IntechOpen

Models and Technologies for Smart, Sustainable and Safe Transportation Systems

Edited by Stefano de Luca,
Roberta Di Pace and Chiara Fiori

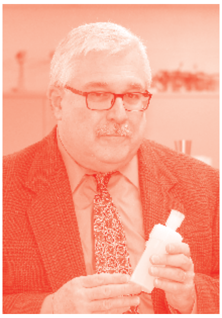
Models and Technologies for Smart, Sustainable and Safe Transportation Systems

Edited by Stefano de Luca,
Roberta Di Pace and Chiara Fiori

Published in London, United Kingdom

IntechOpen

Supporting open minds since 2005



Models and Technologies for Smart, Sustainable and Safe Transportation Systems
http://dx.doi.org/10.5772/intechopen.87681
Edited by Stefano de Luca, Roberta Di Pace and Chiara Fiori

Contributors
Carlos Hugo Criado del Valle, Parichehr Scharifi, Bahman Moghimi, Camille Kamga, Salvatore Antonio Biancardo, Francesca Russo, Nunzio Viscione, Cristina Oreto, Hai L. Vu, Atousa Tajaddini, Geoffrey Rose, Kara M. Kockelman, Pradip Kumar Sarkar, Fabien Leurent, Francesco Viti, Marco Rinaldi, Georgios Laskaris, John Baras, Fatemeh Alimardani, Nilesh Suriyarachchi, Faizan Muhammad Tariq, Joan Harvey, Orlando Giannattasio, Giulio Erberto Cantarella, Stefano de Luca, Margherita Mascia, Chiara Fiori, Roberta Di Pace, Luigi Pariota, Facundo Storani

© The Editor(s) and the Author(s) 2021
The rights of the editor(s) and the author(s) have been asserted in accordance with the Copyright, Designs and Patents Act 1988. All rights to the book as a whole are reserved by INTECHOPEN LIMITED. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECHOPEN LIMITED's written permission. Enquiries concerning the use of the book should be directed to INTECHOPEN LIMITED rights and permissions department (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 these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in London, United Kingdom, 2021 by IntechOpen
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 5 Princes Gate Court, London, SW7 2QJ, United Kingdom
Printed in Croatia

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

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

Models and Technologies for Smart, Sustainable and Safe Transportation Systems
Edited by Stefano de Luca, Roberta Di Pace and Chiara Fiori
p. cm.
Print ISBN 978-1-83880-802-0
Online ISBN 978-1-83880-823-5
eBook (PDF) ISBN 978-1-83880-824-2

Meet the editors

Stefano de Luca is a full professor of Transportation Planning and Transportation Systems Theory, University of Salerno, Italy. Currently he is director of the Transportation Systems Analysis laboratory, and rector's delegate to Transport and Mobility, University of Salerno. His research focus includes transportation planning techniques, choice modelling, signal settings design, traffic assignment models and algorithms, freight/passenger terminal simulation, and optimization. He serves on the Editorial Advisory Board for Transportation Research Part F, the *Journal of Advanced Transportation and Sustainability*. He has authored more than 100 book chapters and journal articles. He is also a consultant for the Italian Ministry of Transportation, the Transport Commission of the Campania Region, and the Salerno and Avellino Transportation Departments. He is a member of the IEEE Intelligent Transportation Systems Society, the Italian Association of Transport Academicians, and the Italian Transport Policy Society.

Roberta Di Pace obtained an MSc and Ph.D. in Transportation Engineering from the University of Naples "Federico II," Naples, Italy, in 2005 and 2009, respectively. She is a professor in Transportation Engineering, Department of Civil Engineering, University of Salerno, Italy. She is an aggregate professor of Technique and Transport Economics and Transportation Systems Design. Since 2010, she has been a member of the Transportation Planning and Modelling Laboratory. Her main research fields include the development of analytical tools for advanced traveller information systems, traffic flow modelling, network signal setting design, and advanced traffic management systems. She is a member of the IEEE Intelligent Transportation Systems Society and IEEE Women in Engineering.

Chiara Fiori is an assistant professor in the Department of Civil Engineering, University of Salerno, Italy. She has international research experience as a visiting scientist at the MobiLab Transport Research Group, University of Luxembourg (2020–2021); the European Commission, Joint Research Center, Directorate for Energy, Transport and Climate Change (2017–2018); Virginia Tech Transportation Institute, USA (2015–2016); and others. Her research interests include testing, modelling, and simulation of electrified vehicles' powertrains (for personal, public, and freight mobility) and their impact assessment on eco-routing and the electric power grid; integration of microscopic traffic flow and energy consumption models, and effects on traffic control systems; well-to-wheels analysis; modelling, simulation, and impact assessment of port operations in urban contexts; and Multi-Vehicle Dynamic Traffic Assignment.

Contents

Preface XIII

Section 1
Travel and Driving Behavior 1

Chapter 1 3
Adaptive Travel Mode Choice in the Era of Mobility as a Service (MaaS): Literature Review and the Hypermode Mode Choice Paradigm
by Stefano de Luca and Margherita Mascia

Chapter 2 27
Recent Progress in Activity-Based Travel Demand Modeling: Rising Data and Applicability
by Atousa Tajaddini, Geoffrey Rose, Kara M. Kockelman and Hai L. Vu

Chapter 3 55
Attitudes and Behaviours in Relation to New Technology in Transport and the Take-Up amongst Older Travellers
by Joan Harvey

Chapter 4 77
Cognitive Profile of Optimistic Offender Drivers Affected by Psychological Interventions for a Sustainable and Safer Driving's Behavior
by Carlos Hugo Criado del Valle and Parichehr Scharifi

Section 2
Transportation System Design and Management 93

Chapter 5 95
Optimal Management of Electrified and Cooperative Bus Systems
by Francesco Viti, Marco Rinaldi and Georgios Laskaris

Chapter 6 119
Models and Methods for Intelligent Highway Routing of Human-Driven and Connected-and-Automated Vehicles
by Fatemeh Alimardani, Nilesh Suriyarachchi, Faizan M. Tariq and John S. Baras

Chapter 7 145
Centralised Traffic Control and Green Light Optimal Speed Advisory Procedure in Mixed Traffic Flow: An Integrated Modelling Framework
by Roberta Di Pace, Chiara Fiori, Luigi Pariota and Facundo Storani

Chapter 8 163
Towards Shared Mobility Services in Ring Shape
by Fabien Leurent

Chapter 9 183
Transit Signal Priority in Smart Cities
by Bahman Moghimi and Camille Kamga

Section 3
Technological and Digital Advances 199

Chapter 10 201
Advanced Vehicles: Challenges for Transportation Systems Engineering
by Orlando Giannattasio and Giulio E. Cantarella

Chapter 11 215
Driver Assistance Technologies
by Pradip Kumar Sarkar

Chapter 12 237
BIM Approach for Smart Infrastructure Design and Maintenance Operations
by Salvatore Antonio Biancardo, Nunzio Viscione, Cristina Oreto and Francesca Russo

Preface

The transportation system plays a significant role in making cities sustainable, inclusive, and safe.

Innovative and smart mobility systems are expected to change the way transportation systems work and move traditional mobility paradigms towards a more self-conscious and sustainable behavior. In this context, a great research effort is necessary to cover those issues related to transportation planning, management and design.

Indeed, it is fundamental to understand and model much more complex travel paradigms, to identify robust and effective strategies for traffic control, and to design transit services or personal mobility services. Furthermore, it is crucial to identify real opportunities that the current technological revolution may present for making transportation systems smarter, safer, and more sustainable.

This book proposes a methodological and technological approach to the aforementioned issues. It proposes a comprehensive framework that may support researchers, analysts/practitioners, decision-makers, industries, and investors.

The first section of the book examines travel and driving behavior, investigating the necessity of more realistic and flexible paradigms for mode choice and modeling complex travel patterns. It also focuses on two very timely issues: attitudes and behavior of old travelers towards new technologies and the role of cognitive profile of offender drivers.

The second section addresses the crucial challenges regarding the management and optimization of the multiple components of a transportation system. It investigates route guidance in connected/autonomous environments, management of electric and cooperative buses, traffic control in hybrid traffic flow conditions, and optimization of transit or shared services.

The third section is an overview of three of the most promising technological advances of transportation systems: advanced vehicles, driving technologies, and the Building Information Modeling revolution.

The book includes the following twelve chapters.

Chapter 1 is a brief introduction to the Mobility as a Service (MaaS) paradigm and the main modeling issues that researchers should address in the near future. Then, it provides an extensive review of the state of the art of mode choice approaches and introduces a bi-level mode choice behavior paradigm that explicitly accounts for real-time events and travelers' adaptive behavior.

Chapter 2 describes the state of art and practice of activity-based modeling approaches, including the ongoing research covering both demand and supply considerations. It proposes possible solutions to improve the modeling approach, as well as existing opportunities for effective spatial transferability to new geographical contexts along with expanding the applicability of ABMs in transportation policy-making.

Chapter 3 considers the needs of older travelers and how new technology can meet some of those needs and what is necessary for it to be appropriate to and usable by older travelers. It covers what happens as travelers get older and how changes in transport systems can be made much more useful and usable for older people. In addition, the chapter considers MaaS in the United Kingdom, especially for older and infirm people.

Chapter 4 investigates the cognitive profile of optimistic offender drivers and proposes possible interventions for sustainable and safer driving behavior. In particular, it highlights the lack of understanding of the true impact that external factors can have on driving and how offender drivers overestimate their abilities in avoiding accidents.

Chapter 5 investigates an integrated management approach exploiting the potentials of the new Cooperative Intelligent Transportation Systems (C-ITS) to meet the requirements of the next generation of Public Transport (PT) for electrified and cooperative bus systems, thus taking into account the additional complexity of periodically recharging electric buses during operation using the dedicated infrastructure. Specifically, the proposed system is tested and evaluated in simulation showing the benefits of electrified and cooperative bus systems.

Chapter 6 explores the integration of two traffic management strategies: ramp metering (RM) and route guidance (RG). It focuses on the interaction between automated vehicles (AVs) and human-driven vehicles (HDVs). Indeed, it seems unrealistic that all HDVs will suddenly be replaced by AVs in the near future. Rather, AVs will be introduced in the presence of HDVs. Therefore, there is a need to consider cases where it becomes necessary to model the interactions between AVs and HDVs. The key areas of interest involve traffic routing and management, optimal highway merging, and intelligent overtaking behaviors.

Chapter 7 investigates the integration of traffic management strategies in the era of Cooperative and Connected Intelligent Transportation Systems. The focus is on strategies to optimize vehicle behavior at junctions. In accordance with the literature, one of the proposed approaches is that of the Green Light Optimal Speed Advisory (GLOSA), which provides a warning to the driver regarding the best speed to maintain when approaching the junction by avoiding stops at junctions. The chapter proposes a whole modeling framework based on the integration of GLOSA and traffic control strategy. The framework is also applied considering a real case study.

Chapter 8 focuses on the "systemic qualities" of shared mobility services adopting a ring format, as well as explores the conditions required to establish a ring system in urban settlements. It evaluates whether a ring system makes it possible to cover a relatively large geographical area while also establishing service cycles for shared vehicles. The modal models share a four-tier architecture that involves: (1) the physical operations of the service and the laws governing its vehicle flow, (2) the balance between supply and demand, (3) the optimized service management, and (4) the strategic positioning of the service in terms of technologies. The chapter applies the model to analyze some scenarios.

Chapter 9 provides a comprehensive review of transit signal priority models. Generally, giving priority to public transport vehicles at traffic signals is one of the traffic management strategies deployed at emerging smart cities to increase the quality of service for public transit users. This is key to breaking the vicious cycle of congestion that threatens to bring cities into gridlock. The chapter presents studies in the following categories: signal priority and different control systems, passive versus active priority, predictive transit signal priority, the priority with connected vehicles, multi-modal signal priority models, and other practical considerations.

Chapter 10 addresses the challenges that AVs introduce for transportation systems engineering. It discusses the most likely positive and negative effects of mixed flow expected in the near future, the main classifying criteria such as ownership, on-board technologies (sensors), and the most effective tools already available for macroscopic analysis of multi-vehicle-type transportation systems to highlight the need to update and/or develop new mathematical models.

Chapter 11 describes Driver Assistance Technology, which is emerging as a new driving technology popularly known as Advanced Driver Assistance Systems (ADAS). The chapter provides a complete overview on this topic explaining the functioning of Driver Assistance Technology with the help of its architecture and various types of sensors.

Chapter 12 provides an overview of a creative process for the digitalization of existing roads. It depicts the approach, known as the reverse engineering method, by (1) modeling the 3D digital terrain model; (2) creating the horizontal alignment, vertical profiles, and editing cross-sections; and (3) modeling the 3D corridor. As a response to long-term development between Big Data, Building Information Modeling (BIM), and road engineering, this chapter offers innovative and practical solutions for integrating road design and pavement analysis for better management and optimization of road pavement maintenance.

Stefano de Luca, Roberta Di Pace and Chiara Fiori
Department of Civil Engineering,
University of Salerno,
Italy

Section 1

Travel and Driving Behavior

Chapter 1

Adaptive Travel Mode Choice in the Era of Mobility as a Service (MaaS): Literature Review and the Hypermode Mode Choice Paradigm

Stefano de Luca and Margherita Mascia

Abstract

Mobility as a Service (MaaS) is becoming a "fashionable" solution to increase transport users' satisfaction and accessibility, by providing new services obtained by optimally integrating sustainable modes, but also guaranteeing mass transport and less sustainable modes, guaranteeing fast and lean access/egress to the mass transport. In this context, the understanding and prediction of travellers' mode choices is crucial not only for the effective management of multimodal transport networks, but also successful implementation of new transport schemes. Traditional studies on mode choices typically treat travellers' decision-making processes as planned behaviour. However, this approach is now challenged by the widely distributed, multi-sourced, and heterogeneous travel information made available in real time through *information and communication technologies* (ICT), especially in the presence of a variety of available mode options in dense urban areas. Some of the real-time factors that affect mode choices include availability of shared vehicles, real-time passenger information, unexpected disruptions, and weather. These real-time factors are insufficiently captured by existing mode choice models. This chapter aims to propose an introduction to MaaS, a literature review on mode choice paradigms, then it proposes a novel behavioural concept referred to as the hypermode. It will be illustrated a two-level mode choice decision architecture, which captures the influence of real-time events and travellers' adaptive behaviour. A pilot survey shows the relevance of some real-time factors, and corroborates the hypothesized adaptive mode choice behaviour in both recurrent and occasional trip scenarios.

Keywords: Mobility as a Service, mode choice, urban transport, Intelligent transportation systems

1. Introduction

MaaS can be considered as a tool to improve users' mobility and, as stated by [1, 2] "MaaS provides an alternative way to move more people and goods in a way that is faster, cleaner, and less expensive than current options". In particular, also according to [1], "MaaS is a user-centric, intelligent mobility distribution model in which all mobility service providers' offerings are aggregated by a sole mobility provider, the MaaS provider and supplied to users through a single digital platform".

Overall, MaaS application allows to optimize a trip for transit factors such as convenience, carbon emissions, and reliability. In general, the following virtuous impacts may be associated to MaaS concept:

1. reducing car ownership, the use of personally owned modes, car use, energy consumption and pollution;

2. increasing accessibility, equity, welfare;

3. pursuing the system optimum.

Currently, MaaS services have been implemented in several countries such as Germany (MOOVEL, Qixxit, BeMobility, HannoverMobil), Netherlands (Mobility Mixx, NS-Business card, Radiuz Total Mobility), Finland (WHIM, Tuup), Sweden (UBIGO), France (EMMA, Optymod), Austria (Smile, WienMobil lab), USA (SHIFT) and Singapore.

A list of key fundamentals to support such a virtuous MaaS ecosystem is reported by the Maas Alliance, a public-private partnership established in 2015 by the UE encouraging and catalyzing pilot projects. Moreover, several contribution give interesting insights, such as [3, 4].

Importantly, the MaaS Alliance, in the White Paper [5], states that: "due to ecological and capacity advantages, the traditional modes of public transport, like bus, tram and metro/underground, should remain as the backbone of MaaS in urban areas". On the other hand, the integration of traditional modes of public transport with other faster, cheap and continuous services can provide new lifeblood and a new look to public transport and can be the key solution to aid sustainable modes of transport [6].

Unfortunately, if the vision of what MaaS should be is clear, the real challenge is what MaaS could become if its development completely relies on private operators. [7] reported that the main risk of a purely commercial approach to MaaS is to disincentivize sustainable trips, stating that "The success in some markets of new services, including apps for private-hire vehicles and ride-sharing, clearly has the potential to disrupt existing urban mobility services and could also encourage a shift towards car use away from more sustainable modes." Therefore, in the same study it has been concluded that "City and regional authorities need to be involved in the development of policy around MaaS at EU and national level, through new models of governance and with public sector leadership, to avoid environmental, economic and social dysfunctions."

According to the four workshops organized for the MAASiFiE project to define a European 2025 MaaS roadmap, the policy & regulation between public and private participants is seen as the most significant driver for MaaS development [8]. Along the same line, the Maas Alliance recognizes that a limited regulation can compromise the way MaaS applications impact on urban environment and compromise the public interest (e. g. decrease congestion, pollution, etc.).

However, if the problem is clear the solution is not so easy to find. One step forward was taken by the Finnish government adopting in April 2017 the Act on Transport Services (also known as Transport Code), the first known regulatory effort on this matter. However, the Code main aim is to boost the establishment of requirements for MaaS services (integration, interoperability, etc.) and the availability of data [9].

In addition, starting from the vision of how MaaS should work, a great effort need to be used in proposing models and methodologies enabling such working.

MaaS platform should be supported by a simulation environment which, starting from historical and observed data, should be able to reproduce the actual state of the multi-modal transportation system and forecast the future states. This is strictly needed to "offer" updated, reliable and personalized MaaS solutions to the users. Current platforms rely on simplified hypotheses on users' travel behavior models, on transport system simulation models and on the short-midterm traffic forecasting models.

Summarizing, the following drawbacks seem to characterize MaaS implementation:

- MaaS development is mainly driven by private operators; hence, they are mainly developed as business solutions in an open market, where different competitors offer their services, without a clear regulatory framework and without a clear vertical/horizontal integration.
- Uncertainty about the way decision makers and governmental agencies may push towards specific, system-oriented transportation planning strategy due to the lack of a clear regulatory framework.
- MaaS platforms do not rely on consistent and reliable modelling framework able to forecast future system state and consistently modify their offer.
- Existing analyses have not clearly demonstrated the environmental and social sustainability of a MaaS service.

It is, therefore, important to preliminarily define the conceptual framework in which any MasS service should be figure out. To this aim, five pillars can be identified.

i. MaaS service should be an open market but regulated by decision makers/governmental agencies and characterized by a specific and organized regulatory framework;

ii. the traditional modes of public transport, like bus, tram and metro/underground, should remain as the backbone of MaaS in urban areas;

iii. MaaS service should guarantee sustainability and equity and lead toward a car-free transportation system;

iv. MaaS service should offer a smart integration of single-step transport modes offered by different providers;

v. MaaS service should be intrinsically dynamic adapting its characteristics to the day-to-day and the en-route travel needs.

In this context, MaaS requires the ability to model mobility and travel choices. The issue is not trivial especially for travel choices [10]. First, it is necessary to model individual choices, secondly, both predictive and adaptive choices must be considered and third, the intrinsic dynamic of the choice behavior must be explicitly considered.

Indeed, the understanding and quantification of travellers' mode choices is crucial for the prediction and management of multimodal traffic networks, and have become an important field of inquiry in cross-disciplinary research spanning transport engineering, computing, mathematics, psychology, and social and behavioural sciences.

The underlying assumption of most existing studies on travel mode choice is that a traveller chooses a specific mode before commencing his/her trip, which is categorized as planned behaviour. However, some studies have identified and demonstrated the influence of real-time events on mode choices,[1] particularly for travellers using public transport; some examples include real-time passenger information, weather, and transport disruptions [11–13]. These real-time events may lead travellers to assess various modes in an adaptive way to the extent that the aforementioned planned behaviour plays a less significant role in the final outcome of the mode choices.

This chapter aims to give a literature overview of the existing approaches, the aims to propose an adaptive mode choice behaviour paradigm which takes into account real-time events, and provides an empirical validation of this mode choice paradigm. The real-time events include, but not limited to, availability of shared bikes at the docking station, real-time information on bus arrival time, scheduled or unexpected local disruptions, and weather conditions. This research is an important undertaking as it not only identifies a set of new factors that influence mode choices, but also presents a novel framework to study mode choice behaviour. This behavioural paradigm may pose interesting challenges from a modelling perspective and may require an integrated modelling approach for both mode choice and traffic assignment to fully capture the adaptive behaviour. The latter statement stems from the observation that many real-time factors identified above have a dynamic and stochastic nature that is related to the evolution of the system (e.g. the dynamic network loading).

The main contribution made by this paper includes:

- A novel adaptive mode choice behavioural paradigm able to incorporate real-time events (both pre-trip and en-route), which advances state-of-the-art modelling approaches that mostly rely on static attributes and simulate mode choices as planned behaviour.

- A pilot survey that shows the viability and validity of the adaptive mode choice behaviour for real-world scenarios, where a number of mode options and real-time events are defined and combined to analyse user responses under different circumstances.

The rest of this chapter is organized as follows. Section 2 provides an extensive review of state-of-the-art mode choice approaches. In Section 3 the bi-level mode choice behaviour paradigm that explicitly accounts for real-time events is proposed. A real-world scenario pertaining to the hypothesized adaptive behaviour is introduced in Section 4, which also presents the pilot survey study, which assesses the behavioural validity of this new concept at a qualitative level, and discusses the survey results. Section 5 introduces some remarks on the main issues of MAAS and the research perspectives regarding the proposed interpretative hypermode paradigm.

2. Literature review

In general, travel mode choices may be updated between different periods of time (period-to-period) or within the same trip (within-day). In the period-to-period choice process all the available transport modes are considered. Users have the option

---

[1] Throughout this paper, real-time information is treated as a special case of real-time events.

to choose among the available modes and their decision-making processes converge towards a stable choice that, once reached, can be considered as habitual.

Almost the totality of the existing scientific contributions assume that travellers choose their mode of transport through a one-step decision as a planned behavior, only few exceptions explore the alternatives. In particular, travellers' mode choices are usually reported to be habitual in several travel behaviour studies [14–16]. In general, habits depend on the perception and preference towards a travel mode and it can hardly be modified. As a matter of fact, it is a common approach to investigate and model the habitual behaviour (holding behaviour), and neglect the dynamic element of the choice process. Within this framework, the mode choice analysis may depend on the different interpretative paradigms that can be assumed for modelling travel demand:

a. Trip-based. It implicitly assumes that the choices relating to each origin–destination trip are made independently of the choices for other trips within the same and other journeys.

b. Trip chaining. It assumes that the choices concerning the entire journey influence each other. In this case, the choice of an intermediate destination, if any, takes into account the preceding or following destinations in the trip chain; the choice of transport modes takes into account the whole sequence of trips in the chain, and so on.

c. Activity-based. It analyses transport demand as the outcome of the need to participate in different activities in different places and at different times. It therefore takes into account the relationships among different journeys made by the same person during the day and, in the most general case, between journeys made by the various members of the same household.

The trip-based paradigm is the most widely adopted, and relies on a consolidated theoretical literature [17] and operational literature, which has predominantly investigated mono-modal transport systems competing with each other (e.g. [18–34]).

Minor attention, yet increasing in the last years, has been paid to individuals' preferences in multi-modal networks where different transport modes are integrated and a possible choice alternative is a combination of them (e.g. [22, 35–40]). Nevertheless, it should be noted that most of them consider public or private transport modes separately or, consider integrated transport modes, for instance when combined with park-and-ride. More Recently, [41] attempt to model the full range of choice options in multimodal network settings using a stated preferences approach, and approach the problem as a route choice problem. But they only investigate pre-trip choices.

Trip-chain and activity-based paradigms model pre-trip behaviour in a more realistic behavioural context, hence may allow a better interpretation and simulation of the travellers' mode choices. However, they are usually rather complex for calibration and implementation. Some examples include:

i. mode and departure time [41, 42];

ii. trip chain [42–47]

iii. activity-based [32, 46–50];

Among the pre-trip choice paradigms, pre-trip switching approaches have also been developed to understand and simulate potential modal shift (e.g. [51–56]).

Finally, different attempts have also been made to model an habitual behaviour (holding decision), but taking into account temporal correlation for the same user, thereby showing how tastes can vary for the same traveller using short-term cross-sectional data [57–59].

The pre-trip and habitual choices have been extensively investigated in the literature; however, some scholars have also dealt with the explicit simulation of mode choice dynamics with regard to both short- or long-term scenarios. In particular, [60–63] study short-term mode choice dynamics using discrete choice method and panel data. With regard to long-term mode choice dynamics, [64] investigate commuting behavior within the traditional maximum utility framework, whereas alternative approaches have tried to take into account more complex behavioural determinants and processes such as habits and learning. In particular, [65] derived decision rules based on neural networks to predict activity scheduling and mode choice; [66] developed a computational process model to mimic travel decision-making process; [67] developed an agent-based process to simulate travel behavior in terms of information acquisition, learning, adaptation and decision heuristics. Recently, the Markov chain approach has been fruitfully adopted to model and interpret the decision-making process [32, 68–72].

With regard to the within-day travel mode choice behavior, it can be assumed that a typical traveller chooses a transport mode (or a combination of transport modes) and may change his/her initial choice by switching to other modes before leaving the origin and/or during the trip. Obviously, such behaviour is reasonable only in a multi-modal or inter-modal context. On the one hand not many contributions can be founded in the literature; on the other hand the landscape of available mode options is evolving particularly at the urban level. Multi-modal networks are rapidly growing, and a new generation of mobile, personalised information systems and intelligent transport systems are ready to support this flexible and adaptive behaviour by providing assistance in the planning and implementation of multimodal trips [73].

As a matter of fact, an increasing number of users may reconsider their initial travel choices. However, not many contributions can be found in literature. From a psychological viewpoint, the study undertaken by [74] considers a two-level approach to simulate the mode choice. At the first level (more related to the person) the authors apply the comprehensive action determination model, which assumes that intentional processes, habitual processes, and normative processes lead to a certain level of propensity to use the private car. The second level of choice (more related to the trip) is characterized by situational influences, where trip purpose, disruptions on public transport, and weather are identified as predictors. The authors conclude that the multi-level approach is a promising alternative to conventional models. These insights from the field of psychology are valuable for the correct interpretation of the decision-making processes of travellers, and will be considered in the hypermode approach proposed in this paper.

Different contributions have analysed mode choice as a sort of path choice in a broader context of a multimodal network [49, 75–77], by considering interconnected networks, one for each different transport mode. Such an issue has been addressed in a multi-modal context through the well-established supernetworks [78, 79]. However, such an approach has been mainly adopted/used for modelling elastic demand assignment problems; it is not very flexible to address possible adaptive behaviour.

Contribution by [80] take the mode availability into consideration, but mainly from an assignment perspective as both model the user decision to change mode at each node. In [81] each option (mode and route) is associated with a probability of immediate availability, which is one for private modes and less than one for public transport, the latter being a specific value affected by service frequency. The author therefore revised the transit assignment problem by taking into account mode availability at each node, which represents a decision point for the users. The proposed assignment model entails sequential choices at each intermediate node in the multimodal network and seeks an equilibrium. [80] proposes the strategy of adaptive multimodal least expected time in order to determine the hyperpath associated with the least cost in a multimodal network. In addition to the modes of walking and driving, each public transport line is considered a separate mode. The authors consider a delay associated with the transfer between modes, and model the users' capability to reassess the costs at each node and determinate if switching mode may be a better option. However, the assessment of switching from one mode to another is merely based on time as this is a reasonable assumption for assignment algorithms, but it is not sufficient to capture the decision making process at the mode choice level. Moreover, users are quite reluctant to have too many transfers and reassess all mode options at every en-route node unless a disruption occurs.

In conclusion, mode choice behaviour may rely on an extensive scientific literature, but it predominantly deals with habitual behaviour including pre-trip behaviour, pre-trip switching behaviour or travellers' behaviour at specific nodes of the transportation network. Most of the existing efforts have been focused on multi-modal networks in which different transit modes are connected or in which individual transport modes (car, motorbike, cycling) and collective transit modes may interact with each other (Park and Ride). However, the choice contexts are always pre-trip and not much can be found with regard to multimodal contexts in which the transport mode can be changed during the trip (transit alternatives and shared modes). For example, the introduction of shared modes (car/motorbike/bike-sharing) and their integration with the various existing transit systems lead to a significant flexibility that cannot be neglected. Instead, they should be carefully analysed and interpreted with behavioural paradigms that are different from the traditional ones. Furthermore, the literature suggests that weather conditions have the potential to influence mode choices [12, 82], and that there is a lack of comprehensive evaluation of costs, time and service quality in multimodal travel choice [41].

In conclusion, this paper focuses on the decision making process that leads users to take a specific mode in the presence of different mode options and real-time information/events. The proposed novel mode choice paradigm satisfies the following requirements:

- it captures a more realistic mode choice behaviour, which is influenced by real-time events, building on the multi-level approach proposed in the psychology field by [74];
- it is able to subsume planned behaviour as a special case, which addresses travellers whose mode choices are not adaptive; this is particularly true for travellers who use their own vehicles (private cars and bikes);
- it is validated with a pilot study through a preliminary qualitative survey to demonstrate the validity of the approach.

3. The hypermode concept

Currently, mode choice is considered a planned behavior and is embedded within traffic assignment procedures only in a static context [83], which obviously does not capture the influence of any real-time events. With regard to dynamic modelling, mode choice is usually considered a fully pre-trip behavior. This paper investigates an adaptive mode choice behaviour and presents the results of an empirical study undertaken to validate the approach. It focuses on the potential effects of real-time events on both pre-trip and en-route mode choices.

For reason that will become clear below, this adaptive mode choice will be hereafter called "*hypermode*", in analogy to the hyperpath concept proposed for the route choice in transit assignment [84]. The hyperpath approach suggests that travellers first identify a set of attractive lines that connect their origin–destination (O-D) pair; then they choose a specific service according to certain strategies. Such strategies can be based on the minimization of travel time, waiting time, walking distance, or the number of transfers A more complex strategy can also consider the influence of real-time information on path choices [85, 86]. In an analogous way, the hypermode concept stipulates that travellers identify a set of feasible modes for their target trip and may make their final decisions later based on real-time events. These adaptive mode choices have been recently facilitated by the development of Information and Communication Technologies (ICT) such as smartphones, as well as Intelligent Transport Systems (ITS) such as vehicle tracking and prediction. For example, travellers can now make informed mode choices based on estimated time of arrival of buses/trains/trams, or the availability of shared bikes at any given docking station. Such adaptive travel behaviour is suitable for dense urban areas, where plenty of mode options and access points are available to travellers, and walking is always an option especially for short trips. Given that 50% of the trips in urban areas in Europe are shorter than 5 km [87], the hypermode concept enjoys wide empirical support. This extra modelling dimension could lead to a significant yet challenging advancement in the modelling of multimodal transport networks.

This section illustrates this notion by proposing a conceptual analytical framework along with a few examples.

3.1 Decision-making architecture

In this section, we formally introduce the *hypermode* concept, which is analogous to the hyperpath concept proposed for the route choice in public transit assignment [84]. The hyperpath approach suggests that a traveller first identifies a set of attractive lines that connect the origin–destination (O-D) pairs. Then, he/she chooses a specific service according to a certain strategy, which can be based on the minimization of travel/waiting time, amount of walking, or number of transfers. A more sophisticated strategy can also take into account the influence of real-time information on path choices [86]. In an analogous way, the hypermode concept stipulates that travellers identify a set of feasible modes for their target trip and may later make their final decisions based on real-time events. These adaptive mode choices have been recently facilitated by the development of Intelligent Transport Systems (ITS), and Information and Communication Technologies (ICT).

The underpinning decision making process involved in the hypermode concept is articulated in two levels.

1. The user identifies a set of feasible travel modes for the trip, which are accessible at the same physical location or nearby. On this level, the decision making is strategic (i.e. not real-time), and is affected by static characteristics such as user preferences, socio-economic characteristics, average/historical travel times, and financial costs of using different modes.

2. Right before a trip is made, the user evaluates real-time events in order to select a specific mode of transport from the aforementioned feasible set. The real-time event includes but is not limited to: availability of vehicles (relevant to

shared modes), weather conditions (relevant to walking and biking), vehicle arrival time information (relevant to scheduled or unscheduled public transport), and disruption or crowdedness.

This adaptive behaviour can occur at the following different stages of the trip:

- The user has not yet left the origin and has a set of modes in mind that could bring him/her to the destination with acceptable time and cost. Just before leaving the origin the user reassesses these modes based on real-time events such as weather, real-time bus information and so forth, which may influence the user's the final choice of mode within the feasible set.

- The user has just left the origin with a specific mode in mind (e.g. tube). He/she then approaches a tube station and notices a disruption or heavy crowding, hence immediately considers another mode from the feasible set.

- The user has chosen an *access point*, which is a specific location where he/she can access several modes that can all serve the trip. The user approaches the access point and then chooses a specific mode based on a combination of his/her preferences (e.g. first coming/least walking/least transfers) and real-time events.

The extent to which the real-time events affect the mode choices varies among individuals. For example, some users may take their preferred modes in any circumstance. This is particular the case for travellers who use their own vehicles, such as private cars or bikes (cyclists who use their own bikes usually stick to the same mode in case of very adverse weather conditions). Such behaviour is referred to as *planned* in this paper since it is not adaptive, and only involves the first level of the decision making process. Such planned behaviour can be subsumed by the proposed two-level decision making paradigm, as it is a special case with decision parameters on the second level being rigid and non-responsive to real-time events.

3.2 Factors affecting adaptive mode choices

Table 1 illustrates the proposed approach and a non-exhaustive list of factors affecting choice probabilities on the two choice levels.

The realization of a specific mode choice is therefore the consequence of the mode first belonging to the feasible set (choice level 1), and then actually chosen within such set with given real-time events (choice level 2).

Figure 1 illustrates, in further detail, individual components of the decision making processes with inputs and outputs of the two levels of choices.

Any of the traditional mode choice models can be applied to calculate the probability at the first level. Once the probabilities of all possible modes are calculated, the set of feasible modes can be formed, which is a quite standard procedure and thus omitted here. In the second level of the decision making process, the feasible modes are subject to re-interpretation and their probabilities are reassessed based on real-time events. For example, if walking is the preferred mode with the highest probability at the first level, and the weather is rainy in real time, the probability associated with walking decreases.

The whole procedure may be easily formalized in a compact formulation coherent with existing assignment models, thus may be implemented for simulation any transportation system (see technical report [88]).

Level of choice	Factors affecting choice probability
(1): Feasible set	Probability of each mode to belong to the *feasible mode set* depends on: • Socio-economic characteristics (age, gender, income, etc.) • Health and/or environmental concern • Financial cost • Average travel time • Number of transfers
(2): Final mode choice	Choice probability of a specific *mode* depends on: • Real-time arrival time (bus, train) • Vehicle availability (bike-sharing, car club) • Weather (walking, cycling) • Disruptions and crowdedness (bus/train/tube stations)

Table 1.
The two choice levels and influencing factors in the hypermode approach.

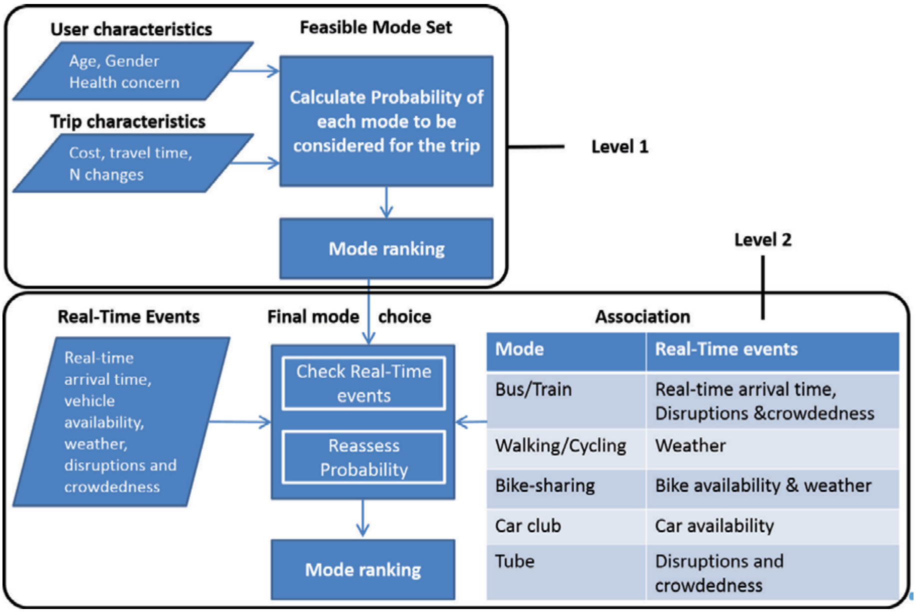


Figure 1.
Flow chart representation of the hypermode concept.

4. Real-world case study

The hypermode concept is illustrated here using a real-world example. The area of interest is part of South Kensington in London.

As shown in **Figure 2**, a traveller starts his trip in O (origin) and wishes to reach the destination D. Before leaving the origin, the user has a set of feasible modes he would consider, namely bus, tube, bike-sharing, and walking, which are all accessible in the vicinity of the origin. These feasible modes are ranked by the user according to his/her own preferences, which are static in nature. For example, the traveller may consider cycling as unsafe, thus bike-sharing may receive a low rank or even is excluded from the feasible set. Moreover, the traveller usually has a

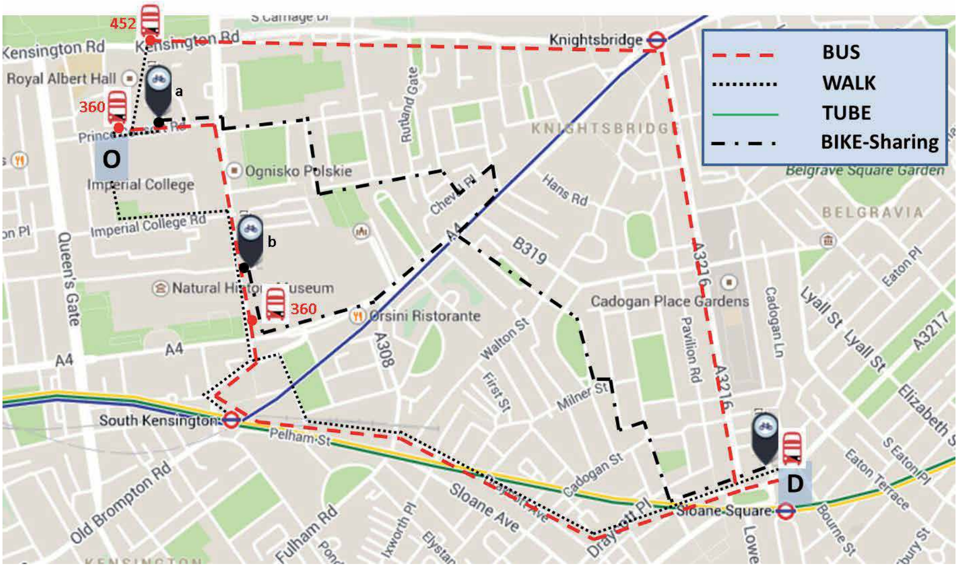


Figure 2.
The study area in South Kensington, with available modes and routes shown.

preferred mode within the feasible set, which is likely to be the one he/she pursues at the first attempt. If this preferred mode is not viable given real-time conditions (e.g. no shared bike is available, or the weather is unsuitable for walking), then the probability of selecting that mode decreases and the user will consider other modes in the feasible set.

Based on **Figure 2**, we describe the following specific scenarios, which are examples of potential adaptive behaviour.

- The user includes walking and bus in his feasible mode set. He prefers to take the no.360 bus at the closest bus stop towards the destination. When he reaches the bus stop, he sees on the digital display board that the next bus will arrive in 10 minutes. Rather than waiting at the bus stop, he switches to walking knowing that the total travel time would be similar (notice that the waking route in this case differs from the one shown in **Figure 2**).

- The user, who eliminates the possibility of walking due to physical conditions, may have bus and tube in his feasible set with bus being the preferred option. Before leaving the origin, he checks his cell phone and finds out the estimated waiting time for the bus is 10 minutes. He then prefers to take the tube at the South Kensington Station instead of waiting for the bus.

- The user has walking, tube and bike-sharing in his feasible set with cycling being the preferred option. He approaches the nearest docking station and cannot find any available bike. In this case he decides to walk or take the tube, depending on which one ranks higher, instead of looking for other docking stations nearby.

- The user has walking and bus in his feasible set, with walking being his preferred option. He is about to leave the origin when it starts raining. He then chooses to take a bus instead.

All of these illustrative examples have one thing in common: The pre-defined feasible modes are re-interpreted and re-ranked with the influence of real-time information, which is dynamic and stochastic in nature. This highlights the key difference between the traditional mode choice model and the adaptive behaviour that we try to demonstrate.

Note that it is possible that the repetitive occurring of a negative real-time event on a day-to-day basis may lead to the exclusion of a mode from the feasible set. For example, if a user constantly finds the bike-sharing station empty, he/she may exclude bike-sharing as one of the feasible modes in his/her planned behaviour. This, however, does not contradict the mode choice behaviour that we propose here. In fact, it still falls within the scope of the proposed two-stage decision-making process, i.e. in the forming of feasible mode choice set (see Level 1 of **Figure 1**). In most cases, the feasible mode set contains more than one element, and the realization of a particular mode choice (or sequence of mode choices) must thus rely on real-time events.

To further support the relevance and likelihood of such adaptive behaviours, we conduct a qualitative survey to validate the behavioural soundness of this subject, as described in Section 4.1.

4.1 Survey study

A pilot survey has been undertaken to explore the validity of the underpinning idea of the proposed hypermode concept. 50 respondents have been interviewed at Imperial College London. The sample includes academic, technicians and administrative staff as well as students, to ensure that behaviour in different user categories is captured. The respondents have been interviewed face-to-face to ensure an in-depth and comprehensive grasp of their decision-making processes.

4.2 Survey design

The respondents were presented with two different scenarios:

SCENARIO 1. The regular commuting trip home from the College at the end of the day, which is a Revealed Preference scenario. The origin is the same for all respondents but the destinations vary, with some at walking distance and others outside of London.

SCENARIO 2. A hypothetical trip from the College to Sloane Square (a shopping destination 2.1 km away from the origin) at the end of the working day. This is a Stated Preference scenario.

In the first scenario the respondent is asked what modes are available for his/her trip. An open question is then asked to describe the decision making process that shortlists the possible mode options or leads to a specific mode choice. Afterwards they are asked if any of the following real-time events may affect their final mode choice:

1. Real-time bus arrival time

2. Bike availability at docking stations for bike-sharing service

3. Disruptions on the tube

4. Weather

5. Other, specify.

If the respondent's explanation of the decision making process at the open question is in line with the adaptive behaviour, as confirmed by answering "yes" to any of the above real-time events (1 to 4), then this user behaviour is related to hypermode.

In the Stated Preference scenario (Scenario 2) the modes available to the user are the same as those shown in **Figure 2** (with possibly different routes and access points), and are associated with given average costs and travel times. The user is asked what would his preferred mode option be in the described scenario. Depending on the preferred mode, a range of real-time events are presented to the respondent, which may lead him/her to reassess the original choice. For example, **Table 2** shows the situation presented to the respondent who selects bus as the preferred mode.

Two different types of trips are considered:

- Leisure (e.g. shopping, visiting friends)
- Important appointment, (on-time arrival is needed)

Since trip purpose is likely to be an influencing factor of mode choice.

Real-time events relevant to other preferred modes are also included in the survey; a few examples are provided below.

- Availability of bike at the docking station if the user chooses bike sharing;
- Wet weather if the user chooses walking;
- Disruptions on the tube once the user reaches the tube station.

The key point for Scenario 2 is to understand if the user would either consider alternative transport modes in a specific situation or stick to the initial mode preference regardless of any real-time events. In the first case the user is associated with the hypermode behaviour. In Scenario 2 both the origin and destination are located in central London, which is not necessary true in Scenario 1. This could have

Preferred mode		**Bus**
Real-time event	You arrive at the bus stop and the information system says that your bus will arrive in 12 minutes.	
Purpose of trip	Leisure (e.g. shopping)	Appointment (on-time arrival is crucial)
What do you do?		
Wait for the bus even if you may arrive later than expected		
Use the 12 minutes for other errands and then go back to the bus stop		
Consider alternative buses		
Consider alternative modes		

Table 2.
Survey scenarios for the bus.

potentially influenced the results as more mode options are available to reach the destinations in Scenario 2, while in the first scenario the users with destination far away may have quite limited mode choices. To avoid potential bias, in Scenario 1 the respondents with destination outside of London are asked to consider the trip from the College to the station in central London from which they take a train; this allows plenty of mode options to be available to all users.

4.3 Survey results

Some characteristics of the sample respondents are reported in **Table 3**.

Table 4 shows the percentages of respondents associated with the hypermode behaviour. We also use the sample size (50) to calculate the 95% binomial proportion confidence interval. Here, the category "Either scenario" accounts for those who show the hypermode behaviour in either Scenarios 1 or 2.

The results of this exploratory survey show that the vast majority of the respondents follow an adaptive behaviour, which is in line with the hypermode concept. In the first scenario, 34% of the respondents consider initially a set of feasible modes for their trips, and their final mode choices are determined/affected by real-time events.

In the second scenario, 86% of the respondents indicate that they would consider alternative modes once adverse real-time events occur. This result refers to the overall responses for the two travel purposes (leisure/appointment) (i.e. respondents who consider alternative modes for at least one of the two travel purposes are associated with hypermode). The analysis of the responses for each travel purpose indicates that:

- [Leisure] 60% of the respondents re-assess the available modes in the presence of adverse real-time events.

- [Appointment] 82% of the respondents re-assess the available modes in the presence of adverse real-time events.

This difference is easily understandable as the urge to reach the destination on time offers another motivation to reconsider other modes and justifies the associated effort.

A more detailed analysis of Scenario 2 identifies the pattern of mode switches under the two different trip purposes. The results are reported in **Table 5**.

Age of respondents		**Gender of respondents**	
N respondent 18–25	24%	N female respondent	38%
N respondent 26–44	48%	N male respondent	62%
N respondent 45–64	28%		

Table 3.
Age and gender of respondents.

	Scenario 1	**Scenario 2**	**Either scenario**
Average percentage	34%	86%	92%
Confidence interval	**[21%, 49%]**	**[73%, 94%]**	**[81%, 98%]**

Table 4.
Survey results with 95% confidence levels for the hypermode behaviour. Sample size: 50.

Mode initially considered	% of switches (Leisure)	% of switches (Appointment)
Walking	50%	41%
Tube	37%	39%
Bus	13%	17%
Bike-sharing	0%	2%
Cycling	0%	0
Taxi	0%	0

Table 5.
Mode switches in Scenario 2.

The adaptive behaviour is more evident in Scenario 2. This may be explained by the fact that the respondents were referring to their regular commuting trip in the first scenario, and were less likely to abandon their preferred mode due to extensive learning of the preferred and alternative modes based on past experience. On the other hand, in the more hypothetical scenario (Scenario 2), the travel environment is new to the commuters, who might be more inclined to consider different modes due to the lack of experience.

The inertia in decision-making may also play a role in the sense that users may be inclined to stick to one specific mode of transport even though it may not appear to be the most rational choice at the moment. This choice behaviour is known as bounded rationality. In the second scenario, despite their familiarity with the area, the users were more prone to consider different modes, as their experience on specific trips is relatively limited. The adaptive behaviour is more evident when on-time arrival at the destination is important.

Table 5 partially illustrates the relevance of users' adaptive behaviour to planning. In particular, it shows the percentage of travellers who abandon their initial (static) mode choice in reaction to real-time events. For example, when there is a interruption/delay of tube service, 39% of travellers will switch to other modes, possibly at nearby access points. Such information is crucial for planning service interruption at tube stations (such as scheduled maintenance or train operation): the planner need to take into account the increase in demand for other modes in the vicinity of the tube station to avoid heavy congestion and/or shortage of supplies.

5. Conclusions

5.1 Remarks on MaaS

Many researchers and stakeholders of the transport sector see Mobility as a Service (MaaS) as the mobility of the future. However, a lot of uncertainty lye under this travel solution. The same first statement is actually uncertain, considering that it depends on MaaS diffusion, which in turn depends on the adopted business model, on its financial convenience and on the membership rate, which in turn depend on what kind of services are offered, their level of service and their price. On the other hand, first MaaS applications have not helped to clarify the financial convenience of a MaaS.

Furthermore, MaaS is also generally associated with many virtuos impacts which can be sinthesized by saying that it goes in the direction of a sustainable mobility, by aiding and supporting intermodality. However, also this statement is not clearly confirmed by the literature. In fact, put different services in the same market place does not suffice to guarantee intermodality: who chooses the service/services to offer to the users? Instead, the literature agrees that if MaaS is developed according to a purely commercial approach, there is a serious risk to dis-incentivize sustainable trips, encouraging instead a shift towards car use. Hence, the literature also agrees on the need of regional authorities to lead MaaS development through new models of governance. However, the way this result can be achieved is not clear.

Consistently with these premises, one of the main necessary research perspective is to investigate and identify possible regulatory frameworks, MaaS schemes and approaches for bidding/tendering the services enabling the public entity to rule the MaaS.

In addition, the way MaaS should work is clear in the literature, but models and methodologies enabling such working are very complex and largely new with respect to the approaches currently used in transportation system modelling.

From the demand side, new approaches are needed to profile users coupling with MaaS requirements and involving also psychological and social science. Moreover, new approaches are needed for the modelling of individual choices, considering the intrinsic dynamicity of the MaaS system and including the effect of service reliability and of the provided information.

From the supply side, new modelling framework are needed to integrate continuous and discontinuous services, to deal with diffused/distributed inter-modality and with dynamically changing conditions and unpredicted temporary disruptions of the service.

5.2 Remarks on the Hypermode paradigm

The widespread of real-time travel information combined with the presence of a variety of travel modes available in dense urban areas could lead travellers to reconsider their planned mode choices based on real-time events, such as real-time passenger information, transport disruptions, overcrowdings, and weather. However, most existing mode choice studies analyse the decision process as planned behaviour, and hence do not capture the influence of these real-time events on mode choices.

This paper aims to address this limitation by presenting an innovative approach to interpret mode choice, which captures an adaptive behaviour of travellers. The underpinning assumption is that the traveller first identifies a set of feasible modes that connect his/her origin to the destination; then he/she evaluates the real-time situation in order to select a specific mode of transport from the feasible set. This novel behaviour paradigm is referred to as hypermode in this paper, in analogous to the hyperpath concept used extensively in transit assignment. A two-level decision-making process is illustrated, which rests on planned and adaptive model choice behaviour, respectively.

A survey has been undertaken to test the proposed approach; it demonstrates the validity of the underlying assumption of hypermode, and serves as a proof of concept. As the next step of this research, we will explore different modelling approaches (e.g. Nested/Mixed Logit and discrete choice model with endogenous attribute threshold/cut-off) along with calibration and validation methods based on a wide variety of data.

For future research, the hypermode concept could be explored alongside dynamic assignment of the multimodal network, which provides feedback to mode choice models in the form of real-time events. For example, the dynamic re-distribution of shared bikes, as a result of multimodal traffic assignment, could affect the availability of bikes at docking stations and hence travellers' mode choices.

Such a feedback mechanism between traffic assignment and model choice presents a research direction not previously investigated, and calls for a more integrated modelling approach driven by the presence of real-time travel information.

Author details

Stefano de Luca[1]* and Margherita Mascia[2]

1 Department of Civil Engineering, University of Salerno, Salerno, Italy

2 ITS Counsult, Bari, Italy

*Address all correspondence to: sdeluca@unisa.it

References

[1] Kamargianni, M., & Matyas, M. (2017). The business ecosystem of Mobility as a Service. Paper presented at the 96th Transportation Research Board Annual Meeting, Washington DC, USA.

[2] Goodall et al., (2017). The rise of Mobility as a Service: Reshaping how urbanites get around Deloitte University Press (2017)

[3] D.A. Hensher, C. Ho, C. Mulley, J. Nelson, G. Smith, Y. Wong, (2020). Understanding Mobility as a Service (MaaS) - Past, Present and Future, Elsevier Science, UK.

[4] M. Kamargianni, M. Matyas, W. Li, J. Muscat, L. Yfantis, (2018). The MaaS dictionary, Retrieved from University College London's website

[5] MaaS Alliance (2017). White Paper: Guidelines & Recommendations to create the foundations for a thriving MaaS Ecosystem, MaaS Alliance, Brussels (2017)

[6] Kamargianni, M., Li, W., Matyas, M., Schäfer, A., 2016. A Critical Review of New Mobility Services for Urban Transport. Transp. Res. Procedia 14, 3294-3303.

[7] Hoadley (2017). Mobility as a Service: Implications for urban and regional transport, S. Hoadley (Ed.), Discussion paper, POLIS, Brussels, Belgium (2017)

[8] L. Nykänen, J. Eckhardt, A. Aapaoja, J. Sochor, and M. Karlsson. "The European Roadmap 2025 for MaaS". In ICoMaaS 2017 Proceedings, pages 266-279. Tampere University of Technology, Transport Research Center Verne, Finland, 2017

[9] I.C.M. Karlsson, D. Mukhtar, Landgren, G. Smith, T. Koglin, A. Kronsell, E. Lund, *et al. (2020),* Development and implementation of Mobility-as-a-Service – A qualitative study of barriers and enabling factors, Transport. Res. Part A. Pol. Pract., 131 (2020), pp. 283-295.

[10] G. Lyons, P. Hammond, K. Mackay, (2020), The importance of user perspective in the evolution of MaaS Trans. Res. Part A Policy Pract., 131 pp. 20-34

[11] Zhang, M. 2005. "Exploring the relationship between urban form and nonwork travel through time use analysis." Landscape and Urban Planning 73 (1):244-61.

[12] Saneinejad, Sheyda, Matthew J. Roorda, and Christopher Kennedy. 2012. "Modelling the impact of weather conditions on active transportation travel behaviour." Transportation Research Part D: Transport and Environment 17 (2):129-37. doi: http://dx.doi.org/10.1016/j.trd.2011.09.005.

[13] Klöckner, Christian A., and Thomas Friedrichsmeier. 2011. "A multi-level approach to travel mode choice – How person characteristics and situation specific aspects determine car use in a student sample." Transportation Research Part F: Traffic Psychology and Behaviour 14 (4):261-77. Doi: http://dx.doi.org/10.1016/j.trf.2011.01.006.

[14] Pendyala, R., A., Parashar, and G., Muthyalagari. 2000. "Measuring day-to day variability in travel characteristics using GPS data". In: Paper presented at the 79th Annual Meeting of the Transportation Research Board, Washington.

[15] Gärling, T., and K. Axhausen. 2003. "Introduction: habitual travel choice". Transportation 30:1-11

[16] Gärling, T., and J., Rise. 2002. "Understanding attitude, intention, and

behavior: a common interest to economics and psychology". C.L. Spash, Anders Biel (Eds.), Social Psychology and Economics in Environmental Research, Cambridge University Press.

[17] Cascetta, E. 2009. "Transportation Systems Analysis: Models and Applications". Springer Verlag.

[18] Koppelman, F.S. 1983. "Predicting transit ridership in response to transit service changes". Journal of Transportation Engineering 109(4):548-564.

[19] Bhat, C.R. 1995. "A Heteroscedastic Extreme Value Model of Intercity Mode Choice". Transportation Research Part B 29(6): 471- 483

[20] Bhat, C.R. 2000. "Incorporating Observed and Unobserved Heterogeneity in Urban Work Mode Choice Modeling". Transportation Science 34(2): 228-238

[21] Bhat, C.R, and R., Sardesai. 2006. "The Impact of Stop-Making and Travel Time Reliability on Commute Mode Choice" Transportation Research Part B 40 (9): 709 -730.

[22] Wardman, Mark. 2001. "A review of British evidence on time and service quality valuations". Transportation Research Part E 37:07-128

[23] Cantarella, Giulio Erberto, and S. de Luca. 2005. "Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility models." Transportation Research Part C 13 (2):121-55. Doi: http://dx.doi.org/10.1016/j.trc.2005.04.002.

[24] Hensher, David A., and John M. Rose. 2007. "Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study." Transportation Research Part A: Policy and Practice 41 (5):428-43. Doi: http://dx.doi.org/10.1016/j.tra.2006.09.006.

[25] Lo, Hong K., and William H.K. Lam. 2008. "Recent advances in travel choice behavior modeling". Transportmetrica 4(2): 79-81.

[26] Greene, W.H., D.A., Hensher, and J.M., Rose. 2006. "Accounting for heterogeneity in the variance of unobserved effects in mixed logit models". Transportation Research Part B, 40:75-92

[27] Hoogendoorn-Lanser, S., R. Van Nes, S.P. Hoogendoorn, and P. Bovy. 2006. "Home-activity approach to multimodal travel choice modeling". Transportation Research Record 1985: 180-187.

[28] Horeni, O., T.A., Arentze, B., Dellaert, H.J.P., Timmermans. 2008. "Design of a computer-assisted instrument for measuring mental representations underlying joint activity-travel choices". In: Proceedings ISCTSC Conference, Annecy, France.

[29] Walker, J.L. 2011. "Beyond Rationality in Travel Demand Models". ACCESS: the Magazine of UCTC 39: 8-15.

[30] Zhou, Jiangping. 2012. "Sustainable commute in a car-dominant city: Factors affecting alternative mode choices among university students." Transportation Research Part A: Policy and Practice 46 (7):1013-29. doi: http://dx.doi.org/10.1016/j.tra.2012.04.001.

[31] Mercado, Ruben G., A. Paez, S. Farber, M.J. Roorda, and C. Morency. 2012. "Explaining transport mode use of low-income persons for journey to work in urban areas: a case study of Ontario and Quebec". Transportmetrica 8(3): 157-179

[32] Akshay, Vij, A. Carrel, and J. L. Walker. 2013. "Incorporating the influence

of latent modal preferences on travel mode choice behavior". Transportation Research Part A 54: 164-178,

[33] Ermagun, A., R., Taha Hossein Rashidi, and A. Samimi. 2015. "A joint model for mode choice and escort decisions of school trips". Transportmetrica A: Transport Science 11(3).

[34] Cartenì, A., E. Cascetta, S., de Luca. 2015. "A random utility demand model for car park&share services and the pure preference for electric vehicles", Transport Policy, 48(C), pages 49-59

[35] Hensher, Davis. A., and J., King. 2001. "Parking demand and responsiveness to supply, pricing and location in the Sydney central business district". Transportation Research Part A, 35: 177-196.

[36] Cherchi, E., and J. De Dios Ortuzar. 2002. "Mixed RP/SP models incorporating interaction effects: modelling new suburban train services in Cagliari". Transportation 29: 371-395

[37] Espino, R., J. De Dios Ortuzar, and C. Roman. 2007. "Understanding suburban travel demand: flexible modelling with revealed and stated choice data". Transportation Research Part A 41: 899-912

[38] Bos, I.D.M., E., Molin, R.E., van der Heijden, H.J.P., Timmermans. 2007. "Cognition and Relative Importance Underlying Consumer Valuation of Park-and-Ride Facilities". Transportation Research Record, 1835:121-127.

[39] Dell' Olio, L., A., Ibeas, P., Cecín, and F. Dell' Olio. 2011. "Willingness to pay for improving service quality in a multimodal area". Transportation Research Part C 19:060-1070

[40] Abrantes, P., and M. Wardman. 2011. "Meta-analysis of UK values of travel time: an update". Transportation Research Part A 45: 1-17

[41] Arentze, Theo A., and Eric J.E., Molin. 2014. "Travelers' preferences in multimodal networks: Design and results of a comprehensive series of choice experiments". Transportation Research Part A 58:15-28.

[42] Bajwa, Shamas, S., Bekhor, M. Kuwahara, E., Chung. 2008. "Discrete choice modeling of combined mode and departure time". Transportmetrica 4(2).

[43] Bhat, C.R., R., Misra. 2001. "Comprehensive activity travel pattern modeling system for nonworkers with empirical focus on organization of activity episodes". Transportation Research Record 1777:16-24

[44] Wen, C.H., and F.S. Koppelman. 2000. A conceptual and methodological framework for the generation of activity-travel patterns. Transportation 27: 5-23.

[45] Susilo, Yusak O., and R., Kitamura. 2008. "Structural changes in commuters' daily travel: The case of auto and transit commuters in the Osaka metropolitan area of Japan, 1980-2000". Transportation Research Part A 42: 95-115.

[46] Sakano, Ryoichi, and J. Benjamin. 2011. "A structural model of mode-activity choice: The case of commuter rail in a medium-size metropolitan area". Transport Policy 18(2): 434-445,

[47] Ho, Chinh Q., and C. Mulley. 2013. "Multiple purposes at single destination: A key to a better understanding of the relationship between tour complexity and mode choice". Transportation Research Part A 49: 206-219.

[48] Habib, Khandker M. Nurul, and Eric J. Miller. 2009. "Modelling activity generation: a utility-based model for

activity-agenda formation". Transportmetrica 5(1): 3-23.

[49] Zhang, Junyi, and H. Timmermans. 2012. "Activity-travel behaviour analysis for universal mobility design". Transportmetrica 8(3).

[50] Bifulco, G.N., A., Cartenì, A., and A. Papola. 2010. "An activity-based approach for complex travel behaviour modelling" European Transport Research Review 2 (4): 209-221.

[51] Ben-Akiva, Moshe, and T. Morikawa. 1990. Estimation of switching models from revealed preferences and stated intentions, Transportation Research Part A: General, 24(6): 485-495.

[52] Diana, M. 2010. "From mode choice to modal diversion: A new behavioural paradigm and an application to the study of the demand for innovative transport services". Technological Forecasting and Social Change 77(3): 429-441.

[53] Chen,C.F., and Wei-Hsiang Chao. 2011. "Habitual or reasoned? Using the theory of planned behavior, technology acceptance model, and habit to examine switching intentions toward public transit". Transportation Research Part F 14(2): 128-137,

[54] Idris, Ahmed Osman, K. M. N. Habib, and A. Shalaby. 2015. "An investigation on the performances of mode shift models in transit ridership forecasting" Transportation Research Part A 78: 551-565,

[55] de Luca, S., and R. Di Pace. 2015. "Modelling users' behaviour in inter-urban carsharing program: A stated preference approach". Transportation Research Part A 71:59-76

[56] Li, Yeun-Touh, Jan-Dirk, Schmöcker, and S.Fujii. 2015. "Demand adaptation towards new transport modes: the case of high-speed rail in Taiwan". Transportmetrica B: Transport Dynamics 3(1): 27-43

[57] Hess, S., K.E., and Train, K.E. 2011. Recovery of inter-and intra-personal heterogeneity using mixed logit models. Transportation Research Part B 45(7): 973-990.

[58] Cherchi, E., and C.A., Guevara. 2012. "A monte carlo experiment to analyze the curse of dimensionality in estimating random coefficients models with a full variance-covariance matrix". Transportation Research Part B 46(2): 321-332.

[59] Kitamura, R. 1990. "Panel analysis in transportation planning: an overview" Transportation Research Part A 24(6): 401-415.

[60] Pas, E.I., F.S., Koppelman. 1987. "An examination of the determinants of day-to-day variability in individuals' urban travel behavior", Transportation 14(1): 3-20.

[61] Cirillo, C., and K.W., Axhausen. 2002. "Mode Choice of Complex Tours: A Panel Analysis". ETH, Eidgenössische Technische Hochschule Zürich, Institut für Verkehrsplanung, Transporttechnik, Strassen-und Eisenbahnbau IVT.

[62] Ramadurai, G., and K.K, Srinivasan. 2006. "Dynamics and variability in within-day mode choice decisions: role of state dependence, habit persistence, and unobserved heterogeneity". Transportation Research Records 1977(1): 43-52

[63] Cirillo, C., and K.W., Axhausen. 2010. "Dynamic model of activity-type choice and scheduling". Transportation 37(1): 15-38.

[64] Srinivasan, K.K., and P., Bhargavi. 2007. "Longer-term changes in mode choice decisions in Chennai: a comparison between cross-sectional

and dynamic models". Transportation 34(3): 55-374.

[65] Pendyala, R., R. Kitamura, and D., Prasuna Reddy. 1998. "Application of an activity-based travel-demand model incorporating a rule-based algorithm". Environmental Planning B 25: 753-772.

[66] Arentze, Theo A., H.J., Timmermans. 2004. "A learning-based transportation oriented simulation system". Transportation Research Part B 38 (7): 613-633.

[67] Xiong, C., and L., Zhang. 2013. "Positive model of departure time choice under road pricing and uncertainty". Transportation Research Records 2345(1): 117-125.

[68] Goulias, K.G. 1999. "Longitudinal analysis of activity and travel pattern dynamics using generalized mixed markov latent class models". Transportation Research Part B 33(8): 535-558.

[69] Ben-Akiva, Moshe. 2010. Planning and Action in a Model of Choice. Choice Modelling: The State-of-the-Art and the State-of-Practice. Emerald, Bingley.

[70] Choudhury, C.F., M. Ben-Akiva, and M., Abou-Zeid. 2010. "Dynamic latent plan models". Journal of Choice Modelling 3(2):50-70.

[71] Xiao, Fu, W.H.K. Lam, and Q. Meng. 2014. "Modelling impacts of adverse weather conditions on activity–travel pattern scheduling in multi-modal transit networks". Transportmetrica B: Transport Dynamics, 2(2):151-167

[72] Xiong, C., and L. Zhang. 2015. "Dynamic travel mode searching and switching analysis considering hidden modal preference and behavioral decision processes". Transportation, 42(6): 985-1002

[73] Zhang, Junyi, T. Arentze, and H., Timmermans. 2012. "A Multimodal transport network model for advanced traveler information system". Journal of Ubiquitous Systems & Pervasive Networks, 4 (1): 21-27.

[74] Klöckner, C.A., Friedrichsmeier, T. (2011). A multilevel approach to travel mode choice—how person characteristics and situation specific aspects determine car use in a student sample, Transportation Research Part F: Traffic Psychology and Behaviour, 14 (2011), pp. 261-277

[75] Rehrl, K., S. Bruntsch, and H. J. Mentz. 2007. "Assisting multimodal travelers: design and prototypical implementation of a personal travel companion". IEEE Transactions on Intelligent Transport Systems 8 (1): 31-4

[76] Zografos, K.G., and K.N., Androutsopoulus. 2009. "Algorithms for Itinerary Planning in Multimodal Transportation Networks". IEEE Transactions on Intelligent Transport Systems, 9(1):75-184

[77] Yin, M., and M. Griss. 2005. "SCATEAgent: context-aware software agents for multi-modal travel". Applications of agent technology in traffic and transportation. Whitestein Series in Software Agent Technologies and Autonomic, Computing: 69-84.

[78] Sheffi, Yosef. 1985. "Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods". Prentice Hall.

[79] Liao, F., T.A. Arentze, and H.J.P. Timmermans. 2011. "Application of supernetworks in modeling activity-travel behaviour". Science and Technology 33:279-286

[80] Opasanon, Sathaporn, and Elise Miller-Hooks. 2001. "LEAST EXPECTED TIME HYPERPATHS IN STOCHASTIC, TIME-VARYING MULTIMODAL NETWORKS." Transportation Research Record 1771 (1):89-96. doi: 10.3141/1771-12.

[81] Leurent, Fabien. 2013. "On modal availability, travel strategies and traffic equilibrium on a multimodal network." In hal.archives-ouvertes.fr

[82] Sabir M., Koetse M. J., and P. Rietveld. 2009. "The Impact of Weather Conditions on Mode Choice: Empirical Evidence for the Netherlands." In Proceedings of the 17th Annual conference EAERE, June 2009, Amsterdam.

[83] Cantarella, G.E., A., Cartenì; and S., de Luca. 2015. "Stochastic equilibrium assignment with variable demand: theoretical and implementation issues" EUROPEAN JOURNAL OF OPERATIONAL RESEARCH 241(2): 330-347

[84] Nguyen, S., and S. Pallottino. 1988. "Equilibrium traffic assignment for large scale transit networks." European Journal of Operational Research 37 (2):176-86. doi: http://dx.doi.org/10.1016/0377-2217(88)90327-X.

[85] Cats, O., H.N., Koutsopoulos, W., Burghout, and T. Toledo. 2012. "Effect of Real-Time Transit Information on Dynamic Path Choice of Passengers".

[86] Trozzi, Valentina, Guido Gentile, Michael Bell, and Ioannis Kaparias. 2013. "Dynamic User Equilibrium in Public Transport Networks with Passenger Congestion and Hyperpaths." Procedia - Social and Behavioral Sciences 80 (0):427-54. doi:http://dx.doi.org/10.1016/j.sbspro.2013.05.024.

[87] The European Commission, DG XI — Environment. 1999. "Cycling: the way ahead for towns and cities."

[88] Mascia, M., Han, K. (2015). "The Hypermode concept", Internal report, Imperial College London.

Chapter 2

Recent Progress in Activity-Based Travel Demand Modeling: Rising Data and Applicability

Atousa Tajaddini, Geoffrey Rose, Kara M. Kockelman and Hai L. Vu

Abstract

Over 30 years have passed since activity-based travel demand models (ABMs) emerged to overcome the limitations of the preceding models which have dominated the field for over 50 years. Activity-based models are valuable tools for transportation planning and analysis, detailing the tour and mode-restricted nature of the household and individual travel choices. Nevertheless, no single approach has emerged as a dominant method, and research continues to improve ABM features to make them more accurate, robust, and practical. This paper describes the state of art and practice, including the ongoing ABM research covering both demand and supply considerations. Despite the substantial developments, ABM's abilities in reflecting behavioral realism are still limited. Possible solutions to address this issue include increasing the inaccuracy of the primary data, improved integrity of ABMs across days of the week, and tackling the uncertainty via integrating demand and supply. Opportunities exist to test, the feasibility of spatial transferability of ABMs to new geographical contexts along with expanding the applicability of ABMs in transportation policy-making.

Keywords: activity-based models, travel demand forecasting, transportation planning, big data, transferability of transport demand models

1. Introduction

In recent years, behaviorally oriented activity-based travel demand models (ABMs) have received much attention, and the significance of these models in the analysis of travel demand is well documented in the literature [1, 2]. These models are found to be consistent and realistic in several fundamental aspects. They possess some significant advantages over the simple aggregated trip-based travel demand models [3]. To achieve this, ABMs consider the linkage among activities and travel for an individual as well as different people within the same household and place more attention to the constraints of time and space. In other words, these models are capable of integrating both the activity, time, and spatial dimensions. The comprehensive advantages of activity-based models in comparison to the trip-based models have been discussed in previous papers [4–8]. Activity-based models are suitable for a wider variety of transportation policies involving individual decisions such as congestion pricing and ridesharing. More especially, enabling the relationship between activity and behavioral pattern of trip making is one of the main reasons for the shift from the aggregate-level in trip based models to disaggregate-level provided by ABMs [9].

Activity-based travel demand models (ABMs) can be classified into two main groups: Utility maximization-based econometric models and rule-based computational process models (CPM). Utility maximization-based econometric models apply different econometric structures such as logit, probit, hazard-based, and ordered response models. While the logit models rely on different assumptions about the distribution of the error terms in the utility functions, hazard-based models use the duration of activity based on end-of-duration occurrence to generate activity schedules [10]. Rule-based computational process models apply different sets of condition-action rules and focus on the implementation of daily travel and ordering activities to mimic individuals' behavior when constructing schedules. In addition to the aforementioned models, other approaches can be employed either in combination with these models or separately to develop activity-based models. Examples include agent-based and time-space prism approaches. While an agent-based approach allows agents to learn, modify, and improve their interactions with other agents as well as their dynamic environment, time-space prisms are utilized to capture spatial and temporal constraints under which individuals construct the patterns of their activities and trips. **Figure 1** exhibits critical elements of ABM such as activity generation, activity scheduling, and mobility choices. It also provides a comparison among the notable existing travel demand models regarding their different elements. The development of activity-based travel demand models has been reviewed comprehensively in previous studies [10, 11]. **Table 1** provides a summary of the literature on the evolution of these models over time by introducing the notable existing developed models and highlighting their limitations.

Despite the existence of many models as listed in **Table 1**, ABM's abilities in reflecting behavioral realism are still limited [40]. The capability of ABM models

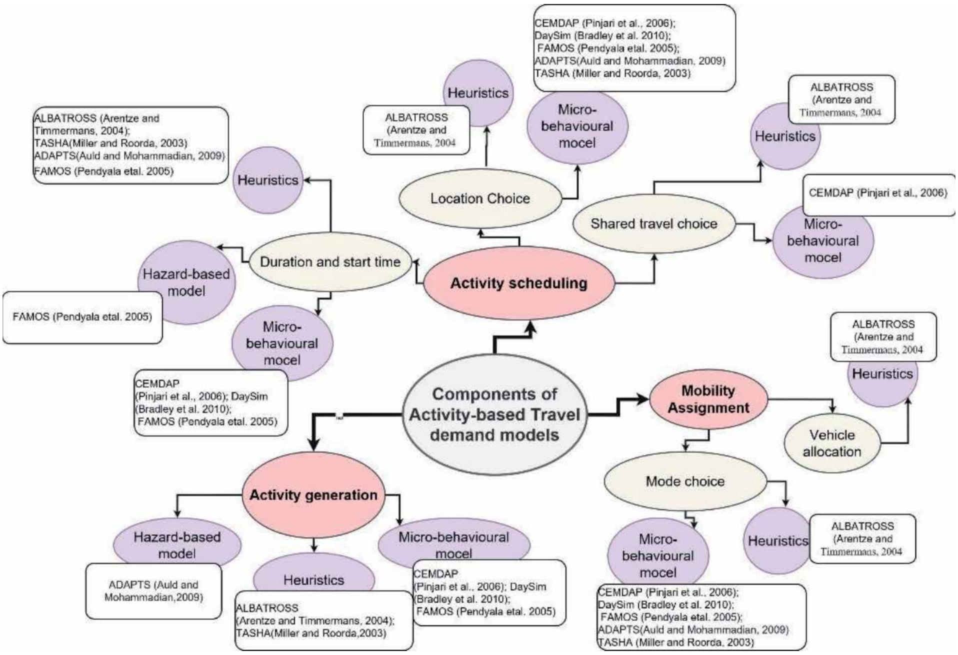


Figure 1.
Components of activity-based travel demand models.

ABM type + year of proposal	Examples	Model limitations
Constraint-based models 1967	PESASP [12]	Consider only individual accessibility, rather than household-level accessibility Some system features, like open hours and travel times, are considered fixed [11]
	CARLA [13]	
	BSP [14]	
	MAGIC [15]	
	GISICAS [16]	
Utility maximization-based models 1978	Portland METRO [17]	• Assume that all decision-makers are fully rational utility maximizers which are not realistic in practice [10] • Unable to reflect latent behavioral mechanisms in the decision processes [11]
	San Francisco SFCTA [18]	
	New York NYMTC [19]	
	Columbus MORPC [20]	
	Sacramento SACOG [21, 22]	
	CEMDAP [23, 24]	
	FAMOS [25]	
	CT-RAMP [26]	
Computational process models 2000	ALBATROSS [27, 28]	Focus more on scheduling and sequencing of activities than the underlying rules in decision-making [11]
	TASHA [29, 30]	
	ADAPTS [31–33]	
	Feathers [34]	
Agent-based modeling 2004	ALBATROSS [27, 28]	• High computational complexity • No transparency in the mechanical process of agents interacting with other agents and environment which depends on the parameters' values • Requires well-defined conditions and constraints • Non-reproducibility due to the non-streamlined process of calibrating and imputing parameters for the models [39]
	Feathers [34]	
	MATSim [35]	
	TRANSIMS [36]	
	SimMobilitiy [37]	
	POLARIS [38]	

Table 1.
ABM evolution over time.

in predicting individual travel movements can be evaluated from two perspectives of input (data) and output (applicability). Activity schedules are an essential input into the ABM model. From an input point of view, the necessity of deriving activity schedules from dynamic resources together with their challenges will be reviewed. From the applicability perspective, the application of ABM output in integration with dynamic traffic assignment (DTA) models, transferring to a new geographical context, and why and how it is applied in transport planning management will also be discussed. To this end, the first part of this paper will review the new real-time data resources revealing the pattern and traces of traveler's mobility at a large scale and over an extended period of time. The big data enables new ABM models to reflect mobility behavior on an unprecedented level of detail while collecting data over a longer period (e.g., more than one typical day) would improve the behavioral realism in trip making [41]. The second part of this paper looks into the applicability of ABM models. This part includes (i) gap investigation in enriching ABMs by integrating time-dependent OD matrices produced by ABMs with dynamic traffic assignment; (ii) investigation of ABMs' applicability in transferring from one region to another; and (iii) enriching the capability of ABMs by moving beyond the transportation domain to other such as environment and management strategies.

The remainder of the paper is organized as follows. Section 2 introduces new data sources such as mobile phone call data records, transit smart cards, and GPS data where the influence of new data sources on the planning of activities, formation, and analysis of the travel behavior of individuals will be investigated. This section also introduces activity-based travel demand models, which generates activity-travel schedules longer than a typical day. Section 3 describes the existing experiences in transferring utility-based and CPM activity-based travel demand models from one geographical area to another. This section also reviews the integration of ABM models with dynamic traffic assignment and other models such as air quality models. The possibility of using activity-based models in travel demand management strategies with a focus on car-sharing and telecommuting are considered as examples. The last section concludes the paper and identifies remaining challenges in the area of activity-based travel demand modeling.

2. ABMs and the emerging of big data

This section provides an overview of the role of big data in replacing the traditional data sources, and the changes in activity-based travel demand models given these newly available data.

2.1 Improvements in activity-based travel demand modeling

It is more than half a century that transportation planners try to understand how individuals schedule their activities and travel to improve urban mobility and accessibility. The evolution of travel demand modeling from trip-based to activity-based highlighted the need for high-resolution databases including sociodemographic and economic attributes of individuals and travel characteristics. Today, with the rapid advancements in computation, technology, and applications, the intelligent transportation systems (ITS) have revolutionized the analysis of travel behavior by having more accurate data, removing human errors, and making use of the vast amount of available data [42]. Tools such as GPS devices, smartphones, smart card data, and social networking sites all have the potential to track the movements and activities of individuals by recording and retaining the relevant data continuously over time. Most of the traditional travel survey data are rich in detail. However, it can result in biased travel demand models because of incomplete self-reports and inaccurate scheduling patterns. Therefore, in this section, the common tools used in collecting big data are introduced and the progress made in the area of extracting big data sources is discussed.

