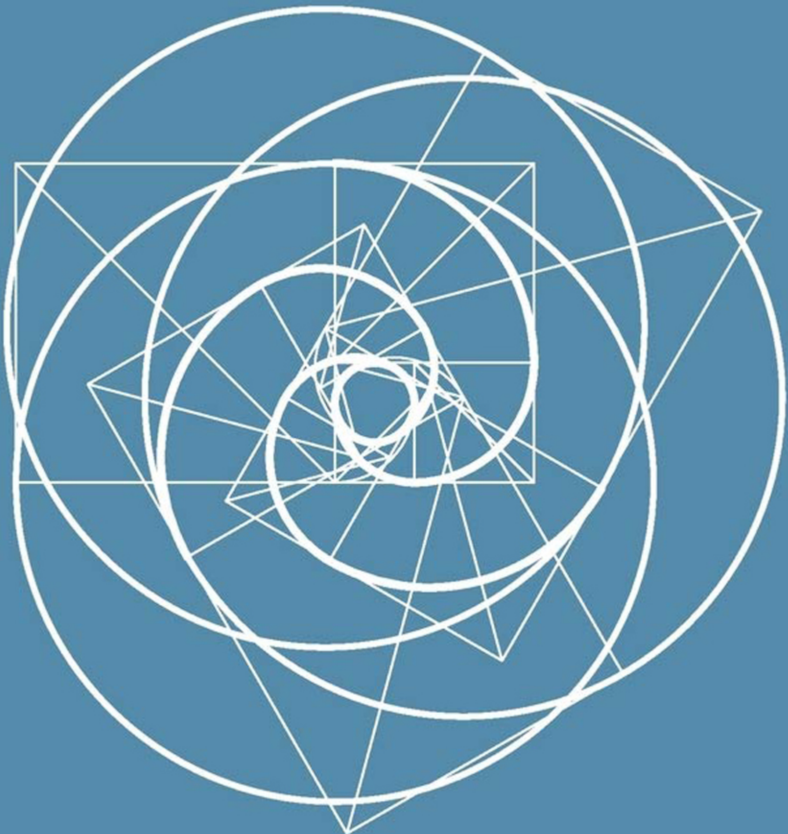
Dynamics and Stochasticity in Transportation Systems



Tools for Transportation Network Modelling

Giulio E. Cantarella
David P. Watling | Stefano de Luca | Roberta Di Pace

ELSEVIER

Dynamics and Stochasticity in Transportation Systems

Dynamics and Stochasticity in Transportation Systems

Tools for Transportation Network Modelling

Giulio Erberto Cantarella

David Paul Watling

Stefano de Luca

Roberta Di Pace

ELSEVIER

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

© 2020 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

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

ISBN: 978-0-12-814353-7

For information on all Elsevier publications
visit our website at https://www.elsevier.com/books-and-journals

Publisher: Joe Hayton
Acquisition Editor: Brian Romer
Editorial Project Manager: Aleksandra Packowska
Production Project Manager: Nirmala Arumugam
Cover Designer: Christian J. Bilbow

Typeset by SPi Global, India

ELSEVIER Book Aid International
Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org

Contributors

Giulio Erberto Cantarella
University of Salerno, Salerno, Italy

Roberta Di Pace
University of Salerno, Salerno, Italy

Stefano de Luca
University of Salerno, Salerno, Italy

David Paul Watling
Institute for Transport Studies, University of Leeds, Leeds, United Kingdom

Preface

Dissidence and controversy are what bring science forward.
Agreement and acceptance rarely stimulate experiments and progress.
Thor Heyerdahl.

Outline. In this chapter the reader may find the contents and the purpose of this book, and a framework general enough to encompass almost all TTT tools. Beside the reader may find a proposal of nomenclature. Main emphasis in this book is on the mathematical features of models and algorithms for travel demand assignment to a transportation network. Implementation and application issues will be the topics of a future companion book (possibly by other authors) as well as control and design tools.

Since hunting gathering era human brains have evolved to be more sensitive to variations in space and/or time of the surrounding environment rather than regularity and uniformity; (mostly unconscious) representations of location over space and evolution over time allowed human beings to survive in challenging conditions. This is still the case: a pedestrian wishing to cross a urban street tries to anticipate evolution over time of the locations of the surrounding vehicles.

Developing a (mathematical) model of real systems, as common in modern applied sciences, is a more conscious way to follow that ancestral attitude.

Even though future were perfectly determined by past, according to Beowulf's well known statement "Fate will unwind as it must!" (but not to authors' opinion), still it may not be perfectly forecasted due to lack of enough information about past, to uncertainty affecting forecasting methods, ... Thus, however desirable, in several cases a precise model providing deterministic description and forecasting of system state cannot be developed, and the most general modelling tools include both dynamic and stochastic features together with space characterisation.

It should be remarked that any kind of representation or model mentioned above is, as beauty, in the mind of the beholder; therefore dynamics or stochasticity are features of (mathematical) models only, a sort of social constructions agreed by the modeller community, not be confused with the object of their applications, such variations in space and/or time in a real system. Along this line of reasoning observations of real world are facts, whilst models are opinions about them.

The focus of this book is the use of mathematical modelling methods to assist in the understanding, prediction, policy assessment and design of transportation systems; but what is a "transportation system", or most pertinently what do we mean by it for the purposes of this book? Firstly, it contains the infrastructure, the pavements to walk on, the roads on which we cycle, drive or may use a bus or taxi, the train tracks, as well as the fleets of buses, airplanes and trains that are used to run services and transport goods. Secondly, it contains the users of the transportation system, namely the people who choose where, when and how to travel, as well as the goods operators and suppliers who decide how and where to transport their goods. Thirdly, it contains the various public and private organisations responsible for planning, operating, pricing and providing information on the infrastructure.

Such a “system” contains many interacting elements. A traveller may decide to drive to their normal place of work during a more busy (congested) period of the day than they would normally, and by doing so contributes additionally to the congestion for that day. This congestion may delay other road users who are using some of the same roads, but perhaps travelling between an entirely different origin and destination to the first traveller. The additional delay experienced may, on the other hand, hold back the second traveller so that traffic is in fact more freely running that it might be on some downstream stretch of road on their intended route. This may cause the responsive traffic signals at an intersection to trigger at a different time, and so influence some other travellers. On the other hand, our second traveller has such a bad experience of travelling that day that they decides to try a different route when they make that trip next time, whereas the first traveller decides that re-adjusting their departure time would be wise in the future. As this happening, a private transport operator decides to introduce a new high-speed train service in the area, which our first traveller then decides to use on some subsequent day, thus alleviating some of the pressure on road capacity. At the same time as all these interactions are on-going, each minute of every day, a transport planner is deciding on how to make adjustments to achieve some policy objective, and as a result introduces on some subsequent day a new high-occupancy vehicle lane for certain hours of the day.

This is only one example. Making sense of such systems is no simple task. Mathematical models are a tool for capturing at least some of this complexity, assisting those who are responsible for planning and designing such systems for the ‘public good’. They allow location-specific data to be systematically used to calibrate a model to particular *locale*. Importantly, the models that will interest us in this book have a clear forecasting capability, allowing the modeller to study “what-if” scenarios. These scenarios may range, for example, from studying the impact of potential future changes to travel demand patterns, the effect of alternative control measures, or the impact of policy measures on travellers’ experience and network performance.

These models are, in some respect, always wrong, as always the case in all applied sciences. They always omit or over-simplify or incorrectly capture some phenomenon, and so when using them it is always important to be as clear as possible on the limitations of the model, what it is not able to capture as well as what it is representing. The modeller has the responsibility to communicate this to the decision-maker, however much the decision-maker may not like to hear it. To do this effectively, we believe we need a clear, standard reference system’ for dynamic transportation analysis, in order that particular models and model assumptions fit into such a unified framework. Developing such a consistent and systematic treatment is a major goal of this book, and therefore we take some time and care in defining and explaining our terminology, assumptions and concepts, and how these relate to both the real-world and to common methods used for modelling dynamic transportation systems.

1 Purpose of this book

Travel and transportation play a central role in the lives of most of the world's population. Transportation provides both a means of trade in moving goods, and a way of moving people to engage in employment, education, social and other activities. If we had observed the same geographical area over a period of past decades, we would likely have seen that the size and structure of employment, production and residential areas had changed over time, and these changes had in turn changed the requirements and pressures on the transportation system. At the same time, these changes will have made environmental, social and economic impacts, some positive and some negative, with some winners and some losers. It is natural then to ask whether we can hold a mirror to the past, and use it to see into the future; at least then we may be able to react in a better way to the inevitable changes the mirror shows us, and thereby as a society expend resources more efficiently (in the sense of less negative and more positive impacts). It is only one more step to then realise that the mirror analogy is limited, that unless we believe the future is pre-determined we may influence it by our actions, both as individuals and as organisations. Understanding such influences and their likely consequences then provides a way of not only 'managing' a transportation system more effectively, but also positively engineering it to improve the lives of the people using it. The current book fits into this wide area of 'transportation planning', and particularly the field of transportation modelling which aims to postulate mathematical systems that broadly approximate the changes and processes underlying such phenomena.

Space and time are two intrinsically important aspects to understanding travellers' needs and what transportation systems can supply. Let us firstly consider space. The type and density of activities are not distributed evenly across a city or region, and there are fixed geographical features (rivers, mountains, valleys, etc.) that influence the feasibility of different transportation options across an area. Dense, 'vertical' residential areas provide very different challenges to more sparsely distributed ones. There also complex interactions that play out in the transport infrastructure; a congested road or overcrowded bus may be partially the result of travellers avoiding overloaded facilities, meaning that a good solution will not be understood without considering system-level interactions between the various travel needs of people/organisations and the services and facilities which are provided. Over the last 50 years the transportation community has developed rather sophisticated ways of representing these kinds of spatial interactions, typically by representing the infrastructure as a network (mathematical 'graph'), and by considering various levels of sophistication in representing the behavioural responses of travellers (e.g. from the perfectly informed traveller to random utility approaches). At the same time, however, it should be mentioned that while the individual fields have developed to a high level, it is relatively rare to find a consistent integration of demand modelling and network modelling.

There are rather well developed (if not always consistent) methods, then, for considering 'space'; so what about 'time'? While travel time, as a disincentive in making travel choices, is a central aspect of transportation planning, by the word 'time' we are instead referring here to changes that occur over time ("dynamics"). As there is considerable potential for confusion, let us very early on make a clear distinction: changes on a 'within-day' time-scale are the kind of changes that we would expect to see as we made a journey on a particular day, or if we compared our travel experience with someone travelling by the same route/service but at a different time on that day (there are many other ways to characterise this kind of time, but these examples suffice for now). On the other hand, changes on a 'between-day' time-scale concern, for example, the way in which we might adapt our travel choice next time we make a journey, based on our travel experiences today. While researchers have been aware of both 'within-day' and 'between-day' effects for several decades, it is only relatively recently that a concerted effort has been made to develop tools and methods to explicitly model them. On the within-day scale, this has been achieved by introducing and adapting methods from traffic flow theory for use in network models. On the between-day scale, it has involved bringing in new techniques from both applied mathematics (for deterministic dynamical systems) and probability theory (for stochastic processes).

The new student to this field is faced with an extremely challenging task to digest and assimilate these developments, with their various assumptions, approaches and need to draw on different aspects of mathematics. In fact, it is also a challenge for the experienced researcher to keep fully abreast of developments in this field, as without detailed reading it is often unclear where a new paper fits within the literature, and how it complements or advances previous developments. The intention of this book is to bring together the theoretical work in this field in an internally consistent way. This means unifying transportation network modelling theory, travel choice theory and traffic flow theory, while drawing on relevant mathematical concepts from operations research, matrix algebra, dynamical systems, statistics and simulation. Our intention is that the book is self-contained, so does not require additional reading, although we suggest further reading for those interested. In order to assist us in this goal of being self-contained, we supply a mathematical supplement which outlines the main mathematical ideas that we will draw on.

Apart from the pedagogical aim, to assist newcomers to the field, it might be asked: why do we wish to set out a consistent theoretical approach to such problems? What advantage does it give? One answer might that it is a purely aesthetic objective, that having a single unifying theoretical frame is satisfying purely in its own right.

Certainly we would admit that personally we do find satisfaction in this aesthetic quality, however we feel that it has much greater significance than this, for several reasons. Firstly, as we shall show, it provides a way of using theory to assess what we might expect from applying such a modelling approach to any case-study, for example in terms of uniqueness, stability or reproducibility of its predictions. These have important implications for project evaluation, and may not be self-evident from computational application alone, especially given the complex nature and interactions involved in such models. Secondly, our book is an attempt to integrate the state-of-the-art in a way that we hope generates a kind of feedback effect with the academic community, whereby our work provokes a reaction either in terms of disagreements with our method of integration, or the stimulus for new research directions, which in turn will impact on the future state-of-the-art. Thirdly, we intend that our approach leads to a kind of taxonomy of models that facilitate easier communication of assumptions, by reference to a common modelling frame. This is important in order that forecasting is not seen as a kind of supernatural divination, akin to 'reading' tarot cards, but so that uncertainty over assumptions made and trends assumed can be properly communicated to decision-makers.

Given the task to bring together several theoretical fields, it has been necessary to keep the focus of the book firmly on the development of forecasting methods. In order that the book be of a manageable size, we therefore do not consider two activities that might be said to 'bookend' the task of forecasting, namely calibration and design. We should mention, however, that the approaches described are, we believe, especially suited to these two activities. In the case of calibration, as we write this book we are in a period of unprecedented data availability, ranging from various kinds of 'tracking' data (from mobile phone activity, GPS or Bluetooth, for example), to archived detector data over long periods, and other means of inferring activity patterns (e.g. shopping transaction data, video surveillance). Such data reveals a complex, time-of-day-varying, daily-varying, heterogeneous pattern of travel behaviour that is not only poorly captured by steady-state approaches, but is actually rather difficult to 'project' onto the conventional ideologies. Moreover the approaches presented that are developed from probability theory provide a means of calibrating such models using formal statistical inference. In terms of design, the process approach we develop allows us to examine how likely it is that the system could be influenced to attain and operate under very different transportation system designs to the present. The traditional, 'comparative statics' approach, on the other hand, only allows us to assess alternative designs on the premise they are attainable from the present.

Even given our decision to focus on the theoretical development of forecasting methods, space limitations mean that it has still been necessary to be selective in terms of the kinds of problem considered. This means we exclude many important areas, such as the routing of freight vehicles, the use of taxis (or taxi-like vehicles), and the behaviour of pedestrians in continuous space, such as a square. Given its rather long association with dynamic processes, and the practical use of disequilibrium forecasts, it would have been interesting to have included a consideration of land-use and transport interaction models. Methodologically, it would have been very relevant to include a consideration of tour-based and activity-based approaches, given their natural association with scheduling during the day, and a more thorough study of microscopic approaches, given their natural relation with stochastic process models. Perhaps the most natural extension to our work would have been to include mode choice (integrated with route/service and time-of-day choice) in a fully multi-modal modelling approach. The list could continue, and we would have sufficient to write several books. No doubt we will irritate many people by the problems we decided to include and exclude, but in the end we satisfied ourselves that our choice of what to include was a kind of metaphor for how a modeller in practice must approach a real-life study: ultimately, not all processes may be represented in a chosen modeller, and so the modeller must always accept (often unwillingly) that there are potentially important aspects that are outside the remit of the model chosen. If the task of a modeller is to provide the tools for a modelling tool-box, then we hope that we can be considered to have provided a good selection of screwdrivers and chisels, and hope that others in the future can add to this box with their own selection of saws and hammers.

Looking to the future, modelling faces unprecedented challenges and opportunities. On the one hand, it must rise to the challenge of representing 'new' forms of transportation, such as autonomous and/or connected vehicles, electric bicycles and shared mobility services, and the seemingly ever more complex ways that lives are organised; for example, if we plan a business meeting in an autonomous vehicle in such a way that the trip ends at a social meeting point at some prescribed time (greater than the minimum time to reach that point), what do we consider the purpose and destination of the trip to be, and what value-of-time and routing principle may be consider operating? How might we model populations of individuals being transported around a city in coordinated, shared transportation?

These challenges are balanced by new opportunities in terms of the richness and extent of the data we may expect. Some have argued that such data may be of such extent that models are no longer required, since we might expect to have almost complete observation of travel movements and related phenomena. Actually, we believe the contrary to be true; this is exactly when models are needed most. The quantity of data from such sources is potentially overwhelming and diverse, and models may then be used to extract key patterns from this morass of information. Furthermore, even with perfect observation of the past, we return to our original question of whether we will simply use it as a mirror for the future. Without forecasting tools we will be unable to consider how the future might change and how we might influence it in a beneficial direction. In an era of "big data", we need even more than before simplified, abstracted models that are able to provide an indication of the likely future trajectory of our transportation system, and the influence that individuals, organisations and policy-makers may have on such a trajectory.

2 Contribution of this book

Most branches of engineering were founded on physics (and/or chemistry) developed from late 19th century, and now are well-established. This actually is the traditional image to common folks of an engineer: a person able to solve practical problems that are well rooted on a specific background, for instance electronics, hydraulics, etc., through specialised mathematical tools. Good examples within transportation engineering are analysis and design of components such as vehicles (and their engines), facilities, etc., and traffic engineering, developed by applying a metaphor derived from fluid dynamics, which deals with the behaviour of several vehicles sharing the same facility and the design of traffic control devices, such as traffic lights, ATC, etc.,.

At the beginning of 50s of the last century a new paradigm was introduced, by linking together the contributions of several authors, leading to the (abstract) systems engineering, where emphasis is on mathematical representation of a problem rather than its physical background. This is a new type of engineer: a person able to solve a practical problem considering it as a whole through an ever increasing box of non specialised mathematical tools.

For what may now be considered a charming synchronicity, John G. Wardrop, in his seminal presentation held on 24 January 1952, and published in June (Wardrop, 1952), founded transportation systems engineering, including both analysis and design. In his paper, he proposed a wide and comprehensive review of traffic engineering, but at the same time, he understood that traffic engineering techniques can be used only to analyse the performance of a single component (cfr p. 344). He also stated that a transportation system cannot be studied on a single element basis, but as a whole system indeed (note that he actually used the word “system”).

Hence, starting from a small example, he proposed his now widely known two principles to model travel demand distribution over alternative routes in a transportation networks (cfr p. 345). Then, he stated that these two criteria must be extended to deal with a whole network, where route are broken into links possibly shared by other routes, even though (cfr p. 348) in the case of a network of roads the theoretical problem becomes very complicated. This way, John G. Wardrop introduced the main elements of any effective model of a transportation system:

- a user behaviour model, which simulate how level-of-service, say journey times, affects user choices, as expressed in his paper by the two criteria (travel demand);
- a performance model, which simulate how user choices, say flows, affect level of service, say journey times; it is made up by a network model representing topological features and, at a link level, by performance – flow relations derived by applying traffic engineering techniques (transportation supply).

Besides, he greatly stressed *The Value of a Theoretical Approach* (cfr p. 326) as the only effective one, thus stating that models within a specific theory should be developed. These models, now referred to as travel demand assignment to transportation networks, are the basic tools to simulate a transportation system. It is also worth noting that, pointing out that the user behaviour is likely selfish and does not lead towards the most efficient pattern, he stated the need of supply network design.

From the seed planted by J.G. Wardrop the still growing tree of the modern Traffic and Transportation Theory (TTT) emerged. A general overview of existing problems and tools of TTT is given below in order to point out the contribution of this book. TTT studies the interactions between the level of service provided by transportation systems and the results of several types of user choice behaviour, which may regard in a hierarchical order from bottom to top:

- driving, concerning interactions between users travelling on the same facility and their effects on travel time, …;
- routing, concerning connections between origin and destination of the journey, parking location and type, possibly departing time, …;
- travelling, concerning transportation mode, time-of-day, destination, frequency, …;
- mobility, concerning car ownership, driving licence acquisition, … .

On top of the above hierarchy there are the kinds of user behaviour addressed by land-use/transport interaction theories.

Tools of TTT have reached a very advanced and sophisticated level, and large-scale applications are a current practice. Most of these tools are based on explicitly behavioural modelling approaches, which grant clear interpretation of parameters. A taxonomy is given in Table 1 below where for brevity's sake *kinds of choice behaviour others than routing and driving have not be explicitly considered*. A brief review of these tools is given in the four sections below to introduce the nomenclature used in this book and the contents of the chapters of this book.

(in round parenthesis the number of the corresponding section below.)

2.1 Traffic analysis

This section briefly discusses methods for traffic analysis, which addresses the effects of driving choice behaviour, and are usually derived from Traffic Theory, described in details in Appendix B, also discussing Queuing Theory for bottlenecks.

Table 1 Tools of TTT.

Modelled behaviour	**Problems methods**	
	Analysis descriptive—predictive	**Decision prescriptive**
Driving	Traffic analysis (2.1)	Traffic control (2.3)
Driving and routing	Transportation systems analysis (2.2)	Transportation systems control and design (2.4)

Under steady-state conditions (introduced in Chapter 1) the most commonly used model to describe vehicles flowing along a street (railway, airway, …) is the so-called fundamental diagram (FD) describing the relations among density, flow and (space average) speed. In particular, in the stable regime speed is a decreasing function of flow, that can be used to specify travel time functions.

When steady state conditions do not hold, within-day dynamics (introduced in Chapter 1) in a link should explicitly be taken into account through three kinds of macroscopic models, described in details in Appendix B:

- continuous in space and time;
- discrete in space and continuous in time;
- discrete in space and time.

The full specification of all the above models requires an equation describing the relation between speed and flow or between speed and density, to be derived from the FD, as well some network equations to lead to within-day dynamic assignment models.

In appendix B mesoscopic and microscopic modelling approaches will also be described.

2.2 Transportation systems analysis

This section briefly discusses methods for transportation systems analysis, which can be distinguished into methods for:

- travel demand analysis,
- transportation supply analysis,
- demand–supply interaction analysis, or assignment.

Before applying any of the above methods some preliminary steps should be carried out. The study area is delimited and divided into zones, where a journey starts or ends, and main infrastructures and services are singled out to support journeys between any pair of them. Then, users are distinguished, following a 5 W approach, with respect to.

WHO: socio-economic characteristics and grouped into categories (or into commodities for freight),
WHY: trip purpose,
WHAT: trip frequency,
WHERE: trip origin and destination [for simplicity's sake we will assume that each journey is defined by a single trip, thus trip-chains or tours are not considered],
WHEN: time of day (morning vs. afternoon peak period, day of week vs. weekend days, winter vs. summer, special events, usual vs. emergency conditions, …).

Once trip origins and destinations have been singled out, itineraries between each pair of origin and destination can be defined, possibly distinguished by category, purpose, …. Then, each itinerary can be broken down into links, each link being a stretch of street, railway, airway, …, with common characteristics. In the most general case an itinerary is a routing strategy including both pre-trip and en-route choices depending on information available to users.

2.2.1 Travel demand analysis

During each time-of-day, users belonging to different user categories, with different trip purposes, and journeying between different origin–destination pairs interact each other and compete to access same infrastructures and services, thus for easy reference any combinations of user category, trip purpose, origin–destination pair is called a user class (u.c.). Given a time-of-day, for each user class the travel demand flow defines.

HOW MANY users are moving in the study area.

Travel demand flows can be estimated through statistic methods giving the so-called origin–destination travel demand flow matrices (or o-d matrices for short). They are assumed constant in the following, since, as already stated above, *kinds of choice behaviour others than routing and driving are not explicitly considered* in this book. Thus, the travel demand model describes:

HOW users reach their destinations from their origins.

Travelling choice behaviour can be modelled through any discrete choice analysis theory (e.g. Random Utility, Fuzzy Utility, Prospect, … theory), as described in Appendix A. Under within-day dynamic conditions, user departure time choice behaviour should also be taken into account through pre-fixed proportions or through a further choice model, where the utility function includes penalty for early or late arrival with respect to desired arrival time.

2.2.2 Transportation supply analysis

Methods for transportation supply analysis combine methods for traffic flow analysis, with methods derived from the Theory of Congested Networks, including synchronic and diachronic networks, described in details in Chapter 1. In this book tools for modelling *continuous supply,* e.g. *pedestrians moving in a square, are not considered.*

The connections between trip origins and destinations are described by an oriented graph, such that each link is described by an oriented arc between two nodes and each itinerary is described by a route, with nodes modelling junctions. Moreover, each origin and each destination is modelled through a further node connected to the main network through connecting arcs (or connectors) not corresponding to links (see Section 1.2).

Under steady-state conditions a flow and a transportation cost are associated to each arc; usually cost is a combination of several attributes regarding time (on-board, waiting, delay due to junctions, …), money (fuel cost, tolls, fees, …), and reliability (dispersion indices of travel time, …

The transportation supply model describes:

HOW MUCH it costs to users reach their destination from their origins.

Within-day dynamics greatly affects the transportation supply model, since arc flows and costs depend on time, moreover the flow of an arc also depends on the position within the arc. The flow entering an arc at a given time depends on travel time to reach the arc, generally through different paths, the travel time of each of these paths depends on the travel time of each arcs previously traversed, which in turn depend on the flow that has traversed them. Therefore, within-day dynamic models require that HOW LONG it takes to users to reach their destination from their origins is explicitly modelled (if travel time is different from transportation cost).

2.2.3 Assignment

Traditional equilibrium assignment searches for mutually consistent arc flows and costs. It was first introduced under steady-state conditions by Wardrop (1952), who named it User Equilibrium (UE).

Equilibrium assignment may effectively be formulated through fixed-point models which can be easily extended to deal with several types of assignment Fixed-point models can easily be specified combining together the three equations of the transportation supply model and the three equations of the travel demand model, as described in Chapter 3, starting from assignment to uncongested networks described in Chapter 2.

Equilibrium assignment may be regarded as a special case of day-to-day dynamics (see Chapter 1) that mainly affects the travel demand model, in this case indeed both the utility function and the choice function are specified through recursive equations. This approach allows to explicitly modelling evolution over time through deterministic and stochastic process models, as described in Chapters 4 and 5.

Methods for within-day dynamic equilibrium assignment (traditionally called Dynamic Traffic Assignment) follow the same structure of steady-state equilibrium assignment. Still user departure time choice behaviour should also be taken into account; moreover, the flow and cost consistency relations are non-linear as discussed above; therefore, several results available for steady-state equilibrium no longer apply. Moreover, each of the three kinds of macroscopic models described in Section 2.1 leads to different assignment methods that can hardly be put under a common framework to highlight their mutual relationships, let alone mesoscopic and microscopic models. Within-day dynamics may be combined with day-to-day dynamics leading to so-called doubly dynamic assignment. These topics are described in Chapter 6.

2.3 Traffic control

Most methods for traffic flow control are based on the description and prediction of traffic flows without any modelling of routing choice behaviour.

They include fixed sign strategies, such as priority junctions or roundabouts designed off-line, as well variable sign strategies, such as ramp metering or traffic lights. The latter may be applied off-line to support transportation planning or in simple cases when there is no need of adaptive control or on-line to support real-time traffic management.

On-line applications requires sensors for flow monitoring and within-day dynamic models for flow prediction, such as those described in Section 2.1, or simple data-driven methods, such as methods from time series analysis, Bayesian networks, Artificial Neural Networks (ANN).

All these topics are out of the scope of this book.

2.4 Transportation systems control and design

Methods for transportation systems control and design provide optimal features of transportation interventions taking into account their effects on travel demand, say on user route (and departure time, if the case) choice behaviour (quite often under the assumption of constant demand flows). This is usually achieved by considering any model for equilibrium assignment as a constraint embedded within the whole optimization model underling the design method.

Methods for transportation systems or better transportation supply control and design may be grouped into:

- Methods for transportation systems control or transportation network capacity design, or for Continuous Network Design, such as network signal setting design with equilibrium constraints, bus frequency design, …;
- Methods for transportation network topology design, or for Discrete Network Design, such as Urban Network Design, including both signal setting design and lane allocation with equilibrium constraints [lane allocation cannot consistently be carried out without including signal setting design too], design of bus lines, … .

On-line transportation systems control methods can also be combined with Intelligent Transportation Systems (ITS) such as Advanced Transportation Information Systems (ATIS), which may provide information or indications. In this case any information or indication provided to users should be consistent with the control strategy and the user reaction leading to closed-loop systems; satisfactory models of such systems are still an open issue. Currently no method is available to design all features of a transportation system, possibly including a ITS.

According to an arguable but long since established tradition, measures regarding restricted area access policy, parking policy, tolls, congestion charge, and the like, are collectively known as Travel Demand Management (TDM) measures. Needless to say travel demand resulting from the free choices made by users cannot be centrally managed, but only be described. These kinds of interventions should be considered part of transportation supply design within a consistent plan also including kinds of interventions mentioned above.

