *Journal of Sedimentary Research*. 1992; 62: 1130-1146.

*Edited by Maged Marghany*

Nowadays, advanced remote sensing technology plays tremendous roles to build a quantitative and comprehensive understanding of how the Earth system operates. The advanced remote sensing technology is also used widely to monitor and survey the natural disasters and man-made pollution. Besides, telecommunication is considered as precise advanced remote sensing technology tool. Indeed precise usages of remote sensing and telecommunication without a comprehensive understanding of mathematics and physics. This book has three parts (i) microwave remote sensing applications, (ii) nuclear, geophysics and telecommunication; and (iii) environment remote sensing investigations.

Photo by bme1ofo / iStock

**IntechOpen**