2.1.1 Cell phone data

A call detail record (CDR) is a data record produced by a telephone exchange and consists of spatiotemporal information on the recent system usage [40], which can track people's movements. This CDR data can be processed and applied in activity-based travel demand modelings to better understand human mobility and obtain more accurate origin-destination (OD) tables [43]. The first attempt using CDR data was a study of Caceres et al. [44], who applied mobile phone data to generate OD matrices. Their concept was then formalized by Wang et al. [45] to obtain transient OD matrices by counting trips for each pair of the following calls from two different telephone (cell) towers at the same hour. Afterward, using the shortest path algorithm, OD trips are assigned to the road network. In the area of urban activity recognition, Farrahi et al. [46] applied two probabilistic methods (i.e., Latent Dirichlet Allocation (LDA) and Author Topic Models, ATM) to cluster CDR trajectories according to their temporal aspects to discover the home and work activities. Considering the spatial aspect of CDR data, Phithakkitnukoon et al. [47] applied auxiliary land use data and geographical information database to find possible activities around a certain cell tower. And considering both the temporal-spatial aspect of CDR, Widhalm et al. [48] used an undirected relational Markov network to infer urban activities. They extracted activity patterns for Boston and Vienna by analyzing cell phone data (activity time, duration, and land use). Their results show that trip sequence patterns and activity scheduling observed from datasets were compatible with city surveys as well as the stability of generated activity clusters across time. In a more recent study, [49] an unsupervised generative state-space model is applied to extract user activity patterns from CDR data. Furthermore, it has been shown that the method of CDR sampling is as significant as survey sampling. For example, in one study [50], CDR and survey data is used during a period of six months to investigate the daily mobility for Paris and Chicago. The result shows that 90% of travel patterns observed in both surveys are compatible with phone data. In another similar study [51], a probabilistic induction was proposed using motifs (daily mobility network), time of day activity sequence, and land use classification to produce activity types. CDR data of Singapore was used by Jiang et al. [52] to produce activity-based human mobility patterns.

In the context of activity-based transport modeling, Zilske et al. [53] replaced travel diaries with CDRs as input data for agent-based traffic simulation. They first generated the synthetic CDR data, then the MATSim simulation software was used to identify every observed person as an agent to convert call information into activity. They fused the CDR data set with traffic counts in their next paper [54], to reduce the Spatio-temporal uncertainty.

In summary, the findings reported from different studies indicated the major implications of mobile phone records on the estimation of travel demand variables including travel time, mode and route choice as well as OD demand and traffic flow estimation; however; in practice, the information generated from CDR data are yet to be used widely in simulation models. This is mainly because of the conflict between either level of resolution or format and completeness of model and data [55].

2.1.2 Smart card data

Smart card systems with on- and off-boarding information gained much popularity in large public transport systems all over the world, and have become a new source of data to understand and identify the Spatio-temporal travel patterns of the individual passengers. The smart card data are investigated in various studies such as activity identification, scheduling, agent-based transport models, and simulation [56]. Besides, in other studies [57–59] smart card data was used as an analysis tool in investigating the passenger movements, city structure, and city area functions. Similarly, in the recent study [60], a visual analysis system called PeopleVis was introduced to examine the smart card data (SCD) and predict the travel behavior of each passenger. They used one-week SCD in the city of Beijing and found a group of “familiar strangers” who did not know each other but had lots of similarities in their trip choices. Zhao et al. [61] also investigated the group behavior of metro passengers in Zhechen by applying the data mining procedure. After extracting patterns from smart card transaction data, statistical-based and clustering-based methods were applied to detect the passengers' travel patterns. The results show that a temporally regular passenger is very probable to be a spatially regular passenger. The disaggregated nature of smart card data represents suitable input to multi-agent simulation frameworks. For example, the smart card data is used to generate activity plans and implement an agent-based microsimulation of public transport in two cities of Amsterdam and Rotterdam [62]. An agent-based transport simulation is developed for Singapore's public transport using MATSim environment [63]. Unlike Bouman's study, they considered the interaction of public transport with private vehicles. The study of Fourie et al. [64] was another research work to present the possibility of integrating big data algorithms with agent-based transport models. Zhu [65] compared one-week transaction data of smart cards in Shanghai and Singapore. They found feasibility in generating continuous transit use profiles for different types of cardholders. However, to have a better understanding of the patterns and activity behaviors, in addition to collecting the data from smart cards, one should integrate them with other data set.

2.1.3 GPS data

In travel demand modeling, it is important to have accurate and complete travel survey data including trip purpose, length, and companions, travel demand, origin and destination, and time of the day. Since the 1990s, the global positioning system (GPS) became popular for civil engineering applications, especially in the field of transportation as it provides a means of tracking some of the above variables. In the literature, methods of processing the GPS data and identifying activities can be classified according to different approaches such as rule-based and Bayesian model [66]; fuzzy logic [67]; multilayer perceptron [68]; and support vector machine learning [69]. Nevertheless, the disadvantages of using GPS data include the cost, sample size limitation, and the need to retrieve and distribute GPS devices to participate. Since smartphones are becoming one of the human accessories while equipped with a GPS module, they can be considered as a replacement of the GPS device to gather travel data. In this regard, CDR from smartphones is used [70] to estimate origin-destination matrices, or a smartphone-based application is used [71] to map the semiformal minibus services in Kampala (Uganda) and to count passenger boarding and alighting [72]. In the Netherlands, the Mobidot application is developed for analyzing the mobility patterns of individuals. To deduce travel directions and modes, this application uses the real-time data gathered by sensors of smartphones including GPS, accelerometer, and gyroscope sensors to compare them with existing databases [73].

Applying smartphones as a replacement of GPS however, holds several restrictions including the draining of smartphone battery and it is not possible to record travel mode and purpose.

2.1.4 Social media data

Today transport modelers, planners, and managers have started to benefit from the popularity of social networking data. There are different kinds of social media data such as Twitter, Instagram, and LinkedIn data, which consist of normal text, hash-tag (#), and check-in data. As hash-tag and check-in data are related to an activity, location or event, they can be used as meaningful resources in analysis of destination/origin of the activity [74]. According to the literature, social media has a great influence on different aspects of travel demand modeling [75]. Using social media instead of traditional data collection methods was investigated in different studies [76]. The way of processing these data to extract useful information is challenging as investigated in different studies [77, 78]. Various studies [79–82] also examined social media data to understand the mobility behavior of a large group of people. Testing the possibility of evaluating the origin-destination matrix based on location-based social data was researched [83] or in another similar studies [84, 85] where Twitter data was used to estimate OD matrices. The comparison between this new OD with the traditional values produced by the 4-step model proved the great potential of using social media data in modeling aggregate travel behavior. Social media data can be used in other areas such as destination choice modeling [86], recognizing activity [87], understanding the patterns of choosing activity [80, 88, 89], and interpreting life-style behaviors via studying activity-location choice patterns [90].

2.2 Dynamic ABM using a multi-day travel data set

Most existing travel demand modelers have applied the household survey data during the period of one day to construct activity schedules. However, longer periods such as one week or one month gained substantial importance during recent years. For simulating everyday travel behavior and generating schedules, a one-week period provides more comprehensive coverage because it includes weekdays and weekends and represents the weekly routines of individuals in making trips. Periods longer than one week can further provide detail on personal behavior as well as various usage of modes in different ways. So far only a few travel demand models covered a typical week as a studied period. For example rhythm in activity-travel behavior based on the capacity of one week was presented by applying a Kuhn-Tucker method [41]. Few works have been concentrating on the generation of multiple-day travel dataset. For example, by using large data and surveys, Medina developed two discrete choice models for generating multi-day travel activity types based on the likeliness of the activity [91]. a sampling method based on activity-travel pattern type clustering [92] was proposed to extract multi-day activity-travel data according to single-day household travel data. The results show similarities in distributions of intrapersonal variability in multi-day and single-day. MATSim is a popular agent-based simulation for ABM research [93, 94], however, it is not appropriate for modeling the multi-day scenarios because MATSim uses the co-evolutionary algorithm to reach the user equilibrium which is a time consuming particularly for multi-day plans. To solve these problems, Ordonez [95] proposed a differentiation between fixed and flexible activities. Based on different time scales, Lee examined three levels of travel behavior dynamics, namely micro-dynamics (24 hours), macro-dynamics (lifelong travel behavior), meso-dynamics (weekly/monthly/yearly basis) by applying different statistical models [96]. A learning day-by-day module in another agent-based simulation software SimMobility is proposed [97]. Furthermore, ADAPTS is one of the few activity-based travel demand models which depends on activity planning horizon data for a longer period than one day, for example, one week or one month [33].

As highlighted by the above literature review, applying one-day observation data in travel demand modeling provides an inadequate basis of understanding of complex travel behavior to predict the impact of travel demand management strategies. So multi-day data are needed to refine this process. Previously, it was not easy to collect multi-day data, however, today thanks to advantages to technology it is possible to extract data from GPS, smartphones, smart cards, etc. with no burden for the respondent. Models built based on GPS data have been found to be more accurate and precise due to having fewer measurement errors. Collecting call detail records from mobile phones provide modelers with large trip samples and origin-destination matrices, while smart card data are more useful in terms of validation.

3. ABM transferability

We now turn to the recent advances and ongoing research in ABM focused on testing and enhancing geographical transferability and capacity to predict a broader range of impacts than flows and performance of the transport network.

3.1 ABM transferability from one geographical context to another

The spatial transferability of a travel demand model happens when the information or theory of a developed model of one region is applied to a new context [98]. Transferability can be used not only as a beneficial validation test for the models but also to save the cost and time required to develop a new model. Validation of a model by testing spatial transferability beside other various methods such as base-year and future-year data set is a test of validity which represents the capability of activity-based models in predicting travel behavior in a different context [99]. The exact theoretical basis and behavioral realism of activity-based travel demand model make them more appropriate for geographic transferability in comparison to traditional trip-based models [100]. Testing the transferability of ABM was first investigated by Arentze et al. [101]. They examined the possibility of transferring the ALBATROSS model at both individual and aggregate levels for two municipalities (Voorhout and Apeldoorn) in the Netherlands by simulating activity patterns. The results were satisfactory except for the transportation mode choice. In the United States, the CT-RAMP activity-based model which was developed for the MORPC region then transferred to Lake Tahoe [102]. In another study, one component of the ADAPTS model showed the potential for having good transferability properties [31]. The transferability of the DaySim model system developed for Sacramento to four regions in California and two other regions in Florida was investigated in [103]. The results show that the activity generation and scheduling models can be transferred better than mode and location choice models. The CEMDAP model developed for Dallas Fort Worth (DFW) region was transferred to the southern California region [104]. Outside the U.S., the TASHA model system developed for Toronto was transferred to London [105], and also in another study [106] the transferability of TASHA to the context of the Island of Montreal was assessed. Activity generation, activity location choice, and activity scheduling were three components of TASHA that transferred from Toronto to Montreal. In general, TASHA provided acceptable results at (macro and meso-level) for work and school activities even in some cases better results for Montreal in comparison to Toronto area. The possibility of developing a local area activity-based transport demand model for Berlin by transferring an activity generation model from another geographical area (Los Angeles) and applying the traffic counts of Berlin was investigated [107]. In their research, the CEMDAP model was applied to achieve a set of possible activity-travel plans, and the MATSim simulation was then used to generate a representative travel demand for the new region. The results were quite encouraging, however, the study indicated a need for further evaluation. In one recent study [108], an empirical method was used to check the transferability of ABMs between regions. According to their investigations, the most difficult problems with transferability caused by parameters of travel time, travel cost, land use, and logsum accessibilities. They suggested that in the transferability of the ABM from another region, agencies should be aware of finding a region within the same state or with similar urban density, or preferably both in order to improve the results. The possibility of transferring the FEATHERS model to Ho Chi Minh in Vietnam is investigated [109]. FEATHERS initially is developed for Flanders in Belgium. After calibration of FEATHERs sub-models, testing results using different indicators confirmed the success of transferring the FEATHER's structure to the new context.

At the theoretical level, a perfect transferable model contributes to the transferability of its underlying behavioral theory, model structure, variable specification and coefficient to the new context. However, perfect transferability is not easy to achieve due to different policy and planning needs as well as the size of the regions, and the availability of data and other resources. Although the results of several transferred ABM model systems seem to have worked reasonably, it is equally important to assess how much accuracy is important in transferring models and how best and where to transfer models from.

3.2 ABM transferability to other non-transport domain

One of the advantages of the activity-based travel demand models over trip-based models is its capability to generate various performance indicators such as emission, health-related indicators, social exclusion, well-being, and quality of life indicators. Application of disaggregate models for the area of emission and air quality analysis was introduced by Shiftan [110] who investigated the Portland activity-based model in comparison to trip-based models. In another study [111], the same author integrated the Portland activity-based model with MOBILE5 emission model to study the effects of travel demand techniques on air quality. Regarding the integration of ABM with the emission model, the Albatross ABM model was coupled with MIMOSA (macroscopic emission model) [112] considering the usage of fuel and the amount of produced emission as a function of travel speed. A study in [113] added one dispersion model (AUROTA) to the previous integration of Albatross and MIMOSA to predict the hourly ambient pollutant. Albatross linked with a probabilistic air quality system was employed [114] in air quality assessment study. TASHA was another activity-based model, which has been extensively employed in air quality studies. For example, this model was integrated [29, 115] with MOBILE6.2 to quantify vehicle emissions in Toronto. In their study, EMME/2 was used in the traffic assignment part. The previous research was improved [116] by replacing EMME/2 with MATSim as an agent-based DTA model. This TASHA-MATSim chain was used in the research [117] with the integration of MOBILE6.2C (emission model) and CALPUFF (dispersion model). OpenAMOS linked with MOVES emission model [118], and ADAPTS linked with MOVES [119] together with Sacramento ABM model [120] are among recent studies which represented the application of activity-based models in analyzing the impacts of vehicular emissions.

Human well-being and personal satisfaction play an important role in social progression [121]. To understand the theory behind human happiness, transport policies concentrated on the concept of utility as a tool to increase activity, goods, and services [122, 123]. The issue of well-being as a policy objective is addressed in the literature and measured through various indicators, which show personal satisfaction and growth. For example, in the study by Hensher and Metz [124, 125], saving time which leads to engagement in more activities was introduced as one of the benefits of measuring transport performance. Spatial accessibility was another benefit of travel that provides a range of activities that can be reasonably reached by individuals [126]. A dynamic ordinal logit model was developed [127] based on the collected data on happiness for a single activity in Melbourne. The authors found different activity types, which have different influences on the happiness that each individual experienced. Well-being can be integrated into activity-based models based on random utility theory. In terms of modeling, a framework was introduced [122] considering well-being data to improve activity-based travel demand models. According to their hypothesis, well-being is the final aim of activity patterns. They applied a random utility framework and considered well-being measures as indicators of the utility of activity patterns, and planned to test their framework empirically by adding well-being measurement equations to the DRCOG's activity-based model.

The above literature review showed the importance of applying traffic models to generate input data to other models such as the air quality model. The accuracy of emission models is highly dependent on the level of detail in transport demand model inputs. Activity-based and agent-based models are supposed to describe reality more accurately by providing more detailed traffic data. Beyond measurement of air quality, well-being and health have drawn increasing attention. The health impact of changes in travel behavior, health inequalities, and social justice can be assessed within the activity-based platform [128]. With the help of geospatial data acquisition technologies like GPS, behavioral information with health data can be integrated into the development of an activity-based model to provide policies that affect the balance of transport and well-being.

3.3 ABM integration with dynamic traffic assignment

In parallel with the travel demand modeling, on the supply side, the conventional supply models used to be static, which import constant origin-destination flows as an input and produce static congestion patterns as an output. Consequently, these models were unable to represent the flow dynamics in a clear and detailed manner. Dynamic traffic assignment (DTA) models have emerged to address this issue and are capable of capturing the variability of traffic conditions throughout the day. It is evident that the shift of analysis from trips to activities in the demand modeling, as well as, the substitution of the static traffic assignment with dynamic traffic assignment in the supply side, can provide more realistic results in the planning process. Furthermore, the combination of ABM and DTA can better represent the interactions between human activity, their scheduling decision, and the underlying congested networks. Nevertheless, according to the study of [11], the integration of ABM with DTA received little attention and still requires further theoretical development. There are different approaches to the integration of ABM and DTA, which started with a sequential integration. In this type of integration, exchanging data between two major model components (ABM and DTA) happens at the end of the full iteration, to generate daily activity patterns for all synthetic population in an area of study, the activity-based model is run for the whole period of a complete day. The outputs of the ABM model which are lists of activities and plans are then fed into the DTA model. The DTA model generates a new set of time-dependent skim matrices as inputs to ABM for the next iteration. This process is continued until the convergence will be reached in the OD matrices output. Model systems applying the sequential integration paradigm can be found in most of the studies in the literature. For example, Castiglione [129] integrated DaySim which is an activity-based travel demand model developed for Sacramento with a disaggregate dynamic network traffic assignment tool TRANSIMS router. Bekhor [130] investigated the possibility of coupling the Tel Aviv activity-based model with MATSim as an agent-based dynamic assignment framework. Hao [116] integrated the TASHA model with

MATSim. Ziemke [107] integrated CEMDAP, which is an activity-based model with MATSim to check the possibility of transferring an activity-based model from one geographic region to another. Lin [131] introduced the fixed-point formulation of integrated CEMDAP as an activity-based model with an Interactive System for Transport Algorithms (VISTA). Based on the mathematical algorithm of household activity pattern problem (HAPP), ABM and DTA were integrated [132] by presenting the dynamic activity-travel assignment model (DATA) which is an integrated formulation in the multi-state super network framework.

In the sequential integration, the ABM and DTA models run separately until they reach convergence. At the end of an iteration, these models perform data exchange before iterate again. Therefore, this kind of integrated framework cannot react quickly and positively to network dynamics and is unable to adapt to real-time information available to each traveler. In addressing this limitation, integrated models that adopt a much tighter integration framework have been developed recently. This approach is quite similar to the sequential approach, however; the resolution of time for ABM simulation is one minute rather than 24 hours (complete day). Relating to this level of dynamic integration, Pendyala [133] investigated the possibility of integrating OpenAMOS which is an activity-travel demand model with DTA tool name MALTA (Multiresolution Assignment and loading of traffic activities) with appropriate feedback to the land-use model system. For increasing the level of dynamic integration of ABM and DTA models, dynamic integration having pre-trip enroute information with full activity-travel choice adjustments has been introduced. In this level of ABM & DTA integration, it is assumed that pre-trip information is available for travelers about the condition of the network. It means that travelers are capable of adjusting activity-travel choices since they have access to pre-trip and Enroute travel information. Another tightly integrated modeling framework was proposed in [134] to integrate ABM (openAMOS) and DTA (DTALite) to capture activity-travel demand and traffic dynamics in an on-line environment. This model is capable of providing an estimation of traffic management strategies and real-time traveler information provision. Zockaie et al. [135] presented a simulation framework to integrate the relevant elements of an activity-based model with a dynamic traffic assignment to predict the operational impacts related to congestion pricing policies. Auld et al. [38] developed an agent-based modeling framework (POLARIS) which integrates dynamic simulation of travel demand, network supply, and network operations to solve the difficulty of integrating dynamic traffic assignment, and disaggregate demand models. A summary of the current literature on ABM and DTA integration is presented in **Table 2**.

The above discussion illustrates that most of the model integration platforms between ABM + DTA work based on sequential integration. This loose coupling platform is the most straightforward and popular approach albeit is not responsive to network short-term dynamics and real-time information. Efforts to develop a comprehensive simulation model that can account for all components of dynamic mobility and management strategies continue. Further developments will have to deal with the implementation of an integrated ABM + DTA platform on a large network to support decision-makers, focus on the integration between activity-based demand models and multimodal assignment [143] as well as reducing computational efforts via better data exchange procedure and improving model communication efficiency. Defining practical convergence criteria is another issue which needs further investigations. Fully realistic convergence is normally never happened in sequential integration due to applying a pre-defined number of feedback loops in order to save model runtime.

Paper	ABM structure	DTA Structure	Method of integration	Insights
[136]	Kutter Model developed for the city of Berlin	Multiagent Simulation (MATSim)	Sequential	Discuss the disadvantages of the integration of ABM and DTA using OD matrices and link travel times
[137]	TASHA model	Multiagent Simulation (MATSim)	Sequential	Show the advantages of the microsimulation approach over conventional methodologies relying heavily on temporal or spatial aggregation
[138]	CEMDAP	(VISTA)	Sequential	Show the impacts of multiple time interval portioning and varying step size on reaching faster and more stable convergence results
[130]	Tel Aviv activity-based model	Multi-agent Simulation (MATSim)	Sequential	Show improved run times, the full activity list can be used directly, without creating origin-destination matrices
[129, 139]	DaySim ABM model developed for the Sacramento and Jacksonville	Disaggregate dynamic network assignment tool (TRANSIMS)	Sequential	Running time limitations prevent the models to realistically represent the impacts of network events or disruptions on activity-travel patterns
[140]	Agent-based Dynamic Activity Planning and Travel Scheduling (ADAPTS) developed for the Chicago region	Disaggregate dynamic network assignment tool (TRANSIMS)	Sequential	Choosing smaller time steps in the interaction of ABM and DTA makes integration more accurate
[133]	Simulator of travel, route, activity, vehicles, emission and land use (SimTRAVEL) that integrates land-use, activity-based travel demand with DTA models		Dynamic integration	Show the proposed model is capable of simulating the behavioral pattern of human activity in space, time, and networks
[134]	ABM (openAMOS) and DTA (DTALite)		Dynamic integration	Show the model is capable of providing an estimation of traffic management strategies and real-time traveler information provision
[132]	Formulation of a dynamic activity-travel assignment (DATA) model in the multi-state supernetwork framework combining ABM and DTA		Dynamic integration	Show the power of the model to capture multi-modal and multi-activity trip chaining at equilibrium states while sensitive to policy interventions
[141]	Integrated ABM-DTA framework to consider congestion pricing in a large-scale network		Dynamic integration	A user-based approach to evaluate equilibrium conditions

Paper	ABM structure	DTA Structure	Method of integration	Insights
[38]	POLARIS, which executes a continuous exchange of information between the ABM and DTA components		Dynamic integration	The resulting gains in computational efficiency and performance allow planning models to include previously separate aspects of the urban system
[92]	Advanced demand models (InSITE ABM)	Time-sensitive traffic network model (DTALite)	Sequential	Show the efficiency of the model over the static assignment-based ABM capturing behavioral changes at a finer time resolution
[142]	The ABM (CT-RAMP)	DTA (DynusT)	Sequential	Evaluate different convergence measurements: ABM demand, DTA in terms of a gap of costs

Table 2.
A summary of the empirical literature on ABM and DTA integration.

3.4 ABM and travel demand management applications

Travel demand management (TDM) strategies are implemented to increase the efficiency of the transportation system and reduce traffic-related emissions. Some examples include mode shift strategies (encouraging people to use public transport) [144], time shift (to ride in off-peak hours, congestion pricing), and travel demand reduction [145] (using shared mobility service or teleworking). Shared transport services including car sharing, bike sharing, and ridesharing have been implemented in most of the transport planning systems across the world. Applying activity-based travel demand models to study the optimal fleet size can be found in different studies in the literature [146, 147]. Parking price policies and their impacts on car sharing were investigated using MATSim in [148]. Results show shared vehicles use more efficient parking spaces in comparison to private vehicles. In the first attempt to model car sharing on more than one typical day [149] the agent-based simulation (mobitopp) was extended with a car-sharing option to study the travel behavior of the population in the city of Stuttgart in one week. In the recent study of [150], car sharing was integrated into an activity-based dynamic user equilibrium model to show the interaction between the demand and supply of car sharing. Among all the TDM strategies, telecommuting can be implemented in a shorter time [151–153]. The results of these studies present a reduction in vehicle-kilometers-traveled (VKT) during peak hours mainly because telecommuters change their trip timetable during these times. This plan rescheduling is also investigated and addressed in different studies [154] based on the statistical analysis of worker's decisions about choice and frequency of telecommuting. While the plan rescheduling leads to reducing commute travel, the overall impacts of telecommuting on the formation of worker's daily activity-travel behavior is challenging. For example, this policy reduced total distance traveled by 75% on telecommuting days while telecommuting could reduce the total commute distance up to 0.8% and 0.7% respectively [151, 155]. Based on the adoption and frequency of telecommuting, a joint discrete choice model of home-based commuting was developed for New York city using the revealed preference (RP) survey [156]. Their results show a powerful relationship among individuals' attributes, households' demographics, and work-related factors, and telecommuting adoption and frequency decisions. A similar study [157] estimated the telecommuting choice and frequency by using a binary choice model and ordered-response model respectively. In terms of using activity-based modeling, [158] POLARIS activity-based framework was applied to research telecommuting adoption behavior and apply MOVES emission simulator model to assess the consequences of implementing this policy on air quality. Their results show that considering 50% of workers in Chicago with flexible working time hours in comparison to the base case with 12% flexible time hour workers, telecommuting can reduce Vehicle Mile Traveled (VMT) and Vehicle Hour Traveled (VHT) by 0.69% and 2.09% respectively. This policy reduces greenhouse gas by up to 0.71% as well. Pirdavani et al. [159] investigated the impact of two TDM scenarios (increasing fuel price and considering teleworking) on traffic safety. In this work, FEATHERS model, which is an activity-based model, was applied to produce exposure matrices to have a more reliable assessment. The results show the positive impacts of two scenarios on safety (**Figure 2**).

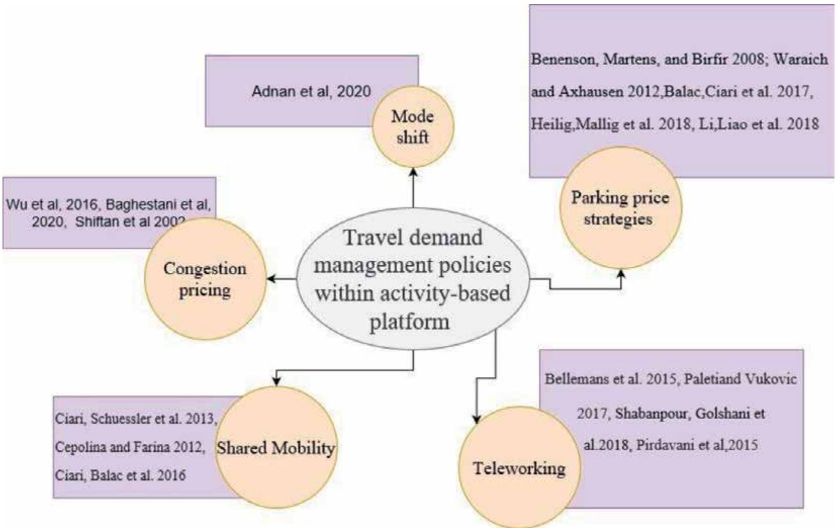


Figure 2.
Travel demand management policies within the activity-based platform.

The above section explores the relationship between transport demand management policies and travel behavior in the ABM context. The use of an activity-based travel demand model provides flexibility to employ a range of policy scenarios, and at the same time, the results are as detailed as possible to obtain the impact of policies on a disaggregated level. The finding highlights the importance of implementing different transportation policies management together to reach the most appropriate effect in terms of improving sustainability and the environment. The discussion emphasizes the need for considering more comprehensive transportation and environmental policies concerning sustainability to tackle travel planning in light of the increasingly diverse and complex travel patterns.

4. Summary and research directions

The use of activity-based models to capture complex underlying human's travel behavior is growing. In this paper, we began by introducing the components of activity-based models and the evolution of the existing developed ABM models. In the first part of this paper, the new resources of data for travel demand analysis were introduced. In the new era of travel demand modeling, we need to deal with a dynamic, large sample, time-series data provided from new devices, and as a result manage observation covering days, weeks, and even months. The outcome of the recent works revealed that since activity-based models originated from the concept of individual travel patterns rather than aggregate flows, they highly suited to these new big data sources. These big datasets, which document human movements, include the information about mobility traces and activities carried out. Based on the in-depth and critical review of the literature, it is clear that while these big datasets provide detailed insight into travel behavior, challenges remain in extracting the right information and appropriately integrating them into the travel demand models. In particular, extracting personal characteristics and trip information like trip purpose and mode of transport are still open problems as these big data resources which provide space-time traces of trip-maker behaviors. Research works along these lines have been started as it was reviewed in the first part; however, further researches should be conducted to handle the uncertainty of big data mobility traces in the modeling process. Also, new methods should be investigated to validate the results for each step of the data analysis and mining. The possibility of fusing data from different available datasets needs further investigation. For instance, to understand the mode inference both data from the smart card and CDRs can be analyzed simultaneously. Another challenging issue regarding the application of this rich new data in transport modeling is that the need for methodologies to extract useful information needed regarding the traveler's in-home and out-of-home activity patterns, which highlights the combination of data science, soft computing-based approaches, and transport research methods. It requires new Different algorithms such as statistical, genetic, evolutionary, and fuzzy as well as different techniques including advanced text and data mining, natural language processing, and machine learning.

The spatial transferability of activity-based travel demand models remains an important issue. Generally, it is found that the transferability of these models is more feasible than trip-based models, especially between two different regions with similar density or even between two areas in the same state. To date, most of the transferability research in activity-based travel demand modeling is motivated by a desire to save time, and very few studies that applied spatial transferability of activity-based models have undertaken rigorous validation of the results. While literature showed successful model transferability in terms of transferring activity/tour generation, time-of-day choice components, more studies are required on the model transferability regarding mode and location choice models as well as the validation test of activity-based models in different levels, i.e., micro, meso, and macro models.

As part of the second section of this study, this paper reviewed the progress made in the integration of activity-based models with dynamic traffic assignment.

Based on the literature, although evolution has occurred in DTA models, the loose coupling (sequential method) between ABM and DATA models still dominate the field. Two main challenges remain, namely poor convergence quality and excessively long run time. Replacing MATSim as a dynamic traffic assignment tool with other route assignment algorithms in recent years was a technical solution to loose coupling, which considered route choice as another facet of a multi-dimensional choice problem. MATSim provides not only an integration between the demand and supply side, but it can also act as a stand-alone agent-based modeling framework. However; MATSim potential drawbacks include being based on unrealistic assumptions of utility maximization and perfect information. To remove these unrealistic rational behavioral assumptions, applying other approaches such as a new innovative method of behavioral user equilibrium (BUE) is needed. This method helps trip-makers to reach certain utility-level rather than maximize the utility of their trip making [160]. Work along this approach has started (e.g., [161]).

The capability of activity-based models in generating other kinds of performance indicators in addition to OD matrices was also reviewed. Literature proved activity-based models generate more detailed results as inputs to air quality models, however; error rises from the accuracy of the information has a relevant impact on the process of integration. So it is necessary to do a comprehensive analysis of the uncertainties in traffic data. Literature proved that despite of the improvements in such disaggregate frameworks and the capability of these models in replicating policy sensitive simulation environment; there is yet to develop the best and perfect traffic-emission-air quality model. While the issue of health has drawn extensive attention from many fields, activity-based travel demand models have proved to have the potential to be used in estimating health-related indicators such as well-being. However, very few studies have been found to investigate the theories required to extend the random utility model based on happiness. While it is proved that mobility and environment have direct impacts on transport-related health [162], investigations on how travel mode preferences and air pollution exposure are related in this context are needed. Another area of research within ABM platform which is yet to be studied is the relationship between individual exposure to air pollution and mobility, especially in space, and time.

In the last part of this paper, the capability of activity-based models in the analysis of traffic demand management was investigated. Generally, the influence of telecommuting on both travel demand and network operation is still incomplete. Very few studies were found in which activity-based framework is used to simulate the potential impacts of telecommuting on traffic congestion and network operation where the real power of activity-based models lie.

In conclusion, while there are still open problems in activity-based travel demand models, there has been a lot of progress being made which is evidenced by the various recent and on-going researches reviewed in this paper. The review showed that by applying different methodologies in the modeling of different aspects of activity-based models, these models are becoming more developed, robust, and practical and become an inevitable tool for transport practitioners, city planners, and policy decision-makers alike.

Acknowledgements

The research work presented in this paper was supported by the Australian Government-Department of Education under Research Training Program (RTP Stipend) award.

Author details

Atousa Tajaddini[1]*, Geoffrey Rose[1], Kara M. Kockelman[2] and Hai L. Vu[1]*

1 Institute of Transport Studies, Monash University, Melbourne, Vic, Australia

2 The University of Texas, Austin, TX, The United States of America

*Address all correspondence to: atousa.tajaddini@monash.edu
and hai.vu@monash.edu

References

[1] Vovsha, P., M. Bradley, and J.L. Bowman. *Activity-based travel forecasting models in the United States: progress since 1995 and prospects for the future*. in *at EIRASS Conference on Progress in Activity-Based Analysis*. 2005. Vaeshartelt Castle, Maastricht.

[2] Joe Castiglione, M.B., and John Gliebe, *Activity-Based Travel Demand Models A Primer*, in *SHRP 2 Report S2-C46-RR-1*. 2015, TRANSPORTATION RESEARCH BOARD: Washington, D.C.

[3] McNally, M.G. and C.R. Rindt, *The activity-based approach*, in *Handbook of Transport Modelling: 2nd Edition*. 2007, Emerald Group Publishing Limited. p. 55-73.

[4] Henson, K., K. Goulias, and R. Golledge, *An assessment of activity-based modeling and simulation for applications in operational studies, disaster preparedness, and homeland security.* Transportation Letters, 2009. **1**(1): p. 19-39.

[5] Davidson, B., P. Vovsha, and J. Freedman. *New advancements in activity-based models*. in *Australasian Transport Research Forum*. 2011.

[6] Chu, Z., L. Cheng, and H. Chen, *A Review of Activity-Based Travel Demand Modeling*, in *CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient*. 2012. p. 48-59.

[7] Zhong, M., et al., *A comparative analysis of traditional four-step and activity-based travel demand modeling: a case study of Tampa, Florida.* Transportation Planning and Technology, 2015. **38**(5): p. 517-533.

[8] Subbarao, S.S. and K. Krishnarao, *Activity based approach to travel demand modelling: An overview.* EUROPEAN TRANSPORT-TRASPORTI EUROPEI, 2016 (61).

[9] Daisy, N.S.M., Hugh; Liu, Lei, *Individuals' Activity-Travel Behavior in Travel Demand Models: A Review of Recent Progress*, in *18th COTA International Conference of Transportation Professionals*. 2018: Beijing, China.

[10] Pinjari, A.R. and C.R. Bhat, *Activity-based travel demand analysis.* A Handbook of Transport Economics, 2011. **10**: p. 213-248.

[11] Rasouli, S., and Harry Timmermans, *Activity-Based Models of Travel Demand: Promises, Progress and Prospects.* International Journal of Urban Sciences 2014. **18(1)**: p. 31-60

[12] Lenntorp, B., *Paths in space-time environment: A time geographic study of possibilities of individuals. The Royal University of Lund, Department of Geography.* Lund Studies in Geography, Series B. Human Geography, 1976. **44**.

[13] Jones, P., et al., *Understanding Travel Behavior Gower Publishing Co.* Ltd. Aldershot, UK Google Scholar, 1983.

[14] Huigen, P.P., *Binnen of buiten bereik?: Een sociaal-geografisch onderzoek in Zuidwest-Friesland*. 1986: na.

[15] Dijst, M. and V. Vidakovic, *INDIVIDUAL ACTION SPACE IN THE CITY.* Activity-based approaches to travel analysis, 1997.

[16] Kwan, M.-P., *GISICAS: AN ACTIVITY-BASED TRAVEL DECISION SUPPORT SYSTEM USING A GIS-INTERFACED COMPUTATIONAL-PROCESS MODEL.* Activity-based approaches to travel analysis, 1997.

[17] Bradley, M.A., Portland Metro, J. L. Bowman, *A system of activitybased*

models for Portland, Oregon, in *USDOT report number DOT-T-99-02*. 1998: Washington, D.C.

[18] Bradley, M., et al. *Estimation of an activity-based micro-simulation model for San Francisco*. in *80th Annual Meeting of the Transportation Research Board, Washington DC*. 2001.

[19] Vovsha, P., E. Petersen, and R. Donnelly, *Microsimulation in travel demand modeling: Lessons learned from the New York best practice model.* Transportation Research Record: Journal of the Transportation Research Board, 2002(1805): p. 68-77.

[20] Consult, P.B., *The MORPC travel demand model: Validation and final report.* Prepared for the Mid-Ohio Region Planning Commission, 2005.

[21] Bowman, J. and M. Bradley, *Activity-based travel forecasting model for SACOG: Technical memos numbers 1-11.* Available at* HYPERLINK" http://jbowman.net"* http://jbowman.net, 2005.

[22] Bradley, M. and J. Bowman. *A summary of design features of activity-based microsimulation models for US MPOs*. in *White Paper for the Conference on Innovations in Travel Demand Modeling, Austin, TX*. 2006.

[23] Bhat, C., et al., *Comprehensive econometric microsimulator for daily activity-travel patterns.* Transportation Research Record: Journal of the Transportation Research Board, 2004(1894): p. 57-66.

[24] Pinjari, A.R., et al., *Activity-based travel-demand analysis for metropolitan areas in Texas: CEMDAP Models, Framework, Software Architecture and Application Results*. 2006.

[25] Pendyala, R., et al., *Florida activity mobility simulator: overview and preliminary validation results.* Transportation Research Record: Journal of the Transportation Research Board, 2005(1921): p. 123-130.

[26] Davidson, W., et al. *CT-RAMP family of activity-based models*. in *Proceedings of the 33rd Australasian Transport Research Forum (ATRF)*. 2010.

[27] Arentze, T. and H. Timmermans, *Albatross: a learning based transportation oriented simulation system*. 2000: Citeseer.

[28] Arentze, T.A. and H.J. Timmermans, *A learning-based transportation oriented simulation system.* Transportation Research Part B: Methodological, 2004. **38**(7): p. 613-633.

[29] Miller, E. and M. Roorda, *Prototype model of household activity-travel scheduling.* Transportation Research Record: Journal of the Transportation Research Board, 2003(1831): p. 114-121.

[30] Roorda, M.J. and E.J. Miller, *Strategies for resolving activity scheduling conflicts: an empirical analysis.* 2005.

[31] Auld, J. and A. Mohammadian, *Framework for the development of the agent-based dynamic activity planning and travel scheduling (ADAPTS) model.* Transportation Letters, 2009. **1**(3): p. 245-255.

[32] Auld, J.A. and A. Mohammadian. *Planning Constrained Destination Choice Modeling in the Adapts Activity-Based Model*. in *Proceedings of the Transportation Research Board 90th Annual Meeting, Washington, DC, USA*. 2011.

[33] Auld, J. and A.K. Mohammadian, *Activity planning processes in the Agent-based Dynamic Activity Planning and Travel Scheduling (ADAPTS) model.* Transportation Research Part A: Policy and Practice, 2012. **46**(8): p. 1386-1403.

[34] Bellemans, T., et al., *Implementation framework and development trajectory of FEATHERS activity-based simulation platform.* Transportation Research Record: Journal of the Transportation Research Board, 2010(2175): p. 111-119.

[35] Balmer, M., K. Axhausen, and K. Nagel, *Agent-based demand-modeling framework for large-scale microsimulations.* Transportation Research Record: Journal of the Transportation Research Board, 2006(1985): p. 125-134.

[36] Nagel, K., R.L. Beckman, and C.L. Barrett. *TRANSIMS for transportation planning*. in *In 6th Int. Conf. on Computers in Urban Planning and Urban Management*. 1999. Citeseer.

[37] Adnan, M., et al. *Simmobility: A multi-scale integrated agent-based simulation platform*. in *95th Annual Meeting of the Transportation Research Board Forthcoming in Transportation Research Record*. 2016.

[38] Auld, J., et al., *POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations.* Transportation Research Part C: Emerging Technologies, 2016. **64**: p. 101-116.

[39] Kagho, G.O., M. Balac, and K.W. Axhausen, *Agent-based models in transport planning: Current state, issues, and expectations.* Procedia Computer Science, 2020. **170**: p. 726-732.

[40] Miller, E.J., *Travel demand models, the next generation: Boldly going where no-one has gone before*, in *Mapping the Travel Behavior Genome*. 2020, Elsevier. p. 29-46.

[41] Nurul Habib, K., E. Miller, and K. Axhausen, *Weekly rhythm in joint time expenditure for all at-home and out-of-home activities: application of Kuhn-Tucker demand system model using multiweek travel diary data.* Transportation Research Record: Journal of the Transportation Research Board, 2008(2054): p. 64-73.

[42] Anda, C., A. Erath, and P.J. Fourie, *Transport modelling in the age of big data.* International Journal of Urban Sciences, 2017. **21**(sup1): p. 19-42.

[43] Zhao, Z., J. Zhao, and H.N. Koutsopoulos, *Individual-Level Trip Detection using Sparse Call Detail Record Data based on Supervised Statistical Learning*. 2016.

[44] Caceres, N., J. Wideberg, and F. Benitez, *Deriving origin–destination data from a mobile phone network.* IET Intelligent Transport Systems, 2007. **1**(1): p. 15-26.

[45] Wang, P., et al., *Understanding road usage patterns in urban areas.* Scientific reports, 2012. **2**: p. 1001.

[46] Farrahi, K. and D. Gatica-Perez, *Discovering routines from large-scale human locations using probabilistic topic models.* ACM Transactions on Intelligent Systems and Technology (TIST), 2011. **2**(1): p. 3.

[47] Phithakkitnukoon, S., et al. *Activity-aware map: Identifying human daily activity pattern using mobile phone data*. in *International Workshop on Human Behavior Understanding*. 2010. Springer.

[48] Widhalm, P., et al., *Discovering urban activity patterns in cell phone data.* Transportation, 2015. **42**(4): p. 597-623.

[49] Yin, M., et al., *A generative model of urban activities from cellular data.* IEEE Transactions on Intelligent Transportation Systems, 2017.

[50] Schneider, C.M., et al., *Unravelling daily human mobility motifs.* Journal of The Royal Society Interface, 2013. **10**(84): p. 20130246.

[51] Jiang, S., et al. *A review of urban computing for mobile phone traces: current methods, challenges and opportunities*. in *Proceedings of the 2nd ACM SIGKDD international workshop on Urban Computing*. 2013. ACM.

[52] Jiang, S., J. Ferreira, and M.C. Gonzalez, *Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore.* IEEE Transactions on Big Data, 2017. **3**(2): p. 208-219.

[53] Zilske, M. and K. Nagel, *Studying the accuracy of demand generation from mobile phone trajectories with synthetic data.* Procedia Computer Science, 2014. **32**: p. 802-807.

[54] Zilske, M. and K. Nagel, *A simulation-based approach for constructing all-day travel chains from mobile phone data.* Procedia Computer Science, 2015. **52**: p. 468-475.

[55] Bassolas, A., et al., *Mobile phone records to feed activity-based travel demand models: MATSim for studying a cordon toll policy in Barcelona.* Transportation Research Part A: Policy and Practice, 2019. **121**: p. 56-74.

[56] Devillaine, F., M. Munizaga, and M. Trépanier, *Detection of activities of public transport users by analyzing smart card data.* Transportation Research Record: Journal of the Transportation Research Board, 2012(2276): p. 48-55.

[57] Ceapa, I., C. Smith, and L. Capra. *Avoiding the crowds: understanding tube station congestion patterns from trip data*. in *Proceedings of the ACM SIGKDD international workshop on urban computing*. 2012. ACM.

[58] da Silva, T.L.C., J.A. de Macêdo, and M.A. Casanova. *Discovering frequent mobility patterns on moving object data*. in *Proceedings of the Third ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems*. 2014. ACM.

[59] Liu, Y., et al., *Intelligent bus routing with heterogeneous human mobility patterns.* Knowledge and Information Systems, 2017. **50**(2): p. 383-415.

[60] Zhang, F., et al., *Spatiotemporal segmentation of metro trips using smart card data.* IEEE Transactions on Vehicular Technology, 2016. **65**(3): p. 1137-1149.

[61] Zhao, J., et al., *Estimation of passenger route choice pattern using smart card data for complex metro systems.* IEEE Transactions on Intelligent Transportation Systems, 2017. **18**(4): p. 790-801.

[62] Bouman, P., et al., *Recognizing demand patterns from smart card data for agent-based micro-simulation of public transport.* 2012.

[63] Chakirov, A. and A. Erath, *Activity identification and primary location modelling based on smart card payment data for public transport.* [Working paper/Transport and Spatial Planning], 2012. **786**.

[64] Fourie, P.J., et al., *Using smartcard data for agent-based transport simulation*, in *Public Transport Planning with Smart Card Data*. 2016, CRC Press. p. 133-160.

[65] Zhu, Y., *Extract the Spatiotemporal Distribution of Transit Trips from Smart Card Transaction Data: A Comparison Between Shanghai and Singapore*, in *Big Data Support of Urban Planning and Management*. 2018, Springer. p. 297-315.

[66] Patterson, D.J., et al. *Inferring high-level behavior from low-level sensors*. in *International Conference on Ubiquitous Computing*. 2003. Springer.

[67] Schuessler, N. and K. Axhausen, *Processing raw data from global positioning systems without additional*

information. Transportation Research Record: Journal of the Transportation Research Board, 2009(2105): p. 28-36.

[68] Rudloff, C. and M. Ray, *Detecting travel modes and profiling commuter habits solely based on GPS data*. 2010.

[69] McGowen, P. and M. McNally. *Evaluating the potential to predict activity types from GPS and GIS data*. in *Transportation Research Board 86th Annual Meeting, Washington*. 2007. Citeseer.

[70] Demissie, M.G., et al., *Inferring passenger travel demand to improve urban mobility in developing countries using cell phone data: a case study of Senegal.* IEEE Transactions on Intelligent Transportation Systems, 2016. **17**(9): p. 2466-2478.

[71] Ndibatya, I., J. Coetzee, and T. Booysen, *Mapping the informal public transport network in Kampala with smartphones: international.* Civil Engineering = Siviele Ingenieurswese, 2017. **2017**(v25i1): p. 35-40.

[72] Saddier, S., et al., *Mapping the Jitney network with smartphones in Accra, Ghana: the AccraMobile experiment.* Transportation Research Record: Journal of the Transportation Research Board, 2016(2581): p. 113-122.

[73] Bisseling, R., et al., *Inferring transportation modes from smartphone sensors.* Proceedings 106th European Study Group Mathematics with Industry, 2016: p. 21-34.

[74] Cheng, Z., et al., *Exploring millions of footprints in location sharing services.* ICWSM, 2011. **2011**: p. 81-88.

[75] Rashidi, T.H., et al., *Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges.* Transportation Research Part C: Emerging Technologies, 2017. **75**: p. 197-211.

[76] Golder, S.A. and M.W. Macy, *Digital footprints: Opportunities and challenges for online social research*. Annual Review of Sociology, 2014. **40**.

[77] Cramer, H., M. Rost, and L.E. Holmquist. *Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare*. in *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*. 2011. ACM.

[78] Maghrebi, M., et al. *Complementing travel diary surveys with Twitter data: application of text mining techniques on activity location, type and time*. in *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*. 2015. IEEE.

[79] Zhu, Z., U. Blanke, and G. Tröster. *Inferring travel purpose from crowd-augmented human mobility data*. in *Proceedings of the First International Conference on IoT in Urban Space*. 2014. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[80] Hasan, S. and S.V. Ukkusuri, *Social contagion process in informal warning networks to understand evacuation timing behavior.* Journal of Public Health Management and Practice, 2013. **19**: p. S68-S69.

[81] Noulas, A., et al., *A tale of many cities: universal patterns in human urban mobility.* PloS one, 2012. **7**(5): p. e37027.

[82] Jurdak, R., et al., *Understanding human mobility from Twitter.* PloS one, 2015. **10**(7): p. e0131469.

[83] Cebelak, M.K., *Location-based social networking data: doubly-constrained gravity model origin-destination estimation of the urban travel demand for Austin, TX.* 2013.

[84] Lee, J.H., et al., *Activity space estimation with longitudinal observations*

of social media data. Transportation, 2016. **43**(6): p. 955-977.

[85] Lee, J.H., S. Gao, and K.G. Goulias. *Can Twitter data be used to validate travel demand models*. in *14th International Conference on Travel Behaviour Research*. 2015.

[86] Hasnat, M.M., et al., *Destination choice modeling using location-based social media data.* Journal of choice modelling, 2019. **31**: p. 22-34.

[87] Lian, D. and X. Xie. *Collaborative activity recognition via check-in history*. in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*. 2011. ACM.

[88] Pianese, F., et al. *Discovering and predicting user routines by differential analysis of social network traces*. in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a*. 2013. IEEE.

[89] Coffey, C. and A. Pozdnoukhov. *Temporal decomposition and semantic enrichment of mobility flows*. in *Proceedings of the 6th ACM SIGSPATIAL International Workshop on Location-Based Social Networks*. 2013. ACM.

[90] Hasan, S. and S.V. Ukkusuri, *Location contexts of user check-ins to model urban geo life-style patterns.* PloS one, 2015. **10**(5): p. e0124819.

[91] Medina, S.A.O., *Inferring weekly primary activity patterns using public transport smart card data and a household travel survey.* Travel Behaviour and Society, 2016.

[92] Zhang, A., et al. *Multi-day activity-travel pattern sampling based on single-day data*. in *97th Annual Meeting of the Transportation Research Board (TRB 2018)*. 2018. TRB Annual Meeting.

[93] Horni, A. and K.W. Axhausen, *MATSim Agent Heterogeneity and a One-Week Scenario.* ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme, 2012.

[94] Ordóñez Medina, S.A., A. Erath, and K.W. Axhausen. *Simulating Urban Transport for a Week Time Horizon Using MATSim*. in *3rd Workshop on Time Use Observatory (TUO 3)*. 2012. Complex Engineering Systems Institute (ISCI).

[95] Ordóñez Medina, S.A. *Recognizing personalized flexible activity patterns*. in *14th International Conference on Travel Behavior Research (IATBR 2015)*. 2015. IVT, ETH Zurich.

[96] Lee, J.H., *Travel Behavior Dynamics in Space and Time*. 2016, University of California, Santa Barbara.

[97] Balac, M., M. Janzen, and K.W. Axhausen. *Alternative Approach to Scoring in MATSim and how it affects Activity Rescheduling*. in *97th Annual Meeting of the Transportation Research Board (TRB 2018)*. 2018. TRB Annual Meeting.

[98] Koppelman, F.S. and C.G. Wilmot, *Transferability analysis of disaggregate choice models.* Transportation Research Record, 1982. **895**: p. 18-24.

[99] Yasmin, F., C. Morency, and M.J. Roorda, *Assessment of spatial transferability of an activity-based model, TASHA.* Transportation Research Part A: Policy and Practice, 2015. **78**: p. 200-213.

[100] Sikder, S., et al., *Spatial transferability of travel forecasting models: a review and synthesis.* International Journal of Advances in Engineering Sciences and Applied Mathematics, 2013. **5**(2-3): p. 104-128.

[101] Arentze, T., et al., *Spatial transferability of the Albatross model*

system: Empirical evidence from two case studies. Transportation Research Record: Journal of the Transportation Research Board, 2002(1805): p. 1-7.

[102] Inc, P.C., *Lake Tahoe Resident and Visitor Model: Model Description and Final Results.* 2007.

[103] Bowman, J.L., et al. *Making advanced travel forecasting models affordable through model transferability*. in *93rd Transportation Research Board Annual Meeting, Washington DC, USA*. 2014.

[104] Goulias, K.G., et al. *Simulator of activities, greenhouse emissions, networks, and travel (SimAGENT) in Southern California*. in *91st annual meeting of the Transportation Research Board, Washington, DC*. 2012.

[105] Sivakumar, A., S. Le Vine, and J. Polak. *An activity-based travel demand model for London*. in *European Transport Conference, 2010Association for European Transport*. 2010.

[106] Yasmin, F., C. Morency, and M.J. Roorda, *Macro-, meso-, and micro-level validation of an activity-based travel demand model.* Transportmetrica A: Transport Science, 2017. **13**(3): p. 222-249.

[107] Ziemke, D., K. Nagel, and C. Bhat, *Integrating CEMDAP and MATSim to increase the transferability of transport demand models.* Transportation Research Record: Journal of the Transportation Research Board, 2015(2493): p. 117-125.

[108] Bowman, J.L. and M. Bradley, *Testing Spatial Transferability of Activity-Based Travel Forecasting Models.* Transportation Research Record: Journal of the Transportation Research Board, 2017(2669): p. 62-71.

[109] Linha, H.T., et al., *Exploring the Transferability of FEATHERS–An Activity Based Travel Demand Model–For Ho Chi Minh City, Vietnam.* Journal of Traffic and Transportation Management, 2019. **1**(2): p. 01-09.

[110] Shiftan, Y., *The advantage of activity-based modelling for air-quality purposes: theory vs practice and future needs.* Innovation: The European Journal of Social Science Research, 2000. **13**(1): p. 95-110.

[111] Shiftan, Y. and J. Suhrbier, *The analysis of travel and emission impacts of travel demand management strategies using activity-based models.* Transportation, 2002. **29**(2): p. 145-168.

[112] Beckx, C., et al., *An integrated activity-based modelling framework to assess vehicle emissions: approach and application.* Environment and Planning B: Planning and Design, 2009. **36**(6): p. 1086-1102.

[113] Dons, E., et al., *Using an activity-based framework to determine effects of a policy measure on population exposure to nitrogen dioxide.* Transportation Research Record: Journal of the Transportation Research Board, 2011(2233): p. 72-79.

[114] Pebesma, E., et al. *Uncertainty in exposure to air pollution*. in *EGU General Assembly Conference Abstracts*. 2013.

[115] Hatzopoulou, M., E. Miller, and B. Santos, *Integrating vehicle emission modeling with activity-based travel demand modeling: case study of the Greater Toronto, Canada, Area.* Transportation Research Record: Journal of the Transportation Research Board, 2007(2011): p. 29-39.

[116] Hao, J., M. Hatzopoulou, and E. Miller, *Integrating an activity-based travel demand model with dynamic traffic assignment and emission models: Implementation in the Greater Toronto, Canada, area.* Transportation Research Record: Journal of the Transportation Research Board, 2010(2176): p. 1-13.

[117] Hatzopoulou, M., J.Y. Hao, and E.J. Miller, *Simulating the impacts of household travel on greenhouse gas emissions, urban air quality, and population exposure.* Transportation, 2011. **38**(6): p. 871.

[118] Vallamsundar, S., et al., *Maternal Exposure to Traffic-Related Air Pollution Across Different Microenvironments.* Journal of Transport & Health, 2016. **3**(2): p. S72.

[119] Shabanpour, R., et al., *Investigating the applicability of ADAPTS activity-based model in air quality analysis.* Travel Behaviour and Society, 2017.

[120] Wu, Y. and G. Song, *The Impact of Activity-Based Mobility Pattern on Assessing Fine-Grained Traffic-Induced Air Pollution Exposure.* International journal of environmental research and public health, 2019. **16**(18): p. 3291.

[121] Shliselberg, R. and M. Givoni, *Motility as a policy objective.* Transport reviews, 2018. **38**(3): p. 279-297.

[122] Abou-Zeid, M. and M. Ben-Akiva, *Well-being and activity-based models.* Transportation, 2012. **39**(6): p. 1189-1207.

[123] Stanley, J.K., et al., *Mobility, social exclusion and well-being: Exploring the links.* Transportation research part A: policy and practice, 2011. **45**(8): p. 789-801.

[124] Hensher, D.A., *Measurement of the valuation of travel time savings.* Journal of Transport Economics and Policy (JTEP), 2001. **35**(1): p. 71-98.

[125] Metz, D., *The myth of travel time saving.* Transport reviews, 2008. **28**(3): p. 321-336.

[126] El-Geneidy, A.M. and D.M. Levinson, *Access to destinations: Development of accessibility measures.* 2006.

[127] de Lima, I.V., et al., *Dynamic Modeling of Activity Happiness: An Investigation of the Intra-activity Hedonic Treadmill*, in *Quality of Life and Daily Travel*. 2018, Springer. p. 95-118.

[128] Guo, W., Y. Chai, and M.-P. Kwan, *Travel-related exposure to air pollution and its socio-environmental inequalities: Evidence from a week-long GPS-based travel diary dataset*, in *Spatiotemporal Analysis of Air Pollution and Its Application in Public Health*. 2020, Elsevier. p. 293-309.

[129] Castiglione, J., et al. *Building an Integrated Activity-Based and Dynamic Network Assignment Model*. in *3rd Transportation Research Board Conference on Innovations in Travel Modeling, Tempe, Ariz*. 2010.

[130] Bekhor, S., C. Dobler, and K. Axhausen, *Integration of activity-based and agent-based models: case of Tel Aviv, Israel.* Transportation Research Record: Journal of the Transportation Research Board, 2011(2255): p. 38-47.

[131] Lin, D.-Y., et al., *Evacuation planning using the integrated system of activity-based modeling and dynamic traffic assignment.* Transportation Research Record: Journal of the Transportation Research Board, 2009(2132): p. 69-77.

[132] Liu, P., et al., *Dynamic activity-travel assignment in multi-state supernetworks.* Transportation Research Part B: Methodological, 2015. **81**: p. 656-671.

[133] Pendyala, R.M., et al., *Integrated Land Use–Transport Model System with Dynamic Time-Dependent Activity–Travel Microsimulation.* Transportation Research Record, 2012. **2303**(1): p. 19-27.

[134] Pendyala, R.M., et al., *Paradigms for integrated modeling of activity-travel*

demand and network dynamics in an era of dynamic mobility management. 2017.

[135] Zockaie, A., et al., *Activity-Based Model with Dynamic Traffic Assignment and Consideration of Heterogeneous User Preferences and Reliability Valuation: Application to Toll Revenue Forecasting in Chicago, Illinois.* Transportation Research Record: Journal of the Transportation Research Board, 2015(2493): p. 78-87.

[136] Rieser, M., et al., *Agent-oriented coupling of activity-based demand generation with multiagent traffic simulation.* Transportation Research Record, 2007. **2021**(1): p. 10-17.

[137] Hao, J.Y., M. Hatzopoulou, and E.J. Miller, *Integrating an activity-based travel demand model with dynamic traffic assignment and emission models: Implementation in the Greater Toronto, Canada, area.* Transportation Research Record, 2010. **2176**(1): p. 1-13.

[138] Lin, D.-Y., et al., *Integration of activity-based modeling and dynamic traffic assignment.* Transportation Research Record, 2008. **2076**(1): p. 52-61.

[139] Hadi, M., et al., *Dynamic, Integrated Model System: Jacksonville-Area Application*. 2014.

[140] Javanmardi, M., J. Auld, and K. Mohammadian. *Integration of TRANSIMS with the ADAPTS Activity-based Model*. in *Fourth TRB Conference on Innovations in Travel Modeling (ITM), Tampa, FL*. 2011.

[141] Zockaie, A., et al., *Activity-based model with dynamic traffic assignment and consideration of heterogeneous user preferences and reliability valuation: application to toll revenue forecasting in Chicago, Illinois.* Transportation Research Record, 2015. **2493**(1): p. 78-87.

[142] Vovsha, P., et al. *Integrated model of travel demand and network simulation*. in *Proceedings of the 6th Conference on Innovations in Travel Modeling (ITM), TRB, Denver, CO*. 2016.

[143] Cipriani, E., et al., *Integration between activity-based demand models and multimodal assignment: some empirical evidences.* Case Studies on Transport Policy, 2020.

[144] Adnan, M., et al., *Integrated agent-based microsimulation framework for examining impacts of mobility-oriented policies.* Personal and Ubiquitous Computing, 2020: p. 1-13.

[145] Baghestani, A., et al., *Evaluating the Traffic and Emissions Impacts of Congestion Pricing in New York City.* Sustainability, 2020. **12**(9): p. 3655.

[146] Cepolina, E.M. and A. Farina, *A new shared vehicle system for urban areas.* Transportation Research Part C: Emerging Technologies, 2012. **21**(1): p. 230-243.

[147] Ciari, F., M. Balac, and K.W. Axhausen, *Modeling Carsharing with the Agent-Based Simulation MATSim: State of the Art, Applications, and Future Developments.* Transportation Research Record: Journal of the Transportation Research Board, 2016(2564): p. 14-20.

[148] Balac, M., F. Ciari, and K.W. Axhausen, *Modeling the impact of parking price policy on free-floating carsharing: Case study for Zurich, Switzerland.* Transportation Research Part C: Emerging Technologies, 2017. **77**: p. 207-225.

[149] Heilig, M., et al., *Implementation of free-floating and station-based carsharing in an agent-based travel demand model.* Travel Behaviour and Society, 2018. **12**: p. 151-158.

[150] Li, Q., et al., *Incorporating free-floating car-sharing into an activity-based dynamic user equilibrium model: A demand-side model.* Transportation

Research Part B: Methodological, 2018. **107**: p. 102-123.

[151] Choo, S., P.L. Mokhtarian, and I. Salomon, *Does telecommuting reduce vehicle-miles traveled? An aggregate time series analysis for the US.* Transportation, 2005. **32**(1): p. 37-64.

[152] Kim, S.-N., *Is telecommuting sustainable? An alternative approach to estimating the impact of home-based telecommuting on household travel.* International Journal of Sustainable Transportation, 2017. **11**(2): p. 72-85.

[153] Zhu, P. and S.G. Mason, *The impact of telecommuting on personal vehicle usage and environmental sustainability.* International Journal of Environmental Science and Technology, 2014. **11**(8): p. 2185-2200.

[154] Paleti, R. and I. Vukovic, *Telecommuting and Its Impact on Activity–Time Use Patterns of Dual-Earner Households.* Transportation Research Record: Journal of the Transportation Research Board, 2017(2658): p. 17-25.

[155] Helminen, V. and M. Ristimäki, *Relationships between commuting distance, frequency and telework in Finland.* Journal of Transport Geography, 2007. **15**(5): p. 331-342.

[156] Pouri, Y. and C. Bhat, *On Modeling Choice and Frequency of Home-Based Telecommunting.* Transportation Research Record: Journal of the Transportation Research Board, 2003(1858): p. 55-60.

[157] Sener, I.N. and C.R. Bhat, *A copula-based sample selection model of telecommuting choice and frequency.* Environment and Planning A, 2011. **43**(1): p. 126-145.

[158] Shabanpour, R., et al., *Analysis of telecommuting behavior and impacts on travel demand and the environment.* Transportation Research Part D: Transport and Environment, 2018. **62**: p. 563-576.

[159] Pirdavani, A., et al., *Traffic Safety Implications of Travel Demand Management Policies: The Cases of Teleworking and Fuel Cost Increase*, in *Transportation Systems and Engineering: Concepts, Methodologies, Tools, and Applications*. 2015, IGI Global. p. 1082-1107.

[160] Zhang, L., et al., *Integrating an agent-based travel behavior model with large-scale microscopic traffic simulation for corridor-level and subarea transportation operations and planning applications.* Journal of Urban Planning and Development, 2013. **139**(2): p. 94-103.

[161] Yang, D., *INTEGRATING ACTIVITY-BASED TRAVEL DEMAND AND DYNAMIC TRAFFIC ASSIGNMENT MODEL: A BEHAVIORAL USER EQUILIBRIUM APPROACH*. 2018.

[162] Frank, L.D., et al., *Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality.* Journal of the American planning Association, 2006. **72**(1): p. 75-87.

Chapter 3

Attitudes and Behaviours in Relation to New Technology in Transport and the Take-Up amongst Older Travellers

Joan Harvey

Abstract

Numbers of older people are increasing and this will continue for several decades to come. With that, there are changes as we age that can affect or impact upon our travelling and transportation needs and behaviour. In addition, there is an almost universal problem that many of all ages people have low levels of computer literacy. Transport may well look very different in the future. Not only automated vehicles, but also new transportation systems, such as Mobility as a Service [MaaS] and the likely developments in public transport that incorporate real time travel information, facilities and ease of use information all mean that older people wishing to travel will necessarily have to engage with some forms of new technology. The new systems will need to be personalisable to individual travellers. This chapter considers the needs of older travellers and how new technology can meet some of those needs and what is necessary for it to be appropriate to, and usable by, older travellers.

Keywords: new technology, older people, travelling, attitudes, behaviour

1. Introduction

This chapter is in three main sections, plus a conclusion at the end.

The first section covers what happens as we get older, in terms of abilities, skills cognition, psychological and social changes and changes in technological 'savvy' or awareness.

The second section looks at how changes in transport systems can be made much more useful and usable for older people; in addition the vision of the future that is MaaS in the UK is considered, especially for the older and more infirm category of older people.

Finally there are a number of considerations [such as ergonomic ones] in terms of the older traveller using new technology to aid their travel, and what the requirements are for that to work as well as possible.

2. What happens as we age?

2.1 Getting older- what changes cognitively?

Being 'older' used to apply to people aged over 50, but in modern times it largely refers to people who are at least past retirement age, which would be between 65 and 70 in most countries, right up to 100+. It is possible now to find large numbers of people aged over 80 who are highly active. However increasing longevity brings not only more fit older people, it also means more older people with limitations and disabilities and more people generally with low levels of computer literacy [1].

Some things do deteriorate with age, including many cognitive functions, for example memory and retention, but also some skills such as navigation and situational awareness. Response times, such as reaction times in the event of an emergency also slow with age, even more so when there are multiple demands on attention and attention-switching and/or distractors (for example see [2]).

2.1.1 Cognitive age-related declines and gains

Between the ages 20–80, there is a decade-by-decade reduction in processing speeds, working memory, cued and free recall: these are real reductions in every decade, although a steeper decline between 70 and 80 [3].

There are also age-related gains:

- habituated skills and sustained attention, past experience allowing better anticipation,

- increased vocabulary and knowledge,

- recognition and other crystallised abilities that rely on culture-related lifelong learning and these increase throughout adulthood and are preserved in healthy ageing; gains may also relate to older adults adopting age-counteractive measures to compensate for losses. The evidence for functional reorganization [the plasticity of the brain] and compensation along with effective interventions does hold some promise for a more optimistic view of neurocognitive status in later life [4–6].

2.1.2 Distracters and slower processing

There is ample research showing the distraction of cellphones causing a deterioration in driving performance in terms of reduced ability to react and respond; whilst this occurs at all ages, it can be a particular problem for older people because it involves switching attention [7]. We also know that visual field declines with age but this is not universal by any means [8]: we cannot assume that all older people necessarily have poor eyesight, and transport or other policies reflecting this that restrict driving privileges for older people have no scientific foundation- in other words, restrict drivers if you wish by testing visual field, but do not do it by age. Decline in situation awareness relates to shrinkage in the field of view but not to cognitive decline: these constitute an issue in perception of travel-related information but mean that training to improve situation awareness may have some real value.

The findings of slower processing, working memory and attention-switching declines, are clear but there is huge variability for all of these: reasons might include health status and fitness being huge positive indicators for self-efficacy and achievement for older people whereas stress levels have negative impacts etc.. Not only that, but pathological age-related changes (such as Alzheimers') can be undiagnosed for up to 10 years, so not only might abilities decline with age, but also there will be declines associated with diseases of which the individual is unaware; for example perceived difficulty in using everyday technology increases in people with mild cognitive impairment [but may be attributed simply to just 'getting older'] and accentuates in mild-stage dementia [9, 10].

Older people displaying lower levels of cognitive skills may actually be due to mechanisms that were present earlier in life that generate *life-long* differences rather than due to ageing, for example due to less-enriched environments; indeed, research on the brain's plasticity implies that changes continue throughout life and thus the option to enrich the environment to facilitate positive changes at any age point presents a distinct possibility [11], such as improving 'situation awareness' by using driving simulators.

From this, an unwillingness to engage with new technology may be a life-long issue but at the same time it is not too late to change, albeit with much training and support. In terms of transport and developments of travel 'apps', many older people may have relatively little experience of journeys, some may have travelled only by plane to a holiday resort and little else by way of organized or unorganized travel (see [12] showing for example lower bus use than planes) and so an 'app' to help with trip-planning may make little or no sense to them.

2.1.3 Other psychological mechanisms that decline

Other psychological mechanisms that decline with age include situation awareness as mentioned above, navigating skills, episodic and autobiographical memory, etc. [13–15]. Losses of episodic memory involve the link between an image [venue] and its name being lost, but better signage linking a name to a picture can help in this situation, and could be adapted to be possible with wearable devices, which could also help those with early dementia be able to travel without worrying about knowing where they need to be. There are many 'visual' or 'conceptual' maps that show imagery linked to names; for example as early as 1968, Fisher developed a 'conceptual map' of Newcastle upon Tyne, using image-name links plus perspective to enable travellers to negotiate the city centre [16]. Sustrans maps showing cycle routes utilize a similar approach. To be amenable to older travellers with memory loss, this approach could be developed for a touch screen with increasing levels of detail.

2.1.4 Age-impaired task performance

Performance of tasks can be impaired or counteracted, enhanced or neutrally effected by age [17]. Older people may learn how to counteract any inability to perform a task or their experience or knowledge might actually enhance what they are doing- examples include driving different routes to avoid difficult situations, driving more slowly to compensate for perceived slower response times, or driving in daylight only, or to avoid glare, or using familiar routes [18, 19]. Age-impaired tasks are not only those that rely on complex switching of attention or speed of response, they may also be impaired by high levels of emotion or stress [such as frustration of being late], both of which directly affect memory. An age-impaired task may be one at a forced pace, whereas an age-counteracted one would be where the older person, aware they may be slower, works at their own pace. Inhibitory responses are less efficient with age so it becomes more difficult to access relevant information and delete old information from our processing; again, many people learn to compensate for this, for example by keeping to simple or more habituated tasks that require fewer or lower inhibitory responses and which appear to be unaffected by ageing [20, 21]. Increased amounts of or new travel-relevant information would mean that processing may be problematic if it interferes with the existing information travellers hold in their memories about the trip (see for example [22–24]); additional information could only be of value in this situation if precisely targeted at the person and the specific journey, and not adding too much new information to process. However older people are very able to maintain sustained attention, so tasks that require this but not divided attention may be relatively age-neutral, and in any event even divided attention tasks may be improved with practice and training in older people [25, 26].

Age-related changes may disadvantage older people in an increasingly screen and button-based world [23]. The over 70s in particular exhibit difficulties with touch screen interfaces and the navigation logic of applications [27]. Too much information can present anybody with a processing dilemma, particularly older people; for example over-complex display systems of travel information, either at sites or on mobile apps, may present them with something they struggle to deal with and thus avoid. So there is the problem that any declines in cognitive processing could lead to reduced or no use of new technology, which in turn leads to exclusion of older people. There are many recommendations that can alleviate or at least mitigate such issues, covered later in this chapter.

2.1.5 Summary of cognitive changes with age

In summary, older people can be slower to respond, can have reductions in working memory capacity and computing span, problems switching attention, decline in visual field; episodic and autobiographical memories decrease most and aids for retrieval are needed; there is difficulty in moving onto new topics as inhibitory mechanisms cling to previous topic; less attention focus especially when tasks increase in number; situation awareness is worse with age; slower navigation skills. Memory is also negatively affected by emotion or stress but is improved by enriched environments. There are real reductions in every decade, although a steeper decline between 70 and 80; in contrast, vocabulary and knowledge-related measures rise slightly with age right up to 70 and then level off, and age makes no difference to recognition [3]. The plasticity of the brain, functional reorganization and compensatory increases in frontal lobe activation mean that the brain can change throughout life and there is evidence that these can be encouraged by interventions so many of these changes can be influenced for the better. There are problem of early undiagnosed pathological ageing diseases, affect working memory. Restricting insurance policies etc. should be done on the basis of testing visual field but not by age.

2.2 What else changes?

The psychological factors and traits that decline/change with age include risk taking, risk awareness, motivation, personality and resistance to change.

2.2.1 Personality

In terms of personality, there is no evidence of changes beyond the age of 30, but it must be said that the age period 16–30 is one of change in most psychological factors and indeed is associated with changes in the brain and hormones. A relevant personality trait in this respect is openness to change, which is likely to be normally distributed and in terms of openness resistance to change is a default position in terms of evolutionary psychology (see for example [28]); and there is evidence of increasing resistance to change with age, although that needs to be unpicked in order to understand it: for example it may be associated with low-involvement whilst high-involvement may be associated with changing attitudes; this is important if we are looking at technology engagement with which older people may exhibit low involvement. In addition, attitudes changed can just the same change back again; so positive changes in attitudes to technology will need to be reinforced if they are not to revert to earlier attitudes [29].

2.2.2 Attitudes and change

More negative attitudes towards computers by older people are related to perception of less comfort, efficacy and control, all of which have been shown to be improved by increased experience [30]. In addition, the literature on change and change management points to low trust, perceived lack of competence, poor communications, not understanding the need for change, exhaustion/saturation and changing the status quo [away from habituated behaviours] as all being culpable. When people are asked to articulate their reasons for resistance, risks outweigh the benefits, they do not have the ability to change, perceive that things will be made more difficult, not meeting their needs and so on are all cited [31]. Many of these issues are relevant to the unwillingness of older people to engage with new technology.

2.2.3 Unwillingness to reduce driving and the role of affect

As already mentioned, older people when driving often engage in age-counteracted behaviours [self-regulation] and are thus less likely to take risks and avoid difficult driving or travelling situations. Less peer pressure and increased self-awareness may influence age-counteracted and compensatory behaviours for older people view limitations [8, 32–34]. Many older drivers will be unwilling to reduce their driving because of the increased inconvenience, loss of social activities and lack of suitable alternative transport modes, loss of independence and increase in social isolation etc. [35–39]. The relationship between many people and their car is not so much an *economic* one as an *affective* one and they make choices about travel mode using the *affect heuristic* [40]; this means that comfort, convenience, feelings of risk and security, etc. all play the larger part in the decision. Older people may be more time-rich and potentially more money-poor than they were when working, but decisions on transport mode are still likely to be affect-driven, and poor public transport provision must surely exacerbate that situation (e.g. [41]).

2.2.4 Risk taking and risk awareness

Risk taking and risk awareness can be critical for all travellers. The reasons for accidents differ considerably between older and younger people and most insurance premia follow a U-shape with age, with by far the highest premia at <25 yrs., dropping to age 30 where they remain low until the late 60s. Accidents and injuries are caused by different elements at different ages: selecting and processing information in a complex task may be causative for older drivers accidents, whereas overestimation of personal skill, sensation-seeking and a preference for risky driving, reaction times and ability are causative for younger drivers [32, 37, 42, 43].

2.2.5 Some health issues can improve

Ageing also has some basic physiological aspects: it is related to changes in the cardiovascular and cardiopulmonary systems; however deficits can be reduced by training, practice, aerobic exercise and these also may improve the efficiency of neural processes, which we can see in healthier older adults' increased take-up of technology [44–46].

2.2.6 Summary of non-cognitive changes

Therefore we can say that increasing age is associated with more positive attitudes and emotions and an optimistic bias; self-awareness of functional decline leads to many age-counteracted or self-regulatory behaviours; resistance to change is the default position and slowly increases through adulthood; older people have a lower self-assessment of skills and abilities; increased lack of confidence, fear and anxiety with regard to new technologies, perceived less comfort with, efficacy or control over computers, dehumanisation all affect motivation to engage with new technology. Attitudes change [to new technology] requires reinforcement otherwise it will decay. Self-efficacy is predictive of better health in older people- relevant to adopting new technology.

2.3 How do people learn new technologies or resist becoming "tech-savvy"

2.3.1 Age related factors

Learning of new technologies is a complex challenge for many older adults if it does not suit individual capabilities; reasons include perceived loss of control, lack of confidence, not seeing the need, wanting to retain the status quo and so on, many articulated in this chapter already [47–51]. Reasons for adoption of new technology *de facto* follow the reverse of these, and would include feeling confident and in control, perceiving a current need and recognising a future need plus past relevant experience [52, 53].

Since much information technology only began to become commercially available in the mid-1980s, so younger people have a 'head start' and often tend to be good at digital technologies; they invariably use them at school or at work in the form of computers, laptops or smartphones, and even internet access in households with an adult aged 65+ have now risen to 80% in 2020, and online shopping has increased massively especially in 2020 [54]. Those older adults whose occupations involved computer use, e.g. engineers [52, 53] are more likely to perceive the need for and be able to learn and use new technologies with confidence.

However the evidence is that most people, across all ages, are a long way from being tech-savvy, and that also includes a large proportion of older people; as technology develops, it may always be ahead of older people, although if the development rate levels off this may be less of an issue in years to come.

2.3.2 The digital divide

An international study by OECD [1] attempted to quantify the differences between the broad population and the technology elite: data collected from over 200,000 people aged 16–65 in 33 countries yielded four levels of technology proficiency. The findings suggest that over 65% of UK adult population are at Level 1 (can do tasks typically requiring the use of widely available and technologies, applications such as email software or a web browser") or below. The OECD average for level 0 or 1 is 69% with Japan 66%, USA 67% and New Zealand 56%. So there is a clear *technology- divide across all ages*, which can be proposed to be larger than any *age-divide*.

Therefore for designers to target a broad consumer audience at any age, they must:

- keep it extremely simple, meaning little/no navigation required to access information or commands required to solve a problem;

- have few steps or operations, few monitoring demands;

- allow identification of content and operators through simple match

- include no need to contrast or integrate information.

More importantly for our considering older people here, the OECD study did not include people aged over 65, where the percentages of being Level 1 or below should be considerably higher. It is clear that there is a responsibility- and benefit- for designers of current and future ICT to make it accessible, affordable, anxiety-free and helpful so older adults can use it [55]. Most ergonomists and gerontologists would agree that designing for the disable user would be similarly useful to all.

Page [50] suggested that whilst older users may show a keen interest in learning and using technology, they often do not feel fully equipped to do so. Motivation to learn may also be a function of utility; this means that over-complexity may present older users with a problem beyond what they can manage [29].

One key area for non-engagement with technology relates to user confidence in own abilities, fear and anxiety [53, 56]. Technology has been shown to be a source of anxiety amongst older users, for example concerning loss of privacy, lack of confidence, a perceived lack of need and an unwillingness to learn through trial and error [55, 57, 58].

Working with computers, tablets or smartphones inherently requires working memory and fluid intelligence, the ability to reason and solve new problems independently of previously acquired knowledge, which are also predictive of each other, along with attention-switching (e.g. [59]). Whilst training on working memory can improve general fluid intelligence, the effect is dosage-dependent, so the more training, the greater the improvement. The implications of this for training older people in new technology are therefore considerable (e.g. [60, 61]).

Instructions and manuals are often not clear, difficult to follow, and need to be improved; there is some research on this already, but more is needed. Examples of where this has been done to very good effect, in the commercial and Government areas, is the work on instructional text of James Hartley (e.g. [62]); examples of this approach can be seen in the design of entirely visual passenger safety information in airplanes and some instruction formats. Too much instructional information in a manual or Web Page, or too much hierarchical and negative searching can lead to cognitive load issues [63, 64].

2.3.3 Summary of section

In summary, *there is and will continue to be a digital divide in tech-savviness-* albeit not total illiteracy on the lower side but a divide nevertheless, which extends currently down to 16–25 year olds. Older people may be less confident and poorly motivated to take on new technology, but there is a strong role for additional training to help with working memory and fluid intelligence and a need to develop good instructional material to support this. Personalised and self-regulated learning should be available for all new technology.

3. How can transport system changes benefit older people?

The UK Government has published details of a MaaS approach [65], which would allow for personalized transport in some ways; this would include ticketing and many other services through single systems – early examples include smartcards that can allow the user to go from mode to mode without any reticketing. In the longer run, it would move towards a more bespoke service that can benefit the less mobile by allowing transport to be summonsed to precise destinations for personalized journeys. This is one example of new technology working to enhance and improve, and for the less mobile actually enable travel to be undertaken.