All these topics are out of the scope of this book.

3 Scope of this book

On deciding on the scope of the book, we decided to make several restrictions, omitting explicit treatment of several important topics, on the grounds that we wish the content of the book to be of a manageable size for the reader to appreciate. This is not, however, intended to suggest a boundary of what is possible with the presented framework; on the contrary, we hope that our work encourages other to cast future modelling advances within the presented conceptual frame. While there are many limitations of our work, which we discuss as the detailed treatment unfolds, there are several aspects that deserve particular mention at the outset. Firstly, we will consider only journeys that occur from a single origin to a single destination. This excludes tours, for example a home-shop-work journey with multiple destinations, as well as good deliveries to multiple locations. Secondly, we consider only networks consisting of a single mode, and make no explicit treatment of inter-modality (e.g. park-and-ride, or a goods delivery combining truck with last-mile delivery by bicycle). Both of these exclusions are significant, but we have made this choice in order to keep the material more manageable, and in any case the study of such issues within a fully dynamic transportation systems is to our knowledge an open research question.

Through Chapters 1–4 main emphasis is on macroscopic modelling, then microscopic methods are discussed in Chapter 5 on time-driven Stochastic Process models for Day-to-day Dynamic assignment, and in Chapter 6 where macroscopic, mesoscopic and (event-driven) microscopic modelling approaches to Within-Day Dynamics are discussed. Data-driven methods are not discussed for brevity's sake; anyhow they seem more useful to solve specific problems, such short-term traffic forecasting, incident-detection, …, rather than providing a wide insight of traffic and transportation systems.

4 Summary

4.1 Major findings

Travel and transportation play a central role in the lives of most of the world's population. Thus travel demand and transportation supply as well as their interaction should carefully be analysed to effectively assess the effects of transportation planning policies and traffic control strategies. This book deals with travel demand assignment to a transportation network, the main tool for transportation system analysis and planning. This preface describes the purpose and the contribution of this book with respect to the current literature about the tools of Traffic and Transportation Theory.

4.2 Further readings

A wide description of most of the tools of TTT may be found in the books by Cascetta (2001, 2009); assignment tools are also described, but without a complete and comprehensive analysis of mathematical features, on the other hand implementation is also discussed. Another widely used book is by de Dios Ortuzar and Willumsen (2011), where implementation and application issues are discussed with great details.

References

Cascetta, E., 2001. Transportation Systems Engineering: Theory and Methods. Kluwer Academic Publisher, p. 2001.

Cascetta, E., 2009. Transportation Systems Analysis: Models and Applications. Springer.

de Dios Ortuzar, J., Willumsen, L.G., 2011. Modelling Transport, fourth ed. Wiley.

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proc. Inst. Civ. Eng. 2 (1), 325–378.

Acknowledgements

Besides the colleagues already mentioned above, I wish to thank the colleagues and friends, too many to be enlisted here, as well as all my students whom I discussed with the topics of this book over the years. I also wish to thank Fulvio Simonelli who carefully reviewed Chapters 1–3. I wish to thanks my university as well for the financial support through several research projects over the years. A special thank goes to Aleksandra Packowska who gently but firmly kept pushing and pulling me until this book was complete, to Nirmala Arumugam, who gently tried to meet all my special requirements about styles, and with them to all the Elsevier's staff who supported me.

This book is dedicate to Rose (whom I married many years ago) who brought and keep bringing arts and beauty in my life, even when I was busy writing this book.

CHAPTER

Introduction

1

Giulio Erberto Cantarella
University of Salerno, Salerno, Italy

The knowledge of first principles, as space, time, motion, number,
is as sure as any of those which we get from reasoning.
And reason must trust these intuitions of the heart,
and must base on them every argument.
Blaise Pascal, Pensées (translated by W. F. Trotter).

Outline. In this chapter the reader may find the basic definitions and assumptions used to develop the models presented in the next chapters. This is consistent with the methodology common to applied sciences whose first steps are a preliminary analysis aimed at providing a simplified description of the system under study and a statement of assumptions about space and time modelling possibly including stochastic approaches.

A model for transportation system analysis tries to describe the state of a transportation system as resulting from the interaction between travellers willing to travel and transportation supply systems providing opportunities to them. In the this section and the following ones we outline the methodology, common to all modern applied sciences such as engineering and economics, applied in this book and the main assumption supporting it. Reader of this book is assumed familiar with fundamentals of Calculus, of Theory of Probability, and of Inferential Statistics.

Through all this book we keep clearly separated features of the real world, say (data from) observations, and those of the modelling tools trying to use different words as far as we can; for instance location over space and evolution over time come from observations of the real word, while spatial and dynamic are adjectives only used to denote models trying to describe these phenomena. Similar considerations hold with regards to observed variations or fluctuations over space or time in real world and to adjectives such as random or stochastic only used to denote models trying to describe these variations as well as other sources of uncertainty. The following examples may help understanding this point and to introduce some basic elements of modelling theory.

Suppose that you are walking a path in a wood looking for seeds of horse-chestnut trees to play conkers with a friend you expect will pay you a visit in the near future (or

Dynamics and Stochasticity in Transportation Systems. https://doi.org/10.1016/B978-0-12-814353-7.00001-7
© 2020 Elsevier Inc. All rights reserved.

for any other reason). You may observe the number of seeds at the foot of each tree along the path in a given day, location over space, and/or at the foot of a tree in consecutive days, evolution over time. The sample of observations can be collected in an array of variables, such that the index describes the tree (space) or the day (time).

If the number of observations is too large to be easily remembered you may compute the observed mean to have a roughly idea of the number seeds you may find together with the observed variance to describe how dispersed the observations are around the observed mean. Both these values, as well as others *indicators from descriptive statistics*, are to be considered as observations since they can be computed from them through (simple algebraic) equations. [This aggregation procedure introduces a source of uncertainty somehow different from the one discussed below.]

A more sophisticated approach is based on *probabilistic models* based on random variables, such that the model mean and the model variance try to reproduce the observed mean or variance, and more generally the probability (density) function tries to mimic the observed frequency distribution; an example are models based on the Poisson random variable, popular models for both location over space and evolution over time. These models are not considered proper spatial or dynamic models since they lack an explicit description of location over space or of evolution over time.

Spatial or dynamic models are those used to explicitly forecast the location over space (in a future day) or of the evolution over time (in a given location). Models including random variables are used when forecasting is affected by uncertainty due to the lack of enough information and possibly other sources of uncertainty, see below so that effective deterministic models cannot be specified; according to these considerations dynamic models are named *deterministic* or *stochastic processes*. [This source of uncertainty due to lack of enough information is somehow different from the one due to the aggregation procedure discussed above.]

After its specification, any kind of model should be calibrated against observed data, say its parameters should be estimated through *inference statistics methods*, before a practical application is possible. This issue is out of the scope of this book.

The modelling approach discussed above can be extended *multi-class models* which also take into account the distribution over other quantitative features, the size of seeds besides their number, or qualitative features, say different types of items to collect such as walnut seeds or mushrooms. *Multi-type models* occur when the distribution of these features is duly modelled, usually through a probabilistic model.

All the above considerations hold in other fields of application as well: if we change the path with a long urban street, the seeds with cars, and the trees with links we get the main elements of Traffic Theory, briefly presented in Appendix B to this book.

Before any (mathematical) model can be developed a preliminary analysis should be carried out aiming at highlighting the most relevant features and providing a simplified (verbal qualitative) description of the system under study, as briefly reviewed in the beginning of Section 2.2 in the Preface for transportation systems. Main elements are repeated below for reader's convenience together with new considerations.

Users can be travellers, or persons, and freight, or goods. This book main emphasis is on travellers, but most described models can relatively easily adapted to freight transportation. In the following a user may mean a walking person, a person riding a bicycle, a vehicle, a ton of freight, a (20 ft long) container, …, thus pronoun "it" is utilised.

Several types of user choice behaviour occur in real life; this book mostly focuses on two of them only [HYP ①]:

- driving, concerning interactions between users moving on the same facility (called congestion) and their effects on travel time, …;
- routing, concerning connections between origin and destination of the journey, parking location and type, possibly departing time, … .

User driving behaviour is usually modelled within the transportation supply models, which describe (if and) how routing user choices affect level of service, say travel time, delay at junctions, monetary cost, …. On the other hand, user routing behaviour is usually modelled within the travel demand models, which describe how provided level of service affects routing user choices.

Transportation supply systems can be distinguished between those providing *continuous* over space and time *services* (walk, bicycles, cars, vans, trucks, …) or *discrete services* (buses, trains, airplanes, ferries, …), often requiring quite different modelling approaches; the geographical scale, urban/metropolitan areas vs. extra-urban connections, is also a relevant features. Discrete service transportation systems operating in urban areas are often called transit systems.

Making a sharp distinction, a discrete service system may be called:

- *frequency-based*: most users arrive at random at stops without any pre-trip planning since they do not precisely know the timetable or the frequency is so high that it does not matter, thus users perceive the system as a set of lines with a given frequency over time, (often urban services, like buses, metro lines, …),
- *schedule-based*: most users arrive at stops at a chosen instant of time after some pre-trip planning since they precisely know the timetable thus users perceive the system as a set of scheduled connections (most often extra-urban services, inter-city trains, air flights, …).

Presented modelling approaches can be applied to either case above.

For each time of day period relevant for the study at the hand, users are assumed grouped into user classes (multi-user class models) with respect to origin-destination pair and possibly to user category and trip purpose. Each group being associated a travel demand flow, assumed constant in the following, and a common set of itineraries, as well as of behavioural parameters. Each itinerary is broken down into links, each link being a stretch of street (railway, airway, …), possibly shared by others itineraries. Both itineraries and links are modelled following a discrete space approach through the elements of a graph as described in the next section [HYP ②].

[In this book, as already stated in the preface, tools for continuously modelling space, e.g. for dealing with pedestrians moving in a square, are not considered.]

Many models presented through this book follow a macroscopic approach [HYP ③] describing the aggregate results of decisions of all the users through flow variables, measured by number of users per time unit, rather than tracing the journey of each single user as in mesoscopic or microscopic modelling approach; some mesoscopic and microscopic are also discussed, but with less details.

1.1 Space modelling: Graphs and networks

In this book, as said above, the space is modelled in a discrete way through graphs and networks; while these two words are often used as synonymous we wish to give them different meanings as shown below; Section 1.1.1 reviews basic definitions (under steady state conditions) useful for this book (for more details and applications of graphs and networks to other applied sciences too, see for instance Barabási, 2016). The way to apply these models to transportation systems analysis will briefly be discussed in Section 1.1.2. Algebraic details will be introduced in Chapter 2.

1.1.1 Basic elements of graph and network theory

A *graph*, G, is mathematical object defined by an order pair of (finite) sets, $G=(N, A)$, the first one usually called the set of *nodes* (or vertices), N, and the second one usually called the set of *arcs*, A, a subset of the Cartesian product between the first one and itself, $A \subseteq N \times N$, with elements $a=(n_1, n_2) \in A$ where $n_1, n_2 \in N$.

If any element in set A is an unordered pair, $(n_1, n_2)=(n_2, n_1)$, the graph is called undirected and its arcs are often called edges. Vice versa, if all the elements of set A are ordered pairs, arcs (n_1, n_2) and (n_2, n_1) are different, the graph is called directed, or a digraph; in this case the first node n_1, is called the tail of arc (n_1, n_2) exiting from node n_1, while the second node n_2 is the head of arc (n_1, n_2) entering to node n_2.

If needed we may assume that a graph does not contain parallel arcs, say different arcs with same head and tail with no loss of generality; indeed if the graph contains some parallel arcs, it suffices to introduce further nodes to get a new (equivalent) graph without parallel arcs. Therefore, if the number of nodes is finite the number of arcs is finite as well.

Graphs used in all applications to model space in transportation systems analysis are always directed, thus for simplicity's sake we will further refer to this kind of graphs only, omitting the word directed.

Fig. 1.1 shows a (directed) graph with 4 nodes: A, B, C, D, and 5 arcs: 1=(A,C), 2=(B,D), 3=(B,C), 4=(A,B), 5=(C,D).

A *path* from origin node n_o to destination node n_m is (a graph) defined by a sequence of m arcs such that the head of an arc is the tail of the successive one

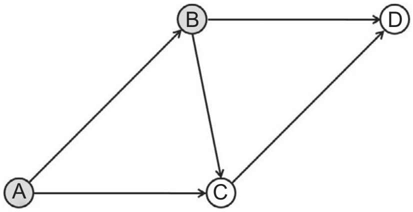


FIG. 1.1

A directed graph with 4 nodes and 5 arcs.

(adjacent arcs), that is (n_O, n_1), (n_1, n_2), …, (n_{m-1}, n_m), a single arc being a special case; the origin node, n_O, and the destination node, n_m, are assumed different. On the other hand, if the origin and the destination are the same node the path is called a loop. In an *elementary path* no node appears twice, thus each node, but the destination, is tail of one arc and head of one arc, with the exemption of the origin; hence an elementary path does not contain loops, and is also called loop-less. Thus, if the number of arcs is finite the number of elementary paths is finite as well. In a *simple path* no arc appears twice, Thus, if the number of arcs is finite the number of simple paths is finite as well.

There 12 pairs of nodes in graph in Fig. 1.1; 6 of them—(B,A), (C,A), (C,B), (D,A), (D,B), (D,C)—are not connected by any path, 3 of them—(A,B), (B,C), (D,C)—are connected, each by one path made by one arc only; last 3 pairs of nodes—(A,C), (A,D), (B,D)—are connected by more than one path, at least one of them composed by more than one arc.

In particular, the pair of nodes (A,D) is connected by 3 paths: 1=(ACD), 2=(ABD) and 3=(ABCD), shown in Fig. 1.2A–C, respectively. [Paths are numbered after arcs 1, 2, 3, respectively.]

An *elementary hyperpath* is an extension of elementary path where some (possibly all) nodes, called *diversion nodes*, may be the tail of one or more arcs, called *diversion arcs*, each of which is given a positive weight summing up to one, called *diversion proportion*; according to this definition an elementary path is an elementary hyperpath where there is no diversion node. An elementary hyperpath (with some diversion nodes) reduces to an elementary path when all diversion proportions are equal to one, and it is called *simple*; otherwise it is called a *composed* elementary hyperpath, made up by several overlapping elementary paths, each given a traversing probability obtained from diversion proportions. It is worth noting that a path may well belong to several hyperpaths, with different traversing probabilities. Thus, if the number, m_P., of elementary paths is finite the number of hyperpaths is finite as well, at most 2^{mp} - 1.

Let nodes A and B be diversion nodes in Fig. 1.1. Between nodes A and B there are 3 simple hyperpaths given by paths: 1=(ACD), 2=(ABD) and 3=(ABCD), shown in Fig. 1.2A–C. Furthermore there are 4 composed hyperpaths:

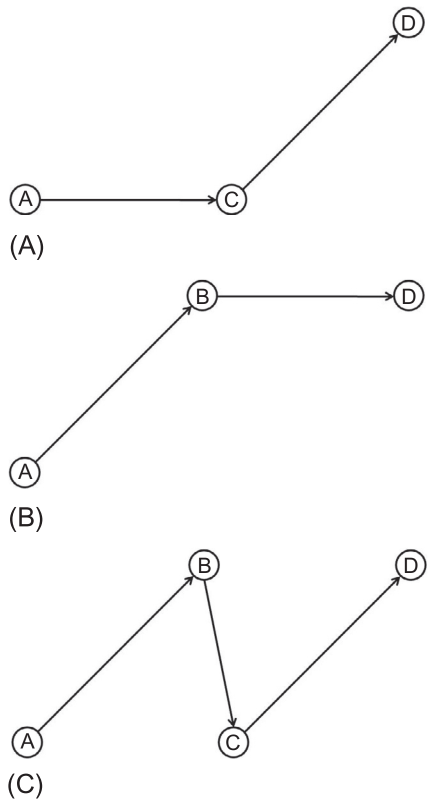


FIG. 1.2

(A) Path 1 = (ACD) between nodes A and D. (B) Path 2 = (ABD) between nodes A and D. (C) Path 3 = (ABCD) between nodes A and D.

- 4 made by paths 1 and 2, assuming that a diversion may occur at node A only;
- 5 made by paths 1 and 3, assuming that a diversion may occur at node A only;
- 6 made by paths 2 and 3, assuming that a diversion may occur at node B only;
- 7 made by paths 1, 2 and 3, assuming that a diversion may occur both at node A and/or node B.

Composed hyperpaths 4, 5, 6, and 7 are shown in Fig. 1.3A–D, respectively. [They are numbered by the sum of the indices of the paths composing them plus 1.]

An ordered pair of (different) nodes, (n_O, n_D) is called *connected* if there is at least one (hyper-)path from the first node, n_O, to the second node, n_D, *single-connected* if there is exactly one (hyper-)path, and *multi-connected* if there are at least two (hyper-) paths; it is worth noting that the inverse pair (n_D, n_O) may not be connected.

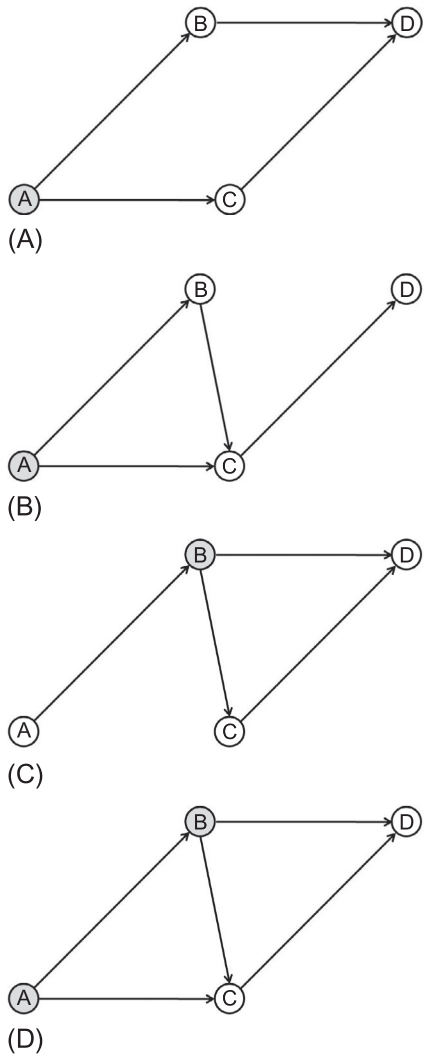


FIG. 1.3

(A) Composed hyperpath 4. (B) Composed hyperpath 5. (C) Composed hyperpath 6. (D). Composed hyperpath 7.

If cost variables are attached to arcs (and possibly to nodes) of a graph it becomes a *network*, as described below. In *discrete* networks variables are attached to arc as a whole, while in *continuous* network variables can be attached to each point along each arc.

A *network* is a graph with a cost, c_a, attached to each arc a. With no loss of generality a cost c' associated to a node n can be transformed into an arc cost by splitting node n into two nodes n_1 and n_2 and introducing arc (n_1, n_2) with cost c'. The arc cost may be a combination of different cost attributes, depending on the application.

In a *Cost Affine Network* (CAN) a cost can be associated to each path through an affine transformation obtained by summing up costs of all arcs belonging to the paths and adding some specific and/or non-additive path costs; then hyperpath costs can be obtained summing up path costs weighted by traversing probabilities.

A *flow network* is a network with a flow, f_a, attached to each arc a. A flow may also be associated to each node equal to the sum of the flows of all arcs entering or exiting the node, unless the node is a source (origin), total exiting flow is greater than total entering flow, or a sink (destination) vice versa. In a *Flow Affine Network* (FAN) the flow associated to each arc can be obtained by an affine transformation of path/hyperpath flows adding base flows due to other phenomena not modelled by paths/hyperpaths (see Chapter 2 for formal details and equation for arc flow conservation at nodes).

A Flow Affine Network that is also a Cost Affine Network such that the Cost Transformation matrix is given by the transpose of the Flow Transformation matrix is called a *Transpose cost flow Affine Network* (TAN). [In this case it is easy to devise simple data structures for moving from path/hyperpath variables to/from arc variables.]

A *congested network* is a flow network where arc costs depend on arc flows through *arc cost* (−flow) *functions*. Quite often the cost attached to an arc can be split in two terms: one including all congested cost attributes depending on flows cost plus one including non-congested cost attributes, independent of arc flows.

In this case path/hyperpath flows depend in turn on path/hyperpath flows. On the other hand, a flow network where arc costs do not depend at all on arc flows is called *uncongested*.

1.1.2 Graphs and networks in transportation system analysis

The practical application of a graph model requires specifying what is modelled by nodes, and which relationship is modelled by arcs, together with the meaning of flows and costs.

In some applications for transportation systems analysis, nodes usually represent locations in space, such as street junctions, airports, …, while arcs represent connecting links, such as streets, airways, …; graphs and networks for this kind of applications are called *synchronic* (or space or 2D graphs or networks). In more detailed modelling approaches further nodes and arcs may be introduced, such as a node for each enter or exit to/fro an intersection, thus each manoeuvre may be represented by an arc, a node for each transit stop, … .

In other applications, for instance to explicitly model scheduled services, it can be useful that a node represents both a location in space and an instant in time; graphs and networks for this kind of applications are called *diachronic* (or space-time or 3D graphs or networks), and are usually acyclic.

In addition, each origin or each destination is modelled through a further node connected to the main network through connecting arcs (or connectors) not corresponding to links, as already stated in the preface. It should be noted the above definitions only refer to the kind of application without affecting the mathematical features of a graph or a network, thus all the above definitions, including those of paths or hyperpaths, hold for any kind of applications.

Paths and hyperpaths, as well as other graphs, can be used to model itineraries available to users say routing alternatives resulting from routing choice behaviour, as described by choice modelling tools in Appendix A. As already stated in the Preface, in the most general case an itinerary is a routing strategy including both pre-trip and en-route choices depending on information available to users, as it occurs for instance if a while-trip information system is in operation, in urban transit networks with overlapping lines, …; these strategies can effectively be modelled through hyperpaths under mild assumptions. In the following the graph used to model an itinerary, in most of the cases a paths or an hyperpath, will be called a *route*, unless further details are needed.

In transportation system analysis generally not all pairs of nodes are relevant, but only those representing an origin and a destination pairs, as defined in Preface. Preface, each of them is called an *OD pair*. Connected, single-connected and multi-connected OD pairs are defined according to definitions introduced earlier with respect to routes.

After the graph has been defined, the practical application of a network model requires specifying what is modelled by costs, flows and cost-flow functions. At this aim it is better to distinguish steady-state conditions vs. within-day dynamics, as discussed in the next section. Anyhow, in transportation applications congested cost attributes usually includes travel time along a street, a stretch of railway, …, waiting time at junctions, …, and possibly other congested costs, such as on-board crowding disutility, penalty for early/late arrival/departure with respect to an indifference time interval; whilst non-congested costs includes monetary costs, due to fuel, tolls, fees, fares, … .

1.2 Time modelling: Dynamic models

Time in nature (at a macro scale at least) follows the time's arrow defined by the second law of thermodynamics - that is our experience of time irreversibility; time in history follows a continuous line from past to present to future as well as.

Time as a social construct to organise daily activities usually follows two evolutions:

- *within-day dynamics*: that occurs over continuous time during the 24 h of a given day (conventionally from 4:00 to 4:00 in transportation systems analysis); often only a part of the day is considered, e.g. the morning peak hour (7:30–8.30); models for transportation systems analysis where this dynamics is explicitly taken into account are called within-day dynamic, otherwise (within-day) steady-state;

- *day-to-day dynamics*: that occurs over discrete time from (a part of) a given day and the next one; models for transportation systems analysis where this dynamics is explicitly taken into account are called day-to-day dynamic, otherwise equilibrium (Box 1.1).

Both these two kinds of dynamics can be observed in a real transportation system and need to be properly modelled through different tools for a complete description of a transportation system. Unfortunately no complete model including both kinds is already available in literature, while only partial modelling approaches have been proposed, named double dynamic models (see Chapter 6). Nonetheless, in this book, as shown in the following Section 1.5, we propose a modelling framework general enough to encompass most existing approaches and to support new ones.

Most effective models used to approach equilibrium analysis are derived from Fixed-Point (FP) theory, as described in Chapter 3 under steady state conditions; these models are very powerful and flexible, and also useful to recognise equilibrium a special case of Day-to-Day Dynamic analysis. Almost all models used to approach Day-to-Day Dynamic analysis are Deterministic Process (DP) models derived from the Non-linear Dynamic Systems theory, as described in Chapter 4, or are derived from the Stochastic Process theory when several sources of uncertainty are explicitly taken into account (see next section), as described in Chapter 5. Following above considerations, under steady-state conditions, Deterministic and Stochastic process models used for Day-to-Day Dynamic analysis are more properly specified in discrete time; they are a very powerful tools for time limit analysis of the evolution over time, DPs suitable for carrying out stability and bifurcation analysis, SPs suitable to carrying out full statistical description through the invariant probability distribution.

On the other hand, Deterministic and Stochastic process models used for Within-Day Dynamic analysis (see Chapter 6) are more properly specified in continuous time following a macroscopic or a microscopic approach, respectively. The application of continuous time or (event-driven) SP models are often called micro-simulation (see

Box 1.1 Diary metaphor.

Consider an instant of time (assumed a measurable quantity at macroscopic level) described in the form: YYYY/MM/DD/sec, where *sec* is a real number in the range [0, 86400]. When you wish to put a new entry in your diary, for example a one-day workshop, first you look for the page corresponding to the year YYYY, the month MM, the day DD, and likely write down on top of the page the title and venue of the congress, then you write down within the page details of the timetable that, in theory, may refer any time, *sec*, within the day. Still, scheduling any event between the day before, the day of the workshop, and the day after is meaningless.

also appendix B). In this case however these tools are used to analyse the evolution over time during a transient, and the time limit analysis is rather meaningless, since main input data, such as demand flows, and some state variables, such as queue length, keep changing over time. These models are often discretised over time and space for solution.

Under steady-state conditions, space can be modelled through a discrete network and a (generalised) transportation cost can be associated to each arc, usually combination of several attributes regarding time (on-board, waiting, delay due to junctions, …) and money (fuel cost, tolls, fees, …), the unit of measure unit being time as for all other costs. Moreover, a flow can be associated to each arc, measured in users per time unit as all other flows. This way, a Transpose cost flow Affine Network (TAN) is obtained, which can be considered synchronic. Cost-flow functions, modelling driving behaviour can be specified through application of Traffic Flow Theory, reviewed in Appendix B, while route choice behaviour can be modelled through tools of Choice Modelling, reviewed in Appendix A.

Since transportation supply can be modelled through a TAN, the theory of travel demand assignment under steady-state conditions is now well developed, both for continuous and discrete (frequency-based) service systems.

Under within-day dynamic conditions different modelling approaches are usually followed to describe transportation supply with continuous or discrete (scheduled) service systems. In the latter case, indeed, a diachronic discrete TAN can effectively be used to model space and time, thus models for steady-state conditions can almost straightforwardly still be applied. In some cases a diachronic discrete TAN could also be used for continuous service systems by duly discretising time, still in several other cases TANs are not suitable for properly modelling Within-day Dynamics.

On the other hand, mostly modelling continuous service systems under within-day dynamic conditions is based on continuous networks since (as already noted in the preface) flows and costs depend on time and on the position within the arc. Moreover, travel time should explicitly be modelled if different from transportation cost.

In addition, the relation between arc and route flows is highly non-linear since the flow entering an arc at a given time depend on travel time to reach the arc, generally through different paths, the travel time of each of these paths depends on the travel time of each arcs previously traversed, which in turn depend on the flow that has traversed them. Hence, within-day dynamic models for transportation supply analysis are highly non-linear including several feedbacks.

Under within-day dynamic conditions demand modelling also requires to include departure time choice behaviour through pre-fixed proportions or explicit choice model through tools of Choice Modelling, reviewed in Appendix A. It is not surprising that no fully consistent unifying general theory is available yet; some existing modelling approaches will be reviewed in Chapter 6, embedding them in the general framework depicted in Section 1.5.

1.3 Uncertainty modelling: Stochastic models

Several sources of uncertainty may prevent a precise description of a transportation system. Some of them are discussed below.