In the UK Government report on engagement with new technology by older travellers [66], ideas were proposed that related to a smart user interface that could enable travelling. However, several researchers have commented that older people have a long list of needs that would be important for their travelling, for example the availability of rest places, toilet facilities and ease of access and egress at all points. Whilst Transport for London [TfL] has gone a long way in enabling wheelchairs on buses and other travel places, this is clearly not a completed exercise as for example many tube stations remain wheelchair-unfriendly and information about when the next bus is arriving is not sufficient to enable trouble-free travelling; nor with much lower levels of per capita spending on transport outside of London, are things looking as good.

The report [66] also proposed that big data and information being provided by users can work towards the data base needed so that the full panoply of needed information is accurate, precise to what is needed and provided in real time for older and less physically able travellers. As an example, you as a disabled might want to go on a journey to eat a meal at a restaurant with your friends, calling on the way back to collect some items from [say] the chemist. The information you will need includes the transport availability, estimated arrival time, whether or not you will be able to board, the availability of facilities at the destination stop, the availability of ramps, or stairs, or seating, whether under cover, etc. at the destination. Then all the same information to get to the next destination, the chemist, then more information again about the homeward part of the journey. You also need to be fairly certain that the service will actually be running with no cancellations, no temporary movement of stops, for all three journeys. In the world of totally personalisable transport, you would be able to book the whole journey using a personalized vehicle from start to destination to destination to home, and do all this booking by touchscreen.

So the question is- where are we in relation to the ideal world of reliable transport, full information provided in real time, integration of services, etc. And further, can all this be achieved using a full blown Maas? And if so, how will this work in rural and small-town environments as opposed to large cities, particularly London, where there is already some integration?

It has been argued by many authors that this requires a political will, some legislation, and planning and infrastructure changes to get there. The recent massive changes to cycling infrastructure following from making transport more Covid-compliant has actually, at the time of writing, moved people out of buses and into cars and some onto cycles. The Cycle superhighways in London are certainly reducing road space for other vehicles; the congestion charging there has also changed how traffic is operating. However these are somewhat piecemeal and need to be even more integrated, which is possible of course, but so far at a much larger expenditure rate [in London] than for any other region or city in the country. We still see bus stops moved temporarily, and other issues that make a journey for a disabled person much more hazardous and difficult.

The UK government has also reported on what a future might be like for more active modes of travel, including walking of course but also electric bikes, scooters, and e-boards. The possibility for electric bikes and scooters is that they will bring more, and older, women into travelling and possibly leaving their cars behind especially for shorter journeys. The evidence is that shorter journeys dominate car use in cities and towns in the UK and elsewhere, and ebikes that can take cargo offer the possibility of local shopping in a more environmentally friendly, healthier and more sustainable way; however these also require some infrastructure changes to ensure they are as safe as possible, for example cycling highways and safe and secure parking (as things such as batteries are valuable) [67]. Ebikes are also likely to require legislation since at the moment they can go quite fast with an unlicensed driver, such as a 70 year old woman who does not hold any driving license can buy and drive one of these.

The problem is however more complicated, as many of these solutions suit shorter journeys, especially in better weather conditions and necessitate their own infrastructure. On top of that, there is how we address longer journeys- so city to city, from 20 to 30 right up to 500–600 miles. In addition to this, there is the need for MaaS for the older traveller who cannot drive or walk with any ease, and who, as we have seen, might have memory problems and need logistical support and real time information to enable the journey.

4. The needs of the older traveller

There are several major over-arching issues here, from which all the requirements may be derived, and these issues are:

- There are design features that are absolutely necessary for older people but which can benefit all users.
- That the primary way forward, especially for older users of transport, will be personalised and bespoke use of technology - for transport, assisted living, health etc., all designed in a user-centred and participatory way.
- That there is always going to be a 'tech-savvy' divide, for at least several decades into the future; therefore it is *not* proposed here that we try to increase 'tech-savviness' as such but adapt the technology instead.
- Using new technology is never going to be intuitive to people on the lower levels of tech-savviness unless it's design is specifically targeted for that level of user.

4.1 Design features that deter users

There are already many examples of what constitute 'good' and 'poor' features from an ergonomic perspective, and to this we can also add design principles and what research on technology acceptance is telling us. Examples of technology design featured include the following with comments about why each may deter users:

- Poorly designed keypads. A major problem for old people particularly is that keypads of some digital devices are too small for accurate operation.

- Complex interfaces. Complexity may introduce errors and slow them down. If the interface has jargon and unfamiliar symbols as well as too many choices, this will put off many people. Perceived ease of use as well as perceived usefulness are critical in technology acceptance.

- Counter-intuitive or difficult navigation. Today's older people were born and grew up in the analogue, not binary, world. For a digital interface to be intuitive to them, the design proposition must come from their more analogue-oriented point of view. "Users often leave web pages in 10-20 seconds if they do not see a clear value" [68]. Features like flashing and alternating pictures that make websites aesthetically pleasing are often at a cost of usability.

- Over-functionality. 'Design for design's sake': the evidence shows that on many products there is more functionality than most people ever need.

- Lack of support in relation to technical issues. Many older people are dependent on a friend or relative to help in set-up and support; without these, many more older people would become lapsed users. In addition, for technologies purchased for the long term, there is also a concern associated with 'upgrades'.

- Trust and belief. Not meeting current needs leads to sceptical views on being unlikely to meet any needs as yet unidentified.

4.1.1 Features of good design

One of the most cited 'good' technology examples is the iPad. From the ergonomic and design perspectives, iPad and other 'good' designs have all or most of the following features:

- Natural and intuitive navigation and transaction with a clear and consistent structure such that related things are together;

- Simple to use for easy and common tasks in plain language;

- Straightforward visualisation or complexity is not at all evident;

- Embedded reversibility and tolerance principle to allow easy corrections through *undo* and *redo*;

- Large keypad or touchpoints to avoid making errors;

- Visible options without distraction;

- Built-in feedback available so the user is informed of actions or changes;

- Tolerating varied inputs and sequences; and

- Maintaining consistency with purpose so the user does not have to rethink and remember.

The ‘good’ technology can encourage engagement as it offers independence, allows the user to understand what they are doing and the needs it will meet. It does not require any special expertise or skills and its navigation elements are completely clear. Evidence is pointing towards some new technologies being poorly designed and not meeting many of these criteria, thus making them distinctly unattractive to older people. There are lessons here for all design of technology.

4.1.2 Inclusive design and Norman’s principles

Inclusive Design is based on an explicit understanding of users, tasks, and environments, who are involved throughout the design and development stages. User-centred evaluation drives and refines the design (e.g. Kansei engineering); and the process is iterative, i.e. design-prototype-test-modify repeated.

In addition, design must take into account Norman’s [69] main principles:

- Visibility – the technology must show its functions to the user, “if instruction is needed the design has failed”.
- Conceptual models – the designer’s model of how the user perceives the operation of the device or technology, if there is none, we make up our own.
- Mapping – relationship between the controls and the resulting effects.
- Feedback – showing the effect of every action: is the effect immediately obvious? Visual is not enough may need auditory.
- Affordance – appropriate actions, provide clues to how the technology is operated, and define how it will be used.
- Constraints – lead to inappropriate actions, difficult to use, choices constrained.

4.1.3 Engaging older people with new technology.

Engaging older people with new and emerging technologies is fundamental. As the proportion of older people grows, there will be a concomitant increase in those people with functional decline, who will have specific needs at a personal level. People with mild but undiagnosed dementia will have different needs to those with some physical incapacities. We need to establish the design parameters of how technology can be bespoke and easily operationalised to meet user needs. For example, algorithms can and should be developed to take the user towards those functions they need and to direct them away from what they do not want or need, thus reducing over-functionality- a much-cited dis-benefit of technology. A particular and potentially problematic issue for design is that solutions for one disability may present problems for another. Only inclusive, user-centred and participatory design can respond to this challenge and has the additional benefit in designing services and products usable by those with the lowest tech-savviness.

What needs to **change** is the way the design process works, involving older people on an inclusive basis, with consultative teams of older people with mixed levels of tech-savviness to ascertain the types and depths of need, the prototypes developed for trial uses, with feedback and a repeated iterative process until the older users are content. Every Government should produce a Code of Practice relating to the design of technology being user-centred to promote simple, intuitive, adaptable and possibly adaptive human machine interactions to meet individual users' needs.

4.2 Personalised and bespoke travelling

"Inclusive travel" means that a trip is door-to-door, usually undertaken by a variety of modes including active ones and usable by everyone. The trip-maker requires knowledge of travel mode changes, parking, walking distance, accessibility at interchanges, facilities at various points in the trip (e.g. for meals, toilets, seating, escalator or lift), what to do in case of disruption, and details of destinations. The key hurdles on a trip include physical access/egress barriers, lack of accessible real-time trip information, route mapping, affordable and accessible technology, availability and reliability of support, and reliable multi-operator trip information.

However no 'seamless' or inclusive multi-modal travel is going to happen until all travel-related data are opened up, from both private and public sectors alike. This includes not only operational data, capacities and on board and in-situ facilities, but also information on:

- Accessibility at all levels including road surfaces, curbs, ramps, cycle lanes, walking distances, ticketing, boarding, alighting, resting places, parking and on-board seating availabilities, reservation choices, and access to facilities, etc.;

- Safety and comfort including access to support, flexible pedestrian/cyclist crossing times, road priority, visibility of vulnerable road users, detailed descriptions of spaces, seats and facilities, smart ticketing, etc.; and

- Costs and payment channels and methods.

This list is not exhaustive and more can be added, which leads to the need for new forms of data becoming increasingly available and integrated. "Citizen data", coming from individual users, can provide ratings, reviews, updates on current status and even personal information. This can help travellers in selection and use of services and products; it can also help service and operation providers, designers and developers to identify accessibility gaps, shortcomings and improvements. Examples of how citizen data are obtained currently include social media, apps and blogs, whereby citizens add information that will be of use to other people.

4.2.1 Intelligent mobility

The Internet of Things (IoT) has the potential to embed the smartness into everyday objects and enable them to send and receive data. Infrastructure underpinned by IoT will make it possible for open data to be fully employed for the future of intelligent mobility.

Intelligent mobility needs to be considered in terms of not only the technology and the solutions to problems such as congestion, pollution and even the "lack of joined up thinking" between different means of travel, but also the focus on users. Psychological issues of users need to be addressed if intelligent mobility is to work, understanding particularly the different needs and preferences represented by the increasing numbers of over 85 yrs. Autonomous vehicles (AVs) operating door-to-door, One-to-X user(s) and demand-responsive transport services could present a better solution for this group than trying to help them undertake multi-mode trips.

MaaS is perhaps most relevant to those aged 85 + yrs., especially those who may be more vulnerable, live alone or are potentially isolated and may be 'dependent passengers' and thus are likely to be those needing a personalised service. However this appears to present the greatest challenge in MaaS. These people can be early adopters once transport policy is changed to facilitate the development of personalised travel. A place to start might be the full integration of patient and health/hospital transport without it taking 9 hours to get people back home and this could then be extended, for example, to include shopping.

Transport operators, network providers and local and national governments are among the many stakeholders and there are indeed many beneficiaries, not just older people. However it has many challenges, including opening up of all travel and user data, both public and private, identifying physical and information gaps in the detail necessary to allow access and egress. There is a need to integrate accessibility-related travel and destination information into personalised travel information necessary to maximise mobility; the information needs to be e more intuitive and creative in its presentation.

We also need to redefine future public transport. With the rise of AVs and other advanced technologies, future changes such as car ownership and dependency must be anticipated in a world where there is a great uncertainty in future mobility patterns.

4.3 Adapting technology to suit all users including the 'not tech-savvy'

Most technologies are not designed with older people in mind. Designers need to understand the criteria that people use to discriminate "good" from "bad" design of technology, consider the actual meaning of utility and relevance (of products, devices) to older users. The development and implementation of freely accessible 'learning' apps and websites that provide location details, images and dimension details to enable real and accurate travel choices to be made for use on both personal devices and in-situ guides is important and can make good use of Citizen data, if only this can be opened up, as mentioned above.

4.3.1 The role of 'nudge'

So long as technology continues to develop, there will always be varying levels of tech-savviness However changing and improving access and mobility in relation to transport does not need to involve making people more 'tech-savvy': we can achieve small behaviour changes by using the behavioural economics concept of 'nudging' [70], which means that new technology must appear, to *ALL* people who are non-tech-savvy, to be working in ways they already recognise, see as easy to use and as useful. Some of the 'nudges' include:

- Start from the base of 'good technology' i.e. devices most people have or may be familiar with, e.g. iPad or similar, that can then be used as a platform for further new developments and apps.

- Encourage people by using their own cohort: for example technology suppliers and services providers could increase the presence of older staff at public transport interchanges.

- Run regular attitude and behaviour change campaigns that nudge by focusing on already identified user needs and likely future needs.

- Invest in innovative and creative forms of human machine interactions in terms of 'learning' functions and accessibility of current 'help' functions for older travellers and ascertain the best methods to improve confidence through use.

- Produce a guide for the physical design attributes of display and control features, especially applicable to those in close geographical areas where all relevant information, such as on-board, interchanges, stops and stations, can be shared; this relates also to 'learning apps' mentioned above.

4.3.2 Ergonomic design features

Many people will get slower, their working memory will decline, they will have more problems with divided attention but they may still want to do many or all the things they used to, and technology can be a major part of that. A lot of good ergonomic knowledge exists but does not often seem to permeate into good or appropriate design, and so there is a list of ergonomic and design features that remain problematic and should be addressed:

- Technology must present in a way that focuses on memory support, allowing for actions made to be reversed and contain sufficient appropriate prompts.

- Algorithms and systems within new technology must be developed that allows users to find the relevant information and not be distracted by irrelevant or unneeded instructions or information. Either that or the technology should be a lot less complex.

- The technology should contain the option to look at previous successful behaviours to aid those with memory problems.

- Over-functionality swamping the usual usage. It may intimidate users with low skill levels. There is a variety of good pictorial style algorithms that do not appear to be used.

- Integrated technology has mixed potential: to be a wonderful game-changer but also complex, difficult to understand and threatening control or independence. We need to understand how older people may choose to disengage and engage with it, when and why.

4.3.3 Some solutions are here now, if we choose to introduce them

Addressing many of the points identified will take time as they require significant additional work, changes in behaviour or amendments to legislation. However there are a number of issues where solutions and improvements seem to be readily available or just require minor changes to current material and legislation:

- More larger and ergonomic displays with touchscreen technology for journeys and destinations, e.g. transport interchanges and stations.

- Make it easy and simple to interact with technology in every day behaviour and in public areas or with public services, so that it does not deter people whether tech-savvy or not.

- Assess support systems in terms of their responsiveness and ease of use and understanding: we already have the knowledge to produce better designed instruction and manuals. Help in "setting up" and operating computers is needed and help services should avoid scripted answers that deter users.

- Make personalised and self-paced learning available when introducing new technologies.

- Provide 'senior' preference settings (e.g. larger pictograms, simpler buttons, reduced complexity) and simplified navigation

- Enabling more personalization of over-functional complex controls on interfaces.

5. Conclusions

We know that as we get older, several cognitive and other psychological aspects decline and a few others improve. In addition there are sometimes other deteriorations of which we may be unaware such as decline due to illness. In addition, there will be more and more both able bodied and incapacitated people for the next two decades, at least in most countries.

These things all put increasing requirements onto transport of all kinds. In addition, psychologically, people like their own space in transport, so the future must allow for both public and private means of transport and must be increasingly accessible if older travellers are to engage.

The future will not be the same as at present, nor will it necessarily be a modified version of the present. For certain, we will see electric vehicles of all types, and may also see as many or even more vehicles than at present. The problem of vehicle emissions will largely go away, and within two decades in most developed countries. There will be AVs in increasing numbers that essentially must be adaptive to humans in control of other vehicles. Public transport will use Big Data and IoT such that individual travellers can link their own information and convenience needs to the transport availability and will use simple and easy apps on iPad-type devices and eventually wearable devices. For older people, there has to be a full development of MaaS so that personalisable public transport is available for those choosing not to own a programmable AV of their own, or for those no longer able to drive but who seek independence.

Author details

Joan Harvey[1,2]

1 Newcastle University, United Kingdom

2 Czech University of Life Sciences, Czech Republic

*Address all correspondence to: joan.harvey@ncl.ac.uk

References

[1] OECD (2016), Skills Matter: Further Results from the Survey of Adult Skills. OECD Skills Studies, OECD Publishing, Paris, France

[2] Craik, FIM and Salthouse, TA (2000). The handbook of aging and cognition. Hillsdale, NJ: Erlbaum.

[3] Park DC and Schwartz N (2012) Cognitive Aging: A Primer. Psychology Press Hove, East Sussex.

[4] Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. Annual Reviews of Psychology, 50, 471-507.

[5] Reuter-Lorenz PA and Lustig C (2005) Brain aging: reorganizing discoveries about the aging mind. Current Opinion in Neurobiology. Vol 15(2): 245-251

[6] Reuter-Lorenz, PA and Cappell, KA (2008). Neurocognitive aging and the compensation hypothesis. Current Directions in Psychological Science, 17, 177.

[7] Lesch MF and Hancock PA (2004) Driving performance during concurrent cell-phone use: are drivers aware of their performance decrements? Accident Analysis and Prevention Vol 36(3): 471-480

[8] Ball, K.et al. (1998) 'Driving avoidance and functional impairment in older drivers', Accident Analysis & Prevention, 30(3), pp. 313-322.

[9] Rosenberg, L, Kottorp, A, Winblad, B and Nygård, L (2009). Perceived difficulty in everyday technology use among older adults with or without cognitive deficits. Scandinavian Journal of Occupational Therapy, 16(4), pp216-226

[10] Nygård, L and Kottorp, A (2014). Engagement in instrumental activities of daily living, social activities, and use of everyday technology in older adults with and without cognitive impairment. The British Journal of Occupational Therapy, 77(11). bjo.sagepub.com

[11] Carlson, MC, Erickson, KI, Kramer, AF, Voss, MW, Bolea, N, Mielke, M, Fried, LP (2009). Evidence for neurocognitive plasticity in at-risk older adults: The experience corps program. Journal of Gerontology A: Biological Sciences and Medical Sciences, 64, 1275-1282.

[12] DfT (2019) National Travel Survey: 2019 . Personal travel statistics within Great Britain covering English residents. URLs: https://www.gov.uk/government/statistics/national-travel-survey-2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/906847/nts-2019-factsheets.pdf, https://www.gov.uk/government/statistical-data-sets/nts08-availability-and-distance-from-key-local-services

[13] Bolstad C A (2001) Situation awareness: does it change with age? Proceedings of the Human Factors and Ergonomics Society Annual General Meeting. Vol 45(4): 272-276.

[14] Head D and Isam M (2010) Age effects on wayfinding and route learning skills Behavioural Brain Research. Vol 209(1): 49-58

[15] Moffat SD, Zonderman AB and Resnick SM (2001) Age differences in spatial memory in a virtual environment navigation task. Neorobiology of Aging. Vol 22(5): 787-796

[16] Fisher GH (1968) *The Frameworks for Perceptual Localisation*. University of Newcastle Upon Tyne

[17] Laflamme L and Menckel E (1995) Aging and occupational accidents a review of the literature of the last three decades Safety Science Vol 21(2): 145-161

[18] Charlton JL, Oxley J, Fildes B, Oxley P, Newstead S, Koppel S, O'Hare M (2006). 'Characteristics of older drivers who adopt self-regulatory driving behaviours', Transportation Research Part F: Traffic Psychology and Behaviour, 9(5), pp. 363-373.

[19] Davidse, R.J. (2006) 'OLDER DRIVERS AND ADAS: Which Systems Improve Road Safety?', IATSS Research, 30(1), pp. 6-20.

[20] DC and Hedden T (2001) Working memory and Aging. In Naveh-Benjamin, M Moscovitch and HL Roediger III *Perspectives on Human Memory and Cognitive Aging: Essays in honour of Fergus Craik*. Pp 148-160.

[21] Healey, MK, Campbell, KL and Hasher, L (2008). Cognitive aging and increased distractibility: Costs and potential benefits. Progress in Brain Research, 169, 353-363

[22] Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. Journal of Experimental Psychology. Learning, Memory, and Cognition, 26, 336-358.

[23] Pangbourne, K., Aditjandra, P.T., Nelson, J.D. (2010). New technology and quality of life for older people: exploring health and transport dimensions in the UK context. IET Intelligent Transport Systems, 4(4), December 2010, pp318-327. IET.

[24] Leung, R (2009). Improving the learnability of mobile device applications for older adults. Proc. CHI '09 Extended Abstracts on Human Factors in Computing Systems, pp3125-3128. 4-9 Boston, Mass

[25] Tuzlukov VP (2001) Signal Detection Theory. Springer Science New York.

[26] Brouwer WH, Waterink W, Van Woffelaar PC and Rothengatter T (1991, online 2013) Divided attention in experienced young and older drivers: lane tracking and visual analysis in a dynamic driving simulator. Human Factors: the Journal of the Human Factors and Ergonomics Society. Vol 32(5): 573-582.

[27] Pauzie, A, Bathsavanis, A and Hamama, H (2010). Elderly travelers interacting with mobile phone for public transport service: human centred design issues. Applied Human Factors and Ergonomics Conf., Miami, USA.

[28] Nicholson N (2001) *Managing the Human Animal*. Crown Business, New York

[29] Cacioppo JT, Petty RE, Kao CF and Rodriguez R (1986) Central and peripheral routes to persuasion: An individual difference perspective. Journal of Personality and Social Psychology Vol 51(5); 1032-43.

[30] Czala SJ and Sharit J (1998) Age Differences in Attitudes Toward Computers. The Journal of Gerontology. Series B, Psychological Sciences and Social Sciences 1998 Sep;53(5):329-P340.

[31] Rick, T (2011). Top 12 Reasons why People rest Change. Change Management May 23rd edition. URL: https://www.torbenrick.eu/blog/change-management/12-reasons-why-people-resist-change/ (accessed on 7 Jan 2017).

[32] Rhodes N and Pivik K (2011) Age and gender differences in risky

driving: The roles of positive affect and risk perception. Accident Analysis & Prevention Vol 43(3): 923-931.

[33] Harvey J, Heslop S and Thorpe N (2011) The categorization of drivers in relation to boredom. Transportation Planning and Technology Vol 34(1): 51-69

[34] Scialfa, T, Kline, DW and Lyman, BJ (1987). Age differences in target identification as a function of retinal location and noise level: Examination of the useful field of view Psychology and Aging, Vol 2(1)14-19.

[35] Jette, A.M. and Branch, L.G. (1992). A ten-year follow-up of driving patterns among the community-dwelling elderly, Human Factors: The Journal of the Human Factors and Ergonomics Society, 34(1), pp. 25-31.

[36] Marottoli, R.A.et al. (1997) 'Driving cessation and increased depressive symptoms: Prospective evidence from the New Haven EPESE', Journal of the American Geriatrics Society, 45(2), pp. 202-206.

[37] Musselwhite, CBA and Haddad, H (2010). 'Mobility, accessibility and quality of later life', Quality in Ageing and Older Adults, 11(1), pp. 25-37.

[38] Musselwhite, C. (2011) 'The importance of driving for older people and how the pain of driving cessation can be reduced', Journal of Dementia and Mental Health, 15(3), pp. 22-26.

[39] Eby, DW and Molnar, LJ (2012). 'Has the time come for older driver vehicle?'. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/89960/102821.pdf (accessed on 7 October 2016).

[40] Finucane, M.L.; Alhakami, A.; Slovic, P.; Johnson, S.M. (2000) The Affect Heuristic in Judgment of Risks and Benefits. Journal of Behavioral Decision Making 13 (1): 1-17

[41] Harvey J, Thorpe N and Fairchild R (2013) Attitudes towards and perceptions of eco-driving and the role of feedback systems. Ergonomics Vol 56(3): 507-521

[42] Matthews ML and Moran AR (1986) Age differences in male drivers' perception of accident risk: The role of perceived driving ability. Accident Analysis and Prevention Vol 18(4): 299-313.

[43] McGwin Jr G and Brown DB (1999) Characteristics of traffic crashes among young, middle-aged and older drivers. Accident Analysis and Prevention. Vol 35(3): 181-198

[44] Colcombe, S. J., Erickson, K. I., Scalf, P. E., Kim, J. S., Prakash, R., McAuley, E., Kramer, A. F. (2006). Aerobic exercise training increases brain volume in aging humans. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 61, 1166-1170.

[45] Gordon, BA, Rykhlevskaia, E, Brumback, CR, Lee, Y, Elavsky, S, Konopack, JF and Fabiani, M (2008). Anatomical correlates of aging, pulmonary fitness level, and education. Psychophysiology, 45, 825-838.

[46] Heart, T and Kalderon, E (2013). Older adults: are they ready to adopt health-related ICT? International Journal of Medical Informatics, Vol. 82, Issue 11: 209-231,. Elsevier

[47] Hanson, V.L. (2009). *Age and web access: the next generation*. In: Proceedings of W4A, Madrid, Spain, April 20-21, pp. 7-15. ACM, New York

[48] Trewin, S., Richards, J., Bellamy, R., John, B.E., Swart, C., Sloan, D.

(2011). Extending predictive models of exploratory behavior to broader populations. In: Stephanidis, C. (ed.) Proceedings of HCI International, LNCS, Vol. 6765, pp. 152-161. Springer, Heidelberg.

[49] Vines, J, Pritchard, G, Wright, P, Olivier, P and Brittain, K. (2015). An age-old problem: examining the discourses of ageing in HCI and strategies for future research. ACM Transactions on Computer-Human Interaction (TOCHI), 22(1), Feb. 2015. DOI: http://dx.doi.org/10.1145/2696867.

[50] Page, T. (2014). Touchscreen mobile devices and older adults: a usability study *International Journal of Human Factors and Ergonomics*, 3(1).

[51] Sochor, J. and Nikitas, A. (2016). Vulnerable users' perceptions of transport technologies, Proceedings of the Institution of Civil Engineers, 169(3), pp154-162.

[52] Selwyn, N. (2004). The information aged: A qualitative study of older adults' use of information and communications technology, Journal of Ageing Studies, 18 (4), 369-84.

[53] Barnard, Y., Bradley, M.D., Hodgson, F., Lloyd, A.D. (2013). Learning to use new technologies by older adults: perceived difficulties, experimentation behaviour and usability. Computers in Human behavior, 29(4): 1715-1724. Elsevier.

[54] ONS (2020) Internet access – households and individuals, Great Britain: 2020. https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2020

[55] Damant, J and Knapp, M (2015). What are the likely changes in society and technology which will impact upon the ability of older adults to maintain social (extra-familial) networks of support now, in 2025 and in 2040? Future of ageing: evidence review, Government Office for Science, UK, London, UK, August 2015

[56] Mitzner, T.L., Boron, J.B., Fausset, C.B., Adams, A.E., Charness, N., Czaja, S.J., Dijkstra, K., Fisk, A.D., Rogers, W.A., Sharit, J. (2010). Older adults talk technology: technology usage and attitudes Computers in Human Behavior Volume 26, Issue 6: 1710-1721 http://dx.doi.org/10.1016/j.chb.2010.06.020

[57] Rashidi, P., Mihailidis, A. (2013). A survey on ambient-assisted living tools for older adults. IEEE Journal of Biomedical and Health Informatics, 17(3), pp 579-590. IEEE

[58] Ofcom (2020) Adults' Media Use & Attitudes report 2020. https://www.ofcom.org.uk/__data/assets/pdf_file/0031/196375/adults-media-use-and-attitudes-2020-report.pdf

[59] Conway ARA, Jarrold CE, Kane MJ, Miyake A and Towse JN (2007). Variation in working memory. Oxford University Press. Oxford.

[60] Jaeggi SM, Buschkuehi M, Jonides J and Perrig W (2008). IM proving fluid intelligence with training on working memory. Proceedings of the National Academy of Sciences of the United States of America Vol 105(19): 6829-6833.

[61] Sternberg R (2008) Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Science* May 13 Vol 105 (19): 6791-92

[62] Hartley, J (1999). What does it say? text design, medical information and older readers. In [eds] DC Park, RW Morrell and K Shifren: *Processing of Medical Information in Aging*

Patients: Cognitive and Human Factors Perspectives. Chapter 14: 257-272.

[63] Hartley KW (2006) Media Overload in Instructional Web Pages and the Impact on Learning. *Educational Media International*. Vol 36(2): 145-150

[64] Wilkinson CR and De Angeli A (2014) Applying user centred and participatory design approaches to commercial product development. Design Studies Vol 35 (6): 614-631

[65] TSC (Transport Systems Catapult) (2016) Mobility as a Service: exploring the opportunity for mobility as a service in the UK. URL: https://ts.catapult.org.uk/wp-content/uploads/2016/07/Mobility-as-a-Service_Exploring-the-Opportunity-for-MaaS-in-the-UK-Web.pdf (accessed on 5 December 2016).

[66] Guo W, Harvey J, Edwards S. (2017) Older Travellers and Technology Engagement. Transport Systems Catapult, DfT London

[67] Harvey J & Guo A (2018). *Cycling and other active modes in the UK: how and why are they changing?* London: Government Office for Science.

[68] Tullis TS (2005) *Web-based Presentation of Information: The Top Ten Mistakes and Why They Are Mistakes.* HCI International conference, Las Vegas. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.107.9769&rep=rep1&type=pdf

[69] Norman D (2013) *The Design of Everyday Things*. [revised and expanded]. New York Basic Books

[70] Berry, G (2011). Can older drivers be nudged? RAC Foundation February 2011. URL: http://www.racfoundation.org/assets/rac_foundation/content/downloadables/older_drivers_nudge-main_report-berry.pdf

Chapter 4

Cognitive Profile of Optimistic Offender Drivers Affected by Psychological Interventions for a Sustainable and Safer Driving's Behavior

Carlos Hugo Criado del Valle and Parichehr Scharifi

Abstract

An empirically verified fact is that the majority of traffic accidents occur as a result of risky behaviours that drivers assume, more or less, voluntarily. Drivers are not aware of the perception of risk and the subjective perception of control that we believe we have. We have delimitated the characteristics of a group of optimistic offender drivers, which reveal, on the hand, a great lack of understanding of the true impact that external factors can have on driving and; on the other hand, they tend to overestimate their abilities and overconfident in their ability to avoid accidents. In addition, these drivers do not usually experience negative emotions when they fail. All this, together is what increases the probability of suffering an accident. The consideration of the different cognitive profiles in the perception of the risk or challenge when facing potential traffic situations may provide us with a better understanding of the true nature of offending drivers. The need to carry out experimental studies using new assessment instruments (i.e. Eye tracking, Bio-Feedback, evoked potentials, etc.) can facilitate a better understanding of the cognitive processes that explain the attitudes and behaviors of drivers; and therefore, achieve a lower rate of car accidents.

Keywords: optimism, prefactual thinking, contrafactual thinking, road safety, offender drivers

1. Introduction

Improving smarter transportations systems making journeys safer and faster with new mobility types shared cars and green mobility need new intervention methods which are affecting the behavior of drivers and mobility users along with the classical educational courses for offender drivers.

Of course the question of how to improve road safety, year after year, requires considering carefully the human factors that guide behaviors such as motivations, risk perception or culture. In fact, technological developments of cars and infrastructures, including road signs and pavement markings, have already reached a very high level. Moreover, some new developments are forecasted to be developed or even generalized: alcohol interlocks (that prevents drink driving), Intelligent Speed Assistance (that prevents speeding), and even autonomous cars (that prevent driving). Despite the considerable efforts of car engineers, and the crucial role of traffic laws to increase road safety with licensing and enforcement conditions, there will always be someone in the car that will have to make some decisions and inappropriate behaviors are often considered as contributing for a large part to accidents [1].

Traffic accidents currently represent one of the biggest health problems in the world. According to the World Health Organization [2], the number of road traffic deaths continues to rise steadily, reaching 1.35 million in 2016. There has also been more progress in reducing the number of road traffic deaths among middle- and high-income countries than low-income countries. There has been no reduction in the number of road traffic deaths in any low-income country since 2013. According to the World Health Organization [3], based on motorization in the developed world, traffic accidents are expected to become the fifth leading cause of death in the world by 2030.

There are numerous causes that can explain traffic accidents and their severity. These can range from external factors such as infrastructure (i.e. road maintenance or design), the weather or those related to the vehicle (i.e. age) to human factors. Our interest is focused on human factors, not so much in the physiological characteristics of the driver (i.e. age, gender, ...), but as in the psychological processes that could explain their behavior.

An empirically verified fact is that the majority of traffic accidents occur as a result of risky behaviors that drivers assume, more or less, voluntarily. Drivers are not aware of the perception of risk, cognitive overload and the subjective perception of control that we believe we have. We wrongly estimate the probabilities of obtaining a desired result. On many occasions, we are unable to learn from failures since we attribute failures to external factors. Awareness of how this cognitive process works and involvement in driving could favor the modification of risk behaviors.

We start from the study of the driver's personality traits, specifically optimism and pessimism [4]. Scheier and Carver [5] have characterized optimism as a powerful predictor of behavior. Optimistic people can pursue risky goals, where the chances of success are minimal and have many factors against them; as long as they believe that in their case they can achieve what they want (i.e., perceived controllability) [6]. These drivers predict future events, and therefore anticipate what results they may obtain. They explain in a reasoned way about their intentional behavior and plan their behaviors to achieve the desired results [7–9]. Our interest starts from the study of a type of thoughts (i.e., prefactuals and counterfactuals) that reflect the intention of the person, based on the causal inferences that are established; and how these thoughts play a prominent role in decision-making.

2. Optimism and pessimism in road safety

In the study of the human factor and road safety, a key component are the driver's own personality traits. Like Hampson [10], we consider personality processes to analyze how personality manifests itself in the thoughts, feelings, and behaviors of people to give rise to consequent results. Different investigations have focused on personality traits, such as optimism and pessimism [4]. From the theory of the self-regulation of behavior proposed by Scheier and Carver [11], it is contemplated that optimistic people are the ones who believe they can achieve a desirable outcome, and strive to do so. Pessimists, on the other hand, consider that the outcomes are unattainable, and either give up or do not commit to the actions that would lead to the desired outcome. The self-regulation process is activated when the person compares their current state with the desired one, where the resulting behavior is the reflection of the feedback control [12]. This process involves continuous adjustments and corrections to achieve the established objective. Even when the person pursues multiple objectives simultaneously, it helps them focus their attention and efforts on those priority objectives, and reduce their participation in those that are not yet a priority [13, 14].

Obviously, both groups differ in how they process information from the environment and in how they maintain their expectations for the future, since they construct future scenarios in a different way [15]. Sharot et al. [16] have found that people tend to maintain an optimistic bias, even though the evidence is showing them contrary information. This effect is due to the fact that people update their beliefs more in response to positive information about the future than to negative information. Sometimes optimistic people will carry out extremely risky projects, where the chances of success are minimal and have many factors against them; as long as they believe that in their case they can achieve what they want. Furthermore, they are convinced of it, because when optimists imagine possible scenarios, they focus on the short-term consequences because when optimists imagine the possible scenarios, they do so in greater detail and see it closer in time. In contrast to negative scenarios, where apart from being more unspecific in the details, they distance them in time [15, 17].

In traffic psychology, optimistic biases and belief in the illusion of control may be two determining variables to explain risk factors in driving [18–20]. Therefore, it is necessary to approach the concept of perceived controllability and the perception of risk in drivers, as described below.

3. Perceived controllability

In road safety, it is the driver's own behavior, more or less voluntary, that causes traffic accidents in most cases [21]. The role of perceived controllability is decisive, since drivers, on the one hand, frequently underestimate the probability that they may experience negative events; and on the other hand, they tend to overestimate that they experience positive events, especially when they believe they have sufficient personal resources to face situations or challenges [22]. A theoretical model focused on the field of driving, such as the Task–Capability Interface (TCI) model [23], analyzes the relationships established between the driving task and the capability of the driver. The model indicates that both elements interact to determine task difficulty and the outcome for the driver in terms of whether control is maintained or lost. Azjen [24] specifically insisted in the driver's control beliefs. So, he contemplates that, "Perceived control is determined by control belief concerning the presence or absence of facilitators and barriers to behavioral performance, weighted by their perceived power (impact of each control factor to facilitate or inhibit the behavior)".

This control belief is what can have a direct relationship with the intention of the driver. In this regard, Montaño and Kasprzyk, [25] give a determining role to perceived control in the Theory of Planned Behavior (TPB) and the Theory of Reasoned Action (TRA), which assumes that the best predictor of action is intention. When a person has the intention of taking an action, and believes that they control the process to carry it out, the chances of that intention turning into action are very high. Furthermore, Like Harris [26], we think that perceived controllability is a powerful and robust psychological variable that can help predict behavior, as it reflects the intentions of the driver.

On the other hand the results of some studies show reveal that the possession of smart car technologies influences on drivers' perception of control and attachment. While the previous studies have dealt with perceived control as a predictor of the traffic safety behavior, new studies [27] examines it as one of the 'effects' of smart car technology. This is because the extent to which a driver feels easy or difficult to perform the function of driving will vary depending on the degree of possession and use of smart car technology. Recent studies show contradictory results on this issue. For example, Alliani et al. [28] have found that parking becomes easier under a smart parking system based on vehicle-to-vehicle communication. Birrell and Fowkes [29] have verified that the use of smartphone applications during vehicle operation is very informative rather than visually distractive. It has also been shown that context-based or simulation technologies such as head- 6 up displays and in-vehicle information systems contribute to driving space recognition and information acceptance [30, 31]. These studies support that smart car technology helps drivers feel easier to control the vehicle than before. As many advertisements claim, smart car technologies enhance driving pleasure and control by reducing the driver's cognitive effort in manipulating the vehicle.

The motivational cognitive theoretical models within the Traffic Psychology model have focused especially on the study of risk perception and decision-making. Ajzen [24] incorporated the construct of perceived control over the performance of the behavior, to the Theory of Planned Action, to explain the risks assumed by the driver. In some cases, perceived control may be linked to situations of assumed risk, in which the driver behaves prudently, safely, etc., as predicted by the Zero Risk Model. This model incorporates motivational factors in driver's decisions making [32]. In other cases, when they face risky situations, they drive showing mastery, skill, technique, etc. These skills are determined by the driver's subjective perception of the risk of suffering a road accident (i.e., perceived risk) and by the level of risk willing to accept or tolerate (i.e., perceived risk level), as detailed in the Theory of Homeostasis of the Risk [33].

We previously noted that, cognitive biases in optimism and risk perception. Now, we have contemplated how perceived control can be understood as a generalized belief (i.e., illusion of control) related to one's own person. From the theory of self-regulation of behavior proposed by Scheier and Carver [11], commented previously. The conception of perceived controllability is also integrated. Either the intention or/and behavior would show a direct relation with the feedback control. Where the perceived control would be a generalized belief more related to oneself than to a specific situation. In contrast to the expectations of self-efficacy [34], which would be related to specific beliefs about one's ability to successfully perform a task in a given situation.

In the context of driving, perceived high control can overstate your own ability. This leads us to consider that both optimism and the perceived controllability of the event are closely related [35, 36]. In fact, people manifest their optimistic biases in their perception of personal risk [37, 38], and when they have an accident, they tend to attribute it to external factors (eg, rain, a blowout, etc.), and not to internal factors related to driving [39, 40]. This is because drivers show a tendency to think that they are more skilled than other drivers [41–43]. In addition, they think that they are more likely to obtain the desired results, regardless of the tasks they have to perform [44]. McKenna [22] pointed out how drivers believe they are less likely, in relation to others, to suffer a traffic accident, if they are the ones who drive (i.e, personal control). But if they were passengers, the chances of suffering an accident would be equal to those of the rest of the people. It is the illusion of control that leads them to attribute the successes of driving to their own ability and not to the influence of external factors [19, 45].

On the other hand, there are also studies that show that smart car technology does not affect or even reduce control. Rajaonah et al. [46] conducted an experimental study, but did not reveal the relationship between driving assistance and the driver's confidence. Larsson [47] shows the more the driver uses ADAS, the more (s) he perceives the limits of the device itself. Stanton and Young [48] also explain that vehicle automation can help in situational awareness, but does not affect control over the vehicle. In a situation where the smart car technology is not yet complete and the driver is not assimilated enough, the smart car technology may cause a burden of cognitive overload or hyper-connection. The fatigue of the operation of the media device may interfere with the control of the vehicle. Featherstone [49] emphasizes the emergence of new risks as the degree of dependence on software is increased, mentioning the driver needs to constantly manage various technical devices and information, like an airplane pilot. Different autors [50–52] also suggest that manipulating a smartphone or a digital device attached to the vehicle during operation increases the accident rate. Concerns about malfunctioning of smart car technology [53] can also weaken the sense of control over automobiles.

We have commented that most traffic accidents are due to risky behaviors that drivers assume, more or less, voluntarily. Drivers are generally unaware of the perception of risk and the subjective perception of control during driving. They erroneously estimate the probabilities of obtaining a desired result and, at these times, are unable to learn from failures as they attribute failures to external factors, beyond their control. Next, we will focus on a type of factual thoughts that capture the intentionality of the drivers.

4. Prefactual and counterfactual thoughts

The ultimate goal of any study focused on the human factor within Road Safety, is to be able to explain or predict what a driver could do in the future. As in previous sections, we continue to focus on intention as a predictor of action. At this moment, we incorporate thinking as an explanatory variable. We believe, like Malle and Tate [54], that the best way to explain a future event is based on reasoned explanations of intentional behavior. In our daily life, we continuously anticipate and predict what possible results we could obtain, and with this we plan what we must do to achieve our objectives [8, 9]. Similarly, thoughts about what could have been or what could have been done are frequent, especially after disappointing results [55]. The thoughts that we simulate before the event are called "prefactual", and those alternative thoughts that appear after the event has occurred or that the results have already been obtained, are known as "counterfactual" [56–60].

On the one hand, prefactual thoughts focus on predicting behavior and have to do with intentions to take future action. These types of thoughts appear before taking an action and, the subject can generate various alternatives to achieve the objective (eg, "If it were at the established speed, then it would avoid a fine"). It is important to note that, at the time the thought is generated, neither the alternatives nor the results have been carried out, and may or may not be carried out in the future [61]. On the other hand, counterfactual thoughts are important because they imagine changing aspects of the mental representation of reality. In this cognitive process, different alternatives are generated and compared with the results obtained [55]. Therefore, counterfactual thinking focuses on those thoughts about what might have been, if other actions had been different [62–64].

In these types of thoughts, the subject's intentionality is reflected in the subjective perception of control it shows, in the choice of alternatives and the probability of achieving the proposed objectives. Under the structure of a conditional proposition ("If …, then …"), a causal relationship is established between an action and a result that, currently is not occurring, but that may (or may not) occur in the future [56, 61, 65, 66].

We can differentiate two components in the structure of this type of thinking. One, showing the different action alternatives (i.e., antecedents); another, the achievements of possible outcomes (i.e., consequent). In the example, "If I were cautious, then could avoid having an accident", we can establish a contingency between "cautious" and "avoid an accident". Petrocelli et al. [65] point out that the concept of "Prefactual Potency" contemplates the relationship between antecedents and consequents in this type of thinking. They point out that there is a possibility of the antecedent occurring (i.e., cautions) and that the probable outcome (i.e., avoid an accident) is due to the antecedent indicated. There is also the possibility that the antecedent is perceived as probable, but not the desired result, since whether to have an accident does not depend entirely on me. However, as a general norm, when an individual considers a specific antecedent probable, they consider that the alternative outcome may occur [67]. In such a way that, the fact of establishing a causal relationship can be the basis for activating the behavioral intention "I to be careful, to avoid accidents". As we are commenting, these types of thoughts help us to know how the driver selects the significant information and establishes the implicit causal relationships, which for him have a high adaptive value in the environment.

In a more detailed analysis of the structure of this type of thinking, we can analyze the subject's perception of control. Thus, we can identify the alternatives or actions that the subject uses to achieve the results. Thus, the perception of control, both in prefactual and counterfactual thoughts, can be explained by external factors (e.g., opportunity for action, obstacles, time, cooperation, etc.) or internal (e.g., perception of ability or skills to perform the task) that facilitate or hinder execution. In such a way that, when a person believes they have the opportunities or resources to carry out a certain behavior, it is more likely that they also have the intention of carrying it out [24]. On the contrary, if the person does not believe they have these opportunities or resources, it is highly unlikely that the intention to carry out the behavior will arise. This approach includes the central concept of the theory of behavioral self-regulation [11], which we have been developing. Therefore, we return to contemplate the personality traits (i.e., optimism, pessimism) indicated at the beginning of the chapter.

As we have commented, the analysis of this type of thinking facilitates access to the causal relationships that the subject contemplates. it also informs us of how the subject searches for and selects information to make decisions about what actions to carry out.

5. Offender drivers profile

An important advance in Road Safe would be to establish differential profiles between those drivers who behave in a risky or challenging way and those who conduct themselves prudently. In Spain, drivers can lose points on their driver's license, when they commit offenses such as speeding, driving under the influence of alcohol, etc. The withdrawal of points depends on the severity of the infraction. In such a way that, the Spain's Directorate General for Traffic [DGT – Dirección General de Tráfico] can suspend your driver's license. With this, we want to point out that the withdrawal of the license is not due to minor penalties for lack of information or circumstantial infractions. The withdrawal of the license is due to a series of serious infractions or various penalties that can be repeated and accumulate over time, which can lead to the total loss of points. Offender drivers have the chance to recover their driving license points by attending rehabilitation courses. These courses are referred to as "Intervention, awareness and road re-education courses in the licence points system of Spain's DGT". These differ if the offender drivers have lost any of the points (i.e., partial lost) or all points (i.e., total loss).

In Spain, different studies have been carried out with drivers, who have lost their driving licences [68–71]. Del Valle et al. in different studies [72–74], have compared the profiles of offender drivers with non-offender drivers. This last group of drivers has not had any points deducted from their licenses. They are drivers who attend refresher courses. In addition, they are trained to know those personal and situational factors that lead them to perfect their driving in different situations. In our studies, we focus on a group of optimistic offender drivers and on the analysis of the role of causal attributions in prefactual and counterfactual thinking and emotions under conditions of induced control. The study of thoughts of this nature focuses on the subjective perception of control that drivers think they have. In turn, it could explain why individuals drive dangerously in a more or less voluntary manner. Awareness of how this cognitive process works and its impact on driving could foster a change in dangerous driving habits.

We have conducted different studies to analyze this type of thoughts under different conditions of induce control. In Del Valle, [72] we set ourselves the objective was to analyze to what extent optimistic offender drivers differ from dispositional pessimists in their prefactual thoughts generated under different conditions of induced control. We found that drivers believe that they have a certain ability to influence events, and these types of thoughts we can identify intentions about future action. When analyzing the type of prefactual thoughts that optimistic offender drivers show, we have observed that they do not consider that the errors committed are due to personal failures, they usually make an external attribution of the causes of the errors. In the event, that these drivers generate thoughts about how they could achieve better results, they consider that if certain external factors come into play they could achieve a better result. This leads us to consider that these drivers show great confidence in their abilities. If they considered possible alternatives, in this case unwanted, they would attribute their cause to external factors, such as bad luck, for example.

As is derived from the above, a direct reference is being made to the perceived controllability or illusion of control that these drivers believe they have. These types of drivers may have greater problems to identify difficult or impossible targets [75]. We think that, at these moments they maintain an unjustified optimism, based on an illusion of control, where they believe they control the uncontrollable [76–78].

We especially see this fact, when drivers drink alcohol. In the study Del Valle and Sucha [73], we found the drivers showed greater confidence in their abilities, and they believe that they have greater abilities than others [79]. There are several reasons that may justify these biases in the perception of control. Among these reasons are those reported by drivers who have previous experience of driving under the influence of alcohol and have experienced no negative consequences (e.g. a citation, arrest, crash, etc.). In this case, these drivers experience success in their illegal behavior, since they avoid punishment, and this acts as reinforcement in the perpetuation of their behavior [80]. Obviously, this fact generates expectations of self-efficacy in the driver under the influence of alcohol [81] and explains the lower perception of risk [82, 83].

But, it is not only the great confidence in their abilities, but in a comparative processes with other drivers, they think they are more skilled than other people [84–88]. It is difficult to find a driver who recognizes that he drives very badly. In fact, drivers have a higher opinion of their skills, and they have a low perception of the risk of having an accident [40]. These drivers do not consider the possibility of an accident happening to them, if they can demonstrate their capacity and skills [89, 90].

In a current study, del Valle [74] analyzed whether there are differences between optimistic offender drivers and non-offender drivers in counterfactual thoughts and emotions, under induced control conditions. The functionality of counterfactual thoughts and negative emotions appears under situations with unfavourable outcomes, where more causal reasoning appears [91]. When optimistic offender drivers generate counterfactual thinking to explain the mistakes made. They may overestimate their abilities and seek different excuses, focusing on external aspects of the situation ("If it hadn't snowed, then I could have avoided the accident") to justify their unwanted results [65, 92]. With this justification, it would be possible to reduce the size of the problem, instead of considering other possibilities (i.e., lack of knowledge for driving in snow) [93]. Overestimating their abilities leads them to ignore, or at least underestimate, the negative feedback provided by the environment [65, 94].

In the study del Valle [74] optimistic offenders drivers recorded the lowest values of negative emotions (i.e., guilt and shame). When a person experiences shame, what they create is a desire to flee and disappear. Whereas, in guilt, the person tends to carry out an action that amends the generated result [95, 96]. Our interest is focused on the emotion of guilt. Echeburúa, Corral and Amor [97] point out that guilt is not an end in itself, but is a regulatory emotion that, in general, leads to repair and the avoidance of future damage. In investigations carried out by Tangney's team [98–100] have commented that emotions such as guilt depend on the person's negative judgement of their action. This emotion tends to appear in situations in which a failure is perceived, there is a perception of controllability in their actions and, therefore, the driver is attributed internal responsibility for it (e.g., "If I had not drunk, I would have avoided the accident"). Some authors [96] have commented that guilt can encourage actions to amend the result generated: on the one hand, these drivers do not feel guilt, and on the other, they attribute responsibility for the result to external aspects (e.g., "If the pedestrian had not crossed, I would have avoided the accident). Although we cannot reach a causal implication because we do not use a causal model, we do think these two separate sets of findings could be related.

6. Conclusion

it is a fact that highly skilled drivers, or those who believe they are, may be at greater risk due to their tendency to take risks on the road. In this chapter, we have delimitated the characteristics of a group of optimistic offender drivers, which reveal, on the hand, a great lack of understanding of the true impact that external factors can have on driving and; on the other hand, they tend to overestimate their abilities and overconfident in their ability to avoid accidents. In addition, these drivers do not usually experience negative emotions when they fail. All this, together is what increases the probability of suffering an accident. The consideration of the different cognitive profiles in the perception of the risk or challenge when facing potential traffic situations may provide the instructors on these courses with a better understanding of the true nature of those attending. It is not the same to draw attention to the limitations in terms of skills and capacity of someone who has a generally optimistic view of situations they perceive to be controllable, as to point out those limitations to someone with a generally pessimistic outlook regarding those self-same situations. The ability to restructure cognitive distortions and dismantle mistaken beliefs might be an important feature of courses of this kind, as well as in the instruction of new drivers. This should therefore be a priority to increase the effectiveness of driver rehabilitation courses following the withdrawal of points, and reduce the likelihood of a relapse, which would mean a further step forward in the prevention of road accidents.

However, underlying cognitive functioning or driving exposure have not been widely studied. To this end, we suggest future research should utilize the advances in neuroscience methods and clinical tests with relevant technologies (like Eye tracking, Bio-Feedback and modern devices) which can understand neuroscience signals and driving behaviors and attitudes more accurately to study how the cognitive profile of drivers will be affected and how cognitive functions may relate to improved driving abilities and therefore, fewer motor vehicle crashes. Developing these lines of research will allow investigators to understand the mechanisms which underlie safer driving behaviors in order to ultimately inform prevention and driver training programs.

Author details

Carlos Hugo Criado del Valle[1]* and Parichehr Scharifi[2]

1 Faculty of Psychology, University of Salamanca, Salamanca, Spain

2 Center for Technology and Society (ZTG), Technische Universität Berlin, Berlin, Germany

*Address all correspondence to: delvalle@usal.es

References

[1] Elgarov, A. (1995) *Road crashes and alcohol abusage in Kabardino-Balkaria. Proceedings of the 13th International Conference on alcohol, drugs and driving safety, 2,* 741-743

[2] World Health Organization (2018*). Global status report on road safety 2018*. Geneva, Switzerland

[3] World Health Organization (2019), Global action plan on physical activity 2018-2030. More active people for a healthier world. Geneva. Switzerland

[4] Del Valle, C. H. C., & Mateos, P.M. (2008). Dispositional pessimism, defensive pessimism and optimism: The effect of induced mood on prefactual and counterfactual thinking and performance. Cognition and Emotion *22*, 1600-1612.

[5] Scheier, M. F., & Carver, C. S. (1992). Effects of optimism on psychological and physical well-being: Theoretical overview and empirical update. Cognitive Therapy and Research*, 16*, 201-228.

[6] Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. Nature*, 450* (7166), 102-105.

[7] Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

[8] Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. Psychological Review, *109*(4), 646-678. https://doi.org/10.1037/0033-295X.109.4.646

[9] Walsh, C. R. & Byrne, R. M. J. (2004). Counterfactual thinking: the temporal order effect. Memory & Cognition*, 32,* 369-378.

[10] Hampson, S. E. (2012). Personality Processes: Mechanisms by Which Personality Traits "Get Outside the Skin". Annual Review of Psychology, *63*:1, 315-339

[11] Scheier, M. F., & Carver, C. S. (1985). Optimism, coping and health: Assessment and implications of generalized outcome expectancies. Health Psychology*, 4*, 219-247.

[12] Carver, C. S., & Scheier, M. F. (2016). Self-regulation of action and affect. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of Self-regulation: Research, Theory, and Applications*. (pp. 3-23).Third Edition. New York: The Guilford Press.

[13] Geers, A. L., Wellman, J. A., & Lassiter, G. D. (2009). Dispositional optimism and engagement: The moderating influence of goal prioritization. Journal of Personality and Social Psychology*, 96*: 913-932. doi 10.1037/a0014830.

[14] Carver, C. S., & Scheier, M. F. (2014). Dispositional optimism. Trends Cognitive Science. *18*, (6): 293-299.

[15] Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. Nature Neuroscience*, 14*(11), 1475-1479.

[16] Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. Nature*, 450* (7166), 102-105.

[17] Garrett, N., & Sharot, T. (2017). Optimistic update bias holds firm: Three tests of robustness following Shah et al. Consciousness and Cognition *50*, 12-22.

[18] Deery, H. A. (1999). Hazard and risk perception among young novice

drivers. Journal of Safety Research*, 30* (4), 225-236.

[19] Hammond, T. B., & Horswill, M. S. (2002). The influence of desire for control on drivers´ risk-taking behaviour. *Transportation Research Part.* F: Traffic Psychology and Behaviour*, 4*, 271-277.

[20] Harré, N., & Sibley, C. G. (2007). Explicit and implicit self-enhancement biases in drivers and their relationship to driving violations and crash-risk optimism. Accident Analysis & Prevention*, 39*(6), 1155-1161. https://doi.org/10.1016/j.aap.2007.03.001.

[21] Montoro, L., Roca, J., & Tortosa, F. (2008). Influencia del permiso de conducción por puntos en el comportamiento al volante [Influence of the penalty point system on driving behaviour: Drivers' point of view.]. Psicothema*, 4*, 652-658.

[22] McKenna, F. P. (1993). It won't happen to me: Unrealistic optimism or illusion of control? British Journal of Psychology*, 84*, 39-50.

[23] Fuller, R., & Santos, J.A. (2002). Psychology and the highway engineer. In: Fuller, R., Santos, J.A. (Eds.), *Human Factors for Highway Engineers*. Pergamon, Oxford, pp. 1-10.

[24] Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes *50*, 179-211.

[25] Montaño, D.E. and Kasprzyk, D. (2015). Theory of Reasoned Action, Theory of Planned Behavior, and the Integrated Behavioral Model. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.) *Health behavior: Theory, Research, and Practice.* San Francisco: John Wiley & Sons.

[26] Harris, P. (1996). Sufficient Grounds for Optimism? The Relationship Between Perceived Controllability and Optimistic Bias. *Journal of Social and Clinical Psychology, 15*, 1, 9-52.

[27] So-Eun, L. & Seongcheol, K. (June 24-27, 2017). *The influence of smart car technologies on driver's perceived control and attachment*. 14th Asia-Pacific Regional Conference of the International Telecommunications Society (ITS): "Mapping ICT into Transformation for the Next Information Society", Kyoto, Japan, 24th–27th June, 2017, International Telecommunications Society (ITS), Calgary

[28] Aliedani, A., Loke, S. W., Desai, A., & Desai, P. (2016, September). *Investigating vehicle-tovehicle communication for cooperative car parking: the CoPark approach*. In Smart Cities Conference (ISC2), 2016 IEEE International (pp. 1-8). IEEE.

[29] Birrell, S. A., & Fowkes, M. (2014). Glance behaviours when using an in-vehicle smart driving aid: a real-world, on-road driving study. Transportation Research Part F: Traffic Psychology and Behaviour*, 22*, 113-125.

[30] Doshi, Cheng & Trivedi, 2009 Doshi, A., Cheng, S. Y., & Trivedi, M. M. (2009). A novel active heads-up display for driver assistance. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), *39*(1), 85-93.

[31] Kim, S. & Dey, A. K. (2009). Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 133-142.

[32] Näätänen, R., & Summala, H. (1974). A model for the role of motivational factors in drivers' decision-making. Accident Analysis and Prevention*, 6*, 243-261.

[33] Fuller, R. (2011). Driver control theory: From task difficulty Homeostasis to Risk Allostasis. In B.E. Porter (Ed.) *Handbook of Traffic Psychology*. San Diego: Academic Press.

[34] Bandura, A. (1991). Social cognitive theory of self-regulation. Organizational Behavior and Human Decision Processes, *50*, 248-287.

[35] DeJoy, D. M., (1992). An examination of gender differences in traffic accident risk perception. Accident Analysis and Prevention, *24* (3), 237-246.

[36] White, M. P., Eiser, J. R., & Harris, P. R. (2004). Risk perception of mobile phone use while driving. Risk Analysis, *24*, (2), 323-344.

[37] Deery, H. A. (1999). Hazard and Risk Perception among Young Novice Drivers. Journal of Safety Research, *30* (4): 225-236.

[38] Harris, A. J. L., & Hahn, U. (2011). Unrealistic optimism about future life events: A cautionary note. Psychological Review, *118*(1), 135-154. https://doi.org/10.1037/a0020997

[39] DeJoy, D. M., (1989). The optimism bias and traffic accident risk perception. Accident Analysis and Prevention, *21* (4), 333-340.

[40] Harré, N., Foster, S., & O'Neill, M. (2005). Self-enhancement, crash-risk optimism and the impact of safety advertisements on young drivers. British Journal of Psychology, *96*, 215-230.

[41] McKenna, F. P. Stanier, R. A., & Lweins. C. (1991). Factors underlying illusory self-assessment of driving skill in males and females, Accident Analysis & Prevention, *23* (1): 45-52.

[42] Glendon, I., Dorn, L., Davies, D. R., Matthews, G., & Taylor, R. G. (1996). Age and gender differences in perceived accident likelihood and driver competences. Risk Analysis *16* (6), 755-762.

[43] Horswill, M. S., Waylen, A. E., & Tofield, M. I. (2004). Drivers' ratings of different components of their own driving skill: A greater illusion of superiority for skills that relate to accident involvement. Journal of Applied Social Psychology *34* (1), 177-195.

[44] Tyler, J. M., & Rosier, J. G. (2009). Examining self-presentation as a motivational explanation for comparative optimism. Journal of Personality and Social Psychology, *97*, 716-727.

[45] Horswill, M. S., & McKenna, F. P. (1999). The effect of perceived control on risk taking. Journal of Applied Social Psychology, *29*(2), 377-391. https://doi.org/10.1111/j.1559-1816.1999.tb01392.x.

[46] Rajaonah, B., Tricot, N., Anceaux, F., & Millot, P. (2008). The role of intervening variables in driver: ACC cooperation. International Journal of Human-Computer Studies, *66*(3), 185-197.

[47] Larsson, A. F. (2012). Driver usage and understanding of adaptive cruise control. Applied Ergonomics, *43*(3), 501-506.

[48] Stanton, N. A., & Young, M. S. (2005). Driver behaviour with adaptive cruise control. Ergonomics, *48(*10), 1294-1313.

[49] Featherstone, M. (2004). Automobilities: an introduction. Theory, Culture & Society, 21(4/5), 279-284.

[50] Haddington, P., & Mirka Rauniomaa (2011). Technologies, Multitasking, and Driving: Attending to and Preparing for a Mobile Phone Conversation in a Car, Human Communication

Research, 37(2), 223-254, https://doi.org/10.1111/j.1468-2958.2010.01400.x

[51] Kadry, S. (2016, February). *Safe drive in smart city. In Smart Solutions for Future Cities,* (pp. 1-7). IEEE.

[52] Seltzer, M. L., Ju, Y. C., Tashev, I., Wang, Y. Y., & Yu, D. (2011). In-car media search. IEEE Signal Processing Magazine*, 28*(4), 50-60.

[53] Bansal, P., Kockelman, K. M., & Singh, A. (2016). Assessing public opinions of and interest in new vehicle technologies: an Austin perspective. Transportation Research Part C: Emerging Technologies*, 67*, 1-14.

[54] Malle, B. F., & Tate, C. (2006). Explaining the past, predicting the future. In L. J. Sanna & E. Chang (Eds.), *Judgments over time: The interplay of thoughts, feelings, and behaviors*. New York: Oxford University Press.

[55] Byrne, R. M. J. (2016). Counterfactual Thought. Annual Review of Psychology*, 67*, 135-157 https://doi.org/10.1146/annurev-psych-122414-033249

[56] Epstude, K. (2018). Counterfactual Thinking. In G. Oettingen. A. Timur Servincer, & P.M. Gollwitzer (Eds), *The Psychology of Thinking about the Future*. New York, The Guilford press.

[57] Roese, N., & Morrison, M. (2009). The Psychology of Counterfactual Thinking. *Historical Social Research/ Historische Sozialforschung, 34* (2 (128), 16-26.

[58] Sanna, L. J. (1998). Defensive pessimism and optimism: The bitter-sweet influence of mood on performance and prefactual and counterfactual thinking. Cognition and Emotion*, 12*, 635-665.

[59] Sanna, L. J, Chang, E. C., & Meier, S. (2001). Counterfactual thinking and self-motives. Personality and Social Psychology Bulletin*, 27*, 1023-1034.

[60] Sanna, L. J., Meier, S., & Wegner, E. A. (2001). Counterfactuals and motivation: Mood as input to affective enjoyment and preparation. British Journal of Social Psychology*, 40*, 235-256.

[61] Epstude, K., Scholl, A., & Roese, N. J. (2016). Prefactual thoughts: Mental simulations about what might happen. Review of General Psychology *20* (1), 48-56.

[62] Byrne, R. M. J. (2005). *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, MA: MIT Press

[63] Miller, D. T., Turnbull, W., & McFarland, C. (1990). Counterfactual thinking and social perception: Thinking about what might have been. Advances in Experimental Social Psychology*, 23*, 305-331.

[64] Smallman, R., & Roese, N. J. (2009). Counterfactual Thinking Facilitates Behavioral Intentions. Journal of Experimental Social Psychology*, 45*, 845-852. doi:10.1016/j.jesp.2009.03.002.

[65] Petrocelli, J. V., Seta, C. E., Seta, J. J., & Prince, L. B. (2012). If only I could stop generating counterfactual thoughts": When counterfactual thinking interferes with academic performance. Journal of Experimental Social Psychology *48,* 1117-1123.

[66] Petrocelli, J. V., Percy, E. J., Sherman, S. J., & Tormala, Z. L. (2011). Counterfactual potency. Journal of Personality and Social Psychology, *100*(1), 30-46. https://doi.org/10.1037/a0021523.

[67] Byrne, R. M. J., & Egan, S. M. (2004). Counterfactual and prefactual conditionals. Canadian Journal of Experimental Psychology *58,* 113-120.

[68] Faílde, J. M., García M. A.; Carrera, M. V., Rodríguez. Y., Lameiras M., & Ruiz L., (2018). A qualitative approach to the study of the characteristics of individuals convicted of road traffic offences, International Journal of Law and Psychiatry*, 59*, 10-19.

[69] Martí-Belda, A.; Pastor, J.C.; Montoro, L.; Bosó, P., & Roca, J. (2019). Persistent traffic offenders. Alcohol consumption and personality as predictors of driving disqualification. The European Journal of Psychology Applied to Legal Context*, 11*(2), 81-92.

[70] Padilla, J. L., Doncel, P., Gugliotta, A., & Castro, C. (2018). Which drivers are at risk? Factors that determine the profile of the reoffender driver. Accident, Analysis and Prevention*, 119*, 237-247. doi: 10.1016/j.aap.2018.07.021.

[71] Valero, S., Bosch, R., Corominas, M., Giannoni, A., Barrau, V., Ramos-Quiroga, J.A., Casas, M., 2017. Psychopathology and traffic violations in subjects who have lost their driving license. Comprehensive Psychiatry*, 76*, 45-55. https://doi.org/10.1016/j.comppsych.2017.03.012.

[72] Del Valle, C. H. C. (2019). Cognitive profile of optimistic and pessimistic drivers attending re-education courses. Transportation Research Part F: Traffic Psychology and Behaviour*, 65*, 598-609. https://doi.org/10.1016/j.trf.2018.04.018

[73] Del Valle, C. H. C., & Matus, M. (2019). Effects of alcohol and perceived controllability in optimistic offender drivers. Transportation Research Part F: Traffic Psychology and Behaviour*, 64*, 58-69. https://doi.org/10.1016/j.trf.2019.04.024

[74] Del Valle, C. H. C. (in press). Guilt and shame over what might-have-been among optimistic offender drivers. *Safe Science*.

[75] Wrosch, C., Scheier, M.F., & Miller, G.E. (2013). Goal adjustment capacities, subjective well-being, and physical health. Social and Personality Psychology Compass*, 7*, 847-860

[76] Gollwitzer, P. M., & Kinney, R. F. (1989). Effects of deliberative and implemental mind-sets on illusion of control. Journal of Personality and Social Psychology*, 56,* 531-542.

[77] Keller, L., & Gollwitzer, P. M. (2017). Mindsets Affect Risk Perception and Risk-Taking Behavior Illusory Optimism and the BART. Social Psychology*, 48*, 135-147. https://doi.org/10.1027/1864-9335/a000304.

[78] Klein, C. T. F., & Helweg-Larsen, M., (2002) Perceived control and the optimistic bias: A meta-analytic review, Psychology & Health*, 17*:4, 437-446, DOI: 10.1080/08870440220000004920

[79] Starkey, N. J., & Charlton, S. G. (2014). The effects of moderate alcohol concentrations on driving and cognitive performance during ascending and descending blood alcohol concentrations. Human Psychopharmacology: Clinical and Experimental. *29* (4):370-383.

[80] Stafford, M., & Warr, M. (1993). A Reconceptualization of General and Specific Deterrence. Journal of Research in Crime and Delinquency, *30* (2): 123-135.

[81] Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

[82] Bingham, C. R., Elliott, M. R., & Shope, J. T. (2007). Social and behavioral characteristics of young adult drink/drivers adjusted for level of alcohol use. Alcoholism: Clinical and Experimental Research*, 31* (4), 655-664. doi 10.1111/j.1530-0277.2007.00350.x.

[83] Stephens, A. N., Bishop, C. A. Liu, S., & Fitzharris, M. (2017). Alcohol consumption patterns and attitudes

toward drink-drive behaviours and road safety enforcement strategies. Accident Analysis and Prevention*, 98*: 241-251.

[84] Măirean, C. & Havârneanu, C. E. (2018). The relationship between drivers' illusion of superiority, aggressive driving, and self-reported risky driving behaviors, Transportation Research Part F: Traffic Psychology and Behaviour*, 55*, 167-174,

[85] Dogan, E., Steg, L., Delhomme, P., & Rothengatter, T. (2012). The effects of non-evaluative feedback on drivers' self-evaluation and performance. Accident Analysis & Prevention*, 45*, 522-528. https://doi.org/10.1016/j.aap.2011.09.004.

[86] Horswill, M. S., Taylor, K., Newnam, S., Wetton, M., & Hill, A. (2013). Even highly experienced drivers benefit from a brief hazard perception training intervention. Accident Analysis & Prevention*, 52*, 100-110. https://doi.org/10.1016/j.aap.2012.12.014.

[87] Stephens, A. N., & Ohtsuka, K. (2014). Cognitive biases in aggressive drivers: Does illusion of control drive us off the road? Personality and Individual Differences*, 68*, 124-129. https://doi.org/10.1016/j.paid.2014.04.016.

[88] Stephens, A. N., & Groeger, J. A. (2014). Following slower drivers: Lead driver status moderates driver's anger and behavioural responses and exonerates culpability. Transportation Research Part F: Traffic Psychology and Behaviour*, 22*, 140-149. https://doi.org/10.1016/j.trf.2013.11.005.

[89] Perrissol, S., Boscher, G., Cerclé, A., & Somat, A. (2005). Alcohol drinking, self-efficacy and selective exposure to alcohol adverts. European Review of Applied Psychology*, 55*, 235-243. https://doi.org/10.1016/j.erap.2004.11.003

[90] Gosselin, D., Gagnon, S., Stinchcombe, A., & Joanisse, M. (2010). Comparative optimism among drivers: An intergenerational portrait. Accident Analysis and Prevention *42*, 734-740.

[91] Boninger, D. S., Gleicher, F., & Strathman, A. (1994). Counterfactual thinking: From what might have been to what may be. Journal of Personality and Social Psychology*, 67*(2), 297-307. https://doi.org/10.1037/0022-3514.67.2.297

[92] Girotto, V., Ferrante, D., Pighin, S., & Gonzalez, M. (2007). Postdecisional counterfactual thinking by actors and readers. Psychological Science *18*(6):510-515.

[93] Sherman, S. J., & McConnell, A. R. (1995). "Dysfunctional implications of counterfactual thinking: When alternatives to reality fail us". *In What might have been: The social psychology of counterfactual thinking,* Edited by: Roese, NJ. and Olson, J. M. 199-232. Hillsdale, NJ: Erlbaum.

[94] Allison, S. T., Mackie, D. M., & Messick, D. M. (1996). Outcome biases in social perception: Implications for dispositional inference, attitude change, stereotyping, and social behavior. Advances in Experimental Social Psychology*, 28*, 53-93.

[95] Barrett, K. C. (1995). A functionalist approach to shame and guilt. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 25-63). Nueva York: The Guilford Press.

[96] Fedewa, B. A., Burns, L. R., & Gomez, A. A. (2005). Positive and negative perfectionism and the shame/guilt distinction: adaptive and maladaptive characteristics. Personality and Individual Differences*, 38,* 1609-1619.

[97] Echuburua, E., Corral, P., & Amor, P. J. (2004). Evaluación del daño

psicológico en las víctimas de delitos violentos. Psicopatología Clínica, Legal y Forense, *4*, 227-244.

[98] Tangney, J. P., & Dearing, R. L. (2003). *Shame and guilt*. Guilford Press

[99] Tangney, J. P., Stuewig, J., & Martinez, A. G. (2014). Two faces of shame: The role of shame and guilt in predicting recidivism. Psychological Science, *25,* 799-805.

[100] Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. Annual Review of Psychology, *58,* 345-372.

Section 2

Transportation System Design and Management

Chapter 5

Optimal Management of Electrified and Cooperative Bus Systems

Francesco Viti, Marco Rinaldi and Georgios Laskaris

Abstract

This chapter presents an integrated management approach exploiting the potentials of the new Cooperative Intelligent Transportation Systems (C-ITS) to meet the requirements of the next generation Public Transport (PT). This approach considers the additional complexity of electrification—for instance electric busses need to periodically recharge during operation using dedicated infrastructure. This not only can impact service level, but also extend operating costs with complex electric charges. We develop new strategies explicitly optimizing the interactions within the PT ecosystem consisting of vehicles, traffic signals, and e-bus charging infrastructure. To achieve these goals, we rely on vehicle control rather than on the use of transit signal priority, which in congested urban scenarios can have negative effects on overall traffic performance. The main research challenges are in formulating and solving complex multi-objective optimization problems and real-time control. The proposed system is tested and evaluated in simulation showing the benefits of electrified and cooperative bus systems.

Keywords: public transport, integrated charging and scheduling, cooperative intelligent transportation systems, real-time control, electrification

1. Introduction

Sustainable urban development motivates investments in environment-friendly and user-centered Public Transport (PT) services. Three trends towards next generation PT systems are observed, namely 1) introduction of greener vehicles such as electric/hybrid busses (e-busses), 2) focus on high service quality (e.g. increased ride comfort via mitigation of stop-and-go driving) and 3) reduction of emissions and operating costs related to fuel/energy consumption and equipment wear and tear. These trends however bring new challenges. The first challenge is posed by different operational characteristics and constraints of e-busses, e.g. they need to periodically recharge batteries at e-charging stations placed in selected stops and terminals. This brings additional constraints into PT operations and its cost dynamics. The existing approaches lack the required degree of modeling detail necessary to capture the complex interactions emerging between bus operations and charging infrastructure. The second challenge is how to guarantee comfort- and cost-effective operations without negatively impacting general traffic performance. Relying solely on strategies such as Transit Signal Priority (TSP), which prioritize PT vehicles at signalized intersections, might cause congestion effects that could backfire on the PT system itself.

The main contribution of this work is that we jointly address constraints and control capabilities of all entities of the PT ecosystem, which consists of signal control, (e-)busses, and e-bus charging infrastructure. The developed methods combine cooperation and negotiation between all actors thanks to connectivity, in order to effectively achieve mutual goals. Thanks to bus real-time positioning systems (Automatic Vehicle Location, AVL) and vehicle-to-infrastructure communication (Signal Phase and Timing, SPaT), multi-objective optimization is employed to determine bus dispatching time, operating speeds, dwell time plans, e-bus charging schedules, and TSP requirements. Regarding the interaction between busses and e-charging infrastructure, the objective is to minimize electricity costs and adhere to the planned bus dispatching times. From the online/operational perspective, the problem is to model and optimize a connected and cooperative system with a set of heuristic tools and actions, such that real-time system disturbances can be addressed, in order to maximize the adherence to the offline plans. For example, busses can use information on upcoming green times to adapt their speeds or hold at a stop in order to avoid stopping at signals. Consequently, stop-and-go is mitigated in an efficient and non-invasive way.

This chapter is structured as follows. Section 2 provides an overview of the e-bus eco-system, and the integrated design approach we developed in this work. Section 3 focuses on the integrated scheduling and charging problem at the planning phase, in particular considering a hybrid fleet of electric and hybrid busses. Section 4 deals with the operational phase, and in particular it shows the benefits of the cooperative ITS-based control strategies. Finally, section 5 provides an outline and the potential future research directions for this research.

2. The electrified and cooperative bus system

The quality and service level of bus systems often rely on the interaction of different lines, in order to provide optimal frequencies and hence acceptable waiting times for the users, and to offer sufficient capacity to accommodate the demand, measured in terms of passenger flows. These flows vary across the network due to the variability of the demand, which differs depending on the origin and the destination of the users, and in time. To match the demand with the supply, bus operators aim to manage efficiently their fleet of vehicles, identifying at any time the most opportune vehicle type and the number of vehicles to be assigned to a line, together with their dispatching times. This decision has consequences on the way lines run smoothly and provide a certain level of service quality to the passengers, as well as it impacts the operational costs (**Figure 1**). In this study we consider design decisions (node density, network density and line density) as given.

Allocating a small number of vehicles limits the service frequency, which affects the waiting time. However, increasing this number will have a negative impact on the operating costs, since more vehicles and drivers will need to be employed. Too many vehicles on the other hand may result in under-utilizing vehicle capacities, reducing the marginal profits for the operators. Hence, optimally allocating fleet resources in the network is a fundamental planning problem that impacts both operators' costs and passengers' experience.

2.1 Electrification opportunities and challenges

Emerging trends in green PT systems offer new benefits: e-busses reduce emissions, energy use, noise as well as offer smoother rides. There are three types of e-busses—hybrid electric, plug-in hybrid electric, and battery electric. The last two

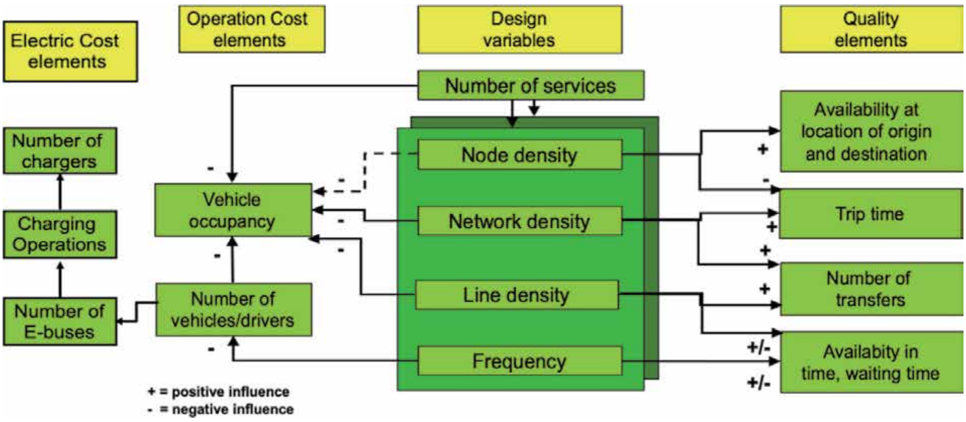


Figure 1.
Integrated design of bus systems.

are able to recharge their batteries from an electric power grid via an *opportunity charging*—a bus periodically charges at bus stops or terminals. This allows to downsize battery and extend bus range to a desirable value. E-bus systems are currently moving from pilot projects to small-scale deployments with single line/operator with very few charging stations. The potentials and needs of large-scale e-bus systems have been investigated by the EU's Zero Emission Urban Bus System (ZeEUS) project [1] as well as Volvo's City Mobility Program [2]. More recent EU projects investigated the impact of fleet mix and configuration parameters to the operation costs [3].

When introducing e-busses, additional costs need in fact to be accounted for, since current battery-capacitated e-busses need to be recharged multiple times a day (e.g. a Volvo 6700 bus can perform a trip in full electric mode for around 30 km, and each vehicle can run distances of a few hundreds of km each day). Current opportunity charging technologies allow a bus to recharge up to 80% in a matter of 6-10 min, while novel flash charging technologies can recharge in less than a minute, but it extends the range of only few more kilometers. An example is the TOSA system in Geneva, a single line that uses both opportunity (3-4 min with low power) and at bus stops e-charging (15-second each 1-1.5 km with high power) [4]. Given the costs of fast and flash charging, bus operators charge their e-busses overnight, when the cost of electricity is lowest, and then use opportunity charging stations, typically located at line terminals, to recharge during the short resting times of the drivers. Flash charging are up to date very rarely implemented, given the very high costs of the relatively small gain in terms of range extension.

The charging infrastructure creates a strong link between infrastructure planning and bus operations [5]. The location and charging operations in fact influences the dispatching times of the vehicles, and in turn irregularities in the operations with recurrent phenomena of bus bunching may result in busses queuing at the charging station, with additional propagation of delays and overall degradation of service levels. Therefore, past research focused on developing a proper system design including strategic locations of e-charging stations [6, 7]. Energy efficiency was also addressed via energy management strategies for the engine [8], and regenerative breaking technologies [9], and taking into account environmental policies such as zero-emission zones [10].

In this study we contribute to this stream of research by focusing on the problem of integrating vehicle scheduling and dispatching times with recharging needs and operations of the e-bus fleet. In particular, we consider the problem of managing a mixed fleet of vehicles, which will be likely to be the case for the next years to come, since full electrification will require heavy investments in the electrical grid and current batteries and chargers are considered a relatively immature technology to completely replace combustion engines. We show in Chapter 3 that optimally assigning vehicle types in the network will provide benefits for both service quality (mitigation of delays due to charging) and operating costs (more e-busses used in daily operations are likely to bring lower energy consumption costs).

2.2 Cooperation opportunities and challenges

The main PT service quality objective is expressed in terms of punctuality (for schedule-based operations) and/or headway regularity (for headway-based operations). Current methods are based upon in-vehicle support systems, managing holding strategies and preferential signal control (TSP) and providing PT vehicles with preferential treatment at intersections via temporary traffic signal timing adjustments [11, 12]. For schedule-based operations, holding strategies (delaying departure of a bus from a bus stop until the scheduled time) ensure punctuality by managing slack times (extra “backup time” inserted into schedules) [13]. The problem of existing methods is that they slow down busses due to the fact that they add delays to the planned trip time [14]. They also address isolated lines and ignore any disturbances observed in real-world PT operations [15]. Headway-based operations are more difficult to control, as the strategies need to account for several busses [16, 17] and multiple interacting lines [18]. Thus, additional ITS systems such as Automated Vehicle Location (AVL), Automated Passenger Count (APC) and a central coordination entity are used to control busses in real time [15, 19].

The core reliability objective is also supported by TSP strategies capable of providing conditional priority. However, since TSP influences the traffic flow reliability [20] its acceptance is limited. Future improvements of TSP exploiting AVL can be achieved. This allows previously unfeasible continuous exchange of information between vehicles and traffic signals [21], allowing cooperation bus-signal through e.g. speed advisory [22, 23]. Such systems are one of the few ITS applications that would provide benefits even at early stages of CV technology [24].

Recent advances in V2I communication enable developing a new promising efficiency-oriented class of driving support systems aiming at improving driving efficiency, comfort and reducing unnecessary stops at signals [25]. Opposite to signal control, which uses CV technology to collect information about the approaching vehicles, in V2I-based systems vehicles use signal control information to optimize their own speeds accordingly. The two SPaT-based DASs researched in literature are the Green Light Optimal Speed Advisory (GLOSA) and Green Light Optimal Dwell Time Advisory (GLODTA). GLOSA provides vehicles with speed guidance, while GLODTA advises additional holding at bus stops. Their main advantage is that these systems improve bus performance with respect to traffic signals, but, unlike TSP, they are non-intrusive (i.e. do not influence signal timings). The two V2I-based advisory systems can be combined to mutually increase their effectiveness [26] and they can be combined with traditional holding strategies. These integrated controls have been shown to meet both objectives of service regularity and reducing the number of stops, as well as they reduce the number of TSP requests [27, 28].

2.3 The eCoBus integrated ecosystem

In this work we adopt a cooperative system approach, following the C-ITS paradigm, reinforced by an energy-aware decision support system. This approach allows to manage the interplay between PT ecosystem actors (vehicles, signals, and e-infrastructure). Secondly, it enables joint optimization and coordination of actions carried out by the different actors, in order to achieve system goals.

Figure 2 provides an overview of the eCoBus integrated system developed in this project. The core module consists of collecting *static* input, namely the location of charging stations, lines timetables, together with the characteristics of the fleet (number of e-busses and hybrid vehicles), the characteristics of the lines (trip lengths) and of the signal infrastructure. We also assume to collect in real time trip times through AVL technology, battery states from the busses, status of each charging (occupied, available) and to have a good estimate of the passenger arrivals at stops (via e.g. APC information). These are input to the *scheduling and charging optimization* module, which is presented in detail in Section 3, whereas the *driver advisory system* combining holding and C-ITS based control and TSP are used at the operational phase to manage the vehicles in real time. The integrated system is shown to provide significant benefits both for planning objectives (better use of the fleet and the charging infrastructure, lower operations costs), and management goals (lower trip time variability and passenger costs, less fuel or energy consumed, less use of TSP requests). These benefits will be showcased in simulations using realistic scenarios in the next sections.

3. Mixed Fleet vehicle scheduling and charging optimization

Vehicle scheduling problems in public transportation have been approached as part of the “full operational planning process” [29]. From a modeling perspective, these problems are usually formulated as Mixed-Integer Linear Programs (MILP), under the name of Single/Multi-Depot Vehicle Scheduling Problem (SDVSP/MDVSP) [30]. The impact of electrification on bus scheduling problems has been recently taken into consideration by researchers, e.g. [31, 32], in preparation and support towards widespread Public Transport electrification. In this Section we

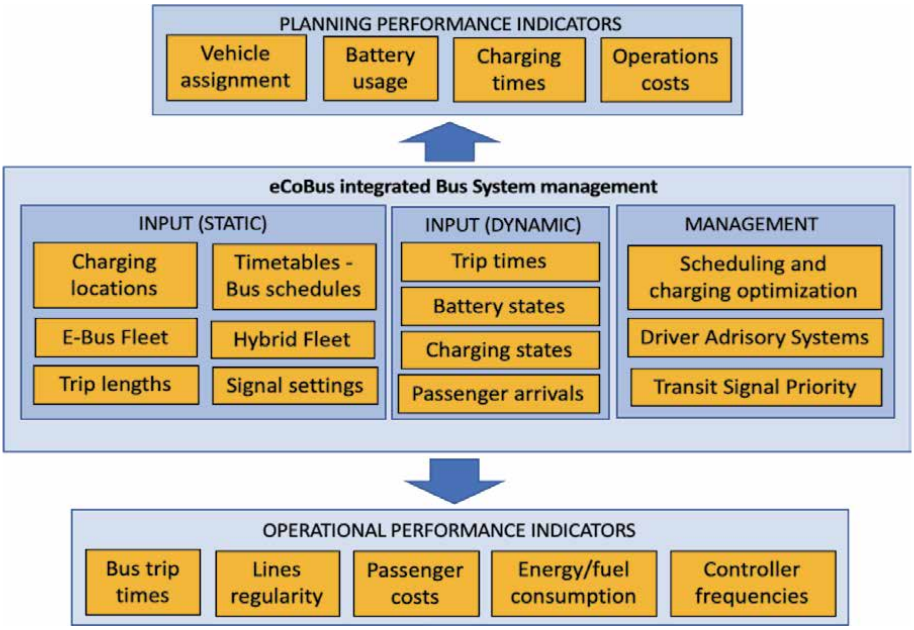


Figure 2.
The eCoBus integrated management ecosystem.

present results stemming from our own recent research efforts [33–35] concerning the development of mixed-fleet vehicle scheduling models and algorithms tailored to the ongoing electrification of the bus fleet in the City of Luxembourg.

Compared to combustion, a fleet of electric or partially electric busses brings novel challenges to transit planning. Within the four decisional stages as discussed in [29] (line planning, timetabling, vehicle scheduling and crew rostering), electrification especially influences scheduling. The problem faces an increase in complexity, as recharging operations must be included without introducing disturbances in the existing schedule, to both ensure that busses have sufficient charge to perform trips and to avoid conflicts at the charging infrastructure. When handling a mixed fleet, optimal scheduling policies should therefore seek to take as much advantage as possible from both coexisting technologies.

In this Section we introduce mathematical models and methodologies to address the problem of scheduling a mixed fleet of conventional and electric busses. We begin by introducing the offline, planning stage optimization problem related to both single and multi-terminal instances. Subsequently we discuss a potential extension towards online, reactive rescheduling in the presence of disturbances, such as delays. Finally, we present a multi-terminal case study based on the city of Luxembourg, in the eponymous country.

3.1 Offline optimization: The SDEVSP and MDEVSP optimization problems

We formulate the problem of assigning a mixed fleet of $I = \{1, \dots, i\}$ electric busses and $H = \{1, \dots, h\}$ hybrid busses to a set of scheduled trips $J = \{1, \dots, j\}$, each characterized by desired departure time $D_j = \{1, \dots, d_j\}$ [time steps], duration $T_j = \{1, \dots, t_j\}$ [time steps] and total energy required $U_j = \{1, \dots, u\}$ [kWh]. Time is subdivided in consecutive steps $\tau = \{0, 1, \dots, N\}$, with a discretization step T_s. For each trip, decision variables $y^t_{i,j}$ and $z^t_{h,j}$describe respectively whether trip j is initiated at time step t by electric bus i or hybrid bus h, while variable $x^t_{i,m}$ represents whether e-bus i is recharging at charging station m at time step t.

Throughout this Section we adopt the assumption that full charging of e-busses happens within a single time step, as will be detailed later. **Table 1** introduces the meaning of each variable and parameter, as well as their domain.

The formulation's objective function, in Eq. (1), is that of minimizing the total operational cost:

$$\begin{aligned} \min \sum_t \sum_i \sum_j \left(y^t_{i,j} \cdot \left(c + r \cdot \left(t - d_j\right)\right)\right) + \sum_t \sum_h \sum_j \left(z^t_{h,j} \cdot \left(\hat{c} + r \cdot \left(t - d_j\right)\right)\right) \\ + \sum_m \sum_i \sum_t q^t_i \cdot x^t_{i,m} \end{aligned} \tag{1}$$

Trip costs c and $\hat{c}$ are determined following Eq. (2), adopting average cost rates per kWh of energy components η_1 and η_2 for e-busses and h-busses respectively:

$$\begin{aligned} c &= \eta_1 \cdot u_j \\ \hat{c} &= \eta_2 \cdot u_j \end{aligned} \tag{2}$$

Energy component η_2 includes a coefficient to represent the difference in consumption rates between electric and conventional combustion (hybrid) busses. The penalty term r [EUR] is applied to trips being performed later than their preferred departure time, to allow, at a cost, trade-offs between schedule adherence and operational performance. We consider $M = \{1, \dots, m\}$ charging stations available at

Var.	Domain	Explanation
$y^t_{i,j}$	$\{0,1\}$	1 if trip j is initiated by e-bus i at time t, 0 otherwise
$z^t_{h,j}$	$\{0,1\}$	1 if trip j is initiated by h-bus h at time t, 0 otherwise
$x^t_{i,m}$	$\{0,1\}$	1 if e-bus i is being recharged at charging station m at time t, 0 otherwise
ε^t_i	$\le E, \in \mathfrak{R}^+$	Total energy in kWh that e-bus i has at time t
u_j	$\mathfrak{R}^+$	Total energy in kWh required to perform trip j, considering e-bus as a means of transport
d_j	$\mathbb{Z}^+$	Preferred departure time step for trip j
t_j	$\mathbb{Z}^+$	Duration of trip j in time steps
s^t_i	$\le E, \in \mathfrak{R}^+$	Slack variable, necessary to ensure that constraint (12) does not violate the domain of ε^t_i
E	$\mathfrak{R}^+$	Total battery capacity in kWh for all electric busses
μ	$\le 1, \in \mathfrak{R}^+$	Minimum battery charge in percentage for each electric bus
A^t_i	$\{0,1\}$	1 if e-bus i is not available to perform any trip nor recharge at time t, 0 otherwise
H^t_h	$\{0,1\}$	1 if h-bus h is not available to perform any trip at time, 0 otherwise

Table 1.
Problem variables and parameters.

selected terminals, and take into account the time dependent cost q^t_i of recharging bus i at time t. System dynamics are captured by constraints (3)–(14):

$$\sum_j y^t_{i,j} + \sum_m x^t_{i,m} \le 1 - A^t_i \forall i, t \tag{3}$$

$$y^t_{i,j} + \frac{1}{t_j - 1} \sum_{\bar{t}=t+1}^{t+t_j-1} \left(\sum_{\bar{j}} y^{\bar{t}}_{i,\bar{j}} + \sum_m x^{\bar{t}}_{i,m} \right) \le 1 \forall i, \forall j : t_j > 1, \forall t : t \ge d_j \tag{4}$$

$$\sum_j z^t_{h,j} \le 1 - H^t_h \forall h, t \tag{5}$$

$$z^t_{h,j} + \frac{1}{t_j - 1} \sum_{\bar{t}=t+1}^{t+t_j-1} \sum_{\bar{j}} z^{\bar{t}}_{h,\bar{j}} \le 1 \forall h, \forall j : t_j > 1, \forall t : t \ge d_j \tag{6}$$

$$\sum_t \left(\sum_i y^t_{i,j} + \sum_h z^t_{h,j} \right) = 1 \forall j \tag{7}$$

$$\sum_{t < d_j} \left(\sum_i y^t_{i,j} + \sum_h z^t_{h,j} \right) = 0 \ \ \forall j \tag{8}$$

$$y^t_{i,j} - \frac{{\varepsilon_i}^t}{u_j + \mu E} \le 0 \forall i, j, \forall t : t \ge d_j \tag{9}$$

$$\sum_i x^t_{i,m} \le 1 \forall m, t \tag{10}$$

$$\varepsilon^0_i = \overline{\varepsilon}_i \forall i \tag{11}$$

$$E \cdot \sum_m x^t_{i,m} - \sum_j y^t_{i,j} \cdot u_j + \varepsilon^t_i - s^t_i = \varepsilon^{t+1}_i \forall i, t \tag{12}$$

$$\sum_m x_{im}^t - \frac{s_i^t}{E} \geq 0 \forall i, t \tag{13}$$

$$\frac{1}{E} \cdot s_i^t - \frac{1}{E} \varepsilon_i^t \leq 0 \forall i, t \tag{14}$$

Constraint (3) ensures that an e-bus can be assigned to at most one trip at a time, or recharge in at most one charger at a time, only in those time steps in which the bus is available. Constraint (4) implies that an e-bus which initiates a trip j whose duration in time steps is greater than one cannot be used to perform any other trip nor recharged throughout the entire duration of trip j. Constraints (5) and (6) enforce similar dynamics for trips being performed by hybrid busses. Note that matrices A_t^i and H_i^t represent exogenous sources of unavailability (e.g., during a scheduled maintenance). Constraint (7) guarantees that each trip be performed exactly once, by either kind of bus, and constraint (8) implies that no trip j can be initiated before its preferred departure time. Constraint (9) guarantees that an e-bus will not perform a trip unless it has enough energy to do so. Constraint (10) implies that a charger can charge at most one e-bus at any given time. Constraint (11) controls the initial state of battery charge of each electric bus, which is set to the exogenously given input value ε_i for all e-busses. Constraint (12) represents recharging and discharging dynamics of electric bus i at time t: if it is assigned to a trip j at time t, its available charge at time $t + 1$ will be reduced by the trip's required energy u_j. Conversely, if the electric bus i is being recharged at time t, total battery capacity E is assumed to be restored at time $t + 1$. We operate under the assumption that a single time step is sufficient to fully recharge an electric bus, although this condition can rather trivially be relaxed by altering this constraint, or by assuming a large enough time step. We consider the availability of charging stations at selected terminals, powerful enough to meet the required electricity demand. To ensure that the total battery capacity is not exceeded during charging, when the residual charge level ε_i^t is greater than zero, a "slack" variable s_i is introduced, with consideration of the domain feasibility for variables ε_i^t and ε_i^{t+1}. When a bus is being recharged, the slack variable s_i must assume a value at least equal to ε_i^t, enforcing that $\varepsilon_i^{t+1} \leq E$. Constraint (13) implies that the slack variable S_i can be non-zero only during recharging operations, and constraint (14) ensures that its maximum value can be ε_i^t. Therefore, the combination of constraints (12)–(14) governs the behavior of the slack variable S_i such that the latter variable is either 0, if bus i is not recharging at time t, or exactly ε_i^t if the bus is recharging.

By supplying a set of lines with accompanying timetables, the model can be employed to determine the optimal scheduling for a mixed-fleet of e-busses and h-busses. Parameters such as fleet size, fleet composition (% of electrics, % of hybrids), charging stations' availability and capacity are supplied exogenously.

3.1.1 MDEVSP: Multi-depot electric vehicle scheduling problem

In order to correctly represent realistic Public Transport services, we improve and extend the model showcased in the previous section to appropriately represent multi-terminal schedules featuring deadheading trips.

For each trip j we therefore introduce a departure terminal α_j and arrival terminal β both within a given set of bus terminals $B = \{1, \dots, b\}$. The set of bus terminals can also include any number of bus depot(s), where busses are stored when not in service. The subset $\overline{B} \subseteq B$ of bus terminals is equipped with charging stations. We assume that each terminal of the $\overline{B}$ subset is equipped with the same amount of m chargers. Deadheading trips are possible between any combination of terminals, with required total energy $\hat{u}_{b1,b2}$ and duration $\hat{t}_{b1,b2}$.

We discretize time in consecutive time steps $\tau = \{0, 1, \dots, N\}$, with a discretization step T_s. For each trip, decision variables $y^t_{i,j}$ and $z^t_{h,j}$ describe respectively whether trip j is initiated at time step t by electric bus i or hybrid bus h, variable $g^t_{i,b}$ controls execution of deadheading trips, and variable $x^t_{i,b,m}$ captures recharging decisions. We adopt the assumption that full charging of e-busses happens within a single time step. Locations of the electric and hybrid busses are captured by variables $\omega^t_{i,b1,b2}$ and $p^t_{h,b}$ respectively.

In this work, we allow deadheading trips for electric busses only. Deadheading is, in fact, critical to optimize usage of electric busses, which have cheaper operational costs, and to optimize their charging dynamics, allowing them to move to terminals equipped with charging stations when needed, while it is not strictly necessary for optimal dispatching of hybrid/conventional combustion busses. The model could anyway be easily extended to consider deadheading for hybrid busses.

The updated formulation's objective function is as follows:

$$\begin{aligned}\min \sum_t \sum_i \sum_j c \cdot \left(1 + r \cdot \left(t - d_j\right)\right) \cdot y^t_{i,j} + \sum_t \sum_h \sum_j \hat{c} \cdot \left(1 + r \cdot \left(t - d_j\right)\right) \cdot z^t_{h,j} \\ + \sum_t \sum_i \sum_{b_1} \sum_{b_2} \bar{c} \cdot \omega^t_{i,b_1,b_2} + \sum_t \sum_i \sum_b \sum_m q^t_i \cdot x^t_{i,b,m} \qquad (15)\end{aligned}$$

Cost vectors c, $\hat{c}$ and $\bar{c}$ are computed as shown in Eq. (16), considering average cost rates per kWh of energy components η_1, η_2 and η_3 for e-busses, h-busses and deadheading trips respectively:

$$\begin{aligned} c &= \eta_1 \cdot u_j \\ \hat{c} &= \eta_2 \cdot u_j \qquad (16) \\ \bar{c} &= \eta_3 \cdot \hat{u}_{b_1,b_2} \end{aligned}$$

Energy component η_2 includes an adaptation coefficient to consider the difference in consumption rates between e-busses and h-busses. A penalty term r [EUR] is applied to trips being performed later than their preferred departure time, to evaluate trade-offs between schedule adherence and operational performance. Regarding the cost of recharging, we take into account the time dependent cost q^t_i of recharging bus i at time t as part of the operational cost needing minimization. Updated system dynamics are captured by constraints (17)–(41) as follows:

$$\sum_j y^t_{i,j} + \sum_{b_1, b_2 \in B} \sum \omega^t_{i,b_1,b_2} + \sum_b \sum_m x^t_{i,b,m} \leq 1 - A^t_i \quad \forall i, t \qquad (17)$$

$$\begin{aligned} y^t_{i,j} + \frac{1}{t_j - 1} \sum_{\bar{t}=t+1}^{t+t_j-1} \left(\sum_{\bar{j}} y^{\bar{t}}_{i,\bar{j}} + \sum_{b_1, b_2 \in B} \sum \omega^t_{i,b_1,b_2} + \sum_{b \in \overline{B}} \sum_m x^{\bar{t}}_{i,b,m} \right) \qquad (18) \\ \leq 1 \forall i, \forall j : t_j > 1, \forall t : d_j \leq t \leq d_j + \theta \end{aligned}$$

$$\begin{aligned} \omega^t_{i,b_1,b_2} + \frac{1}{\hat{t}_b - 1} \sum_{\bar{t}=t+1}^{t+\hat{t}_b-1} \left(\sum_j y^{\bar{t}}_{i,j} + \sum_{\bar{b}_1, \bar{b}_2} \sum \omega^{\bar{t}}_{i,\bar{b}_1,\bar{b}_2} + \sum_{\bar{b} \in \overline{B}} \sum_m x^{\bar{t}}_{i,\bar{b},m} \right) \qquad (19) \\ \leq 1 \forall i, \forall b_1 \in B, \forall b_2 \in B, \forall t : \hat{t}_b > 1 \end{aligned}$$

$$\sum_{j} z^{t}_{h,j} \leq 1 - H^{t}_{h} \forall h, t \tag{20}$$

$$z^{t}_{h,j} + \frac{1}{t_j - 1} \sum_{\bar{t}=t+1}^{t+t_j-1} \sum_{\bar{j}} z^{\bar{t}}_{h,\bar{j}} \leq 1 \forall h, \forall j : t_j > 1, \forall t : d_j \leq t \leq d_j + \theta \tag{21}$$

$$\sum_{t} \left(\sum_{i} y^{t}_{i,j} + \sum_{h} z^{t}_{h,j} \right) = 1 \forall j \tag{22}$$

$$\sum_{t < d_j \cup t > d_j + \theta} \left(\sum_{i} y^{t}_{i,j} + \sum_{h} z^{t}_{h,j} \right) = 0 \quad \forall j \tag{23}$$

$$\sum_{i} x^{t}_{i,b,m} \leq 1 \forall m, t, \forall b \in \overline{B} \tag{24}$$

$$y^{t}_{i,j} - \frac{{\varepsilon_i}^{t}}{u_j + \min\limits_{\beta_j \notin \overline{B}, b_2 \in \overline{B}} \left(\hat{u}_{\beta_j, b_2} \right) + \mu E} \leq 0 \forall i, j, \forall t : d_j \leq t \leq d_j + \theta \tag{25}$$

$$\omega^{t}_{i,b_1,b_2} - \frac{{\varepsilon_i}^{t}}{\hat{u}_{b_1,b_2} + \mu E} \leq 0 \forall i, \forall t, \forall b_1, \forall b_2 \tag{26}$$

$$\varepsilon^{0}_{i} = \overline{\varepsilon}_i \forall i \tag{27}$$

$$E \cdot \sum_{b \in \overline{B}} \sum_{m} x^{t}_{i,b,m} - \sum_{j} y^{t}_{i,j} \cdot u_j - \sum_{b_1, b_2 \in B} \sum \omega^{t}_{i,b_1,b_2} \cdot \hat{u}_{b_1,b_2} + \varepsilon^{t}_{i} - s^{t}_{i} = \varepsilon^{t+1}_{i} \forall i, t \tag{28}$$

$$\sum_{b \in \overline{B}} \sum_{m} x^{t}_{i,b,m} - \frac{s^{t}_{i}}{E} \geq 0 \forall i, t \tag{29}$$

$$\frac{1}{E} \cdot s^{t}_{i} - \frac{1}{E} \varepsilon^{t}_{i} \leq 0 \forall i, t \tag{30}$$

$$\sum_{m} x^{t}_{i,b,m} - g^{t}_{i,b} \leq 0 \forall t, i, \forall b \in \overline{B} \tag{31}$$

$$\sum_{j : \alpha_j = b_1} y^{t}_{i,j} + \sum_{b_2} \omega^{t}_{i,b_1,b_2} - g^{t}_{i,b_1} \leq 0 \forall i, b_1, t \tag{32}$$

$$\sum_{j : \beta_j = b_2} y^{t}_{i,j} + \sum_{b_1} \omega^{t}_{i,b_1,b_2} - g^{t+1}_{i,b_2} \leq 0 \forall i, b_2, t \tag{33}$$

$$\sum_{j : \beta_j = b_2} y^{t}_{i,j} + \sum_{b_1} \omega^{t}_{i,b_1,b_2} - \left(g^{t+1}_{i,b_2} - g^{t}_{i,b_2} \right) \geq 0 \forall i, b_2, t \tag{34}$$

$$\sum_{b} g^{t}_{i,b} = 1 \forall i, t \tag{35}$$

$$g^{0}_{i,b} = \begin{cases} 1 & if \quad b = G_i \\ 0 & otherwise \end{cases} \forall i, b \tag{36}$$

$$\sum_{j : \alpha_j = b} z^{t}_{h,j} - p^{t}_{h,b} \leq 0 \forall h, b, t \tag{37}$$

$$\sum_{j : \beta_j = b} z^{t}_{h,j} - p^{t+1}_{h,b} \leq 0 \forall h, b, t \tag{38}$$

$$\sum_{j:\beta_j=b} z_{h,j}^{t} - \left(p_{h,b}^{t+1} - p_{h,b}^{t}\right) \geq 0 \forall h, b, t \tag{39}$$

$$\sum_{b} p_{h,b}^{t} = 1 \forall h, t \tag{40}$$

$$p_{h,b}^{0} = \begin{cases} 1 & \textit{if } b = P_h \\ 0 & \textit{otherwise} \end{cases} \forall h, b \tag{41}$$

Constraints (17)–(21) avoid conflicts in the usage of resources. Constraints (22)–(23) model when the scheduled trips should be executed. Constraints (24)–(30) control the charging and discharging dynamics and ensure that the trip execution is consistent with battery status. Constraints (31)–(41) control the location dynamics of each bus.

3.2 Online optimization: Decomposition scheme and model predictive control

The optimization model described in the previous section is aimed at determining full day bus schedules in the Public Transport planning stage, i.e. assuming a specific trip timetable and considering no deviations arising from operations. However, the model has been designed with the explicit objective of enabling the application of time-based decomposition schemes, for the sake of scalability in seeking solutions at the planning stage. In this Section we showcase how this decomposable nature can be further exploited, in combination with a Model Predictive Control scheme, to compute real-time rescheduling in case of major disruptions arising from operations (e.g. delays due to overcrowding, bunching, congestion, …).

3.2.1 Time-wise decomposition scheme for the SDEVSP/MDEVSP models

The two models described earlier can be rather straightforwardly decomposed along the time variable $\tau = \{0, 1, \dots, N\}$, arbitrarily choosing both the frequency of sub-problem (defined henceforth as time-lapse) definition and the effective points in time where decomposition should happen. In order to ensure that the decomposed time-lapses effectively capture the original formulation, coupling constraints must be added to the formulation, as follows:

$$A_i^t\big|_f = A_i^{t+l_{(f-1)}}\Big|_{f-1} + \lambda_i^t\big|_{f-1} \forall i, f, t \tag{42}$$

$$H_i^t\big|_f = H_i^{t+l_{(f-1)}}\Big|_{f-1} + \zeta_i^t\big|_f \forall i, f, t \tag{43}$$

$$\lambda_i^t\big|_f = \begin{cases} 1 & \forall i, t : \exists j, \bar{t} : y_{i,j}^{\bar{t}}\Big|_f = 1 \wedge \left(l_f < t \leq \bar{t} + t_j\right) \\ 0 & \textit{otherwise} \end{cases} \tag{44}$$

$$\zeta_i^t\big|_f = \begin{cases} 1 & \forall i, t : \exists j, \bar{t} : z_{i,j}^{\bar{t}}\Big|_f = 1 \wedge \left(l_f < t \leq \bar{t} + t_j\right) \\ 0 & \textit{otherwise} \end{cases} \tag{45}$$

$$\begin{aligned} \overline{\varepsilon_i}\big|_f &= \varepsilon_i^{l_{f-1}}\big|_{f-1} \forall i, f \\ \overline{\varepsilon_i}\big|_0 &= E \forall i \end{aligned} \tag{46}$$

These constraints communicate the status of the fleet along different time-lapses, informing the later time periods on both availability and battery status of the busses as a result of the scheduling decisions performed in the earlier time periods.

By correctly configuring the frequency of time decomposition (at each time step), the width of the time lapses (chosen equal to the desired prediction horizon) and the appropriate bus availability data (busses are marked available only after the effective trip completion, rather than following a pre-determined trip duration), a Model Predictive Control application can be devised, as shown in **Figure 3**.

3.3 Multi-terminal mixed-fleet scheduling in the city of Luxembourg

The proposed model and solution framework have been implemented in Matlab™, employing IBM's ILOG Cplex 12.7 as optimization software. We validated our multi-terminal model against a real-life instance arising in the city of Luxembourg, considering several urban bus lines, as shown in **Figure 4**. Four of the terminals are currently equipped with two opportunity charging stations each. We employ our model on two different sets of tests: one addressing a subset of bus lines (lines 1, 16, 9 and 14, comprising 536 daily trips across 5 bus terminals, 2 of which are equipped with chargers), and one representing the complete instance (10 bus lines, 1034 daily trips across 12 terminals, 4 of which are equipped with chargers).

The results shown in **Figures 5** and **6** show consistently that, as the fleet transitions towards full electrification, the overall operational cost decreases and the number of total recharges increases accordingly.

It is interesting to note that the rate at which operational costs decrease and the total amount of recharging operations increase both exhibit an inflection point: in the set of tests addressing all the 10 lines, the gradient decreases at about 30% of electrified fleet, while in the reduced problem addressing 4 bus lines it becomes actually flat at about 70% of electrified fleet.

These results showcase that a diminishing returns effect might arise when approaching full electric operations. The effect is however less impactful in the full-scale scenario, implying that complex instances might lead to larger potential gains to be attained through electrification.

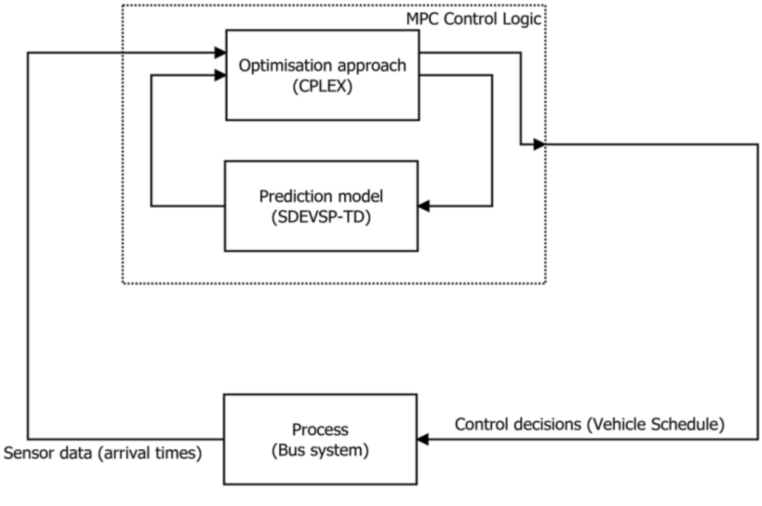


Figure 3.
MPC scheme.

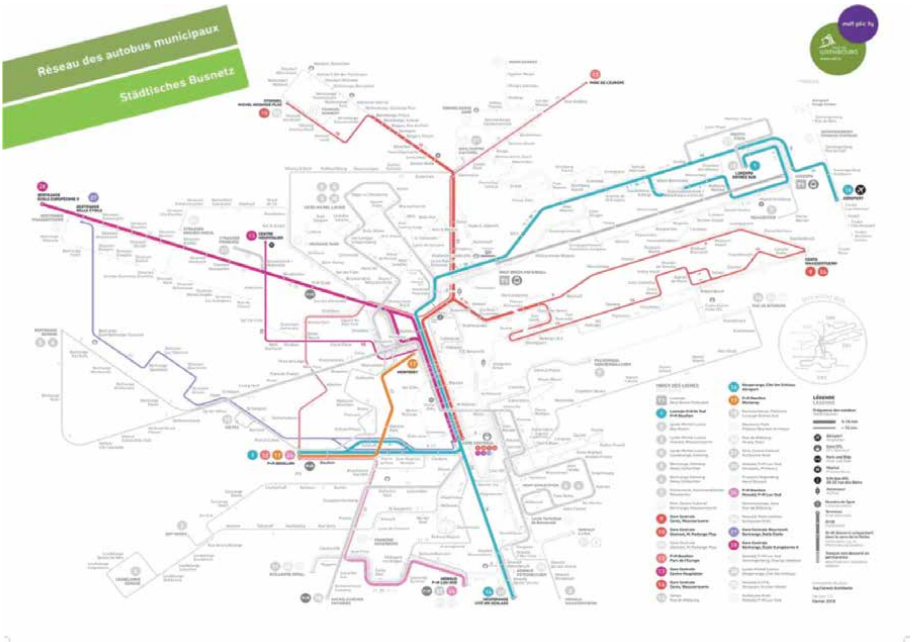


Figure 4.
Case study: 10 bus lines in the City of Luxembourg.

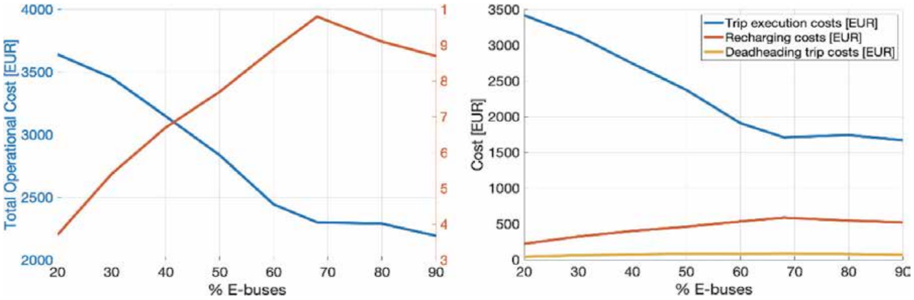


Figure 5.
Bus lines, 536 trips – Total operational costs and recharge operations (left); distinct cost factors (right).

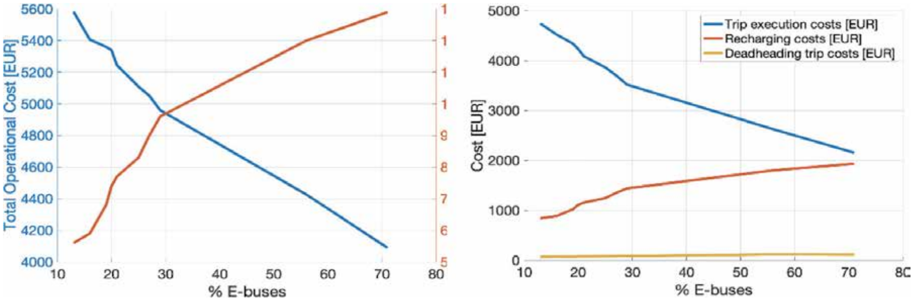


Figure 6.
10 bus lines, 1034 trips – Total operational cost and recharge operations (left); distinct cost factors (right).

4. Real-time cooperative control

Operation is the last pillar, following design and planning. The nature of public transport operations is stochastic, with disruptions occurring due to irregularities in travel times and variation in passenger demand. Thanks to the advances in Intelligent Transportation Systems (ITS), the performance of a transit network can be monitored in real time, and corrective actions can be applied to restore the targeted level of service. All different applications have widened the spectrum of real time control strategies that can be deployed [13]. Until now, C-ITS Driver Advisory Systems have exclusively focused on assisting vehicles traverse signalized intersections and reducing the number of TSP requests, disregarding the consequences of their control actions to the regularity of the transit line [27]. The regularization of a line is the main objective of many real time strategies for public transport, with holding to be one of the thoroughly investigated in literature and applied in practice [17, 36]. We investigate how C-ITS can complement holding strategy and achieve a synergy to address both the objectives of regularity and the mitigation of the number of stops at signalized intersections.

We combine two DAS, namely GLOSA and GLODTA, with a rule-based holding criterion at stops prior to signalized intersections, to provide a pair of holding time and speed advisory or a holding time to achieve both objectives. The combined controllers are presented in the following sections, followed by the results obtained from a real-world case study.

4.1 Regularity based driver advisory systems

4.1.1 Reliability green light optimal speed adaptation (R-GLOSA)

The first regularity based advisory system is R-GLOSA. At the bus stops applied, it instructs a vehicle to be held to regulate the operation and depart with the speed needed to traverse the next green phase. After the arrival of a vehicle at a bus stop prior to a signalized intersection and the completion of dwell time, its position subject to the preceding and the succeeding vehicle is checked. If the headway from the preceding vehicle is short enough, then the vehicle will be held until the consecutive headways are even. We use the same rule-based holding criterion with [36], which regulates the departure time of a vehicle and limits the maximum allowed headway based on the planned headway.

After holding time is calculated, the departure time from the stop is updated and the expected arrival to the first downstream signalized intersection is estimated. The expected arrival time at the first signalized intersection downstream $t_{ijk}^{arr,tl}$ is estimated by adding to the updated exit (departure) time t_{ijk}^{exit}, the time the bus needs between the stop and the intersection. The time corresponds to the expected running time derived by the ratio of the distance $d_{j,tl}$ between current bus stop j and the signalized intersection tl and V_k the average speed of vehicle k at the link downstream of stop j. The expected arrival time is expressed by Eq. (43):

$$t_{ijk}^{arr,tl} = t_{ijk}^{exit} + \frac{d_{j,tl}}{V_k} \tag{47}$$

After the expected arrival time is calculated, information of the signal timing and phasing are transmitted, to estimate if the vehicle will stop or not by the time of the arrival at the intersection. If the current indication is red then the remaining time for red $t^{Red,remain}$ is estimated and added to the expected arrival time $t_{ijk}^{arr,tl}$. Then the recommended speed is calculated using Eq. (44):

$$V_k^{RGLOSA} = \frac{d_{j,tl}}{\left(t_{ijk}^{arr,tl} - t_{ijk}^{exit} + t^{\mathrm{Re}d,remain}\right)} \tag{48}$$

In case of green, the vehicle should either accelerate to catch the current phase or wait for the next green phase. Therefore, two candidate speeds can be recommended, one for the estimated arrival time $t_{ijk}^{arr,tl}$ and one for the expected arrival at the next green phase, given by Eqs. (45) and (46), respectively.

$$V_1^{RGLOSA} = \frac{d_{j,tl}}{\left(t_{ijk}^{arr,tl} - t_{ijk}^{exit}\right)} \tag{49}$$

$$V_k^{RGLOSA} = \frac{d_{j,tl}}{\left(t_{ijk}^{arr,tl} - t_{ijk}^{exit} + t^{Green,remain} + t^{\mathrm{Re}d}\right)} \tag{50}$$

where t^{Red} the red time of the cycle of the current traffic light.

In case of two candidate speeds, the one respecting the speed limits is selected. If both speeds are within the speed limits, V_1^{RGLOSA} is selected since vehicle accelerates to arrive during current green phase. If both speeds are outside the speed limits, no speed advisory is given by the controller. In contrast, if there is no need to restore regularity, the controller is treated as GLOSA.

4.1.2 Reliability green light optimal dwell-time adaptation (R-GLODTA)

R-GLODTA is the second hybrid controller, combining holding and GLODTA. In principle, holding and GLODTA are using the same control logic, by extending the time at stop to achieve their objectives to restore regularity and mitigate stops at traffic lights respectively. Therefore, with this controller, the prolongation of dwell time at stops aims to satisfy both objectives. After the vehicle arrives at the stop and completes dwell time, two candidate holding times are calculated to restore regularity. Then, the expected arrival time to the next signalized intersection is estimated using Eq. (43).

If the expected arrival time is during green phase, then no GLODTA time is needed. In contrast, if the vehicle is expected to arrive during red, then the waiting time at traffic light $t_{ijk}^{wait,tl}$ is calculated by subtracting the current red time $t^{Red,c}$ from the red time t^{Red} as in Eq. (51):

$$t^{wait,tl} = t^{\mathrm{Re}d} - t^{\mathrm{Re}d,c} \tag{51}$$

The waiting time at the traffic light corresponds to the GLODTA time t^{GLODTA}. The waiting time at traffic light is transferred at the bus stop and utilized as dwell time for the passengers. GLODTA time t^{GLODTA} with the duration of green phase define a time interval, within a vehicle will traverse the downstream signalized intersection without stopping (Eq. 52).

$$\left[t^{GLODTA}, t^{GLODTA} + t^{Green}\right] \tag{52}$$

The hybrid controller can work as holding or GLODTA alone depending on the current performance and needs of the system. If both candidate holding times (for regularity and GLODTA) meet the criteria, then the shorter time is selected. If with both holding times, the vehicle is expected to arrive during red, then the holding time with the less estimated remaining time at the traffic light is selected and the controller counts simply as a regularity controller:

$$t^{hold} = \min\left(t_{ijk}^{exit} + t_{ijk,1}^{hold} + \frac{d}{V_k}, t_{ijk}^{exit} + t_{ijk,2}^{hold} + \frac{d}{V_k}\right) \tag{53}$$

In case of on time or late arrival, the vehicle will depart after t^{GLODTA} in order to recover by saving time at traffic light, again if needed. This joint strategy, which we name R-GLODTA.

4.2 Cooperative control in the City of Luxembourg

The two hybrid controllers are tested for one of the busiest lines of the city of Luxembourg, AVL Line 16. Line 16 is the backbone of the bus network of the city of Luxembourg. As depicted in **Figure 7**, The line consists of 19 stops, among which there are stops in the city center, the central business district of Kirchberg and the new activity zone of Cloche d'Or at the south. Additionally, Line 16 connects the central railway station, the airport and the Kirchberg multimodal transport hub. The line is running in high frequency and double articulated busses are used. In addition, the busses run in dedicated lanes and are equipped with AVL technology. We assume that all traffic lights have the same signal program with cycle of 120 s (80 green and 40 red) with the red indication first at the simulation environment. No coordination has been considered between signals.

Two case studies, one for each of the newly introduced controller, were conducted. In both cases, a do-nothing scenario is used as a benchmark scenario. In addition, the hybrid controllers are compared with a holding strategy and the individual application of GLOSA and GLODTA. Moreover, different levels of TSP are put into test. For the R-GLOSA scenarios, three different levels are tested. The

Figure 7.
Line 16 in Luxembourg City.

first level, referred as weak TSP, the scenario in which both green extend and green recall are up to 5 s. With strong TSP, green phase can be modified by 15 s In the R-GLODTA scenarios only strong TSP is tested. Lastly, in the R-GLODTA scenarios, the hybrid controller is combined with GLOSA and TSP.

The main performance indicators used in this study are the adherence of headway of the line as well as the total trip time and its variability. Moreover, we will also analyze the delay at the signalized intersections and the times the vehicles managed to pass through a green phase. Finally, for the performance of the joint controller, the number of times requested is given and the share or each subcontroller are recorded. In summary, these are the performance indicators selected for the study:

- Regularity indicators: Coefficient of variation of headways; bunching;
- Passengers' cost indicators: in-vehicle time; waiting time at stops;
- Link performance indicators: stop frequency and delay at traffic light, average speed and running time;
- Controller performance: share of control requests and of controller choice.

4.2.1 Results

All regularity indicators are summarized in **Table 2**. It is clear from the results that the control schemes, the objective of which is to regulate the operation, dominate the regularity indicators. The coefficient of variation of headway and the level of bunching are chosen are regularity indicators. It should be noted that R-GLOSA has a minor difference from holding control since it is based on the same criterion to calculate holding time. The additional gaining comes from the speed recommendation given by the GLOSA part of the controller. Among strategies there are no significant differences in waiting time of passengers at stops. The independent application of the two DASs has no effect on system's regularity. Both have the same performance with the benchmark scenario. The regularity indicators remain unchanged regardless the TSP strength and similar to the do-nothing scenario. R-GLOSA manages to integrate the performance of holding strategy in terms of regularity and GLOSA in terms of cycle time. The cycle time with R-GLOSA is better than weak TSP and results to the least variable cycle time among all

	CV Line	Bunching	Waiting time [s]	In vehicle time [s]	Cycle time [s]	Cycle time deviation [s]
NC	0.599	0.372	302.98	204.74	4096.91	415.61
HOLDING	0.486	0.269	302.38	211.90	4042.55	415.61
GLODTA	0.628	0.382	302.40	212.49	4166.16	505.49
GLOSA	0.597	0.351	302.63	200.66	4050.49	480.16
RGLOSA	0.466	0.254	302.30	212.73	4042.09	394.56
TSP5	0.607	0.378	303.14	204.00	4060.26	472.07
TSP10	0.590	0.358	302.22	203.25	4013.18	472.07
TSP15	0.613	0.370	301.45	198.51	4012.75	490.94

Table 2.
Regularity performance indicators.

	Frequency of stop at traffic lights	Total average delay at traffic lights [s]	Running time	Average speed [km/h]
NC	0.309	1778.8	2821.0	18.8
HOLDING	0.302	1751.0	2817.0	18.8
GLODTA	0.237	947.6	2790.6	19.1
GLOSA	0.305	942.3	2808.3	19.0
RGLOSA	0.374	465.2	2828.7	18.6
TSP5	0.223	1265.9	2781.0	19.2
TSP10	0.152	876.5	2757.2	19.4
TSP15	0.076	435.5	2738.2	19.7

Table 3.
Link performance indicators.

strategies, giving the operator the opportunity to administer more efficiently the available resources and construct a more robust schedule.

The performance indicators for the links are documented in **Table 3**. It is worth noting that R-GLOSA reports the highest frequency of stops at traffic lights. However, the total average delay at traffic lights is comparable to strong TSP, which has the best performance in these two indicators. GLOSA and GLODTA perform better than holding in reducing the number of stops and the delay at traffic signals. The running time on the signalized links is also lower, meeting the objectives of both GLODTA and GLOSA. R-GLOSA reduces the running time at signalized links at the same level of weak TSP. The average speed of the vehicles increases only at the scenarios with TSP.

Figure 8 shows the trade-off between the average delay at traffic lights and the additional time due to control. When holding is applied, the travel time increases and the additional delay at signalized intersections is not taken into account. TSP heavily prioritizes PT neglecting the impact on regularity by increasing bunching. Obviously, the application of TSP or GLOSA do not introduce any control delay at stops. GLODTA and GLOSA results to similar performance as with intermediate TSP. In contrast to TSP and GLOSA, holding is not causing any delay at traffic lights but increases significantly the additional time added due to control at stops. The delay of R-GLOSA is similar to the one holding, but delay at traffic signals is

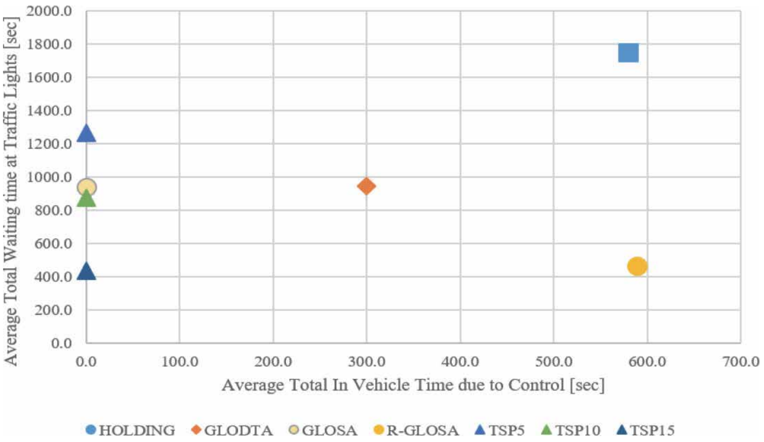


Figure 8.
Tradeoff between waiting time at traffic light and holding time at stop.

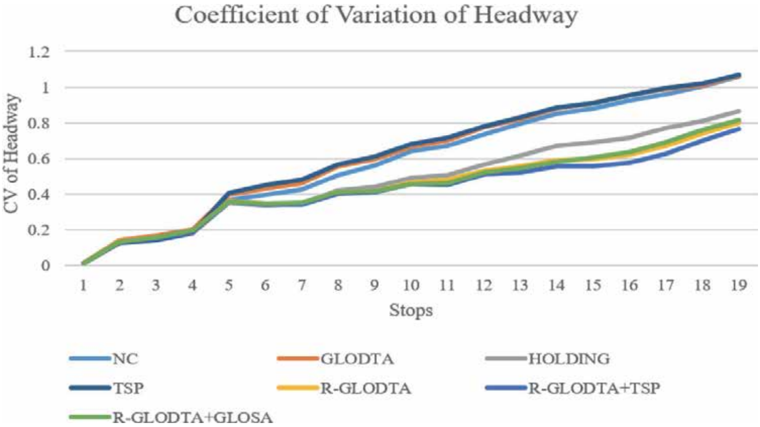


Figure 9.
Coefficient of variation of headway per stop.

significantly reduced to the level of strong TSP. Therefore, the savings obtained in running time can compensate the additional delay at stops. The results can vary subject to the chosen holding criterion.

In **Figure 9**, the coefficient of variation (CV) of headway of all R-GLODTA case study scenarios is depicted. Strategies that target the mitigation of stops at traffic lights neglect the regularity of the line. Between GLODTA or TSP scenarios can be found with minor differences compared to the benchmark scenario, reporting high level of variability which propagates along the line. On the other hand, holding, All the R-GLODTA scenarios show significant improvement on maintaining the propagation of headway low. R-GLODTA outperforms holding and its performance improves further with weak TSP. Although R-GLODTA with GLOSA performs better than GLODTA and TSP, the combination is not the most effective compared to R-GLODTA and TSP.

Regularity performance indicators at line level are summarized in **Table 4**. Similarly to the results in terms of coefficient of variation per stop, R-GLODTA outperforms the other strategies with minor differences from holding and R-GLODTA with TSP. GLOSA has a significant impact on the regularity of the line This can be explained by the fact the GLOSA adjusts the speed in order to traverse green. Acceleration and deceleration can shorten the headway between consecutive vehicles and cause platoons. Again, R-GLODTA has the lowest level of bunching between all scenarios. Passenger indicators are also recorded during simulation. As expected, differences between strategies can be observed in in-vehicle times.

	CV of Headway	Bunching	Waiting time [s]	In vehicle time [s]
NC	0.59	0.37	300.03	204.87
GLODTA	0.62	0.37	300.98	211.2
HOLDING	0.48	0.27	300.08	212.71
R-GLODTA	0.44	0.20	299.96	215.00
R-GLODTA + TSP	0.42	0.19	301.9	212.36
R-GLODTA + GLOSA	0.43	0.21	301.64	226.26
TSP	0.62	0.38	302.75	202.77

Table 4.
Regularity performance indicators.

	Stop at traffic light frequency per segment	Total waiting time at traffic light per segment [s]	Total running time [s]	Average speed [km/h]	Times GLOSA triggered per segment	Number of TSP requests per segment
NC	5.6	113.9	2160.3	18.8	0.0	0.0
GLODTA	4.3	60.7	2135.7	19.0	0.0	0.0
HOLDING	5.4	109.2	2154.2	18.8	0.0	0.0
TSP	1.3	26.1	2084.6	19.7	0.0	4.1
R-GLODTA	4.7	69.4	2132.4	19.0	0.0	0.0
R-GLODTA+TSP	2.9	52.7	2115.9	19.3	0.0	1.7
R-GLODTA+GLOSA	4.7	49.6	2172.1	18.7	2.1	0.0

Table 5.
Link performance indicators.

The additional time added due to control actions increases the time passengers spend on board. The higher in-vehicle time can be compensated with a more robust travel time and the overall improved performance of the line.

One of the objectives of the proposed scheme is the mitigation of stop and go at signalized intersections, therefore the performance of each scenario at a link level is assessed. The results are summarized in **Table 5**.

Unquestionably, providing unconditional signal priority to PT can reduce dramatically the number of stops at signals and the corresponding delay at signalized intersections. However, this reduction will potentially penalize the rest of the traffic. R-GLODTA shows slightly increased number of stops compared to GLODTA alone. This can be explained by the fact that the combined controller prioritizes regularity over stopping at signals. Therefore, it will not exchange holding for regularity to secure passing during green. Weak TSP improves substantially the performance of R-GLODTA in terms of frequency of stops and delay at intersections. Speed adjustment with GLOSA transfers waiting time at traffic lights to running times to the links. A GLOSA advices to decelerate in order to arrive at the intersection during green, prolongs the running time between stops. All R-GLODTA scenarios result in lower total running time compared to an independent application of GLODTA or holding but higher than TSP, but they compensate with their regularity indicators, especially bunching. Among scenarios the differences of the speed are marginal.

We compare the number of TSP requests between the TSP and the R-GLODTA with TSP scenarios. The number of TSP requests is halved with R-GLODTA and with the combination of weak TSP can achieve comparable results with TSP in reducing stop and go actions at traffic lights while it contributes to the regularity of the line.

A final analysis is performed to check how many times the strategies are adopted in the simulated scenarios. **Table 6** shows the share of each control decision, i.e. when each control was needed. Fixing regularity is prioritized over reducing stops at traffic lights. Controlling actions are reduced when R-GLODTA is combined with TSP. R-GLODTA aims to address both objectives and the number of independent applications of holding or GLODTA. On the other hand, the combination with TSP or GLOSA reinforces the objective of GLODTA. The need of holding alone intensifies in these scenarios to restore regularity. With GLOSA, holding is triggered more than half of the times a controller was requested. If the changes of speed do

	Control request	Controller choice		
		GLODTA	**Holding**	**R-GLODTA**
R-GLODTA	61%	38%	42%	19%
R-GLODTA + TSP	58%	37%	49%	14%
R-GLODTA + GLOSA	62%	37%	51%	13%

Table 6.
Controller frequency.

not account for the sequence of vehicles, undesired phenomena as formation of platoons are more likely to occur and impact the performance of a bus line.

5. Conclusions

This chapter has presented an integrated approach to manage electrified bus systems using Cooperative ITS. We first discussed the challenges and opportunities brought by next generation public transport systems, which require to manage the system in an integrated way. Then we introduced novel optimization methods for joint bus scheduling and charging, and real-time operational control strategies. Results in realistic simulations show how the integrated systems achieves cost effective, reliable and energy efficient operations.

Acknowledgements

The authors acknowledge Marcin Seredynski (Volvo E-Bus Competence Center), Erika Picarelli and Andrea D'Ariano(University of Rome Tre). This project has been carried on under to the FNR-CORE Grant C16/IS/11349329 (eCoBus).

Author details

Francesco Viti*, Marco Rinaldi and Georgios Laskaris
University of Luxembourg, Esch-sur-Alzette, Luxembourg

*Address all correspondence to: francesco.viti@uni.lu

References

[1] ZeEUS project (FP7), http://zeeus.eu

[2] City Mobility program, http://news.volvogroup.com/2014/06/13/city-mobility-team-is-the-link-between-stakeholders/

[3] Electric Mobility Europe https://www.electricmobilityeurope.eu/projects/

[4] Auge, O.: TOSA concept: A full electric large capacity urban bus system. In: Proc. 17th European Conference on Power Electronics and Applications. (2015)

[5] Rogge, M., Wollny, S., Sauer, D.U.: Fast charging battery buses for the electrification of urban public transport-a feasibility study focusing on charging infrastructure. Energies 8(5) (2015) 4587-4606

[6] Fusco, G., Alessandrini, A., Colombaroni, C., Valentini, M.P.: A model for transit design with choice of electric charging system. Procedia - Social and Behavioral Sciences 87 (2012) 234-249

[7] Pternea, M., Kepaptsoglou, K., Karlaftis, M.: Sustainable urban transit network design. Transportation Research Part A: Policy and Practice 77 (2015) 276-291

[8] Peng, J., He, H., Xiong, R.: Rule based energy management strategy for a series parallel plug-in hybrid electric bus optimized by dynamic programming. Applied Energy 185, pp. 1633-1643 (2016)

[9] Li, L., et.al.: Model predictive control-based efficient energy recovery control strategy for regenerative braking system of hybrid electric bus. Energy Conversion and Management 111 (2016)

[10] Seredynski M. and Viti F.: Towards Dynamic Zero Emission Zone Management for Plug-in Hybrid Buses. 26th ITS World Congress, Singapore, 21-25 October 2019.

[11] Y.Li: Transit Signal Priority Research. U.S. DOT Transit Administration (2008)

[12] Smietanka, P., Lipinski, B., Szczypiorski, K., Seredynski, M.: Simulation environment for testing transit signal priority. In: Proc. 3rd IEEE International Conference of Cloud Networking (2014)

[13] Ibarra-Rojas, O., Delgado, F., Giesen, R., Munoz, J.: Planning, operation, and control of bus transport systems: A literature review. Transportation Research Part B 77 (38-75) (2015)

[14] Xuan, Y., Argote, J., Daganzo, C.F.: Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis. Transportation Research Part B: Methodological 45(10) (2011)

[15] Argote-Cabanero, J., Daganzo, C.F., Lynn, J.W.: Dynamic control of complex transit systems. Transportation Research Part B: Methodological 81 (2015) 146-160

[16] Hounsell, N., Shrestha, B.: A new approach for co-operative bus priority at traffic signals. IEEE Transactions on ITS 13(1) (2012) 6-14

[17] Laskaris, G., Cats, O., Jenelius, E., Viti, F.: A real-Time holding decision rule accounting for passenger travel cost. IEEE Conference on ITS, pp. 2410-2415 (2016)

[18] Laskaris, G., Cats, O., Jenelius, E., Rinaldi, M., Viti, F.: Multiline holding based control for lines merging to a shared transit corridor.

Transportmetrica B 7(1), pp. 1062-1095 (2017)

[19] Cats, O., Larijani, A., Koutsopoulos, H., Burghout, W.: Impacts of holding control strategies on transit performance bus simulation model analysis. Journal of the TRB 2216 (2011)

[20] Liu, G., Qiu, T.Z.: Trade-offs between bus and private vehicle delays at signalized intersections: case study of a multi-objective model. TRB Meeting (2016)

[21] Smietanka, P., Szczypiorski, K., Seredynski, M., Viti, F.: Distributed automated vehicle location (AVL) system based on connected vehicle technology. 18th IEEE Conference on Intelligent Transportation Systems (ITSC). (2015) 1946-1951

[22] Hu, J., Park, B., Parkany, A.E.: Transit signal priority with connected vehicle technology. Transportation Research Record 4 (2014) 20-29

[23] Hu, J., Park, B.B., Lee, Y.J.: Coordinated transit signal priority supporting transit progression under connected vehicle technology. Transportation Research Part C: Emerging Technologies 55 (2015)

[24] Duncan, L., Head, L.K., Puvvala, R.: Multi-modal intelligent traffic signal system - safer and more efficient intersections through a connected vehicle environment. IMSA Journal LII (5) 12-16 (2014)

[25] Seredynski, M., Khadraoui, D., Viti, F.: Signal phase and timing (SPaT) for cooperative public transport priority measures. 22nd ITS World Congress (2015)

[26] Seredynski, M., Khadraoui, D.: Combining speed and dwell time advisories for improving bus ride comfort. In: Proc. 21st ITS World Congress. (2014)

[27] Seredynski, M., Laskaris, G., Viti, F.: Analysis of Cooperative Bus Priority at Traffic Signals. IEEE Transactions on ITS 21(5), pp. 1929-1940 (2020)

[28] Laskaris, G., Seredynski, M., Viti, F.: Enhancing Bus Holding Control Using Cooperative ITS. IEEE Transactions on ITS 21(4), pp. 1767-1778 (2020)

[29] Ceder, A.: Urban transit scheduling: framework, review and examples. J. Urban Plan. Dev. 128, 225–244 (2002)

[30] Daduna, J.R., Pinto Paixão, J.M.: Vehicle Scheduling for Public Mass Transit — An Overview, in: Daduna, J. R., Branco, I., Paixão, J.M.P. (Eds.), Computer-Aided Transit Scheduling. Springer, pp. 76–90 (1995)

[31] Lajunen, A.: Energy consumption and cost-benefit analysis of hybrid and electric city buses. Transp. Res. Part C Emerg. Technol. 38, 1–15 (2014)

[32] Xylia, M., Leduc, S., Patrizio, P., Silveira, S., Kraxner, F.: Developing a dynamic optimization model for electric bus charging infrastructure. Transp. Res. Procedia, 27, 776–783 (2017)

[33] Picarelli, E., Rinaldi, M., D'Ariano, A., Viti, F.: Model and Solution Methods for the Mixed-Fleet Multi-Terminal Bus Scheduling Problem. Transp. Res. Procedia 275–282 (2020)

[34] Rinaldi, M., Picarelli, E., D'Ariano, A., Viti, F.: Mixed-fleet single-terminal bus scheduling problem: Modelling, solution scheme and potential applications. Omega 96, 102070 (2020)

[35] Rinaldi, M., Picarelli, E., Laskaris, G., d'Ariano, A., Viti, F.: Mixed hybrid and electric bus dynamic fleet management in urban networks: a model predictive control approach. 6th IEEE Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS) pp. 1–8 (2019)

[36] Cats, O., Larijani, A., Ólafsdóttir, Á., Burghout, W., Andréasson, I., Koutsopoulos, H., 2012. Bus-Holding Control Strategies. Transportation Research Record: Journal of the Transportation Research Board 2274, 100–108 (2012)

Chapter 6

Models and Methods for Intelligent Highway Routing of Human-Driven and Connected-and-Automated Vehicles

Fatemeh Alimardani, Nilesh Suriyarachchi,
Faizan M. Tariq and John S. Baras

Abstract

Connected and automated vehicles (CAVs) have seen a rapid surge in interest over the past few years. A lot of focus is being placed on improving the efficiency and robustness of transportation systems by leveraging the sensors and capabilities of CAVs. However, the integration of CAVs into existing traffic infrastructure would give rise to certain issues that must be addressed before the CAVs can be seen ubiquitously on public roads. Since the highway networks are considered permanent investments that are expensive to build and maintain, the priority is to improve the efficiency of the current traffic system. This chapter explores the integration of two of the most common traffic management strategies, namely, ramp metering (RM) and route guidance (RG), into existing highway networks with human-driven vehicles (HDVs). The introduction of CAVs to public roads will engender issues pertaining to safe interactions between CAVs and HDVs. The later part of the chapter addresses the specific problems of improving highway on-ramp merging efficiency by optimally coordinating CAVs. The chapter concludes by presenting a scenario that requires an explicit consideration of interactions between HDVs and CAVs.

Keywords: ramp metering, route guidance, merging behavior, overtaking behavior, human-driven vehicles, connected-and-automated vehicles

1. Introduction

The ultimate goal of automating the driving process is to improve safety by reducing accidents caused by human errors. If all vehicles in a network are human–driven, the efficiency of traffic networks can be improved by the control of traffic signal lights and the routes that drivers can choose. Studying the literature on freeway traffic control for HDVs demonstrate that the integration of traffic control strategies such as ramp metering (RM) and route guidance (RG) improve the network performance in regards to travel time, travel distance, throughput and emissions.

Moreover, it seems unrealistic that all the HDVs will suddenly be replaced by the AVs in the near future. Rather, what seems more plausible is that the AVs will be introduced onto the roads in the presence of the HDVs. Therefore, there is a need to consider cases where it becomes necessary to model the interactions between AVs and HDVs. Delays caused at on-ramps and off-ramps are some of the major contributors to overall system efficiency degradation. In addition to the increase of congestion in the merge lane and outer freeway lanes, merging lanes can have an overflow effect which causes the entire freeway to become congested. However, with the advent of CAVs, a lot more information has been made available for improving this overall process.

Moving on to mixed–autonomy highway networks, as a specific example of the interaction between HDVs and CAVs, the overtaking behavior performed by a CAV is chosen as the target driving behavior for the last section of this chapter. The reason for this choice is that it is one of the more challenging driving behaviors when compared to car following and lane changing as it encompasses the combination of these behaviors.

This chapter is organized as follows: Section 2 reviews the integration of RM and RG, that have shown significant improvements on different control measures for highway networks with HDVs. Section 3 addresses the specific problem of improving freeway on-ramp merging efficiency by optimally coordinating CAVs. Finally, Section 4 explores the overtaking behavior accomplished by a CAV in the presence of HDVs.

2. Integration of ramp metering and route guidance for HDVs

This section will focus on providing a review on the combined RM and RG control as two of the most common traffic control management techniques. To do so, first, a review on traffic flow models will be provided and then, the most common RM and RG strategies will be explained, respectively. At the end of this section, a review on the studies with the focus on the integration of RM and RG will be presented.

2.1 Traffic flow models

Traffic flow models can be categorized into first order and second order models. The most frequently used models are first order models, such as Lighthill-Whitham-Richards (LWR) model [1], which is a continuous model, and the cell-transmission model (CTM) [2], which is a discretized version of the LWR model. The second-order traffic flow models, besides considering the dynamics of the traffic density, introduce a dynamic equation for the mean velocity. The most famous second order model is the Modèle d'Écoulement de Trafic sur Autoroute NETworks (METANET) model [3, 4]. In this section, a review on the CTM and METANET model, as the two most used discrete traffic flow models in the literature, will be provided. The notations adopted in this section are adopted from [5]. **Tables 1** and **2** describe the model variables and parameters of these two models with their symbols, definitions, and units.

2.1.1 The cell transmission model (CTM)

The CTM was first developed by Daganzo [2] in 1992 and then, through out the following years, many other extensions of it were developed. The following version is the original version of the CTM with some minor modifications from [6] and the notations are borrowed from [5]. The CTM is characterised by the following equations:

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L}\left(\Phi_i^+(k) - \Phi_i^-(k)\right) \tag{1}$$

$$\Phi_i^+(k) = \phi_i(k) + r_i(k) \tag{2}$$

$$\Phi_i^-(k) = \phi_{i+1}(k) + s_i(k) \tag{3}$$

$$s_i(k) = \frac{\beta_i(k)}{1-\beta_i(k)}\phi_{i+1}(k) \tag{4}$$

The dynamic equation of the on-ramp queue length is:

$$l_i(k+1) = l_i(k) + T(d_i(k) - r_i(k)). \tag{5}$$

The mainline flows and on-ramp flows are:

$$\phi_i(k) = \min\left\{(1-\beta_{i-1}(k))v_{i-1}(\rho_{i-1}(k) + r_{i-1}(k)), w_i\left(\rho_i^{max} - \rho_i(k) - r_i(k)\right), q_i^{max}\right\} \tag{6}$$

Symbol	**Description**	**Unit/Range**
T	Sampling time	$[h]$
N	Number of cells	int
i	Cell index	$i = \{1, \ldots, N\}$
K	Time horizon	int
k	Time index	$k = \{0, \ldots, K-1\}$
L	Length of each cell	$[km]$
v_i	Free-flow speed	$[km/h]$
ω_i	Congestion wave speed	$[km/h]$
q_i^{max}	Cell capacity (Maximum flow rate)	$[veh/h]$
ρ_i^{max}	Jam density	$[veh/km]$
ρ_i^{cr}	Critical density	$[veh/km]$
l_i^{max}	Maximum on-ramp queue length	$[veh]$
$r_i^{C,max}$	Maximum ramp metering rate	$[veh/h]$
$\rho_i(k)$	Traffic density	$[veh/km]$
$\Theta_i^+(k)$	Total flow entering cell i	$[veh/h]$
$\Theta_i^-(k)$	Total flow exiting cell i	$[veh/h]$
$\phi_i(k)$	Mainstream flow entering cell i from cell $i-1$	$[veh/h]$
$r_i(k)$	Flow entering cell i from its on-ramp	$[veh/h]$
$s_i(k)$	Flow exiting cell i through its off-ramp	$[veh/h]$
$\beta_i(k)$	Split ratio	$\in [0, 1]$
$l_i(k)$	Queue length in the on-ramp	$[veh]$
$d_i(k)$	Flow accessing the on-ramp	$[veh/h]$
$r_i^C(k)$	Ramp metering control variable	$[veh/h]$

Table 1.
CTM model variables and parameters of cell i during interval [kT,(k+1)T).

Symbol	Description	Unit/Range
T	Sampling time	$[h]$
K	Time horizon	int
k	Time index	$k = \{0, \dots, K-1\}$
M	Number of mainline links	int
m	Mainline link index	$m = \{1, \dots, M\}$
N_m	Number of sections of mainline link m	int
i	Section index	$i = \{1, \dots, N_m\}$
O	Number of origin links	int
o	Origin link index	$i = \{1, \dots, O\}$
L_m	Length of each mainline link m	$[km]$
λ_m	Lane numbers of each mainline link m	int
O_n	Set of exiting mainline links from node n	-
I_n	Set of entering mainline links to node n	-
$\bar{I}_n$	Set of entering origin links to node n	-
J_m	Set of destinations reachable from mainline link m	-
$\bar{J}_o$	Set of destinations reachable from origin link o	-
$\bar{\bar{J}}_n$	Set of destinations reachable from node n	-
$\nu^f_{m,i}$	Free-flow speed in section i of link m	$[km/h]$
ρ^{cr}_m	Critical density	$[veh/km]$
ρ^{max}_m	Jam density	$[veh/km]$
q^{max}_o	Capacity of origin link o	$[veh/h]$
τ	Model parameter	-
η	Model parameter	-
χ	Model parameter	-
ϕ	Model parameter	-
a_m	Model parameter	-
δ_{on}	Model parameter	-
$\rho_{m,i,j}(k)$	Partial density	$[veh/km]$
$\rho_{m,i}(k)$	Total density	$[veh/km]$
$\nu_{m,i}(k)$	Mean traffic speed	$[km/h]$
$q_{m,i}(k)$	Traffic flow leaving section i of link m	$[veh/h]$
$\gamma_{m,i,j}(k)$	Composition rate	$\in [0,1]$
$d_{o,j}(k)$	Partial origin demand at origin link o	$[veh/h]$
$d_o(k)$	Total origin demand at origin link o	$[veh/h]$
$l_{o,j}(k)$	Partial queue length at origin link o	$[veh]$
$l_o(k)$	Total queue length at origin link o	$[veh]$
$\gamma_{o,j}(k)$	Composition rate	$\in [0,1]$
$\theta_{o,j}(k)$	Portion of demand originating in origin link o	$\in [0,1]$

$q_o(k)$	Total traffic volume leaving origin link o	$[veh/h]$
$r_o^C(k)$	Ramp metering control variable	$[veh/h]$
$Q_{n,j}(k)$	Flow entering node n	$[veh/h]$
$\beta_{m,n,j}(k)$	Split ratio	$\in [0, 1]$

Table 2.
METANET model variables and parameters of mainline link m, section i, node n, origin link o, destination j during interval [kT,(k+1)T).

$$r_i(k) = \begin{cases} \min\left\{l_i(k) + d_i(k), \rho_i^{max} - \rho_i(k)\right\} & \text{Uncontrolled On} - \text{Ramps} \\ \min\left\{l_i(k) + d_i(k), \rho_i^{max} - \rho_i(k), r_i^{C,max}\right\} & \text{Controlled On} - \text{Ramps} \end{cases} \tag{7}$$

Metering rate variables $r_i^C(k)$ come from the RM control law which will be mentioned in detail in Section 2.2. All variables are bounded between zero and their maximum possible value.

Many extensions of the original CTM have been proposed in the literature in the last two decades. The CTM in a mixed-integer linear form [7], the CTM including capacity drop phenomena [8, 9], the CTM for a freeway network [10], the asymmetric CTM [6], the link-node CTM [11], and the variable-length CTM [12] are some of these extended versions. Although these models have been proposed in different years and are suitable for different networks and applications, the original CTM [2] is the underlying model in all of them and it proves how powerful the original CTM is.

2.1.2 The METANET model

The METANET model presented here is an improved version [4] of the original that was first presented in [3]. However, the notation has been adopted from [5] in order to agree with the other notations of this section.

Freeway Links

$$\rho_{m,i,j}(k+1) = \rho_{m,i,j}(k) + \frac{T}{L_m\lambda_m}\left[\gamma_{m,i-1,j}(k)q_{m,i-1}(k) - \gamma_{m,i,j}(k)q_{m,i}(k)\right] \tag{8}$$

$$\rho_{m,i}(k) = \sum_{j\in J_m} \rho_{m,i,j}(k) \tag{9}$$

$$\gamma_{m,i,j}(k) = \frac{\rho_{m,i,j}(k)}{\rho_{m,i}(k)} \tag{10}$$

$$\begin{aligned} \nu_{m,i}(k+1) = {} & \nu_{m,i}(k) + \frac{T}{\tau}\left[V\left(\rho_{m,i}(k)\right) - \nu_{m,i}(k)\right] + \frac{T}{L_m}\nu_{m,i}(k)\left[\nu_{m,i-1}(k) - \nu_{m,i}(k)\right] \\ & - \frac{\nu' T\left[\rho_{m,i+1}(k) - \rho_{m,i}(k)\right]}{\tau L_m\left[\rho_{m,i}(k) + \chi\right]} \end{aligned} \tag{11}$$

$$q_{m,i}(k) = \rho_{m,i}(k)\nu_{m,i}(k)\lambda_m \tag{12}$$

$$V\left(\rho_{m,i}(k)\right) = \nu_m^f \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k)}{\rho_m^{cr}}\right)^{a_m}\right] \tag{13}$$

The speed reduction caused by merging phenomena near on-ramps (possible additional term to Eq. (11)):

$$-\delta_{on} T \frac{\nu_{m,1}(k) q_o(k)}{L_m \lambda_m \left[\rho_{m,1}(k) + \chi\right]} \tag{14}$$

The speed reduction due to weaving phenomena in case of lane reductions in the mainstream (possible additional term to Eq. (11)):

$$-\phi T \Delta\lambda \frac{\nu_{m,N_m}(k)^2 \rho_{m,N_m}(k)}{L_m \lambda_m \rho_m^{cr}} \tag{15}$$

The virtual downstream density at the end of the link (for node n at the end of link m with more than one outgoing link):

$$\rho_{m,N_m+1}(k) = \frac{\sum_{\mu \in O_n} \rho_{\mu,1}(k)^2}{\sum_{\mu \in O_n} \rho_{\mu,1}(k)} \tag{16}$$

The virtual upstream speed at the beginning of the link (for node n at the beginning of link m with more than one entering freeway link):

$$\nu_{m,0}(k) = \frac{\sum_{\mu \in I_n} \nu_{\mu,N_\mu}(k) q_{\mu,N_\mu}(k)}{\sum_{\mu \in I_n} q_{\mu,N_\mu}(k)} \tag{17}$$

Origin links

$$l_{o,j}(k+1) = l_{o,j}(k) + T\left[d_{o,j}(k) - \gamma_{o,j}(k) q_o(k)\right] \tag{18}$$

$$l_o(k) = \sum_{j \in \bar{J}_o} l_{o,j}(k) \tag{19}$$

$$\gamma_{o,j}(k) = \frac{l_{o,j}(k)}{l_o(k)} \tag{20}$$

$$d_{o,j}(k) = \theta_{o,j}(k) d_o(k) \tag{21}$$

For uncontrolled on-ramps:

$$q_o(k) = \min\left\{d_o(k) + \frac{l_o(k)}{T}, q_o^{max}, q_o^{max} \frac{\rho_m^{max} - \rho_{m,1}(k)}{\rho_m^{max} - \rho_m^{cr}}\right\} \tag{22}$$

For controlled on-ramps:

$$q_o(k) = \min\left\{d_o(k) + \frac{l_o(k)}{T}, q_o^{max}, r_o^C(k), q_o^{max} \frac{\rho_m^{max} - \rho_{m,1}(k)}{\rho_m^{max} - \rho_m^{cr}}\right\} \tag{23}$$

where $r_o^C(k)$ come from the RM control law which will be mentioned in detail in Section 2.2.

Nodes

$$Q_{n,j}(k) = \sum_{\mu \in I_n} q_{\mu,N_\mu}(k) \gamma_{\mu,N_{\mu,j}}(k) + \sum_{o \in \bar{I}_n} q_o(k) \gamma_{o,j}(k) \tag{24}$$

$$q_{m,0}(k) = \sum_{j \in J_m} \beta_{m,n,j}(k) Q_{n,j}(k) \tag{25}$$

$$\gamma_{m,0,j}(k) = \frac{\beta_{m,n,j}(k) Q_{n,j}(k)}{q_{m,0}(k)} \tag{26}$$

In presence of RG control, the splitting rates become the control variables and are calculated based on the RG control law. It will be mentioned in detail in Section 2.3.

Few research studies have developed different versions of the METANET model due to the complexity and non-linearity of the second-order models. However, the extensions of the METANET for a freeway network [13], and the multi-class METANET both for a freeway stretch [14] and for a freeway network [15] are examples of the extensions of this second-order traffic model developed in the recent years.

Both the first-order and second-order models are capable of developing the evolution of traffic flow in both urban and non-urban network. However, to highlight their difference, it is necessary to emphasize that first order models focus on the evolution of the density while the second-order traffic flow models, besides considering the dynamics of the traffic density, explicitly introduce a dynamic equation for the mean speed. Second order models have the distinct advantage over first order models that they can reproduce the capacity drop, which is the observed difference between the freeway capacity and the queue discharge rate. First order models, because they do not capture this phenomenon, are incapable of exploiting the benefits of increasing bottleneck flow. They can only reduce travel time by increasing off-ramp flow. The obvious disadvantage to second order models is that they lead to more complex optimization problems.

The focus of the rest of this section will be on ramp metering and route guidance control schemes as two of the most famous traffic management techniques.

2.2 Ramp metering

Ramp metering is achieved by placing traffic signals at on-ramps to control the flow rate at which vehicles enter the freeway. The ramp metering controller computes the metering rate to be applied. Ramp metering has various goals [16]: to improve or remove congestion, to alleviate freeway flow, traffic safety and air quality, to reduce total travel time and the number of peak-period accidents, to regulate the input demand of the freeway system so that a truly operationally balanced corridor system is achieved. Although the ramp metering provides many advantages, at the same time, it can have disadvantages too. The following are two of the most plausible ones [5]: (1) drivers may use parallel routes to avoid ramp meters which may lead to increased travel time and distance, (2) it can shift the traffic congestion from one location to another.

Ramp metering control strategies can be classified in the following categories [16]: (1) local system where the control is applied to a single on-ramp, (2) coordinated system where the control is applied to a group of on-ramps, considering the traffic conditions of the whole network, (3) integrated system where a combination of ramp metering, signal timing, and route guidance is applied as the control system. Also, from another point of view, there are two types of RM control schemes [16]: (1) pre-timed or isolated where metering rates are fixed and predefined, (2) traffic-responsive control where real time freeway measurements are used to determine the control variables.

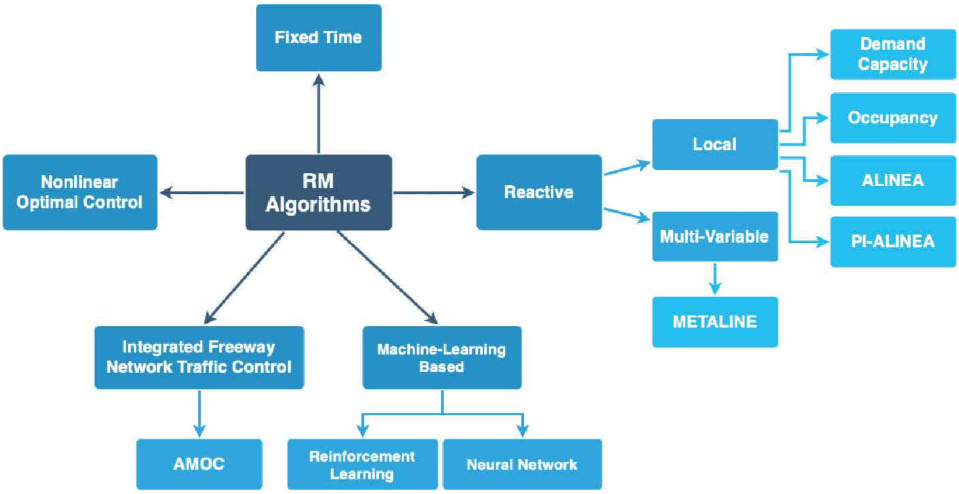


Figure 1.
Ramp metering algorithms classification.