- User perception errors: users may take wrong decisions since they wrongly perceived or weight attributes such as travel time or money affecting the set and the utility values of the available options.
- User heterogeneity: aggregation is necessary to keep any model at a manageable level of complexity but it introduces modelling errors:
  - over space, for example during study area delimitation and zoning;
  - over time, for example neglecting difference among days of the week;
  - over type, for example grouping users with respect to class of income, age, education degree.
- Missing attributes: due lack of data modeller may decide to exclude some attributes who affect users' behaviour, for example weather conditions, or may ignore them.
- Attribute measurement errors: attribute measurements may be affected by errors due to for example data collection procedures, different conditions during collection.

All these sources of uncertainty, as well as others not described above, support the use of uncertain numbers to model (at least some of) the relevant variables of an effective user choice behaviour modelling. As already noted in Preface, even though future were perfectly determined by past, still it may not be perfectly forecasted due to lack of enough information about past, as well as to uncertainty affecting forecasting methods.

Several approaches are available to the skilled modeller to catch the many elusive facets of uncertainty, including fuzzy sets theory, evidence theory, possibility theory leading to fuzzy numbers, probability theory leading to random (numbers or) variables, each of them with strengths and weaknesses, and generally modelling only some of the many facets of uncertainty (some of these approaches are described in Appendix A with reference to choice modelling).

In this book we will mostly use random variables from probability theory since they are well established in transportation systems analysis, calibration of models based on them can be consistently carried out through techniques from inferential statistics, they can easily be casted in a day-to-day dynamic framework leading to stochastic process models. On the other hand, the formal structure of assignment models presented in Chapters 2–4 following a macroscopic approach do not need random variables and may well include other approaches to uncertainty. Needless to say Stochastic process model for day-to-day dynamic assignment, described in Chapter 5, are meaningful under the assumptions of Probability Theory only.

Within-day Dynamics can be modelled, as shown in Chapter 6, through different approaches: macroscopic, mesoscopic and microscopic, the last leading requiring a sort of stochastic modelling. As already noted in the previous section, micro-simulation or discrete simulation usually refers to the application of continuous time or (event-driven) SP models, as in Chapter 6 for Within-day Dynamic analysis and in appendix B for traffic flow analysis.

1.4 Founding conceptual equations

This section describes the main framework used in this book to develop models for travel demand assignment to transportation networks, and will applied in Chapters 2–6. It is general enough to encompass other approaches that are already available literature, but not mentioned in this book, or to support the development of new ones.

The proposed framework for assignment models requires to specify six equations among six vectors, thus any model consistent with it is named SEAM, acronym of Six Equation Assignment Model, also meaning that it joins together the transportation supply model and the travel demand model, as described below.

Given the network modelling connections and cost function, the transportation supply model describes how user choice behaviour affects network level of service through three equations:

EQN 1. arc-route flow consistency relation;
EQN 2. arc cost(−flow) function, modelling effects of driving behaviour (through a macroscopic approach), say congestion (cfr Chapter 2), in uncongested networks arc costs do not depend on arc flows, thus Eq. 2 is not present; similar concepts may be defined for mesoscopic and microscopic approaches; in some cases the arc cost may be different from travel time, and both need to be separately modelled;
EQN 3. route-arc cost consistency relation.

The arc cost(−flow) function is always non-linear. Under steady state conditions both the flow and the cost consistency relations can be specified through affine transformations. On the other hand under within-day dynamic conditions, flow (and possibly cost) consistency relations are highly non-linear.

Given the travel demand flows, assuming that each itinerary is described by a route, the travel demand model this model describes how network level of service affects user route choice behaviour, usually through three equations:

EQN 4. route utility function, between route utility and costs,
EQN 5. route choice function, between route utilities and route choice proportions,
EQN 6. route-demand flow consistency relation, involving demand flows and route choice proportions.

The route-demand flow consistency equation is linear in any case, under the assumption of constant demand flows. The utility function is specified through an affine transformation almost always in research analysis, as well as in practical applications. The route choice function, derived from any discrete choice analysis theory (e.g. Random Utility, Fuzzy Utility, Prospect, … theory) is non-linear; together the utility function and the choice function make the choice model (cfr Appendix A).

Full exploitation of the SEAM framework is currently available for steady-state assignment only, still this conceptual framework may be applied in a broader sense to within-day dynamic assignment too.

Once all the six equations have been specified the resulting SEAM is usually simplified into a model with two non-linear equations between two vectors only with respect to:

- route costs and flows: Eq. 1, Eq. 2 and Eq. 3 are combined together into the route cost function, while Eq. 4, Eq. 5 and Eq. 6 are combined together into the route flow function, or
- arc costs and flows: Eq. 2 stays alone, and Eq. 3 Eq. 4, Eq. 5 Eq. 6 and Eq. 1 are combined together into the arc flow equation.

Either model it is named TEAM, acronym of *Two Equation Assignment Model*. This modelling approach is very effective when a TAN can be used to model transportation supply, since highlight the role of non-linear Eqs. 2 and 5.

1.5 Summary

1.5.1 Major findings

This introduction reports the main hypothesis and definitions underling the models proposed in this book for travel demand assignment to a transportation network, also helping classifying them. At the end of the chapter a general modelling framework is proposed, named SEAM, acronym of Six Equation Assignment Model. This framework is an original contribution of the authors who already used them in some of their recent papers.

Models for *within-day static assignment*, where no kind of dynamics is explicitly addressed, are described in full details in Chapter 2, for uncongested networks, and Chapter 3 where fixed-point (FP) models for equilibrium assignment to congested network are discussed. In Chapters 4 and 5 deterministic (DP) and stochastic process (SP) models for day-to-day dynamic assignment are respectively presented. Models for *within-day dynamic assignment* are described in full details in Chapter 6. Definitions and notations are introduced in a progressive way chapter by chapter.

The reader may find useful to move from this point to Appendix A to review Choice Modelling Theories and to Appendix B to review Traffic Flow Theory.

1.5.2 Further readings

As already stated implementation and application issues are out of the scope of this book, mainly focusing on mathematical features. For details on these issues see Cascetta (2009). Further considerations on within-day static vs. dynamic flows can be found in Köhler et al. (2009). A companion book (possibly by other authors) discussing these issues has already been programmed for 2020.

References

Barabási, A.-L., 2016. Network Science. Cambridge University Press. Cfr the web-site, http://networksciencebook.com.

Cascetta, E., 2009. Transportation Systems Analysis: Models and Applications. Springer.

Köhler, E., Möhring, R.H., Skutella, M., 2009. Traffic networks and flows over time. In: Lerner, J., Wagner, D., Zweig, K.A. (Eds.), Algorithmics of Large and Complex Networks. In: Lecture Notes in Computer Science, vol. 5515. Springer, Berlin, Heidelberg, pp. 166–196.

CHAPTER

Assignment to uncongested networks 2

Giulio Erberto Cantarella
University of Salerno, Salerno, Italy

The White Rabbit put on his spectacles.
"Where shall I begin, please, your Majesty" he asked.
"Begin at the beginning," the King said gravely,
"and go on till you come to the end; then stop."
Lewis Carroll, Alice in Wonderland.

Outline. This chapter introduces basic definitions and notations common to all kinds of assignment models proposed in the next chapters, and a comprehensive modelling approach to static assignment to uncongested networks useful as such or as a part of models for assignment to congested network; presented models are consistent with the SEAM modelling framework presented in Chapter 1, actually being a special case, called Five Equation Assignment Model since the arc cost function is missing; the arc flow function as a flexible model is introduced and discussed.

Methods for travel demand assignment to transportation networks play a central role in transportation system analysis describing how travel demand and transportation supply interact each other. These methods allow to compute performance level and user flow for each supply element (network arc), resulting from origin-destination demand flows, user choice behaviour, and the interactions between supply and demand in the current or any design scenario. Their results, in turn, are the inputs for the design and/or assessment evaluation of transportation projects.

This chapter discusses within-day static assignment to uncongested transportation networks, the simplest kind of assignment, relevant as such and as the first step of a long journey through increasingly more general kinds of assignment discussed in the following Chapters 3–5, before moving to within-day dynamic assignment in Chapter 6. In any case travel demand is assumed given [HYP ④].

The topic of travel demand assignment to uncongested transportation networks has been rarely been discussed as such in literature. The book by Sheffi (1985) was the first to provide a specific analysis of this kind of assignment assuming that the route choice behaviour is described by a Random Utility Model (cfr Chapter 3), naming it Stochastic

Dynamics and Stochasticity in Transportation Systems. https://doi.org/10.1016/B978-0-12-814353-7.00002-9
© 2020 Elsevier Inc. All rights reserved.

Network Loading (SNL); since then this name has been widely used - assignment with deterministic route choice being traditionally called All-or-Nothing (AoN). Strangely, afterwards the name Dynamic Network Loading (DNL) has been introduced for defining the relationship between route and arc flows and costs (cfr EQN 1 and EQN 2 of SEAM) under within-day dynamic conditions, thus leading to ambiguous meaning of Network Loading. Recently, the book by Cascetta (2009) introduced clearer names and acronyms: Stochastic/Deterministic assignment to Uncongested Networks (SUN/DUN).

In this chapter, we introduce and discuss a comprehensive modelling approach to assignment to uncongested networks, including all those mentioned above as well as all those that result from most route choice modelling approaches (cfr Appendix B to the book). Hence we named it Comprehensive assignment to Uncongested Networks (CUN). It is described for steady-state conditions, but it also applies to any TAN. Presented models are consistent with the SEAM modelling framework presented in Chapter 1, actually being a special case, called Five Equation Assignment Model (FEAM) since EQN 2, arc cost function, is missing. Emphasis is models and their features, whilst solution algorithms are only briefly addressed.

A special case occur assuming that all users follow a maximum utility or minimum cost routes, this kind of assignment is traditionally called All-or-Nothing (AoN) assignment and this denomination is followed in this book since it may be derived from several theories, such as Deterministic Utility Theory, Expected Utility Theory, or as a limit case of Random or Fuzzy Utility Theory (see Remarks in the Summary at the end of this chapter).

First we introduce basic notations and definitions in Section 2.1, then we discuss single-class assignment, when of all users belonging to the same class, in Section 2.2, and multi-class assignment in Section 2.3, when users are grouped into classes with different characteristics; in Section 2.4 independent route formulation is discussed and finally in Section 2.5 the arc flow function is introduced for assignment to uncongested networks.

2.1 Basic notations and definitions

Connections are described by an oriented graph $G(N, A \subseteq N \times N)$, an order pair of a set of nodes N and a set of arcs A. Each origin and each destination is modelled through a further node connected to the main network through connecting arcs (or connectors) not corresponding to real infrastructures. Each route connecting an origin - destination pair (o-d pair) is described by an acyclic sub-graph depending on the application: a path, when the route to destination is completely chosen at origin, or a hyperpath, when the route chosen is result of travelling strategy including both pre-trip and en-route choices.

Transportation supply is modelled through a (synchronic or diachronic) flow network, that is a graph with a transportation cost, c_a, and a flow, f_a, associated to each arc, a. Moreover, a transportation cost, w_r, and a flow, h_r, is associated to each route, r.

Main sets and vectors used in the following are enlisted below in alphabetical order (sets come first) for reader's convenience:

A is the set or arcs, assumed non empty and finite, with $m=|A|$ elements;
R is the set or routes assumed non empty and finite, with $n=|R|$ elements;
$\mathbf{c}\geq\mathbf{0}$ is the $m\times 1$ (column) vector of *arc costs* with entries c_a;
$\mathbf{f}\geq\mathbf{0}$ is the $m\times 1$ (column) vector of *total arc flows* with entries f_a;
$\mathbf{h}\geq\mathbf{0}$ is the $n\times 1$ (column) vector of *route flows* with entries h_r;
$\mathbf{w}\geq\mathbf{0}$ is the $n\times 1$ (column) vector of *total route costs* with entries w_r.

Under steady state conditions, consistency equations hold among the above variables expressed by affine transformations (TAN):

$$\mathbf{f}=\mathbf{B}\cdot\mathbf{h}+\mathbf{f_Z} \tag{2.1}$$

$$\mathbf{w}=\mathbf{B}^{\mathrm{T}}\cdot\mathbf{c}+\mathbf{w_Z} \tag{2.2}$$

where

$\mathbf{B}$ is the $m\times n$ *arc-route generalised incidence matrix* (ARGIM) with entries $b_{ar}\in]0,1]$ if route r uses arc a, $b_{ar}=0$ otherwise; meaning of entries b_{ar} depends on the definition of route, see below for details and examples;
$\mathbf{f_Z}\geq\mathbf{0}$ is the $m\times 1$ (column) vector of *other arc flows* with entries $f_{Z,a}$ not due to route flows, for instance route flows $\mathbf{h}$ regard car flows, while arc flows in vector $\mathbf{f_Z}$ regard lorries, taxis, …;
$\mathbf{w_Z}\geq\mathbf{0}$ is the $n\times 1$ (column) vector of *other route costs* with entries $w_{Z,r}$ that are not additive over generic arc costs, for instance tolls, fees, … with respect to travel time.

Eq. (2.1) also expresses the arc flow conservation at each node for each o-d pair. [Thus, no further equation is need like in fluid or electric networks where route flows are not introduced as variables.]

The arc-route generalised incidence matrix in Eqs (2.1) and (2.2) is very useful for compact matrix notations, but there is no need to explicitly define it for computation and application. [Examples below show it for pedagogical purpose only.]

Above equations and the meaning of matrix of $\mathbf{B}$ are discussed below for two cases: routes are modelled by paths or by hyperpaths. Further useful definitions of route variables will be introduced and discussed at the end of this chapter. Let

Δ be the arc-path incidence matrix with entries $\delta_{ak}=1$ if arc a belongs to path k, $\delta_{ak}=0$ otherwise.

Table 2.1 shows the arc-path incidence matrix for graph in Fig. 1.1 with respect to the 3 paths connecting nodes A and B, already shown in Fig. 1.2.

The cost of path k is given by the sum of arc costs over all arcs belonging to path k plus the other path cost:

$$X_k=\sum\nolimits_{a\in k}c_a+X_{Z,k} \qquad \forall k$$
$$\text{or } X_k=\sum\nolimits_{a}\delta_{ak}c_a+X_{Z,k} \qquad \forall k \tag{2.3}$$
$$\text{or } \boldsymbol{X}=\Delta^{\mathrm{T}}\cdot\mathbf{c}+\mathbf{x_Z}$$

Table 2.1 Incidence matrix (cfr graph in Fig. 1.1)

		routes		
		1	2	3
arcs	1	1	0	0
	2	0	1	0
	3	0	0	1
	4	0	1	1
	5	1	0	1

Table 2.2 Arc-path cost consistency

		routes			
		1	2	3	c
arcs	1	1	0	0	15
	2	0	1	0	12
	3	0	0	1	8
	4	0	1	1	24
	5	1	0	1	15
	w	30	36	47	

where

$x_k \geq \mathbf{0}$ is the cost of path k;
$\mathbf{x} \geq \mathbf{0}$ is the path cost vector, with entries x_k;
$x_{Z,k} \geq \mathbf{0}$ is the other cost of path k;
$\mathbf{x_Z} \geq \mathbf{0}$ is the path other cost vector, with entries x_{Zk}.

Table 2.2 shows an example of the arc-path cost consistency equation, with respect to the arc-path incidence matrix in Table 2.1, other costs are not considered.

Analogously the flow of arc a is given by the sum of the path flows over all paths traversing arc a plus the other arc flow:

$$f_a = \sum\nolimits_{k:a\in k} y_K + f_{Z,a} \qquad \forall a$$
$$\text{or } f_a = \sum\nolimits_{k} \delta_{ak} y_K + f_{Z,a} \qquad \forall a \tag{2.4}$$
$$\text{or } \mathbf{f} = \boldsymbol{\Delta} \cdot \mathbf{y} + \mathbf{f_Z}$$

where

$y_k \geq \mathbf{0}$ is the flow of path k;
$\mathbf{y} \geq \mathbf{0}$ is the path flow vector, with entries y_k.

If routes are paths then $\mathbf{h} = \mathbf{y}$, $\mathbf{w} = \mathbf{x}$, and the arc-route generalised incidence matrix is given by the arc-path incidence matrix, $\mathbf{B} = \boldsymbol{\Delta}$, thus (2.1) and (2.2) are proved.

Table 2.3 shows an example of the arc-path cost consistency equation, with respect to the arc-path incidence matrix in Table 2.1, other flows are not considered.

Table 2.3 Path-arc flow consistency

		routes 1	routes 2	routes 3	f
h		2380	1010	210	
arcs	1	1	0	0	2380
arcs	2	0	1	0	1010
arcs	3	0	0	1	210
arcs	4	0	1	1	1220
arcs	5	1	0	1	2590

On the other hand, if routes are hyperpaths, they includes some diversion nodes and each arc exiting from any diversion node is given a diversion proportion (cfr Chapter 1). Let $\psi_{ar} \in [0,1]$ be the diversion proportion given to arc a with reference to hyperpath (route) r, it may have the following values:

$\psi_{ar} \in]0,1]$, if arc a is a diversion arc belonging to hyperpath r, $\psi_{ar}=1$ meaning that only one diversion arc exits from the diversion node,
$\psi_{ar}=1$, if arc a is not a diversion arc, and belongs to hyperpath r,
$\psi_{ar}=0$, otherwise, that is if arc a does not belong to hyperpath r.

All diversion proportions from the same (diversion) node have to sum up to 1. The values of diversion proportions at a diversion nodes depend on the application.

For instance for an urban transit system with overlapping lines they are proportional to the frequencies of the bus lines considered within the hyperpath among all lines available at the bus stop corresponding to the diversion node. Fig. 2.1 shows an example of transit network (cfr Chapter 1), each arc is a bus line, the attached number is the bus frequency, A and B may be diversion nodes.

Corresponding diversion proportions for the composed hyperpaths 4, 5, 6, and 7—cfr Fig. 1.3A–D—are shown in Fig. 2.2A–D.

The proportion λ_{kr} that user follows path k having chosen hyperpath r is given by:

$$\lambda_{kr} = \prod_{a \in k} \psi_{ar}$$

thus

$\lambda_{kr} \in]0,1]$ if path k is in hyperpath r,
$\lambda_{kr}=0$ otherwise, that is path k is not in hyperpath r.

Let

$\boldsymbol{\Lambda}$ be the path-hyperpath proportion matrix with entries λ_{kr}, with $\sum_k \lambda_{kr}=1\ \forall r$.

Table 2.4 shows an example of the path-hyperpaths proportion matrix considering the simple hyperpaths 1, 2, and 3, and the composed hyperpaths 4, 5, 6, and 7.

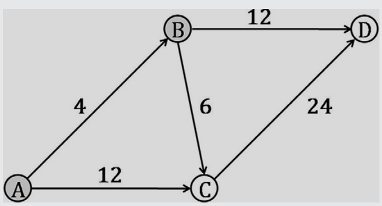


FIG. 2.1

Transit network with line frequencies.

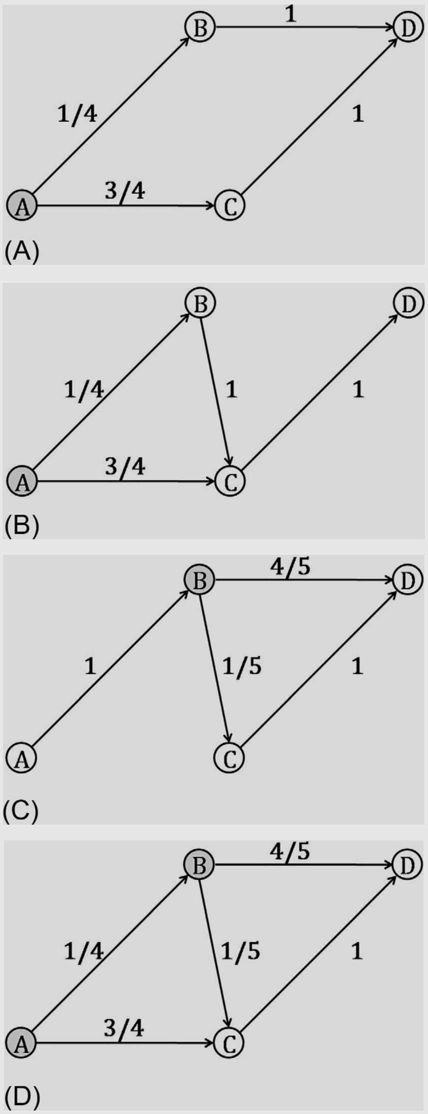


FIG. 2.2

(A) Diversion proportions for composed hyperpath 4. (B) Diversion proportions for composed hyperpath 5. (C) Diversion proportions for composed hyperpath 6. (D) Diversion proportions for composed hyperpath 7.

Table 2.4 Path-hyperpath matrix

		hyperpaths						
		1	2	3	4	5	6	7
paths	1	1.00	0.00	0.00	0.75	0.00	0.00	0.75
	2	0.00	1.00	0.00	0.25	0.75	0.80	0.20
	3	0.00	0.00	1.00	0.00	0.25	0.20	0.05

The cost of hyperpath r is given by the weighted sum of the path costs over all paths belonging to hyperpath r plus the other hyperpath cost not including other path cost:

$$w_r = \sum_k \lambda_{kr} x_k + w_{Z,r}' \quad \forall r$$

or

$$\mathbf{w} = \boldsymbol{\Lambda}^{\mathrm{T}} \cdot \mathbf{x} + \mathbf{w_Z}' \tag{2.5}$$

where

$\mathbf{w_Z}' \geq \mathbf{0}$ is the vector of *other route costs* with entries $w_{Z,r}'$ not including other path costs.

Analogously the flow of path k is given by the weighted sum of the hyperpath flows over all hyperpaths including path k:

$$y_k = \sum_r \lambda_{kr} h_r \quad \forall k$$

or

$$\mathbf{y} = \boldsymbol{\Lambda} \cdot \mathbf{h} \tag{2.6}$$

If routes are hyperpaths combining Eqs (2.3) and (2.5) yields:

$$\mathbf{w} = \boldsymbol{\Lambda}^{\mathrm{T}} \cdot \boldsymbol{\Delta}^{\mathrm{T}} \cdot \mathbf{c} + \mathbf{w_Z} \tag{2.7}$$

where $\mathbf{w_Z} = \boldsymbol{\Lambda}^{\mathbf{T}} \cdot \mathbf{x_Z} + \mathbf{w_Z}'$. Moreover, combining Eqs. (2.4) and (2.6) yields:

$$\mathbf{f} = \boldsymbol{\Lambda} \cdot \boldsymbol{\Delta} \cdot \mathbf{h} + \mathbf{f_Z} \tag{2.8}$$

Thus assuming $\mathbf{B} = \boldsymbol{\Lambda} \cdot \boldsymbol{\Delta}$ (2.1) and (2.2) are proved.

2.2 Basic assignment models

This section presents models for assignment to uncongested networks if users are only distinguished with respect to o-d pair i they are travelling from/to with a common set of routes, called the route choice set assumed non empty and finite. Main vector notations used in the following are enlisted below in alphabetical order for reader's convenience (sets come first, then Roman letters, at last Greek letters); these notations exploit the block structure of most vectors and matrices introduced in the previous section.

A is the set of arcs, with $m = |A|$ elements;
R_i is the set of (elementary) routes for o-d pair i, with $n_i = |R_i|$ elements;
$n = \sum_i n_i$ is the number of routes connecting all o-d pairs;

$S_{hi} \triangleq \{\mathbf{h}_i \geq \mathbf{0}: \mathbf{1}^{\mathbf{T}}\, \mathbf{h}_i = d_i\} \subseteq \mathbb{E}^{n_i}$ is the *feasible route flow set* for o-d pair i according to route-demand flow consistency Eq. (2.13);

$S_h \triangleq \{\mathbf{h}$ with blocks $\mathbf{h}_i \in S_{hi}\} = \{\mathbf{h}$ with blocks $\mathbf{h}_i \geq \mathbf{0}: \mathbf{1}^{\mathbf{T}}\, \mathbf{h}_i = d_i\} \subseteq \mathbb{E}^n$ be the *feasible route flow set*;

$\mathbf{B}_i$ is the $(m \times n_i)$ block of the *ARGIM* for o-d pair i;

$\mathbf{c} \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *arc costs*;

$d_i \geq 0$ is the demand flow for o-d pair i;

$\mathbf{f} \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *total arc flows*;

$\mathbf{f}_{\mathbf{Z}} \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *other arc flows*;

$\mathbf{h}_i \geq \mathbf{0}$ is the $n_i \times 1$ (column) block of the vector of *route flows* for o-d pair i;

$\mathbf{h}_i(\cdot)$ is the $n_i \times 1$ (column) *route flow function* for o-d pair i;

$\mathbf{p}_i \geq \mathbf{0}$ is the $n_i \times 1$ (column) block of the vector of *route choice proportions* for o-d pair i, with $\mathbf{1}^{\mathbf{T}}\, \mathbf{p}_i = 1$;

$\mathbf{p}_i(\cdot)$ is the $n_i \times 1$ (column) *route choice function* for o-d pair i;

$\mathbf{v}_i$ is the $n_i \times 1$ block of the (column) vector of *route systematic utility* for o-d pair i;

$\mathbf{w}_i \geq \mathbf{0}$ is the $n_i \times 1$ block of the (column) vector of *total route costs* for o-d pair i;

$\mathbf{w}_{\mathbf{Z}i} \geq \mathbf{0}$ is the $n_i \times 1$ block of the (column) vector of *other route costs* for o-d pair i;

$\mathbf{w}_i(\cdot)$ is the $n_i \times 1$ (column) *route cost function* for o-d pair i;

$\boldsymbol{\theta}_i > 0$ is the vector of the *route choice function parameters* for o-d pair i;

$\psi_i > 0$ is the *utility scale parameter* in the route choice model, for o-d pair i.

Arc and route and demand flows are assumed measured by a common unit: users per time unit, where a user may be a person walking or riding a bicycle, or a vehicle (a car, a bus, a truck, …), a ton of freight, a TEU, …. Arc and route costs are assumed measured by a common unit, usually travel time or money, through duly homogenization of different attributes not explicit introduced to simplify notations.

This section focuses on transportation systems with a single transportation mode [HYP ⑤] and single vehicle type [HYP ⑥].

2.2.1 Supply model

Transportation supply models express how user behaviour affects network performances. This section describes the two equations that according to the FEAM framework specify the transportation supply model for a transportation system in steady-state conditions in the case of an uncongested network, when users are distinguished with respect to o-d pair only.

- arc-route flow consistency relation

Under the steady-state assumption the total arc flows due to all o-d pairs can be obtained from the route flows through an affine transformation from the route space to the arc space defined by the arc-route generalised incidence matrix; highlighting the block structure in Eq. (2.1) it yields:

$$\mathbf{f} = \sum\nolimits_i \mathbf{B}_i \cdot \mathbf{h}_i + \mathbf{f}_{\mathbf{Z}} \tag{2.9}$$

- arc cost function

No arc cost function since the network is uncongested, thus the arc cost vector, **c**, is an input data.

- route-arc cost consistency relation

Under the steady-state assumption the route costs for o-d pair i, can be obtained from the arc costs through an affine transformation from the arc space to the route space defined by the transpose of arc-route generalised incidence matrix; highlighting the block structure in Eq. (2.2) it yields:

$$\mathbf{w}_i = \mathbf{B}_i^{\mathrm{T}} \cdot \mathbf{c} + \mathbf{w}_{\mathbf{Z}i} \quad \forall i \tag{2.10}$$

2.2.2 Demand model

Travel demand models express how network performances affect user choice behaviour. Only route choice behaviour, and possibly vehicle type choice behaviour are explicitly modelled assuming constant demand flows. This section first describes the three equations that according to SEAM framework specify the travel demand model with given demand flows (constant demand).

- route utility function

The utility function for o-d pair i is assumed specified through a linear transformation of route costs [HYP ⑦], almost always in research analysis as well as in practical applications:

$$\mathbf{v}_i = -\psi_i \, \mathbf{w}_i \quad \forall i \tag{2.11}$$

where $\psi_i > 0$ is the utility scale parameter, such that the term $\psi_i \, \mathbf{w}_i$ is dimensionless to be consistent with utility unit. A constant term has not been introduced since it plays the same role of the other route cost vector. [More generally the utility function may be any continuous non-linear strictly decreasing separable function, since the utility of a route only depends on the cost of that route.]

- route choice function

Route choice behaviour for o-d pair i can be described by applying any discrete choice modelling theory (see Appendix A2 to the book) so that route choice proportions depend on route systematic utility:

$$\mathbf{p}_i = \mathbf{p}_i(\mathbf{v}_i; \boldsymbol{\theta}_i) \quad \forall i \tag{2.12}$$

where $\boldsymbol{\theta}_i$ is the choice function parameter vector, whose meaning depends on the choice model specification. If a utility scale parameter is present, it is considered included in the utility parameter ρ_i (or vice versa).