A classification of ramp metering algorithms based on a study by Papageorgiou and Kotsialos [17] is presented in **Figure 1**. Fixed time metering is the simplest strategy which is usually adjusted based on historical data and applied during particular times of day. Reactive ramp metering techniques are based on real time traffic metrics. Local ramp metering uses traffic measures collected form the ramp vicinity. Demand-capacity, and occupancy-based strategies allow as much traffic inflow as possible to reach the freeway capacity. ALINEA and PI-AlINEA offer a more complex and more responsive strategy that, unlike capacity and occupancy strategies, generates smoother responses towards changes in metrics. Multi-variable regulator strategies perform the same as local strategies, but more comprehensively and independently on a set of ramps and usually outperform local strategies. METALINE can be viewed as a more general and extended form of ALINEA. Nonlinear optimal control strategy considers local traffic parameters and metrics as well as nonlinear traffic flow dynamics, incidents, and demand predictions in a freeway network and outputs a consistent control strategy. Knowledge-based control systems are developed based on historical data and human expertise. Integrated freeway network traffic control is a more general approach to nonlinear control that extends application of optimal control strategies to all forms of freeway traffic control. In case of knowledge based systems, inability to learn and adapt to temporal evolution of the system being controlled can be an issue, so knowledge based systems need to be periodically updated to remain efficient. Artificial intelligence and machine learning approaches like reinforcement learning (RL) and artificial neural networks (ANN) are new techniques being implemented recently for RM control [18].

Two of the most common ramp metering strategies are described in the following. Here, the flow that can enter section i of a freeway from the on-ramp during time interval $[kT, (k+1)T)$ is shown by $r_i^{C}(k)$ where k is the time index and T is the sampling time.

2.2.1 ALINEA

ALINEA [19] is an I-type controller in which the metering rate is given by

$$r_i^{C}(k) = r_i^{C}(k-1) - K_R\left[\rho_i^* - \rho_i^{down}(k)\right] \tag{27}$$

where $\rho_i^{down}(k)$ is the density measured downstream the on-ramp, ρ_i^* is a set-point value for the downstream density, and K_R is the integral gain. Note that, in case the main objective of the traffic controller is to reduce congestion and to maximise the throughput, a good choice for the set-point is $\rho_i^* = \rho_i^{cr}$.

2.2.2 PI-ALINEA

A very famous extension of ALINEA is the PI-ALINEA, in which a proportional term is added to result in a PI regulator. The metering rate is given by

$$r_i^{\mathrm{C}}(k) = r_i^{\mathrm{C}}(k-1) - K_P\left[\rho_i^{down}(k) - \rho_i^{down}(k-1)\right] + K_R\left[\rho_i^* - \rho_i^{down}(k)\right] \tag{28}$$

where K_P is another regulator parameter.

Based on the stability analysis of the closed-loop ramp metering system provided in [20], it can be stated that PI-ALINEA is able to show a better performance than ALINEA.

2.3 Route guidance

Route guidance (RG) is an efficient technique to distribute the traffic demand over the network, by providing information about alternative paths to drivers. The variable message signs (VMSs) are one of the main actuators which can provide route information to drivers in the RG control scheme. In the RG control, the concepts of equilibrium play an important role. Wardrop has offered the two following principles [21]: (1) The *system optimum (SO)* is achieved when the vehicles are guided such that the total costs of all drivers (typically the TTS) is minimized, (2) The traffic network is in *user equilibrium (UE)* when the costs on each utilized alternative route is equal and minimal, and on routes that are not utilized, the cost is higher that on the utilized routes.

If the goal of a control strategy is defined as the travel time, it is typically defined as the *predicted travel time* or as the *instantaneous travel time*. The predicted travel time is the time that the driver will experience when he drives along the given route, while the instantaneous travel time is the travel time determined based on the current speeds on the route. In a dynamic setting, the instantaneous travel time may be different from the predicted travel time [5].

In route guidance control strategies, the control variable is the splitting rate at a given node. Considering the simple case of only two alternative paths [22], originating from node n, let us denote with m and m' the two links exiting node n, corresponding respectively to the primary and secondary path. The primary path is the one characterised by the shortest travel time, in case of regular traffic conditions. In particular, the control variable is the splitting rate $\beta_{m,n,j}^{\mathrm{C}} \in [0,1]$, representing the portion of flow present in node n at time instant kT which should choose link m to reach destination j. The other control variable is $\beta_{m',n,j}^{\mathrm{C}}$, referred to link m', where it is easily computed: $\beta_{m',n,j}^{\mathrm{C}} = 1 - \beta_{m,n,j}^{\mathrm{C}}$. The following feedback regulators of P-type or PI-type are the most used strategies for route guidance systems in the literature [5, 23]. According to a proportional control law, the portion of flow present in node n at time instant kT which should choose link m to reach destination j is computed as

$$\beta_{m,n,j}^{\mathrm{C}}(k) = \beta_{m,n,j}^{N}(k) + K_P \Delta\tau_{n,j}(k) \tag{29}$$

where $\beta^{N}_{m,n,j}(k)$ is the nominal splitting rate, K_P is a gain, $\Delta\tau_{n,j}(k)$ is the instantaneous travel time difference between the secondary and primary direction from n to j. In proportional-integral regulators, the splitting rate is

$$\beta^{C}_{m,n,j}(k) = \beta^{C}_{m,n,j}(k-1) + K_P\left[\Delta\tau_{n,j}(k) - \Delta\tau_{n,j}(k-1)\right] + K_I\Delta\tau_{n,j}(k) \qquad (30)$$

where K_P and K_I are other controller gains.

Another possible class of RG strategies is *iterative strategies*, where the splitting rate is computed by iteratively running different simulations in real time with different RG, in order to achieve conditions of either user equilibrium or system optimum [24, 25]. Iterative strategies are very beneficial, however, their high computational effort is a major drawback for this category of RG control techniques.

It is also interesting to describe how drivers react to travel time information and how they adapt their route choice. A well-known behavior model used for this purpose is the logit model [26], which is used to model all kinds of consumer behavior based on the cost of several alternatives. The lower the cost of an alternative, the more consumers will choose that alternative. In the case of traffic management, consumers are the drivers, and the cost is the comfort, safety, or travel time of the alternative routes to reach the desired destination. The logit model calculates the probability that a driver chooses one of more alternatives based on the difference in travel time between the alternatives. Assume that we have two possible choices m_1 and m_2 at node n to get to destination j. For the calculation of the split rates out of the travel time difference between two alternatives, the logit model results in:

$$\beta_{m,n,j}(k) = \frac{\exp\left(\sigma\theta_{n,m,j}(k)\right)}{\exp\left(\sigma\theta_{n,m_1,j}(k)\right) + \exp\left(\sigma\theta_{n,m_2,j}(k)\right)} \qquad (31)$$

for $m = m_1$ or $m = m_2$, where $\theta_{n,m_1,j}(k)$ is the travel time shown on the DRIP at node n to travel to destination j via link m. The parameter σ describes how drivers react on a travel time difference between two alternatives. The higher σ, the less travel time difference is needed to convince drivers to choose the fastest alternative route.

2.4 Integration of ramp metering and route guidance

The RM and RG controllers are both feedback and predictive controllers as they apply not only on the real-time measurements of the system to calculate the control actions, but they also use information about the prediction of the system evolution. The combination of RM and RG controllers has shown promising results in network performance from the point of view of different performance measures. The focus of this section is on providing a review on the related studies on the combination of these two main traffic management techniques.

In 1990, a study performed by Iida et al. [27] considered the development of an improved on-ramp traffic control technique of urban expressway. They extended the conventional LP control method to consider the multiple paths between on-ramps and off-ramps of the test case network and also the route choice behavior of drivers. They assumed that in the future, the drivers would have the travel information offered by the route guidance system. Their formulation combined the user equilibrium with the available LP traffic control formulation at the time. In their problem statement, the goal was to determine the optimal metering rate so that the system measures of the network would be maximized, while the drivers would choose the path provided by the RG control. They discussed their mathematical formulation in detail and provided the solution finding algorithm.

In 1999 and then with some modifications in 2002, Apostolos Kotsialos et al. in [28, 29], considered the design of an integrated traffic control system for motorway networks with the use of ramp metering, motorway-to-motorway control, and route guidance. They offered a generic problem formulation in the format of a discrete time optimal control problem. They assumed that both RM control measures and RG are available. The METANET model was used for the description of traffic flow. A hypothetical test network was considered to evaluate the performance of the proposed control system. The control measure considered was the minimisation of the travel time spent (TTS). The results showed the high efficiency of the proposed control system.

In 2004, Karimi et al. [30] considered the integration of dynamic RG and RM based on MPC. They used the dynamic route guidance panels (DRIPs) as both a control tool and an information provider to the drivers, and ramp metering as a control tool to spread the congestion over the network. This resulted in a control strategy that reduced the total time spent by optimally re-routing traffic over the available alternative routes in the network, and also kept the difference between the travel times shown on the DRIPs and the travel times actually realized by the drivers as small as possible. The simulations done for the case study showed that rerouting of traffic and on-ramp metering using MPC has lead to a significant improvement in performance.

In 2015, Yu Han et al. [31] proposed an extended version of the CTM first proposed in [2] with the ability to reproduce the capacity drop at both the on-ramp bottleneck and the lane drop bottleneck. Based on this model, a linear quadratic model predictive control strategy for the integration of dynamic RG and RM was offered with the objective of minimizing the TTS of a traffic network. In this paper, a RG model based on the perceived travel time of each route by drivers was also offered. If the instantaneous travel time of each route is provided to travelers, the perceived travel time is assumed to be the same as the instantaneous travel time. If not, the perceived travel time is assumed to be the free flow travel time. The splitting rates at a bifurcation are determined by the well-known Logit model [26, 30]. A test case network containing both on-ramp bottlenecks and lane drop bottlenecks was used to investigate the effectiveness of the proposed framework and the results showed the improvement the proposed control strategy brought for the network performance.

In 2017, Cecilia Pasquale et al. [22] offered a multi-class control scheme for freeway traffic networks with the integration of RM and RG in order to reduce the TTS and the total emissions in a balanced way. Their two controllers were feedback predictive controllers and it was shown how this choice for their controllers can benefit the performance of the controllers. They applied the multi-class METANET model and the multi-class macroscopic VERSIT+ model for prediction of the traffic dynamics in the network. In addition, they designed a controller gain selector to compute the gains of the RM and RG controllers. The simulation results showed significant improvements of the freeway network performance, in terms of reduction of the TTS and the total emissions.

In 2018, Hirsh Majid et al. [32] designed integrated traffic control strategies for highway networks with the use of RG and RM. The highway network was simulated using the LWR model. A control algorithm was designed to solve the proposed problem, based on the inverse control technique and variable structure control (super twisting sliding mode). Three case studies were tested in the presence of an on-ramp at each alternate route and where there was a capacity constraint in the network. The objective was to avoid congestion on the main road and to balance the traffic flow on the alternate routes. The obtained results showed that the proposed algorithms could establish user equilibrium between two alternate routes even when the on-ramps have different traffic demands.

In 2019, Martin Gregurić et al. [33] proposed the approach of coordination between controlling on-ramp flows with ramp metering (RM) and dynamic route guidance information systems (DRGIS), which reroute vehicles from congested parts of the motorway. DRGIS is used to inform drivers about current or expected travel times and queue lengths so that they may reconsider their choice for a certain route. It can be seen that DRGIS can directly impact on traffic demand at the urban traffic system by informing the drivers about travel times on its crucial segments. Reduced traffic demand on congested urban motorway section or at congested on-ramp in coordination with the adequate ramp metering control strategies can prevent "spill-back effect" and increase overall throughput of the urban motorways.

2.5 Summary

To conclude, in this section, a review on the integration of RM and RG controllers was presented. The section started by describing the two most commonly used discrete first-order and second-order traffic flow models. Then, it continued with an overview of the popular RM and RG control strategies and it finished by discussing most of the important studies on the integration of RM and RG. Most of them considered TTS as the performance metric and applied an MPC optimization framework since it reduces the computation efforts required to solve the optimization problem specially if the traffic evolution model used was the METANET model since it makes the formulation non-linear and non-convex. Overall, all the studies reviewed here have shown improvements in the network performance in comparison with the case of having either of these controllers alone.

The discussion so far has centered around the macro-level control of intelligent highways. However, an integral component of the transportation system of a smart city is the micro-level control of autonomous vehicles. It is, therefore, imperative that we cover some micro-level details pertaining to the CAVs. The following sections address problems related to the micro-level vehicle coordination. Commonly found interactions include highway merging, off-ramp exit, vehicle overtaking and lane changing. Focus is placed on on-ramp merging and overtaking in presence of incoming traffic, as these two tasks combined encompass most of the complexities involved in inter-vehicle interactions.

The selection of the on-ramp merging task (see Section 3) as a key area to be explored is due to both the extent of variables involved in coordinating this process and the dynamic nature of the process itself. In fact this is one of the tasks that today's autonomous vehicles find difficult to carry out due to the need of reactive control and precise planning. The role of inter-vehicle coordination in efficient merging is also discussed in detail. While coordination of human-driven vehicles is mostly reactive, CAVs can be assigned goals proactively so as to optimize the merging process. Similarly, a detailed discussion is provided on the car overtake problem (see Section 4) to emphasize the need for explicit modelling of human behavior when designing algorithms for CAVs. This seemingly simple problem is specifically chosen to draw attention to the complexities that could arise due to the presence of human drivers on the road. A naive data-driven algorithm can fail catastrophically in scenarios where humans may behave unpredictably so an overview of algorithms that explicitly take human behavior into consideration is provided later on in this chapter.

3. Merging behavior at on-ramps and off-ramps for CAVs

In transportation networks, overall highway system efficiency can be severely reduced due to delays caused at on-ramps and off-ramps. If the merge process is incorrectly handled, merging lanes can have an overflow effect which causes the entire highway to become congested. This effect is caused by slower moving vehicles facing congestion in outer lanes near the merge junction deciding to switch into inner lanes in order to move faster. Therefore, even vehicles in the inner high speed lanes have to slow down. Overtime with continuing merge lane congestion, the entire freeway can become blocked.

In fact, in human-driven vehicles, this issue is further compounded due to the lack of cooperation and limited visibility for decision making. However, the introduction of CAVs has led to a lot more information becoming available for improving this overall merging process. In addition to the improved local sensing on-board modern CAVs such as 360° radar and vision based sensors, most of the increased information comes from improvements in Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communication methods. These communication protocols enable individual vehicles to broadcast their intent and status, receive command velocities to enable optimal flow and also collaborate with each other to self-organize in such a way that freeway traffic flow is optimized.

Effectively managing the highway merge problem has multiple benefits, both to the end users and the entire transportation system as well. Reduced time in traffic at merge junctions means that overall throughput of the highway is increased, individual waiting time is reduced and the wastage of fuel (energy) in idling vehicles stuck in traffic is also reduced. Therefore, it is evident that improvements to intelligent highways will have an impact on the economy as well as helping combat environmental issues caused by excessive fuel consumption.

3.1 Standard problem formulation

Most approaches to solving the highway merging problem use a similar structure to model the physical highway on-ramp. **Figure 2** shows an abstracted model of a highway on-ramp. A control zone is defined, where all vehicles in this zone communicate with a central controller and each other to decide individual optimum

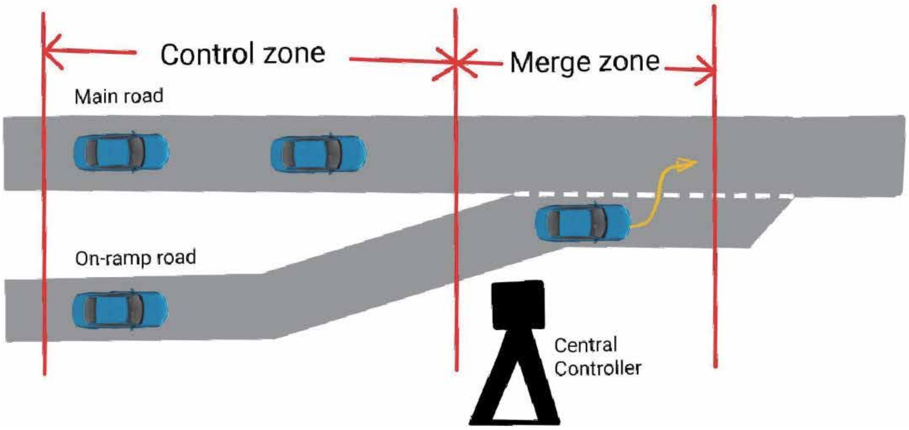


Figure 2.
On-ramp merging regions and infrastructure model.

velocities and paths to be followed. The control zone encompasses both the main road and the on-ramp. Vehicles are allowed to merge from the on-ramp onto the main road in the merge zone, which is located at the end of the control zone.

The vehicles involved are modelled based on simplistic second order dynamics given by,

$$\begin{aligned} \dot{p}_i &= v_i(t) \\ \dot{v}_i &= u_i(t) \end{aligned} \tag{32}$$

where $p_i(t)$, $v_i(t)$, $u_i(t)$ denotes the position, velocity and acceleration/ deceleration (control input) respectively for each vehicle. The vehicle state is then defined as,

$$x_i(t) = \begin{bmatrix} p_i(t) \\ v_i(t) \end{bmatrix} \tag{33}$$

Furthermore, assuming that lateral control keeping the vehicle in lane is managed elsewhere, the vehicle can be modelled as a point mass moving along the center of the lane with the following state equation, where time 0 is the point at which the vehicle enters the control zone.

$$\dot{x}_i = f(t, x_i, u_i), \qquad x_i\left(t_i^0\right) = x_i^0 \tag{34}$$

Additionally, all the methods discussed have the shared assumption that, *the vehicle speed inside the merging zone is constant*.

To compare these algorithms, the two main performance indicators are *throughput* (maximum number of vehicles that can merge onto the highway in an hour) and *delay* (average delay experienced by vehicles compared to the ideal travel time). In addition to these parameters, some research in this area also takes into consideration the savings in fuel consumption due to improvements in the highway merging process.

3.2 Various methods

So, let's now explore some of the methods used in handling the highway merging problem in greater detail. Here, multiple approaches and methodologies to address this problem are discussed. Most of the work done in finding an optimal solution to automated freeway merging is based on posing the problem in the form of an optimization problem [34] with centralized control [35], virtual slot-based dynamics [36] problem or as broadcast communication [37] problem.

3.2.1 Optimal control method

This problem is posed as an unconstrained optimization problem [34] and then further extended to also consider the impact of fuel consumption [35]. The output was the ability to derive online an optimal closed-form solution for vehicle coordination at a merge intersection. The importance of safety constraints is also stressed here. Constraints in positioning, maximum velocity and maximum accelerations are imposed. Additionally, the algorithm only allows one vehicle to enter into the merging zone at any time.

The algorithm calculates the time at which each vehicle would enter the merge zone and requires that this time does not conflict with any of the other vehicles.

This ensures that a lateral collision can never happen. Additionally, the algorithm also directs that vehicles maintain at least a specified gap between each other which ensures that rear-end collisions do not occur. Hamiltonian analysis is then used to convert the optimization problem into a system of four equations that can be solved in real time to output the optimal control for each vehicle.

Simulation of this system was then carried out to show that the algorithm performs as desired. It was found that compared to a baseline situation where on-ramp vehicles always give way to vehicles on the freeway this algorithm performs significantly better. An improvement of 52% in fuel consumption when compared to the baseline situation was also reported.

Disadvantages: Only one lane of the freeway is in use and the benefits obtained from allowing/forcing vehicles to switch lanes in the freeway are ignored (i.e. full capacity of the freeway is not used). Additionally, vehicles are given merging rights based on a simplistic FIFO (First In First Out) queue which can cause additional delays and is definitely sub-optimal.

3.2.2 Slot based method

This method [36] primarily relies on creating virtual slots for each vehicle that moves along the freeway at a constant velocity. Then all changes to this behaviour such as switching lanes, on-ramp merging and exiting the highway on an off ramp are modelled as a switch from one virtual slot to another. A virtual slot S is defined with five properties as is denoted by $S = \{z, p, t, b, o\}$, where z is the size of the slot, p is the position, t is the time, b is the behaviour of the slot and o is the density status of the slot.

Slots are created by a central slot controller and vehicles can request to change from one slot to another. This change will then be approved by the slot controller as long as the slot is not already occupied or there is no other vehicle requesting to switch to that same slot. This approach was also shown to work in the absence of infrastructure at the merge junction since vehicles can use V2V communication to find an unoccupied virtual slot and perform the merging task.

An additional benefit of this slot based system is the ease by which the method can be extended to allow the entire bandwidth of the freeway to be used in order to further improve efficiency. For example, if the slot controller realizes that there are a lot of empty slots in the central lanes of the freeway, the controller can request that vehicles in the outer lane prior to the merge point to move into the empty inner lane slots. This creates more empty slots in the outer lane and provides more opportunities for vehicles on the on-ramp to merge successfully. This type of cooperative behaviour has been proven to drastically improve the throughput of vehicles through these freeway merge zones.

Using simulations, it has been shown that throughput at merge intersections can be increased and delay can be decreased drastically (throughput: 230% increase and delay: 452% decrease) vs. human driven vehicles under heavy traffic conditions.

Disadvantages: Many limitations brought about by having slot based systems include, difficulty in robustly handling emergency/breakdown situations, lack of flexibility in catering to different needs such as different vehicles requesting different speeds, inefficiency in heavy traffic density situations where not enough free slots are available to facilitate lane changes etc. Moreover, the slot based method places a lot of restrictions on the way vehicles can move about and position themselves on the freeway. Communication between vehicles and infrastructure also needs to be extremely good for this system to work and this type of perfect communication is rarely available in practice.

3.2.3 Broadcast communication based method

Some of the major issues in coordinating automated vehicles at merge intersections are the problems caused by imperfect communication. In these type of applications, delays of even a few seconds can have devastating results. Work on using Pseudo-perturbation based broadcast communication (PBC) [37] instead of unicast communication focuses on reducing the overhead on the V2I communication systems and leveraging the capabilities of V2V communication to handle any short term changes. In this method, a global controller (infrastructure) broadcasts an identical message to all vehicles in its vicinity. Each of these vehicles uses this information along with V2V data from other vehicles to select a suitable control strategy to safely perform the coordinated merging task. Vehicles send updates back to the central controller and the controller uses this feedback to decide its next broadcast message.

The key focus of this research was to extend available PBC capabilities to handle the multi-state vehicle dynamics and the multi-objective optimization required to solve a freeway merging coordination problem. The capability of this system to handle both CAVs as well as a mix of CAVs and human driven vehicles was also showcased. Here, the CAVs are controlled by the coordination system and are shown to be able to function even in the presence of human driven vehicles. The smoothness of the actual merging process was also evaluated through simulation. The main output of this research was to show that complex coordination problems such as freeway merging can be successfully solved with the use of minimal communication bandwidth.

Disadvantages: Lack of focus on the actual merging algorithm. Based on the broadcast signals, the decisions individual vehicles make may be sub-optimal and cause an unnecessary delay in the system. Also, very little work has been done on seeing whether this system actually has a effect on throughput.

3.2.4 Temporal logic based method

This method involves formalizing traffic rules using temporal logic [38] in order to ensure safety and robustness of automated highway vehicle control. This method helps with formulating existing traffic rules in a mathematical way in order to be easily applied in CAVs. While this method does not solely focus on the merging problem, the merge window is addressed in the metric temporal logic (MTL) formulas included. When the merge operation is formulated as a MTL formula, it leads the way to specifying safety guarantees in autonomous vehicles. Furthermore, the legality of trajectories generated by a motion planner can be easily checked using these MTL formulae. It also allows the use of standard verification and validation methods in order to ensure that there are no loopholes or issues in the generated logic.

3.3 Summary

While there are many methods to improve intelligent merging behavior, the core fundamentals of these algorithms are quite similar. They look into minimizing gaps or under-utilized space on the road while minimizing the control inputs (acceleration and braking) needed to achieve this. All algorithms also prioritize safety and fairness in the merging process. Additionally, there are extensions that focus on maximum capacity utilization by moving vehicles already on the highway to less congested lanes, which further improves the efficiency of the merging process. Each of the algorithms discussed in the section has its own advantages and disadvantages. Therefore, it falls on the highway regulatory agencies to decide which of these are most suitable to the conditions of each individual highway system.

4. The overtaking behavior with the combination of HDVs and CAVs

The final section of this chapter focuses on providing an in-depth analysis of a complex scenario in which the autonomous vehicle (AV) has to perform maneuvers in the presence of HDVs. Upon a cursory glance at the current state of AV research [39], it becomes obvious that not a lot of emphasis is being placed on explicitly modelling the varying behavior patterns of HDVs on the road. One such instance that highlights the need to model the varying HDV driving patterns is the car overtaking problem in a bidirectional traffic flow setting. In this problem, a scenario with three vehicles is considered; two HDVs and one AV (ego vehicle), as shown in **Figure 3**. The HDVs are travelling in opposite directions in adjacent lanes and the ego vehicle is following one of the HDVs. The objective of the AV is to safely overtake the vehicle travelling ahead while maintaining safety distances to the HDV in the adjacent lane, the HDV travelling ahead and the boundaries of the road.

This is a particularly hard problem to solve because it involves a scenario composed of both human-driven and autonomous vehicles. The first major complication is the lack of global information because the V2X communication protocols cannot be leveraged in this scenario due to the presence of HDV. Then, there is the problem of uncertainty that arises due to the varying driving patterns so the algorithm needs to be robust enough to handle the different driving patterns of human drivers. Moreover, the traditional Supervised Learning based approaches cannot be applied directly to this problem due to a lack of labeled training data. Finally, the model-free Reinforcement Learning (RL) based techniques [40] cannot be employed due to a lack of safety guarantees while the model-based RL or Control techniques [41] cannot be employed since they require an accurate representation of system model and cannot capture uncertainties that arise due to varying driving patterns well.

In this chapter, a simplified stochastic control based formulation taken from [42] is laid out to provide a mathematical description of the problem. The formulation is followed by a brief discussion of a couple of algorithms to give the readers a glimpse of the possible avenues that could be taken to reach a solution. The references to detailed resources are also provided for the interested readers to explore further.

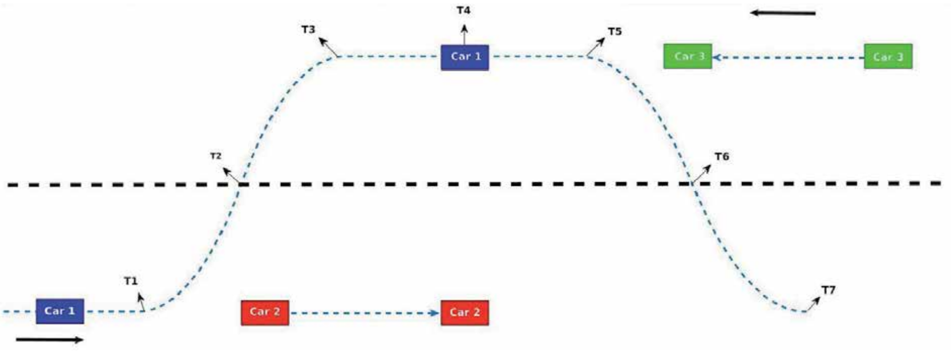


Figure 3.
Overview of the car overtaking problem.

4.1 Stochastic control formulation

In this approach, the ego vehicle (Car 1, from **Figure 3**) has to first decide if it is feasible to overtake the HDV ahead (Car 2) once it gets "close enough". In this simplified formulation, it is assumed that the Car 1 is able to measure its own relative velocity with respect to the HDVs with some additive noise. If the decision to overtake is made, then the AV has to generate the exact trajectory it will take to perform the overtake maneuver. An overview of these steps is outlined henceforth.

In the formulation below, it is assumed that the width of each of the lanes is defined to be a constant value d and the minimum safety distance between cars is L. A collision is defined in terms of AV violating the minimum safety distance threshold with respect to the centroids of the cars. The positive velocities are defined towards the right in **Figure 3** and positive θ is defined counterclockwise relative to the velocity vector of the car. Finally, the sets of admissible linear and angular speeds for the cars are considered to be finite.

4.1.1 Modelling

The states of Car i are x_i, y_i and θ_i which respectively correspond to the longitudinal coordinate, lateral coordinate and orientation of Car i. The inputs to Car i are v_i and w_i, the linear and angular velocities respectively of Car i respectively. An index k is used to denote the time step.

Considering the initial states of the vehicles, it is assumed that initially at $k = 0$, Car 1's longitudinal coordinate is a random variable distributed normally with $\mu = 0$ and $\sigma = \Sigma_1$ while the lateral coordinate is fixed at the center of the bottom lane i.e. $y_1(0) = d/2$ and facing forward i.e. $\theta = 0$. As for Car 2, it is assumed that initially at $k = 0$, x_2 is distributed normally with $\mu_2 = \tilde{x}_2$ and $\sigma = \Sigma_2$ where Σ_2 is s.t. $x_2(0) > x_1(0)$ while the lateral coordinate is at $y_2(0) = d/2$ and facing forward i.e. $\theta = 0$, identical to Car 1. Finally, for Car 3, it is assumed that initially at $k = 0$, x_3 is distributed normally with $\mu_3 = \tilde{x}_3$ and $\sigma = \Sigma_3$ where Σ_3 is s.t. $x_3(0) > x_2(0)$ while the lateral coordinate is fixed at the center of the top lane i.e. $y_3(0) = 3d/2$ and facing reverse (since $v_3 < 0$) i.e. $\theta = 0$.

As for the dynamics of the vehicles, it is assumed that cars 2 and 3 keep travelling along the same lane i.e. ω_2 and ω_3, the angular velocities of cars 2 and 3 respectively, are identically zero for all time steps $k \geq 0$.

Based on the assumptions above, the dynamics for the cars are defined by the equations below:

$$x_1(k+1) = x_1(k) + v_1(k)\cos(\theta_1(k)) \tag{35}$$

$$y_1(k+1) = y_1(k) + v_1(k)\sin(\theta_1(k)) \tag{36}$$

$$\theta_1(k+1) = \theta_1(k) + \omega(k) \tag{37}$$

$$x_2(k+1) = x_2(k) + v_2(k) \tag{38}$$

$$y_2(k) = d/2 \tag{39}$$

$$\theta_2(k) = 0 \tag{40}$$

$$x_3(k+1) = x_3(k) + v_3(k) \tag{41}$$

$$y_3(k) = 3d/2 \tag{42}$$

$$\theta_3(k) = 0 \tag{43}$$

To keep the problem as general as possible, it is assumed that at time step k, the AV has all the history of its past states, linear velocity and its relative position and velocity w.r.t the other two cars with some additive white Gaussian noise (AWGN). Therefore,

$$\mathcal{I}(k) = \left\{x_1(n), y_1(n), \theta_1(n), v_1(n), z_1(n), z_2(n), z_3(n), z_4(n)\right\}_{n=0}^{n=k} \tag{44}$$

where

$$z_1(k) = x_2(k) - x_1(k) + n_1(k) \tag{45}$$

$$z_1(k) = x_3(k) - x_1(k) + n_2(k) \tag{46}$$

$$z_1(k) = v_2(k) - v_1(k) + n_3(k) \tag{47}$$

$$z_1(k) = v_3(k) - v_1(k) + n_4(k) \tag{48}$$

Here, n_i(k) are white Gaussian Processes with mean 0 and variances $\sigma_i^2(k)$. There is an assumption placed on the independence of $\{n_i(k)\}_{k\geq 1}$ from the distribution of initial longitudinal coordinates of the three cars i.e. $x_1(0)$, $x_2(0)$ and $x_3(0)$.

4.1.2 Control problem

Upon consideration of the control problem that the AV has to solve, it becomes apparent rather quickly that the AV simply doesn't have to decide on linear and angular velocities. If that were the case, then the Car 1 could simply wait till Car 3 passes and then overtake Car 2 by having a velocity greater than Car 2. That is not an optimal solution for all possible scenarios. Therefore,the first thing that the AV needs to do is to get better estimates of position and velocity of Cars 2 and 3 rather than using the raw noisy data. With the better estimates of positions and velocities of the HDVs, the AV can perform a feasibility analysis to see if it is feasible to overtake the Car 2 or not. If the Car 1 deems the overtake maneuver to be infeasible, it can resort to the waiting strategy. If, however, Car 1 decides to overtake, then it has to decide on the time to start overtaking. Once the time to start overtaking is finalized, then the AV has to plan a trajectory that it will take to perform the maneuver such that it will not violate any safety margins. Finally, it needs to generate control commands i.e. linear and angular velocities to execute the overtaking maneuver. One typical problem that could arise is that the vehicle may have to return to its original lane, after starting the overtaking maneuver, but this scenario is beyond the discussion of this chapter.

4.1.3 Sample trajectory

A sample trajectory with a constant linear velocity for the AV is displayed in **Figure 4**. Between $k = 0$ and $k = T1$, Car 1 is approaching Car 2. At $k = T1$, Car 1 decides to run the feasibility analysis and decides to overtake Car 2. Between $k = T1$ and $k = T2$, Car 1 has a positive constant angular velocity resulting in motion towards the adjacent lane. At $k = T2$, Car 1 has reached the divider between the lanes so it switches to a negative angular acceleration having the same magnitude as

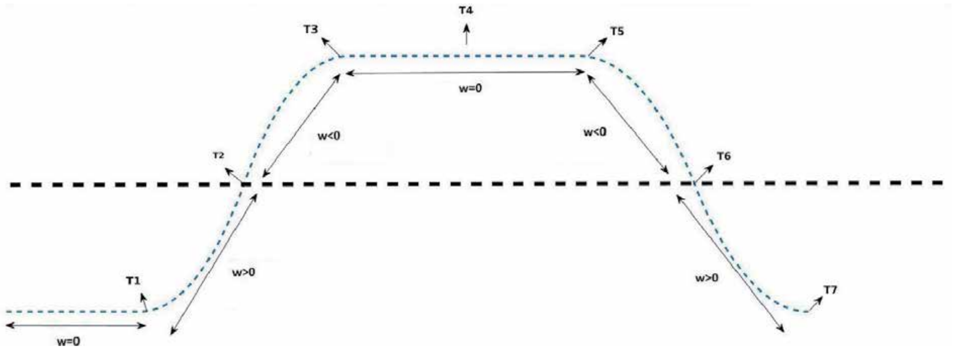


Figure 4.
Sample trajectory for car overtaking problem.

before, until the car reaches the center of adjacent lane at $k = T3$. Between $k = T3$ and $k = T5$, the angular acceleration remains at 0 for a straight line motion in opposite lane for overtaking. A negative angular acceleration having the same magnitude as before is applied between $k = T5$ and $k = T6$ followed by positive angular acceleration between $k = T6$ and $k = T7$. Car 1 continues its motion in a straight line after $k = T7$ with zero angular acceleration.

4.2 Solution methods

There were quite a few assumptions made to obtain the simplified model discussed above and some of those assumptions could be removed in order to obtain a rather complicated yet general framework. For instance, it might not be possible for the AV to store all the history of past states and actions so an attempt at modelling with limited information could be made. With this disclaimer, a brief overview of some of the possible solution methods for this problem is presented below.

4.2.1 Minimizing probability of collision

In this approach, taken from [42], it is assumed that the AV will travel at a constant speed throughout the overtaking maneuver. The estimates of positions and velocities of HDVs are obtained using Kalman filtering and a constant N is introduced to characterize the behavior of the AV with higher values corresponding to higher level of aggressiveness. Two feasible sets for the linear and angular velocities respectively of Car 1 are obtained by ensuring that the estimated position of AV after performing the maneuver will stay outside the estimated minimum safety region around the HDVs. Using the feasibility sets, probability of collision of Car 1 with the HDVs is obtained and the decision to overtake is based on those probabilities. If the decision to overtake is made, the linear and the angular velocities are chosen to minimize the probability of collision and the maneuver is performed as detailed in Section 4.1.3.

4.2.2 Reachability analysis-based with martingale-based HDV modelling

In this approach, taken from [43], the focus is on obtaining safety guarantees while overtaking. In this approach, the restriction on the constant speed of HDVs is also lifted, which was alluded to previously in Section 4.2. There are two different reachability analysis-based algorithms presented: one is a robust time-optimal algorithm such that it provides strict guarantees in regards to collision avoidance while the other is a stochastic algorithm that that yields a small collision probability with the advantage of shorter overtaking time. Moreover, the expected behavior of the human driver is modeled using a stochastic model based on martingales. It is shown that if the human driver is non-aggressive, the stochastic algorithm will yield a shorter overtaking time and if the driver is aggressive, the behaviors for stochastic and the robust algorithms will be identical.

4.3 Summary

The key takeaway from this section is that there is a need to place focus on explicitly considering the role of human drivers on the road while developing algorithms for autonomous vehicles. It was shown by the study of the simple car overtaking example that there are scenarios where the need to model human drivers on the road increases manifold. The algorithm presented in Section 4.2.2 explicitly models the behavior of human-drivers with martingales and provides overtaking algorithms with safety guarantees. The solution complexity of this approach is rather high due to the reachability analysis-based solutions so further research could be directed at improving the solution complexity or coming up with approaches that yield lower complexity while maintaining the safety guarantees. Furthermore, there is a prospect of exciting research in the direction of modeling human behavior with other approaches which may lead to various other interesting algorithms. The aim of this section was to provide motivation to the reader to explore research avenues that incorporate explicit modelling of human driving patterns yielding algorithms that will expedite the introduction of Autonomous Vehicles onto our roads.

5. Conclusions

The advent of improved communication, sensing and control in modern day vehicles and infrastructure creates a lot of opportunities to improve the efficiency and safety in many highway processes. Key areas of interest involve traffic routing and management, optimal highway merging and intelligent overtaking behaviours. This chapter examined some of the methods used in these areas and discussed the various improvements and shortcomings of each of them. The implementation of the algorithms discussed in this chapter, would lead to modern transportation systems becoming more effective, productive and safe. While there are many other methods worthy of merit not discussed in this chapter, the areas covered should give the reader a broad understanding of the extent of possibilities in this field and also spark further thinking which may lead to the generation of innovative new solutions.

Author details

Fatemeh Alimardani, Nilesh Suriyarachchi, Faizan M. Tariq
and John S. Baras*
Department of Electrical and Computer Engineering, Institute for Systems
Research, University of Maryland, College Park, Maryland, USA

*Address all correspondence to: baras@umd.edu

References

[1] Lighthill MJ, Whitham GB. On kinematic waves I. Flood movement in long rivers. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences. 1955 May 10;**229**(1178):281-316

[2] Daganzo C. The cell transmission model Part I: a simple dynamic representation of highway traffic. PATH Report, 93-0409. 1993;3.

[3] Papageorgiou M, Blosseville JM, Hadj-Salem H. Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris. Transportation Research Part B: Methodological. 1989 Feb 1;**23**(1):29-47

[4] Lu XY, Qiu TZ, Horowitz R, Chow A, Shladover S. METANET model improvement for traffic control. In 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC) 2011 Oct 5 (pp. 2148-2153). IEEE.

[5] Ferrara A, Sacone S, Siri S. Freeway traffic modelling and control. Berlin: Springer; 2018 Apr 12

[6] Gomes G, Horowitz R. Optimal freeway ramp metering using the asymmetric cell transmission model. Transportation Research Part C: Emerging Technologies. 2006 Aug 1; **14**(4):244-262

[7] Ferrara A, Sacone S, Siri S. Event-triggered model predictive schemes for freeway traffic control. Transportation Research Part C: Emerging Technologies. 2015 Sep 1;**58**:554-567

[8] Srivastava A, Geroliminis N. Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model. Transportation Research Part C: Emerging Technologies. 2013 May 1;**30**: 161-177

[9] Roncoli C, Papageorgiou M, Papamichail I. Traffic flow optimisation in presence of vehicle automation and communication systems–Part I: A first-order multi-lane model for motorway traffic. Transportation Research Part C: Emerging Technologies. 2015 Aug 1;**57**: 241-259

[10] Daganzo CF. The cell transmission model, part II: network traffic. Transportation Research Part B: Methodological. 1995 Apr 1;**29**(2):79-93

[11] Muralidharan A, Dervisoglu G, Horowitz R. Freeway traffic flow simulation using the link node cell transmission model. In2009 American Control Conference 2009 Jun 10 (pp. 2916-2921). IEEE.

[12] Canudas-de-Wit C, Ferrara A. A Variable-Length cell road traffic model: Application to ring road speed limit optimization. In2016 IEEE 55th Conference on Decision and Control (CDC) 2016 Dec 12 (pp. 6745-6752). IEEE.

[13] Messner A, Papageorgiou M. METANET: A macroscopic simulation program for motorway networks. Traffic engineering & control. 1990;**31** (8-9):466-470

[14] Pasquale C, Sacone S, Siri S. Two-class emission traffic control for freeway systems. IFAC Proceedings Volumes. 2014 Jan 1;**47**(3):936-941

[15] Pasquale C, Sacone S, Siri S, De Schutter B. A multi-class model-based control scheme for reducing congestion and emissions in freeway networks by combining ramp metering and route guidance. Transportation Research Part C: Emerging Technologies. 2017 Jul 1;**80**: 384-408

[16] Zhang M, Kim T, Nie X, Jin W, Chu L, Recker W. Evaluation of on-ramp control algorithms.

[17] Papageorgiou M, Kotsialos A. Freeway ramp metering: An overview. IEEE transactions on intelligent transportation systems. 2002 Dec;**3**(4): 271-281

[18] Sanandaji A, Ghanbartehrani S, Mokhtari Z, Tajik K. A Novel Ramp Metering Approach Based on Machine Learning and Historical Data. arXiv preprint arXiv:2005.13992. 2020 May 26.

[19] Papageorgiou M, Hadj-Salem H, Blosseville JM. ALINEA: A local feedback control law for on-ramp metering. Transportation Research Record. 1991 Jan;**1320**(1):58-67

[20] Wang Y, Kosmatopoulos EB, Papageorgiou M, Papamichail I. Local ramp metering in the presence of a distant downstream bottleneck: Theoretical analysis and simulation study. IEEE Transactions on Intelligent Transportation Systems. 2014 Mar 31; **15**(5):2024-2039

[21] Wardrop JG. Road paper. some theoretical aspects of road traffic research. Proceedings of the institution of civil engineers. 1952 Jun;**1**(3):325-362

[22] Pasquale C, Sacone S, Siri S, De Schutter B. A multi-class model-based control scheme for reducing congestion and emissions in freeway networks by combining ramp metering and route guidance. Transportation Research Part C: Emerging Technologies. 2017 Jul 1;**80**: 384-408

[23] Pavlis Y, Papageorgiou M. Simple decentralized feedback strategies for route guidance in traffic networks. Transportation science. 1999 Aug;**33**(3): 264-278

[24] Mahmassani HS, Peeta S. Network performance under system optimal and user equilibrium dynamic assignments: implications for ATIS. Transportation Research Board; 1993 Jan 10.

[25] Wang Y, Messmer A, Papageorgiou M. Freeway network simulation and dynamic traffic assignment with METANET tools. Transportation Research Record. 2001; **1776**(1):178-188

[26] Theil H. A multinomial extension of the linear logit model. International Economic Review. 1969;**10**(3):251-259

[27] Iida Y, Hasegawa T, Asakura Y, Shao CF. A formulation of on-ramp traffic control system with route guidance for urban expressway. IFAC Proceedings Volumes. 1990 Sep 1;**23**(2): 229-236

[28] Kotsialos A, Papageorgiou M, Messner A. Integrated optimal control of motorway traffic networks. In Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251) 1999 Jun 2 (Vol. 3, pp. 2183-2187). IEEE.

[29] Kotsialos A, Papageorgiou M, Mangeas M, Haj-Salem H. Coordinated and integrated control of motorway networks via non-linear optimal control. Transportation Research Part C: Emerging Technologies. 2002 Feb 1; **10**(1):65-84

[30] Karimi A, Hegyi A, De Schutter B, Hellendoorn H, Middelham F. Integration of dynamic route guidance and freeway ramp metering using model predictive control. In Proceedings of the 2004 American Control Conference (ACC 2004) 2004 Jun 30 (pp. 5533-5538).

[31] Han Y, Yuan Y, Hegyi A, Hoogendoorn SP. Linear quadratic MPC for integrated route guidance and ramp metering. In2015 IEEE 18th International Conference on Intelligent Transportation Systems 2015 Sep 15 (pp. 1150-1155). IEEE.

[32] Majid H, Lu C, Karim H. An integrated approach for dynamic traffic

routing and ramp metering using sliding mode control. Journal of Traffic and Transportation Engineering (English Edition). 2018 Apr 1;**5**(2):116-128

[33] Gregurić M, Mandžuka S, Vidović K. Advanced Applications for Urban Motorway Traffic Control. InInternational Conference "New Technologies, Development and Applications" 2019 Jun 27 (pp. 395-400). Springer, Cham.

[34] Rios-Torres J, Malikopoulos A, Pisu P. Online optimal control of connected vehicles for efficient traffic flow at merging roads. In 2015 IEEE 18th international conference on intelligent transportation systems 2015 Sep 15 (pp. 2432-2437). IEEE.

[35] Rios-Torres J, Malikopoulos AA. Automated and cooperative vehicle merging at highway on-ramps. IEEE Transactions on Intelligent Transportation Systems. 2016 Aug 5; **18**(4):780-789

[36] Marinescu D, ÄŒurn J, Bouroche M, Cahill V. On-ramp traffic merging using cooperative intelligent vehicles: A slot-based approach. In2012 15th International IEEE Conference on Intelligent Transportation Systems 2012 Sep 16 (pp. 900-906). IEEE.

[37] Ito Y, Kamal MA, Yoshimura T, Azuma SI. Multi-vehicle coordination on merging roads based on pseudo-perturbation-based broadcast control. In2018 Annual American Control Conference (ACC) 2018 Jun 27 (pp. 4008-4013). IEEE.

[38] Maierhofer S, Rettinger AK, Mayer EC, Althoff M. Formalization of Interstate Traffic Rules in Temporal Logic. Proc. of the IEEE Intelligent Vehicles Symposium. 2020

[39] Badue C, Guidolini R, Carneiro RV, Azevedo P, Cardoso VB, Forechi A, Jesus L, Berriel R, Paixao T, Mutz F, Veronese L. Self-driving cars: A survey. arXiv preprint arXiv:1901.04407. 2019 Jan 14.

[40] Li X, Xu X, Zuo L. Reinforcement learning based overtaking decision-making for highway autonomous driving. In2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP) 2015 Nov 26 (pp. 336-342). IEEE.

[41] Petrov P, Nashashibi F. Modeling and nonlinear adaptive control for autonomous vehicle overtaking. IEEE Transactions on Intelligent Transportation Systems. 2014 Mar 3; **15**(4):1643-1656

[42] Raghavan A, Wei J, Baras JS, Johansson KH. Stochastic control formulation of the car overtake problem. IFAC-PapersOnLine. 2018 Jan 1;**51**(9):124-129

[43] Gao Y, Jiang FJ, Johansson KH, Xie L. Stochastic Modeling and Optimal Control for Automated Overtaking. In 2019 IEEE 58th Conference on Decision and Control (CDC) 2019 Dec 11 (pp. 1273-1278). IEEE.

Chapter 7

Centralised Traffic Control and Green Light Optimal Speed Advisory Procedure in Mixed Traffic Flow: An Integrated Modelling Framework

Roberta Di Pace, Chiara Fiori, Luigi Pariota and Facundo Storani

Abstract

The paper aims to develop an integrated modelling framework for urban network traffic control in the presence of connected and autonomous vehicles (CAVs). The framework is further composed of two sub models: the first of which focuses on the traffic control problem in the case of hybrid flow conditions (unequipped vehicles and connected vehicles) and the second aims to control the automated vehicles in terms of speed optimisation. The traffic control strategy drew on the hybrid combination between the centralised approach based on a multi-objective optimisation and a link metering based on a single control function; whilst with reference to the speed guidance, the GLOSA (Green Light Optimal Speed Advisory) procedure was considered. Furthermore, the presence of connected vehicles has also been considered to support the estimation procedure of location and speed of unequipped vehicles. In terms of traffic flow modelling the microscopic approach has been applied. The proposed framework was applied by considering a simple real network (in the city centre of Naples, in the Southern of Italy) that was composed by one origin–destination pair and two alternative paths. The network layout is characterised by one diversion node and two alternative paths connecting the same origin - destination pair; three scenarios were tested: the first was only based on a centralised traffic control procedure, the second on speed guidance optimisation and the third was based on the combination of both sub-models. Finally, the framework effectiveness was analised in terms of within-day dynamics with respect to the travel times and queue length performance indices.

Keywords: centralised traffic control, speed guidance, multi-criteria optimisation, microscopic traffic flow modelling

1. Introduction and motivation

In general terms, the Intelligent Transportation Systems (ITS) have historically been introduced to increase the transportation networks performances allowing for the optimisation of several indicators which are strictly related such as travel time, emissions, consumption and safety. The effectiveness of all ITS proposed strategies is mainly based on the idea of traffic congestion predictions and drivers'/travellers' behaviour anticipation. Indeed, all relevant policies such as driving guidance, information systems design and traffic management are based on the consistency between the decision/control variables design and the actual traffic conditions (degree of congestion and travel times estimation). ITS solutions may generally be distinguished for urban and non-urban network applications; in particular, in the case of urban networks, the traffic control is one of the most suitable solutions to be applied, especially in the case of on-line traffic management.

However, in more recent times the ITS field has been integrated with cooperative services (Cooperative ITS; C-ITS). The main contribution is on the communication between vehicle and infrastructure (V2I), able to further optimise the vehicle driving behaviour along arterials in uninterrupted flow conditions and at junctions in the case of interrupted flow situations. The chapter's main focus is on strategies able to optimise the vehicles' behaviour at junctions, indeed, in accordance with literature, one of the proposed approaches is that of the GLOSA (Green Light Optimal Speed Advisory) which provides a warning to the driver regarding the best speed to maintain while approaching the junction by avoiding stops at junctions. A further classification of the GLOSA may be defined in literature in terms of MULTI-SEGMENT GLOSA (MS-GLOSA) if the optimisation strategy is applied at several/successive junctions and in terms of SINGLE-SEGMENT GLOSA (S-GLOSA) if the optimisation is applied only to the next traffic light that the driver will encounter along his/her trajectory.

The chapter aims to propose the integration between the GLOSA and the Traffic control strategy. The proposed framework has been applied at real case study; the considered subnetwork is composed by successive junctions and in terms of driving control the S-GLOSA procedure has been applied and this has been combined with the traffic control method. Regarding the traffic management a hybrid approach, suitable for urban network management, has been applied combining the centralised traffic control for urban networks (interacting junctions) and the link metering. The proposed simulation environment, based on Matlab/Simulink and SUMO, is a modular platform that considers the vehicle, the driver, the infrastructure and the traffic, the driving assistance systems, and finally the communication systems for cooperative driving; all components are simultaneously simulated in a whole environment.

As already anticipated the whole framework was analysed by considering an application to simple real network (in the city centre of Naples, in the Southern of Italy) that was composed by one origin–destination pair and two alternative paths.

Three scenarios were tested: the first was only based on a traffic control procedure, the second one concerned the speed guidance optimisation and the third was focused on the combination of both sub-models.

The remainder of the paper is organised as follows: in section 2 a brief overview of the literature is proposed; the whole modelling framework and the implementation settings, the traffic control problem and the Green Light Optimal Speed Advisory (GLOSA) procedure are displayed in section 3; in section 4 is presented the numerical application whilst results and future perspectives are discussed in section 5.

2. State of play

In this section an overview of the literature review regarding the Vehicle to Vehicle communication (V2V) and the vehicle to infrastructure communications (V2I) and in particular the driving assistance services, and the urban traffic control problem is provided.

2.1 Vehicle to vehicle (V2V) /to infrastructure communication (V2I): the driving assistance

The vehicle to vehicle communication is based on the idea that vehicles may exchange information about position, speed and location. In general, most relevant enhancements in the research field of the driving assistance refer to the cooperative awareness aiming to support the active road safety and the traffic efficiency to guarantee the speed management and the road navigation. A more detailed description is provided in the following.

Firstly, the Hazardous Location Notifications (HLN) category may be identified; this kind of services aims to provide road' users about hazardous situations in particular in terms of location, type, expected duration, etc. These services may be further classified in terms of Emergency electronic Brake Light (EBL) for warning drivers of hard braking by vehicles ahead; Emergency Vehicle Approaching (EVA) for providing an early warning of approaching emergency vehicles; Slow or Stationary Vehicle (SSV) for warning drivers about slow or stationary/broken down vehicles ahead; Traffic Jam ahead Warning (TJW) able to provide an alert to the driver that in traffic jam conditions reaches the end of the queue tail; Road Works Warning (RWW) aiming to inform drivers about works on the roads; Intersection movement assist (IMA) that warms drivers of vehicles approaching from a lateral position to the junction.

Further services within HLN category concern the collision risk minimisation (i.e. Cooperative Collision Risk Warning, CCRW) and the drivers of motorcycles warning (i.e. Motor Cycle Approaching indication, MCA).

Other kinds of applications refer to the vehicle to infrastructure communications and in particular to the signage; as the in-Vehicle SiGNage (VSGN) aiming at providing users' with road signs advanced information in the vehicle surroundings (this may facilitate drivers' gap at the signalised junctions), the in-Vehicle SPeeD limits (VSPD), aiming to provide users' with speed limits as well the ShockWave Damping (SWD) service able to recommend drivers about the optimal speed to be adopted by displaying the information in the vehicle. More in general there are the vulnerable road user (VRU) applications aiming at targeting crashes in case of vulnerable situations (for instance work areas, pedestrian detections, presence of emergency vehicles).

Other enhanced applications in case of urban contexts are Green Light Optimal Speed Advisory (GLOSA), Signal Violation/Intersection safety (SigV), Traffic Signal Priority etc.

Concerning the Green Light Optimal Speed Advisory (GLOSA) this is able to provide drivers with speed advice when they are approaching the traffic lights in order to uniformly mitigate the driving conditions by reducing the impact of acceleration/braking. With reference to the Signal Violation/Intersection safety (SigV) and the Traffic Signal Priority (TSP) these are respectively a safety-critical task focusing on the reduction of the number and severity of collisions at signalised intersections and a service able to guarantee the priority at signalised junctions of specific vehicles as emergency vehicles, public transport, etc.

Finally, other services are also referred to the in vehicle - infotainment applications that may be adopted in order to provide drivers with different kinds of information not only in terms of routes but also in terms of available services as parking or charging stations.

An overview of the main V2V and V2I applications is provided in the following table (see **Table 1**).

In general, it may be argued that in general vehicles are already connected devices; the development of an integrated framework combining the above

V2V	V2I
Emergency electronic Brake Light	in-Vehicle SiGNage
Emergency Vehicle Approaching	in-Vehicle SPeeD limits
Slow or Stationary Vehicle	ShockWave Damping
Traffic Jam ahead Warning	Vulnerable Road User applications
Road Works Warning	
Intersection movement assist	

Table 1.
V2V & V2I applications.

described services in which the vehicles will be able to interact each other and with the road infrastructures, is defined within the domain of Cooperative Intelligent Transport Systems (C-ITS). The C – ITS will be able to guarantee the road network management by synchronising all services and all shared information.

In conclusion in terms of driver guidance this research focuses on the implementation of GLOSA algorithm aiming to improve the traffic efficiency. The algorithm firstly calculates the distance and the travel time to the front traffic signal, then estimate the target speed constrained to the rules that were predefined considering different signal phases at the estimated arrival time.

2.2 V2I – intersection applications: the urban traffic control

The first criterion of classification of proposed methods in literature refers to the level of aggregation of input variables suitable for consideration in the optimisation procedure; in general two different kinds of variables may be adopted: the aggregate *flow* variables or the disaggregate *arrival* variables; therefore methods based on aggregate variables are also called flow based whereas methods based on disaggregate variables are also called arrival based methods.

Within *flow based methods* a further categorisation may be introduced in terms of junctions interaction; in particular the methods may be divided under single junction and networks depending on the degree of interaction between successive junctions that is isolated junction and interacting junctions [1]; then in case of interacting junctions the urban networks methods have to be applied whilst in case of isolated junction the single junction methods have to be considered. On the methodological point of view in case of interacting junctions the delay of the downstream approaches is influenced by delay of upstream furthermore the set of decision variables needed is also composed not only by stage durations and cycle time but also by an additional variable represented by the offset. Indeed, the offsets are introduced to describe the leg between the green stage at upstream and the green stage at downstream on the same flow direction. In this paper the interacting junctions' approaches are considered. It must be clarified that in case of urban traffic control the interaction between successive junctions may not be neglected therefore methods referring to the interacting junctions are needed in case of urban traffic control.

Alternatively, the *arrival based methods* may be considered in which starting from the number of arriving vehicles collected through loop detectors, the timing plans may be dynamically adapted to the traffic changes by allocating different green timings durations (extend/shorten) and by optimising the cycle length.

In terms of time dependency, it may be argued that flow based methods may be stationary or dynamic over time differently from arrival based methods that are intrinsically dynamic; this paper mainly focuses on dynamic approaches in order to provide a method suitable for on-line traffic management.

More in general two main traffic control paradigms may be related to the traffic flow input variables: the *centralised* and the *decentralised* approaches [2]. Indeed in case of centralised paradigms the traffic measurements are supposed to be received by a single central control agent which is responsible for deriving and implementing all control actions system considered consisting of three components respectively for regulating green splits, offsets, and cycle time; in case of decentralised paradigms the controller does not require information about global network inflow and the controller locally adjusts the traffic signal decision variables. In the last case depending on the adopted method variables adjustment may depend on both upstream and downstream local measurements (e.g. queue length) at each junction.

In summary two main dynamic approaches may be distinguished: the *planning-based traffic signal control* within centralised paradigms and the *actuated traffic signal control* within decentralised paradigms. In the first case, optimisation method starting from observed data and a traffic flow prediction model in forward time horizon, the actual input flows are estimated [3]. In general, the approach is oriented to decision variables design every control interval. Concerning the actuated traffic signal control, starting from the number of arriving vehicles collected through loop detectors, the timing plans may be dynamically adapted to the traffic changes by allocating different green timings durations (extend/shorten) and by the cycle length optimisation.

Alternatively, to these methods are approaches are also discussed in literature in particular:

- the control of some sensitive links, arterials [4–7],
- parts of the urban network through the implementation of gating control at the perimeter of the protected network (e.g. link metering or gating control; LM; see [8–11].

All these methods usually adopted in presence of unequipped vehicles must be extended to the case of connected vehicles. One of the most limiting points in case of centralised traffic control is the traffic flow prediction necessary to guarantee the consistency between traffic flow inputs and decision variables optimisation every control interval. The presence of connected vehicle may be useful in terms of estimation of location and speed of unequipped vehicles supporting the traffic flow prediction robustness. In terms of traffic flow modelling a microscopic approach has been considered.

In conclusion in terms of traffic control the paper aim is twofold:

- To apply a hybrid implementation of the centralised traffic control method and the link metering approach;
- To integrate these approaches with a procedure for traffic flow estimations.

In particular, one of the main problems in case of centralised control is the queue spillback and propagation in oversaturation conditions and queue may not be properly managed with respect to the longitudinal capacity.

To this aim a further refinement of the optimisation criteria is herein introduced: the queue equidistribution. A multi - objective optimisation procedure has been considered based on the combination of two criteria: the queue length optimisation and the queue equidistribution and a proper metaheuristics algorithm has been applied. Regarding the link metering control as further discussed in sub-Section 2, this is based on occupancy as a control variable.

3. Modelling framework and implementation settings

3.1 Overview of the proposed control framework

The proposed framework is composed by two sub-models: the first one aims at the traffic lights decision variables optimisation whilst the second one aims at the vehicle control trough speed optimisation.

Furthermore, in terms of traffic management an on-line procedure based on the combination of a centralised method and a link metering approach is adopted.

Regarding the driver guidance this paper focuses on the implementation of GLOSA algorithm aiming to improve the traffic efficiency. The algorithm firstly calculates the distance and travel time to the front traffic signal, then calculate the target speed constrained to the traffic signal decision variables and then to the estimated travel times.

Two sub-models operate simultaneously, and an overview of the framework is displayed in **Figure 1**.

In particular, the vehicle control is actuated depending on the vehicles distance from the infrastructure, whilst the traffic control procedure operates every control interval as it will be further discussed in the following Section 3 focusing on the implementation settings.

As already anticipated, the whole framework is composed by two sub- models:

- The first one aims at the traffic control decision variables design;

- The second one aims at the vehicle control decision variables optimisation.

In **Figure 2** a further overview of the whole framework including the vehicle control and in particular the traffic management, is shown then in the following a detailed description of each sub-model is provided.

Regarding the traffic control framework, this operates as a predictive control in which the network traffic control is the optimisation procedure, the proposed

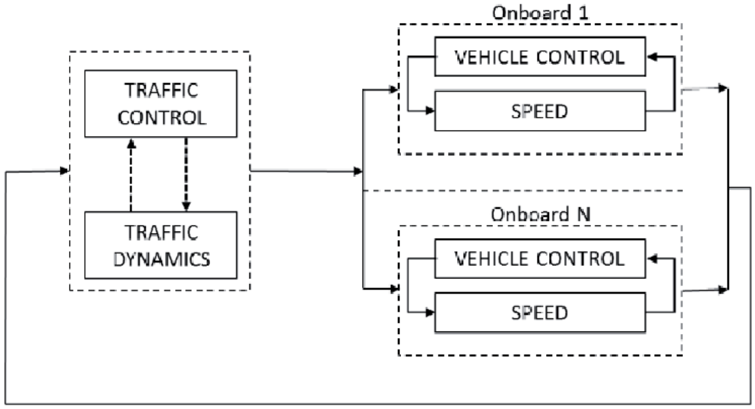


Figure 1.
Overview of the proposed control framework.

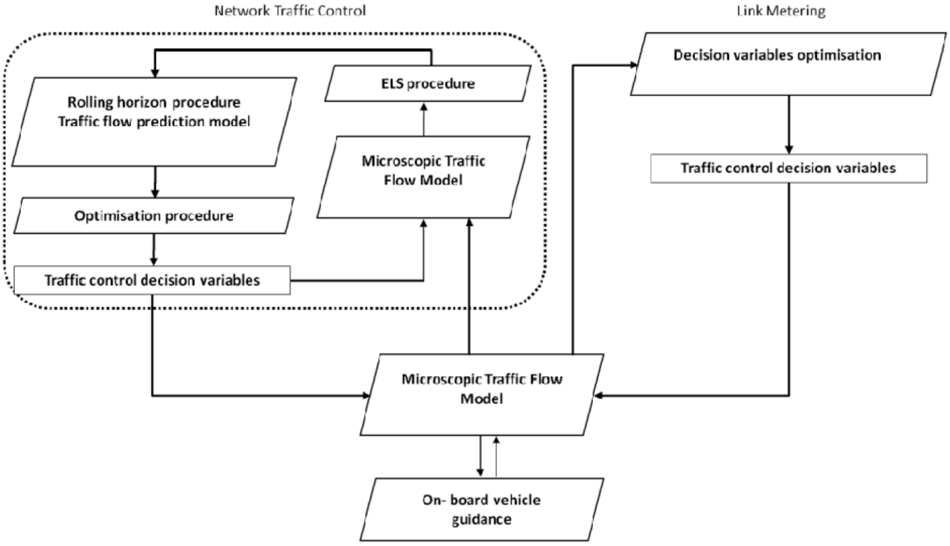


Figure 2.
Further description of the framework overview.

traffic flow models are the plant models, the Kalman Filter acts as prediction model and the ELS algorithm [12] for unequipped vehicles location and speed estimation.

Then this framework is composed by:

1. The microscopic traffic flow model;
2. The traffic flow prediction and estimation model providing input flows for the implementation of a traffic signal centralised approach;
    a. The rolling horizon approach;
    b. The KF;
    c. The unequipped vehicles status estimation;
3. The traffic control procedures.

As already anticipated, in order to guarantee the consistency between the traffic signals decision variables and the traffic flow two prediction terms are applied: the first one is related to the traffic flow model which is predicted with reference to the prediction horizon (e.g. fifteen min) the second one is the rolling horizon of the control. Concerning the rolling horizon, it must be clarified that the optimisation procedure works every control interval and the traffic information are updated every roll period (e.g. five minutes). Finally, the traffic information is collected in general every sub-interval (e.g. five seconds).

The second sub - model is represented by the on-board vehicle control procedure and operates depending on the vehicle distance as it will be discussed in more detail in Section 5 about the algorithm explication. As already anticipated in the introduction in this paper the S-GLOSA algorithm has been implemented. Therefore, the considered traffic control method is able to consider the interaction among junctions, whilst the vehicles control is applied only to the vehicles approaching each junction.

3.2 Centralised traffic control strategy

The Network Traffic Control (NTC) may be classified as a continuous linear optimisation problem and a multi - objective approach is pursued combining the total delay minimisation and the minimisation of the queue equidistribution criterion [13].

The parameters and constraints used in the model are listed below

- k approach
- j stage
- Δ approach-stage incidence matrix with entries $\ddot{a}_{kj} = 1$ if k receives green during j and 0 otherwise
- $C > 0$ the cycle length
- $s_j \in [0, C]$ the length of stage j as an optimisation variable; if no minimum length constraint is introduced
- $AR \in [0, C]$ the so-called all red period at the end of each stage
- $l_k \in [0, c]$ the lost time for approach k, assumed known
- g_k the effective green for approach k
- $g_{\min} \forall_k$ the minimum value of the effective green
- $q_k > 0$ the arrival flow for approach k, assumed known
- $sat_k > 0$ the saturation flow for approach k, assumed known
- $b \in [0,1]$ and eventually $t \in [1,,2,,3]$ the discrete variables for stage sequence definition, as decision variables
- i the generic links
- tr the turning rates
- $\beta_{l,r}$ the split ratio of the traffic demand in the l^{th} link and r^{th} movement
- $\xi_{l,r}$ the number of lanes assigned to the r movements in the l^{th} link
- m_i^{in} the total inflow
- $m_{(i,tr)}^{out}$ the discharging capacity expressed as vehicles/hour/lane
- $s_{(i,tr)}$ the sum of the signal phase ratios for the r^{th} movement in the l^{th} link

Finally, for each junction the node offset is needed; it represents the period between the start of a reference stage of junction i and the start of the reference stage of the first junction used as master for clock.

Regarding the queue equidistribution the following objective function (*of*, see Eq. 1) has been considered:

$$of\left(g_k,\varphi_i\right)=\sum_{i=1}^{n}\sum_{tr=1}^{3}\left(\max\left(\beta_{i,tr}m_i^{in}-\xi_{l,r}m_{i,tr}^{out}s_{i,tr},0\right)\right)^2 \quad (1)$$

The criterion aims to balance the rates of queue growth (or equalise them in an ideal case) in a network and then minimises the spill-over risk; it is based on traffic control decision variables design able to minimise the difference between the discharging capacity and the traffic demand at each link.

Summing up the on-line synchronisation [14] is obtained by combining together:

- the continuous variables needed to completely define the signal plan, that is are: (i) the stage lengths, constrained by the consistency among the stage lengths and the cycle length, (ii) the node offsets;
- the objective functions defined by the total delay and the queue equidistribution.

The procedure is able to simultaneously optimise the green timings and the offsets.

Regarding the solution algorithm in this paper the meta-heuristic Multi - objective Simulated Annealing [15, 16] has been applied. As a matter of fact, meta-heuristic algorithms can effectively address even optimisation problems with objective function not expressed in closed form, so that derivatives are not easily available, as it occurs for the scheduled synchronisation.

In particular the basic Simulated Annealing algorithm is a neighbourhood based meta-heuristic, which is inspired by the statistical mechanics to find solutions for both discrete and continuous optimisation problems.

Regarding the link metering (LM), is a feedback method implemented in accordance with the proportional integral type proposed by [6, 17–19] and it is based on occupancy as a control variable.

We list here the parameters used in the model

- k be the time step
- s be the section
- $\hat{o}$ be the desired downstream occupancy
- q_s be the gated flow
- o_s be the observed occupancy at downstream
- K_p be the proportional gain
- K_I be the integral gain

Regarding the control function, a proportional-integral-type (PI) feedback controller (2) aiming to maintain the observed occupancy around the desired value as in following displayed has been applied:

$$q_s(k)=q_s(k-1)-K_p\left[o_s(k)-o_s(k-1)\right]+K_I\left[\hat{o}-o_s(k)\right] \quad (2)$$

3.3 GLOSA (green light optimal speed advisory)

The Green Light Optimal Speed Advisory (GLOSA) is a Traffic LightS (TLS) time information system for advising drivers by means of using V2I communication. Messages are received by the vehicles on the times of the next TLS. Additionally, an on-board system calculates an ideal approach speed. The convenience on using this system relies on an increasing in safety, reducing the consumption and increasing efficiency of the unction.

In particular, the system adopted provides information about recommended speed level to the vehicle at 300 m from the TLS. If the current speed allows it to cross the intersection without stopping, the vehicle maintains the speed. Otherwise, the speed value allowing it is calculated. If the specific speed value detected is higher or lower than the maximum speed or less than the minimum speed allowed, no communication is provided to the driver and he will stop at the intersection. If the GLOSA is not active, the vehicle stops at signalised junctions; with the presence on-board of the GLOSA the vehicle travels through the same path and stops only 1 time.

It must be clarified that, the research does not focus on the type of communication between infrastructure and vehicle assuming that the messages are always delivered.

Furthermore, the algorithm has been designed aiming to guarantee that the vehicle will be able to cross the unction as soon as possible preferring, therefore, the travel time at fuel consumption or emission. The considered algorithm and the adopted variables are summarised in the following:

- D is the communication distance between the On-Board Units (OBU) and Road-Side-Units (RSU);
- T_{attr} is the crossing time for the vehicle;
- T_{switch} is the remaining time for green phase;
- $T_{next,phase}$ interval of the next green phase $[T_{initial}, T_{final}]$;
- V_{set} is the set of tested speed to cross the junction in the next phase. The possible speeds are defined with a 10 steps interval between the minimum and maximum speed;
- $T_{attr,Vset,i}$ is the crossing time for each speed of V_{set}.

The main successive steps of the algorithm are summarised in the following:

- Traffic signal control (TS) may be green or red;
    - ○ If red, then the optimisation of the next green intervals is activated;
    - ○ Otherwise, if the phase is green, it is verified if the vehicle speed is able to guarantee the vehicle crossing the section without stopping;
        a. If yes, the algorithm does nothing
        b. Otherwise, it is calculated a new green interval for the TS and a set of speeds is calculated (composed by teen values); it is calculated the crossing time and the consistency with the initial value of the green (within green interval) [the optimal value of the speed will guarantee the minimum value of the crossing time].

4. Application

4.1 Case study

The area identified for the case study is the city centre of Naples (regional capital of Campania, southern Italy). This area is characterised by a population of 978,399 and a population density of 8220 per km^2. Moreover, the metropolitan area has a number of inhabitants is around 3,118,000 and the population density is 2645 per km^2. Also, the total number of internal systematic yearly trips is around 685,000.

Two main roads are connected, Via Francesco Caracciolo and Via Riviera di Chiaia, from the West to the East side of the city. Current traffic rules and the connection between these two sides with two concurrent paths are implemented. Two paths are identified, path 1 goes through the Galleria Vittoria, and path 2 is composed of Via Chiatamone, Via Nazario Sauro and Via Acton. The sub-network layout is reported in **Figure 3**.

The network comprises four signalised junctions, among them traffic signals in Section 1, 2 and 5 are pedestrian traffic signals.

In terms of implementations remarks it must be highlighted that the whole traffic control procedure operates every control interval (every five minutes).

To optimise the traffic signal decision variables, the rolling horizon approach is adopted combined with a traffic flow prediction model. In particular, the rolling horizon itself is characterised by two terms the roll period (equal to five seconds) and the look ahead period (starting at the end of the roll period and ending at the upper bound of the prediction); in order to further guarantee the consistency with the traffic flow, traffic information are collected every roll period.

It must be distinguished that traffic signals in Section 1 and 2 are managed through LM whereas traffic signals in sections 4 and 5 are optimised through NTC. In general, the duration of the cycle length is 110 seconds and the stages 1 2 and 3 are respectively equal to 19 seconds, 65 seconds, and 26 seconds.

Regarding the origin - destination flows (and then the entry exit matrix definition) these have been obtained by combing the results of a macroscopic static traffic assignment procedure (PUMS - NAPOLI) [20] with a traffic counts survey done in 2017; in

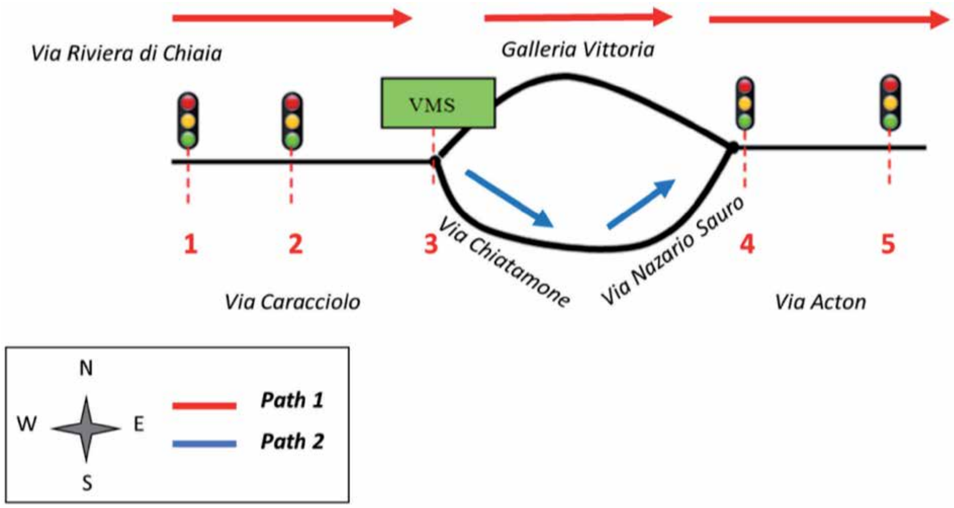


Figure 3.
Topology of the tested network.

particular traffic counts were collected at the beginning of 2017 in two peak hours of the day (mooring, from 7. 30 until 10.30 and afternoon, from 17.30 until 20.00).

The traffic flow was microscopically model through SUMO which is able to guarantee the on-line consistency of the procedures by adopting the TraCI interface, and the supporting code was developed in MATLAB (the R2018b was adopted). The input of the optimisation procedures are the travel times (TT) and the queue lengths (QL) that are collected through specific detectors located on the network. Due to the stochastic nature of the microsimulation approach, each simulation is run twenty successive times and the final values are provided by averaging the value of each simulation.

In terms of the goodness-of-fit function for model calibration, the Geoffrey E. Havers statistic [21] was adopted considering the observed and modelled data and the correspondence is less than 5 for 75% of the pairs (in accordance with the guidelines provided in [22].

4.2 Numerical results

In **Table 1**, the travel times obtained during peak hour simulation after model calibration are shown as well as the RSME. **Table 2** shows the queue lengths obtained during peak hour simulation after model calibration are shown. Let us observe that with reference to the results displayed in **Table 2** regarding the simulation of the current scenario, the main critical points are identified on Via Caracciolo (see **Figure 3**, junctions 1 and 2 are along Via Caracciolo) and Galleria Vittoria, direction W-E. For these two roads it was possible to reconstruct (in the condition of full network loading) the mean maximum queue lengths for intervals of 900 seconds or 15 minutes that fluctuate between 200 and 300 m on Via Caracciolo and between 300 and 400 m on Galleria Vittoria W-E. In the same table the root mean square (RSME) is summarised as a goodness of fit indicator of the calibration procedure (**Table 3**).

The considered path are two urban roads thus the speed limit is constrained to the 50 km/h; the difference in terms of length is around 650 m (indeed the alternative path is 1300 m); both paths diverge from the same node and merge to the same junction (Section 4).

Path 1 [min]	**Path 2 [min]**	**RSME**
44	35	3.14

Table 2.
Calibration results (TT) of the simulation model.

	Simulation intervals [s]				
	300–1200	**1200–2100**	**2100–3000**	**3000–3900**	**RSME**
Road	**Queue lengths [m]**				
Caracciolo	296.54	294.95	243.08	234.4	6.72
Galleria Vitt. $^{\text{E–W}}$ (1)	81.93	83.7	95.76	98.53	2.76
Galleria Vitt. $^{\text{W–E}}$ (2)	277.41	339.13	390.39	384.56	6.18

(1) *From Via Acton to Piazza Vittoria (junction 3 in* ***Figure 3****).*
(2) *From Piazza Vittoria (junction 3 in* ***Figure 3****) to Via Acton.*

Table 3.
Calibration results (QL) of the simulation model.

The reader may refer to **Figure 3**, showing the traffic signal controllers (sections 4 and 5) and the link controllers (sections 1 and 2) are represented.

Concerning the simulation results, it must be specified that all numerical applications were run on a server machine Intel(R) Xeon(R) CPU E5–1620 v3, clocked at 3.50GHz and with 8GB of RAM. As already anticipated in the introduction three scenarios were considered in all:

- the first was only based on a traffic control procedure;
- the second on speed guidance optimisation;
- the third was based on the combination of both sub-models.

The simulation interval considered for each scenario is equal to one hour and for each simulation a warm-up period taking teen minutes is considered.

The tested scenarios are listed below:

1. the traffic control scenario [TC], the hybrid traffic control strategy combining the centralised traffic control and the link metering are applied; this strategy has been tested considered a successive bi-level mono-criterion optimisation [TCMONO] and a simultaneously multi-criteria optimisation [TCMULTI];
2. the speed guidance scenario [S-GLOSA], providing a warning to the driver concerning the optimal speed to be maintained while approaching the junction;
3. the mixed scenario [TC& S-GLOSA], both the TC strategy and the MS-GLOSA are implemented.

To preliminarily compare the achieved results a further scenario has introduced as baseline. In particular, in the baseline scenario an Adaptive Signal Control [A - SC] strategy has been considered [23]. Furthermore, it must be also clarified that in all scenarios, the considered penetration rate of CAV (Connected Autonomous Vehicles) equals 50%, and the impact of the penetration rate of connected and autonomous vehicles has not yet been tested. Indeed, it will be remarked as future perspective then in terms of further issues to be investigated. In the following the results of each scenario are displayed.