Definition 1 A choice function is defined *regular*, if:

- it is continuous and monotone increasing with respect to systematic utility,
- it is continuously differentiable with symmetric positive semi-definite (with respect to real vectors) Jacobian, formally $\nabla \mathbf{p}_i(\mathbf{v}_i) \succeq 0$,
- resulting choice proportions depend on differences between systematic utility values only [HYP ⑧].

Most often Random Utility Theory (RUT) is applied, assuming that (i) each user, travelling between o-d pair i, associates to each available route a *perceived utility*, (ii) the perceived utility is modelled by a continuous random variable, with mean given by the *systematic utility*, due to several sources of uncertainty regarding the users or the modeller, and (iii) chooses the maximum perceived utility route; thus the choice probability of an alternative is given by the probability that its perceived utility is equal to maximum among all alternatives; hence the route choice proportions are assumed defined by the route choice probabilities.

When the perceived utility co-variance matrix is non singular, a *probabilistic route choice function* is obtained; it is also called *strictly positive* if each alternative gets a strictly positive probability, whichever are the systematic utility values; examples of strictly positive probabilistic route choice functions are the Logit, Weibit, Probit, Gammit choice functions, usually adopted for route choice modelling. [Strictly positive choice functions may sound somehow unrealistic, as any model they have to be considered suitable mathematical approximations.] If the parameters of the perceived utility pdf do not depend on systematic utility values, the resulting choice function is called *invariant*, if continuous and continuously differentiable, it is regular.

Anyhow Eq. (2.12) is generally enough to include choice models derived from other discrete choice theories, such as Fuzzy Utility Theory, Bounded Rationality, Prospect Theory, … (some of them are described in Appendix B). In any case, a choice function combined with an utility function gives a choice model.

An example of choice function (2.12) derived from RUT is the well known Logit choice function, often used as benchmark. For each o-d pair i, connected by routes in the route choice set R_i, the choice proportion/probability of using route r is given by:

$$p_r = \exp(\mathrm{v}_r/\theta_i) / \sum_{k \in Ri} \exp(\mathrm{v}_k/\theta_i) \quad \forall r \in R_i$$

where $\theta_i = (6^{0.5}/\pi)\,\sigma_i \cong 0.78\,\sigma_i > 0$ is a dispersion parameter proportional to the standard deviation σ_i common to the perceive utilities of all routes connecting o-d pair i; the above Logit function is invariant if the route choice set R_i and the dispersion parameter θ_i do not depend on systematic utility values. Combing the above choice function with the utility function: $\mathrm{v}_r = -\psi_i\,\mathrm{w}_r$, leads to:

$$p_r = \exp(-\mathrm{w}_r/\theta_i) / \sum_{k \in Ri} \exp(-\mathrm{w}_k/\theta_i) \quad \forall r \in R_i$$

The utility scale parameter ψ_i is included in the dispersion parameter θ_i.

- route-demand flow consistency relation

Flow conservation for o-d pair i can be expressed as:

$$\mathbf{h}_i = d_i\,\mathbf{p}_i \quad \forall i \tag{2.13}$$

It assures that the sum of the flows of all routes connecting the o-d pair i sum up to the demand flow, that is $\mathbf{1}^{\mathbf{T}}\,\mathbf{h}_i = d_i$, since $\mathbf{1}^{\mathbf{T}}\,\mathbf{p}_i = 1$, and non-negative, $\mathbf{h}_i \geq \mathbf{0}$, since $d_i \geq 0$ and $\mathbf{p}_i \geq \mathbf{0}$. A constant term has not been introduced since the resulting arc flows may be considered in the other arc flow vector. Let

n_i be the number of routes connecting o-d pair i;
$n = \sum_i n_i$ be the number of routes connecting all o-d pairs;
$S_{hi} \triangleq \{\mathbf{h}_i \geq \mathbf{0}\text{: } \mathbf{1}^{\mathbf{T}}\,\mathbf{h}_i = d_i\} \subseteq \mathbb{E}^{n_i}$ is the *feasible route flow set* for o-d pair i according to route-demand flow consistency Eq. (2.13); this set

- has a finite dimension if the number of routes available to o-d pair i is finite (as it occurs considering all or some elementary routes),
- is non empty if o-d pair i is connected by at least one route,
- is compact, since closed and bounded [in the Euclidean space],
- is convex.

$S_h \triangleq \{\mathbf{h}$ with blocks $\mathbf{h}_i \in S_{hi}\} = \{\mathbf{h}$ with blocks $\mathbf{h}_i \geq \mathbf{0}\text{: } \mathbf{1}^{\mathbf{T}}\,\mathbf{h}_i = d_i\} \subseteq \mathbb{E}^n$ be the *feasible route flow set*, with same features of sets S_{hi} since their number is finite.

Fig. 2.3 shows the feasible route flow set for an o-d pair connected by 3 routes (cfr pair of nodes (A, D) in Chapter 1), with demand flow d. It is described by a triangle in a 3-dimensional space defined by 3 axis, one for reach route flow, h_1, h_2, h_3, each vertex representing the case of all demand flow, d, using one route only.

Set in Fig. 2.3 can also be drawn in a 2-dimensional space (plane), as in Fig. 2.4. It is still described by a triangle each vertex representing the case of all demand flow, d, using one route only (cfr De Finetti diagram). On each axis the wide tick is at flow equal to $0.50d$, narrow ticks at $0.25d$ or $0.75d$; the three axes meet in point representing the case of all 3 route flows being equal, $h_1 = h_2 = h_3 = 1/3$ (the blue circle).

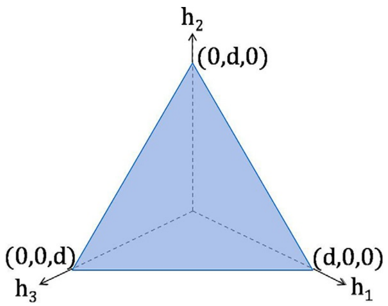


FIG. 2.3

Feasible route set, in 3D space, for 3 routes.

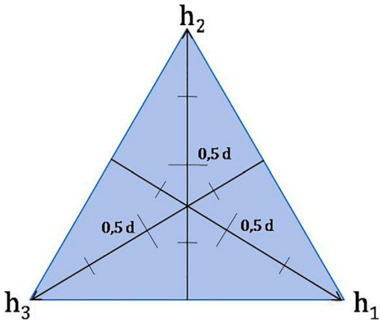


FIG. 2.4

Feasible route set, in 2D space, for 3 routes.

Table 2.5 shows an example of demand model, say Eqs (2.11), (2.12), and (2.13), from route costs to route flows, through route utility and route choice proportions, with reference to the 3 routes in Tables 2.2 and 2.3. Choice functions is the Logit function in the above example, the utility scale factor is included in dispersion parameter $\theta=7$; the demand flow $d=3600$.

Table 2.5 Demand model.

w	30	36	47
v	–30	–36	–47
p	0.66	0.28	0.06
h	2380	1010	210

2.2.3 Parking choice behaviour

Parking choice behaviour jointly with route choice behaviour can easily be casted in the previously described model, defined by Eqs (2.9)–(2.13), by describing each destination with three nodes:

1. the access node, it models the access to parking facilities available in the destination zone and it is connected through a parking arc for each available parking type (free, metered, illegal, on street, on dedicated areas, …) to
2. the egress node, it models the egress from the parking facility and it is connected through walking arcs to
3. the final destination node, and possibly to the final destination nodes in other zones.

The cost associated to each parking arc includes time to find a slot, fare (for metered slots) or mean fine (for illegal parking), and can be differentiated by od pair i, user category j and vehicle type m to model restricted parking policies. This modelling approach is based on Bifulco (1993).

2.3 Independent route formulations

The above presented FEAM (2.9)–(2.13) can be reformulated with respect to independent route variables. Indeed one route choice proportion or flow is redundant because it may easily be obtained from the others, assuming all o-d pairs connected by at least two routes with no loss of generality; indeed arc flows due to o-d pairs connected by one route can be added to the vector of other arc flows. With reference to the standard form for simplicity's sake, for each user class i any of the first $\tilde{n}_i = n_i - 1$ routes is called an independent route, *i-route* for short. To get a compact matrix formulation, let

- n_i be the number of routes connecting o-d pair i;
- $\tilde{n}_i = n_i - 1$ be the number of i-routes connecting o-d pair i;
- $\mathbf{I}_i$ be the $(n_i \times n_i)$ identity matrix;
- $\mathbf{E}_i$ be the $(\tilde{n}_i \times n_i)$ matrix obtained by dropping the last row from the identity matrix $\mathbf{I}_i$;
- $\mathbf{e}_i = [0, 0, \ldots, 1]^{\mathrm{T}}$ be the n_i-*th* versor, say a $(n_i \times 1)$ column vector given by the transpose of the last row of the identity matrix $\mathbf{I}_i$;
- $\mathbf{1}_i = [1, 1, \ldots, 1]^{\mathrm{T}}$ be an $(n_i \times 1)$ column vector with all entries equal to one;
- $\mathbf{L}_i = (\mathbf{I}_i - \mathbf{e}_i \cdot \mathbf{1}_i^{\mathrm{T}}) \cdot \mathbf{E}_i^{\mathrm{T}}$ be a $(n_i \times \tilde{n}_i)$ the route - i-route matrix, obtained from the $(\tilde{n}_i \times \tilde{n}_i)$ identity matrix by adding at the bottom one more row given by the $(1 \times \tilde{n}_i)$ row vector $-\mathbf{1}_i^{\mathrm{T}}$ with all entries equal to -1.

Table 2.6 shows how matrix **L** can be obtained from matrix **I** for 3 routes.

Table 2.6 Route—i-route matrix.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \mathbf{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{e}^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{e} \cdot \mathbf{1}^{\mathrm{T}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{I} - \mathbf{e} \cdot \mathbf{1}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{bmatrix}$$

$$\mathbf{L} = (\mathbf{I} - \mathbf{e} \cdot \mathbf{1}^{\mathrm{T}}) \cdot \mathbf{E}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}$$

The equations below hold between route and i-route choice proportion and flow vectors in any case:

$$\tilde{\mathbf{p}}_i = \mathbf{E}_i \cdot \mathbf{p}_i \quad \forall i \tag{2.14}$$

$$\tilde{\mathbf{h}}_i = \mathbf{E}_i \cdot \mathbf{h}_i \quad \forall i \tag{2.15}$$

$$\mathbf{p}_i = \mathbf{L}_i \cdot \tilde{\mathbf{p}}_i + \mathbf{e}_i \quad \forall i \tag{2.16}$$

$$\mathbf{h}_i = \mathbf{L}_i \cdot \tilde{\mathbf{h}}_i + \mathbf{h}_{\mathbf{Z}i} \quad \forall i \tag{2.17}$$

where

- $\tilde{\mathbf{h}}_i \geq \mathbf{0}$ is the $\tilde{n}_i \times 1$ (column) block of the vector of *i-route flow*s for o-d pair i;
- $\tilde{\mathbf{h}}_i(\cdot)$ is the $\tilde{n}_i \times 1$ (column) *i-route flow function* for o-d pair i;
- $\tilde{\mathbf{p}}_i \geq \mathbf{0}$ is the $\tilde{n}_i \times 1$ (column) block of the vector of *i-route choice proportions* for o-d pair i, with $\mathbf{1}^{\mathbf{T}}\ \mathbf{p}_i < 1$;
- $\tilde{\mathbf{p}}_i(\cdot)$ is the $\tilde{n}_i \times 1$ (column) *i-route choice function* for o-d pair i;
- $\tilde{\mathbf{w}}_i$ is the $\tilde{n}_i \times 1$ (column) *i-route cost (differences) function* for o-d pair i.
- $\mathbf{h}_{\mathbf{Z}i} = d_i\ \mathbf{e}_i$ is a $(n_i \times 1)$ column vector with all entries equal to 0, but the last one equal to the demand flow.

In most cases, such as for invariant choice functions derived from RUT, the route choice proportions $\mathbf{p}_i$ for user class i do not actually depend on systematic utilities $\mathbf{v}_i$ but on their differences only [HYP ⑧]. Let $\tilde{\mathbf{v}}_i$ be the vector of i-route systematic utility differences, with an entry for each i-route given by the i-route systematic utility minus the systematic utility of the last route, that is:

$$\tilde{\mathbf{v}}_i = \mathbf{L}_i^{\mathrm{T}}\ \mathbf{v}_i \quad \forall i \tag{2.18}$$

Thus the i-route choice proportions $\tilde{\mathbf{p}}_i$ may be specified as a function of the i-route systematic utility differences, that is:

$$\tilde{\mathbf{p}}_i = \tilde{\mathbf{p}}_i\left(\tilde{\mathbf{v}}_i; \boldsymbol{\theta}_i\right) \quad \forall i \tag{2.19}$$

See Section 2.2 for details about its specifications and features.

Definition 2 A choice function referring to independent alternatives only is defined *strictly (positive) regular*, if:

- it gives strictly positive choice proportions for any values of systematic utility,
- it is continuous and monotone strictly increasing with respect to systematic utility differences,
- it is continuously differentiable with symmetric positive definite (with respect to real vectors) Jacobian, $\nabla\, \tilde{\mathbf{p}}_i(\tilde{\mathbf{v}}_i) \succ 0$,
- resulting choice proportions depend on systematic utility differences only.

Instances of strictly regular choice functions are invariant strictly positive probabilistic choice functions derived from the Random Utility Theory (such that the perceived utility pdf is strictly positive over an unbounded set with non null measure). Examples are the Logit, Weibit, Probit, Gammit choice functions, usually adopted for route choice modelling. Instances also exist from other discrete choice theories (see Appendix A1).

Moreover, let

$\tilde{\mathbf{w}}_i$ be the vector of i-route cost differences, with an entry for each i-route given by the i-route cost minus cost of the last route, that is:

$$\tilde{\mathbf{w}}_i = \mathbf{L}_i{}^{\mathrm{T}} \cdot \mathbf{w}_i \quad \forall i \tag{2.20}$$

The supply model (2.9)–(2.10) repeated below for reader's convenience:

$$\mathbf{f} = \sum\nolimits_i \mathbf{B}_i \cdot \mathbf{h}_i + \mathbf{f}_{\mathbf{Z}} \tag{2.9}$$

$$\mathbf{w}_i = \mathbf{B}_i{}^{\mathrm{T}} \cdot \mathbf{c} + \mathbf{w}_{\mathbf{Z}i} \quad \forall i \tag{2.10}$$

can be formulated with respect to the i-route variables including Eq. (2.17) into (2.9) and (2.20) into (2.10):

$$\mathbf{f} = \sum\nolimits_i \mathbf{B}_i \cdot \left(\mathbf{L}_i \cdot \tilde{\mathbf{h}}_i + \mathbf{h}_{\mathbf{Z}i}\right) + \mathbf{f}_{\mathbf{Z}} \tag{2.21}$$

$$\tilde{\mathbf{w}}_i = \mathbf{L}_i{}^{\mathbf{T}} \cdot \left(\mathbf{B}_i{}^{\mathrm{T}} \cdot \mathbf{c} + \mathbf{w}_{\mathbf{Z}i}\right) \quad \forall i \tag{2.22}$$

or

$$\mathbf{f} = \sum\nolimits_i \mathbf{B}_i{}' \cdot \tilde{\mathbf{h}}_i + \mathbf{f}_{\mathbf{Z}}{}' \tag{2.23}$$

$$\tilde{\mathbf{w}}_i = \mathbf{B}_i{}'^{\mathrm{T}} \cdot \mathbf{c} + \tilde{\mathbf{w}}_{\mathbf{Z}i} \quad \forall i \tag{2.24}$$

where

$$\mathbf{B}_i{}' = \mathbf{B}_i \cdot \mathbf{L}_i$$
$$\mathbf{f}_{\mathbf{Z}}{}' = \mathbf{B}_i \cdot \mathbf{h}_{\mathbf{Z}i} + \mathbf{f}_{\mathbf{Z}}$$
$$\tilde{\mathbf{w}}_{\mathbf{Z}i} = \mathbf{L}_i{}^{\mathbf{T}} \cdot \mathbf{w}_{\mathbf{Z}i}$$

Furthermore, the demand model (2.11)–(2.13) can be formulated with respect to the i-route variables as:

$$\tilde{\mathbf{v}}_i = -\psi_i\, \tilde{\mathbf{w}}_i \quad \forall i \tag{2.25}$$

$$\tilde{\mathbf{p}}_i = \tilde{\mathbf{p}}_i\left(\tilde{\mathbf{v}}_i; \boldsymbol{\theta}_i\right) \quad \forall i \tag{2.26}$$

$$\tilde{\mathbf{h}}_{i'}' = d_i\, \tilde{\mathbf{p}}_i \quad \forall i \tag{2.27}$$

Eq. (2.27) assures that the sum of the flows of all routes connecting the o-d pair i are upper bounded the demand flow, that is $\mathbf{1}^{\mathbf{T}}\, \tilde{\mathbf{h}}_i \leq d_i$, since $\mathbf{1}^{\mathbf{T}}\, \tilde{\mathbf{p}}_i \leq 1$, and non-negative, $\tilde{\mathbf{h}}_i \geq \mathbf{0}$, since $d_i \geq 0$ and $\tilde{\mathbf{p}}_i \geq \mathbf{0}$.

Eqs (2.23)–(2.27) describe the assignment to uncongested network with respect to arc and i-route variables, they are equivalent to (2.9)–(2.13) since they lead to the same arc flows. This model is useful for some relevant features of the i-route flow function as discussed below, these features will be useful in Chapter 3.

Moreover, this formulation allow to clearly indentify the actual size of the assignment problem, that is the number of independent variables, say the number of i-routes. It is worth stressing that this modelling approach can only be applied with linear utility functions [HYP ⑦ mentioned above] and with route choice functions such that choice proportions depend on differences between systematic utility values only [HYP ⑧ mentioned above].

2.4 Multi class assignment

In this section users are distinguished with respect to o-d pair they are travelling from/to, i, with a common set of routes, called the route choice set, R_i, as in the previous section, but they are also distinguished with the respect the *user class* they belong, j, with specific arc costs, route utility and choice functions, as well as any behavioural parameter. User classes can be used to distinguish users with different socio-economic characteristics, such as age, occupation, household size, …, type of freight, shipping size.

This section presents models for multi-class assignment to uncongested networks assuming that users are distinguished with respect to o-d pair i and user class j with a common set of routes, called the route choice set, R_{ij}, assumed non empty and finite, extending equations presented in the previous section. Main vector notations used in the following are enlisted below in alphabetical order for reader's convenience (sets come first, then Roman letters, after Greek letters); these notations highlight the need of a further block structure of most vectors and matrices introduced in the Section 2.2.

A is the set of arcs, with $m=|A|$;
R_{ij} is the set of (elementary) routes for o-d pair i and user class j, with $n_{ij}=|R_{ij}|$;
$\mathbf{B}_{ij}$ is the $m\times n_{ij}$ block of the *ARGIM* for o-d pair i and user class j;
$\mathbf{c}\geq\mathbf{0}$ is the $m\times 1$ (column) vector of *arc generic costs*, common to all user classes;
$\mathbf{c}^j\geq\mathbf{0}$ is the $m\times 1$ (column) vector of *arc total costs* for user class j;
$\mathbf{c}_\mathbf{S}^j\geq\mathbf{0}$ is the $m\times 1$ (column) vector of *arc specific costs* for user class j, such as additive tolls, …;
$d_{ij}\geq 0$ is the demand flow for o-d pair i and user class j;
$\mathbf{f}\geq\mathbf{0}$ is the $m\times 1$ (column) vector of *total arc flows*;
$\mathbf{f}^j\geq\mathbf{0}$ is the $m\times 1$ (column) vector of *arc flows* of user class j;
$\mathbf{f}_\mathbf{Z}\geq\mathbf{0}$ is the $m\times 1$ (column) vector of *other arc flows*;

$\mathbf{h}_{ij} \geq \mathbf{0}$ is the $n_{ij} \times 1$ (column) block of the vector of *route flows* for o-d pair i and user class j;

$\mathbf{h}_{ij}(\cdot)$ is the $n_{ij} \times 1$ (column) *route flow function* for o-d pair i and user class j;

$\mathbf{p}_{ij} \geq \mathbf{0}$ is the $n_{ij} \times 1$ (column) vector of *route choice proportions* for o-d pair i and user class j, with $\mathbf{1}^{\mathbf{T}}\, \mathbf{p}_{ij} = 1$;

$\mathbf{p}_{ij}(\cdot)$ is the $n_{ij} \times 1$ (column) *route choice function* for o-d pair i and user class j;

$\mathbf{v}_{ij}$ is the $n_{ij} \times 1$ block of the (column) vector of *route systematic utility* for o-d pair i and user class j;

$\mathbf{w}_{ij} \geq \mathbf{0}$ is the $n_{ij} \times 1$block of the (column) vector of *total route costs* for o-d pair i and user class j;

$\mathbf{w}_{\mathbf{Z}ij} \geq \mathbf{0}$ is the $n_{ij} \times 1$ block of the (column) vector of *other route costs* for o-d pair i and user class $j;$

$\eta_{ij} > 0$ is the *flow equivalence parameter* for o-d pair i and user class j;

$\boldsymbol{\theta}_{ij} > 0$ is the vector of the *route choice function parameters* for o-d pair i and user class j;

$\psi_{ij} > 0$ is the *utility scale parameter* in the route utility function for o-d pair i and user class j;

$\chi_{ij} > 0$ is the *cost equivalence parameter* for o-d pair i and user class j, modelling for example different on-board comfort, speeds, ...

Arc and route and demand flows for each user class j are assumed measured by a specific common unit, say users per time unit, the flow equivalent parameters enable to combined them together. Arc and route costs are assumed measured by a common unit, usually travel time or money, through duly homogenization of different attributes not explicit introduced to simplify notations.

2.4.1 Supply model

This section describes the extension of the two equations that according to the FEAM framework specify the transportation supply model for a transportation system in steady-state conditions in the case of an uncongested network, to the case of users distinguished with respect to both o-d pair and user class.

- arc-route flow consistency relation

Under the steady-state assumption the arc flows due to each user class j (and all o-d pairs) can be obtained from the route flows through an affine transformation:

$$\mathbf{f}^j = \sum\nolimits_i \mathbf{B}_{ij} \cdot \mathbf{h}_{ij} \quad \forall j \tag{2.28a}$$

Then the total arc flows can be obtained by summing up over all user classes:

$$\mathbf{f} = \sum\nolimits_j \eta_{ij}\, \mathbf{f}^j + \mathbf{f}_{\mathbf{Z}} \tag{2.28b}$$

where

$\eta_{ij} \geq 0$ is the equivalence flow parameter for o-d pair i and user class j such that all arc flows of any user class are measured in the same unit.

- arc cost function

No arc cost function since the network is uncongested, thus the arc cost vector, **c**, is an input data.

- route-arc cost consistency relation

The arc costs are generally different with respect to the o-d pair i and user class j to reflect different performances and can be defined through an affine transformation of common arc costs:

$$\mathbf{c}^j = \chi_{ij}\,\mathbf{c} + \mathbf{c_S}^j \quad \forall j \tag{2.29a}$$

where

$\mathbf{c_S}^j \geq \mathbf{0}$ is the $A \times 1$ (column) vector of *arc specific costs* for user class j, such as additive tolls, …; these costs may be used for modelling congestion charge tolls, limited access, … differentiated per user class;

$\chi_{ij} \geq 0$ is the equivalence cost parameter for o-d pair i and user class j such that all arc costs of any user class are measured in the same unit.

Under the steady-state assumption the route costs for o-d pair i and user class j can be obtained from the corresponding arc costs through an affine transformation:

$$\mathbf{w}_{ij} = \mathbf{B}_{ij}{}^{\mathrm{T}} \cdot \mathbf{c}^j + \mathbf{w_Z}_{ij} \quad \forall i,j \tag{2.29b}$$

2.4.2 Demand model

This section describes the straightforward extension of the three equations that according to the FEAM (or SEAM) framework specify the transportation demand model with given demand flows (constant demand) to the case of users distinguished with respect to user class too.

- route utility function

The utility function for o-d pair i is assumed specified through a linear transformation of route costs, almost always in research analysis as well as in practical applications:

$$\mathbf{v}_{ij} = -\psi_{ij}\,\mathbf{w}_{ij} \quad \forall i,j \tag{2.30}$$

where $\psi_{ij} > 0$ is the utility scale parameter for o-d pair i and user class j.

- route choice function

Route choice behaviour for o-d pair i and user class j can be described by applying any discrete choice modelling theory so that route choice proportions depend on route systematic utility:

$$\mathbf{p}_{ij} = \mathbf{p}_{ij}\left(\mathbf{v}_{ij}; \boldsymbol{\theta}_{ij}\right) \quad \forall i,j \tag{2.31}$$

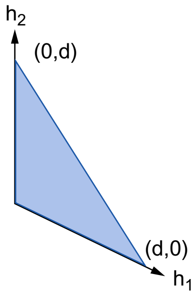


FIG. 2.5

Feasible i-route set, for 2 i-routes.

where $\boldsymbol{\theta}_{ij}$ is the choice function parameter vector for o-d pair i and user class j; see Section 2.2 for details about route choice function specification and features.

- route-demand flow consistency relation

Flow conservation for o-d pair i and user class j can be expressed as:

$$\mathbf{h}_{ij} = d_{ij}\,\mathbf{p}_{ij} \quad \forall i, j \tag{2.32}$$

It assures that flows of all routes connecting the o-d pair i for user class j sum up to demand flow, that is $\mathbf{1}^{\mathbf{T}}\,\mathbf{h}_{ij} = d_{ij}$, since $\mathbf{1}^{\mathbf{T}}\,\mathbf{p}_{ij} = 1$ and non-negative, $\mathbf{h}_{ij} \geq \mathbf{0}$, since $d_{ij} \geq 0$ and $\mathbf{p}_{ij} \geq \mathbf{0}$.

Fig. 2.5 shows the feasible i-route flow set for an o-d pair connected by 2 i-routes, with demand flow d. It is described by a triangle in a 2-dimensional space defined by 2 axis, one for reach i-route flow, $\tilde{h}_1$ and $\tilde{h}_2$, each vertex representing the case of all demand flow, d, using one i-route only.

If the distribution among user classes is given, the above three equations are enough to model the travel demand. On the other hand, if the distribution among user classes is explicitly modelled further equations are needed. This issue is out of the scope of this book.

2.4.3 Reduction to the standard form

The above model for multi-class assignment to uncongested networks can easily be reformulated in a form similar to the single-class case, which is called in the following the *standard form*. At this aim, let

$i' = (i, j)$ denote the combination of o-d pair i and user class j,

$\bar{d}_{i'} = \eta_{ij}\, d_{ij}$,

$\bar{\mathbf{h}}_{i'} = \eta_j\, \mathbf{h}_{ij}$,

$\bar{\mathbf{w}}_{i'} = \left(1/\chi_{ij}\right) \mathbf{w}_{ij}$,

$\bar{\mathbf{w}}_{\mathbf{Z}i'} = \left(1/\chi_{ij}\right)\left(\mathbf{w}_{\mathbf{Z}ij} + \mathbf{B}_{ij}{}^{\mathrm{T}}\,\mathbf{c}_{\mathbf{S}}{}^{j}\right),$

$\bar{\psi}_{i'} = \psi_{ij}\,\chi_{ij},$

the above presented Eqs (2.28)–(2.32) become

$$\mathbf{f} = \sum\nolimits_{i'} \mathbf{B}_{i'} \cdot \bar{\mathbf{h}}_{i'} + \mathbf{f}_{\mathbf{Z}} \tag{2.28$'$}$$

$$\bar{\mathbf{w}}_{i'} = \mathbf{B}_{i'}{}^{\mathrm{T}} \cdot \mathbf{c} + \bar{\mathbf{w}}_{\mathbf{Z}i'} \quad \forall i' \tag{2.29$'$}$$

$$\mathbf{v}_{i'} = -\bar{\psi}_{i'}\,\bar{\mathbf{w}}_{i'} \quad \forall i' \tag{2.30$'$}$$

$$\mathbf{p}_{i'} = \mathbf{p}_{i'}\left(\mathbf{v}_{i'}; \boldsymbol{\theta}_{\bar{i}'}\right) \quad \forall i' \tag{2.31$'$}$$

$$\mathbf{h}_{i'} = \bar{d}_{i'}\,\mathbf{p}_{i'} \quad \forall i' \tag{2.32$'$}$$

formally equal to Eqs (2.9)–(2.13), although some variables have a slightly different meaning; it is worth noting that parameters η_j still play a role, since they are hidden in variables $\bar{d}_{i'}$, whilst parameters χ_{ij}, which in this new formulation cannot be distinguished from parameters $\bar{\psi}_{i'}$, still play a role, since they are hidden in variables $\bar{\mathbf{w}}_{\mathbf{Z}i'}$.