In particular, in order to evaluate the effectiveness of the proposed strategy, the [A - SC] scenario was compared with each one of three scenarios [YY], and the relative difference (i.e. A - SC vs. YY) between the mean value of actual travel times of two alternative paths is then performed [TT_{pathx}] (see **Table 4**) as well as the mean value of the relative difference of the queue lengths [QL] at significant sections (see **Table 5**; sections are identified in accordance with **Figure 3**). The results

YY - SCENARIO	Path 1	Path 2
TC_{MONO}	−18.75	−26.70
TC_{MULTI}	−25.31	−33.28
S-GLOSA	−15.12	−21.27
TC_{MULTI}&S-GLOSA	−29.11	−37.22

Table 4.
Mean TTS rel. Diff. [%] of [a - SC] scenario w.r.t [YY] scenario.

YY - SCENARIO	Section 1	Section 2	Section 3	Section 4	Section 5
TC_{MONO}	−88.70	−75.28	−84.46	−68.13	−32.18
TC_{MULTI}	−92.22	−81.31	−87.08	−72.15	−35.27
S-GLOSA	−83.18	−69.07	−80.22	−65.21	−27.06
TC_{MULTI}&S-GLOSA	−97.04	−84.43	−91.18	−75.13	−36.22

Table 5.
Mean QLS rel. Diff. Of [a - SC] scenario wrt [YY] scenario.

highlight three main considerations: the first one is about the TC and in particular it is confirmed that TC based on multi-criteria optimisation outperforms that TC based on mono-criterion optimisation and the result was not intuitively expected due to the further constrain that is introduced in case of multi-criteria optimisation. Secondly it must be observed that TC mono-criterion and S-GLOSA provide

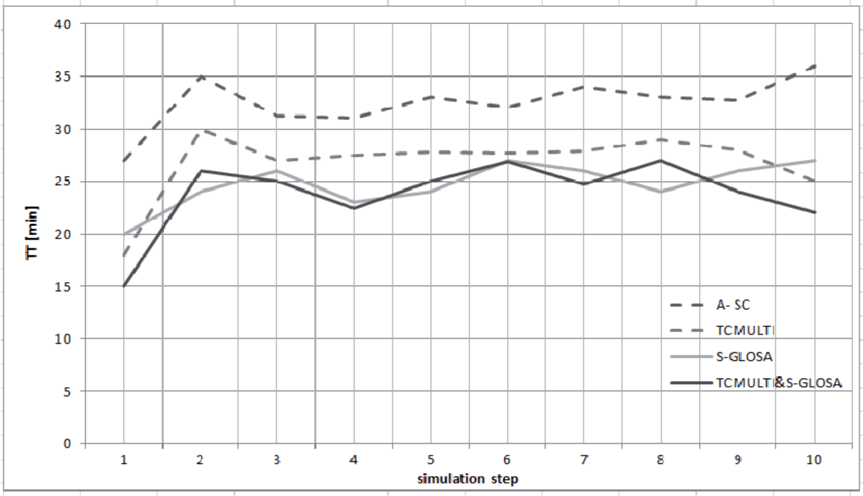


Figure 4.
Results for each scenario: Mean TTs [min] against simulation step for each scenario.

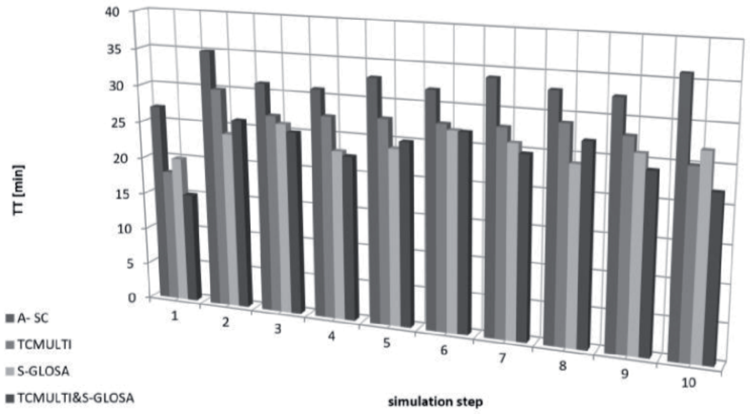


Figure 5.
Results for each scenario: Rel diff [%] of the mean TT wrt the A-SC scenario, against simulation step for each scenario.

similar results therefore better network performances may be achieved through the implementation of TC based on multi-criteria optimisation; finally, as expected, the combination of TC based on multi-criteria and S-GLOSA provides final best performances.

In order to provide a further comparison among all scenarios the results in terms of mean TT of the alternative route 2 are also displayed against simulation step for each scenario (see **Figure 4**) as well as the relative difference of the mean TT of each scenario with respect to the baseline scenario, that the Adaptive Signal Control scenario (see **Figure 5**).

5. Conclusions and future perspectives

The paper illustrates a unified framework which embeds a simultaneous traffic control strategy and the automated vehicle control. In particular the traffic control strategy is composed by two sub models: one is referred to the centralised traffic management the other one is characterised by the link metering strategy; regarding the vehicle control, the speed optimisation procedure based on Green Light Optimal Speed Advisory (GLOSA) has been applied in particular with reference to the next single junction approached by the vehicles (S-GLOSA). A microscopic traffic flow modelling has been adopted and all models were run in a SUMO simulation environment.

The integrated framework was then tested on a real case study consisting of a highly congested sub-network in the city centre of Naples (Italy). The network layout is represented by one diversion node and two alternative paths connecting the same origin - destination pair.

In order to evaluate the effectiveness of the proposed framework, three scenarios were tested: the first was only based on a centralised traffic control procedure [TC] that was further analysed considering the bi-level mono-criterion implementation and the multi-criteria approach; the second one was based on speed guidance optimisation [S-GLOSA] and the third was based on the combination of both sub-models the multi-criteria traffic control and the speed optimisation [TCMULTI & S-GLOSA]. Finally, the framework effectiveness was evaluated in terms of within-day dynamics with respect to the travel times and queue length performance indices.

Three main considerations have arisen: the first one is about the TC strategy and in particular it was tested that multi-criteria optimisation outperforms the mono-criterion approach; the second one refers to the comparison between TC_{MULTI} and S-GLOSA therefore it is verified that S-GLOSA provides worse performances than the TC_{MULTI} method; finally the combination between TC_{MULTI} and S-GLOSA provide as expected best results.

Regarding future research perspectives, some preliminary modelling considerations may be summarised. First of all, the authors would like to test the proposed framework on different networks characterised by more complex topologies. Secondly, the sensitivity at different penetration rates of CAV must be analysed. Thirdly, further refinements are needed for the implementation of the S-GLOSA strategy and, for completeness, in future researches the environmental impact will be also analysed.

Finally some further technological and operational perspectives may be discussed. The situation described and analysed in the chapter has shown the benefit of the cooperation among infrastructures and vehicles control. It is worth noting that this situation is one of the possible results that technological development on one side, and normative evolution on the other, will enable in next years.

For example, the implementation of S-GLOSA (or even MS-GLOSA) in urban environments will be strongly affected by the communication technologies used, with an evident advantage for this use case of the Automotive LTE with respect of Dedicated Short Range Communication (DSRC). In summary, for this and many other reasons, the concrete future implementation of cooperative scenarios has some kinds of uncertainty, but this last observation make even probably more meaningful the kind of experiments discussed here.

Acknowledgements

We would like to thank the Municipality of Naples for the support provided.

Conflict of interest

No potential conflict of interest was reported by the authors.

Funding

This research has been partially supported by the University of Salerno, under local grants n. ORSA165221–2016 and n. ORSA180377–2018; and by the Italian Program PON AIM - Attraction and International Mobility, Linea 1 under Grant AIM1877579–3-CUP- D44I18000220006.

Author details

Roberta Di Pace[1]*, Chiara Fiori[1], Luigi Pariota[2] and Facundo Storani[1]

1 University of Salerno, Fisciano (SA), Italy

2 University of Napoli Federico II (NA), Italy

*Address all correspondence to: rdipace@unisa.it

References

[1] G. E. Cantarella, S. de Luca, R. Di Pace, and S. Memoli, “Network Signal Setting Design: meta-heuristic optimisation methods,” *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 24-45, 2015.

[2] R. Sha and A. H. Chow, “A comparative study of centralised and decentralised architectures for network traffic control,” *Transportation Planning and Technology*, vol. 42, no. 5, pp. 459-469, 2019.

[3] M. Di Gangi, G. E. Cantarella, R. Di Pace, and S. Memoli, “Network traffic control based on a mesoscopic dynamic flow model,” *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 3-26, 2016.

[4] D. Bretherton, G. Bowen, and K. Wood, “Effective urban traffic management and control: recent developments in SCOOT,” 2003.

[5] J. Luk and D. Green, “Balancing traffic density in a signalised network,” 2010.

[6] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville, “ALINEA: A local feedback control law for on-ramp metering,” *Transportation Research Record*, vol. 1320, no. 1, pp. 58-67, 1991.

[7] K. Wood, D. Bretherton, A. Maxwell, K. Smith, and G. Bowen, “Improved Traffic Management and Bus Priority with SCOOT,” *TRL STAFF PAPER PA 3860/02*, 2002.

[8] K. Aboudolas and N. Geroliminis, “Feedback perimeter control for multi-region and heterogeneous congested cities,” 2013.

[9] N. Geroliminis and C. F. Daganzo, “Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings,” *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759-770, 2008.

[10] N. Geroliminis and J. Sun, “Properties of a well-defined macroscopic fundamental diagram for urban traffic,” *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 605-617, 2011.

[11] M. Ramezani, J. Haddad, and N. Geroliminis, “Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control,” *Transportation Research Part B: Methodological*, vol. 74, pp. 1-19, 2015.

[12] Y. Feng, K. L. Head, S. Khoshmagham, and M. Zamanipour, “A real-time adaptive signal control in a connected vehicle environment,” *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 460-473, 2015.

[13] K. Jang, H. Kim, and I. G. Jang, “Traffic signal optimization for oversaturated urban networks: Queue growth equalization,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2121-2128, 2015.

[14] S. Memoli, G. E. Cantarella, S. de Luca, and R. Di Pace, “Network signal setting design with stage sequence optimisation,” *Transportation Research Part B: Methodological*, vol. 100, pp. 20-42, 2017.

[15] S. Bandyopadhyay, S. Saha, U. Maulik, and K. Deb, “A simulated annealing-based multiobjective optimization algorithm: AMOSA,” *IEEE transactions on evolutionary computation*, vol. 12, no. 3, pp. 269-283, 2008.

[16] R. Di Pace, “A traffic control framework for urban networks based on within-day dynamic traffic flow models,” *Transportmetrica A: Transport Science*, vol. 16, no. 2, pp. 234-269, 2020.

[17] M. Keyvan-Ekbatani, X. Gao, V. V. Gayah, and V. L. Knoop, "Traffic-responsive signals combined with perimeter control: investigating the benefits," *Transportmetrica B: Transport Dynamics*, vol. 7, no. 1, pp. 1402-1425, 2019.

[18] A. Kotsialos, M. Papageorgiou, and F. Middelham, "Local and optimal coordinated ramp metering for freeway networks," *Journal of Intelligent Transportation Systems*, vol. 9, no. 4, pp. 187-203, 2005.

[19] I. Papamichail and M. Papageorgiou, "Traffic-responsive linked ramp-metering control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 111-121, 2008.

[20] "Piano Urbano della Mobilità Sostenibile (PUMS) - NAPOLI." [Online]. Available: https://www.comune.napoli.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/28525.

[21] J. Ma, H. Dong, and H. M. Zhang, "Calibration of microsimulation with heuristic optimization methods," *Transportation Research Record*, 2007, doi: 10.3141/1999-22.

[22] Highway Agency, *Design Manual for Roads and Bridges, Volume 12: Traffic Appraisal of Road Schemes Section 2: Traffic Appraisal Advice. Part I: Traffic Appraisal in Urban Areas*. The Stationery Office, London, 1996.

[23] R. P. Roess, E. S. Prassas, and W. R. McShane, *Traffic Engineering*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2004.

Chapter 8

Towards Shared Mobility Services in Ring Shape

Fabien Leurent

Abstract

A shared mobility service (SMS) under ring shape would combine the principle of service cycle along a fixed route (as in a transit line) and a fairly important territorial coverage, assuming that every user would accept to walk on some length to and from the service. Thus, service availability can be optimised, detours are avoided, vehicles achieve higher productivity. The synergy between the ring-shaped infrastructure and the vehicle fleet enables to optimise the quality of service in terms of access time and ride time, and also to reduce production costs - and therefore the tariff fares, under suitable regulation. The chapter aims to reveal these 'systemic qualities' of ring-shaped SMSs by providing a mathematical model called 'Orbicity'. It has a four-fold architecture: (i) traffic operations, (ii) supply-demand equilibrium under elastic demand, (iii) service management with endogenous fleet size and fare rate, (iv) service policy in terms of technology (vehicle type, number of places, energy vector, driving technology) and also the regulation regime. After outlining the model for ring-shaped shuttle services, we explore a set of scenarios along two axes of technological generation and regulation regime. It appears that ring-shaped shuttle services could be supplied at very affordable prices, while achieving profitability and requiring no public subsidies.

Keywords: ring shape, traffic model, fleet sizing, pricing, service regulation

1. Introduction

Conducting research for the Institute for Sustainable Mobility (a partnership between the Renault Group and ParisTech), we have modelled an array of shared mobility services using different forms of transportation (taxis, shuttles, vehicle hire schemes) arranged in a ring-shaped system [1–3].

For passenger mobility services in urban areas, the ring-shape principle is aimed to combine the axiom of service cycles (like a public transport line) with a broad geographical coverage, assuming that users are willing to walk a certain distance before and after each journey they make on the ring. By keeping the vehicles in service to run along the ring, availability can be optimised and detours avoided, ensuring that every vehicle is genuinely productive. Achieving greater synergy between the ring infrastructure and the vehicle fleet makes it possible to optimise the quality of service in terms of access time and driving time, while also pushing down production costs and thus enabling for affordable fares.

This chapter is aimed to highlight the 'systemic qualities' of shared mobility services adopting a ring format, as well as exploring the conditions required to establish a ring system in urban settlements.

We will first examine recent technological advances in mobility services and mobility-adjacent services, with reference to the fundamental spatial components of transportation: vehicles, stations, lines and networks. We will then demonstrate how a ring system makes it possible to cover a relatively large geographical area while also establishing service cycles for shared vehicles. A simple geographical model will be provided to quantify the geographical potential of demand, with reference to a few examples from France.

Of course, such services still need to be attractive, offering decent quality of service at an affordable price. These factors have been represented in a specific technical and economic model [1–4]. The 'Orbicity' generic model can be tailored to different types of service. The modal models share a four-tier architecture which involves, from bottom up, (i) the physical operations of the service and the laws governing its vehicle flow, (ii) the balance between journey supply and demand, (iii) optimised service management in terms of fleet size and fare price, (iv) the strategic positioning of the service in terms of technologies, conditionally to the applicable regulation regime.

Using this model, we will examine a number of scenarios which incorporate two key analytical dimensions: the generation of technology used and the applicable regulatory framework. We will demonstrate that not only does technological progress considerably expand the scope of possibilities, but also that regulation plays a vital role. It is entirely possible to imagine a shuttle service offering very reasonable fares, and possibly even without public subsidies.

2. Technical and spatial forms of urban mobility

2.1 Sweeping technological change

The 2010s were a decade defined by the confluence of multiple technological advances, with a strong focus on mobility [5]. Various technologies developed over the preceding years were combined with a new sense of synergy: GIS, GPS and smartphones. Geographical Information Systems (GIS) allowing for the processing, mapping and administration of geographical databases, are utilised by Google (Maps, Earth, StreetView) and others. GPS tracking, providing precise geographical location data in real time, became available for each individual mobile entity (individuals as well as vehicles), with relatively inexpensive portable devices. Mobile telephone services mean that individuals are always connected, anywhere and anytime (network coverage). Touch-screen smartphones have become the ideal tool for user interaction with any service. Individual users have the power to organise their travel plans, see their position on a dynamic map, enrich that map with information of interest to them (addresses, traffic conditions, public transport routes and stations), get recommendations for accessing transport and planning routes, and even receive directions in writing or in audio form in the language of their choice. Moreover, these services can be combined with the vast array of multimedia functions offered by modern smartphones.

The rise of individual mobility management has been remarkable. To get an idea of this, we need only consider the task of planning a complex itinerary on a metropolitan public transport network, before and after the advent of online route search services.

Individual users are now masters of their own 'customer experience', designers of their own transport services [6]. Consider the familiar Plan-Book-Ticket steps of the 'customer experience' from a marketing theory perspective:

- Planning to purchase a product: in this case, a travel itinerary;

- Booking: reservations for public transport, where necessary;

- Ticketing: both to provide easy information on the commercial conditions including fares and for payment and invoicing.

These are mobility-adjacent services which add much value to the travel experience as a whole, a value felt more keenly for public transport trips than car trips. The provision of information, the capacity to search massive databases and the customisation features 'make up for' the dissociation between vehicle and user which is an inherent feature of public transport (a dissociation which is necessary at this higher level of organisation, but which represents a fundamental handicap for collective transport solutions in comparison with private vehicles).

The benefits on the demand side are not limited to these mobility-adjacent services. Operators and innovators have seized upon the opportunities offered by advanced technologies to invent (or reinvent) new mobility services and new vehicles:

- Reinventing the bicycle, with more electric options and 'shared' vehicles in the form of short-term cycle-hire services with designated stations (cf. 'Boris bikes' in London or Velib systems in Paris and elsewhere) or without (i.e. free-floating services such as Jump etc). In Paris, the summer of 2018 saw a rapid proliferation of these free-floating cycle-share services, with each new player adopting a different colour for its fleet of bicycles: almost all of them had disappeared by the end of that autumn!

- Scooters have experienced a similar overhaul: modernised, reinforced and equipped with electric motors and batteries, they are also being offered by a variety of free-floating services in big cities. Parisians witnessed a sudden influx of new scooter services in spring 2019, followed by a period of consolidation which left only 3 or 4 companies standing by that autumn.

- Something similar has happened with cars: the renaissance of the electric car in the 2000s was followed in the 2010s by their deployment in urban car-sharing services (short-term rentals where the vehicle must be returned to its point of origin) and free-floating car-sharing services, with stations (Autolib) or without (Car2Go, now ShareNow).

- A similar vehicle sharing system has been developed for electric motorcycles: for example, CityScoot had over 6000 mopeds in circulation in the Greater Paris region at the end of 2019.

Each of these sharing services depends upon a two-sided ***digital platform***: a customer interface which handles the commercial operations, while the production side centralises the management and optimisation of resources.

If operators are capable of mobilising a fleet of vehicles and a team of service and maintenance personnel, they may also offer door-to-door services not dissimilar to a classic taxi service: Uber has emerged as the champion of so-called 'ride hailing' services accessed via mobile phone, combining the Booking, Planning and Ticketing functions into an extremely fluid user experience enshrined in a mobile app.

Other platforms offering car-sharing services (e.g. Drivy) or car-pooling, which are thriving for inter-urban travel (Blablacar), are yet to hit upon the magic formula for urban users.

All of these changes add up to form a new ecosystem, whose components mutually reinforce one another. While in transit, vehicles record information on the urban traffic conditions in real time, and this information is used to adjust the services on offer. The remarkable development of Uber has revolutionised taxi services in major metropolises and beyond (even as far as the distant suburbs of North America, cf. [7]).

2.2 Questions of form

The typical form of the ***vehicle*** as a mechanised, often motorised, form of transportation has been reinforced, with a diversification of models allowing for adaptations to local conditions: small electric vehicles in very dense urban areas where space is at a premium and pollutant emissions need to be kept to a minimum.

The typical form of the ***road network*** as a medium for multiple uses, a circulation infrastructure connecting different places, has also been confirmed. Shared services make use of the road network as their means of circulation and parking, and also recharging for electric vehicles.

Our notion of what constitutes a ***station*** has been diversified. Stations such as railway stations are fixed hubs with large numbers of users which serve as landmarks in the urban landscape. The positioning of bus and coach stops across the road network has become more visible thanks to mobile applications highlighting their location. Cycle hire stations are now being adapted to serve electric mobility options: the concentration achieved by grouping together the available spaces increases the probability of finding a free vehicle when it is needed (yet it puts a parking constraint at the trip destination). Meanwhile, free-floating services are challenging the very notion of stations: in fact, they are making every available parking space on the road network a potential station.

Ultimately, it is the form of the transport ***line*** which has been most affected by these changes. On the one hand, designated itineraries on the public road network are being challenged by automatic route-planning search engines which throw up any number of unlikely variants (such as cutting through a hospital courtyard, or taking a side street between two sections of much busier roads). Put simply, customised itineraries are taking over from established routes. On the other hand, public transport lines such as bus routes are now in competition with shared services, especially those routes where passenger numbers are not sufficient to justify a high-frequency service. The competition is greater still in less dense zones where collective transport services form only a loose network with large blind spots, too much distance between stops and a lack of effective connections to other segments of the network.

Some authors have even questioned the pertinence of rail transportation for long-distance travel from one suburb to another, floating the hypothesis that self-driving taxis could be used to collect passengers travelling from similar starting points to similar destinations, dropping them off at their respective destinations [8, 9].

2.3 Origins of the ring form

Certain public transport lines take the form of a ring: Line 6 of the Madrid metro, the Circle line on the London Underground, the Singapore metro, the systems in place in various Chinese cities, and even the ‘Circulator’ bus route in Washington. The main function assigned to these lines is to fulfil both the transmission and distribution of passenger flows [10].

The principle is similar to that of urban ring roads such as the Paris peripheral boulevard, which are more about distributing traffic flows than circumventing the city altogether.

The key to using a loop transit system to transport products (e.g. water in buckets, to put out a fire) or people (e.g. a cable car) is to effectively combine the form of the circuit with the service cycle of the containers. The result should be a homogenous spread of load between the containers, with regularisation and intensification of the overall rhythm.

Transport lines which are relatively linear in shape are typically used for radial connections, linking dense zones (urban centres) to less densely-packed areas (the suburbs). The flow pressure is high on the dense side, and low on the less dense side. The ring form is more specifically suited to a relatively homogenous spatial milieu: the homogeneity of the milieu contributes to the effectiveness of the ring-shaped service; reciprocally, the effectiveness of the service contributes to the homogenisation of the milieu.

3. Geographical potential and technical advantages of a ring-shaped service

3.1 Geographical principles: potential locations

What is the potential user base for a ring-shaped service operating in a relatively homogenous territory?

To answer, let us build up a simple model. Let A represent the surface area of the territory in question, and P its population. The mean density is P/A people per unit of surface area. Let μ represent the mobility rate per individual and per day, typically somewhere between 3 and 4 trips. In spatial terms, the hypothetical 'outbound density' is $\mu P/A$ trips per unit of surface area.

Assume further, on a provisional basis, that the ring is a circle of radius R, and that along its whole circumference it attracts passengers from within a band 2ℓ wide – i.e. from a distance of at most ℓ from the pathway of the circular transport service. The total drainage basin for this infrastructure is thus equivalent to a surface area of $4\pi\ell R$ (**Figure 1**).

This drainage basin can be represented as a proportion of the total surface area (A) of the territory:

$$\theta = \frac{4\pi\ell R}{A} \ (\textit{limited to } 1) \tag{1}$$

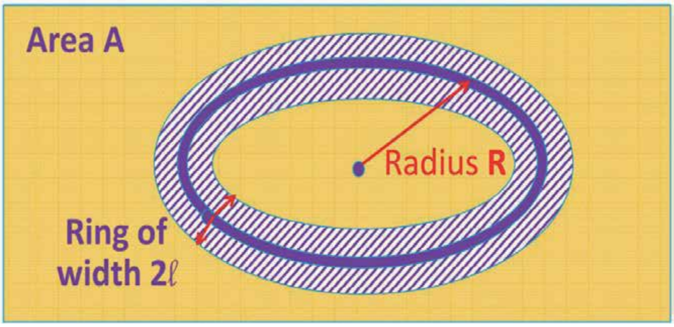


Figure 1.
Area served by a ring-shaped line.

Based on the hypothesis of a homogenous spatial distribution of activities throughout this territory, the ratio θapplies to both the point of origin of trips and their destination. The market potential of the ring service is thus θ^2 multiplied by the city's total mobility, i.e.

$$\begin{aligned}\tilde{Q} &= \mathrm{P}.\mu.\left(\frac{4\pi\ell\mathrm{R}}{\mathrm{A}}\right)^2 \\ &= \frac{\mathrm{P}.\mu}{\mathrm{A}}.\frac{(4\pi\ell\mathrm{R})^2}{\mathrm{A}} \\ &= \frac{\mathrm{P}.\mu}{\mathrm{A}}.4\pi.(2\ell)^2\left(\frac{\mathrm{R}}{\mathrm{R_A}}\right)^2\end{aligned} \tag{2}$$

The terms of the formula can thus be rearranged to show:

- Mobility generation density $\mu\mathrm{P}/\mathrm{A}$,
- An equivalent radius for the urban settlement $\mathrm{R_A}$, such as $\pi.\mathrm{R_A^2} = \mathrm{A}$.
- The width of the band 2ℓ, with exponent 2,
- The role of the radii: the ratio $\mathrm{R}/\mathrm{R_A}$, squared, determines the potential demand.

For a given territory, the goal is to find the 'natural' proportion between R and $\mathrm{R_A}$. **Figure 2a,b** shows two examples from the cities of Rennes and Saint-Malo, in the Brittany region of France. **Figure 2c,d** shows two examples in the Ile-de-France region: one for the Paris-Saclay area, and another for the Greater Paris conurbation, here labelled Grand Paris.

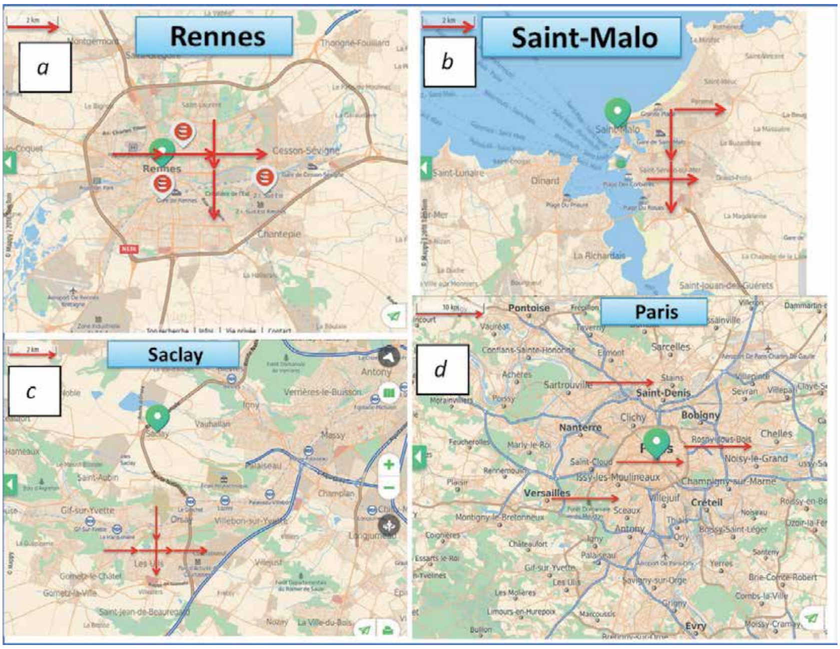


Figure 2.
Examples for (a) Rennes, (b) St Malo, (c) Paris-Saclay, (d) Grand Paris (source: Mappy, modified by the author).

City	St Malo	Rennes	Grand Paris	Ville de Paris
Surface A (km^2)	80	200	2000	100
Eqt radius R_A (km)	5	8	25	6
Population P (thousands)	50	300	11,000	2000
P/A (k people/km^2)	0.63	1.50	5.50	20,0
Ratio $3/R_A$	59%	38%	12%	53%
Ratio θ^2 if R = 3 km	74%	30%	3%	59%
$\tilde{Q}$ (thousands trips/day)	38.9	37.3	13.7	995
Variant if R = 5 km	107.9	103.6	38.0	2763

Table 1.
Quantification of geographical potential.

Table 1 shows a high degree of similarity between the results (Q-tilde) for the two radius values envisaged: between 10,000 and 12,000 journeys per day for a radius of 3 km, or 29,000 to 34,000 journeys per day for a radius of 5 km. To put it another way, the urban conditions found in a variety of French cities all represent interesting levels of potential demand. Nevertheless, the quality of service would need to be attractive to users, in conjunction with attractive prices.

3.2 Quality of service: four components

Whatever the means of transportation - shuttle, taxi, car share, moped, bicycle, scooter etc. – users expect a satisfactory quality of service. While the quality of service has been adequately defined and described for collective transit systems in the Transit Capacity and Quality of Service Manual [11], for shared mobility services that are more diverse a more generic definition is required. To that end, an analysis framework comprising four components was put forward as follows [4]:

1. ***Maintenance and manners***: the vehicle needs to be appropriate for this service, in terms of its mechanical condition and energy supply; it must be clean; if it is to be used collectively (shuttle), each user has a right to expect a minimum level of courtesy from other passengers on board.

2. ***'Pleasure', including Protection and comfort***: four-wheeled vehicles offer shelter from the elements (weather hazards) and potential impacts (shocks), along with more comfortable seating arrangements than mopeds or bicycles. At the other end of the spectrum, scooters require passengers to be standing up (although this does offer a certain degree of excitement).

3. ***'Conductance', including Efficiency, mobility and speed***: 'conductance' is the term we use to describe the aptitude of a vehicle to fulfil its transportation role, from proximity of access to arrival at final destination. A taxi offers maximum conductance. The mobility of a vehicle depends upon its mechanical capacities, and particularly its power source. Speed depends upon the mobility and agility of the vehicle (the latter being greater for two-wheeled vehicles) as well as the fluidity of traffic conditions.

4. ***Ease of use, including Availability***. The less time required to access a service, the more available it is. Shuttles and taxis pick up passengers directly, they

come to them, so the access time depends on the operating speed. For self-service solutions, users must arrive at the vehicle, which implies walking pace. In both cases, the initial distance between the user and the service plays an important role: it will depend on the number of vehicles in service, their level of occupancy and the size of the ring network.

This Availability category might also include the transaction operations between users and the service: information, pricing, payment and invoicing, reserving a vehicle or selecting a destination for taxi and shuttle services. It falls to the operator to simplify the corresponding tasks required of users as far as possible.

3.3 Ring-shaped services and quality of service

Conductance. Running vehicles on a ring-shaped line avoids wasting time on detours, pick-ups and set-downs outside the designated circuit. This allows us to optimise the ***availability*** of each vehicle in service, and thus maximise the availability of the service for potential users.

Avoiding detours helps to increase the ***speed*** of journeys made on the ring. Speed also depends upon infrastructure design on the ground, with certain measures which may be taken to improve the fluidity of travel. Considering the best interests of the circuit as a whole, it is advisable to decide upon a suitable speed of travel, determined with reference to the urban environment and in order to satisfy all users and occupants of the space.

Pleasure. ***Protection and comfort*** depend first and foremost on the vehicles used: when determining which vehicles to acquire for the service, a number of specifications should be outlined in this respect. The design of the infrastructure also plays a role in ensuring the safety of individuals, ensuring that there is sufficient space for vehicles to move, guaranteeing sufficient visibility, smoothing out potential problem points and using appropriate signage to draw attention to them.

Maintenance. This depends on the preventive and curative measures taken by service operators, as well as the behaviour of passengers, other road users and residents. The ring format facilitates the logistics of curative interventions. It also lends itself to CCTV surveillance and on-site surveillance: these functions could well be integrated into the broader system of surveillance for the city's traffic and parking.

Ease of use. It is worth focusing in particular on the ***availability*** of the service: it will depend upon the ratio between the level of demand and the size of the fleet, among other factors including the circumference of the ring, the average distance travelled by passengers and the operating speed. The complex interplay of these factors is rendered more complex still by the fact that quality of service will have an influence on the level of demand. This is why we have developed specific technical and economic models for mobility services in ring form.

3.4 Technical principles

Combining a ring-shaped format with a fleet of vehicles provides significant opportunities for synergy. The ring circuit connects a number of points distributed along its circumference (C), attracting users from a band which is 2ℓ wide, with ℓ the range of attraction either side of the circuit. This ring has the potential to attract a certain amount of demand: let us call it Q number of journeys for each day the service is operational. The ring connects locations on a point to point basis: this function is clear when shown on a map, and must be obvious (legible) on the ground. Along the route, we can distribute logistical functions such as parking bays, recharging stations and cleaning stations, located together or separately as required.

As for the fleet of vehicles, let us assume that it represents modal homogeneity. Each vehicle is capable of travelling at an average speed of v_0. The size of the vehicle will depend on the mode of transportation: let K represent the number of places on board. For two-wheeled vehicles $K = 1$, whereas $K = 4$ or thereabouts for cars and $K = 12$ for shuttle buses. Each vehicle is shared, in that it constitutes a component of a fleet whose total size is N.

Keeping the fleet of vehicles attached to the ring circuit ensures that:

- Each vehicle runs productively, in cyclical fashion, with an average workload of Q/N trips per day.
- Each point receives a guaranteed frequency of service (for shuttles), or else the availability across the whole zone is homogenous (for shared vehicles). This means that any user can access the service at any point.
- Multi-passenger vehicles can be pooled without the need for a detour: the only condition is that the vehicle must stop to let passengers come on and off. These elementary logistical tasks are distributed between the vehicles in circulation.

The average distance of the rides made by passengers, which we can denote as L_R for ride length, is a factor in the rate of occupancy of vehicles, and also of the exposure of passengers to the delays required for other passengers to board and alight the service.

Average operating speed v depends on all of the factors mentioned above:

- In terms of infrastructure, the circumference C and the speed of travel v_0,
- In terms of demand, the volume Q and the average ride length L_R,
- In terms of the vehicles themselves, the passenger capacity K and the time taken for passengers to alight/board the service t_S,
- In terms of the overall service, the size of the fleet N and the number of hours H for which the service runs during the day.

4. Technical and economic modelling

There exists a substantial body of knowledge, theoretical as well as methodological, to design transportation networks and services. The classical textbook [12] provides travel demand models for traffic simulation, therefore enabling for traffic and revenue forecasting. Its domain of application encompasses roadway networks and public transport networks involving lines. The more advanced book [13] also considers more diverse forms of public transport, including on-demand services (chapter 8): on representing service operations, the production costs can be modelled, too, and compared to service revenues in order to assess service profitability.

As the said models of travel demand and transportation supply involve spatial finesse, they are solved numerically. The set of analytical conditions to depict e.g. an optimal system state is typically very large so that little insight might be gained from its inspection. By contrast, the models presented hereafter are analytical, owing to the postulates of ring shape and, more generally, homogeneity in space and time. Then, the models are endowed with circular symmetry, which is crucial to characterise the system state in a simple way and to obtain the few system state variables as easy-to-interpret analytical formulas of the model parameters.

4.1 The Orbicity modelling family

We have developed a family of models called Orbicity, designed to study ring-shaped mobility services in urban environments. As of early 2020 the family comprises three models:

- a ***taxi service***: vehicles take individual clients, picking them up from their start point and dropping them off at their destination [1].
- a ***sharing service for individual vehicles***, typically two-wheeled vehicles or small cars which do not require driving licences: users pick up the vehicle from its parking spot, make their journey and leave the vehicle at their destination [2].
- a ***shuttle service***: vehicles with a capacity of K circulating on a ring-shaped route, with vehicles running in both directions. They stop to pick up and drop off passengers at their requested locations [3].

4.2 Architecture of an Orbicity model

Each model has a four-tiered structure which covers, from bottom up, [i] Traffic conditions and service operations, [ii] Demand in equilibrium to supply, [iii] Service management, [iv] Technology and regulation.

i. ***The technical performance of the service***, with a fixed fleet size N, in order to serve a fixed level of demand Q which is compatible with N. This model gives us the mean journey time t_R and access time t_A for each journey, along with indicators useful for designing the service offered.

ii. ***The balance of supply and demand***: with a fixed offer in terms of fleet size and pricing, the volume of demand Q can be modelled as a function of the fares charged and the waiting t_A and journey t_R times. But these times are themselves dependent on Q, as per the traffic model. These two causal systems are interlinked, and together they determine the state of equilibrium between supply and demand in terms of volume Q and time t_A and t_R.

iii. ***Service management***. Production of this service aims to optimise a certain objective function, which is determined by the regime of regulation. In cases involving an unregulated monopoly, the objective function will be the net profit accrued by the operator, equal to the difference between commercial revenue and production costs. The producer determines the fleet size and pricing policy with a view to maximising this objective function. Maximisation is realised by taking into account both the technical requirements of the service and elasticity in demand.

iv. ***Service policy depending on regulation regime***. It is at this upper level that the structuring characteristics of the service are decided:

    a. The regulatory regime corresponds to one of several available options – unregulated monopoly (MO), optimal system with no budget constraints (SO), optimal system with the constraint that the producer must balance their budget (S2) or return on public funds (S1).

b. The technology corresponding to the mode of transportation: vehicle capacity (factor K), type of motor (combustion engine or electric), driving technology (human or robotic).

c. Territorial conditions: size of the ring (circumference C or radius R), average speed of travel (parameter v_0), demand function.

This upper level thus determines the strategic conditions of the service. For the service operator, this is a matter of strategic marketing. For a public authority responsible for coordinating sustainable mobility, it is a matter of strategic planning.

Figure 3 offers a schematic representation of this four-tier architecture.

4.3 Indications regarding the traffic model

Figure 4 shows the chain of causation which determines the traffic conditions. The traffic model establishes a 'stochastic equilibrium' for the availability of

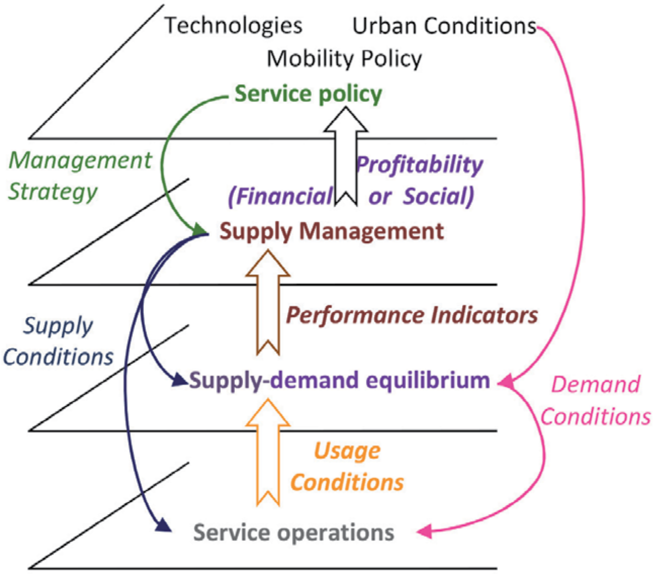


Figure 3.
Four-tier architecture.

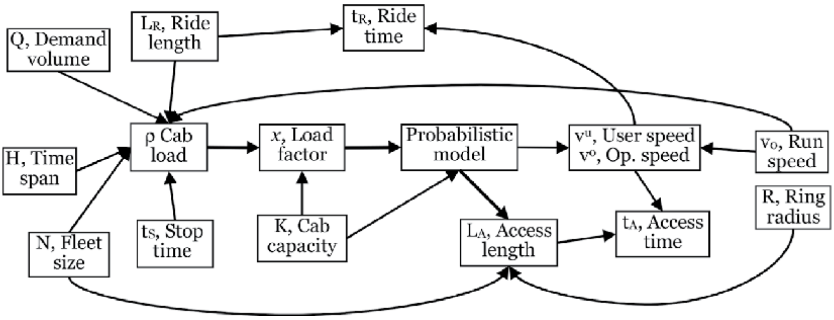


Figure 4.
Systemic diagram of the traffic model for a ring-format shuttle service.

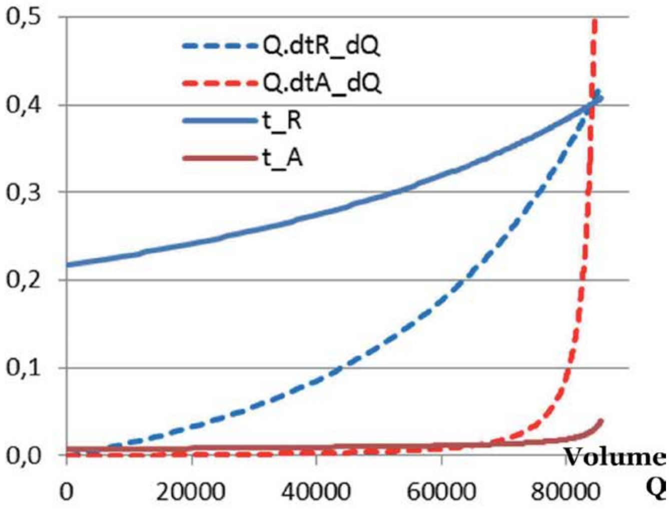


Figure 5.
Traffic laws for a ring-format shuttle service (under N = 200).

vehicles, taking into account the components of demand and their fluctuation along the ring route, over time and in terms of ride length. This gives us the average speed of the service, with the operating speed per vehicle, which is of interest to the operator, as well as the average usage speed, of interest to users. These speeds will also determine the average ride time and access time.

Figure 5 shows the variations in average access time t_{A} and ride time t_{R} in response to the volume of demand, for a fixed-size shuttle fleet.

4.4 Indications regarding the supply-demand equilibrium

Figure 6 shows the variations in the volume of demand in response to price, or rather the generalised cost (the sum of the ticket price plus access and ride times, weighted by their respective 'values of time'), for a fleet of fixed size and for a demand function with constant elasticity of -2 with respect to generalised cost. The blue curve shows the 'original' demand function for exogenous traffic conditions, while the red curve illustrates the 'adjusted' demand function which takes into account the interaction between the volume of demand and the traffic conditions (simulated in the preceding tier of the model).

4.5 Indications regarding optimised service management

Table 2 shows the objective function assigned to the production of the service, for each of the different operating regimes available as options in the strategic phase. Demand function $Q(\tau, N)$ represents the volume of demand in response to trip price τ and fleet size N: it summarises both the demand model and the traffic model. Cost function $C(N, Q)$ represents the cost of producing the service, on a daily basis: this depends on the number of vehicles and number of trips required. The underlying parameters, for example the capacity of the shuttles or the energy costs involved, are to be determined at the strategic level.

Figure 7 provides a graphic representation of the solution to the production optimisation problem, in this case for a shuttle service. The primary unknown is price per ride τ, on the y axis. The variable x is a load factor: this is the key factor to resolving the traffic model. The pink function models the behaviour of demand: demanded price

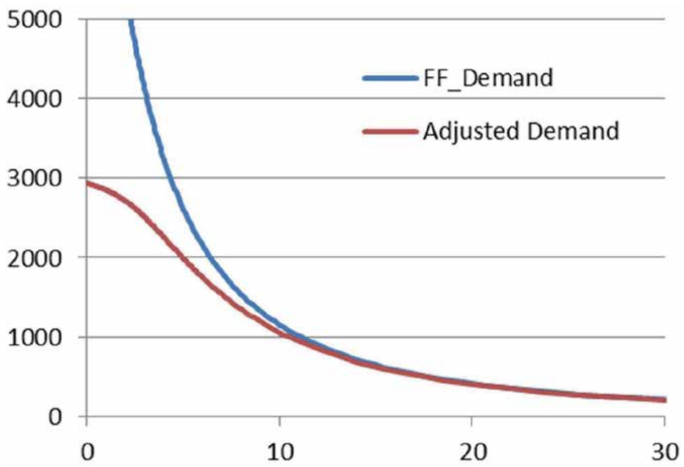


Figure 6.
Volume of demand with respect to generalised cost.

Pattern	Objective function and side constraint	Pricing rule	Notation
MO, monopoly	$\mathrm{P^o}(\tau, N) \equiv \tau.\mathrm{Q}(\tau, N) - C(N, \mathrm{Q}(\tau, N))$	$\tau = \frac{\varepsilon\left(\dot{C}_Q + \mathrm{Q}\dot{g}_Q\right)}{\varepsilon+1}$	$\tau = \hat{\tau}^{\mathrm{o}}_{\mathrm{MO}}$
SO, system optimum	$\mathrm{P^{ou}} \equiv \mathrm{P^o} + \mathrm{P^u}$ with $\mathrm{P^u} \equiv \int_0^Q \mathrm{D}^{(-1)}(q).\mathrm{d}q - Q.g$	$\tau = \dot{C}_Q + \mathrm{Q}\dot{g}_Q$	$\tau = \hat{\tau}^{\mathrm{o}}_{\mathrm{SO}}$
S1 = SO under min return	$\mathrm{P^{ou}}$ under $\mathrm{P^u}/(1+b) + \mathrm{P^o} \geq 0$ Min benefit per € of public funds, b	$\frac{C}{Q} - \frac{g_T + C/Q}{1-(1+\varepsilon)(1+b)}$	$\tau = \hat{\tau}^{\mathrm{o}}_{\mathrm{S1}}$
S2 = SO under budget balance	Min $\mathrm{P^{ou}}$ subject to $\mathrm{P^o} \geq 0$	$\tau \geq C(N, Q)/Q$	$\tau = \hat{\tau}^{\mathrm{o}}_{\mathrm{S2}}$

Table 2.
Objective function and pricing rule depending on regulation pattern.

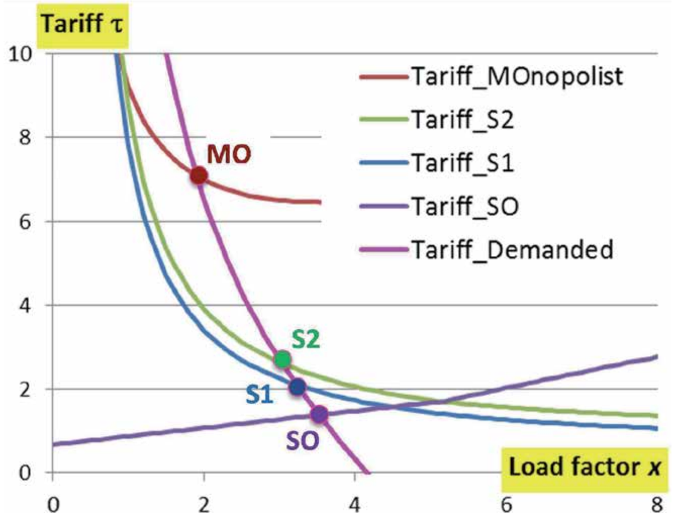


Figure 7.
Supply-demand state is the solution to a fixed-point problem.

$\tau^{u}(x)$. Each of the other curves represents a specific pricing policy adopted by the producer: proposed price $\tau^{o}(x)$. Optimising production consists of levelling up demanded price and proposed price: for each producer behaviour, the corresponding solution is the point of intersection (x^{*}, τ^{*}) which ensures that $\tau^{o}(x) = \tau^{u}(x)$.

5. Exploratory study

Using a systems model allows us to simulate different scenarios, varying the values assigned to exogenous parameters. In this section, we will explore in greater detail the potential parameters of a shuttle service operating on a ring basis, focusing on two major investigative variables: technology and regulation.

5.1 The strategic domain

These two investigative variables represent the two dimensions of our strategic domain. We will now cross-compare four different levels of technology and three potential regulatory regimes, giving 12 possible scenarios.

These levels of technology in fact represent successive technological generations which succeed one another in a process of progressive accumulation. Level zero, the 'pre-platform' generation, corresponds to the situation which prevailed before the emergence of platform-based services. For each trip, we take into account a user transaction time of 2 minutes as an additional ride time in the generalised cost. The first level corresponds to the platform era, defined by the advent of platform technologies and the radical simplification of transaction operations for customers. We use the abbreviation PF to represent this level. The 2nd generation of the platform uses electric vehicles, better suited to the urban environment and more economical when used intensively. We can abbreviate this as PF + EV, where EV stands for Electric Vehicles. The 3rd generation of the platform might incorporate self-driving shuttles, drastically reducing the primary cost of production. We can abbreviate this to PF + EV + AD, where AD stands for Autonomous Driving.

The regulatory axis comprises three possible regimes: monopoly (MO), first-best system optimum (SO), second-best system optimum (S2). Abbreviation MO stands for an unregulated monopoly, with a single operator dedicated to maximising profit. Abbreviation SO stands for a socio-economic system optimum of both supply and demand; then the objective function includes the net profit to the operator and the net surplus for users (the total reserve price less the generalised cost). Abbreviation S2 stands for an optimised system subject to the constraint of balancing the production budget. The objective function is the same as for SO, but optimisation is restricted to all values of the paired parameters (fleet size and price) which ensure a non-negative return for the service provider. As such, S2 is an intermediate solution between SO and MO.

5.2 The projected service

We have chosen to focus on a collective service running shuttles whose capacity $K = 12$ places. The ring format is consistent with the service cycle principle, and the vehicle capacity corresponds to the shared occupancy associated with public transport.

The limited capacity of the vehicles offers several advantages. First, as for the physical form of the vehicles, limiting their size should allow for a certain degree of agility on the road network, and reduce the disruption of other traffic caused by the shuttles stopping to pick up/drop off passengers. Second, per user ride, limiting capacity enables us to reduce the number of stops required to pick up/drop off other passengers, thus increasing the usage speed. Third, in order to satisfy a given volume of ridership, the fleet size will need to be bigger than it would if using vehicles with a larger capacity (e.g. around fifty passengers in a standard city bus); this makes it possible to increase the frequency of the service, and thus to reduce the time users wait to access it. Fourth, on an industrial level, the manufacturing of larger fleets of vehicles allows for greater economies of scale and drives down the cost price per vehicle. We might also expect to see greater flexibility in the internal logistics of the service: recharging, cleaning, maintenance and repairs. Fifth, regarding the environment, the impact of the fleet of vehicles over its whole life cycle may be reduced: this applies to the construction phase, due to the industrial efficiency gained, and the usage phase, thanks to the increased operating speed.

Nevertheless, using collective vehicles of a smaller size presents at least two disadvantages. The first is that the cost of driving the vehicles is increased proportionally to $1/K$. The second is the sacrifice of potential economies of scale for vehicles powered by combustion engines. However, these disadvantages could feasibly be swept away by the advent of self-driving vehicles and the rise of electric motors.

For all of the simulation scenarios envisaged here, the common parameters are as follows. The territorial aspects involve:

- A ring route with a radius of R = 2.3 km.
- Driving speed $v_0 = 20$ km/h.

For demand:

- An average journey distance of $L_R = 3.7$ km.
- Benchmark volume of $Q_0 = 26{,}000$ journeys per day at a generalised cost of €2 per trip. The elasticity of the volume of demand to this generalised cost is set at -2.
- Values of time are set at €12/h for running time and €8/h for access time.

Production parameters:

- For each day of activity, service in operation for H = 14 hours.
- For each shuttle, a daily cost price of €600 (resp. 500) for electric (resp. combustion) vehicles with drivers and €100 for self-driving electric vehicles.
- Energy cost: €0.05/km for an electric motor, or €0.15/km for a combustion engine.

5.3 Results and discussion

Table 3 presents the main results for each of the 12 scenarios simulated, with one scenario per column. The scenarios are grouped by technological generation, in ascending order of technological progress. For each generation, the three regulatory regimes are presented in the order MO – S2 – SO.

Taking all of these scenarios into consideration, the following indicators emerge. The number of trips per day varies from 455 to 8000 depending on the scenario. The size of the fleet varies from 4 vehicles, for an MO regime in the pre-platform

Technology	**0: preplatform**			**1: platform**			**2: PF + EV**			**3: PF + EV + AD**		
Regulation	**MO**	**S2**	**SO**	**MO**	**S2**	**SO**	**MO**	**S2**	**SO**	**MO**	**S2**	**SO**
Load factor x	1.72	2.82	3.5	1.9	3.08	3.8	1.85	3.1	4.0	1.4	2.0	2.05
Ride price τ	8.75	3.45	1.59	8.02	3.2	1.55	9.0	3.8	1.7	4.4	1.1	0.7
Gen. cost g	14.6	8.3	5.93	13.7	7.5	5.45	15	8.5	5.8	7.4	4.0	2.6
Demand Q	455	1510	2913	554	1858	3500	450	1500	3100	1900	6800	8000
Fleet size N	4.0	8.9	14.4	4.5	10.1	16.4	4	8	14	20	50	60
Costs C	2600	5210	8280	2855	5860	9300	2800	5500	9400	3000	7100	8200
Revenues R	3980	5210	4640	4450	5860	5200	3400	5500	5060	8000	7100	6000
System profit	8270	12550	13760	9193	13900	15040	8100	12200	13500	19000	26400	29000

Table 3.
Principal results for the simulated scenarios.

era, to 60 for an SO system running PF + EV + AD: this fifteen-fold increase needs to be seen in light of the respective passenger numbers. The number of trips provided by vehicle and by day varies from 95 for self-driving technologies to 220 for the generation PF + EV. In the simulated range of traffic conditions, the 'Load factor' is almost equivalent to the average number of passengers per shuttle. The values range from 1.4 to 4.0, suggesting that the capacity of 12 seats per shuttle is excessive and a 6 or 8-seater alternative would be sufficient.

Average access time varies from one hundredth of an hour to one tenth of an hour. It is systematically lower in SO and S2 regimes than in MO systems using the same technology. Average journey time does not vary so substantially, ranging from 0.22 to 0.28 hours. For each technology, this time is systematically lower in MO systems than it is in SO regimes. Fare per trip varies from €0.7 to €9. For each generation of technology, it is systematically much higher in MO than in SO and S2, from the platform generation onwards. Furthermore, the pre-platform fares in S2 systems are lower than those charged in the MO regime with the most advanced technology PF + EV + AD, which goes to show that offering an efficiently-organised service, combining a ring format with a suitable regulatory regime, is more important than the generation of technology deployed. The generalised cost per trip ranges from €2.6 to €15, obtained by adding the fare and the average ride and access times (weighted by their specific time values).

The commercial revenue is large, varying from €3400 to €8000 per day. Both of these extremes are found in the most advanced technological generation. Daily production costs are also substantial, ranging from €2600 to €9400 and depending on the generation of technology used and the operating system (via fleet size). Daily operating profit is nil for S2 regimes, highly positive for MO (increasingly so as technology advances) and negative for SO, with lower profits for the most advanced technologies compared with intermediate technologies. The daily surplus of demand is very large: double the total generalised cost to passengers, due to the specific elasticity of the demand function. For each generation of technology, the SO and S2 operating regimes are much more beneficial to demand than the MO system, while the difference between SO and S2 is relatively slight. The socio-economic benefits of the service ('system profit Puo') are massive: these benefits increase with each technological generation, and within each generation of technology the profit is greater for system optimal regimes than for the unregulated monopoly (in keeping with the definition of the system optimum).

In summary, the configuration of the service requires a large fleet of vehicles, indeed a very large fleet to make the system 'optimal'. Optimising operations gives us a number of trips per shuttle and per day ranging from one to two hundred. This gives us an idea of scale. A vehicle capacity of 6 or 8 places should be sufficient.

Generally speaking, technological progress leads to quantitative and qualitative improvement of the service. However, switching to fully self-driving technology alters the economics of production and reconfigures certain parameters: for some indicators, the variation in performance as technology advances is a U-shaped curve rather than a steady increase. The regulatory regime has a major influence, which becomes increasingly important with each successive generation of technology. S2 systems are most compatible with achieving financial equilibrium in the production of the service, while also getting as close as possible to the system's first-best socio-economic optimum. They allow for relatively moderate tariffs compared with monopoly scenarios (which operate at the same rate as pre-platform era taxis).

The target value for the most advanced generation of technology is €0.7 per trip, for an average length of 3.7 km. For intermediate technological generations, S2 prices are around €3.5 per trip. These values provide points of reference which can help us to gauge the current economics of public transport services.

6. Conclusion

When designing shared mobility services, ring-shaped configurations allow us to cover a relatively large territory with a cyclical service provided by shared vehicles. These advantages mean that we may expect a relatively high level of demand, as well as a service which is economically attractive. By combining them we can hope to achieve a symbiosis between these two qualities, forming a virtuous circle: increased demand, more profitable production, expansion of the service allowing for greater quality of service and reduced fares, and so on and so forth.

We have also considered a number of potential geographical configurations, and the technical and economic conditions required to make them profitable. A ring-shaped system, combined with platform management and the use of electric vehicles, allows for prices in the order of €3.2/veh.km while maintaining the budgetary equilibrium of the service provider. By switching to fully self-driving vehicles, this equilibrium price could fall to as little as €0.6/veh.km.

These simulations are based upon the 'Orbicity' technical and economic model, focusing on a ring-shaped shuttle service. The model holds that demand is elastic to the trip generalised cost, including the fare and the access time weighted by their respective values of time. In practical terms, this elasticity means that the service must provide a competitive alternative to travelling by car, in terms not just of price but also of user comfort: comfort on board the vehicle, ease of access, quality of information, and the smoothness of transactions.

This exploratory study will need to be followed up with further research focusing on urban integration. This research should take several geographical levels into consideration, from local to district and up to metropolitan. The specific local level involves the layout of the ring roadway, as well as the adaptation to junctions to avoid traffic hindrance. At the district level, the ring can be targeted to a specific function within the city, for example protecting the epicentre by providing a system for passenger flow distribution and access. Finally, at the metropolitan level, the issue is to embed ring-shaped on-demand shuttle services in a coherent transport strategy. Various schematic solutions could be envisaged: for example, a star-shaped network with a central ring connecting to radial routes, or else multiple rings forming a patchwork across the whole urban area (with or without points of intersection).

Acknowledgements

To Jean Grébert (expert leader on mobility systems for Renault) for his suggestion of the loop transit system, and to Jaafar Berrada (Védécom) for our stimulating discussions during his doctoral thesis, which I had the pleasure to supervise on related issues.

Conflict of interest

The author declares no conflict of interest.

Author details

Fabien Leurent
Ecole des Ponts ParisTech, Champs-sur-Marne, France

*Address all correspondence to: fabien.leurent@enpc.fr

IntechOpen

References

[1] Leurent F. (2019a) Microeconomics of a taxi service in a ring-shaped city. hal-02047269. https://hal.archives-ouvertes.fr/hal-02047269

[2] Leurent F. (2019b) Modèle probabiliste d'un service de trottinettes. ENPC Course document, 7 p.

[3] Leurent F. (2019c) On a Ring-Shaped Service of Collective Taxis, Part 1: Traffic Physics and Service Quality. https://hal.archives-ouvertes.fr/hal-02090947

[4] Leurent, F. (2020). Vers une mobilité partagée en forme olympique ? Communication au congrès RDMI 2020, Atelier A3: Mobilité partagée. Congrès annuel ATEC-ITS: RDMI 2020, Jan 2020, 92120 MONTROUGE, France. https://hal.archives-ouvertes.fr/hal-02461695/

[5] Leurent F., Haxaire O. et Lesteven G. (2018) La Smart Mobility: un paysage en voie d'aménagement. Chapitre 21 dans Peuportier B., Leurent F. et Roger-Estrade J. (coord.) *« Ecoconception des ensembles bâtis et des infrastructures, tome 2 »*. Presses des Mines, Paris, pp. 487–540.

[6] Andreasson, I., Leurent, F. and Rossetti, R. (2016). Future Developments and Research Topics. In Gentile, G. & Noeckel, K. (eds): *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*, Springer Tracts on Transportation and Traffic, Vol. 10: 561–641.

[7] Boutueil, V. (2018). New Mobility Services. Chapter 2 in Aguiléra A. & Boutueil V. (eds) *Urban Mobility and the Smartphone*, Elsevier, pp. 39–78.

[8] BCG (2016) Will Autonomous Vehicles Derail Trains? Strategic study report, September 2016, 20 p.

[9] Delons, J. (2017). Sur la massification et la rentabilisation d'un service de covoiturage domicile-travail en Ile de France. Communication à l'Université d'été de la Chaire Ecoconception, Paris, 20 septembre 2017.

[10] https://en.wikipedia.org/wiki/Circle_Line

[11] TCQSM (2013), Transit Capacity and Quality of Service Manual 3rd edition. Transit Cooperative Research Program, Transportation research board (TRB) http://www.trb.org/Publications/Blurbs/169437.aspx

[12] Ortuzar, J.D., Willumsen, L.G. (2011). Modelling Transport (fourth ed.). Chichester: John Wiley & Sons Ltd.

[13] Gentile, G. & Noeckel, K. (eds): *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*, Springer Tracts on Transportation and Traffic, Vol. 10.

Chapter 9

Transit Signal Priority in Smart Cities

Bahman Moghimi and Camille Kamga

Abstract

Giving priority to public transport vehicles at traffic signals is one of the traffic management strategies deployed at emerging smart cities to increase the quality of service for public transit users. It is a key to breaking the vicious cycle of congestion that threatens to bring cities into gridlock. In that cycle, increasing private traffic makes public transport become slower, less reliable, and less attractive. This results in deteriorated transit speed and reliability and induces more people to leave public transit in favor of the private cars, which create more traffic congestion, generate emissions, and increase energy consumption. Prioritizing public transit would break the vicious cycle and make it a more attractive mode as traffic demand and urban networks grow. A traditional way of protecting public transit from congestion is to move it either underground or above ground, as in the form of a metro/subway or air rail or create a dedicated lane as in the form of bus lane or light rail transit (LRT). However, due to the enormous capital expense involved or the lack of right-of-way, these solutions are often limited to few travel corridors or where money is not an issue. An alternative to prioritizing space to transit is to prioritize transit through time in the form of Transit Signal Priority (TSP). Noteworthy, transit and specifically bus schedules are known to be unstable and can be thrown off their schedule with even small changes in traffic or dwell time. At the same time, transit service reliability is an important factor for passengers and transit agencies. Less variability in transit travel time will need less slack or layover time. Thus, transit schedulers are interested in reducing transit travel time and its variability. One way to reach this goal is through an active intervention like TSP. In this chapter a comprehensive review of transit signal priority models is presented. The studies are classified into different categories which are: signal priority and different control systems, passive versus active priority, predictive transit signal priority, priority with connected vehicles, multi-modal signal priority models, and other practical considerations.

Keywords: transit signal priority, traffic system communication, adaptive control, connected vehicles, smart cities

1. Introduction

Transit-Oriented Development (TOD) has been widely considered as a wise approach to make the public transit more attractive. Transit Signal Priority (TSP) is a promising and key tool in support of TOD strategies. The first research about TSP dates back to the 1970s [1]. TSP has evolved, together with its application tested and implemented through the advances in intelligent transportation technologies. TSP is an operational treatment that facilitates the movement of public transit,

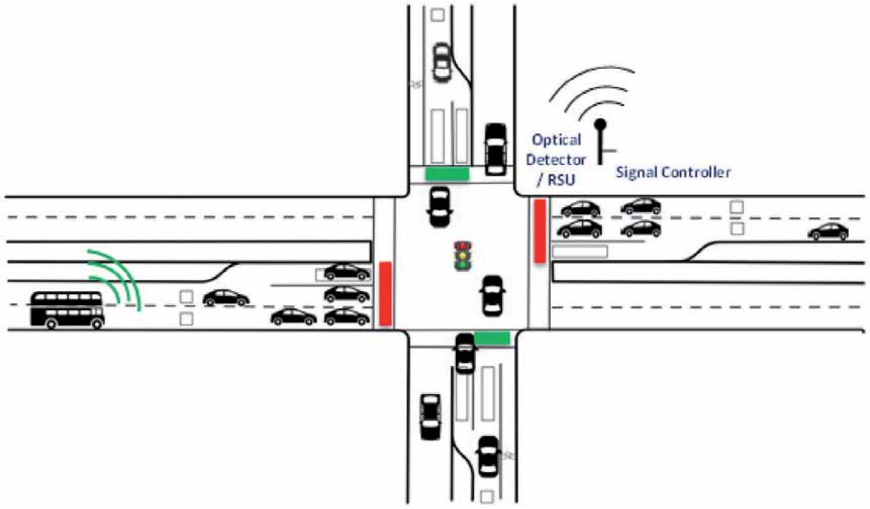


Figure 1.
Transit signal priority framework.

either busses, tramways, or streetcars, through signalized intersections without the interruption of red signal lights when possible (see **Figure 1**). It can be an effective strategy for lowering transit vehicle delays at signalized intersections as well as passengers delays, reducing fuel consumption and vehicle emissions, preventing bus-bunching, and significantly enhancing transit reliability which results in transit with lower passenger waiting time and operating costs (smaller transit fleet size).

The Transit Capacity and Quality of Service Manual [2] described bus preferential treatments at intersections including transit signal priority, queue jumping, curb extension, and boarding islands. Among the stated transit treatments, TSP has the lowest cost and can easily be implemented in dense urban transportation networks, while other treatments require more capital and physical spaces to be considered. In practice, transit agencies are more interested in making use of limited resources in an efficient manner and TSP could satisfy such needs. However, the impact of TSP will prove more effective with the use of extensive evaluation, ongoing performance monitoring, and adjustment after the initial implementation [3]. By 2015 [4], 109 cities around the world, mostly in North America and Europe, have implemented TSP. However, the majority of attempts used simple and easy-to-implement logics in practice. Portland, Los Angeles, Chicago, and New York City in the United States have been the pioneer cities in implementing TSP in order to enhance their transit system. Various technologies have been deployed for TSP including: the Loopcom System in Los Angeles, CA, the Amtech System in Seattle, WA, the TriMet in Portland, OR, and wireless local area network in Minnesota, MN and in New York City, NY [5–7].

2. Transit signal priority tactics

The goal of all tactics is to grant green time when transit vehicles are approaching intersections, which would offer them shorter delays. To do so, a combination of TSP tactics is given the best result. Thus far, a lot of tactics have been used; most of them are listed below.

Green Extension is one of the main TSP tactics that has been used in most TSP studies. It extends the green time when the transit vehicle is expected to arrive a few seconds after the end of green time, which opens a short green-window for the transit vehicle. It provides a relatively large bus delay-reduction, but only for a small fraction of busses that arrive at the signals approximately the end of green phase [8, 9]. The maximum green extension is an input value in the TSP logic. Based on that, and together with bus speed, a maximum point/time at which transit can be detected will be measured [10].

Early Red or Green Truncation is another main TSP tactic scheme that has been widely implemented. It is granted to accommodate transit vehicle that would arrive a few seconds before the start of green. Green truncation cuts down the green time of the conflicting or non-transit phases so that the signal cycles faster and provides a green window sooner to the transit vehicle that stopped behind the traffic lights. This tactic serves all busses caught at red light, providing a small benefit to many [8, 9]. Since the early red technique cuts down the green of the conflicting phases, while there are vehicles in the queue, it causes greater disruption to the conflicting traffic and it is recommended to be applied to intersections with low to medium interruptible bus requests.

Phase Rotation is a TSP technique that has been recently emerged in studies. It changes the sequence of the phases in the ring barrier signal system to match the bus's arrival time with the green light. It mostly changes a leading left phase to a lagging left phase or vice versa. Thus, the application of phase rotation tactic comes into play when there is a left-turn phase in the ring barrier with more than two critical phases. For example, it assumes that the initial signal timing design is set to be leading left and lagging through. Assuming that the bus runs in the through-phase, the phase rotation tactic changes the sequence of the phase in such a way as to make the through-phase lead and the left-turn phase lagged. The effect of phase rotation is ineffective where bus volume is high [9].

Phase Insertion is another tactic used to provide green time for transit vehicles. This tactic strategy is applied mostly when there are several turning movement phases, and the logic is not flexible enough to handle such a case when the transit vehicle arrives in the middle of a red light, making the logic unable to apply other TSP techniques like Green Extension or Early Green. Instead, it creates a temporary green phase within a cycle. Sometimes it is called Double Realization because the transit-phase receives green time twice in one cycle. Double realization is used for a left-turn phase at an intersection with high bus requests [9]. The phase sequence used was leading left - through movement - lagging left (for the second time), which all were applied in one cycle, at a signalized intersection near a major bus terminal [9].

Queue Jump is another transit preferential treatment that stretches a short bus-lane at traffic signals, together with a specific queue jumper leading phase interval. This allows busses to receive green light sooner in order to be ahead of the queues backed-up in the adjacent lanes [11]. Queue jump was applied in a few studies and the results have shown that the combination of queue jump treatment and signal priority tactics have yielded the highest benefit in terms of transit travel time, speed, and delay as compared to scenarios in which each one is deployed separately [12, 13]. Cesme and Altun [14] studied the effect of queue jump lane on a hypothetical intersection. The results revealed that a queue jump lane corresponding to the 95th percentile queue has reduced the transit delay 1.3 times more than that of queues measured by the 35th percentile.

There are other transit signal tactics that have been introduced, which in a sense is the combination of the above listed tactics; some of which include: transit phase truncation and queue dissipation [15], early red, flush-and-return [16], expedited return [17, 18], etc.

3. Signal control systems and TSP

To better understand how to treat transit differently, diverse types of signal control systems need to be explained. Traffic signal system is one of the most significant parts of transit signal priority application in the emerging system of smart cities. Signals have been designed to control demand so as to improve traffic flow. There have been many developments in signal control systems. In this section, signal control models are categorized into four generations including: fixed time, coordinated-actuated, fully actuated, and adaptive signal control.

Pre-timed (fixed) signal control means that each phase has a fixed split length, resulting in a signal with a fixed cycle length. To be more responsive to traffic changes, one approach could be to use different plans according to the time of day (a.m. and p.m. peak, midday, nighttime, etc.). This way, the historical traffic demands will be used to determine signal timing plans.

The TSP on the fixed time control logic indicates whether the transit is projected to pass through the signal at a green light. If so, no alternative is made; but if the bus is projected to be at the stop line just after the end of green, green extension tactic will extend the green time until the transit vehicle can pass through or before the allowable maximum green. The green time required for an extension is taken from the next phase or other conflicting phases (if there are more than two critical phases). However, if the bus is at the signal while traffic from another approach is being served, the TSP logic truncates the active green phase, after the minimum green of that phase is satisfied. In the fixed-time control logic, applying TSP will reduce bus delay substantially, whereas it may increase the delay of the conflicting phases. Lack of compensation in the fixed time signal control does not allow it to recover from interruptions like TSP. That is part of the reason why many developed models have limited applying TSP over the fixed-time control at every cycle [17].

Coordinated-actuated signal control is another controlling system that is mostly being used for signals placed along a corridor. The logic provides all signals with a fixed cycle length and let non-coordinated phases behave like actuated control, aiming to enlarge green bandwidths and allow all slacks to run in the coordinated phases. Cycle length, force-off points, offsets, and phase sequences are mostly the signal timing parameters that are being optimized widely through available signal optimization software like Synchro and Transit-7F Error! Reference source not found.. There have been some optimization models developed over coordinated-actuated control to make its performance even better which can be found in [19, 20].

Applying TSP to coordinated-actuated logic is done by granting green extension to the coordinated/transit phase, early green to the actuated phases (non-coordinated phases), and TSP phase rotation whenever it is needed. As the cycle length is fixed, like the fixed-time control, the granted green extension time is taken from other conflicting phases. Actuated control with absolute priority can result in near zero delay for busses, but it sometimes causes long delays for the general traffic [21].

Actuated signal control relies on traffic data from sensors embedded in the infrastructure including loop-detectors, video detectors, or radar, to make controlling decisions. Actuated control better captures the real-time dynamic of traffic system since traffic demand may fluctuate from time to time. The fully actuated signal control run as fast (snappy) as possible to have less slack time, cycle length, and thereby less overall delay at intersections [22]. It matches supply to demand in real time. It has a feature of compensation which means if the controller gives more green time to a phase due to considerations such as TSP; the logic automatically will compensate and provide more green time to the conflicting phases in the next cycle. The faster it runs, the more efficient it will be [23–25]. The actuated traffic signal functions approximately as a fixed signal when the degree of saturation is too high (oversaturated condition).

Adaptive signal control gets feedback from detectors based on the latest update of the past 5 or 10 minutes to update and re-optimize the control plans. Adaptive control can also be designed to predict traffic flow and optimize in the anticipation of the flows expected to arrive in the next few minutes (e.g. 2 to 5 min horizon). There are several adaptive control systems being developed; some of which include: SCATS [26], OPAC [27], TRANSYT-7F [28], UTOPIA [29], SCOOT [30], RHODES [31], ACS-Lite [32], MOTION [33], and more programs are also coming on to the market. Many of them are adaptive signal systems with a centralized controller. In adaptive signals with centralized control, the complexity increases as the number of traffic lights and contributing variables increase. Adaptive control with decentralized approach has also been developed, e.g. self-organizing system, that functioned well with signal priority [18].

The use of TSP on top of adaptive signal control was developed by many scholars. TSP applied to SCATS includes green extension, special phase sequences, and compensation to the non-transit phases [34]. Transit priority on TRANSYT-7F benefited bus delay-saving by 6 s/intersection/bus [35]. TSP on UTOPIOA reported a 20 percent increase in the average bus speeds [36]. TSP with SCOOT reached bus delay-savings ranging from 5 to 10 s/signal [35]. Priority on RHODES has also increased traffic speed and reduced average and variance of bus-delay significantly [37, 38]. TSP with self-organizing system result in a very low bus delay [18, 39].

4. Passive versus active

Passive or inactive TSP refers to an initial method of signal priority which adjusts the signal timing offline while relying on the historical data. This adjustment mainly changes signal time parameters including split length, offset, and cycle length. The objective of signal setting with respect to passive TSP is to increase the probability of transit vehicles arriving at the intersection during the green interval. However, passive TSP is inflexible in adapting to the dynamic flow of traffic and bus conditions. The reason is that passive priority always provides a green light to transit even if there is no transit vehicle; not to mention about the delay it would cause to the other conflicting phases by giving ineffective green to the bus-phases. Passive priority becomes more effective when the traffic volume is light or moderate, with high transit frequencies, and predictable transit travel time [40]. Passive priority is cheap and easy-to-implement; both are advantages, since the transit detection and communication equipment are not required. It is worth noting here that preemption priority applies priority tactics abruptly. This is sometimes done by interrupting signal operation by skipping phases or terminating pedestrian clearance time, in order to permit a specific vehicle (e.g. ambulance) pass through the traffic light. Preemption can be considered as the highest level of priority, which is frequently used for emergency vehicles [41].

Contrary to passive priority, active TSP is about granting priority tactics in real time and only to those transit vehicles that are present or about to approach the signalized intersections. In an active priority system, the real-time information regarding transit vehicles' speed and location should be detected. Some standard vehicle/bus detection techniques are inductive loops, infrared, and radio based systems which are considered as static detection or selective vehicle detection (SVD) [42, 43]. On the other hand, the automatic vehicle location (AVL) system is another transit detection approach that provides dynamic monitoring of transit location. Taking into account the use of detectors, TSP logic is activated when the transit vehicle passes the check-in detector, which is located upstream of the signal. Where to put the check-in detector is not deterministic and its optimal location is mostly related to traffic demand, and signal timing. The result demonstrated that putting a detector between 450 ft. (150 m) and 900 ft. (300 m) upstream of the intersection can output better results [43]. Meanwhile, the detection should cancel out the priority request when transit passes the stop-line detectors (check-out detectors). Those are located just after the traffic light, indicating the transit vehicle received priority, could pass through, and it is the time to start compensating the amount of time taken from the conflicting phases. Active TSP has been demonstrated as a better approach to improve transit performance, to better accommodate uncertain arrival time, and make on-street transit more reliable, faster, and cost-effective [42, 44]. Active TSP has been taken into consideration worldwide. For instance, applying active signal priority was studied on the two old and large street-car systems in Melbourne, Australia, and Toronto, Canada [45]. The results confirmed that such an approach is a cost-effective approach to manage traffic systems.

Song et al. [46] compared the GPS-based TSP and traditional TSP on two corridors in Utah, and it was found that GPS-based TSP reduced the same delay and travel time similarly to the traditional TSP. Surprisingly, the GPS-based signal priority system was effective in the flexible detection zone and could bring conditional priority into its logic while causing smaller impact on the side-street traffic. Active priority has recently focused its attention not only on the presence of transit vehicles, but rather on applying priority logic based on some conditions.

Unconditional priority means granting TSP tactics to the upcoming transit vehicles regardless of cross-street traffic or queue length, state of signal, or transit arrival time. It is more of an aggressive approach toward granting priority. In other words, unconditional priority is beneficial in improving bus delays, travel time, and reliability when the bus frequency is low, and when the traffic demand over signal is low. On the other hand, conditional priority grants transit signal priority only if the state of signal and bus arrival meet some defined requirements. For instance, conditional TSP can be applied if some of the following criteria are met: transit is behind schedule (e.g. let us say 5 min behind as being late), transit passenger-occupancy is more than a defined threshold, the intersection is under saturated level, no queue spillback is happening, the signal did not have a priority request in the previous cycle, etc. It is more complicated than the unconditional priority because it needs more updated information about transit and intersections. Conditional priority will improve bus headway irregularity, crowding, and mean running time to almost the same levels as what absolute TSP. More importantly, conditional TSP makes transit running time less varied (less standard deviation of running time), which indeed improves the reliability of transit scheduling service. The performance of conditional TSP was studied and found that it is more effective for bus routes experiencing more severe lateness [47–49]. Meanwhile, person-based signal priority approach has been recently introduced optimizes signals and applies conditional transit priority based on transit and vehicle passenger-occupancy conditions [50].

5. Transit arrival consideration

Predictive transit priority approaches give more flexibility to the controlling logic to enlarge the scope of signal timing to adjust itself more ahead of the transit arrival time and makes less adverse impact on the conflicting traffic. An accurate transit prediction can be a hint to passengers' departure time from point to point, so as to create more successful transfers at stops. It helps transit agencies to control and monitor their systems with more responsiveness through real time dispatching and scheduling, and making the transportation system a more resilient one. However, there are many parameters involved in the prediction of transit travel time, some of which are listed by [51] including: stochastic traffic flow uncertainties along the route, queue length in front of a traffic light, route length, uncertainties in dwell time (caused by the variation of passengers getting on and off at bus stops), weather conditions, times of day, statistical fluctuation in historical data (with large standard deviation), and GPS data error.

Uncertainty in transit running time and dwell time is mostly pronounced and has been conducted by many scholars who have been developing predictive transit arrival models. Bus dwell time itself consists of passenger boarding and alighting time, door opening and closing time, and clearance time. Hence, predicting the dwell time is cumbersome and a good prediction of transit dwell time, specifically when there is a nearside bus stop, increases the precision of transit arrival time at the target intersection. One of the primary studies about dwell time prediction at nearside bus stops was presented by Kim and Rilett [52]. They used a regression model to come up with an upper and lower bound for dwell time with respect to the bus load, headway, and schedule adherence. Such prediction was included in the improved TSP algorithm which then was applied over a fixed-time signal control. It benefited the operation of bus systems well. Lee et al. [15] developed a predictive model for dwell time at a nearside bus stop, based on headway and passenger arrival rate. Ekeila et al. [53] presented a linear regression and applied empirical Bayesian and Kalman filtering refinement to improve the prediction performance. The developed model, applied to LRT's arrival time (including running time and dwell time) to predict its boundary length, was one standard deviation from its arrival time.