This formulation will be useful in the below Section 2.5 and in Chapters 3–5, since enable to make reference to basic Eqs (2.9)–(2.13) without any loss of generality.

Equations for multi-class assignment can straightforwardly be reformulated with respect to i-route flows and costs, as in the case of single class assignment; details are not reported for brevity's sake.

2.4.4 Multi-vehicle and multi-mode assignment

User classes may also be used to further distinguish users with respect the vehicle type, slow vs. fast cars, traditional vs. advanced cars, … and/or the transportation mode, walk, bicycle, car, … they use to travel. In this case flow and cost equivalence parameters, η_{ij} and χ_{ij}, permit to characterise vehicle type and/or transportation mode.

The degree of occupancy, say the average number of travellers per vehicle, per user class j may be considered included in the flow equivalence parameter η_{ij} with no loss of generality. On the other hand, if the choice proportions among vehicle types and/or transportation modes are explicitly modelled further equations are needed, and these parameters may play a very relevant role. This issue is out of the scope of this book.

2.5 Arc flow function and arc feasible set

In this section the arc flow function is introduced and discussed as a model for the Comprehensive assignment to an Uncongested Network. Main vector notations used in the following are enlisted below in alphabetical order (sets come first, then Roman letters, at last Greek letters) for reader's convenience:

$S_f \subseteq \mathbb{E}^n$ is the feasible arc flow set;
$\mathbf{d}$ is the vector of demand flows, with entries d_i;
$\mathbf{f}(\cdot)$ is the arc flow function.

All equations describing the supply and the demand modes put in standard form, say with formal reference to the basic Eqs (2.9)–(2.13), can be combined together to define the *arc flow function*, that is the relation between arc flows and arc costs:

$$\mathbf{f}(\mathbf{c}; d_i, \psi_i, \boldsymbol{\theta}_i \forall i) \triangleq \sum\nolimits_i d_i \, \mathbf{B}_i \cdot \mathbf{p}_i\left(-\psi_i\left(\mathbf{B}_i^{\mathrm{T}} \cdot \mathbf{c} + \mathbf{w}_{\mathbf{Z}i}; \boldsymbol{\theta}_i\right)\right) + \mathbf{f}_{\mathbf{Z}} \tag{2.33}$$

Thus (omitting parameters for simplicity's sake):

$$\mathbf{f} = \mathbf{f}(\mathbf{c}; \mathbf{d}) \quad \forall \mathbf{c} \geq \mathbf{0} \tag{2.34}$$

Fig. 2.6 shows a data-flow diagram of the arc flow function, highlighting the roles of the main variables.

The very same function is obtained with respect to i-route flows and costs with reference to Eqs (2.23)–(2.27). A formally similar function is obtained in the case of multi-class assignment; all considerations below still hold in any case.

The arc flow function gets values in the *feasible arc flow set:*

$$S_f \triangleq \left\{\mathbf{f} \geq \mathbf{0} : \mathbf{f} = \sum\nolimits_i \mathbf{B}_i \, \mathbf{h}_i + \mathbf{f}_{\mathbf{Z}}, \mathbf{h}_i \geq \mathbf{0} : \mathbf{1}^{\mathrm{T}} \, \mathbf{h}_i = d_i \forall i\right\} \subseteq \mathbb{E}^m$$

where $\mathbb{E}^m$ is the set of real $m \times 1$ (column) vectors with Euclidean distance.

This set is a linear transformations of the route feasible set, thus it:

- has a finite dimension if the number of arcs is finite,
- is non empty if each o-d pair is connected by at least one route,

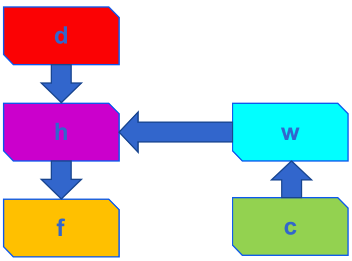


FIG. 2.6

Data-flow diagram of the arc flow function.

- is compact, since closed (or its closure is considered if open) and bounded, the latter if the number of routes available to each o-d pair is finite (as it occurs considering all or some elementary routes),
- is convex.

The *arc flow function* share all features of the route flow functions if the number of routes available to each o-d pair is finite. Accordingly it is defined *regular* if:

- it is continuous and monotone decreasing with respect to arc costs:

$$(\mathbf{f}(\mathbf{c}') - \mathbf{f}(\mathbf{c}''))^{\mathrm{T}} \cdot (\mathbf{c}' - \mathbf{c}'') \leq 0 \quad \forall \mathbf{c}' \neq \mathbf{c}''$$

- it is continuously differentiable with symmetric negative semi-definite (with respect to real vectors) Jacobian, $\nabla \mathbf{f}(\mathbf{c}) \preccurlyeq 0$.

It is worth noting it is not strictly monotone / regular even if all the route choice functions are strictly monotone/regular, thus it is not invertible, and different arc cost vectors giving the same arc flow vector may exist.

Since the arc flow function is defined by the sum over all o-d pairs, it is homogenous of degree 1 with respect to demand flows:

$$\mathbf{f}(\mathbf{c}; \alpha\, \mathbf{d}) = \alpha\, \mathbf{f}(\mathbf{c}; \mathbf{d}) \quad \forall \alpha > 0 \tag{2.35}$$

The arc flow function can easily be computed if route can explicitly be enumerated. Computation algorithms that avoid explicit route enumeration are available for some probabilistic choice functions derived from RUT (and choice functions described in the appendix). Eq. (2.33) implies that arc flows due to each o-d pair i can be computed one by one and then summed up together, independently of the order. These algorithms are usually based on shortest (hyper-)path algorithms, or some extensions of them; they share a similar structure starting from an origin (destination):

- forward (backward) step: arc weights are computed defining how much of the demand will traverse each arc;
- backward (forward) step: back tracing from each destination (forward tracing from each origin) demand flows are added to each arc taking into account arc weights computed in the first step.

In some cases the solution is guaranteed within a finite number of steps, in others Montecarlo techniques are needed that can only provide an unbiased estimation of the searched arc flow vector within a finite number of steps. Details are out of the scope of this book (see Cascetta, 2009).

Table 2.7 shows an example of the arc flow function (2.34), from arc costs to arc flows, summing up Tables 2.2, 2.5, and 2.3. Choice functions is the Logit function in the above example, the utility scale factor ψ is included in dispersion parameter $\theta = 7$; the demand flow is $d = 3600$.

Table 2.7 Arc flow function

arcs	routes 1	routes 2	routes 3	c
1	1	0	0	15
2	0	1	0	12
3	0	0	1	8
4	0	1	1	24
5	1	0	1	15

	1	2	3
w	30	36	37
v	–30	–36	–47
p	0.66	0.28	0.06
h	2380	1010	210

arcs	routes 1	routes 2	routes 3	f
1	1	0	0	2380
2	0	1	0	1010
3	0	0	1	210
4	0	1	1	1220
5	1	0	1	2590

With reference to the route feasibility set shown in Fig. 2.4, Fig. 2.7 shows the effects on the route flows **h** of doubling or halving the dispersion parameter θ, that is θ ∈ {14.0, 7.0, 3.5}. [As already noted it includes the utility scale factor ψ].

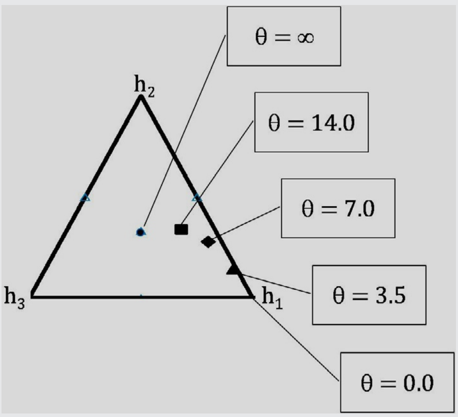


FIG. 2.7

Route flows with θ = ∞ (●), 14.0 (■), 7.0 (♦), 3.5 (▲).

As expected increasing θ move **h** towards the conditions of users uniformly spreading among all routes, that is all route flows being equal to one third of demand flow, *d*/3, say θ = ∞ (350 is a great enough value in this case). On the other hand, decreasing θ move **h** towards the conditions of all users concentrating on the shortest route 1, say θ=0 (0.7 is a small enough value in this case).

In this examples, as in the following chapters, we make reference to route flows since they can described in a plane figure, cfr Fig. 2.4, but any computation can be carried out without explicit enumeration of routes, as already stated.

2.6 Summary

2.6.1 Major findings

This chapter presented several models for Comprehensive assignment to Uncongested Networks, within Five Equation Assignment Modelling (FEAM) framework, obtained from the SEAM framework dropping the equation relative to the arc cost function. All of them can be reduce to the standard form and then to the arc flow function. Results hold under any assumptions leading to a Transpose Affine Network (TAN), being synchronic or diachronic.

The proposed approach is very general, can easily extended under other assumptions, and enables to specify models for assignment to congested networks, as shown in the next chapters. It should be remarked that optimization models are available for some very particular instances of assignment to uncongested networks, but they cannot be generalised, nor used for general models for assignment to congested networks (see appendix to Chapter 5 in Cascetta, 2009, for a review).

All parameters introduced above are to be calibrated against real/simulated data, this relevant issues is out the scope of this book. As already stated implementation and application issues are out of the scope of this book, mainly focusing on mathematical features. (For details on these issues see Cascetta (2009). A companion book (possibly by other authors) discussing these topics is under planning.

2.6.2 Further readings

Hyperpaths and their relationship with pre-trip en-route route choice strategies have been introduced by Nguyen and Pallottino (1988) and Spiess and Florian (1989). For details on schedule-based assignment to diachronic networks see the recent book edited by Gentile and Nökel (2016), even though the contents mostly refer to deterministic assignment only (see below). Some references on assignment with fuzzy utility see next chapter.

2.6.3 Remarks

Some approaches to assignment assume that all sources of uncertainty are negligible, thus all users travelling between o-p pair i follow maximum utility routes, and do not use at all any of the other routes. This user choice behaviour assumption (cfr Wardop, 1952) may be obtained from Deterministic Utility Theory, or as a limit of Random/Fuzzy Utility Theory when dispersion goes to zero, as well as from the Expected Utility Theory. (See Appendix A for more details.) This approach leads to the so-called All-or-Nothing (AoN) assignment for uncongested networks.

In this case, the route choice function (2.12) is actually a multi-valued function (also called a one-to-many or a point-to-set function or a map), since it may well be the case that the systematic utility values of two or more routes are equal to the maximum. Thus the arc function (2.34) turns out a multi-valued function as well.

A different approach, often followed to avoid this kind of functions, is useful for equilibrium assignment to congested networks described in this chapter. Let

$\mathbf{p}_{\mathbf{D}i}(\mathbf{v}_i) \geq \mathbf{0}$ be any of the route deterministic choice proportion vectors corresponding to systematic utility vector $\mathbf{v}_i$ for o-d pair i with $\mathbf{1}^{\mathbf{T}}\ \mathbf{p}_{\mathbf{D}i}(\mathbf{v}_i) = 1$; for any route r

if $v_{ir} < v_{imax}$ then $p_{ir} = 0 \Leftrightarrow$ if $p_{ir} > 0$ then $v_{ir} = v_{imax}$

[The case $p_{ir} = 0$ with $v_{ir} = v_{imax}$ it is not ruled out by this condition.]

From the above condition for any systematic utility vector v_i $\mathrm{p}_{\mathrm{D}i} = \mathrm{p}_{\mathrm{D}i}(\mathrm{v}_i)$ is equivalent to the following condition (see Appendix A for more details):

$$\mathbf{v}_i^{\mathbf{T}} \cdot (\mathbf{p}_{\mathbf{D}i} - \mathbf{q}_i) \geq 0\ \forall \mathbf{q}_i \geq \mathbf{0} \quad \text{with} \quad \mathbf{1}^{\mathbf{T}}\ \mathbf{q}_i = 1 \tag{2.36}$$

If condition (2.37), instead of Eq. (2.12), is combined with Eqs (2.9), (2.10), (2.11) and (2.13) the following condition is obtained, equivalent to the arc flow function with deterministic route choice behaviour:

$$\mathbf{c}^{\mathbf{T}} \cdot (\mathbf{f} - \mathbf{f}_{\mathbf{D}}) \geq 0\ \forall \mathbf{f} \in S_f \tag{2.37}$$

where $\mathbf{f}_{\mathbf{D}}$ is any of the arc flow vectors corresponding to arc cost vector $\mathbf{c}$. [Other flows and other costs are omitted to avoid awkward equations.] The (linear variational) inequality (2.37) is equivalent to the following linear optimization model:

$$\mathbf{f}_{\mathbf{D}} = \operatorname{argmin}_{\mathbf{f} \in Sf}\ \mathbf{c}^{\mathbf{T}} \cdot \mathbf{f} \tag{2.38}$$

See remarks at the end of the next chapter for further comments.

References

Bifulco, G.N., 1993. A stochastic user equilibrium assignment model for the evaluation of parking policies. Eur. J. Oper. Res. 71, 269–287.

Cascetta, E., 2009. Transportation Systems Analysis: Models and Applications. Springer.

Gentile, G., Nökel, K., 2016. Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems: COST Action TU1004 (TransITS). Springer.

Nguyen, S., Pallottino, S., 1988. Equilibrium traffic assignment for large scale transit networks. Eur. J. Oper. Res. 37, 176–186.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods. Prentice-Hall, Inc, Englewood Cliffs.

Spiess, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. Transp. Sci. 23, 83–102.

Wardop, J.G., 1952. Some theoretical aspects of road traffic research. Proc. Inst. Civ. Eng. 2 (1), 325–378.

CHAPTER

Assignment to congested networks: User equilibrium—Fixed points

3

Giulio Erberto Cantarella
University of Salerno, Salerno, Italy

Phil: Do you know what today is?
Rita: No, what?
Phil: Today is tomorrow. It happened.
from movie Groundhog Day.

Outline. This chapter describes a comprehensive modelling approach to steady-state user equilibrium assignment to congested networks through Fixed-Point (FP) models; presented models are consistent with the SEAM modelling framework presented in Chapter 1; first the route cost and flows functions are also introduced and discussed, then fixed-point models with respect to flows and/or costs are introduced and discussed.

As already noted, methods for travel demand assignment to transportation networks play a central role in transportation system analysis, they allow to compute flows and costs for each supply element, resulting from origin-destination demand flows, user choice behaviour, congestion and their interactions in any scenario.

This chapter discusses Fixed-Point (FP) models for steady-state equilibrium assignment to congested transportation networks, one of the most used kind of assignment, implemented in several commercial software applications. They will turn out a special case of the Deterministic Process models described in Chapter 4.

User Equilibrium assignment searches for mutually consistent arc flows and costs. It was first introduced under steady-state conditions by Wardrop (1952), who named it User Equilibrium (UE), following a modelling approach to route choice behaviour that we may now consider an application of the Deterministic Utility Theory. A more general kind of equilibrium based on application of the RUT was introduced by Daganzo and Sheffi (1977), who named it Stochastic User Equilibrium

Dynamics and Stochasticity in Transportation Systems. https://doi.org/10.1016/B978-0-12-814353-7.00003-0
© 2020 Elsevier Inc. All rights reserved.

(SUE). The book by Sheffi (1985) provided several optimisation models for both UE and SUE. Afterwards Daganzo (1983) introduced fixed-point models using the inverse cost function. A general and flexible framework was proposed by Cantarella (1997), still based on FP models, but without the need of the inverse cost function (see also Chapters 5 and 6 in Cascetta, 2009).

This chapter introduces and discusses a comprehensive fixed-point modelling approach to equilibrium assignment to congested networks, including all those mentioned above as well as all those that result from most route choice modelling approaches (cfr Appendix A1 to the book). Hence it is named Comprehensive User Equilibrium assignment to congested networks (CUE). It is described for steady-state conditions, but it also applies to any TAN. Presented models are consistent with the SEAM modelling framework presented in Chapter 1, leading to fixed-point models.

A special case occur assuming that all users follow a maximum utility or minimum cost routes, this kind of assignment is most often called just User Equilibrium (UE) assignment (see above) and this denomination is followed in this book; it may be derived from several theories, such as Deterministic Utility Theory, Expected Utility Theory, or as a limit case of Random or Fuzzy Utility Theory (see Remarks at the end of this chapter for details and comments).

Section 3.1 introduces basic equations. Section 3.2 discusses fixed-point models. Existence and basic uniqueness conditions, solution algorithms, and conditions for their convergence are presented (implementation issues are not discussed); advanced uniqueness and convergence conditions are discussed in Section 3.3.

3.1 Basic equations

This section presents the basic equations for (within-day static) user equilibrium assignment adding the arc cost function to the five equations in the standard form introduced in the previous Chapter 2, reference is made to Eqs (2.9)–(2.13) and (2.23)–(2.27), but the presented approach can straightforwardly be applied to multi-class assignment (2.29)–(2.32) as well; all definitions and assumptions introduced in the previous chapter still hold, unless otherwise stated. Main vector notations from Chapter 2 as well as few new ones used in the following are enlisted below in alphabetical order for reader's convenience (sets come first, then Roman letters, at last Greek letters).

A is the set of arcs, with $m = |A|$ elements;
$\mathbb{E}^m$ is the set of real $m \times 1$ (column) vectors with Euclidean distance;
m is the number of arcs;
n_i is the number of routes connecting o-d pair i;
$n = \sum_i n_i$ is the number of routes connecting all o-d pairs;

$\tilde{n}_i = n_i$ - 1 is the number of i-routes connecting o-d pair i;
$\tilde{n} = \sum_i \tilde{n}_i$ is the number of i-routes connecting all o-d pairs;
R_i is the set of routes for o-d pair i, with $n_i = | R_i |$ elements;
$R = \cup_i R_i$ is the set of routes for all o-d pairs, with $n = | R | = \sum_i n_i$ elements;
S_f is the feasible arc flow set;
S_{hi} is the feasible route flow set for o-d pair i;
S_h is the feasible route flow set;
$S_{\tilde{h}}$ is the feasible i-route flow set;

$\mathbf{B}$ is the $(m \times n)$ *ARGIM*;
$\mathbf{B}_i$ is the $(m \times n_i)$ block of the *ARGIM* for o-d pair i;
$\mathbf{c} \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *arc costs*;
$\mathbf{c}(\cdot)$ is the $m \times 1$ (column) *arc cost function*;
$d_i \geq 0$ is the *demand flow* for o-d pair i;
$\mathbf{d} \geq \mathbf{0}$ is the *demand flow* vector with entries d_i;
$\mathbf{f} \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *total arc flow*s;
$\mathbf{f_Z} \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *other arc flows*;
$\mathbf{f}(\cdot)$ is the $m \times 1$ (column) *arc flow function*;
$\mathbf{h} \geq \mathbf{0}$ is the $n \times 1$ (column) vector of *route flow*s for all o-d pairs;
$\mathbf{h}_i \geq \mathbf{0}$ is the $n_i \times 1$ (column) block of the vector of *route flow*s for o-d pair i;
$\mathbf{h}_i(\cdot)$ is the $n_i \times 1$ (column) *route flow function* for o-d pair i;
$\mathbf{h}(\cdot)$ is the $n \times 1$ (column) *route flow function* for all o-d pairs;
$\tilde{\mathbf{h}}$ is the $\tilde{n} \times 1$ (column) vector of i-*route flow*s for all o-d pairs;
$\tilde{\mathbf{h}}(\cdot)$ is the $\tilde{n} \times 1$ (column) *route flow function* for all o-d pairs;
$\mathbf{p}_i \geq \mathbf{0}$ is the $n_i \times 1$ (column) block of the vector of *route choice proportions* for o-d pair i, with $\mathbf{1}^{\mathbf{T}} \, \mathbf{p}_i = 1$;
$\mathbf{p}_i(\cdot)$ is the $n_i \times 1$ (column) *route choice function* for o-d pair i;
$\mathbf{v}_i$ is the $n_i \times 1$ block of the (column) vector of *route systematic utility* for o-d pair i;
$\mathbf{w}_i \geq \mathbf{0}$ is the $n_i \times 1$ block of the (column) vector of *total route costs* for o-d pair i;
$\mathbf{w} \geq \mathbf{0}$ is the $n \times 1$ (column) vector of *total route costs*;
$\mathbf{w_{Zi}} \geq \mathbf{0}$ is the $n_i \times 1$ block of the (column) vector of *other route costs* for o-d pair i;
$\mathbf{w_Z} \geq \mathbf{0}$ is the $n \times 1$ (column) vector of *other route costs* for all o-d pairs;
$\mathbf{w}(\cdot)$ is the $n \times 1$ (column) *route cost function* for all o-d pairs;
$\mathbf{w}_i(\cdot)$ is the $n_i \times 1$ (column) *route cost function* for o-d pair i;
$\tilde{\mathbf{w}}$ is the $\tilde{n} \times 1$ (column) vector of i-*route cost*s for all o-d pairs;
$\tilde{\mathbf{w}}(\cdot)$ is the $\tilde{n} \times 1$ (column) i-*route cost function* for all o-d pairs;
$\boldsymbol{\theta}_i > 0$ is the vector of the *route choice function parameters* for o-d pair i;
$\kappa_a > 0$ is the *capacity* of arc a;
$\boldsymbol{\kappa} > 0$ is the $m \times 1$ (column) vector of the *arc capacities*, with entries κ_a;
$\psi_i > 0$ is the *utility scale parameter* in the route choice model, for o-d pair i.

3.1.1 Supply models

Transportation supply models express how user behaviour affects network performances. This section describes the three equations that according to the SEAM framework specify the transportation supply model for a transportation system in steady-state conditions in the case of a congested network.

- arc-route flow consistency relation

Under the steady-state assumption the total arc flows due to all o-d pairs can be obtained from the route flows through an affine transformation from the route space to the arc space defined by the arc-route generalised incidence matrix:

$$\mathbf{f} = \sum_i \mathbf{B}_i \cdot \mathbf{h}_i + \mathbf{f_Z} \tag{3.1}$$

or

$$\mathbf{f} = \mathbf{B} \cdot \mathbf{h} + \mathbf{f_Z}$$

- arc cost function

Due to congestion, say driving user behaviour, arc costs depend on the arc total flows:

$$\mathbf{c} = \mathbf{c}(\mathbf{f}; \boldsymbol{\kappa}) \geq \mathbf{0} \quad \forall \mathbf{f} \in S_f \tag{3.2}$$

where

$S_f \triangleq \{\mathbf{f} \geq \mathbf{0}: \mathbf{f} = \sum_i \mathbf{B}_i\, \mathbf{h}_i + \mathbf{f_Z}, \mathbf{h}_i \geq \mathbf{0}: \mathbf{1}^{\mathbf{T}}\, \mathbf{h}_i = d_i\, \forall i\} \subseteq \mathbb{E}^m$ is the feasible arc flow set; this set, as already stated in Chapter 2,

- has a finite dimension if the number of arcs is finite,
- is non empty if each o-d pair is connected by at least one route,
- is compact, since closed and bounded, the latter if the number of routes available to each o-d pair is finite (as it occurs considering all or some elementary routes),
- is convex;

$\boldsymbol{\kappa} > 0$ is the vector of the *arc capacities*, with entries κ_a, say the maximum flow that may traverse arc a, measured consistently with arc flows; in most functions the arc cost actually depends on the ratio between the arc flow and the capacity, f_a / κ_a, in this case the capacity plays the role of arc flow scale factor.

Other parameters of the arc cost function are not explicitly introduced. [Function (3.2) is considered a vector function, thus singular is used; sometime plural "arc cost functions" is used to stress that each arc has its own cost function.] The cost function is called:

- *separable*, the cost of each arc a, c_a, only depends on the corresponding flow, f_a,
- *non-separable*, the cost of at least an arc a, c_a, depends on the flow of another arc, $f_{a'}$.

Most often cost functions are derived from Traffic Flow Theory for arcs modelling moving along a street (a railway, an airway, …) or from Queuing Theory for arcs modelling waiting at a bottleneck, such as a junction approach (see Appendix A).

All usually adopted cost functions can be assumed:

- continuous over the set of feasible arc flows S_f; this assumption implies that the function may be defined for values of flow greater than capacity too, and no vertical asymptote is present, according to within-day static assumption;

- continuously differentiable with respect to arc flows over the set of feasible arc flows S_f; [or better over a suitable (open) superset of S_f such that all point of S_f are interior points of it], with Jacobian matrix $\mathbf{J_C}(\mathbf{f}) = \nabla \mathbf{c}(\mathbf{f})$, thus $\mathbf{J_C}(\mathbf{f}) = \mathbf{0}$ means that the network is uncongested, and $\mathbf{J_C}(\mathbf{f}) \neq \mathbf{0}$ that it is congested (at least for some arcs); a continuously differentiable arc cost function is:
  - *separable*, if $\mathbf{J_C}(\mathbf{f})$ is a diagonal matrix;
  - *non-separable*, if $\mathbf{J_C}(\mathbf{f})$ otherwise; in this case it is useful to distinguish two cases:
    - o $\mathbf{J_C}(\mathbf{f})$ is symmetric,
    - o $\mathbf{J_C}(\mathbf{f})$ is asymmetric.

In the following $\mathbf{M} \succ 0$ ($\succeq 0$) means that matrix $\mathbf{M}$ is positive (semi-)definite, $\mathbf{M} \prec 0$ ($\preccurlyeq 0$) that is negative (semi-)definite, with respect to real vectors, thus matrix $\mathbf{M}$ may be not symmetric. Another relevant feature is monotonicity:

- the arc cost function is strictly increasing monotone, a sufficient condition for this is that the Jacobian matrix is positive definite (for real vectors), $\mathbf{J_C}(\mathbf{f}) \succ 0$, but not necessarily symmetric:

$$(\mathbf{c}(\mathbf{f}') - \mathbf{c}(\mathbf{f}''))^{\mathbf{T}} \cdot (\mathbf{f}' - \mathbf{f}'') > 0 \quad \forall \mathbf{f}' \neq \mathbf{f}''$$

- the arc cost function is increasing monotone, a necessary and sufficient condition for this is that the Jacobian matrix is positive semi-definite (for real vectors), $\mathbf{J_C}(\mathbf{f}) \succeq 0$, but not necessarily symmetric:

$$(\mathbf{c}(\mathbf{f}') - \mathbf{c}(\mathbf{f}''))^{\mathbf{T}} \cdot (\mathbf{f}' - \mathbf{f}'') \geq 0 \quad \forall \mathbf{f}' \neq \mathbf{f}''$$

This feature can easily be checked for separable arc cost functions, and strict monotonicity holds for all usually adopted separable cost functions. On the other hand, this may not be the case for non-separable cost functions.

An example of separable arc cost functions is given by the often used BPR-like function (Fig. 3.1 and Table 3.1):

$$c_a = c_{o,a}\left(1+\nu_1\left(f_a/\kappa_a\right)^{\nu}{}_2\right).$$

where

$\nu_1 > 0$ is the congestion multiplier: how much greater is the arc cost when flow is equal to capacity with respect to zero flow cost; this parameter is to be calibrated against data, in urban applications 2 is an often used values [in the original BPR function it was set to 0.15 for extra urban highways];

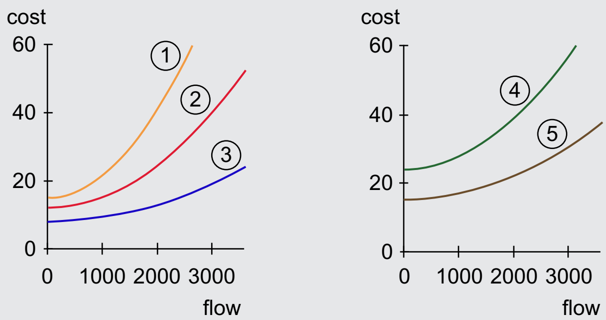


FIG. 3.1

Arc cost functions.

Table 3.1 Parameters of arc cost functions; $\nu_2 = 2$ for all arcs.

a		$C_{o,a}$	κ_a	ν_{1a}
1	AC	15	2400	2.5
2	BD	12	2400	1.5
3	BC	8	3600	2.0
4	AB	24	3600	2.0
5	CD	15	2400	1.5

$\nu_2 > 0$ is the congestion exponent: how fast the arc cost increases against flow; this (integer) parameter too is to be calibrated against data, in urban applications 2 is an often used values; as this value increases the shape of the function tends to a vertical asymptote [in the original BPR function it was set to 4];

$c_{o,a} > 0$ is the cost when flow is zero.

- route-arc cost consistency relation

Under the steady-state assumption the route costs for o-d pair i, can be obtained from the arc costs through an affine transformation from the arc space to the route space defined by the transpose of arc-route generalised incidence matrix:

$$\mathbf{w}_i = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c} + \mathbf{w}_{\mathbf{Z}i} \qquad \forall i \tag{3.3}$$

or

$$\mathbf{w} = \mathbf{B}^{\mathbf{T}} \cdot \mathbf{c} + \mathbf{w}_{\mathbf{Z}}.$$

- route cost function

The three Eqs (3.1)–(3.3) describing the supply model can be combined to define the *route cost function* for each o-d pair i which expresses the relation between route costs and route flows, say how user choice behaviour affects network performances:

$$\mathbf{w}_i(\mathbf{h};\, \boldsymbol{\kappa}) \triangleq \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}(\mathbf{B} \cdot \mathbf{h} + \mathbf{f}_{\mathbf{Z}};\, \boldsymbol{\kappa}) + \mathbf{w}_{\mathbf{Z}i} \forall i.$$

that are blocks of the block vector function: $\mathbf{w}(\mathbf{h};\, \boldsymbol{\kappa})$, thus.

$$\mathbf{w} = \mathbf{w}(\mathbf{h};\, \boldsymbol{\kappa}) \geq \mathbf{0} \qquad \forall \mathbf{h} \in S_h \tag{3.4}$$

where

n_i is the number of routes connecting o-d pair i;
$n = \sum_i n_i$ is the number of routes connecting all o-d pairs;
$S_{hi} \triangleq \{\mathbf{h}_i \geq \mathbf{0}\colon \mathbf{1}^{\mathbf{T}}\, \mathbf{h}_i = d_i\} \subseteq \mathbb{E}^{n_i}$ is the *feasible route flow set* for o-d pair i; according to route-demand flow consistency Eq. (2.13) in Chapter 2, repeated below as (3.8), it (see Fig. 2.3)

- has a finite dimension if the number of routes available to o-d pair i is finite (as it occurs considering all or some elementary routes),
- is non empty if o-d pair i is connected by at least one route,
- is compact, since closed and bounded [in the Euclidean space],
- is convex.

$S_h \triangleq \{\mathbf{h}$ with blocks $\mathbf{h}_i \in S_{hi}\} = \{\mathbf{h}$ with blocks $\mathbf{h}_i \geq \mathbf{0}\colon \mathbf{1}^{\mathbf{T}}\, \mathbf{h}_i = d_i\} \subseteq \mathbb{E}^n$ is the *feasible route flow set*, with same features of sets S_{hi} since their number is finite.

The route cost function share most of the features of the arc cost function, but generally is increasing monotone, with a positive semi-definite Jacobian, $\nabla \mathbf{w}(\mathbf{h}) \succeq 0$, both for increasing or strictly increasing arc cost functions.

Similar equations can be defined with respect to i-route variables (details are omitted) leading to the *i-route cost function*:

$$\mathbf{w} = \mathbf{w}(\mathbf{h};\, \boldsymbol{\kappa}) \geq \mathbf{0} \quad \forall \mathbf{h} \in S_h \tag{3.5}$$

where

$\tilde{n}_i = n_i$ - 1 is the number of i-routes connecting o-d pair i;
$\tilde{n} = \sum_i \tilde{n}_i$ s the number of i-routes connecting all o-d pairs;
$S_{\tilde{h}} \triangleq \{\tilde{\mathbf{h}}$ with blocks $\tilde{\mathbf{h}}_i \geq \mathbf{0}$: $\mathbf{1}^{\mathbf{T}}\, \tilde{\mathbf{h}}_i \leq d_i\} \subseteq \mathbb{E}\tilde{n}$ is the *feasible i-route flow set*, it

- has a finite dimension if the number of routes available to each o-d pair i is finite (as it occurs considering all or some elementary routes),
- is non empty if each o-d pair i is connected by at least one route,
- is compact, since closed and bounded [in the Euclidean space
- is convex,
- has interior points.

The route cost function share most of the features of the arc cost function, but generally is increasing monotone, with a negative semi-definite Jacobian, $\nabla \tilde{\mathbf{w}}(\tilde{\mathbf{h}}) \preccurlyeq 0$, both for increasing or strictly increasing arc cost function.

3.1.2 Demand models

Travel demand models express how network performances affect user choice behaviour. This section first describes the three equations that according to SEAM framework specify the travel demand model as already introduced in the previous chapter and repeated here for reader's convenience.

- route utility function

The utility function for o-d pair i is assumed specified through a linear transformation of route costs, almost always in research analysis as well as in practical applications:

$$\mathbf{v}_i = -\psi_i\, \mathbf{w}_i \quad \forall i \tag{3.6}$$

where $\psi_i > 0$ is the utility scale parameter, such that the term $\psi\, \mathbf{w}_i$ is dimensionless to be consistent with utility unit.

- route choice function

Route choice behaviour for o-d pair i can be described by applying any discrete choice modelling theory (see appendix A2 to the book) thus route choice proportions depend on route systematic utility:

$$\mathbf{p}_i = \mathbf{p}_i(\mathbf{v}_i; \boldsymbol{\theta}_i) \quad \forall i \tag{3.7}$$

where $\boldsymbol{\theta}_i$ is the choice function parameter vector, whose meaning depends on the choice model specification. If a utility scale parameter is present, it is considered included in the utility parameter ψ_i (or vice versa).

- route-demand flow consistency relation

Flow conservation for o-d pair i can be expressed as:

$$\mathbf{h}_i = d_i\,\mathbf{p}_i \quad \forall i \tag{3.8}$$

It assures that flows of all routes connecting the o-d pair i sum up to demand flow, that is $\mathbf{1}^{\mathbf{T}}\,\mathbf{h}_i = d_i$, since $\mathbf{1}^{\mathbf{T}}\,\mathbf{p}_i = 1$, and non-negative, $\mathbf{h}_i \geq \mathbf{0}$, since $d_i \geq 0$ and $\mathbf{p}_i \geq \mathbf{0}$.

- route flow function

The three Eqs (3.6)–(3.8) describing the demand model can be combined to define the *route flow function* for o-d pair i, which expresses the relation between route flows and route costs, say how network performances affect user choice behaviour:

$$\mathbf{h}_i(\mathbf{w}_i; d_i, \psi, \boldsymbol{\theta}_i) \triangleq d_i\,\mathbf{p}_i(-\psi\,\mathbf{w}_i; \boldsymbol{\theta}_i) \quad \forall i$$

that are blocks of the block vector function (omitting parameters): $\mathbf{h}(\mathbf{w};\,\mathbf{d})$, thus.

$$\mathbf{h} = \mathbf{h}(\mathbf{w}; \mathbf{d}) \in S_h \quad \forall \mathbf{w} \geq \mathbf{0} \tag{3.9}$$

with values in set S_h introduced above. Since demand flows are non-negative the route flow function has the same features of the route choice proportion functions; in particular if each of them is regular, that is.

- it is continuous and monotone increasing with respect to systematic utility values,
- it is continuously differentiable with symmetric negative semi-definite (with respect to real vectors) Jacobian,

the route flow function is regular:

- it is continuous and monotone decreasing with respect to route costs,
- it is continuously differentiable with symmetric negative semi-definite (with respect to real vectors) Jacobian, $\nabla\,\mathbf{h}(\mathbf{w}) \preccurlyeq 0$.

Similar equations can be defined with respect to i-route variables (details are omitted) leading to the *i-route flow function*:

$$\tilde{\mathbf{h}} = \tilde{\mathbf{h}}(\tilde{\mathbf{w}}; \mathbf{d}) \in S_{\tilde{h}} \quad \forall \tilde{\mathbf{w}} \geq \mathbf{0} \tag{3.10}$$

with values in set $S_{\tilde{h}}$ introduced above.

Since demand flows are non-negative the i-route flow function features can easily be derived from those of the i-route choice proportion functions; in particular if each of them is strictly regular, that is.

- it gives strictly positive choice proportions for any values of systematic utility,
- it is continuous, continuously differentiable and monotone strictly increasing with respect to systematic utility differences,
- it has symmetric negative definite (with respect to real vectors) Jacobian, $\nabla \tilde{\mathbf{h}}(\tilde{\mathbf{w}}) \prec 0$,
- resulting choice proportions depend on systematic utility differences only,

the i-route flow function is strictly regular:

- it is continuous and monotone strictly decreasing with respect to i-route costs,
- it is continuously differentiable with symmetric negative definite (with respect to real vectors) Jacobian,
- resulting route flows depend on i-route costs, say on differences between route costs only;

moreover, in this case the feasible i-route flow set is an open set with interior points only, $S_{\tilde{h}} \triangleq \{\tilde{\mathbf{h}}$ with blocks $\tilde{\mathbf{h}}_i > \mathbf{0}$: $\mathbf{1}^{\mathbf{T}}\ \tilde{\mathbf{h}}_i < d_i\} \subseteq E\tilde{n}$, as already noted above.

3.2 Fixed-point models for equilibrium assignment

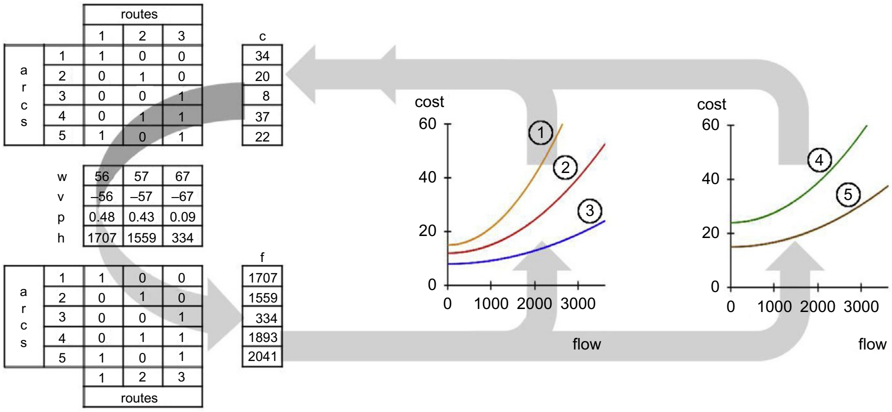
The set of six Eqs (3.1)–(3.3) and (3.6)–(3.8) defines a fixed-point (FP) model with respect to all the six basic variables, describing the comprehensive user equilibrium (CUE) state, as a consistent condition between costs and flows.

Fixed-points

Let $\boldsymbol{\varphi}(\mathbf{x})$ be a vector function from set S to set $\boldsymbol{\varphi}(S)$,
any point $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*) \in S$ is a fixed-point of this function.

Fig. 3.2 shows a data-flow diagram of the fixed-point model (3.1)–(3.3) and (3.6)–(3.8), highlighting the roles of the main variables. The loop between flows and costs is a graphical illustration of the fixed-point consistency. This figure also highlights that a fixed-point model for CUE can be obtained by combining together the arc flow function. Which models the comprehensive assignment to uncongested networks (CUN), and the arc cost function, which models the congestion due to users sharing the same transportation facility; formal details are given in Section 3.2.2 below.

Table 3.2 Equilibrium assignment.



arcs \ routes	1	2	3	c
1	1	0	0	34
2	0	1	0	20
3	0	0	1	8
4	0	1	1	37
5	1	0	1	22

	1	2	3
w	56	57	67
v	–56	–57	–67
p	0.48	0.43	0.09
h	1707	1559	334

arcs \ routes	1	2	3	f
1	1	0	0	1707
2	0	1	0	1559
3	0	0	1	334
4	0	1	1	1893
5	1	0	1	2041

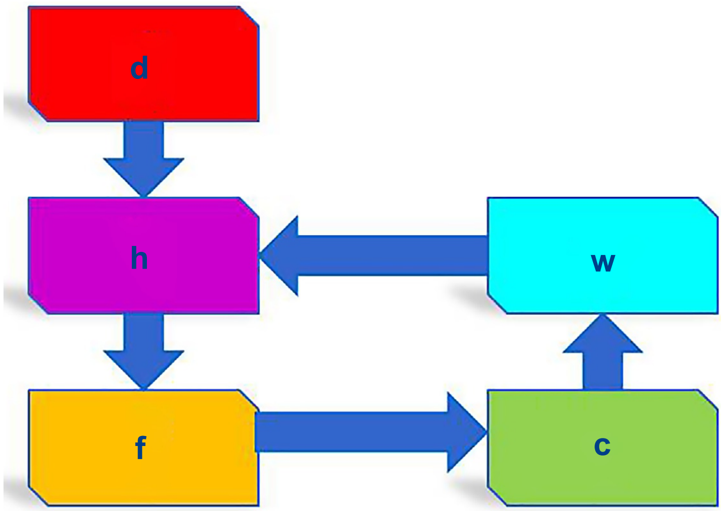


FIG. 3.2

Data-flow diagram of the Fixed-Point models for equilibrium assignment.

Table 3.2 shows an example of the equilibrium assignment obtained by combining the arc cost functions in Fig. 3.1 with the arc flow function in Table 2.7. Choice functions is the Logit function, the utility scale factor ψ is included in dispersion parameter $\theta = 7$; the demand flow is $d = 3600$, as in Table 2.7.

Continued

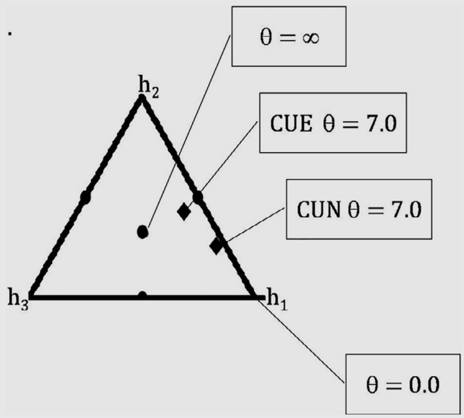


FIG. 3.3

CUE vs. CUN route flows.

With reference to the route feasibility set shown in Fig. 2.4, Fig. 3.3 compares the CUE and the CUN route flow patterns. As expected they are quite different; indeed, the effect of congestion is quite relevant thus routes 2 and 3 (and arcs 2 and 3) are more used in the CUE pattern. Still when the dispersion parameter goes to infinity, $\theta = \infty$ (350 is a great enough value in this case), the two flow patterns are equal since all route flows are one third of demand flow, *d*/3, in both cases.

In this examples, as in the following chapters, we make reference to route flows since they can described in a plane figure, cfr Fig. 2.4, but any computation can be carried out without explicit enumeration of routes, as already stated.

To further analyse the model it is better to reduce the number of equations and variables, as shown in the following. In any case all described models are equivalent since they provide the same solution(s); nevertheless as discussed below, each of them may be useful for different purposes.

3.2.1 Two equation assignment models

Fixed-point models described below are made by two equations with respect to two variables, a flow vector and a cost vector, or Two Equation Assignment Models (TEAMs), since only two equations of the above discussed six equations are non-linear: the arc cost function and the route choice function.

- route costs and flows

Models based on route costs and flows are made by Eq. (3.4) describing the supply model and (3.9) the demand model, repeated below for reader's convenience:

$$\mathbf{w}^* = \mathbf{w}(\mathbf{h}^*; \boldsymbol{\kappa}) \geq \mathbf{0} \tag{3.11}$$

$$\mathbf{h}^{*}=\mathbf{h}(\mathbf{w}^{*};\mathbf{d})\in S_{h} \tag{3.12}$$

These models are useful as the base for developing general day-to-day dynamic process models described in Chapter 4.

- i-route costs and flows

Models based on i-route costs and flows are made by Eq. (3.5) describing the supply model and (3.10) the demand model, repeated below for reader's convenience:

$$\widetilde{\mathbf{w}}^{*}=\widetilde{\mathbf{w}}\left(\widetilde{\mathbf{h}}^{*};\boldsymbol{\kappa}\right)\geq\mathbf{0} \tag{3.13}$$

$$\widetilde{\mathbf{h}}^{*}=\widetilde{\mathbf{h}}(\widetilde{\mathbf{w}}^{*};\mathbf{d})\in S_{\widetilde{h}} \tag{3.14}$$

These models are useful for the specification of solution algorithms based on explicit route enumeration. Their main usefulness is for advanced uniqueness analysis in Section 3.3. It is worth stressing that this modelling approach can only be applied with linear utility functions [HYP⑦ in Chapter 2.] and with route choice functions such that choice proportions depend on differences between systematic utility values only [HYP⑧ in Chapter 2.].

- arc costs and flows

Fixed-point models with respect to arc flows and/or costs are the most used for specifying basic uniqueness conditions as well as solution algorithms. All of them are based on the arc cost function, $\mathbf{c}=\mathbf{c}(\mathbf{f};\boldsymbol{\kappa})$, (3.2) and the arc flow function, $\mathbf{f}=\mathbf{f}(\mathbf{c};\mathbf{d})$, introduced in Chapter 2, as a model for assignment to uncongested networks:

$$\mathbf{c}^{*}=\mathbf{c}(\mathbf{f}^{*};\boldsymbol{\kappa})\geq\mathbf{0} \tag{3.15}$$

$$\mathbf{f}^{*}=\mathbf{f}(\mathbf{c}^{*};\mathbf{d})\in S_{f} \tag{3.16}$$

where $\boldsymbol{\kappa}$ is the arc capacity vector, $\mathbf{d}$ the demand flow vector and $S_f\subseteq\mathbb{E}^m$ is the feasible arc flow set, which.

- has a finite dimension if the number of arcs is finite,
- is non empty if each o-d pair is connected by at least one route,
- is compact, since closed and bounded,
- is convex.

If all the route choice functions, and the route flow functions as well, are regular, the arc flow function $\mathbf{f}=\mathbf{f}(\mathbf{c};\mathbf{d})\in S_f$ is regular:

- it is continuous and monotone decreasing with respect to arc costs,

$$(\mathbf{f}(\mathbf{c}')-\mathbf{f}(\mathbf{c}''))^{\mathbf{T}}\cdot(\mathbf{c}'-\mathbf{c}'')\leq 0\quad\forall\mathbf{c}'\neq\mathbf{c}''$$

- it is continuously differentiable with symmetric negative semi-definite (with respect to real vectors) Jacobian, $\mathbf{J_f}(\mathbf{c}) = \nabla \mathbf{f}(\mathbf{c}) \preccurlyeq 0$.

The arc flow function can also be obtained combing the route flow function (3.9) with the arc-route flow consistency relation (3.1) and the route-arc cost consistency relation (3.3), or with reference to the i-route flows and costs as shown in Section 2.3.

The TEAMs (3.15) and (3.16) are useful for basic and advanced uniqueness analysis, as shown in "Basic uniqueness conditions" section and Section 3.3.

3.2.2 One equation assignment models

It is often worth to further reduce the number of the equations and the variables to one, by including one equation into the other with reference to any TEAM described above, leading to One equation Assignment Models (OEAMs). Considering TEAM (3.15) and (3.16) with respect to arc costs and flows only yields to:

$$\mathbf{f}^* = \mathbf{f}(\mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa}); \mathbf{d}) \in S_f \tag{3.17}$$

$$\mathbf{c}^* = \mathbf{c}(\mathbf{f}(\mathbf{c}^*; \mathbf{d}); \boldsymbol{\kappa}) \geq \mathbf{0} \tag{3.18}$$

The OEAM (3.17) is useful for existence analysis and algorithm convergence analysis if the Jacobian of the arc cost function is symmetric, as shown in "Existence conditions" section and Section 3.2.4. The OEAM (3.18) is useful for algorithm convergence analysis if the Jacobian of the flow cost function is symmetric, as it occurs for regular arc flow functions, even if the Jacobian of the arc cost function is asymmetric, as shown in Section 3.2.4. OEAMs with respect to route or i-route variables are not reported for brevity.

The data-flow diagram in Fig. 3.2 above is also a description of the OEAMs (3.17) and (3.18). Indeed, as already noted, this figure also highlights that a fixed-point model for CUE can be obtained by combining together the arc flow function. Which models the comprehensive assignment to uncongested networks (CUN), and the arc cost function, which models the congestion due to users sharing the same transportation facility.

3.2.3 Existence and basic uniqueness analysis

This subsection presents sufficient conditions for existence (at least one solution exists) or uniqueness (at most one solution exists) of the equilibrium assignment flows and costs with reference to the fixed-point models presented above. Those models are equivalent, say from the solution of one of them can easily be obtained the solution of the others, as well as any other variables through Eqs (3.1)–(3.6); thus it suffices to state conditions with reference to one fixed-point model only, not necessarily the same for existence and uniqueness analysis.

Existence conditions

Existence conditions. Sufficient conditions for existence of solutions can be easily derived with reference to OEAM (3.17), involving the composed function $\mathbf{f}(\mathbf{c}(\cdot))$ defined over set S_f with values in the same set (see Section 2.5), as:

1. each o-d pair i is connected by at least one route,
2. the arc cost function $\mathbf{c}(\mathbf{f}; \boldsymbol{\kappa})$ is continuous with respect to the arc flows $\mathbf{f}$,
3. the arc flow function $\mathbf{f}(\mathbf{c}; \mathbf{d})$ is continuous with respect to the arc costs $\mathbf{c}$ (as for a regular arc flow function), since all the route choice functions are continuous.

Proof is based on the Brouwer's theorem.

Necessary fixed-point existence conditions

Let $\boldsymbol{\varphi}(\mathbf{x})$ be a vector function from set S to set $\boldsymbol{\varphi}(S)$, and $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*)$ one of its fixed-points, necessary conditions for the existence of at least one fixed-point are that

– the domain S is non empty,
– the co-domain $\boldsymbol{\varphi}(S)$ is a subset of the domain, that is $\boldsymbol{\varphi}(S) \subseteq S$.

Indeed, necessary existence conditions hold, since set S_f is non empty for hypothesis 1, and the composed function $\mathbf{f}(\mathbf{c}(\cdot))$ defined over set S_f has values in set S_f by definition.

Sufficient fixed-point existence conditions - Brouwer's theorem

Let $\boldsymbol{\varphi}(\mathbf{x})$ be a vector function from set S to set $\boldsymbol{\varphi}(S)$, and $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*)$ one of its fixed-points, if necessary conditions for fixed-point existence hold, sufficient conditions for the existence of at least one fixed-point are that

– set S is compact and convex,
– function $\boldsymbol{\varphi}(\mathbf{x})$ is continuous.

In addition, both the assumptions of Brouwer's theorem hold, since set S_f is compact and convex, and the composed function $\mathbf{f}(\mathbf{c}(\cdot))$ is continuous for hypotheses 2 and 3.

The very same conditions can be derived with reference to OEAMs with respect to route or i-route flows.

Basic uniqueness conditions

Uniqueness may be stated with respect arc costs and flows with reference to TEAMs (3.15)–(3.16) or to i-route costs and flows with reference to TEAMs (3.13)–(3.14). Basic sufficient uniqueness conditions are reported below. Advanced sufficient conditions and references are discussed in Section 3.3.

Uniqueness conditions/arc. Sufficient conditions for uniqueness of solutions can be easily derived with reference to TEAM (3.15)–(3.16) as:

1. the arc cost function $\mathbf{c}(\mathbf{f};\boldsymbol{\kappa})$ is strictly increasing monotone with respect to the arc flows $\mathbf{f}$; if it is continuously differentiable a sufficient condition is that the Jacobian matrix is positive definite (for real vectors), but not necessarily symmetric, $\mathbf{J_C}(\mathbf{f}) \succ 0$;
2. the arc flow function $\mathbf{f}(\mathbf{c};\mathbf{d})$ is decreasing monotone with respect to the arc costs $\mathbf{c}$; if it is continuously differentiable a necessary and sufficient condition is that the Jacobian matrix is negative semi-definite (for real vectors), but not necessarily symmetric, $\mathbf{J_f}(\mathbf{c}) \preccurlyeq 0$, (as for a regular arc flow function, in this case the Jacobian matrix is also symmetric).

Proof is based on *reductio ad absurdum*.

Assuming that two different fixed-points exist:$(\mathbf{c}^*, \mathbf{f}^*) \neq (\mathbf{c}^{**}, \mathbf{f}^{**})$, Eqs (3.15) and (3.16) yield:

$$\mathbf{c}^* = \mathbf{c}(\mathbf{f}^*) \quad \mathbf{c}^{**} = \mathbf{c}(\mathbf{f}^{**})$$

$$\mathbf{f}^* = \mathbf{f}(\mathbf{c}^*) \quad \mathbf{f}^{**} = \mathbf{f}(\mathbf{c}^{**})$$

for hypothesis 1:

$$(\mathbf{c}(\mathbf{f}') - \mathbf{c}(\mathbf{f}''))^{\mathbf{T}} \cdot (\mathbf{f}' - \mathbf{f}'') > 0 \quad \forall \mathbf{f}' \neq \mathbf{f}''$$

for hypothesis 2:

$$(\mathbf{f}(\mathbf{c}') - \mathbf{f}(\mathbf{c}''))^{\mathbf{T}} \cdot (\mathbf{c}' - \mathbf{c}'') \leq 0 \quad \forall \mathbf{c}' \neq \mathbf{c}''$$

Thus

$$(\mathbf{c}(\mathbf{f}*) - \mathbf{c}(\mathbf{f}**))^{\mathbf{T}} \cdot (\mathbf{f}^* - \mathbf{f}^{**}) > 0 \quad \text{since } \mathbf{f}^* \neq \mathbf{f}^{**}$$

$$(\mathbf{f}(\mathbf{c}*) - \mathbf{f}(\mathbf{c}**))^{\mathbf{T}} \cdot (\mathbf{c}^* - \mathbf{c}^{**}) \leq 0 \quad \text{since } \mathbf{c}^* \neq \mathbf{c}^{**}$$

leading to a contradiction. Hence either one of (or both) assumptions 1 and 2 does not hold or there are not two different fixed-points.

Weaker uniqueness conditions can be stated with respect to i-route costs and flow, if the utility function is linear [HYP⑦] and choice proportions depend on differences between systematic utility values only [HYP⑧].

Uniqueness conditions /i-route. Sufficient conditions for uniqueness of solutions can be easily derived with reference to TEAM (3.13)–(3.14) as:

1. the arc cost function $\mathbf{c}(\mathbf{f};\boldsymbol{\kappa})$ is increasing monotone with respect to the arc flows $\mathbf{f}$; if it is continuously differentiable a sufficient condition is that the Jacobian matrix is positive semi-definite (for real vectors), but not necessarily symmetric, $\mathbf{J_C}(\mathbf{f}) \succeq 0$; thus the i-route cost function $\tilde{\mathbf{w}} = \tilde{\mathbf{w}}(\tilde{\mathbf{h}};\boldsymbol{\kappa})$ is increasing monotone with respect to the i-route flows $\tilde{\mathbf{h}}$;

2. the i-route flow function $\tilde{\mathbf{h}} = \tilde{\mathbf{h}}(\tilde{\mathbf{w}}; \mathbf{d})$ is strictly decreasing monotone with respect to the i-route costs $\tilde{\mathbf{w}}$, as it occurs for a strictly regular function based on strictly regular i-route choice functions; if it is continuously differentiable a sufficient condition is that the Jacobian matrixis negative definite (for real vectors) but not necessarily symmetric, $\nabla \tilde{\mathbf{h}}(\tilde{\mathbf{w}}) \prec 0$.

Proof is based on *reductio ad absurdum*, as for uniqueness conditions/arc, and it is omitted for brevity's sake.

3.2.4 Solution algorithms and convergence analysis

Algorithms based on the Method of Successive Averages (MSA) (introduced for equilibrium assignment by Powell and Sheffi, 1982; Daganzo, 1983) are the most used to solve fixed-point models for equilibrium assignment, since they can accommodate any choice model from RUT as well as other choice modelling approaches, and are suitable for large scale applications.

With reference to OEAM (3.17) or (3.18), basic iteration of MSA algorithms requires the computation of the arc cost function to get arc costs from arc flows and the computation of the arc flow function to get arc flows from arc costs. Similar algorithms may be specified with respect to route or i-route variables.