Some researchers have attempted to predict the transit arrival time further ahead of the target signalized intersection. Their focuses were mostly on the prediction of the transit travel time (running time) together with the dwell time consideration. Zlatokovic et al. [54] developed predictive priority for LRT in Salt Lake City, Utah which could reduce train travel time by 20–30%. The logic used almost all TSP tactics along with peer-to-peer communication between intersections. With the peer-to-peer communications, the logic activated priority when the train was stopped at the adjacent intersection, pointing to the target intersection to be prepared for the arriving train. Wadjas and Furth [55] used advanced detectors in order to detect the light-rail's arrival more ahead of the target intersection. Once the transit is detected, the logic applies the cycle length adaptation which lengthens or shortens the cycles/phases in such a way as to find the best match aligned with the TSP tactics. The logic applied to the signals with the fully actuated control. They found that the logic functioned better in enhancing transit travel time and regularity compared to the simple preemption, with a negligible impact on general vehicular traffic and pedestrians. Moghimi et al. [18] developed a model that calculates the expected remaining dwell time, added to the adaptive bus running time estimation, at each time-step of the simulation in order to predict the bus arrival time. Their presented model reduced bus net delay to less than 5 seconds per intersection. Moghimi et al. [56] also developed a look-ahead TSP to include a longer range of prediction in order to provide transit priority far in advanced. Their presented model outperformed the conventional TSP model and improved the reliability of travel time significantly.

6. Signal priority with connected vehicles

Some of the advanced TSPs are based on the wireless communications using Global Positioning System (GPS). These systems only report instantaneous vehicle location data. Also, with the advances in emerging technologies, vehicles can communicate with each other (V2V) and with the infrastructure (V2I), through 5.9 GHz dedicated short-range communication (DSRC). Using this technology, each vehicle is equipped with an on-board unit (OBU) that broadcasts the vehicle speed, and acceleration at 10 times a second. A road side unit (RSU) is installed at the intersection to broadcast traffic signal status and also intersection geometry maps. The RSU can receive messages from surrounding vehicles and can provide better traffic resolution.

As soon as a transit vehicle enters the Dedicated Short-Range Communication (DSRC) range of intersection, it receives the map of the intersection and determines its location. Then, it broadcasts a request message and asks for priority. The RSU at the intersection receives the request and provides treatment. If the vehicle's speed changes dramatically, (e.g. the vehicle joins a queue) or if the transit vehicle stops at the bus stop, an updated request is broadcasted. The updated request is received by RSU and proper actions are planned. Connected vehicle technology can provide countable data because it updates vehicle dynamical traffic-related information like speed, acceleration, location, and other vehicle data in real time [57]. Such technology also provides the information about passenger counts (sitting/standing on transit) and at stops which can be transmitted to the signal controller in real time which makes dwell time prediction more accurate.

Hu et al. [5] used TSP with connected vehicles (TSP-CV) technology and compared their logic with conventional TSP and no-TSP scenarios. Results reported that the proposed logic reduced bus delay between 9–84% as compared to conventional TSP, as well as outperforming the no-TSP scenario. Meanwhile, it was shown that as volume-to-capacity ration increases (approaching to v/c equal to 1), the difference between TSP using connected vehicle and conventional TSP decreases. Hu et al. [58] continued their studies on conditional TSP-CV and proposed a person-based optimization method along with recommending a desired speed to the bus. The conditional logic was applied only to busses that are behind schedule and it was tested on two closely spaced intersections with fixed-time control. The results revealed that conditional TSP-CV performed better, specifically when the v/c is under saturated, and again it was found that as demand increases, when it approaches capacity, the benefit of TSP decreases. Lee et al. [59] tested the application of TSP in connected vehicle technology over a smart road test bed in Blacksburg, Virginia. The experiment results confirmed that the TSP-CV logic provided bus green extension with a 100% success rate, together with reducing bus delay between 32 to 75% as compared to No TSP scenario.

7. Multimodal traffic signal

In transportation systems, there are many users at signalized intersections that consist of passenger cars, commercial vehicles, busses, streetcars, emergency vehicles, snowplows, bikes, and pedestrians. The idea of providing TSP is not just to prioritize public transit but making other modes of transportation remain relevant in any developed TSP system. The conventional traffic controlling system treats all vehicles of different class/mode as an aggregate flow of traffic into a signal flow. This approach does not adequately consider each mode based on weight per se and is not well-aligned with the system operating objective [60]. On the other hand, treating each mode separately would result in a sub-optimal system performance [61]. The better manner of treatment is to come up with an algorithm that can consider all traffic modes based on their weight in order to reach the overall objective function. Recently, many new researchers have been developing algorithms that can better quantify such objective function including all modes and then make the TSP algorithm more holistic.

He et al. [62] formulated a mixed integer linear program (MILP) for robust multiple priority requests with different modes including busses, pedestrians, and cars. The mixed-integer nonlinear program was used by Christofa and Skabardonis [63] to minimize total person delays, while assigning priority to transit based on the passenger occupancy. It is true that at one signalized intersection, pedestrian is a dominant mode, at another one, bus or truck is a dominant mode. Hence, a better approach would be to make the signals friendly with respect to the relative importance of each signal. Zamanipour et al. [64] developed a new approach that capture a relative importance of each section of the traffic signal and then establish a priority policy for that. They enhanced the work done by [62] and used such a flexible implementation algorithm that considers real time actuation on top of the MILP [60]. The developed model was designed to be utilized under a connected vehicle environment and was tested in San Mateo, California and Anthem, Arizona to confirm the better performance of the multimodal control over the fully actuated control. Later, Beak et al. [65] improved the MILP model to consider peer-to-peer signal priority control in a corridor. They designed a signal priority control framework to address the limitation of the effective range of DSRC and the extent of the intersection map message.

In today's complex transit network, multiple priority requests at intersections occur frequently. Although applying a first-come-first-serve policy is a widely acceptable approach in many cities, it cannot be the best option and sometimes can perform worse compared to providing no priority [66]. Therefore, some scholars tried to challenge this approach and take better advantage of priority logic to make it more beneficial in terms of minimizing total transit delay. Head et al. [67] developed a mixed integer programming for multiple-priority requests problems based on a precedence graph. Their findings demonstrated that their model performed better in minimizing total priority delay compared to the first-come-first-serve policy. He et al. [68] developed a fast heuristic algorithm to provide priority to simultaneous multiple requests in real time using V2I communication. Ma and Bai [69] proposed a decision tree for optimizing TSP-requests sequence. Ma et al. [70] used a dynamic programming model to generate an optimal sequence for the conflicting requests.

8. Other practical considerations

Another approach being mentioned in the literature relates to providing facilities with exclusive or dedicated bus lanes and giving transits/busses exclusive right of way. The advantage of dedicated/dedicated bus lanes is to free busses from traffic interference and benefit transit operation. However, taking a lane from general traffic and assigning it to transit may increase general traffic travel time, specifically when congestion is high [71, 72]. Eicher and Daganzo [73] proposed a bus lane with intermittent priority which allows general traffic to use the dedicated bus lane dynamically when it is not used by busses. Their idea is applicable on bus route with low frequencies. Indeed, TSP associated with exclusive bus lane [13, 74] and TSP with intermittent priority [75] revealed improvement in bus travel time and its reliability. An example of a transit system with dedicated lane is bus rapid transit (BRT), which is an integrated system of facilities that plays a significant role in today's urban transportation and can reasonably improve bus speed, travel time, reliability, as well as serving as a catalyst for redevelopment [76]. TSP with BRT can produce significant enhancement, because in such a system, TSP can be applied to any time without being worried about queue length ahead of the transit [77]. In addition, simulation results demonstrate that applying TSP with BRT was the most effective scenario in reducing travel times (up to 26%) and delays (up to 64%), as well as increasing bus speed (up to 47%), when it was compared to scenarios without TSP and BRT [78].

The location of bus stops along the corridor is frequently under discussion and their placement depends on the traffic demand, geometry, and policy constraints. Bus stop locations can be far-side, near-side, or mid-block. With regard to the stop setback, the near-side bus stop can change into midblock or far-side bus stop as the setback distance increases. A far-side bus stop is mostly better than a near-side bus stop [14, 79, 80]. It can either cause a very low delay or zero net-delay, whereas a near-side bus stop can reduce a delay in a few cases like reserved bus lane, but it will cause increased bus delay depending on factors like cycle length, red ratio, dwell time, and stop setback distance [79]. Cesme et al. [14] compared three main transit preferential treatments including queue jump lane, transit signal priority, and stop location evaluation over an isolated test-intersection with a fixed cycle length, and bus headway of six minutes. After extensive simulation runs under various scenarios, the results indicated that relocating the bus stop from near-side to far-side resulted in the most delay-reduction per intersection when it was compared to the two other scenarios. Results showed that a far-side bus stop was superior to the near-side one with zero setback distance. Far-side relocation's delay-saving became smaller as the dwell time at the near-side stop increased. This can be interpreted as the signal's red time and dwell time have a lot of overlap. Bus delay with near-side stops can be reduced by lowering vehicular queue interaction through increasing setback distance or decreasing signal cycle length [14, 80].

9. Summary

In this chapter a comprehensive literature review of transit signal priority models was presented. The review was classified into various categories including: *a)* TSP tactics, *b)* transit priority and different signal control systems including fixed-time, coordinated-actuated, fully actuated, and adaptive system, *c)* passive versus active signal priority, and conditional TSP, *d)* transit arrival time and predictive signal priority, *e)* TSP with connected vehicle, *f)* multi-modal signal priority models, and *g)* other practical consideration. There is no one-size-fits-all approach in term of applying TSP. Each transit route has different characteristics and presents various challenges which needs to be addressed differently. Each TSP treatment must be evaluated on a case by case basis. Applying a combination of different transit signal priority models that can work with the state-of-the-art technology would subtly facilitate the movement of transit and provide some performance improvement in transit operations. Most recently, the concept of Complete Streets is being introduced in emerging smart cities. The application of TSP tactics to improve the efficiency of transportation network with all users in mind will integrate well with the Complete Streets approaches. In addition, to make a transit system run faster with higher ridership, specifically busses, a strategy like transit signal priority needs to be applied together with other strategically chosen improvements. These improvements are increasing service, prioritizing transit on city streets (priority in space), redesigning bus networks, balancing bus stops, upgrading technology like fare payment system, etc. Any type of operational improvement should make sense first in order to be implemented in reality.

Author details

Bahman Moghimi* and Camille Kamga
The City College of New York, New York, NY, USA

*Address all correspondence to: smoghim000@citymail.cuny.edu

IntechOpen

References

[1] R. J. Salter and J. Shahi, "PREDICTION OF EFFECTS OF BUS-PRIORITY SCHEMES BY USING COMPUTER SIMULATION TECHNIQUES," in Transportation Research Record, 1979.

[2] K. & Associates, U. S. F. T. Administration, T. C. R. Program, and T. D. Corporation, Transit capacity and quality of service manual. Transportation Research Board, 2003.

[3] T. Kimpel, J. Strathman, R. Bertini, and S. Callas, "Analysis of transit signal priority using archived TriMet bus dispatch system data," Transp. Res. Rec. J. Transp. Res. Board, no. 1925, pp. 156-166, 2005.

[4] Y. Lin, X. Yang, N. Zou, and M. Franz, "Transit signal priority control at signalized intersections: a comprehensive review," Transp. Lett., vol. 7, no. 3, pp. 168-180, Jun. 2015.

[5] K. Hu, S. Skehan, and R. Gephart, "IMPLEMENTING A SMART TRANSIT PRIORITY SYSTEM FOR METRO RAPID BUS IN LOS ANGELES," presented at the Transportation Research Board 80th Annual Meeting, 2001.

[6] J. J. Dale, R. J. Atherley, T. Bauer, and L. Madsen, "A Transit Signal Priority Impact Assessment Methodology—Greater Reliance on Simulation," in 78th Annual Meeting of the Transportation Research Board, 1999.

[7] D. T. Crout, "Evaluation of Transit Signal Priority at the Tri-County Metropolitan Transportation District of Oregon (TriMet)," in 12th World Congress on Intelligent Transport Systems, 2005.

[8] S. Altun and P. Furth, "Scheduling buses to take advantage of transit signal priority," Transp. Res. Rec. J. Transp. Res. Board, no. 2111, pp. 50-59, 2009.

[9] P. Furth, B. Cesme, and T. Rima, "Signal Priority near Major Bus Terminal," Transp. Res. Rec. J. Transp. Res. Board, vol. 2192, pp. 89-96, Dec. 2010.

[10] P. Koonce, J. Ringert, T. Urbanik, W. Rotich, and B. Kloos, "Detection range setting methodology for signal priority," J. Public Transp., vol. 5, no. 2, p. 6, 2002.

[11] TCRP, "Bus and Rail Transit Preferential Treatments in Mixed Traffic: A Synthesis of Transit Practice," Transportation Research Board, Transit Cooperative Research Program, Washington, DC, 2010. 83.

[12] M. Zlatkovic, A. Stevanovic, and R. Z. Reza, "Effects of queue jumpers and transit signal priority on bus rapid transit," in Transportation Research Board 92nd Annual Meeting, 2013.

[13] L. T. Truong, G. Currie, M. Wallace, and C. De Gruyter, "Does Combining Transit Signal Priority with Dedicated Bus Lanes or Queue Jump Lanes at Multiple Intersections Create Multiplier Effects," 2017.

[14] B. Cesme, S. Z. Altun, and B. Lane, "Queue Jump Lane, Transit Signal Priority, and Stop Location Evaluation of Transit Preferential Treatments Using Microsimulation," Transp. Res. Rec. J. Transp. Res. Board, no. 2533, pp. 39-49, 2015.

[15] J. Lee, A. Shalaby, J. Greenough, M. Bowie, and S. Hung, "Advanced transit signal priority control with online microsimulation-based transit prediction model," Transp. Res. Rec. J. Transp. Res. Board, no. 1925, pp. 185-194, 2005.

[16] M. Janos and P. Furth, "Bus priority with highly interruptible traffic signal control: simulation of San Juan's Avenida Ponce de Leon," Transp. Res. Rec. J.

Transp. Res. Board, no. 1811, pp. 157-165, 2002.

[17] F. Dion, H. Rakha, and Y. Zhang, "Evaluation of potential transit signal priority benefits along a fixed-time signalized arterial," J. Transp. Eng., vol. 130, no. 3, pp. 294-303, 2004.

[18] Moghimidarzi, SayedBahman, Peter G. Furth, and Burak Cesme. "Predictive–tentative transit signal priority with self-organizing traffic signal control." Transportation Research Record 2557.1 (2016): 77-85.

[19] I. Yun and B. Park, "Stochastic optimization for coordinated actuated traffic signal systems," J. Transp. Eng., vol. 138, no. 7, pp. 819-829, 2012.

[20] B. Park and J. Lee, "Optimization of coordinated-actuated traffic signal system: Stochastic optimization method based on shuffled frog-leaping algorithm," Transp. Res. Rec. J. Transp. Res. Board, no. 2128, pp. 76-85, 2009.

[21] P. G. Furth, "Public Transport Priority for Brussels: Lessons from Zurich, Eindhoven, and Dublin," Rep. Bruss. Cap. Reg. Univ. Libre Brux., 2005.

[22] B. Cesme and P. Furth, "Multiheadway gap-out logic for actuated control on multilane approaches," Transp. Res. Rec. J. Transp. Res. Board, no. 2311, pp. 117-123, 2012.

[23] P. Furth, B. Cesme, and T. Muller, "Lost time and cycle length for actuated traffic signal," Transp. Res. Rec. J. Transp. Res. Board, no. 2128, pp. 152-160, 2009.

[24] Moghimi, B., Safikhani, A., Kamga, C., & Hao, W. (2017). Cycle-length prediction in actuated traffic-signal control using ARIMA model. Journal of Computing in Civil Engineering, 32(2), 04017083.

[25] Moghimi, B., Safikhani, A., Kamga, C., Hao, W., & Ma, J. (2018). Short-term prediction of signal cycle on an arterial with actuated-uncoordinated control using sparse time series models. IEEE Transactions on Intelligent Transportation Systems.

[26] A. G. Sims and K. W. Dobinson, "The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits," IEEE Trans. Veh. Technol., vol. 29, no. 2, pp. 130-137, 1980.

[27] N. H. Gartner, OPAC: A demand-responsive strategy for traffic signal control. 1983.

[28] M.-T. Li and A. Gan, "Signal timing optimization for oversaturated networks using TRANSYT-7F," Transp. Res. Rec. J. Transp. Res. Board, no. 1683, pp. 118-126, 1999.

[29] V. Mauro and C. D. Taranto, "UTOPIA. IFAC Control," Paris Fr., 1989.

[30] B. Bing and A. Carter, "SCOOT: The world's foremost adaptive traffic control system," TRAFFIC Technol. Int., 1995.

[31] P. Mirchandani and L. Head, "A real-time traffic signal control system: architecture, algorithms, and analysis," Transp. Res. Part C Emerg. Technol., vol. 9, no. 6, pp. 415-432, 2001.

[32] F. Luyanda, D. Gettman, L. Head, S. Shelby, D. Bullock, and P. Mirchandani, "ACS-Lite algorithmic architecture: applying adaptive control system technology to closed-loop traffic signal control systems," Transp. Res. Rec. J. Transp. Res. Board, no. 1856, pp. 175-184, 2003.

[33] W. Brilon and T. Wietholt, "Experiences with adaptive signal control in Germany," Transp. Res. Rec. J. Transp. Res. Board, no. 2356, pp. 9-16, 2013.

[34] P. R. Cornwell, J. Y. K. Luk, and B. J. Negus, "Tram priority in SCATS," Traffic Eng. Control, vol. 27, no. 11, 1986.

[35] A. Skabardonis, "Control strategies for transit priority," Transp. Res. Rec. J. Transp. Res. Board, no. 1727, pp. 20-26, 2000.

[36] J. D. Nelson, D. W. Brookes, and M. G. Bell, "Approaches to the provision of priority for public transport at traffic signals: a European perspective," Traffic Eng. Control, vol. 34, no. 9, 1993.

[37] P. B. Mirchandani and D. E. Lucas, "Integrated transit priority and rail/emergency preemption in real-time traffic adaptive signal control," J. Intell. Transp. Syst., vol. 8, no. 2, pp. 101-115, 2004.

[38] P. Mirchandani, A. Knyazyan, L. Head, and W. Wu, "An approach towards the integration of bus priority, traffic adaptive signal control, and bus information/scheduling systems," Comput.-Aided Sched. Public Transp., pp. 319-334, 2001.

[39] Moghimi, B., Zamanipour, M., Furth, P. G., Kamga, C., Vadakpat, G., & Head, L. (2019). Signal Priority: Comparison of Two Recent Traffic Signal Control Models (No. 19-03804).

[40] A. Shalaby, J. Lee, J. Greenough, S. Hung, and M. D. Bowie, "Development, evaluation, and selection of advanced transit signal priority concept directions," J. Public Transp., vol. 9, no. 5, p. 6, 2006.

[41] X. Qin and A. M. Khan, "Control strategies of traffic signal timing transition for emergency vehicle preemption," Transp. Res. Part C Emerg. Technol., vol. 25, pp. 1-17, 2012.

[42] H. R. Smith, B. Hemily, and M. Ivanovic, "Transit signal priority (TSP): A planning and implementation handbook," 2005.

[43] H. Liu, A. Skabardonis, W. Zhang, and M. Li, "Optimal detector location for bus signal priority," Transp. Res. Rec. J. Transp. Res. Board, no. 1867, pp. 144-150, 2004.

[44] K. Balke, C. Dudek, and T. Urbanik II, "Development and evaluation of intelligent bus priority concept," Transp. Res. Rec. J. Transp. Res. Board, no. 1727, pp. 12-19, 2000.

[45] G. Currie and A. Shalaby, "Active transit signal priority for streetcars: experience in Melbourne, Australia, and Toronto, Canada," Transp. Res. Rec. J. Transp. Res. Board, no. 2042, pp. 41-49, 2008.

[46] Y. Song, M. Zlatkovic, and R. J. Porter, "A Corridor-Level Evaluation of GPS-Based Transit Signal Priority," in International Conference on Transportation and Development 2016, 2016, pp. 455-466.

[47] P. Furth and T. H. Muller, "Conditional bus priority at signalized intersections: better service with less traffic disruption," Transp. Res. Rec. J. Transp. Res. Board, no. 1731, pp. 23-30, 2000.

[48] E. Albright and M. Figliozzi, "Factors influencing effectiveness of transit signal priority and late-bus recovery at signalized-intersection level," Transp. Res. Rec. J. Transp. Res. Board, no. 2311, pp. 186-194, 2012.

[49] F. McLeod and N. Hounsell, "Bus Priority at Traffic Signals—Evaluating Strategy Options," J. Public Transp., vol. 6, no. 3, p. 1, 2003.

[50] Y. Z. Farid and E. Christofa, "Real-Time Signal Control with Transit Priority Window: A Person-Based Approach," 2017.

[51] V. Ngan, T. Sayed, and A. Abdelfatah, "Impacts of various parameters on transit signal priority effectiveness," J. Public Transp., vol. 7, no. 3, p. 4, 2004.

[52] W. Kim and L. Rilett, "Improved transit signal priority system for networks with nearside bus stops," Transp. Res. Rec. J. Transp. Res. Board, no. 1925, pp. 205-214, 2005.

[53] W. Ekeila, T. Sayed, and M. El Esawey, "Development of dynamic transit signal priority strategy," Transp. Res. Rec. J. Transp. Res. Board, no. 2111, pp. 1-9, 2009.

[54] M. Zlatkovic, P. Martin, and A. Stevanovic, "Predictive priority for light rail transit: University light rail line in Salt Lake County, Utah," Transp. Res. Rec. J. Transp. Res. Board, no. 2259, pp. 168-178, 2011.

[55] Y. Wadjas and P. Furth, "Transit Signal Priority Along Arterials Using Advanced Detection," Transp. Res. Rec. J. Transp. Res. Board, vol. 1856, pp. 220-230, Jan. 2003.

[56] Moghimi, B., Kamga, C., & Zamanipour, M. (2020). Look-Ahead Transit Signal Priority Control with Self-Organizing Logic. Journal of Transportation Engineering, Part A: Systems, 146(6), 04020045.

[57] Y. Feng, K. L. Head, S. Khoshmagham, and M. Zamanipour, "A real-time adaptive signal control in a connected vehicle environment," Transp. Res. Part C Emerg. Technol., vol. 55, pp. 460-473, 2015.

[58] J. Hu, B. B. Park, and Y.-J. Lee, "Coordinated transit signal priority supporting transit progression under connected vehicle technology," Transp. Res. Part C Emerg. Technol., vol. 55, pp. 393-408, 2015.

[59] Y.-J. Lee, S. Dadvar, J. Hu, and B. B. Park, "Transit Signal Priority Experiment in a Connected Vehicle Technology Environment," J. Transp. Eng. Part Syst., vol. 143, no. 8, p. 05017005, 2017.

[60] M. Zamanipour, K. L. Head, Y. Feng, and S. Khoshmagham, "Efficient priority control model for multimodal traffic signals," Transp. Res. Rec. J. Transp. Res. Board, no. 2557, pp. 86-99, 2016.

[61] Q. He, K. L. Head, and J. Ding, "PAMSCOD: Platoon-based arterial multi-modal signal control with online data," Transp. Res. Part C Emerg. Technol., vol. 20, no. 1, pp. 164-184, 2012.

[62] Q. He, K. L. Head, and J. Ding, "Multi-modal traffic signal control with priority, signal actuation and coordination," Transp. Res. Part C Emerg. Technol., vol. 46, pp. 65-82, 2014.

[63] E. Christofa and A. Skabardonis, "Traffic signal optimization with application of transit signal priority to an isolated intersection," Transp. Res. Rec. J. Transp. Res. Board, no. 2259, pp. 192-201, 2011.

[64] M. Zamanipour, L. Head, and J. Ding, "Priority System for Multi-modal Traffic Signal Control," in Transportation Research Board 93rd Annual Meeting, 2014.

[65] Beak, B., Zamanipour, M., Head, K. L., and Leonard, B. Peer-to-peer Priority Signal Control Strategy in a Connected Vehicle Environment, Transp. Res. Rec. J. Transp. Res. Board, 2018.

[66] M. Zlatkovic, A. Stevanovic, and P. Martin, "Development and evaluation of algorithm for resolution of conflicting transit signal priority requests," Transp. Res. Rec. J. Transp. Res. Board, no. 2311, pp. 167-175, 2012.

[67] L. Head, D. Gettman, and Z. Wei, "Decision model for priority control of traffic signals," Transp. Res. Rec. J. Transp. Res. Board, no. 1978, pp. 169-177, 2006.

[68] Q. He, K. Head, and J. Ding, "Heuristic algorithm for priority traffic signal control," Transp. Res. Rec. J. Transp. Res. Board, no. 2259, pp. 1-7, 2011.

[69] W. Ma and Y. Bai, "Serve sequence optimization approach for multiple bus priority requests based on decision tree," in Plan, Build, and Manage Transportation Infrastructure in China, 2008, pp. 605-615.

[70] W. Ma, Y. Liu, and X. Yang, “A dynamic programming approach for optimal signal priority control upon multiple high-frequency bus requests,” J. Intell. Transp. Syst., vol. 17, no. 4, pp. 282-293, 2013.

[71] G. Currie, M. Sarvi, and B. Young, “A new approach to evaluating on-road public transport priority projects: balancing the demand for limited road-space,” Transportation, vol. 34, no. 4, pp. 413-428, 2007.

[72] A. S. Shalaby and R. M. Soberman, “Effect of with-flow bus lanes on bus travel times,” Transp. Res. Rec., no. 1433, 1994.

[73] M. Eichler and C. F. Daganzo, “Bus lanes with intermittent priority: Strategy formulae and an evaluation,” Transp. Res. Part B Methodol., vol. 40, no. 9, pp. 731-744, 2006.

[74] W. Ma, W. Ni, L. Head, and J. Zhao, “Effective coordinated optimization model for transit priority control under arterial progression,” Transp. Res. Rec. J. Transp. Res. Board, no. 2356, pp. 71-83, 2013.

[75] N. Chiabaut, X. Xie, and L. Leclercq, “Road capacity and travel times with bus lanes and intermittent priority activation: analytical investigations,” Transp. Res. Rec. J. Transp. Res. Board, no. 2315, pp. 182-190, 2012.

[76] H. Levinson, S. Zimmerman, J. Clinger, and J. Gast, “Bus Rapid Transit: Synthesis of Case Studies,” Transp. Res. Rec. J. Transp. Res. Board, vol. 1841, pp. 1-11, Jan. 2003.

[77] W. Ma and X. Yang, “A passive transit signal priority approach for bus rapid transit system,” in Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE, 2007, pp. 413-418.

[78] A. H. Alomari, H. Al-Deek, A. Sandt, J. H. Rogers, and O. Hussain, “Regional Evaluation of Bus Rapid Transit With and Without Transit Signal Priority,” Transp. Res. Rec. J. Transp. Res. Board, vol. 2554, pp. 46-59, Jan. 2016.

[79] P. Furth and J. SanClemente, “Near side, far side, uphill, downhill: impact of bus stop location on bus delay,” Transp. Res. Rec. J. Transp. Res. Board, no. 1971, pp. 66-73, 2006.

[80] D. S. Terry and G. J. Thomas, “Farside bus stops are better,” Traffic Eng. Inst Traffic Engr, 1971.

Section 3

Technological and Digital Advances

Chapter 10

Advanced Vehicles: Challenges for Transportation Systems Engineering

Orlando Giannattasio and Giulio E. Cantarella

Abstract

Automatic vehicles represent one of the most active research fields within engineering. Among transportation systems engineering research topics, we highlight the need to update and/or develop new mathematical models, computer science methods and electronic technologies that contribute to the development of more effective, accurate and robust tools. In order to develop more effective models, it is advisable to consider the opportunity to interact with other specialists from sectors different of the transportation systems engineering to provide solutions to problems that may arise during the modeling and further new points of view. The main goal of this paper is discussing the most likely positive and negative effects of mixed flow expected in the near future, analyzing the main classifying criteria such as ownership, on-board technologies (sensor), and reviewing the most effective tools already available for macroscopic analysis of multi vehicle type transportation systems.

Keywords: advanced vehicles, sensors, classification, transportation systems, macroscopic analysis, mixed flow

1. Introduction

Evolution has always been a fundamental component of life and technology developments have always represented a fundamental step in human civilization evolution. In fact, it is possible to say that technology speeds up everyday life thanks to the continuous introduction of innovative techniques, new devices and new perspectives. This results in turn in huge modifications in human activities perception, due to the time involved, their safety and in general their degree of difficulty.

Among the various cases of the daily life, one of the most discussed and interesting issue of the last years is related to automated or self-driving vehicles (AVs). The introduction of such vehicles resulted in the fact that nowadays car company cannot think of just projecting the mechanical parts or the basic electronic used in the vehicle, but are also involved in the project of sensors, integration issues and in the development of the related software.

In this way the integration of electronic components has become a necessary part in the vehicle development, in order to obtain an increase in safety and in easiness of the driving experience (e.g. Anti-Breaking Systems, Hydraulic Break Assist, Electronic Stability Control, etc.), as well as a reduction of impacts such as air pollution, noise, fuel consumption. This resulted in a major cooperation between companies coming from different backgrounds and in the creation of partnerships and joint ventures among them.

These effects result in a renewed impulse in the research related to the vehicles, due to the growing mingling among the various sectors of the industry. The common goal is to develop new models or update the old ones to simulate real life situations and predict the future developments of the automotive and, in general, of transportation research field. Considering the field of transportation systems engineering, it is possible to state that right now the main focus is to achieve a deeper understanding of the hardware and software involvement in this new kind of vehicles. This, in turn, allows to obtain better models developments and more reliable scenario assessment. These issues are often accentuated by the impossibility of retrieving commercial standards, due to the fact that, in some cases, they are not yet available, or the vehicles are still prototypes.

New technologies for automated vehicles look very appealing, but, it may easily be anticipated that the time needed to turn the existing stock of traditional vehicles (TVs) into the new ones will last several years during which mixed traffic is expected, requiring ad hoc tools for analysis and design [1].

Moreover, the likely effects on congestion, and more generally traffic management and control, are not necessarily positive. Some of them are enlisted below, together with indications of needs of new models:

- increase of max density, since AVs may be shorter on the average,
- increase of capacity, since safety distance may be shorter,
- decrease of speed dispersion, useful for more effective traffic control,
- decrease of the number of vehicles on streets with shared vehicles, since the number of trips per vehicle increases, and a decrease of the number of vehicles parked along the sidewalk might be anticipated, leading to an increase of capacity.
- increase of vehicle x km due to empty movements, if shared vehicles spread, due to longer paths and of change of effective origin and destination,
- discrete space / time access for AVs vs. continuous access for TVs,
- decrease of the number of users per car, if shared vehicles spread, leading to an increase of vehicle (car) demand flows,
- change of modal split leading to an increase of car flows with respect to transit (need of new model split models)
- increase of generated flows, due to increase of mobility index and reference population, likely including mobility-impaired people (need of new trip generation and distribution models)

On the other hand, some positive social impacts may also be anticipated such as

- less pollution and noise (greater effect with electric powered vehicles),
- increase of safety,

- increase of generated flows, due to increase of mobility index and reference population,

- easier integration with medium/long-distance transportation.

The main goals of this paper are:

- discussing the most likely positive and negative effects of mixed flow expected in the near future, as briefly outlined above,

- providing an analysis on sensors used in the automotive field,

- analyzing the main classifying criteria,

- reviewing the most effective tools already available for macroscopic analysis of multi vehicle type transportation systems.

A vehicle classification based on available sensors is surely helpful to support specification and calibration of models for transportation systems analysis, thus a sensor critical analysis is carried out in Section 2, supporting the discussion of classification issues in the same Section 2. Section 3 discusses the main tools for the analysis of transportation systems with mixed flow. Section 4 reports some conclusions.

2. Sensors analysis and vehicle classification

During the last years, electronic systems assumed a growing importance in the development of assisted and automated systems. In particular, the research has been focused around two main topics:

- the development of new algorithms and the improvement of the already existing ones, at the end of obtaining more compact, less consuming and, above all, faster implementations, in order to fit the needs of the automotive industry;

- the development of "ad hoc" sensors capable of operating in the various environments and having performances suitable to detect, intervene and advice the driver in time, making it possible for the human or autonomous driver to take the best decision due to the particular conditions.

This work starts from the sensors evolution to give an insight of the different techniques and sensors families used in nowadays autonomous vehicles, in order to gain insight on the different issues related to their development and provide ideas useful to better integrate them into the traffic models, obtaining more fitting models and better results.

First of all, it is possible to define four types of sensors and related systems used in nowadays smart vehicles:

- Image sensors and processors,

- Radar sensors and systems,

- Lidar sensors and systems,

- Ultrasonic sensors and systems.

Moreover, the data coming in from the different sensors have to be considered as a whole, in order to have a redundant and robust system and to discriminate in the case of contrasting detections from different sensors. This exchange of information between the different systems is done using an internal network, such as a CAN bus, to avoid the instantiation of several sensors of the same type. Ideally, the aim of the data from and to these control units is to describe the status of the car and formulate the actions to be taken at any time. It is interesting to note, that with the new smart vehicles also systems to process and take into account the data coming from other vehicles have to be considered, causing a huge increase in the amount of data to be processed and, thus, an increase in the research for fast computing, stream processing systems and diagnostic protocols [2].

A careful examination of the various techniques used has to be developed, together with evaluation of the differences between the different types of sensors and how and when to use them, in order to achieve both a better understanding of their behavior and develop new models.

2.1 Image sensors and processors

The use of image sensors in automotive field goes back to vision guided auto-parking systems which has been around for the past twenty years [3]. Nevertheless, some of the main techniques are still used nowadays; the main topic has shifted from parking problems to active detection during driving. Several studies have been conducted on these applications, e.g. dedicated to pedestrian crossing and signal recognition. In particular in this type of problems the image sensor has to acquire the images from the environment at resolutions and acquisition rates suitable for the detection problem. Together with the sensor also the image processing methods have to be considered. In particular, to enhance the performances of the sensors in terms of processed frame rate per second (fps) hardware accelerators closely coupled to the image sensor are used. The accelerators are mainly dedicated to filtering operations on the incoming images from the sensors, aiming to send to the processing units only the significant data to be taken into account in the detection. Several systems have been developed, like the ones dedicated to edge detection and recognition, segmentation and object recognition. However, the trend is shifting to the use of Artificial Neural Networks capable of calibrating the various detection parameters according to the incoming stimulus and their previous history. Neural Networks hardware accelerators have been developed in the last years to make the systems capable of operating in real-time. While the training of the Neural Networks to determine the optimal parameters is usually done offline, some parameters could also be varied on-the-fly to achieve a resilience of the system to false detections or to improve the driving performances.

The challenge is nowadays to develop faster Neural Networks systems operating together with the image sensors, in order to process higher amount of data while operating with higher resolution frames; on the other hand, the development of new dedicated algorithms could also allow a complexity reduction of the problem which, in turn, could cause lower latencies in the system and better detection performances, both from a correctness and a speed standpoint.

2.2 Radar sensors

Radar systems are used in automotive to help solve problems like auto-parking, pre-crash sensing and adaptive cruise control, since one of the main characteristics of a radar system is the capability of inferring the velocity of the detected objects in its range. Radar systems used in automotive applications usually operate at 77 GHz frequency. The measurement time could be reduced to 10 ms allowing controlling the scene 100 times per second [4].

Radar systems are used together with other systems and results particularly efficient for detections at distances up to approximately 200 m, with a range resolution of 1 m [4]. These peculiarities make the radar system particularly appealing for mid-range applications in condition of non-optimal visibility, which could cause major problems to systems based on image sensors. However, radar systems have got poor object detection issues in medium ranges (30 m to 60 m) for angles larger than $\pm\pi/6$; for that reason, they need to be supported by other detection systems, based on different technologies [5]. Moreover, it has to be highlighted that a typical radar system shows difficulties in taking into account targets having different azimuth values and, thus, more sophisticated data processing systems are needed, like tracking ones. These systems could show problems in the case of wrong modeling assumptions for the targets, causing detection problems, like target loss.

2.3 Lidar sensors

Lidar systems are capable of good performances in presence of adverse environmental conditions, such as fog, heavy rain or snow, but also in the case of scenes having low illumination levels.

However, also lidar systems show some drawbacks, due to the fact that in some cases the dynamic range of the lidar system could be exceeded. When it happens two extremely severe types of error occur: the loss of targets and the generation of ghost targets, which worsen the detection performances unacceptably in some cases [5].

2.4 Ultrasonic sensors

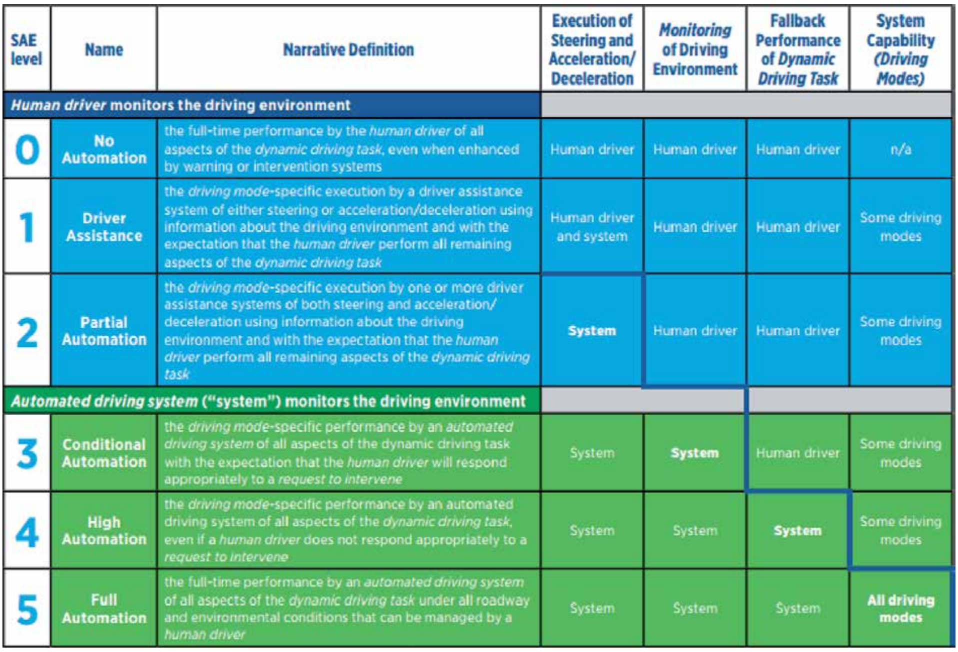
Ultrasonic systems are mainly used for parking assist systems, where they are capable of insuring low costs and good performances. Several threshold systems are designed to achieve the correct detection of corners and edges, while other systems take care of the steering angle and wheel speed; all is coordinated by a network, like the CAN bus previously mentioned [6]. Moreover, this kind of systems demonstrated to work well in cooperation with laser parking systems or to substitute them [7].

Finally, it has to be highlighted that great importance has to be put on decision algorithms, which could be usually implemented in software and which could represent the bottleneck in some systems. In what follows we concentrate on the effect related to the sensors, assuming that the performances of the software are the same for all the considered systems in order to do not undermine the comparison results.

2.5 Classification issues

The schemes that is internationally used as standard for the classification of “Automated Vehicles” is the “SAE Levels” [8], written by “SAE INTERNATIONAL” (U.S.-based, globally active professional association and standards developing organization for engineering professionals in various industries) (**Figure 1**).

As shown in **Figure 1**, a six levels taxonomy proposed; these six levels are collected in a meta-classification based on two macro categories (Human Driver e Automated Driving System) that identify the technology inside the vehicle and the type of driver.



SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	**No Automation**	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	**Driver Assistance**	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	**Partial Automation**	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* ("system") monitors the driving environment**						
3	**Conditional Automation**	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	**High Automation**	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	**Full Automation**	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

Figure 1.
SAE levels.

To understand the differences between each SAE level, guidelines written by AdaptIVe (a project co-founded by European Commission) [9] have been taken into account. In the following a brief explanation for each SAE level is reported:

- Level 0: no electronic device and human driver,
- Level 1: some electronic devices but the driver is still human,
- Level 2: more electronic devices for security, the driver is helped in some maneuvers, but it still has the main control of the vehicle,
- Level 3: in a smart road, the vehicle is supervised by human driver that takes control of the vehicle in case of emergency,
- Level 4: in a smart road, the vehicle is still supervised by the human driver, but the driver can make other actions,
- Level 5: completed automated vehicle.

In current research and practice SAE levels are grouped into 2 classes: 012 and 345 as it can be seen within the **Figure 1**.

Still, several considerations about the relationships between kinds of sensors available on a vehicle and SAE Levels support a 3-class meta-classification useful to effectively support specification and calibration of models for transportation systems analysis.

Table 1 shows the relationship between SAE Levels and types of sensors, supporting the 2-class meta-classification in **Figure 1**: Human driving vs. Automated driving. It could be noticed indeed that there is a huge difference between the type

	Human driving			**Automated driving**		
	SAE 0	**SAE 1**	**SAE 2**	**SAE 3**	**SAE 4**	**SAE 5**
Image sensors and processors	N	N	Y	Y	Y	Y
Radar sensors and systems	N	Y	Y	Y	Y	Y
Lidar sensors and systems	N	N	N	Y	Y	Y
Ultrasonic sensors and systems	N	Y	Y	Y	Y	Y

Table 1.
Relationship between SAE levels and types of sensors.

of sensors installed in vehicles of each level. Note that the ultrasonic sensors are used only for parking maneuvers and not for drive assistance, thus they will not be further considered within the following.

From these considerations, it is possible to see why many authors choose to use two macro classes, to group SAE Levels [10]:

- Human Driver: as shown in **Figure 1**, the vehicle certified within SAE Level 0 to Level 2 is in this macro class. Even if they show some technological differences between each level, in this type of vehicles the driving function is still human related. Hence, the currently available models for transport systems analysis [11] can almost straightforwardly be applied to these vehicles;

- Automated Driving System: the other levels are defined this class; the types of vehicles in this class are still at a prototypal stage, but it can easily be anticipated that new models need to be studied and developed.

This meta-classification is probably over-simplified: at a closer look, vehicles in the “Automated Driving System” class cannot be conceived as homogeneous, as shown below.

First, looking at the relationship between Automated Driving SAE Levels (3 to 5) and ranges of sensors, as shown in **Table 2**, it can easily find out that application ranges of sensors for SAE levels 3 and SAE Level 4 and quite different from those for SAE level 5.

In addition to sensors application range, by studying the SAE Levels definitions [10], it’s possible to see that SAE Level 3 and SAE Level 4 vehicles still need a human driver compare to SAE Level 5. The human driver function within this level represents only a safeguard measure in emergency conditions or both SAE Levels: both presents different reaction times for the takeover, but the description used for the possible action that can be made by the driver during the travel are too permissive for the human driver and their reaction time. On the other hand, SAE Level 5 vehicles do not need a human driver at all, not even a steering wheel (as show in some prototype vehicle that have the possibility to hidden it), leaving all the control function to the vehicle. Starting from those considerations about the control of

	SAE 3	**SAE 4**	**SAE 5**
Radar sensors and systems	200 [m]	200 [m]	200 [m]
Image sensors and processors	250 [m]	250 [m]	200 [m]
Lidar sensors and systems	150 [m]	150 [m]	100 [m]

Table 2.
Relationship between automated driving SAE levels (3 to 5) and ranges of sensors.

the vehicle, it's possible to see that "Automated Driving System" class present a non-homogeneous control, but these criteria it's not the only one.

Another criterion is represented by the "type of ownership". Different from SAE Levels 0 to SAE Levels 2 that present a huge percentage of private owned vehicle; it's possible to notice that:

- SAE Level 3 and SAE Level 4 can be considered a technological evolution of traditional vehicles aiming at reducing effort of human drivers and will likely be mostly privately owned,
- SAE Level 5 vehicles should be considered an evolution of taxi, other vehicles available on demand and public transport system and will likely be mostly not privately owned such as the so-called robotaxi).

Therefore different "user definitions" [11] must be considered in the two cases. Moreover, the new models needed to be studied and developed should be differentiated, at least with respect to parameters.

For all these reasons, a new approach to vehicle meta-classification should be formulated. Assuming that:

- the sensors related software and involved algorithms are the same for each vehicle and level,
- the sensors have their functional range as described in **Table 2**,
- the sensors functional range as line of vision for the user,
- the SAE Level Certification remains as fundamental definition for the type of vehicles,
- the SAE Level can be used for an "ownership" definition by the user (Starting for SAE Level 0 to be private to arrive at SAE Level 5 as Shared Vehicle)

The new meta-classification proposed is the following including three classes:

- Human Driver: the same definition as the previous 2-class meta-classification,
- Advanced Driving System: this class includes the SAE Level 3 and SAE Level 4 vehicles; these vehicles can be modeled as the same type of vehicle for the "user definition", since a human driver is needed to control the vehicle and the considered sensors have the same application range,
- Automated Driver: this class includes the SAE Level 5 vehicles only.

This meta-classification allows the analysis of the following future scenarios:

- short-term scenarios with both Human Driver and Advanced Driving System vehicle classes only,
- medium-term scenarios as above with low percentage of Automated Driver class vehicles,
- long-term scenarios with Advanced Driving Systems and Automated Driver classes only and no Human Driver class.

3. Analysis of transportation systems with mixed flow

A change so great may be not technology-driven only, but also requires a carefully analysis of its several impact through well designed enhancements of tools of Traffic and Transportation Theory (TTT) already available to the transportation systems modelers and planners (see the comprehensive book by Cascetta [11]).

The analysis of transportation systems with several types of vehicles require a generalization of existing models and algorithms for travel demand assignment to transportation networks, as described in Cantarella and Di Febbraro [12], Cantarella et al. [13, 14]; the proposed approach can be applied to real size networks. It is briefly reviewed in the following.

Users are partitioned into o-d pairs they are traveling from/to, user categories (with common socio-economic and behavioral features) and types of used vehicle, such as traditional, connected, automated, autonomous, ...; fossil fuel vs. electrical powered; privately owned vs. shared; .... Demand flows are assumed constant and known.

Transportation supply is modeled through a flow network, say a graph with a transportation cost and a flow associated to each arc. All costs are assumed measured by a common unit, usually travel time or money, through duly homogenization of different attributes, if the case. A route connecting an Origin Destination pairs is described by a path. (Presented results still hold if more general definitions of routes are used, such as hyperpaths).

3.1 Assignment to uncongested networks

In uncongested networks the arc flows depend on the arc costs, through the arc-flow function obtained as described below; its structure is shown in **Figure 2**.

The arc costs (**c**) may be different among the vehicle types to reflect different performances, and we assume that the arc cost per vehicle type are given by an affine transformation of the arc generic costs.

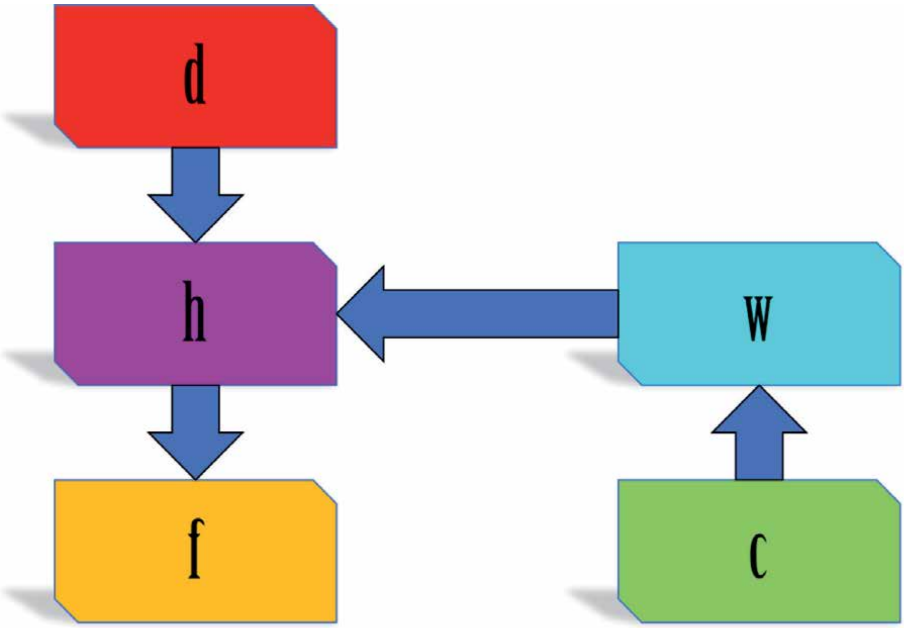


Figure 2.
The arc-flow function for assignment to uncongested networks.

Then, the route costs (**w**) for each o-d pair, user category and vehicle type can be obtained from the corresponding arc total costs through an affine transformation from the arc space to the route space defined by the transpose of arc-route incidence matrix.

The utility function for each o-d pair, user category and vehicle type is almost always specified through an affine transformation of costs both in research analysis and in practical applications.

Route choice behavior for users of each o-d pair, user category and vehicle type m can be modeled by applying any discrete choice modeling theory, such the well-established Random Utility Theory. In this case the choice proportion of an alternative is given by the probability that its perceived utility is equal to maximum among all alternatives. When the perceived utility co-variance matrix is non-singular, probabilistic route choice functions are obtained.

Demand conservation flow relation for each o-d pair, user category, vehicle type assures that flows of all connecting routes (**h**) sum up to demand flow (**d**).

The arc flows (**f**) due to each o-d pair, user category and vehicle type can be obtained from the route flows through a linear transformation from the route space to the arc space defined by the arc-route incidence matrix. Having assumed that all arc flows are measured in TVs per time unit, the arc total flows are given by the sum over all o-d pairs, user categories and vehicle types.

Main input data of the arc flow function are arc costs, and demand flows. Vehicle types may be distinguished with respect to:

- flow equivalence
- mean number of users on board
- cost equivalence
- specific arc, cost, e.g. monetary cost (and VoT), access cost, …
- route utility parameter and route choice function parameters
- route choice function

The arc flow function is monotone non-increasing with respect to arc costs under mild assumptions. It can be computed for large scale applications through algorithms derived from network theory, avoiding explicitly path enumeration.

3.2 Equilibrium assignment to congested networks

In congested transportation networks arc flows depend on arc costs, and user equilibrium assignment searches for mutually consistent arc flows and costs. Arc generic costs depend on the arc total flows through the arc cost function, which models user driving behavior at macroscopic level.

Equilibrium assignment can effectively be described through fixed-point (FP) models obtained combining the arc-flow function and the arc cost function. These models can be solved for large scale applications through algorithms based on the Method of Successive Averages, which avoid the use of matrix algebra and computation of derivatives. Their structure is shown in **Figure 3**.

Existence is guaranteed if both the arc flow function and the arc cost function are continuous (and the network is connected), applying Brouwer theorem. For a monotone decreasing arc flow function, if the arc cost function is monotone strictly increasing uniqueness is guaranteed. Uniqueness of arc flows also guarantees

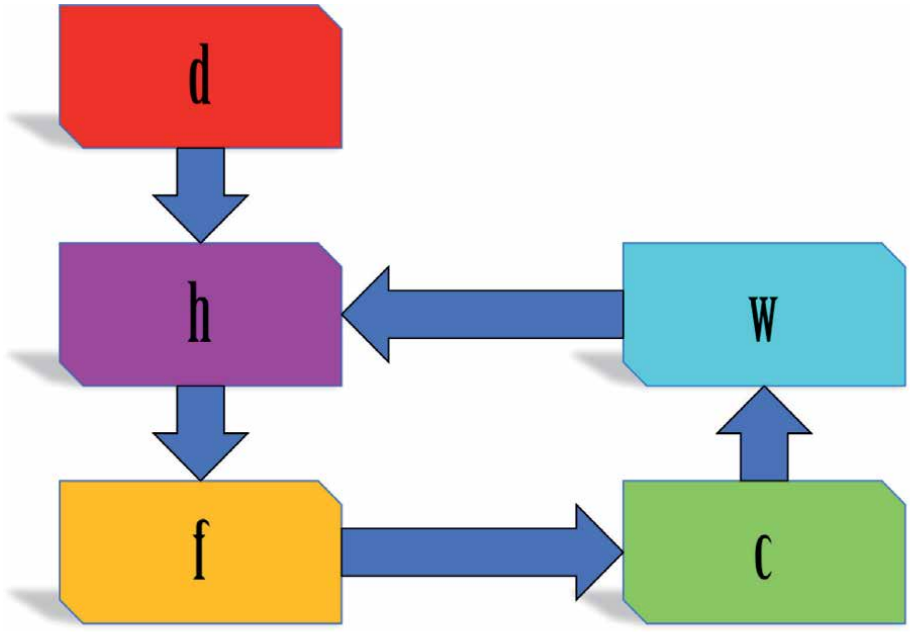


Figure 3.
Fixed point models for equilibrium assignment to congested networks.

uniqueness of arc costs as well as route flows and costs, and of flows and cost per o-d pair, user category, vehicle type.

3.3 Day-to-day dynamic assignment to congested networks

The evolution over time of arc flows and costs can effectively be described through day-to-day dynamic assignment models. The specification of these models requires an extension of models for the equilibrium assignment by including sub-models of

- User memory and learning: how users forecast the level of service that they will experience today, from experience and other sources of information, such as informative systems, about previous days;
- User habit and inertia to change: how users make a choice today, possibly repeating yesterday choice to avoid the effort needed to take a decision, or reconsidering it according to the forecasted level of service.

The arc cost updating relation, modeling user memory and learning, gives the today forecasted route costs with respect to previous day costs. It extends the arc cost function. In the simplest instance, this relation can be specified by an exponential smoothing (ES) filter, say a convex combination of yesterday route forecasted costs and yesterday actual route costs, given by an affine transformation of the yesterday arc costs.

The arc flow updating relation, modeling user habit and inertia to change, gives the today arc flow with respect to forecasted costs and previous day flows. It extends the above arc flow function. In the simplest instance, this relation too can be specified by an exponential smoothing (ES) filter, say a convex combination of yesterday arc flows due to users who do not reconsider their yesterday choice and today arc flows due to users who reconsider their yesterday choice.

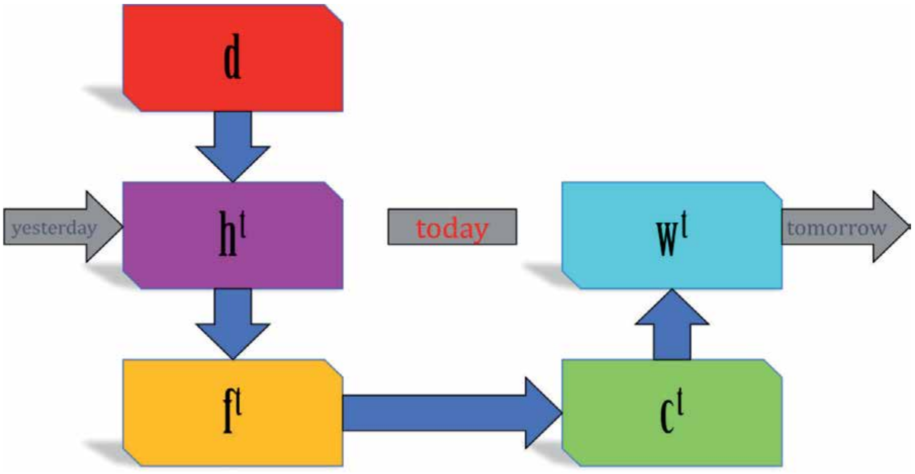


Figure 4.
Dynamic process models for day to day dynamic assignment to congested networks.

The arc flow updating relation can be combined with the arc cost updating relation to specify Deterministic Process (DP) models for day-to-day dynamic assignment. Their structure is shown in **Figure 4**.

The fixed-point states of the DP model specified by the ES filters are equivalent to the equilibrium states as defined by FP model mentioned above). Applying techniques from the theory of discrete-time non-linear dynamic systems, the above DP models can be used to study the local stability of each fixed-point state, say whether it is an attractor. Moreover, a bifurcation analysis can be carried to single out which attractor is reached by the evolution over time when an input data and or a parameter is changed.

The DP model specification can be used as a base to specify time discrete Stochastic Process models, which may provide full statistical characterization.

3.4 Within-day dynamic assignment to congested network

All the above models can be extended to cope with Within-day Dynamics requiring highly non-linear specification. This kind of models are useful to describe while-trip re-routing due to interaction with information and/or control system as well as queuing phenomena.

4. Conclusions

At first this paper has reviewed the analysis of the most likely effects of the introduction of new type of vehicles including vehicles with different level of automation, differently powered, privately owned or shared, ... Then, this paper has discussed the main technological and modeling issues for the analysis of transportation systems with mixed flow.

Some issues are worth of further research work, such as:

- specification of new models of Traffic Flow Theory to deal with congestion in mixed traffic,
- new path choice models for travel demand assignment within the general framework of Transportation System Analysis,

- parameter calibration from real data, or laboratory studies (survey, driving simulator, etc.).

Acknowledgements

The authors want to fulfill their gratitude Dr. Carmine Cappetta for the support and the patience in developing and publishing this chapter in regard to the analysis of the sensors shown within Section 2.

Authors wish to thank University of Salerno for financial support under local grants ORSA187110 and ORSA192198.

Author details

Orlando Giannattasio[1]* and Giulio E. Cantarella[2]

1 Studio Tecnico S.In.Tec, Salerno, SA, Italy

2 University of Salerno, Salerno, SA, Italy

*Address all correspondence to: o.giannattasio@studiosintec.it

References

[1] Cantarella GE, Di Febbraro A, Transportation systems with autonomous vehicles: Modeling issues and research perspectives. In the 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017, Proceedings 8005580, pp. 1-6.

[2] Schweppe H, Zimmermann A, Grill D, Flexible On-Board Stream Processing for Automotive Sensor Data, IEEE Transactions on Industrial Informatics, 2010, 6(1): 81-92.

[3] Xu J, Chen G,Xie M, Vision-Guided Automatic Parking for Smart Car, IEEE Intelligent Vehicles Symposium 2000Dearborn (MI), USA October, 2000: 725-730.

[4] Rohling H, Meinecke MM, Waveform Design Principles for Automotive Radar Systems, Radar, 2001 CIE International Conference on, Proceedings, 2001, 4: 1-4.

[5] Rasshofer RH, Gresser K, Automotive Radar and Lidar Systems for Next Generation Driver Assistance Functions, Advances in Radio Science, 2005, 3: 205-209.

[6] Park WJ, Kim BS, Seo DE, Kim DS, Lee KH, Parking Space Detection Using Ultrasonic Sensor in Parking Assistance System, 2008 IEEE Intelligent Vehicles Symposium Eindhoven University of Technology, 2008: 1039-1044.

[7] Jeong SH, Choi CG, Oh, JN, Yoon PJ, Kim BS, Kim M, Lee KH, Low Cost Design of Parallel Parking Assist System Based on an Ultrasonic Sensor, International Journal of Automotive Technology, 2010, 11(3):409-416.

[8] SAEdocument J3016, (2014), Taxonomy and Definitions for Terms Related to On-Road Automated Motor Vehicles.

[9] Bartels A, Eberle U, Knapp A, System Classification and Glossary. H2020 AdaptIVe Project Deliverable D2.1, 2015.

[10] Di Febbraro A, Sacco N, Open Problems in Transportation Engineering with Connected and Autonomous Vehicles, Transportation Research Procedia, 2016, 14: 2255-2264.

[11] Cascetta E, Transportation System Analysis: Models and Applications, New York: Springer, 2009.

[12] Cantarella GE, Di Febbraro A, Transportation Systems with Autonomous Vehicles: models and algorithms for equilibrium assignment. In Transportation Research Procedia 27 (2017) 349-356.

[13] Cantarella G.E., Di Febbraro A., Di Gangi M., Giannattasio O. (2019a). Solving Stochastic Assignment to Transportation Networks with TVs and AVs. To appear in Transportation Procedia 42, 7-18 (an early version in ETC 2018 on-line proceedings).

[14] Cantarella G.E., Di Febbraro A., Di Gangi M., Giannattasio O. (2019b). Stochastic Multi-Vehicle Assignment to Urban Transportation Networks. In IEEE Xplore (an early version presented at MT-ITS 2019).

Chapter 11

Driver Assistance Technologies

Pradip Kumar Sarkar

Abstract

Topic: Driver Assistance Technology is emerging as new driving technology popularly known as ADAS. It is supported with Adaptive Cruise Control, Automatic Emergency Brake, blind spot monitoring, lane change assistance, and forward collision warnings etc. It is an important platform to integrate these multiple applications by using data from multifunction sensors, cameras, radars, lidars etc. and send command to plural actuators, engine, brake, steering etc. ADAS technology can detect some objects, do basic classification, alert the driver of hazardous road conditions, and in some cases, slow or stop the vehicle. The architecture of the electronic control units (ECUs) is responsible for executing advanced driver assistance systems (ADAS) in vehicle which is changing as per its response during the process of driving. Automotive system architecture integrates multiple applications into ADAS ECUs that serve multiple sensors for their functions. Hardware architecture of ADAS and autonomous driving, includes automotive Ethernet, TSN, Ethernet switch and gateway, and domain controller while Software architecture of ADAS and autonomous driving, including AUTOSAR Classic and Adaptive, ROS 2.0 and QNX. This chapter explains the functioning of Assistance Driving Technology with the help of its architecture and various types of sensors.

Keywords: sensors, ADAS architecture, levels, technologies

1. Introduction

In order to enhance road safety as well as to satisfy increasingly stringent government regulations in western countries, automobile makers are confronted with incorporating a range of diverse technologies for driver assistance to their new model. These technologies help drivers to avoid accidents, both at high speeds and for backward movement for parking. This system can be placed into the category of advanced driver-assistance systems (ADAS). Besides increasing safety, ADAS [1] applications are concerned with to enhancing comfort, convenience, and energy efficiency. It is emerging as new driving technology supported with Adaptive Cruise Control, Automatic Emergency Brake, blind spot monitoring, lane change assistance, and forward collision warnings etc. It is an important platform to integrate these multiple applications by using data from radar, lidar, and ultra sound sensors etc. The vehicle engine related to hardware such as actuators, engine, brake, steering get the commands from the above sensors to enable the ADAS to take desired actions with respect to alerting the driver for detection of hazardous object or location or stopping the vehicle if necessary. For example, the recognition of black spot warning, lane change assistance and forward collision warning are extremely becoming useful in the ADAS.

During the gradual emergence of Connected and Automated vehicle (CAV), driver behavior modeling (DBM) coupled with simulation system modeling appears to be an instrumental in predicting driving maneuvers, driver intent, vehicle and driver state, and environmental factors, to improve transportation safety and the driving experience as a whole. These models can play an effective role by incorporating its desired safety-proof output into Advanced Driver Assistance System (ADAS. To cite an example, it could be said with confidence that the information generated from all types of sensors in an ADAS driven vehicle with accurate lane changing prediction models could prevent road accidents by alerting the driver ahead of time of potential danger. It is increasingly felt that DBM developed by incorporating personal driving incentives and preferences, with contextual factors such as weather and lighting, is still required to be refined, calibrated and validated to make it robust so that it turns into more better personalized and generic models. In regard to the modeling of personalized navigation and travel systems, earlier studies in this area have mainly considered ideal knowledge and information of the road network and environment, which does not seem to be very realistic. More researches are required to be conducted to address this real life challenges to make ADAS more acceptable to society.

There are an increasing evidences from the various literatures that a single vehicle making inferences based on sensed measurement of the driver, the vehicle, and its environment is mostly focused for DBM where there is any hardly attempt made to develop DBM in the traffic environment in the presence of vehicle to vehicle (V2V), and vehicle to infrastructure (V2I) scenario- communications system. It would be interesting to develop DBM with respect to connected and automated vehicle (CAV) to leverage information from multiple vehicles so that more global behavioral models can be developed.. This would be useful to apply the output of the CAV modeling in the design of ADAS driven vehicle to create a safety proof driving-scenario for diverse applications.

2. Architecture of ADAS

There are a number of sensors which are increasingly being used. These are namely cameras, medium and long-range radar, ultrasonic, and LIDAR.. Data generated from these sensors go through fusion process to authenticate the data so as to enable the computer software perform the necessary tasks to activate the driver assistance system to take correct decisions. These decisions are related to parking assistance, automatic emergency breaking, pedestrian detection, surrounding view, and even drowsiness of the driver. The functional components such as various types of sensors collecting data from immediate surrounding environment are related to ADAS architecture that helps to perform necessary tasks as shown in the **Figure 1**. The forward collision-avoidance ECU module is located in the windshield, supported with the blind spot ultrasonic sensors and related ADAS processor may be located in the side mirrors or other location areas.

The **architecture** [2–4] of the electronic control units (ECUs) is responsible for executing advanced driver assistance systems (**ADAS**) in vehicles which is changing for its response during the process of driving. Automotive system architect integrates multiple applications into **ADAS** ECUs that serve multiple functions of ITS architecture as shown in the **Figure 2**. **Figures 3** and **4** show Architecture for other functions related to Forward Collision and Parking Assistance respectively.

Hardware architecture of ADAS and autonomous driving, includes automotive Ethernet, TSN, Ethernet switch and gateway, and domain controller while Software architecture of ADAS and autonomous driving, including AUTOSAR Classic and Adaptive, ROS 2.0 and QNX.

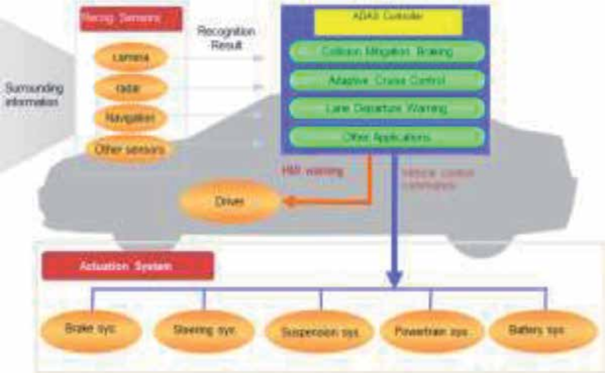


Figure 1.
Functional components and various types of sensors. Source: http://www.hitachi-automotive.us/Products/oem/DCS/ADAS/index.htm

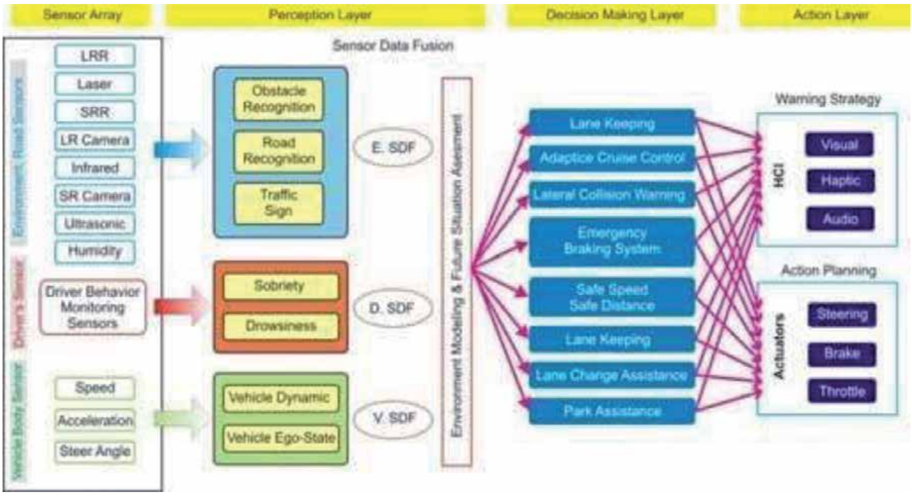


Figure 2.
Architecture of ADAS, source: Ref [3].

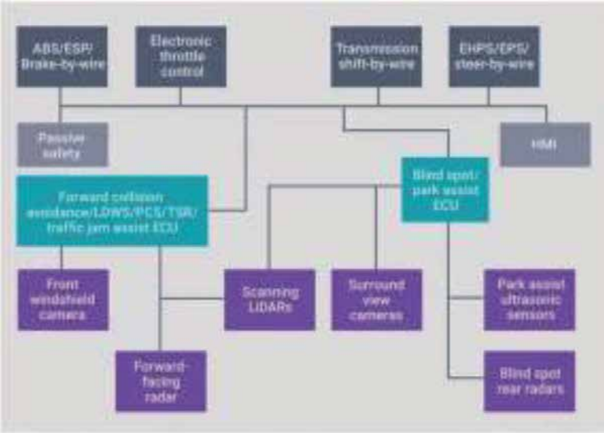
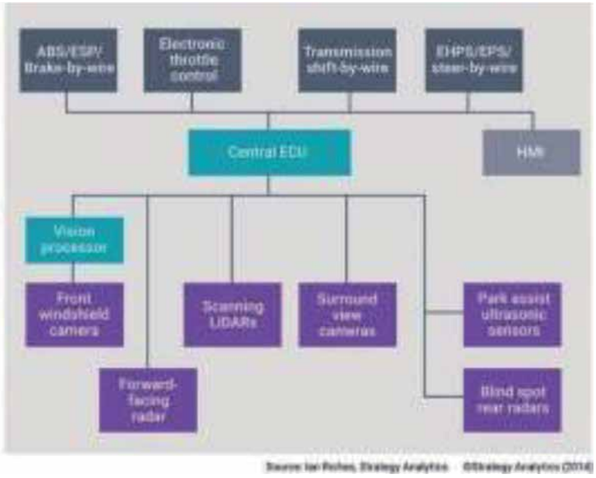


Figure 3.
Architecture of forward collision avoidance & blind spot avoidance. Source: Ian Riches, strategy analytics.



Source: Ian Riches, Strategy Analytics ©Strategy Analytics (2014)

Figure 4.
Architecture of ADAS -Parking Avoidance & Blind. Source: http://www.techdesignforums.com/practice/technique/managing-the-evolving-architecture-of-integrated-adas-controllers/.

3. Functioning of ADAS

Advanced driver assistance systems (ADAS) need a number of integrated sensors to accurately determine situational assessment and action implementation. In ADAS technologies [5–7] sensors such as video, radar, LIDAR, ultrasonic and infrared (IR) sensors are being increasingly utilized. Sensor fusion with advanced algorithms and computing power, connectivity and data transmission, contextual awareness and processing, and virtual sensors is extremely important for success of ADAS.

There are six levels of vehicle automation as shown in **Figure 5** defined by the Society of Automotive Engineers (SAE) [8] with a span from Level 0, which has no automation, to Level 5, which involve fully autonomous vehicles. As automation expands, driver assistance and ADAS plays an increasingly important role.

Level 0: Driver only: the driving is controlled by the human driver using with driving aids independently including steering, throttle, brakes, etc.

Level 1: Assisted driving: driver needs assistance during vehicle operation with respect to Cruise Control, ACC.

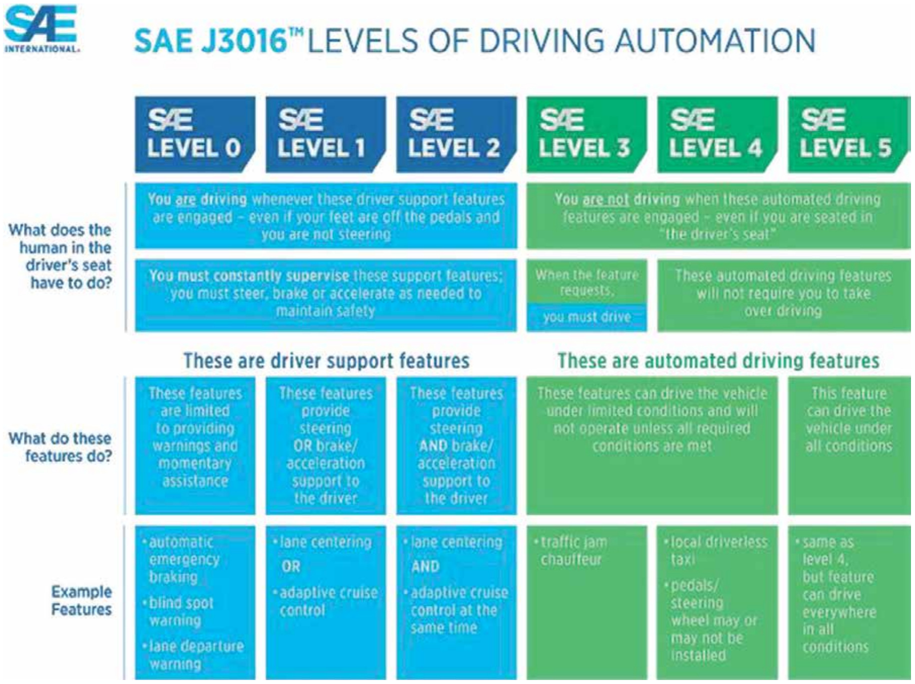
Level 2: Partial automation: the system is monitored during driving. At least one system, such as cruise control and lane centering, is fully automated.

Level 3: Conditional automation: the system is monitored by the operator and can intervene when it is necessary. Safety-critical functions, under certain circumstances, are shifted to the vehicle.

Level 4: High automation: there is no monitoring required by the driver. Vehicles are designed to operate safety-critical functions and monitor road conditions for an entire trip. However, the functions do not cover all every driving scenario and are limited to the operational design of the vehicle.

Level 5: Full automation: it ensures operator-free driving without any intervention.

As of today, no car manufacturer has achieved level 3 or higher in production, although several have produced demonstration vehicles. The legislature of some countries is working on a possible admission of "Level 3" vehicles, which is expected to be available in 2020/21. Driver assistance systems enabling autonomous driving from level 3 onwards will require at least three types of sensor systems: camera, radar, and LIDAR systems. As can be seen in **Figure 5**, several of each type of sensor operates at



SAE J3016™ LEVELS OF DRIVING AUTOMATION

	SAE LEVEL 0	SAE LEVEL 1	SAE LEVEL 2	SAE LEVEL 3	SAE LEVEL 4	SAE LEVEL 5
What does the human in the driver's seat have to do?	You **are** driving whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			You **are not** driving when these automated driving features are engaged – even if you are seated in "the driver's seat"		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
	These are driver support features			These are automated driving features		
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering OR brake/acceleration support to the driver	These features provide steering AND brake/acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met		This feature can drive the vehicle under all conditions
Example Features	•automatic emergency braking •blind spot warning •lane departure warning	•lane centering OR •adaptive cruise control	•lane centering AND •adaptive cruise control at the same time	•traffic jam chauffeur	•local driverless taxi •pedals/steering wheel may or may not be installed	•same as level 4, but feature can drive everywhere in all conditions

Figure 5.
Various levels of ADAS, source: https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving- automation%E2%80%9D-standard-for-self-driving-vehicles.

various locations on the vehicle. The development of the LIDAR system is still posing the bigger and most dynamic challenge in technical and commercial terms.

4. Data fusion

4.1 Fusion of DATA at ECU

There are a number of sub systems associated in performing various tasks of ADAS. A vehicle's movement detected by the ADAS can be seen in the main system inside the vehicle when the driver is present. This system interacts with the environment. There are different functions of the system as can be clearly distinguished in **Figure 6**. The following distinctive features of fusion are mentioned as under:

1. Information has to be gathered;

2. Information needs to be evaluated;

3. A safety measure need to be taken;

These functions are synonymous to as Sense (1), Think(2), and Act(3). Only the Sense sensors are reviewed and only the systems in which the driver is inside the loop. **Figure 6** shows the process of 'multi sensor processing', starting with the sensor data acquisition. Next, the sensors processing, divided into several tasks, as 'Calibration, 'Feature Extraction', 'Object Detection', etc., begins to analyze the sensors data and, in the end, serves the application with a more or less detailed model of the environment [4].

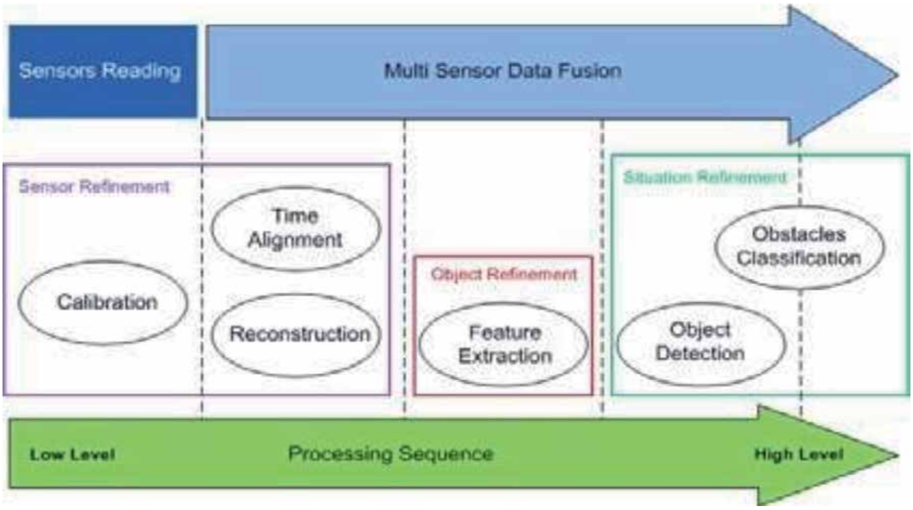


Figure 6.
Fusion of data at ECU received from various types of sensors housed in ADAS, Source: Ref No: [3].

Fusion of data received from complementary and independent sources place the data into a single description. Data association and data assimilation are two important components to be addressed for data fusion as a part of the process that matches sensor data with the description of the environment that requires synchronization of the sensor data and the associated object state (e.g., position and velocity).

5. Various sensors of ADAS

It is extremely important to know which sensors are required for autonomous driving from Levels 1 to 5. As already mentioned, there are three main groups of sensor systems camera-, radar-, and LIDAR-based systems. Although, for parking, ultrasonic sensors are available today and are widespread, they are of minor importance for autonomous driving. Camera and radar systems are in the Level 1 and 2 vehicles today and are prerequisite for all further levels of automation.

5.1 Sensor camera for ADAS

This advanced Camera (digital HDR CMOS cameras) with large dynamic range is well suited to poor light conditions and primary differences are due to its brightness.

A large number of digital interfaces are available with camera for automobiles along with digital signal processor and internal memory capacity. The camera generates processed video images for evaluation using software algorithm. It also help images transformed in to signals to merge with other sensor signals such as other as radar and lidar etc. Due to the inherent intelligence of the camera, all the signals are processed in the fusion mode to enable the ADAS to take correct decision. The camera used as sensor [9] is required to go through the quality management (ISO/TS/16949 in the automobile industry and are suited for adaptability which is quick and flexible. Current digital camera system is continuously receiving raw data that is then processed and forwarded to the display unit for image display. This procedure is shown in **Figure 7**.

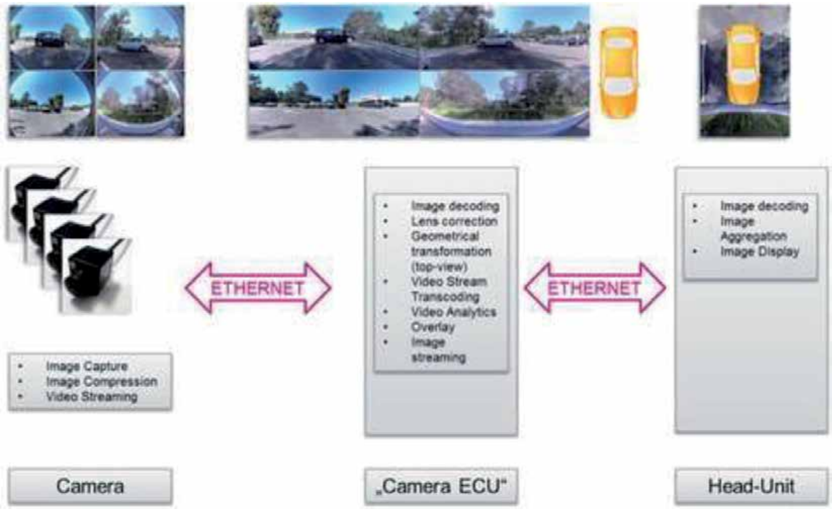


Figure 7.
Video data transfer to head unit of camera through Ethernet, source: https://www.fierceelectronics.com/components/three-sensor-types-drive-autonomous-vehicles

Besides this, the **i**nfra**r**ed (IR) camera consists of several components. It is important to distinguish 2 different versions of the IR camera:

The **i**nfra**r**ed (IR) camera consists of several components. It is important to distinguish 2 different versions of the IR camera:

1. **N**ear **I**nfra **R**ed(NIR);
2. Far **I**nfra-**R**ed(FIR);

In both systems a camera plays an important role in identifying radiation of objects. It may be mentioned that NIR technology offer an extra illumination by IR-headlights while the FIR systems is not characterized with special headlights. The primary difference between the two is picking up the extra-radiated objects by the NIR systems while FIR only accepts only the regular radiation of objects.

Table 1 presents transmission of data rate from sensors [10]. **Figure 8** shows the functioning of Lidar.

5.2 LIDAR systems

For purpose of measuring distance and creation of three-dimensional images of the environment, LIDAR system [11] is fitted and integrated ever more frequently into vehicles and mobile machines. A pulsed laser beam assesses the signal's transit time from the object back to the detector as shown in **Figure 8**. A highly sensitive

Sensor	Data rate required to transmit raw data
Camera	1Gb/sec to 24Gb/sec
Radar	5Gb/sec to 120Gb/sec
Lidar	2 Mb/sec to 10Gb/sec

Table 1.
Data rate required for transmission of data.

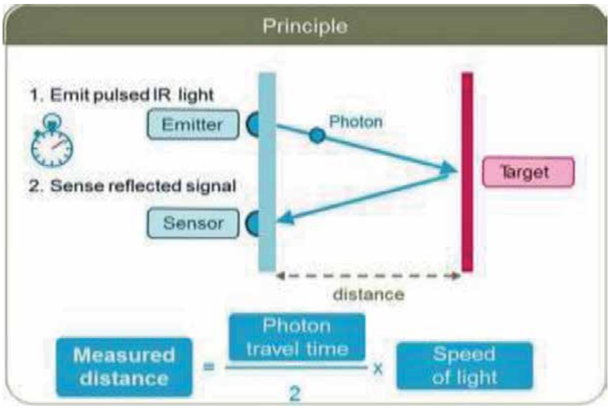


Figure 8.
Principle of the functioning of LIDAR. Source: https://www.fierceelectronics.com/components/three-sensor-types-drive-autonomous-vehicles.

technique using Avalanche Photodiodes along with internal amplification measure the light pulses in the nanosecond range across wider bandwidths. Lidar optical system requires the high spatial resolutions. Therefore sensor has the capability to develop APD arrays comprising with multiple sensor elements. The APD arrays from sensor addresses the effect of temperature due to its high voltage. Their highly accurate amplification offers excellent APD signal quality. The modules can be adapted to as per the specific application. Development boards with digital output signal and Low Voltage Differentiating Signal (LVDS) is interfaced. With the help of Lidar and Radar System, the object of the road can easily be identified. But in addition to these, there is a necessity for a camera for classification and detection of an object in a correct way. With the development of point density cloud from the reflections from radar and lidar, the distance and closing speed of the object can easily be measured. It may be mentioned that due to lower resolutions from these sensors as compared to camera, the detection of the objects are not easily made. To optimize the detection at varying ranges with lower resolution, a number of units are installed from a medium-range unit for emergency brake assist to long-range radar for adaptive cruise control although LIDAR & radar, functions in a similar way at longer ranges with lower point-density.

5.3 Radar

RADAR is meant to define its full form “Radio Detection And Ranging.”. By this sensor, the object is detected with the identification of localization of objects using radio waves with a frequency range from 24 to 77 GHz. It is noteworthy to mention that the higher measurement of accuracy with respect to distance and speed along with precise angular resolution depends on high intensity of radio wave frequency. Generally the frequency over 24 GHz is used for the smaller antenna size with the lower interference problem. The examples of various types of frequency band [12] used for different sensors are as under:

Short-range radio applications include:

- Blind Spot Detection (Blind Spot Monitoring)
- The lane and the lane-change assistant
- Rear end radar for collision warning or collision avoidance

- Park Assist
- Brake Assist
- Emergency braking,
- Automatic distance control

Radar configurations can be broadly categorized into three categories namely short-range radar with a maximum distance of about 30 meters, medium range radar with about 60 meters and long- range radar with about 250 meters. It may be mentioned that the use of Short Range Radar is increasingly seen with the detection of blind spot, rear and forward mitigation, parking assist etc. On the other, there are a number of detection system namely forward collision warnings, cross traffic alert, stop & go etc. operated by Medium Range Radar. So far there is no specific distinction made between SRR and MRR by the industry. It is seen now a days that ultrasonic sensors and highly automated driving are gradually replaced by the SRR. We do not have as such specific definitions and distinctions between the SSR and MDR as formulated by the industries. As far as the placement of sensors in the vehicles, the forward looking sensor for long range detection is generally placed in the front of the vehicle.

For a ‘cocoon’ radar system, extra sensors are placed on each side mid-body. Ideally, these radar sensors work on the 79-GHz frequency band with a 4-GHZ bandwidth. It may be mentioned that, global frequency specifications so far allow only 1 GHZ bandwidth at 77 GHz. Now a days a radar MMIC (monolithic microwave integrated circuit) comprises of three transmission channels (TX) and four-receiver channel (RX) to be monolithically integrated. Whether it creates a sense to integrate base band processing in the monolithic microwave integrated circuit (MMIC) or whether it is better to concentrate on a raw data radar sensor, it is a matter of debate.

The difference is that the output of the baseband processor provides so called pre-targets. In this case, data is pre-processed such as unverified information on speed, distance, signal strength, horizontal angle, and vertical angle for each detected object. The raw data radar sensor presents unfiltered raw data, to the ECU for processing. **Figure 9** demonstrates the architecture of such a raw data radar sensor. The radar sensor used as partitioned simplifies the data fusion of the video and radar data, and LIDAR data since the same communication interface can be used A prerequisite for the development of MWICs (Millimeter Wave Integrated Circuit)

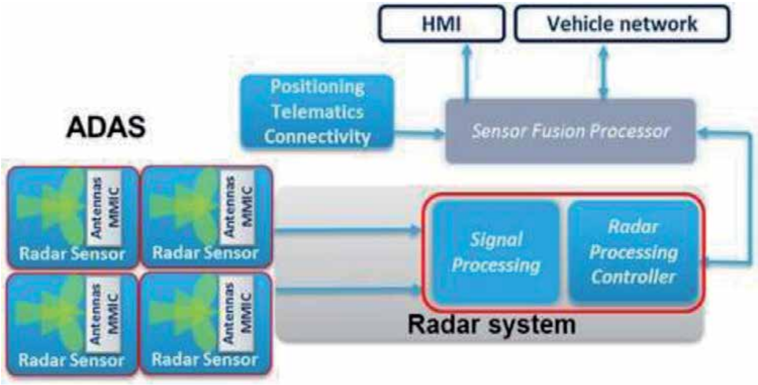


Figure 9.
Radar architecture for processing of raw data. Source: (https://www.sensorsmag.com/components/three-sensor-types-drive-autonomous-vehicles

is dedicated high-frequency (HF) technologies to realize the frequencies (24 GHz or 77 GHz) and the corresponding output power. **Table 2** presents *summary table of the properties of a radar sensor in certain ADAS.*

Multiple transmitters and receivers are generally are in-built to determine range, angle, and velocity of objects in their field of view. As various sensors are concerned, it consists of ultra-short-range- radar (USRR), short-range-radar (SRR), medium-range-radar (MRR), and long-range-radar (LRR) sensors or systems.

Property	Present in systems	Comment
Frequency: 76–77 GHz Range: 1 to 200 m Search Area: 12° Speed measurement precision: < 0.2 km/h	• Adaptive Cruise Control • Lane Change Assistant	Long range, Pulse Doppler, Active sensor,
Angular Precision: < 0.3°		
Frequency: 24.125 Ghz Distance range: 10 m Velocity range: 60 m/s Field of view: • ± 50 0Horizontal • ± 0,5 m/s Accuracy: • ± 0,05 m • ± 0,5 m/s Smallest object: Metal bar 10 mm dia (vertically placed at 1,5 m) Dimensions: 90 x 40 x 15 mm	• Adaptive Cruise Control • ACC/Stop &Go • ACC/Stop & Go +Foresight • Lane Change Assistant • Automatic Parking	Short range
Frequency: 24 GHz	• Forward Collision Warning	Forward looking, long Range
Frequency: 24 GHz Frequency: 5.8 GHz	• Side Obstacle Detection	Side looking, short range Side looking, short Range
F = 76.5 GHz Resolution = 100 cm Bandwidth = 100–500 MHz Range = 7–150 m		Long range, Pulse Doppler, Active sensor,
Radar	• Near field Collision Warning • Lane Keeping Assistant • Obstacle & Collision Warning • Rural Drive Assistance • Obstacle and Collision Avoidance	Active sensor,
Frequency: 24 GHz UWD (Ultra Wide Band) Resolution = 3 cm Bandwidth = 5 GHz Range: 0.3–30 m	• Pre-Crash Collision and Mitigation System	Short range, Active sensor,
Transmission Power = −41.3 dBm/ MHz		
Infrared Radar	• Pre-Crash Collision and Mitigation System	Near Infra Red (NIR), Far Infra Red (FIR),

Table 2.
Summary table of the properties of a radar sensor in certain ADAS, source: Ref. [12].

5.4 Ultrasonic sensing system

The primary philosophy of working with the ultrasonic technology is to transmit short bursts of sound waves that return back after hitting objects for which the measurement are to be taken in terms of time required to bounce back with speed of approximately 346 m/s which is the speed of the sound. For detection of short distance range obstacle, Ultrasonic sensors are increasing being used in the automobile industries which is generally characterized by with a sound pressure kHz and detection covering range of one to three meters supported by horizontal beam width of maximum100°and60°vertical. The ultrasonic and radar technology complements each other to determine the higher degree of accuracy,

Ultrasonic sensing is generally meant for short-distance applications at low speeds, such as park assist, self-parking, and blind-spot detection. For maximum coverage, an automotive ultrasonic system typically performs with multiple sensors placed in the wing mirror and front and rear bumpers. Ultrasonic sensing is a more cost-effective approach than cameras, which have poor close- distance detection. Though infrared sensing is cheaper than ultrasonic, it's less accurate and cannot function properly in direct sunlight. Objects closer to the transmitter generate a stronger echo than an object with more distantly located. In order to avoid false

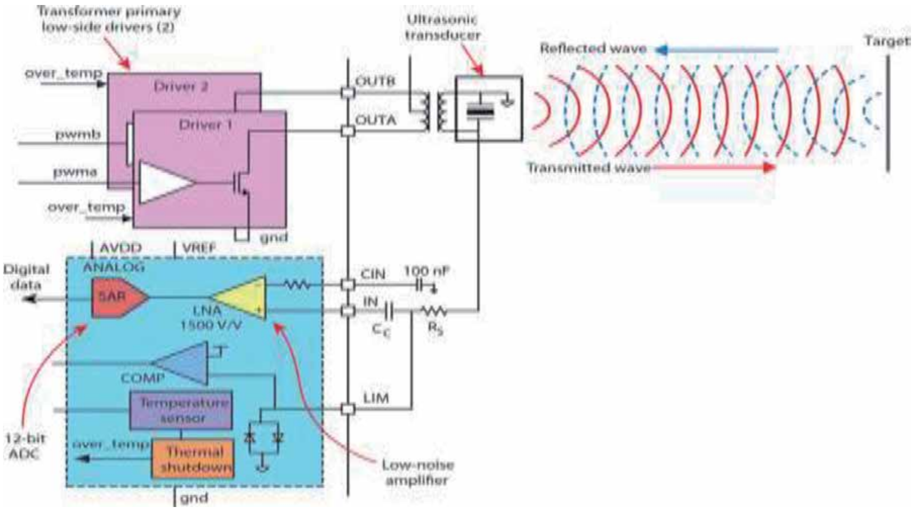


Figure 10.
This ultrasonic system features a PGA450 analog front end (source: Author/PGA450-Q1 PDF).

Property	Present in systems	Comment
F = 40 kHz Distance range: 0 to 3 meter Distance accuracy: 10 cm Angular range: 120° Angular accuracy: +/− 5° Response time: 60 ms	• Pre-Crash Collision and Mitigation System • Obstacle & Collision Warning	In adverse weather conditions

Table 3.
Summary table of the properties of an ultrasonic sensor in certain ADAS, source: https://www.embedded.com/how-smart-sensors-enhance-adas-designs/

positives, the system neglects all inputs that are less than that of the noise. The important parameters related to the specifications of ultrasonic sensor are the frequency, sensitivity, and directivity. The system is further characterized by the tunable transformer that is required to excite the transducer.

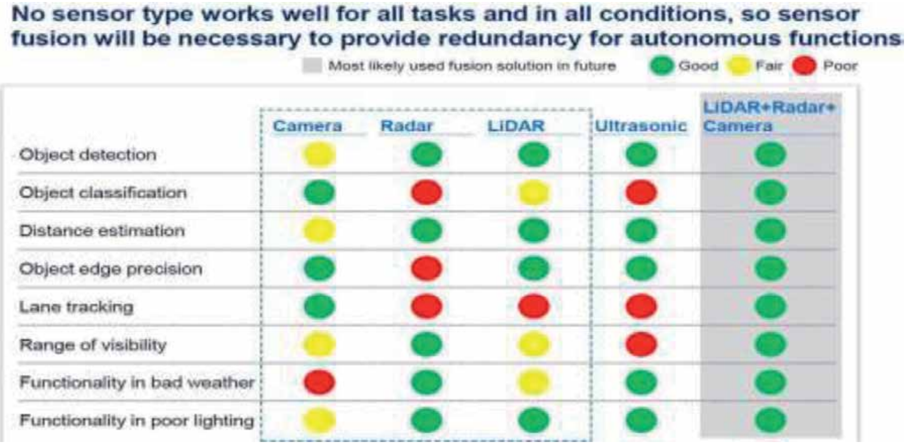
A tuning capacitor built into the system is concerned with matching the resonant frequency between the transducer and transformer. The speed of sound in air is affected by air temperature, humidity, and wind. If multiple sensors are applied, they must be placed in sufficient space so that the sensor signals do not interfere. **Figure 10** shows the features of ultrasonic system (**Table 3**).

6. Understanding the design of ADAS

It is realized that in order to make the ADAS commercially viable, three aspects on designing, testing and validating are of great importance and challenge to researchers/ scientists and manufacturer. The processing and sharing of information requiring a huge computation effort, within its fusion system in real time situation is a complex and difficult task in view of the computational load and the time-constraints placed on the system.

The inertial navigation systems identify, measures position, orientation, and velocity measurements. The sensor of RT-Ranges [13] is responsible for creating a real-time network, which is capable of tracking multiple targets, calculating distance, time to collision, and other relative measurements. Targets include primarily road vehicles, vulnerable road users (VRUs) such as cyclists or pedestrians. Euro NCAP (The **European New Car Assessment Programme,)** targets traffic assets and more. Euro NCAP is a European car safety performance assessment programme. Data is available in real-time on a software dashboard captured to verify test outcomes. Vehicle-to-vehicle measurements can be made over a 1 km range. Many similar systems in different parts of the world are increasingly seen, all with a slightly different name.

Various system of ADAS associated with various sensors is presented in the **Table 4**. A number of sensors developed during the process of development of ADAS are briefly discussed below.



No sensor type works well for all tasks and in all conditions, so sensor fusion will be necessary to provide redundancy for autonomous functions

▨ Most likely used fusion solution in future 🟢 Good 🟡 Fair 🔴 Poor

	Camera	Radar	LiDAR	Ultrasonic	LiDAR+Radar+ Camera
Object detection	🟡	🟢	🟢	🟢	🟢
Object classification	🟢	🔴	🟡	🔴	🟢
Distance estimation	🟡	🟢	🟢	🟢	🟢
Object edge precision	🟢	🔴	🟢	🟢	🟢
Lane tracking	🟢	🔴	🔴	🔴	🟢
Range of visibility	🟡	🟢	🟡	🔴	🟢
Functionality in bad weather	🔴	🟢	🟡	🟢	🟢
Functionality in poor lighting	🟡	🟢	🟢	🟢	🟢

Table 4.
Various sensors related to their applications. Source: Automotive ADAS Systems, ST Developers Conference, Sep, 12, 2019, Santa Clara Convention Centre, Mission City Ballroom, Santa Clara, CA.

6.1 Night vision

During night vision, one is more concerned with the proper visibility where the camera plays an important role. Therefore the camera for this purpose is designed with the use of near or far infrared to improve the perception of the driver in dark conditions. The improved sight vision created by the above near or far- infrared camera is displayed in the monitors of the vehicle. Human Machine interface though poses an issue for correctly showing the road-side picture for timely intervention plays an important role to the driver to enhance the safety to the driver so that the driver is not distracted. **Table 5** *presents available sensors with their properties for night vision.*

6.2 Lane departure warning

Lane departure warning mechanism works on the principles of certain thresholds with respect to distance, time to lane crossing. It is based on the decision made out from the data fusion analysis supported with computer software algorithm to warn the driver that he or she is about commit mistake in departing traffic lane. For example, sensors such as acoustic, optic means continuously generate and analyze the data along with the video image processing data created by the vehicle cameras results in the detection of warning to the vehicle. In order to make the warning system effective, the carriage way would have to be laid with Good visible lane markings system. These influence the complexity of the system on the roadside. This system aims to prevent involuntary lane departure, which constitutes a relevant cause of road accidents. With real-time measurement and positional accuracy which is generally at less than 2 cm, the system captures the data that the sensor performs the task of lane departure action as shown in **Figure 11**. This warns the Lane Departure Warning system if the vehicle suddenly decides to change the lane without proper indication. The camera used for the lane detection system is low cost generally mounted on the windscreen near the rear view. The position of this location of the camera helps continuously capture the image of solid lane line marking of the road towards the front side of driving. it also works along with the front (adaptive cruise control and, ii) forward collision warning), side (lane departure warning), and iii) rear side (blind spot detection).

6.3 Near field collision warning

There are multiple collision warning systems (12) mentioned on the **Table 6**.

The finest example of the application of near field collision warning is the detection of blind spot, which takes very close proximity of the presence of vehicle. Lidar, radar or vision based sensors are generally used. It may also be acoustical, haptical or optical also. In many cases, the frequency of this kind sensor is found to be 24 GHz. To test and develop blind-spot detection systems, it is necessary to

Sensor	Property	Comment
Infrared camera	l = 800 nm	**N**ear **I**nfra**R**ed (NIR)
(CMOS)	= 7–14 µm	**F**ar **I**nfra**R**ed (FIR)
		Both systems are mono-camera, mono-camera,

Table 5.
Available sensor and properties in night vision systems, source: [12].

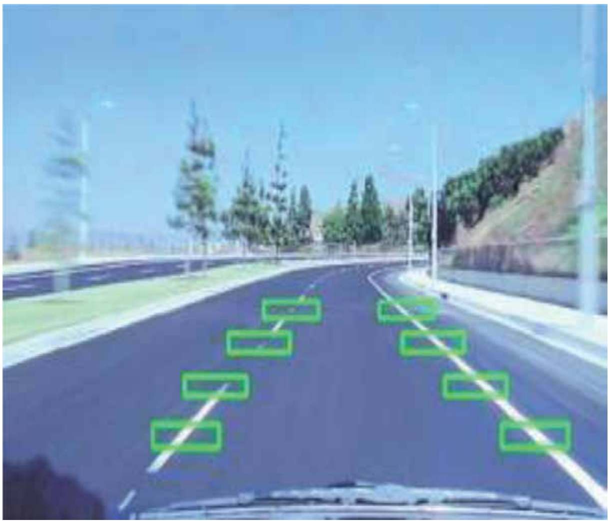


Figure 11.
How it works: Windshield camera tracks lane markings. Source:https://www.extremetech.com/extreme/165320-what-is-lane-departure-warning-and-how- does-it-work.

Sensor	Property	Comment	
Infrared camera	l = 800 nm	**N**ear **I**nfra**R**ed (NIR)	
(CMOS)	= 7–14 μm	**F**ar **I**nfra**R**ed (FIR)	
		Both systems are mono-camera,	mono-camera,

Table 6.
Available sensor and properties in night vision systems.

accurately measure the position and trajectory of targets relative to the vehicle under test (VUT). The system may require the following protocol accuracy:

- Relative accuracy 2 cm
- Heading accuracy 0.1°
- Free post-processing software
- Ability to track multiple objects in real-time
- Perfectly suited to open-road testing

To evaluate blind-spot detection systems, an RT inertial navigation system and RT- Range S [7] are installed in the vehicle under test. This powerful system is designed to work in conjunction with GNSS-aided inertial navigation products. Automobiles can be equipped with GNSS receivers, which display moving maps and information about location, speed, direction, and nearby streets and points of interest. The manner in which sensor works is based the measurement of real-time distance between the sensor and the identified object. It may include any type of vehicle, blind corner of a junction, pedestrian and bicycle etc.

For real-time testing, range measurements from the RT-Range S Hunter can be used as output via Ethernet or CAN (Controller Area Network) which is a communication hardware that allows communication between parts of a system without the intermediary of a central computer. Or data can be logged internally and analyzed back at base where it can be post-processed and exported in CSV file format ((“Comma Separated Values”) which is often used to exchange data between differently similar applications).

6.4 Forward collision warning

Warning system developed by EATON-VORAD in the USA for trucks and busses [13] as the first step towards the **A**dvanced **D**river **A**ssistance **S**ystems (ADAS) can be considered as Forward Collision Warning System. Forward Collision warning with a frequency of 24 Ghz is first seen in the USA market in 1995. It used to detect the object with signal emitted through either optical or acoustical method to the driver when the object happens to be close to the path of collision.

6.5 Side obstacle detection

This system addresses a side looking short-range radar that operates at 24 GHz. This sensor identifies and detects side obstacles that are signaled with a proper display. As a further option, the system can also be linked to engine control with a view to controlling speed. This function is called “Smart Cruise”. More recently, the side obstacle detection System has been introduced also on Volvo cars based on camera sensor and image processing.

6.6 Curve & speed limit information

This system communicates with the driver about speed limits and informed the recommended speed at curves. There are a number of relevant information generated from digital maps, image processing or communication system between the interactions of vehicles and road infrastructure. That is the reason that updated real time data is important to the driver generated from the above which helps in recognizing the speed limit of the road where the vehicle is traveling. It may be mentioned that the details of the road features such exact location of traffic marking, position of street light etc. are available in the form of digital map in ADAS that helps in identifying and recognizing the speed limit.

6.7 Adaptive cruise control (ACC)

This system was introduced firstly inside Japan, and then in Europe for the car market. ACC systems are based on a front looking sensor designed with laser radar, (LIDAR) or microwave radar with a maximum detection range of around 100 m. The microwave radar sensor operates in the 76–77 GHz bands that have been reserved for application of automotive obstacle detection. Based on front vehicle information, mainly distance and speed, the ACC system regulates own vehicle speed by acting on engine control and braking system. The ACC is an extension of the standard Cruise.

Control system, with the extra capability to adapt the speed of the vehicle to the speed of the preceding one. This function was firstly introduced in Japan on 1995 based on LIDAR technology.

Europe experienced the emergence of lidar and microwave technology in the following years which led the introduction of these technologies in the Mercedes car during the year 1999. It is noteworthy to mention that the automatic cruise control system (ACC) was seen fitted with truck manufactured by Mercedes automobile industry. Presently around twenty automobile manufacturers are producing this type car and truck.

It is based on a high performance GNSS/INS for dynamic applications developed on the convenience of a conventional cruise control system by automatically changing speed to match the vehicular flow in front. It's important to determine precisely when and how the system intervenes, how well it acquires and then it tracks the targets and how it performs in a number of different real-world scenarios [6]. Measurements such as target bearing, distance, relative velocity and time-to-collision are key to the evaluation of these systems. Sensors with RT and RT range for ACC offers the following characteristics:

- Relative accuracy 2 cm
- Heading accuracy 0.1°
- Real-time birds eye view showing measurements
- Ability to track multiple objects in real-time
- Perfectly suited to open-road testing

In order to get accurate vehicle-to-vehicle measurements, an RT inertial navigation system and RT-Range S [7] are installed in the vehicle under test (VUT) and any target vehicles. An RT inertial navigation takes into account a number of parameters for operation. These include position with respect to latitude, longitude, altitude distance and its coordinate position. Besides the position of these, velocity, acceleration, orientation, angular rates and acceleration and slip angle are also taken into account. RT-XLAN Wi-Fi radios then send real-time information from target vehicles back to the VUT where the RT-Range S calculates, logs and outputs real- time measurements about the relative position of the target vehicles. The measurements being the output include the position of both the Hunter and target vehicles, orientation and velocity. The current status of the ACC hardware can also be logged with the data via a CAN bus interface, which is a robust vehicle bus standard designed to allow microcontrollers and devices to communicate with each other in applications without a host computer. It can also be or later synchronized with the measurements via a GPS time stamp. Moreover, from some manufacturers, ACC is given in combination with lane warning system.

It will have frequency allocation for 24 GHz sensors. The properties of various sensors associated with the functioning this ACC are presented in **Table 7** as under:

6.8 ACC/stop & go

Adaptive cruise control (ACC) permits a driver to travel with the flow in traffic. In this situation, a radar sensor monitors the situation in front of the vehicle. As the road is observed to be clear, ACC operates with the desired speed. If the radar sensor finds a slower vehicle ahead of it, ACC automatically maintains and adjusts the speed a preset distance. In the Stop & Go version, the system results in slowing the car down in a traffic jam, or even comes to a halt it completely. If the car has an automatic transmission, Stop & Go also restarts the engine once traffic gets moving again after a brief pause.

6.9 Stop & go

In this system the driver continues to receive support from this sensor with respect longitudinal control for the formation of queue. During the stop & go of the vehicle facing the front side, longitudinal control is carried out by the system for detecting the near side objects.

Sensor	Property
LIDAR	Wavelength l: 850 nm
Radar	Frequency: 76–77 GHz Range: 1 to 200 m Resolution: 100 cm Search Area: 12° Speed measurement precision: < 0.2 km/h Angular Precision: < 0.3° Frequency: 24.125 Ghz Distance range: 10 m Velocity range: 60 m/s Field of view: • ± 50 0Horizontal • ± 0,5 m/s Accuracy: • ± 0,05 m • ± 0,5 m/s Smallest object: Metal bar 10 mm diam (vertically placed at 1,5 m) Dimensions: 90 x 40 x 15 mm

Table 7.
Available sensors with their properties in ACC (source: Ref No. [12]).

6.10 Lane keeping assistant

The function of a lane keeping assistant system includes the lane detection and the feedback to the driver if he is leaving a defined trajectory within the lane. Lane departure warning systems merely alert the driver when the car is leaving its lane, while lane-keeping assist actually works to keep the car from moving out of the lane. An active steering wheel can help the driver with a force feedback to keep on this trajectory. The lane is detected by a video image processing system. Additionally to the lane departure warning aspects especially regarding the infrastructure, the HMI becomes more important.

The driver gets all assistance through his touch with steering and other devices for taking decisions for the vehicular movement linking with the controller that also helps to lane keeping assistance to adhere to lane driving.

The Protocol accuracy requirements [12] for this are as under:

- Axes to be in ISO 8855:1991 orientation
- Longitudinal speed to 0.1 km/h
- Update rate at least 100 Hz
- Time is required as a synchronization DGPS (Differential GPS)
- Position to 0.03 m
- Yaw velocity to 0.1°/s
- Acceleration to 0.1 m/s^2
- Vehicle edge to lane edge measurements

For the LSS (Lane support System) LKA tests, the key measurements are the distance between the outer-edge bulge of the front tires and the inside edge of the lane markings when any intervention is triggered.

6.11 Local hazard warning

If a hazard occurs far away in front of the vehicle, so that the driver cannot see it, this system will warn him. By the means of communication, it is possible, to transfer this information over long distances. A usable frequency has to be allocated. Local Hazard Warning [14, 15] is a system that uses short-range communication between cars, and between a car and its surroundings, to give drivers early warning of safety hazards. For example, a car equipped with Local Hazard Warning might issue a warning to other vehicles if it had broken down in the middle of a carriageway or had been involved in a collision. Similarly, emergency vehicles equipped with such a system might send a signal to nearby vehicles to warn them of their presence, or temporary roadwork barriers could issue for such warnings. As well as transmitting such warnings, cars equipped with Local Hazard Warning can also receive these signals and use them to alert the driver to the danger [16].

6.12 Automatic parking

The automatic parking is a function that helps the driver entering into a parking slot in a parallel maneuver by automatically acting on the steering wheel and engine control. The sensors measure [12] the object with following accuracy:

- Relative accuracy 2 cm
- Heading accuracy 0.1°
- Real-time birds eye view showing measurements
- Ability to track multiple objects in real-time

The vehicle is fitted with a GNSS-aided inertial navigation system (GNSS/INS). In most cases (because of the low speeds involved), a dual-antenna model is fitted to maintain the best headway accuracy at all times. The properties of various sensors [12] are presented in **Table 8**.

Sensor	Property	Comment
Laser	Beam deflection: horizontal Range: 0–80 m Range: 0–35 m @ Rr = 5% Resolution: 20 mm Accuracy: ± 50 mm Frequency: 10–40 Hz Cycle time: 25–100 ms Vertical opening angle: ~ 3,5° Horizontalangularfield:+ – 120° Lateral resolution: 0,25° - 1°	
Radar	Frequency: 24.125 Ghz	Short range
	Distance range: 10 m Velocity range: 60 m/s Field of view: • ± 50 0Horizontal • ± 0,5 m/s Accuracy: • ± 0,05 m	
	• ± 0,5 m/s Smallest object: Metal bar 10 mm diam (vertically placed at 1,5 m) Dimensions: 90 x 40 x 15 mm	

Table 8.
Various sensors available for automatic parking and their properties.

6.13 Pre-crash collision and mitigation system

Pre Crash Safety Systems identify an imminent crash and deploy safety devices such as seat belt pretensions.

Pre Crash Safety Systems identify an imminent crash and deploy safety devices such as seat belt pretensions. In order to reduce the damages of an accident, this system has been designed that is capable of applying brake automatically after identification of imminent occurrence of collision. As discussed earlier, various sensors such as Lidar, Camera etc. play an important role in identifying the hindrance for an imminent collision. This feature is primarily designed to address the problem of safety, which integrates the sensitivity of seatbelt. If one happens to wear the seat belt during the occurrence of road accident, the chances of being injured is quite less. Most of the seat belts now available in the car are very sensitive, as the vehicle will not move if car users or someone does not wear seat belt.

6.14 Obstacle & collision warning

The driver will be warned if a potential collision is detected with e.g. another car or obstacle. This warning can be, for example acoustic, visual. The functional limits of these systems have to be clearly pointed out.

In city environments, collision between vehicles and pedestrians or cyclists often result in serious injuries as there is a little time for either party to react. Protocol accuracy requirements [12] for this kind of collision are the following.

- Update rate at least 100 Hz
- Lateral path error
- Time is required as a synchronization DGPS
- Position to 0.03 m
- VUT (Vehicle under test) Speed to 0.1 km/h
- Yaw velocity to 0.1°/s
- Acceleration to 0.1 m/s^2
- Polygon perimeter shapes

6.15 Intersection safety

In an intersection situation especially in cities, a driver has to fulfill several tasks in parallel. In order to assist the driver in such situations, it is necessary to support certain tasks like approaching a stop sign/traffic light or right of way of crossing traffic. The complexity of the possible intersection scenarios leads to the high risk probabilities of causing accidents. As any intersections are designed to address a number of turning movements of automobile traffic coupled with the non-motorized and pedestrian traffic, the detection and recognition are not as simple as on a straight section of a road. Due to these complexities, the safety of the road intersection would have to be taken into all possible scenarios to make hazard free zone.

7. Autonomous driving

The driving of **vehicle** is **controlled by a** computer **algorithm in e**ach situation. It is presently viewed **that this** fully AV **cannot be reached** at the present situation **in the actual road net**work immediately. There is an expectation that true Level – 5 of AV to attain full autonomy is about ten-plus years away. It is also expected that geo-fenced applications of autonomous vehicles (AVs) would reach in the next three to five years. The progress on the hardware as well as software has actually been very significant. The cost of LIDARs [light detection and ranging sensors], for example, has dropped by a factor of ten over the last five years. Similarly, the amount of computational capacity that the GPUs [graphic processing unit] has also increased significantly. ISO 39003 is now working on Guidance on Safety Ethical Considerations for Autonomous Vehicles in order to ensure that this vehicle is absolute safe and smooth from operational point.

8. Conclusions

Although there are many demonstration seen on advanced vehicles up to Level 3 or more, so far automobile manufacturers have not been able to commercialize to the high level automated vehicle which requires detailed and comprehensive legislation in the countries.. International Standard Organization is presently working on the standards for this automated vehicle. A number of fundamentals aspects of ADAS that are a part of the complex process of the system have been discussed. ADAS with level-2 are becoming increasingly available in the market in western countries with implication of increase in its cost. It may be mentioned that the manufacturers of ADAS driven vehicles have not been able to make any significant impact on the sale of this type of vehicle. It may be mentioned that there is not significant negative values experienced so far. The R&D into ADAS is increasingly being accelerated to enhance safety.

Though the ADAS driven vehicle is yet to find its place in the market in spite of apprehension raised by many sections of people on the safety related issues, it would be important to appreciate when it turns into Cooperative Road Vehicle Highways System reducing the probability of accident to almost zero level.

The European Community (16) is leading by investing significantly in R&D into ADAS in Europe. Many countries such as France, the Netherlands and the UK are increasingly taking an active role by participating in research activities and promoting successful implementation. The most important issues of ADAS have two key factors: i) a high level of usability and ii) a low financial risk to the manufacturer. It seems for the time being, ADAS user benefits are not clear yet and financial risks still exists.

As far as legal aspects are concerned, the relation between ADAS and product liability is very important. The product liability for ADA systems will address specific additional requirements, in particular taking into account the interaction of the drivers/users with the product in view of the current legal framework. The Code of Practice is being developed by the European car manufacturers by addressing these requirements. Presently the ADASE II technology roadmap for ADAS confirmed the expectation that ADAS will have potential benefits on safety, throughput and comfort, ranging from positive to very positive. Related technology development, R&D is still required to improve the performance of ADAS to cover wider ranges of traffic scenarios and to bring down costs. Political motivations and intervention may be needed to advise to the different decision makers to accelerate and facilitate (or regulate) the market the introduction of ADAS.

Therefore the government should come forward along with the concerned stake holders like road operators, car manufactures, users etc. by jointly setting up proper conducive environment in order to promote the advances of ADAS. The ADAS driven vehicles should be commercially viable in the market by addressing concerned legal issues in the society.

Author details

Pradip Kumar Sarkar
Department of Transport Planning, School of Planning and Architecture,
New Delhi, India

*Address all correspondence to: pradipsarkar799@gmail.com

IntechOpen

References

[1] https://www.acea.be/uploads/publications/20090831_Code_of_Practice_ADAS.pdf

[2] http://www.hitachi-automotive.us/Products/oem/DCS/ADAS/index.htm (ARCHi

[3] Mahdi Rezaei Ghahroudi^1 and Reza Sabzevari" Multisensor Data Fusion Strategies for Advanced Driver Assistance Systems", https://www.researchgate.net/publication/221787813

[4] D$_R$. MADHU B K1, KARTHIK KOTI2, K SURABHI3, NIKHIL U4,YASHWANTH M." VEHICLE COLLISION AVOIDANCE SYSTEM", International Research Journal of Engineering and Technology(IRJET,) e-ISSN: 2395-0056,Volume: 07 Issue: 06 | June 2020

[5] https://www.kpit.com/insights/architectural-concepts-for-autonomous- driving-applications/ –(Sensor_Architecture)

[6] https://pdfs.semanticscholar.org/4bce/e9bcdfae49c3654429f03be03fb8c5a7836e.pdf (all sensors-twent)

[7] https://www.oxts.com/industry/adas-testing/

[8] https://www.sae.org/news/press-room/2018/12/sae-international- releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving- automation%E2%80%9D-standard-for-self-driving-vehicles

[9] https://www.first-sensor.com/en/applications/mobility/

[10] https://www.embedded.com/how-smart-sensors-enhance-adas-designs/

[11] An Introduction to Automotive LIDAR, Texas Instruments, Motaz Khader Temperature and Humidity Sensing Samir Cherian High-Speed Amplifiers

[12] Capita Selecta: Virtual Reality, Sensors in ADAS, 8th Jan, 2007, University of Twent, Martijn Pijpers, s9709401

[13] https://support.oxts.com/hc/en-us/articles/115002772345-RT-Range-Measurements

[14] www.sae.org/publications/technical-papers/content/2009-01-2911/

[15] https://www.autocarpro.in/news-international/volkswagen-local-hazard- warning-car2x-safety-tech-wins-euro-ncap-award-55905

[16] Information and communication technologies for safe and intelligent vehicles, Communication from the Commission to the Council and the European Parliament, SEC, 2003) 963

Chapter 12

BIM Approach for Smart Infrastructure Design and Maintenance Operations

Salvatore Antonio Biancardo, Nunzio Viscione,
Cristina Oreto and Francesca Russo

Abstract

In the age of the Internet-of-Things and Big Data, Building Information Modeling (BIM) is being expanded into sectors for which it was not originally designed, such as the infrastructure sector, and becomes a necessity for the planning and management of smart cities. The digitization of the urban environment, its building and infrastructural heritage and its services is at the center of the concept of smart city, and this appears strongly linked to the use of BIM on an increasingly extended scale as an enabling tool for planning cities that are increasingly intelligent, sustainable, interconnected and above all liveable. In this chapter a creation process for the digitalization of existing roads, as well-known as reverse engineering method, will be shown as follows: a) modeling 3D digital terrain model; b) creating the horizontal alignment, vertical profiles and editing cross-sections; c) modeling the 3D corridor. As a response to long-term development between BIM and road engineering, this chapter will contribute also by offering innovative and practical solutions for integration of road design and pavement analysis, for a better management and optimization of road pavement maintenance.

Keywords: building information modeling, smart infrastructure, road pavements, computational design

1. Introduction

Smart city is to lead the transformation of urban development with innovation, comprehensively promote the new generation of information and communication technology and the new urbanization development strategy, deeply integrate and improve the modernization level of urban governance capability [1].

A self-respecting smart city cannot ignore serious and far-sighted planning that bases urban landscape design on Geographic Information Systems (GIS) and integrated modeling, which Building Information Modeling (BIM) is able to ensure; this is because only having a clear vision, implemented with a precise planning, of what is going to be built is it possible to avoid the destructive effects that a construction practice without adequate tools can cause. The use of GIS and BIM together therefore allows you to plan, design, build and manage infrastructure resources more efficiently and save time and money.

BIM is widely recognized as a fundamental methodology for relaunching the global economy: this is why in many countries digitalization processes have been started in the AEC sector [2].

BIM starts at the planning and conceptual design stage and continues throughout the lifecycle of the asset. It is important that intelligent information is not lost as the project progresses through the various stages of a BIM Infrastructure Project.

The entire process of developing, executing and managing infrastructure projects can be transformed—initial surveying and data collection, environmental review, public participation, design and documentation, bidding, construction, and operations and maintenance [3, 4]. The model-centric approach enables planners, engineers, and designers to explore and validate innovative design ideas and what-if scenarios with project investors.

To model a smart infrastructure, it is necessary to find a set of variables and parameters essential for the analysis and prediction of the performance of built objects [5].

Data modeling can be performed by procedural, also known as parametric, modeling that provides object-oriented n-dimensional information or generative model information containing objects created through algorithmic processes [6].

Parametric and procedural 3D geometrical models can be represented by graphs in order to define relationships and dependencies between geometric entities and allow its reuse in similar design scenarios or to adapt it to different scenarios [7, 8].

The models created for BIM are not just 3D geometry; they are data-rich objects which are: intelligent - parametric engines help define relationships between objects and keep changes consistent and coordinated; knowledge-based - can be constrained by things like AASHTO codes, design criteria, and company standards; scalable - able to aggregate huge amounts of data from multiple sources; visual - enable better analysis, simulation and communication [9].

In the last few years, researchers have been focusing their attention on assessing the benefits of using digital tools and processes to support effectively the entire life of transportation facilities and road infrastructures, from strategic planning, design and construction [10–17] to performance management and maintenance [18–21].

Marzouk and Othman [22] proposes an inclusive framework for integrating Building Information Modeling (BIM) and Geographical Information System (GIS) to plan and forecast the utility infrastructure needs for expanding and emerging cities to highlight the concept of "smartness" during the planning stage.

As highlighted by Sankaran et al. [23], BIM is an efficient method for collecting and updating as-built data for creating a digital archive of information to facilitate management and future project development.

For example, Tang et al. [24] created a platform for the integration of Building Information Modeling (BIM) based road design and pavement structural analysis, allowing to establish a conversion between the three dimensional (3D) model and the finite element method software ABAQUS, providing quality data and powerful technical support and minimizing the uncertainty factors in the road design and maintenance processes.

Also, the design process was supported through the implementation of an empirical model for the analysis of permanent deformation of the asphalt pavement, which allows selecting the pavement that best suits the desired service life [25].

There is a need for such an approach to assist decision makers to ensure enterprise's objectives and targets are maximized with given budget and planned shutdown time [26, 27].

Interoperable BIM model has been adapted to perform complex multi-physical studies and simulations in several technical fields (including noise exposure, wind comfort, artificial and natural lighting, energy consumption, environmental impacts and global comfort) [28].

However, the existing BIM-based decision-support methods have primarily focused on building design and construction. Therefore, they are limited in their ability to provide an appropriate methodology for master planning of large-scale development projects [29].

It is of great significance to promote the application of BIM technology in the life cycle management of projects in the context of smart cities, ensure the consistency and interoperability of BIM deliverables at all engineering stages, and realize the comprehensive management of the construction industry in smart cities [30, 31].

2. Goals definition

Analysis procedure presented here aims to offer an innovative and practical methodology for integration of road design and pavement analysis, for a better management and optimization of road pavement maintenance.

The work phases are shown in **Figure 1** and basically are carried out as follows:

1. Building the existing ground surfaces; surfaces are used to derive alignments and profiles, and for corridor grading;

2. Designing horizontal-vertical alignment; alignment are used by corridor as its centerline while profiles use existing ground profiles and design finished grade profile (vertical alignments);

3. Create the required assemblies; subassemblies are used to build the required assemblies;

4. Create the 3D corridor;

5. Information management for a decision support system for the management of maintenance processes.

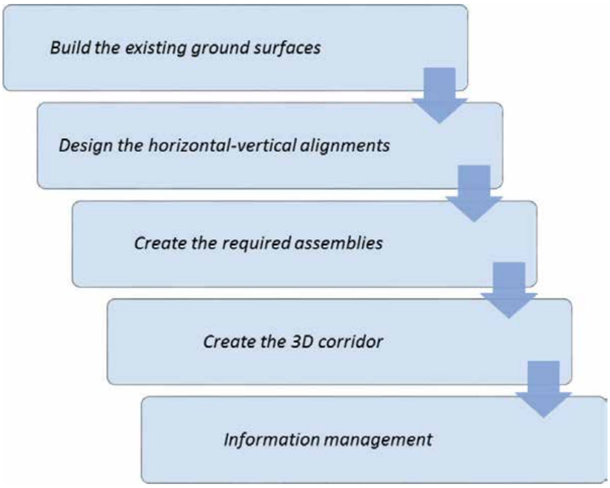


Figure 1.
Methodological approach.

3. Road modeling

3.1 Digital terrain model

Digital Terrain Modeling (DTM) is a concept that underlies all calculations in Civil Engineering involving elevation or slope - profiles, cross sections, grading and volume calculations.

The process of DTM involves the creation of a data structure that the software can instantly "touch" to retrieve elevations or slopes, representing either existing or proposed conditions.

DTM mathematically completes all interpolation possible between the data supplied, and stores the result in a digital file for easy retrieval.

Surfaces can certainly be produced from other data types, including point data.

There are certain data types that are universally applicable to any DTM effort in Civil Engineering and Surveying. These data types are constant in any program: Civil 3D, Open Roads Designer, ArcGIS, etc.

The three data types which can be used in constructing a DTM are Point Data, Breakline Data, and Contour Data.

- Point Data - Point Data for DTM consist of individual discrete X, Y and Z locations, without connecting features between them. Typically, these will be spot elevations in a contour drawing, or the mass points themselves in a Mass Points and Breaklines drawing. Critically, the Point Data must have an elevation or Z component that can be processed in some fashion in building the elevation model. Spot elevation text at elevation 0 in a drawing can be used and processed by Map into an ASCII file, and ASCII files of XYZ format can be used as well.

- Breakline Data - Breaklines are also referred to as Faults, or Features. Breaklines, as used in this context, represent the linear edges of site features along which there is a noticeable change in grade. Successfully applied, a breakline forces a deflection in a contour to show a grade change. Examples are edges of pavement, shoulders, toes or tops of slope, toes or tops of wall, water features, etc. λ Contour Data - The definition of contour Data for Digital Terrain Modeling is very specific, and not necessarily what one would expect.

- Contour Data are strings of point data connected by segments in complex objects; the CAD representation is a polyline. Contour Data do not have to be at constant elevation, as one typically thinks of contours. Contour Data are a fast means of selecting and processing point data, utilizing the vertices of the objects. Most Digital Terrain Modeling applications will also process the segments between the vertices as breakline data, and can filter out vertices too close together or add interpolated vertices if required. Contour Data must be at a correct Z elevation to be processed in a Terrain Model. Polylines must be at a correct Z, either constant as a 2D polyline, or varying, as a 3D polyline. GIS data can again be used, and CAD Map can read elevation attributes from GIS Contour Data and apply them to polylines through a Property Alteration Query.

Most Civil Engineering and Surveying applications will utilize some combination of data types in a Terrain Model; having two types present is common and all three is not unusual at all.

Triangular irregular networks (TIN) are a representation of a continuous surface consisting entirely of triangular facets, used mainly as Discrete Global Grid in primary elevation modeling.

TINs can be constructed using three types of vector information: altitude measurements (mass points), surface continuity breaklines, surface continuity break polygons (polygon surfaces).

The points contain the X, Y coordinates and the Z value. All points are used to establish a connection with the two closest points to create triangles. Surface triangulation is based on the Delaunay algorithm, which ensures that no points are within the circle of a triangle.

The Discontinuity Lines represent the characteristics of a linear infrastructure such as curbs, retaining walls, etc. These lines also define the edges of the triangles. Breaklines can be created from linear entities such as line, polyline, arc, circle.

Contours are the characteristics to define bounded 3D surfaces. Surface contours can be created with closed polylines. Defining external contours on a large 3D surface improves the performance of a TIN Surface plane.

Figure 2 shows the triangles that result from Digital Terrain Modeling where the elevation value is retrieved from the digital surface and displays it in the Tooltip, instantly, and anywhere on the surface.

3.2 Horizontal-vertical alignment

Creating and defining a horizontal alignment is one of the first steps in infrastructure design.

In **Figure 3** a workflow to design and edit alignments is shown.

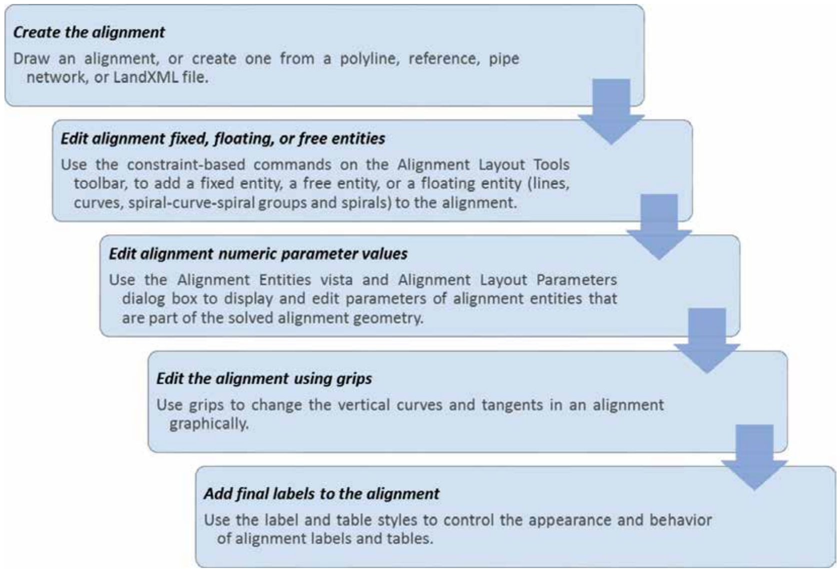
You can draw the alignment geometry as a polyline, and then create the named alignment from that geometry. For greater control, you can create an alignment object or You can also make edits to alignments using grips.

Create alignments in many ways, such as creating them from polylines, from pipe networks, and from LandXML data.

The alignment can be created using fixed, floating, and free elements:

- Fixed elements have its position totally defined by specifying a combination of start/end points or center, length, bearing or angle, and radius. However, as the fixed position of a computator is defined by points that are dependent (referenced) on other elements, a fixed computator is actually free to move as the referenced elements move. It is "fixed" in respect of its location to the referenced element;

Figure 2.
Digital terrain model.



Create the alignment

Draw an alignment, or create one from a polyline, reference, pipe network, or LandXML file.

Edit alignment fixed, floating, or free entities

Use the constraint-based commands on the Alignment Layout Tools toolbar, to add a fixed entity, a free entity, or a floating entity (lines, curves, spiral-curve-spiral groups and spirals) to the alignment.

Edit alignment numeric parameter values

Use the Alignment Entities vista and Alignment Layout Parameters dialog box to display and edit parameters of alignment entities that are part of the solved alignment geometry.

Edit the alignment using grips

Use grips to change the vertical curves and tangents in an alignment graphically.

Add final labels to the alignment

Use the label and table styles to control the appearance and behavior of alignment labels and tables.

Figure 3.
Workflow: To design and edit alignments.

- Floating elements have one unknown, which becomes the “floating” part. The unknown part can be the length, angle, point/centre, or radius. The other parts (one for lines, two for arcs) are fixed in position;

- Free elements are totally unconstrained and will be defined by the adjoining elements. Whilst an arc has two unknowns with only the radius, or one point, or length defined.

Once it is determined which element type best suits the design context, it can be selected the appropriate line, curve, transition, or combination based on available design data, such as whether you have a known through point, length, or radius.

When you create an alignment, you can use the criteria-based design feature to ensure that your alignment design meets minimum local standards and consequently easily identify and report standards violations.

The alignment is an interactive line with profiles, both existing ground and planned work.

Using profiles, you can view changes in elevation along a horizontal alignment. In addition to the centerline profile, you can create offset profiles for features such as waterway or ditch banks. On a profile view, you can also superimpose the profile of a different horizontal alignment that is in the same area. And like LandDesktop you can create a temporary profile that can help you view information at locations where there is no alignment (i.e. line, polyline, feature, or along a series of points you select).

The horizontal and the vertical alignments need to match in length exactly or else the corridor will not be created properly.

Figure 4 shows an example of road alignment with its relative ground and vertical profile.

Once both alignments are created, the next step is to create a section type, with surface depths, sub-earth depth, kerbing, banking, etc.

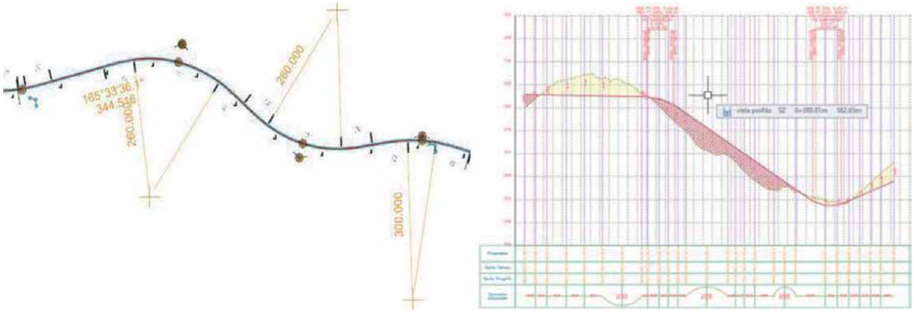


Figure 4.
Road alignment and its vertical profile.

3.3 Section type

Assembly objects contain and manage a collection of subassemblies that are used to form the basic structure of a 3D corridor model.

An assembly is an 3D drawing object that manages a collection of subassembly objects. Together, assemblies and subassemblies function as the basic building blocks of a roadway or other alignment-based design.

Adding one or more subassembly objects, such as travel lanes, curbs, and side slopes, to an assembly baseline creates an assembly object. This forms the design for a corridor section. The subassemblies are provided in a set of catalogs.

It is also possible to create more advanced assemblies referred to as conditional assemblies. A conditional assembly contains one or more conditional subassemblies, which apply subsequent subassemblies when specified conditions at a given station are met.

In **Figure 5** is shown a typical section type for fill and in presence of a bridge.

Specific BIM-based tools as Subassembly Composer/ Generative Components provide an interface for composing and modifying complex subassemblies, without the need for programming. Without the need to be an expert in programming, users can create custom subassemblies to meet their specific needs, making corridors have endless possibilities.

For example, in presence of a retaining wall characterized by a variation of the geometric characteristics in terms of height/weight along the road layout, it is possible to create a flowchart (see **Figure 6**) set with decision variables that change as the boundary conditions vary.

In the case in question, the section changes dimensional characteristics as the distance between the road surface and the ground surface changes.

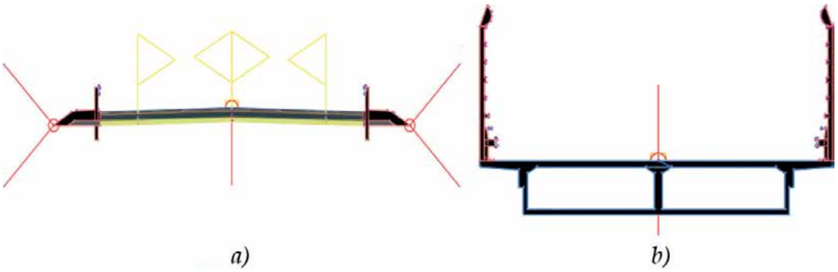

a)
b)

Figure 5.
Section types. (a) Fill, (b) bridge.

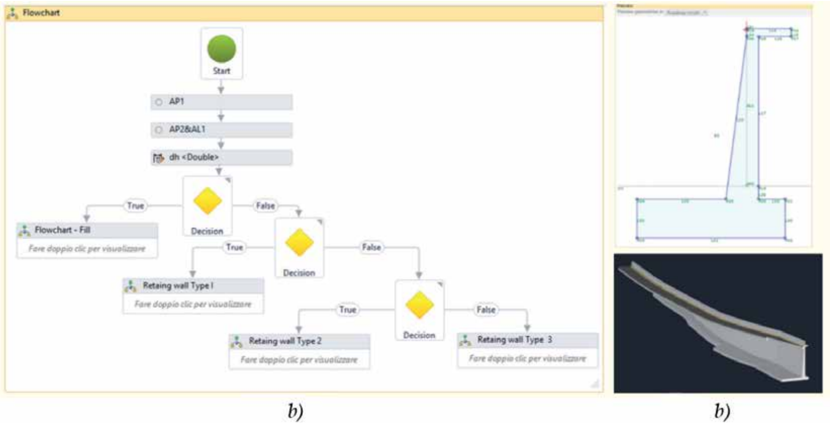


Figure 6.
Modeling retaining walls using subassembly composer. (a) Workflow, (b) result.

Once your assembly is built you need to apply this to your alignment using the corridor function and hey presto, you will have a corridor and basic road design.

3.4 Corridor modeling

Before create corridors, you must have existing data, such as existing ground surfaces, alignments (centerlines), profiles (vertical alignments), and typical sections (assemblies).

All calculations should be finalized before they are applied to the corridor model. Changes in a corridor baseline alignment are not reflected in calculations. Changing the design criteria does not update the corridor model.

In **Figure 7** is shown a generic 3D Corridor model.

3.5 Information management

Once a road network is correctly modeled and parameterized following the above procedure, a number of shared parameters describing the main features of pavement materials can be created to match the information contained in an external database.

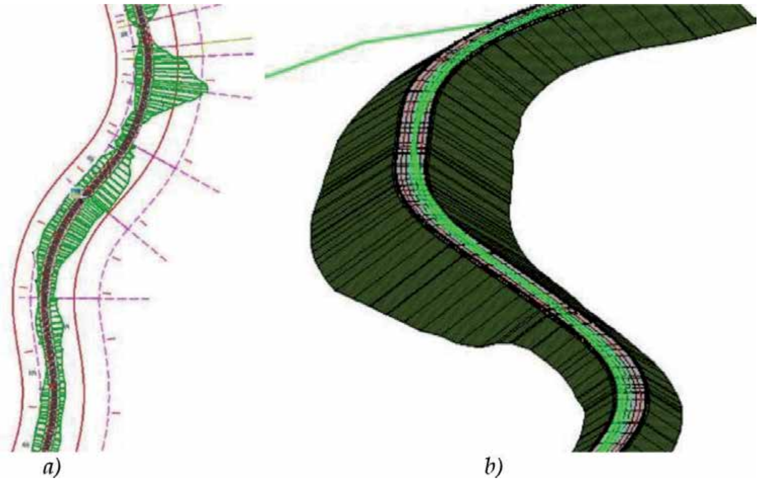


Figure 7.
Modeling corridor. (a) Plan view, (b) 3D view.

There are several visual programming tools (VPL) (i.e. Dynamo) that give to users the possibility to visually script and define custom pieces of logic using various textual programming languages.

The shared parameters imported in the current project as materials features, were the road name, the road administration authority, the year in which the material was layed in place during routine maintenance operations and the physical and mechanical features of the wearing course mixtures, namely bitumen content, air voids percentage calculated with bulk specific gravity determined by means of the dimensional procedure, SSD procedure or sealed specimen procedure and Marshall stability.

Then, the material codes were exported to Excel with the programming flow reported in **Figure 8**, then matched with the materials names in the worksheet and finally imported back into VPL with assigned values.

A worksheet was created using the code block "Data.ExportExcel" (5), whose file path, sheet name and position of the exported data were defined respectively with the code blocks (4A), (4B) and (4C). The worksheet contained a list (4D) of materials identifiers (3A) and names (3B) selected from the list of elements (2) of the materials category (1).

The above mentioned operations allowed visualizing and managing the physical and mechanical features of the wearing course model and updating the information once the input worksheet is integrated with different data. The visualization of the imported data is visible in the material parameters interface, as shown in **Figure 9**.

Then, in the same way, is possible the implementation of a ranking algorithm to evaluate the durability of the wearing course material basing on the material characterization, according to current Regulation [32]. In the specific:

- % bitumen (%B) in the range 4.5 ÷ 6.1% to meet both economic and environmental needs;

- Stability>900 (daN) to respond to mechanical problems;

- % air voids determined by means of the dimensional procedure >3% to improve shear strength;

- The difference between air voids determined by means of the sealed specimen and SSD procedure (Δ) is equal or lower than 1% to ensure that there are no anomalies in the database due to technical errors.

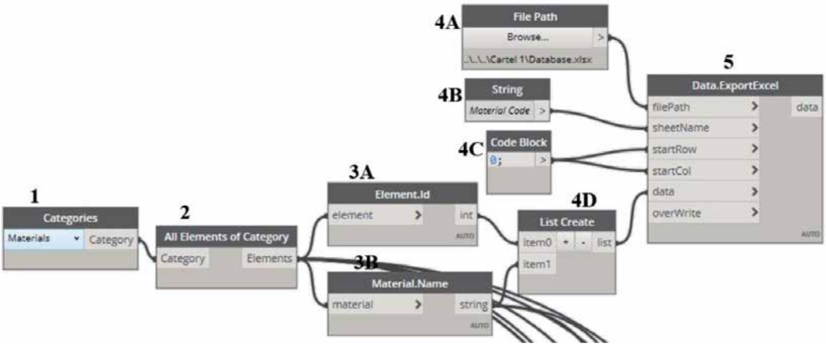
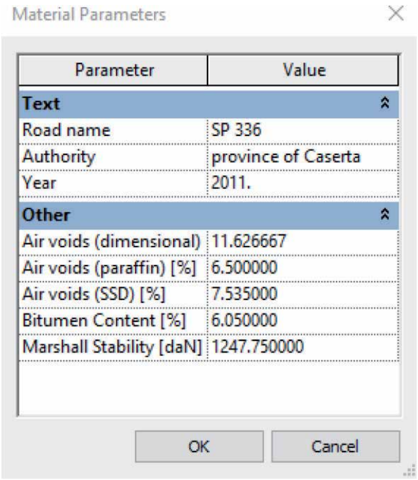


Figure 8.
Materials code creation workflow.



Parameter	Value
Text	
Road name	SP 336
Authority	province of Caserta
Year	2011.
Other	
Air voids (dimensional)	11.626667
Air voids (paraffin) [%]	6.500000
Air voids (SSD) [%]	7.535000
Bitumen Content [%]	6.050000
Marshall Stability [daN]	1247.750000

Figure 9.
Example of material parameters after the association of worksheet data to the model.

In **Figure 10** is shown the workflow for identifying the road pavements that satisfy the first condition.

In the specific: box 1A answers the question if x (%B) is greater than or equal to y (%B lower-limit equal to 4.5%); box 1B answers the question if x (%B) is less than or equal to y (%B upper limit equal to 6.1%); Box 2 “List.Join” concatenates the two lists into one list; Box 3 “List.AllTrue” determines if all the elements of the list are Boolean values with true value; Box 4 “List.Join” merges all the lists associated to other pavement sections of the road network; box 5 “SelectModelElement” for selecting the pavement sections under analysis; box 6 “ListCreate” for merging all the selected pavements in the previous step in a single list; box 7 “*List.FilterByBoolMask*” to filter the list of elements codes by looking up for corresponding indices in the list of Boolean variables, identifying the sections that comply with the technical specifications.

In the same way, the workflow can be adapted to the remaining Regulation conditions, with the possibility then to create combined filters among the mentioned

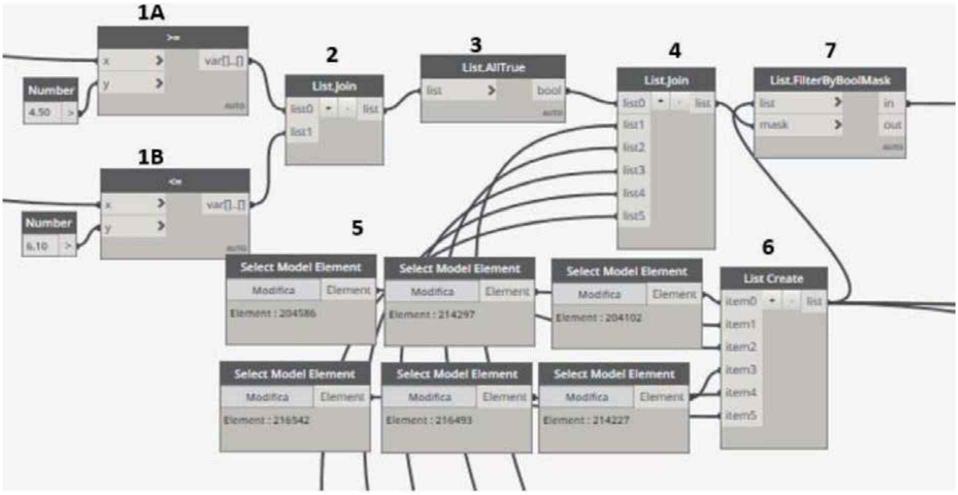


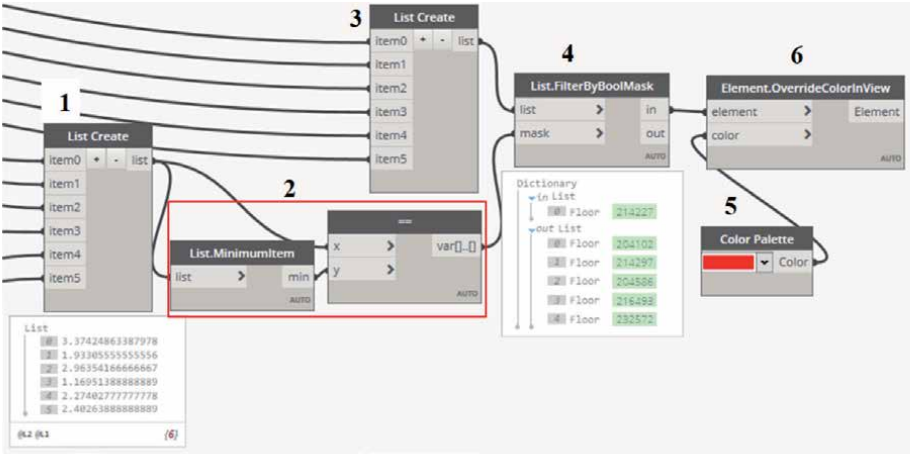
Figure 10.
Workflow for identifying pavements with bitumen content in the range 4.5%–6.1% by the weight of the mixture.

conditions, for visualizing on the road network map, with different color, the pavements with best, worst or intermediate performance.

For example, in **Figure 11**, the list containing the overall scores of the road surfaces under analysis (1) was matched with the list of identification codes of the corresponding elements of the model (3) using again the code block "*List.FilterByBoolMask*" (4). In the present study, the list of Boolean variables was obtained by looking for the road surface with the minimum score (2), obtained from the combination of several physical and mechanical indicators and their upper and lower limit imposed by the Regulation. Lastly, the element code that met condition (2) was emphasized in the model element with the color red (5) by using the code block "*Element.OverrideColorInView*" (6).

As a simplified application to show the impact of information update on the model output modification, two different road sections were considered with bituminous mixtures for wearing course characterized in terms of bitumen content, percentage air voids and Marshall stability.

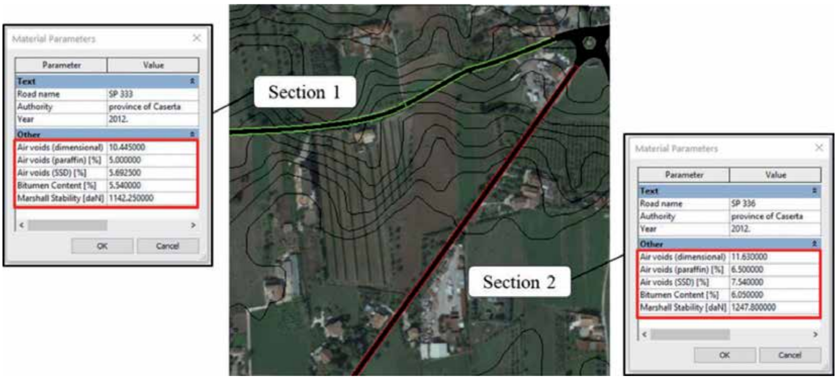
The test results are updated as material parameters in the model and a ranking algorithm is implemented in order to identify the road section with a need for maintenance. As shown in **Figure 12**, the critical road section that requires routine maintenance before the other is highlighted in red.



```
List
  0 3.37424863387978
  1 1.93305555555556
  2 2.96354166666667
  3 1.16951388888889
  4 2.27402777777778
  5 2.40263888888889
@L2 @L1                {6}
```

```
Dictionary
  in List
    0 Floor 214227
  out List
    0 Floor 204102
    1 Floor 214297
    2 Floor 204586
    3 Floor 216493
    4 Floor 232572
```

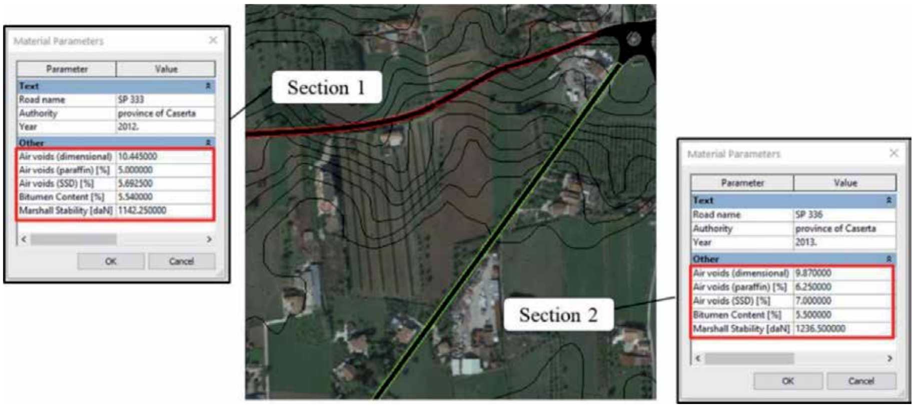
Figure 11.
Workflow for identifying pavements with best/worst performance on the road network.



Parameter	Value
Text	
Road name	SP 333
Authority	province of Caserta
Year	2012.
Other	
Air voids (dimensional)	10.445000
Air voids (paraffin) [%]	5.000000
Air voids (SSD) [%]	5.692500
Bitumen Content [%]	5.540000
Marshall Stability [daN]	1142.250000

Parameter	Value
Text	
Road name	SP 336
Authority	province of Caserta
Year	2012.
Other	
Air voids (dimensional)	11.630000
Air voids (paraffin) [%]	6.500000
Air voids (SSD) [%]	7.540000
Bitumen Content [%]	6.050000
Marshall Stability [daN]	1247.800000

Figure 12.
Example of identification of the critical section.



Parameter	Value
Text	
Road name	SP 333
Authority	province of Caserta
Year	2012.
Other	
Air voids (dimensional)	10.445000
Air voids (paraffin) [%]	5.000000
Air voids (SSD) [%]	5.692500
Bitumen Content [%]	5.540000
Marshall Stability [daN]	1142.250000

Parameter	Value
Text	
Road name	SP 336
Authority	province of Caserta
Year	2013.
Other	
Air voids (dimensional)	9.870000
Air voids (paraffin) [%]	6.250000
Air voids (SSD) [%]	7.000000
Bitumen Content [%]	5.500000
Marshall Stability [daN]	1236.500000

Figure 13.
Example of identification of the critical section after routine maintenance operations.

After routine maintenance operations, quality controls are performed on the wearing course to assess the compliance of the material with the performance required by the Regulation. As a consequence, the model is updated with new information and the raking algorithm immediately finds out the new critical section (**Figure 13**).

4. Conclusions

Computer-aided drafting (CAD) transformed the way professionals created infrastructure designs.

If oriented to the management phase, a BIM model becomes a real simulation, planning and implementation tool for the facility manager, also thanks to the three-dimensional approach related to the parametric objects that populate the environment, representing a valid tool to guarantee control and interoperability of data in an intelligent way.

The implemented methodology allowed creating an integrated model that contains and analyzes data produced by the quality controls of the bituminous materials after laying and compaction.

The applied methodology resulted in a dynamic model that updates its information package and modifies the output of the analysis every time the data worksheet is integrated with new test results.

A tool as such is intended on the one hand to support the prioritization of any existing Pavement Management System that is currently adopted by administrations to plan maintenance operations on the road network and on the other to provide information as an alert system identifying what does not work in maintenance operations.

BIM systems are therefore destined not only to radically change the paradigms of the real estate market but will be able to make a fundamental contribution to the future of the planet, capable of creating IT models from an architectural, urban, environmental and not simply a single infrastructure point of view but of entire inhabited centers: BIM is at the basis of the creation of sustainable cities and Smart Cities. Not only that, with the help of BIM it will be possible to achieve improvements in terms of more sustainable, inclusive and secure cities.

In this way, the proposed framework can also serve as a decision support tool for better planning and management of smart city infrastructure requirements, taking in account as further perspectives other key factors as energy, estimating/cost simulation and mobility analysis.

Conflict of interest

No potential conflict of interest was reported by the authors.

Author details

Salvatore Antonio Biancardo*, Nunzio Viscione, Cristina Oreto
and Francesca Russo
University of Naples Federico II, Napoli, Italy

*Address all correspondence to: salvatoreantonio.biancardo@unina.it

References

[1] Kumar H, Singh MK, Gupta MP. Moving towards smart cities: Solutions that lead to the Smart City Transformation Framework. Technological Forecasting and Social Change. 2018.

[2] Azhar S. Building information Modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry. Leadership and Management in Engineering. 2011; 11:134-146.

[3] Marmo R, Polverino F, Nicolella M, Tibaut A. Building performance and maintenance information model based on IFC schema. Automation in Construction. 2020; 118:103275.

[4] Singh P, Sadhu A. System Identification-Enhanced Visualization Tool for Infrastructure Monitoring and Maintenance. Frontiers in Built Environment. 2020; 6:76.

[5] Ozturk GB. Interoperability in building information modelling for AECO/FM industry. Automation in Construction. 2020; 113:103122.

[6] Al Sayed K, Bew M, Penn A, Palmer D, Broyd T. Modelling dependency networks to inform data structures in BIM and smart cities. In: Proceedings of the 10th International Space Syntax Symposium; 13-17 July 2015; London. United Kingdom; 2015.

[7] Vilgertshofer S, Borrmann A. Using graph rewriting methods for the semi-automatic generation of parametric infrastructure models. Advanced Engineering Informatics. 2017; 33:502-515.

[8] Biancardo SA, Capano A, Guerra De Oliveira S, Tibaut A. Integration of BIM and Procedural Modeling Tools for Road Design. Infrastructures. 2020; 5(37).

[9] McGraw-Hill Construction. The Business Value of BIM for Infrastructure: Addressing America's Infrastructure Challenges with Collaboration and Technology. Smart Market Report. 2012. Available from: http://images.autodesk.com/adsk/files/business_value_of_bim_for_infrastructure_smartmarket_report__2012.pdf

[10] Chong H, Lopez, R, Wang J, Wang, X, Zhao Z. Comparative analysis on the adoption and use of BIM in road infrastructure projects. Journal of Management in Engineering. 2016; 32:1-13.

[11] Sankaran B, O'Brien W, Goodrum P, Khwaja N, Leite F, Johnson J. Civil integrated management for highway infrastructure: Case studies and lessons learned. Transportation Research Record: Journal of the Transportation Research Board. 2016; 2573:10-17.

[12] Abdelwahab HT. Intelligent design (4D, 5D and beyond) for road design and construction projects: Two case studies. International Road Federation (IRF) Examiner. 2018; 12:21-25.

[13] Lee SS, Kim KT, Tanoli WA, Seo JW. Flexible 3D Model Partitioning System for nD-Based BIM Implementation of Alignment-Based Civil Infrastructure. Journal of Management in Engineering. 2020; 36(1): 04019037.

[14] Abbondati F, Biancardo SA, Sicignano G, Guerra de Olivera S, Tibaut A, Dell'Acqua G. BIM parametric modelling of a railway underpass. Ingegneria Ferroviaria. 2020; 6:443-459.

[15] Biancardo SA, Viscione N, Oreto C, Veropalumbo R, Abbondati F. BIM Approach for Modeling Airports Terminal Expansion. Infrastructures. 2020; 5(41).

[16] Abbondati F, Biancardo SA, Palazzo S, Capaldo FS, Viscione N.

I-BIM for Airport Infrastructures. Transportation Research Procedia. 2020; 45:596-603.

[17] Biancardo SA, Viscione N, Cerbone A, Dessì E Jr. BIM-Based Design for Road Infrastructure: A Critical Focus on Modeling Guardrails and Retaining Walls. Infrastructures. 2020; 5(59).

[18] Tschickardt T, Krause D. BIM in highway construction using the example project Availability model A 10/A 24. Bautechnik. 2019; 96(3):259-268.

[19] Bosurgi G, Celauro C, Pellegrino O, Rustica N, Sollazzo G. The BIM (building information modeling)-based approach for road pavement maintenance. Lecture Notes in Civil Engineering. 2020; 48:480-490.

[20] Jing W, Hao G, Li C, Wei W, Cheng J. BIM Application Approach on Highway Maintenance and Management. In: Proceedings of the 19th COTA International Conference of Transportation Professionals; 6-8 July 2019; Nanjing. China; 2016. p. 776-786.

[21] Biancardo SA, Russo F, Veropalumbo R, Vorobjovas V, Dell'Acqua G. Modeling Roman Pavements using Heritage-BIM. The Baltic Journal of Road and Bridge Engineering. 2020; 15(3): 34-46.

[22] Marzouk M, Othman A. Planning utility infrastructure requirements for smart cities using the integration between BIM and GIS. Sustainable Cities and Society. 2020; 57: 102120.

[23] Sankaran B, Nevett G, O'Brien WJ, Goodrum PM, Johnson J. Civil Integrated Management: Empirical study of digital practices in highway project delivery and asset management. Automation in Construction. 2018; 87:84-95.

[24] Tang F, Ma T, Y. Guan, Zhang Z. Parametric modeling and structure verification of asphalt pavement based on BIM-ABAQUS. Automation in Construction. 2020; 111:103066.

[25] Tang F, Ma T, Zhang J, Guan Y, Chen L. Integrating three-dimensional road design and pavement structure analysis based on BIM. Automation in Construction. 2020; 113:103152.

[26] Costin A, Adibfar A, Hu H, Chen SS. Building Information Modeling (BIM) for transportation infrastructure–Literature review, applications, challenges, and recommendations. Automation in Construction. 2018; 94:257-281.

[27] Tam ASB, Price JWH. A maintenance prioritisation approach to maximise return on investment subject to time and budget constraints. Journal of Quality in Maintenance Engineering. 2008; 14(3):272-289.

[28] Delval T, Geffroy B, Rezoug M, Jolibois A, Oliveira F, Carrè S, Trual M, Soula J. BIM to Develop Integrated, Incremental and Multiscale Methods to Assess Comfort and Quality of Public Spaces. Lecture Notes in Civil Engineering. 2021, 98: 160-179.

[29] Kim JI, Kim J, Fischer MA, Orr R. BIM-based decision-support method for master planning of sustainable large-scale developments. Automation in Construction. 2015; 58: 95-108.

[30] Li YW, Cao K. Establishment and application of intelligent city building information modelbased on BP neural network model. Computer Communications. 2020, 153: 382-389

[31] Wang J, Jiang C, Zhang K, Quel TQS. Vehicular sensing networks in a smart city: Principles, technologies and applications. IEEE Wireless Communications. 2018; 25 (1): 122-132.

[32] Azienda Nazionale Autonoma delle Strade (ANAS). Capitolato speciale d'appalto-Norme Tecniche. 2016.

Edited by Stefano de Luca,
Roberta Di Pace and Chiara Fiori

Innovative and smart mobility systems are expected to make transportation systems more sustainable, inclusive, and safe. Because of changing mobility paradigms, transport planning and design require different methodological approaches. Over twelve chapters, this book examines and analyzes Mobility as a Service (MaaS), travel behavior, traffic control, intelligent transportation system design, electric, connected, and automated vehicles, and much more.

Published in London, UK

IntechOpen

ISBN 978-1-83880-824-2