Algorithms based on the Method of Successive Averages (MSA)

Let $\boldsymbol{\varphi}(\mathbf{x})$ be a vector function from a non-empty convex set S to set $\boldsymbol{\varphi}(S) \subseteq S$, with a unique fixed-point $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*)$ in set S, MSA algorithms provide a succession of solutions belonging to S, $\mathbf{x}^k \in S$, possibly converging to required fixed-point, according to the following recursive equation:

$$\mathbf{x}^k = \mathbf{x}^{k-1} + (1/k)\left(\boldsymbol{\varphi}\left(\mathbf{x}^{k-1}\right) - \mathbf{x}^{k-1}\right) \quad \text{with } \mathbf{x}^0 = \mathbf{x_0} \in S$$

or $$\mathbf{x}^k = (1/k)\,\boldsymbol{\varphi}\left(\mathbf{x}^{k-1}\right) + ((k-1)/k)\,\mathbf{x}^{k-1} \quad \text{with } \mathbf{x}^0 = \mathbf{x_0} \in S$$

If at any iteration the succession provides the fixed-point the algorithm, the it will stop and vice versa. Sufficient conditions for theoretical convergence of such a succession may be stated through Blum's theorem, which also applies when only an unbiased estimation of the function $\boldsymbol{\varphi}(\mathbf{x})$ is available.

- MSA Flow Averaging algorithm

Applying the Method of Successive Averages to model (3.17) the MSA-FA algorithm is obtained based on the recursive equation:

$$\mathbf{f}^k = \mathbf{f}^{k-1} + (1/k)\left(\mathbf{f}\left(\mathbf{c}\left(\mathbf{f}^{k-1}\right) - \mathbf{f}^{k-1}\right)\right) \quad \text{with } \mathbf{f}^0 = \mathbf{f_0} \in S_f \tag{3.19}$$

It can be proved asymptotically converging to $\mathbf{f}^* = \mathbf{f}(\mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa}); \mathbf{d}) \in S_f$ if the existence and uniqueness/arc conditions hold and the Jacobian of the arc cost function is symmetric.

- MSA Cost Averaging algorithm

On the other hand, applying the MSA to model (3.18) the MSA-CA algorithm is obtained based on the recursive equation:

$$\mathbf{c}^k = \mathbf{c}^{k-1} + (1/k)\left(\mathbf{c}\left(\mathbf{f}\left(\mathbf{c}^{k-1}\right) - \mathbf{c}^{k-1}\right)\right. \quad \text{with } \mathbf{c}^0 = \mathbf{c}(\mathbf{f_0}) \text{ and } \mathbf{f_0} \in S_f \tag{3.20}$$

It may be proved asymptotically converging to $\mathbf{c}^* = (\mathbf{c}(\mathbf{f}(\mathbf{c}^*; \mathbf{d}); \boldsymbol{\kappa})$ if the existence and uniqueness/arc conditions hold and the Jacobian of the arc flow function is symmetric.

Convergence conditions of both algorithms can be stated through Blum's theorem (see appendix). If the choice proportions from a RUM cannot be evaluated exactly, an unbiased estimation of the arc flow function can be obtained through Monte Carlo techniques (see Section 2.5); in this case, only *almost sure* convergence can be assured.

Since MSA algorithms only provide a succession of feasible solutions, in practical applications the algorithm is stopped when a convergence index is below a given error threshold, ε, or a maximum number of iterations is reached. A convergence index often used for MSA-FA is the average absolute difference over flows:

$$\left(\sum\nolimits_a \left(|f_a\left(\mathbf{c}\left(\mathbf{f}^{k-1}\right) - f_a{}^{k-1}\,|\,/f_a{}^{k-1}\right)/m.\right.\right.$$

A similar index, based on arc costs, may be defined for MSA-CA. Others indices may be defined based on the maximum difference, possibly excluding arcs with very low flows.

3.3 Advanced uniqueness and convergence analysis

This section presents advanced sufficient conditions for uniqueness (at most one solution exists) of the equilibrium assignment flows and costs with reference to the fixed-point models presented above, and in some cases of convergence of MSA algorithms or of other kinds of algorithms. Each set of conditions presented below has a counterpart based on features of Jacobian matrices for differentiable functions.

This section is based on an elaboration of the content of.

- *Uniqueness of Stochastic User Equilibrium*, paper prepared by Cantarella G.E., Gentile G., and Velonà P. for the Proceedings of 5th IMA conference on Mathematics in Transportation, London, UK, April 2010,

that were never published. Unless otherwise stated, earlier versions of most conditions below have first been proposed in a paper and a book available in Italian only:

- *Condizioni di unicità dei flussi e dei costi di equilibrio stocastico*, by Cantarella G.E. in *Metodi e Tecnologie dell'Ingegneria*, a cura di G.E. Cantarella e F. Russo, 376–391. Franco Angeli Editore, 2001;
- *Assegnazione a reti di trasporto: modelli di punto fisso*, by Cantarella, G.E. & Velonà, P. Collana Trasporti - Franco Angeli Editore, 2010.

Uniqueness conditions discussed below are identified by a letter code followed by / and the kind of variables used. For conditions requiring differentiable functions a capital D is added; in this case it is also assumed that conditions hold over a suitable (open) superset of S_f or $\mathbf{c}(S_f)$ such that all point of S_f or $\mathbf{c}(S_f)$ are interior points of it. All proofs of main conditions are by based on *reductio ad absurdum* and omitted for brevity' sake; proof of D conditions results from features of Jacobian matrices, unless otherwise stated. An example below shows a comparison among them.

Uniqueness Conditions Ar/arc:

$$(\mathbf{c}(\mathbf{f}') - \mathbf{c}(\mathbf{f}''))^{\mathbf{T}} \cdot (\mathbf{f}' - \mathbf{f}'') > 0 \qquad \forall \mathbf{f}' \neq \mathbf{f}''$$

$$(\mathbf{f}(\mathbf{c}') - \mathbf{f}(\mathbf{c}''))^{\mathbf{T}} \cdot (\mathbf{c}' - \mathbf{c}'') \leq 0 \qquad \forall \mathbf{c}' \neq \mathbf{c}''$$

These conditions have already been discussed in Section 3.2.3 (first in Cantarella, 1997).

Uniqueness Conditions Ar/arc - D:

$$\nabla \mathbf{c}(\mathbf{f}) \succ 0.$$

$$\nabla \mathbf{f}(\mathbf{c}) \preccurlyeq 0 \text{ (as it occurs for invariant choice functions)}.$$

These conditions are the differentiable counter part of the above (a different proof in Daganzo, 1983, with respect to a different fixed-point model requiring the inverse of the arc cost function; another proof in Sheffi, 1985, based on optimization models).

Uniqueness Conditions A/arc:

$$(\mathbf{c}(\mathbf{f}') - \mathbf{c}(\mathbf{f}''))^{\mathbf{T}} \cdot (\mathbf{f}' - \mathbf{f}'') > (\mathbf{f}(\mathbf{c}(\mathbf{f}')) - \mathbf{f}(\mathbf{c}(\mathbf{f}'')))^{\mathbf{T}} \cdot (\mathbf{c}(\mathbf{f}') - \mathbf{c}(\mathbf{f}''))$$

$$\forall \mathbf{f}' \neq \mathbf{f}'' \in S_f : \mathbf{c}(\mathbf{f}') \neq \mathbf{c}(\mathbf{f}'')$$

These conditions are a generalisation of the Ar/arc conditions. Condition A/arc also support the convergence of MSA algorithms (as noted in *SUE: conditions for solution uniqueness and MSA-based algorithm convergence*, by Cantarella G.E. and Velonà P., in Preprints of XIII Meeting of Euro Working Group on Transportation, Padua, Italy, September 2009).

A special of conditions A/arc occur for invertible arc cost functions, as shown below.

Uniqueness Conditions A/arc for invertible cost functions:

If the arc cost function is invertible with $\mathbf{q}(\mathbf{c}) = \mathbf{c}^{-1}(\mathbf{c})$ conditions A/arc become:

$$(\mathbf{c}'-\mathbf{c}'')^{\mathbf{T}}\cdot(\mathbf{q}(\mathbf{c}')-\mathbf{q}(\mathbf{c}''))>(\mathbf{f}(\mathbf{c}')-\mathbf{f}(\mathbf{c}''))^{\mathbf{T}}\cdot(\mathbf{c}'-\mathbf{c}'')\forall\mathbf{c}'\neq\mathbf{c}''\in\mathbf{c}(S_f)$$

with $\mathbf{f}'=\mathbf{q}(\mathbf{c}')\neq\mathbf{q}(\mathbf{c}'')=\mathbf{f}''\ \ \forall\mathbf{c}'\neq\mathbf{c}''\in\mathbf{c}(S_f)$.

Uniqueness Conditions A/arc - D for invertible cost functions:

$$\nabla\mathbf{c}(\mathbf{f}=\mathbf{q}(\mathbf{c}))^{-1}-\nabla\mathbf{f}(\mathbf{c})\succ 0\forall\mathbf{c}\in\mathbf{c}(S_f)$$

[It is noteworthy that this expression appeared in Daganzo, 1983 only to prove conditions Ar above.]

Other conditions, without any evident relationship with the previous ones, can be derived from a corollary of the Banach's theorem.

Sufficient fixed-point existence and uniqueness conditions - corollary to Banach's theorem

Let $\boldsymbol{\varphi}(\mathbf{x})$ be a vector function from non empty set S to set $\boldsymbol{\varphi}(S)$ (that is necessary conditions for existence hold), it has exactly one fixed-point $\mathbf{x}^*=\boldsymbol{\varphi}(\mathbf{x}^*)$ if

– set S is compact,
– function $\boldsymbol{\varphi}(\mathbf{x})$ is strictly non-expansive:

$$\|\boldsymbol{\varphi}(\mathbf{x}')-\boldsymbol{\varphi}(\mathbf{x}'')\|_2<\|\mathbf{x}'-\mathbf{x}''\|_2\forall\mathbf{x_1}\neq\mathbf{x_2}\in S$$

REMARK. If function $\boldsymbol{\varphi}(\mathbf{x})$ is continuously differentiable with Jacobian $\nabla\boldsymbol{\varphi}(\mathbf{x})$, a sufficient conditions for being strictly non expansive is that the second norm of its Jacobian is less than one, $\|\nabla\boldsymbol{\varphi}(\mathbf{x})\|_2<1\ \forall\mathbf{x}\in S$. [Any other vector norm may be used as well.]

REMARK. According to the Banach's theorem it suffices that set S is complete, but in this case function $\boldsymbol{\varphi}(\mathbf{x})$ has to be a contraction, that is uniformly strictly non-expansive.

Banach's theorem and its corollary can also be used to state convergence conditions of fixed-point algorithms based on the Method of Repeated Approximations (MRA), more efficient than those base on the MSA.

Algorithms based on the Method of Repeated Approximations (MRA)

Let $\boldsymbol{\varphi}(\mathbf{x})$ be a vector function from a non-empty convex set S to set $\boldsymbol{\varphi}(S)\subseteq S$, with a unique fixed-point $\mathbf{x}^*=\boldsymbol{\varphi}(\mathbf{x}^*)$ in set S, MRA algorithms provide a succession of solutions belonging to S, $\mathbf{x}^k\in S$, possibly converging to the required fixed-point, according to the following recursive equation:

$$\mathbf{x}^k=\boldsymbol{\varphi}(\mathbf{x}^{k-1})\quad\text{with }\mathbf{x}^0=\mathbf{x}_0\in S$$

Sufficient conditions for theoretical convergence of such a succession are those of the Banach's theorem or its corollary for existence and uniqueness of the fixed-point.

Uniqueness Conditions Br/arc:

$$\|\mathbf{c}(\mathbf{f}(\mathbf{c}'))-\mathbf{c}(\mathbf{f}(\mathbf{c}''))\|_2<\|\mathbf{c}'-\mathbf{c}''\|_2\forall\mathbf{c}'\neq\mathbf{c}''\in\mathbf{c}(S_f).$$

meaning that the composed function $\mathbf{c}(\mathbf{f}(\mathbf{c}))$ is strictly non-expansive. Similar conditions hold w.r.t. the composed function $\mathbf{f}(\mathbf{c}(\mathbf{f}))$. Condition Br may rarely be applied, since generally involved functions are not strictly non-expansive.

Uniqueness Condition Br/arc - D:

$$\| \nabla[\mathbf{c}(\mathbf{f}=\mathbf{f}(\mathbf{c}))] \cdot \nabla[\mathbf{f}(\mathbf{c})] \|_2 < 1 \forall \mathbf{c} \in \mathbf{c}(S_f).$$

These conditions are the differentiable counter part of the above.

Uniqueness Conditions B/arc:

$$(\mathbf{c}'-\mathbf{c}'')^{\mathrm{T}}(\mathbf{c}'-\mathbf{c}'') > (\mathbf{c}(\mathbf{f}(\mathbf{c}'))-\mathbf{c}(\mathbf{f}(\mathbf{c}'')))^{\mathbf{T}} \cdot (\mathbf{c}'-\mathbf{c}'') \forall \mathbf{c}' \neq \mathbf{c}'' \in \mathbf{c}(S_f)$$

$$\text{or } (\mathbf{c}'-(\mathbf{c}(\mathbf{f}(\mathbf{c}'))) - (\mathbf{c}''-\mathbf{c}(\mathbf{f}(\mathbf{c}''))))^{\mathbf{T}} \cdot (\mathbf{c}'-\mathbf{c}'') > 0 \forall \mathbf{c}' \neq \mathbf{c}'' \in \mathbf{c}(S_f)$$

meaning that function $\mathbf{c} - \mathbf{c}(\mathbf{f}(\mathbf{c}))$ is strictly monotone increasing. These conditions are a generalisation of the Br/arc conditions.

Uniqueness Conditions B/arc - D:

$$(\mathbf{I} - \nabla\mathbf{c}(\mathbf{f}=\mathbf{f}(\mathbf{c})) \cdot \nabla\mathbf{f}(\mathbf{c})) \succ 0 \forall \mathbf{c} \in \mathbf{c}(S_f)$$

These conditions are the differentiable counter part of the above.

Uniqueness Condition C/arc. Both conditions A-arc and B-arc are instances of the general uniqueness condition requiring that:

$$\text{the composed function } (\mathbf{c}-\mathbf{c}(\mathbf{f}(\mathbf{c}))) \text{ is invertible for } \mathbf{c} \in \mathbf{c}(S_f) \tag{3.21}$$

These are most general conditions currently available with respect to arc cost and arc flow functions, encompassing all the conditions presented above. It should be noted that conditions Ar/arc require features for the arc cost function and the arc flow function separately, while all the others refer to features of some combination of them.

If both the arc cost function, $\mathbf{c}(\mathbf{f})$, and the arc flow function, $\mathbf{f}(\mathbf{c})$, are continuous, condition (3.21) implies that the composed function $(\mathbf{c} - \mathbf{c}(\mathbf{f}(\mathbf{c})))$ is strictly monotone, thus either of the following two conditions holds:

$$\text{function } (\mathbf{c}-\mathbf{c}(\mathbf{f}(\mathbf{c}))) \text{ is strictly increasing monotone} \tag{3.22}$$

$$\text{function } (\mathbf{c}-\mathbf{c}(\mathbf{f}(\mathbf{c}))) \text{ is strictly decreasing monotone} \tag{3.23}$$

Uniqueness Condition C/arc-D.

$$|\mathbf{I} - \nabla\mathbf{c}(\mathbf{f}=\mathbf{f}(\mathbf{c})) \cdot \nabla\mathbf{f}(\mathbf{c})| \neq 0 \forall \mathbf{c} \in S_c^* \tag{3.24}$$

where S_c^* is over a suitable open superset of set $\mathbf{c}(S_f)$ such that all point of $\mathbf{c}(S_f)$ are interior points of it; this condition can be proved a sufficient condition for function $(\mathbf{c} - \mathbf{c}(\mathbf{f}(\mathbf{c})))$ being invertible, through the global inverse function theorem. Indeed since the arc flow function $\mathbf{c}(\mathbf{f})$ has been assumed continuously differentiable, it is also continuous thus set $\mathbf{c}(S_f)$ is connected and bounded since S_f is connected (convex) and bounded (compact). If arc cost function is strictly positive, set S_c^* can be chosen as a subset of the set of strictly positive real vectors.

Global inverse function theorem

Let $\boldsymbol{\varphi}(\mathbf{x})$ be a vector function over the non empty connected and bounded set S, and S' be an open superset of S such that (it has interior points and) all point of S are interior points of it, if

- function $\boldsymbol{\varphi}(\mathbf{x})$ is continuously differentiable over set S',
- its Jacobian matrix is non singular over set S', $|\nabla \boldsymbol{\varphi}(\mathbf{x})| \neq 0 \ \forall \mathrm{x} \in S'$, thus it is invertible,

then function $\boldsymbol{\varphi}(\mathbf{x})$ is invertible over set S.

Conditions C /arc and C/arc-D includes as special cases conditions Ar, A, Br, B.

Uniqueness Condition C/arc-D. Remark 1.

Since the arc cost function, $\mathbf{c}(\mathbf{f})$, and the arc flow function, $\mathbf{f}(\mathbf{c})$, are assumed continuously differentiable functions, the determinant $|\ \mathbf{I} - \nabla \mathbf{c}(\mathbf{f}{=}\mathbf{f}(\mathbf{c})) \cdot \nabla \mathbf{f}(\mathbf{c})\ |$ is a continuous function of the vector of arc costs, $\mathbf{c}$, thus due to the *Sign-Preserving property of Continuous Functions* (Bolzano's theorem) condition (3.24) implies that either of the following two conditions holds, C1 or C2, but not both:

$$|\mathbf{I} - \nabla \mathbf{c}(\mathbf{f} = \mathbf{f}(\mathbf{c})) \cdot \nabla \mathbf{f}(\mathbf{c})\,| > 0 \quad \forall \mathbf{c} \in S_c^* \tag{3.25a}$$

$$|\mathbf{I} - \nabla \mathbf{c}(\mathbf{f} = \mathbf{f}(\mathbf{c})) \cdot \nabla \mathbf{f}(\mathbf{c})\,| < 0 \quad \forall \mathbf{c} \in S_c^* \tag{3.25b}$$

Uniqueness Condition C/arc-D. Remark 2.

Condition expressed by (3.24) is equivalent to assuming that all eigenvalues of matrix $\nabla \mathbf{c}(\mathbf{f}{=}\mathbf{f}(\mathbf{c})) \cdot \nabla \mathbf{f}(\mathbf{c})$ are different from 1, since.

- the determinant of a matrix is equal to the product of its eigenvalues, and.
- each of the m eigenvalues λ_a of matrix $\mathbf{I} - \nabla \mathbf{c}(\mathbf{f}{=}\mathbf{f}(\mathbf{c})) \cdot \nabla \mathbf{f}(\mathbf{c})$ are given by $1 - \omega_a$, where ω_a is one of the m eigenvalues of matrix $\nabla \mathbf{c}(\mathbf{f}{=}\mathbf{f}(\mathbf{c})) \cdot \nabla \mathbf{f}(\mathbf{c})$.

Eigenvalues of the product of two matrices

Let $\mathbf{M}$ be a $k \times r$ matrix and $\mathbf{N}$ be a $r \times k$ matrix, thus both the products $k \times k$ $\mathbf{M} \cdot \mathbf{N}$ and $r \times r$ $\mathbf{N} \cdot \mathbf{M}$ are well-defined:

if $k < r$, the $r \times r$ matrix $\mathbf{N} \cdot \mathbf{M}$ has all the k eigenvalues of the $k \times k$ matrix $\mathbf{M} \cdot \mathbf{N}$ plus r - k zero eigenvalues;

if $k = r$, $\mathbf{M}$ and $\mathbf{N}$ are both $r \times r$ square matrices, and have the same eigenvalues;

if $k > r$, the $k \times k$ matrix $\mathbf{M} \cdot \mathbf{N}$ has all the r eigenvalues of the $r \times r$ matrix $\mathbf{N} \cdot \mathbf{M}$ plus k - r zero eigenvalues;

summing up, the two matrices $\mathbf{M} \cdot \mathbf{N}$ and $\mathbf{N} \cdot \mathbf{M}$ have the same non zero eigenvalues.

Moreover, the two square $m \times m$ matrices $\nabla \mathbf{c}(\mathbf{f}) \cdot \nabla \mathbf{f}(\mathbf{c})$ and $\nabla \mathbf{f}(\mathbf{c}) \cdot \nabla \mathbf{c}(\mathbf{f})$ have the same eigenvalues, thus the order of the product is not relevant to check condition (3.24), and it can be also be expressed as:

$$|\mathbf{I} - \nabla \mathbf{f}(\mathbf{c}(\mathbf{f})) \cdot \nabla \mathbf{c}(\mathbf{f})\,| \neq 0 \quad \forall \mathbf{f} \in S_f \tag{3.26}$$

As shown below conditions (3.24) and (3.26) can also be expressed with respect to route cost and flow functions or i-route cost and flow functions leading to same result:

$$|\mathbf{I} - \nabla\mathbf{w}(\mathbf{h} = \mathbf{h}(\mathbf{w})) \cdot \nabla\mathbf{h}(\mathbf{w})\,| \neq 0 \quad \forall \mathbf{w} \in S_w \tag{3.27}$$

$$|\mathbf{I} - \nabla\mathbf{h}(\mathbf{w}(\mathbf{h})) \cdot \nabla\mathbf{w}(\mathbf{h})\,| \neq 0 \quad \forall \mathbf{h} \in S_h \tag{3.28}$$

$$|\mathbf{I} - \nabla\mathbf{w}(\mathbf{h} = \mathbf{h}\mathbf{w})) \cdot \nabla\mathbf{h}(\mathbf{w})\,| \neq 0 \quad \forall \mathbf{w} \in S_w \tag{3.29}$$

$$|\mathbf{I} - \nabla\mathbf{h}(\mathbf{w}(\mathbf{h})) \cdot \nabla\mathbf{w}(\mathbf{h})\,| \neq 0 \quad \forall \mathbf{h} \in S_h \tag{3.30}$$

Same considerations hold for conditions (3.25a) and (3.25b).

Thus, conditions C/−D are equivalent whichever are the variables and the functions used.

First it is worth noting that the two $n \times n$ matrices $\nabla\mathbf{w}(\mathbf{h}) \cdot \nabla\mathbf{h}(\mathbf{w})$ and $\nabla\mathbf{h}(\mathbf{w}) \cdot \nabla\mathbf{w}(\mathbf{h})$ have the same eigenvalues. Moreover, Eqs (3.1), (3.9), and (3.3) imply that:

$$\nabla\mathbf{f}(\mathbf{c}) = \mathbf{B} \cdot \nabla\mathbf{h}\left(\mathbf{w} {=} \mathbf{B}^{\mathrm{T}} \cdot \mathbf{c} + \mathbf{w}_{\mathbf{Z}}\right) \cdot \mathbf{B}^{\mathrm{T}}.$$

and Eqs (3.3), (3.2), (3.1) imply that:

$$\nabla\mathbf{w}(\mathbf{h}) = \mathbf{B}^{\mathrm{T}} \cdot \nabla\mathbf{c}(\mathbf{f} {=} \mathbf{B} \cdot \mathbf{h} + \mathbf{f}_{\mathbf{Z}}) \cdot \mathbf{B}.$$

Thus the $n \times n$ square matrix.

$$\nabla\mathbf{w}(\mathbf{h}) \cdot \nabla\mathbf{h}(\mathbf{w}) {=} \mathbf{B}^{\mathrm{T}} \cdot \nabla\mathbf{c}(\mathbf{f} {=} \mathbf{B} \cdot \mathbf{h} + \mathbf{f}_{\mathbf{Z}}) \cdot \mathbf{B} \cdot \nabla\mathbf{h}(\mathbf{w}).$$

has the same non zero eigenvalues of $m \times m$ matrix.

$$\nabla\mathbf{c}(\mathbf{f} {=} \mathbf{B} \cdot \mathbf{h} + \mathbf{f}_{\mathbf{Z}}) \cdot \mathbf{B} \cdot \nabla\mathbf{h}(\mathbf{w}) \cdot \mathbf{B}^{\mathrm{T}} = \nabla\mathbf{c}(\mathbf{f} = \mathbf{f}(\mathbf{c})) \cdot \nabla\mathbf{f}(\mathbf{c}).$$

and the $n \times n$ square matrix.

$$\nabla\mathbf{h}(\mathbf{w}) \cdot \nabla\mathbf{w}(\mathbf{h}) = \nabla\mathbf{h}(\mathbf{w}) \cdot \mathbf{B}^{\mathrm{T}} \cdot \nabla\mathbf{c}(\mathbf{f} {=} \mathbf{B} \cdot \mathbf{h} + \mathbf{f}_{\mathbf{Z}}) \cdot \mathbf{B}.$$

has the same non zero eigenvalues of the $m \times m$ matrix.

$$\mathbf{B} \cdot \nabla\mathbf{h}(\mathbf{w}) \cdot \mathbf{B}^{\mathrm{T}} \cdot \nabla\mathbf{c}(\mathbf{f} {=} \mathbf{B} \cdot \mathbf{h} + \mathbf{f}_{\mathbf{Z}}) = \nabla\mathbf{f}(\mathbf{c}) \cdot \nabla\mathbf{c}(\mathbf{f}).$$

Similar considerations apply for i-route cost and flow functions.

On the other hand, formally similar Ar, A, Br, B uniqueness conditions can be stated with respect to i-route (or route) variables and functions, but they are not equivalent to those with respect to arc variables. Ar, B, C / i-route conditions only are explicitly reported below, and D counterparts are not explicitly reported for brevity's sake.

Uniqueness Conditions Ar/i-route:

$$(\mathbf{w}(\mathbf{h}') - \mathbf{w}(\mathbf{h}''))^{\mathbf{T}} \cdot (\mathbf{h}' - \mathbf{h}'') \geq 0 \quad \forall \mathrm{h}' \neq \mathrm{h}''$$

$$(\mathbf{h}(\mathbf{w}') - \mathbf{h}(\mathbf{w}''))^{\mathbf{T}} \cdot (\mathbf{w}' - \mathbf{w}'') < 0 \quad \forall \mathrm{w}' \neq \mathrm{w}''$$

These conditions are a generalisation of the Ar/arc conditions.

Uniqueness Conditions B/i-route:

$$(\mathbf{h}' - \mathbf{h}'')^{\mathbf{T}} (\mathbf{h}' - \mathbf{h}'') > (\mathbf{h}(\mathbf{w}(\mathbf{h}')) - \mathbf{h}(\mathbf{w}(\mathbf{h}'')))^{\mathbf{T}} \cdot (\mathbf{h}' - \mathbf{h}'') \forall \mathbf{h}' \neq \mathbf{h}'' \in S_h.$$

$$\text{or } (\mathbf{h}' - (\mathbf{h}(\mathbf{w}(\mathbf{h}'))) - (\mathbf{h}'' - \mathbf{h}(\mathbf{w}(\mathbf{h}''))))^{\mathbf{T}} \cdot (\mathbf{h}' - \mathbf{h}'') > 0 \forall \mathbf{h}' \neq \mathbf{h}'' \in S_h.$$

meaning that function $\tilde{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\mathbf{w}}(\tilde{\mathbf{h}}))$ is strictly monotone increasing.

Uniqueness Condition C/arc. Both conditions A/i-route and B/i-route are instances of the general uniqueness condition requiring that:

the composed function $\tilde{\mathbf{h}} - \tilde{\mathbf{h}}(\tilde{\mathbf{w}}(\tilde{\mathbf{h}}))$ is invertible for $\tilde{\mathbf{h}} \in \in S_{\tilde{h}}$.

This is the most general condition currently available with respect to i-route variables, encompassing all the other i-route conditions; its D counterpart is given by (3.30), already introduced above.

All the /arc-D uniqueness conditions are compared in Fig. 3.4 below for a two-arc network, where $\mathbf{f} = \mathbf{h}$, $\tilde{h} = h_1$, $\mathbf{c} = \mathbf{w}$, $\tilde{w} = c_1 - c_2$. A separable cost function is associated to each arc $a = 1, 2$: $c_a(f_a)$ with derivative $x_a = \partial c_a(f_a) / \partial f_a$. Thus the Jacobian $\nabla \mathbf{c}(\mathbf{f})$ of the arc cost vector function is given by the diagonal matrix:

$$Vc(f) = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}$$

The choice function is an invariant Logit:

$p_a(c_1, c_2) = \exp.(-c_a/\theta) / (\exp(-c_1/\theta) + \exp.(-c_2/\theta)) \ \forall a = 1, 2.$

with $\theta \ \mu \ \sigma > 0$, and $\partial\theta / \partial c_a = 0$, $a = 1, 2$. Let $y = (d/\theta)\, p_1(c_1,c_2)\, p_2(c_1,c_2)$, the Jacobian of the arc flow functions is given by the (singular) matrix.

$$Vf(c) = -(d/\theta) \begin{bmatrix} p_1(1-p_1) & -p_1\, p_2 \\ -p_2\, p_1 & p_2(1-p_2) \end{bmatrix} = -y \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

For differentiable functions, each of the arc uniqueness conditions Ar, A, Br, B, C1, and C2 defines a region over the plane $z_1 = x_1 \cdot y$ and $z_2 = x_2 \cdot y$, as shown in Fig. 3.4 to support comparison among them. The white line between C1 and C2 sub-regions corresponds to condition: $| \mathbf{I} - \nabla \mathbf{c}(\mathbf{f} = \mathbf{f}(\mathbf{c})) \cdot \nabla \mathbf{f}(\mathbf{c}) | = 0 \ \forall \mathbf{c} \in S_c$, thus the two regions are separated. It is noteworthy that the uniqueness region for A/arc conditions is not connected, thus it actually breaks down into 3 sub-regions, only one sub-region may actually be considered due to Bolzano (sign-preserving) theorem. Graphs confirm implications among uniqueness conditions as already discussed above.

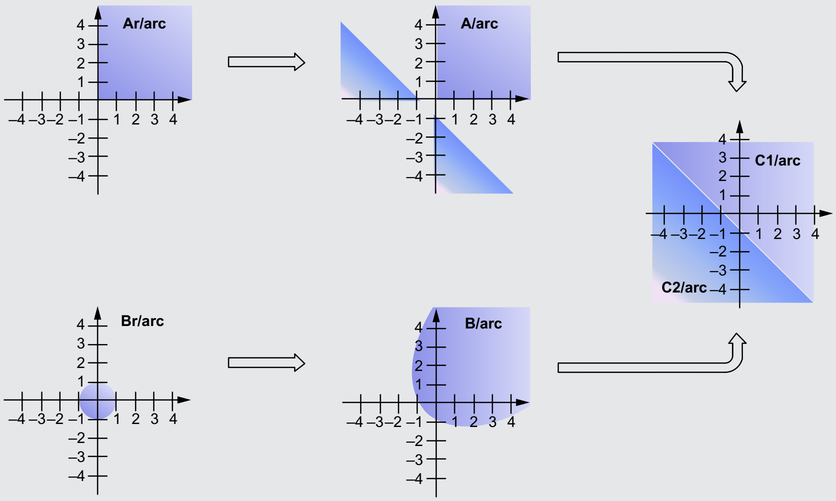


FIG. 3.4

Arc uniqueness regions.

Following on the above example, each of the /i-route-D uniqueness conditions Ar, A, Br, B, C1, and C2 defines a region over the plane z_1 and z_2 as shown in Fig. 3.5 to support comparison among them. Uniqueness regions are generally different from their arc counterparts, apart from conditions C as expected from above considerations. It is noteworthy that the uniqueness regions for conditions A, B and C1 equal. Moreover they.

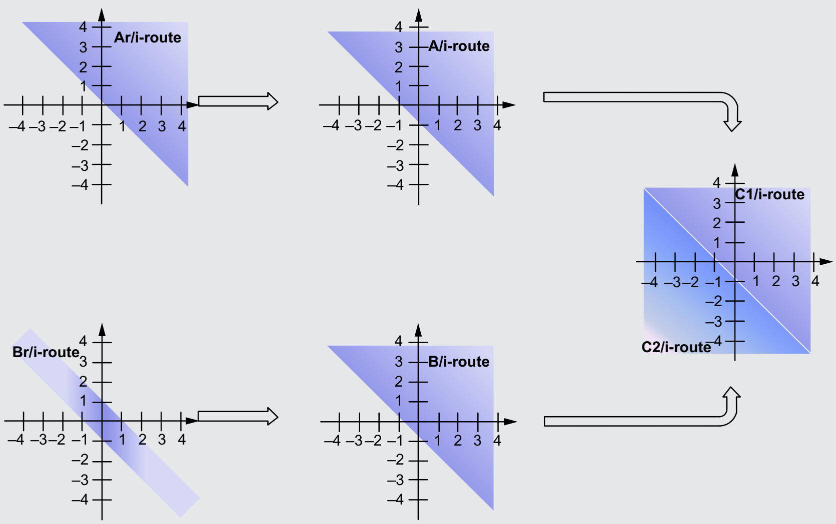


FIG. 3.5

I-route uniqueness regions.

It is should be noted that uniqueness conditions A, B and C allow for non monotone arc cost functions, which may well be the case for non-separable arc cost functions. [These conditions also allow for non-differentiable arc cost functions, such as piecewise linear functions, as it may occur when equilibrium assignment is embedded within a transportation supply design model.]

It is, therefore, worthwhile to introduce some requirements for non monotone arc cost functions:

– strict positivity.

$c_a(f_a) > 0 \quad \forall f_a \geq 0.$

– monotonicity after capacity κ_a.

$(c_a(f_a') - c_a(f_a'')) (f_a' - f_a''') \geq 0 \quad \forall f_a', f_a'' \geq \kappa_a.$

– consistency with capacity κ_a.

$c_a(f_a = \kappa_a) = \nu\, c_a(f_a = 0) \quad \text{with}\, \nu \geq 1.$

Monotone strictly increasing arc cost functions surely meet these requirements if the null-flow cost is strictly positive, $c_a(f_a = 0) > 0$, as it is always the case.

A further relationship between the arc and i-route uniqueness conditions can be exploited through the following uniqueness conditions, which hold under the assumption of invariant strictly positive choice functions.

Uniqueness Condition G. Uniqueness is guaranteed under the following condition about the Jacobian of the cost function, $\mathbf{\nabla c(f)}$:

$(\nabla \mathbf{c}(\mathbf{f}) + \gamma\, \mathbf{I})\, 0 \succ 0 \qquad \forall \mathbf{f} \in S_f$

where $0 < \gamma \leq \gamma^*$ and $\gamma^* = -\text{MAX}\{(\mathbf{x} / \|\mathbf{B\,L\,x}\|_2)^T (\nabla \tilde{\mathbf{h}}(\tilde{\mathbf{w}} = \tilde{\mathbf{w}}\,(\mathbf{h}\tilde{}\,)))^{-1} (\mathbf{x} / \|\mathbf{B\,L\,x}\|_2) > 0$, with $\mathbf{x} \in \{\mathbf{x}\text{:}\ \|\mathbf{x}\| = 1,\ \|\mathbf{B\,L\,x}\|_2 > 0\}\}$; $\mathbf{L}$, $\tilde{\mathbf{h}}$, $\tilde{\mathbf{w}}$ have been defined in Section 2.3.

This condition first appeared in

- Guido Gentile/(2003) *Sufficient conditions for the uniqueness of the solution to the Stochastic User Equilibrium problem*, Internal report available at http://w3.uniroma1.it/guido.gentile.

Further details and a proof are in:

- Gentile G., Velonà P., Cantarella G.E. (2014). Uniqueness of stochastic user equilibrium with asymmetric volume-delay functions for merging and diversion. In *EURO J. of Transportation and Logistics* 3, p309–331.

Condition G implies condition A/iro-D conditions (not explicitly mentioned above). Moreover condition Ar/arc-D implies conditions G since the sum of positive definite matrices is a positive definite matrix. It is also noteworthy that condition G does not require features of any composed function.

3.4 Summary

3.4.1 Major findings

This chapter presented several fixed-point models for Comprehensive User Equilibrium assignment to congested Networks, within Six Equation Assignment Modelling (SEAM) framework; further models have been presented within Two Equation Assignment Modelling (TEAM) or One Equation Assignment Modelling (OEAM) framework. Results hold under any assumptions leading to a Transpose Affine Network (TAN), being synchronic or diachronic.

The proposed approach is very general, can easily extended under other assumptions about route choice behaviour modelling, and enables to specify models for day-to-day dynamic assignment to congested networks, as shown in the next chapters. Optimization models are available for some very particular instances of User Equilibrium assignment to congested networks, but they cannot be generalised, nor used for general models for day-to-day dynamic assignment to congested networks (see appendix to Chapter 5 in Cascetta, 2009, for a review).

This approach can rather easily be extended to assignment with demand flows variable with respect to costs, and/or multi-type or multi-mode assignment, where the choice behaviour among vehicle types or transportation modes is explicitly described by choice models. These extensions are out of the scope of this book (and will possibly be described in a future book on advanced topics, some details can be found in Chapter 6 in Cascetta, 2009).

Fixed-point models for CUE assignment can easily be embedded within methods for Transportation Supply Design, such Urban Network Design where decisional variables are signal settings and street directions, while equilibrium assignment is a constraint.

All parameters introduced above are to be calibrated against real /simulated data, this relevant issues is out the scope of this book. As already stated implementation and application issues are out of the scope of this book, mainly focusing on mathematical features. (For details on these issues see Cascetta (2009). A companion book (possibly by other authors) discussing these topics is under planning.

3.4.2 Further readings

Wardrop (1952) introduced the User Equilibrium (UE) assignment, whose behavioural assumptions are often referred to as Wardrop's I principle, and introduced as well as the System Optimum (SO) assignment, assuming that users cooperate to minimize total system cost, according to the so-called II Wardrop's principle. Generally for congested networks UE and SO assignment lead to different flow and cost patterns, apart from very special cases. SO assignment, often useful as a reference for transportation planning, may be applied when users do not have autonomous route decision capability, such as during controlled people evacuation or freight transportation. An extension to Stochastic System Optimum has been proposed by Maher et al. (2005).

SUE with RUMs over a finite support have recently been by Watling et al. (2015, 2018), see also Rasmussen et al. (2015). SUE with Value of Time (VoT) randomly distributed among users has been discussed in Cantarella and Binetti (1998). SUE for explicit modelling of parking choice behaviour has been introduced by Bifulco (1993), as already noted in Section 2.2.3. For details on schedule-based assignment to diachronic networks see the recent book edited by Gentile and Nökel (2016), even though the contents mostly refer to deterministic assignment only, as already noted in the previous Chapter 2. A few authors have addressed the User Equilibrium assignment with route fuzzy utility, for instance Henn (2000), Ridwan (2004), Henn and Ottomanelli (2006), Ghatee and Hashemi (2009).

3.4.3 Remarks

As noted above Wardrop (1952) introduced the User Equilibrium (UE) assignment, whose behavioural assumptions are often referred to as Wardrop's I principle. This approach to equilibrium assignment is equivalent to those assuming that all sources of uncertainty are negligible, thus all users travelling between o-p pair i follow maximum utility routes, and do not use at all any of the other routes. As already noted in the previous Chapter 2 and in the beginning of this chapter this user choice behaviour assumption may be obtained from several assumptions.

In this case, the route choice function (3.7) [cfr (2.12)] and the arc function (2.34) turns out multi-valued functions. Even though general fixed-point theory also includes fixed-point of multi-valued functions, a different approach is often followed to avoid this kind of functions. Assuming a Wardropian user route choice behaviour, from the previous chapter any arc flow vectors $\mathbf{f_D}$ corresponding to arc cost vector $\mathbf{c}$ must satisfies the following condition [cfr (2.37)]:

$$\mathbf{c}^{\mathbf{T}} \cdot (\mathbf{f} - \mathbf{f_D}) \geq 0 \quad \forall \mathbf{f} \in S_f \tag{3.31}$$

Following the same line of reasoning leading to OEAMs, condition (3.31) may be combined with the arc cost function (3.2) to get a variational inequality (VI) model:

$$\mathbf{c}(\mathbf{f_D})^{\mathbf{T}} \cdot (\mathbf{f} - \mathbf{f_D}) \geq 0 \quad \forall \mathbf{f} \in S_f \tag{3.32}$$

This model has at least one solution if the arc cost function (3.2) is continuous, and at most one if it is monotone strictly increasing. As for any VI model, if the Jacobian matrix of the arc cost function is symmetric there exist convex optimisation models equivalent to VI model (3.30). Further details can be found in Cascetta (2009) and in Sheffi (1985), a useful reference on this topic is:

Patriksson M. (1994). The Traffic Assignment Problem: Model and Methods. VSP, Utrecht, The Netherlands.

Below UE assignment is compared with SUE assignment, as a special instance of CUE, to evidence all the many advantages of SUE with respect to UE, thus motivating why UE is only been briefly discussed in this book. On the other hand, it is should be remember that solving UE assignment is less computer demanding than SUE assignment or other kinds of CUE assignment, thus UE assignment may still be a useful tool for very large scale application if computing time is an issue.

1. SUE includes a more realistic description of user route choice behaviour modelling

— dispersion among users (heterogeneity), users' perception errors, dispersion of a user behaviour over days, …
— aggregation errors (due for instance to zoning), missing attributes, attribute measurement errors (say errors in the supply model), … .

All above issues can hardly be neglected. In other words when applying UE we are assuming that users do not commit any error in forecasting LoS, and that we do not commit any error in modelling their behaviour and in modelling network LoS. This looks as a rather unrealistic assumption, in particular when other choice dimension are involved, such as departure time, apart from route.

UE can be considered just a limit case of SUE when uncertainty goes to zero. It should be also noted that in any modelling approach in Engineering, Economics, Applied Sciences, … a step-wise map, as the flow-cost map in UE, is usually approximated by continuous functions such as logistic or other smooth functions. In this sense SUE-Logit should at least be considered a useful smooth approximation of UE(see point 2.2 below), apart from any behavioural considerations.

We better say SUEs, SUE indeed may be specified through several different route choice models derived from RUT, such as any from the Logit family, Weibit, Probit, Gammit or a mix of these, just to mention a few. Each of these models includes at least one behavioural parameter, sometimes several, to be calibrated, which (try to) model all sources of uncertainty, regarding both the users and the modeller, mentioned above. [Some scholars thinks that having no behavioural parameter to be calibrated is an advantage of UE; clearly this is not the case, since behavioural parameters make a model more flexible and more amenable to real-world application.] Moreover in some recent papers approaches to model route choice set too have been proposed.

Flexibility of SUE is a highly relevant features in particular when dealing with multi-user assignment where user are grouped into classes (see also point 2.1 below), such as commuters vs. non-commuters, ATIS-equipped vs. non-equipped vehicle, and in the near future autonomous/automated vs. traditional vehicles.

Furthermore, the SUE arc flow pattern is less sensible to input data such as demand flows with respect to UE pattern, thus SUE is better suited for project assessment giving more robust results (requiring less precise solution algorithms, see below).

2. SUE shows some very useful **mathematical features** with respect to UE
   1. uniqueness of link flows also implies uniqueness of link flows per user class, and of (i-)route flows and costs;
   2. flows depend on costs through a continuous, c. differentiable function **f**(**c**) with symmetric Jacobian, under very mild assumptions met by all choice models used in current practice.
3. **Fixed-point** models for SUE show several useful mathematical features with respect to any kind of models available for UE as enlisted below
   1. Fixed-point models available for SUE are simpler and require a much simpler mathematical background (see also point 4 below).
   2. These models for SUE allow for weak uniqueness and convergence conditions, including non necessarily increasing cost functions **c**(**f**); these conditions cannot be extended to UE, however modelled.
   3. These models can be solved through simple and feasible algorithms proved converging under very mild assumptions even for cost functions with asymmetric Jacobian; these conditions cannot be extended to UE, however modelled.
   4. An effective and efficient algorithm-independent indicator to measure how far a flow pattern **f** is from the search SUE flow pattern is any metric between the two vectors **f** and **f**(**c**(**f**)). This approach cannot easily be applied for UE since in this case **f**(**c**) is a point-to-set map, thus any metric between the values of a duly defined gap function of two successive intermediate solutions are used instead; but all gap functions proposed until now are very flat close to the equilibrium flow pattern thus this indicator is rather ineffective, moreover it is not efficient since it also requires the computation of the gap function, and it is based on two solutions, which in turn depends on the solution algorithm. Regarding algorithm convergence it is also relevant noting that the SUE arc flow pattern is less sensible to costs with respect to UE pattern (cfr previous comment on sensitivity with respect to demand flows), thus there is no need of a high convergence threshold, 10^{-3} being enough in most cases, to be compared with 10^{-6}, or even less, often required for UE solution. [Most engineering applications do not require more than three significant digits.]
   5. Fixed-point models for SUE, as well as all the related analysis, can easily be extended to deal with VoT distributed among users (in much simpler way than UE with VoT distributed among users according to a r.v., which also looks rather inconsistent).

6. Fixed-point models for SUE, as well as all the related analysis, can easily be extended to deal with SUE with variable demand including any kind of non-separable demand models, whilst models for UE require the inverse demand function, not available in the most general case, and anyway hard to define and compute, apart from other limiting assumptions. Thus, SUE approached through fixed-point models is the only general option for equilibrium assignment with variable demand.
7. Solutions from a fixed-point model for SUE can easily be compared with fixed-point states of a day-to-day dynamic (deterministic) process model (see next chapter).
8. Fixed-point models for SUE are expressed by a (square) system of non-linear equations to be included in optimization models for transportation supply design, such as signal setting and street direction design, possibly including stability constraints derived from a day-to-day dynamic (deterministic) process model.

4. As said above SUE may effectively be modelled through highly flexible fixed-point models. Even though some **optimization models** are also available **for SUE**, (see appendix to Chapter 5 in Cascetta, 2009; Sheffi, 1985) as well for UE, the use of optimization models is somehow misleading, since the nature of equilibrium assignment models is descriptive/predictive (aiming at describing real world), not prescriptive (aiming at providing which decision is best to implement). Moreover, optimization models show several disadvantages with respect to fixed-point models:
   - the mathematical background of KKT conditions for non-linear continuous optimization is much more complicated with respect to (square) systems of non-linear equations and fixed-point models;
   - optimization models are harder to be formalised, that is: a proof should be provided that the optimization models actually describe the assignment problem at the hand, this kind of proof may not be easy to state and/or to understand;
   - their formulation may require additional hypotheses, such differentiability and/or symmetry of Jacobian matrix of cost function;
   - they are hard to be analysed [e.g. compare the proof of uniqueness through a FP model or an optimization one, see for instance Sheffi, 1985];
   - they can hardly be extended to general assignment problems, such as cost functions with asymmetric Jacobian, variable demand, multi-user multi-mode assignment, … .

Similar considerations also apply to VI models for SUE, as proposed in some recent papers. It should be noted these models are improper VI since solution is always an interior point of the set of solution and the VI always holds as an equality condition.

Appendix: Proofs of convergence conditions for MSA-based fixed-point algorithms

Convergence of MSA algorithms through a corollary of Blum's theorem for compact sets

Generally MSA-based algorithms do not provide the equilibrium arc flows in a finite number of iterations, but only a succession of arc flow patterns. They are a special instance of the Method of Successive convex Combinations (MSC); sufficient conditions for convergence of MSC-based methods may be stated through Blum's theorem, which also applies when only an unbiased estimation of arc flow function is available.

First use of MSA-based algorithm to solve equilibrium assignment with separable arc cost functions (with diagonal Jacobian) through an optimisation problem is in: Powell W.B. and Sheffi Y. (1982). The Convergence of Equilibrium Algorithms with Predetermined Step Sizes. Transportation Science, 16, 45–55.

Other step size strategies have been proposed to improve practical convergence of MSA-based algorithms, such as restarting strategies: index k is updated at each

Sufficient MSC convergence conditions - corollary of Blum's theorem over compact sets

Let $\boldsymbol{\varphi}(\mathbf{x})$ be a vector function from a non-empty compact convex set S to set $\boldsymbol{\varphi}(S) \subseteq S$, with a unique fixed-point $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*)$ in set S,

IF there exists

- a non negative real value function $\psi(\mathbf{x}) \geq 0$ defined over set S, with continuous first $\nabla\psi(\mathbf{x})$ and second $\nabla^2\psi(\mathbf{x})$ derivatives,

such that

a. $|\psi(\mathbf{x}) - \psi(\mathbf{x}^*)| > 0 \qquad \forall\, \mathbf{x} \in S, \mathbf{x} \neq \mathbf{x}^*$

b. $\nabla\psi(\mathbf{x})^T [\boldsymbol{\varphi}(\mathbf{x}) - \mathbf{x}] < 0 \qquad \forall\, \mathbf{x} \in S, \mathbf{x} \neq \mathbf{x}^*$

with $\nabla\psi(\mathbf{x}^*)^T [\boldsymbol{\varphi}(\mathbf{x}^*) - \mathbf{x}^*] = 0$

THEN the sequence $\mathbf{x}^k = \mathbf{x}^{k-1} + \alpha^k (\boldsymbol{\varphi}(\mathbf{x}^{k-1}) - \mathbf{x}^{k-1}) \in S$, with $\mathbf{x}^0 \in S$ and $\alpha^k \in]0,1]\ \forall k$ such that $\Sigma_k \alpha^k = \infty$ and $\Sigma_k (\alpha^k)^2 < \infty$, converges to the unique fixed-point $\mathbf{x}^*$.

Remark. Assessing convergence only requires that function $\psi(\mathbf{x})$ exists, but the sequence $\mathbf{x}^k$ is actually obtained without computing $\psi(\mathbf{x})$.

Remark. From any function bounded below a non negative function $\psi(\mathbf{x})$ can be obtained.

Remark. The most natural choice for function $\psi(\mathbf{x})$ is a strictly convex function that attains its minimum at the fixed-point $\mathbf{x}^*$.

Remark. Using weights $\alpha^k \in]0,1]$ assures that the sequence, $\mathbf{x}^k$, is inside set S, $\mathbf{x}^k \in S$, if it starts from a point $\mathbf{x_o} \in S$, since set S has been assumed convex.

Remark. The sequence of weights $\alpha^k = (1/k)$ is the sequence with largest elements $\alpha^k \in]0,1]$ such that $\Sigma_{k>0}\, \alpha^k = \infty$ and $\Sigma_{k>0}\, (\alpha^k)^2 < \infty$; it gives the Method of Successive Averages (MSA).

Remark. If $\mathbf{x}^k$ is an unbiased realisation of a sequence of random variables almost surely convergence is assured.

[For the original statement of Blum's theorem see: Blum J.R. (1954). Multidimensional Stochastic Approximation Methods. Ann. Math. Stat. 25, 737–744.]

iteration as $k = k+1$ starting from $k = k_o$, with $k_o = 1$ at first iteration, then index k is restarted from k_o after k^* iterations (with $k^* = \infty$ we get MSA), and k_o is either equal to 1 again, $k_o = 1$, or increased by one, $k_o = k_o + 1$.

The Method of Repeated Approximations (MRA) is obtained with $\alpha^k = 1$, but in this case convergence conditions of the Blum's theorem are not satisfied, since $\Sigma_k (\alpha^k)^2 = \infty$. In this case sufficient convergence conditions can be stated through the Banach's theorem and its corollaries, see p.14; conditions on function $\boldsymbol{\varphi}(\mathbf{x})$ are very strong and not often satisfied.

Convergence of the MSA-FA algorithm

As noted in the main text (p. 12) applying the Method of Successive Averages to model (3.17) the MSA-FA algorithm is obtained based on the recursive equation:

$$\mathbf{f}^k = \mathbf{f}^{k-1} + (1/k)\left(\mathbf{f}\left(\mathbf{c}\left(\mathbf{f}^{k-1}\right) - \mathbf{f}^{k-1}\right) \in S_f \quad \text{with}\, \mathbf{f}^0 = \mathbf{f_0} \in S_f (3.19)$$

It can be proved asymptotically converging to $\mathbf{f^*} = \mathbf{f}(\mathbf{c}(\mathbf{f^*}; \boldsymbol{\kappa}); \mathbf{d}) \in S_f$ if the existence and uniqueness /arc conditions hold and the Jacobian of the arc cost function is symmetric by applying the corollary of Blum's theorem over compact sets as shown below.

The proof shows that the assumptions of the corollary of Blum's theorem over compact sets hold, with reference to the MSA, if

- for existence conditions:

E1. each o-d pair i is connected by at least one route,

thus the feasible arc flow set S_f is non-empty, beside being compact and convex,

E2. the arc cost function $\mathbf{c}(\mathbf{f})$ is continuous with respect to the arc flows $\mathbf{f}$,
E3. the arc flow function $\mathbf{f}(\mathbf{c})$ is continuous with respect to the arc costs $\mathbf{c}$;

- for uniquenss conditions /arc:

U1. the arc cost function $\mathbf{c}(\mathbf{f})$ is s. monotone increasing with respect to the arc flows $\mathbf{f}$,
U2. the arc flow function $\mathbf{f}(\mathbf{c})$ is monotone decreasing with respect to the arc costs $\mathbf{c}$;

- moreover:

S1. the arc cost function $\mathbf{c}(\mathbf{f})$ is continuously differentiable with symmetric positive definite for real vectors Jacobian matrix $\nabla \mathbf{c}(\mathbf{f}) \succ 0$.

Let $\boldsymbol{\varphi}(\mathbf{x}) = \mathbf{f}(\mathbf{c}(\mathbf{x}))$ as in the corollary of Blum's theorem over a compact set; this fuction has exactly one fixed-point $\mathbf{f^*} \in S_f$ for existence and uniqueness conditions; let $\mathbf{c}^* = \mathbf{c}(\mathbf{f^*})$.

The vector valued function $\mathbf{c}(\mathbf{x})$ - $\mathbf{c}^*$ is continuous over set S_f (according to existence conditions E2) with cont. Symmetric Jacobian $\nabla[\mathbf{c}(\mathbf{x}) - \mathbf{c}^*] = \nabla \mathbf{c}(\mathbf{x})$ (for hypothesis S1). Thus, there exists a real valued function $\psi(\mathbf{x})$, defined over set S_f, for which gradient and Hessian matrix are $\nabla \psi(\mathbf{x}) = \mathbf{c}(\mathbf{x}) - \mathbf{c}^*$ and $\nabla^2 \psi(\mathbf{x}) = \nabla[\mathbf{c}(\mathbf{x}) - \mathbf{c}^*] = \nabla \mathbf{c}(\mathbf{x})$, respectively. Hence, function $\psi(\mathbf{x})$ is twice differentiable.

Since the arc cost function $\mathbf{c}(\mathbf{x})$ is strictly monotone increasing (according to uniqueness conditions U1), $\nabla \psi(\mathbf{x}) = \mathbf{c}(\mathbf{x}) - \mathbf{c}^*$ is a vector-valued strictly monotone increasing function. Thus, function $\psi(\mathbf{x})$ is strictly convex over set S_f. Since $\nabla \psi(\mathbf{f}^*) = \mathbf{c}(\mathbf{f}^*) - \mathbf{c}^* = \mathbf{0}$, function $\psi(\mathbf{x})$ has a unique

Continued

minimum at $\mathbf{f}^*$, $\psi(\mathbf{x}) > \psi(\mathbf{f}^*)\ \forall \mathbf{x} \neq \mathbf{f}^*$, $\mathbf{x} \in S_f$, and $\psi(\mathbf{x})$ is bounded from below. Thus, assumption **a** holds for function $\psi(\mathbf{x})$.

Moreover, condition $\mathbf{f}^* = \mathbf{f}(\mathbf{c}(\mathbf{f}^*))$ yields:

- for $\mathbf{x} = \mathbf{f}^* \in S_f$

$$\nabla\psi(\mathbf{f}*)^{\mathrm{T}}[\varphi(\mathbf{f}^*) - \mathbf{f}^*] = \nabla\varphi(\mathbf{f}*)^{\mathrm{T}}[\mathbf{f}(\mathbf{c}(\mathbf{f}^*)) - \mathbf{f}^*] = 0$$

- for $\mathbf{x} \neq \mathbf{f}^* \in S_f$

$$\nabla\psi(\mathbf{x})^{\mathrm{T}}[\varphi(\mathbf{x}) - \mathbf{x}] = [\mathbf{c}(\mathbf{x}) - \mathbf{c}*]^{\mathrm{T}}[\mathbf{f}(\mathbf{c}(\mathbf{x})) - \mathbf{x}] = [\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{f}*)]^{\mathrm{T}}[\mathbf{f}(\mathbf{c}(\mathbf{x})) - \mathbf{f}(\mathbf{c}(\mathbf{f}^*)) + \mathbf{f}^* - \mathbf{x}] =$$
$$= [\mathbf{f}(\mathbf{c}(\mathbf{x})) - \mathbf{f}(\mathbf{c}(\mathbf{f}*))]^{\mathrm{T}}[\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{f}^*)] - [\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{f}*)]^{\mathrm{T}}[\mathbf{x} - \mathbf{f}^*]$$

The arc cost function is strictly monotone increasing (uniqueness condition U1) and the arc flow function is monotone decreasing (uniqueness conditions U2), therefore.

$$[\mathbf{f}(\mathbf{c}(\mathbf{x})) - \mathbf{f}(\mathbf{c}(\mathbf{f}*))]^{\mathrm{T}}[\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{f}^*)] \leq 0 \text{ and } -[\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{f}*)]^{\mathrm{T}}[\mathbf{x} - \mathbf{f}^*] < 0 \text{ thus } \nabla\psi(\mathbf{x})^{\mathrm{T}}[\varphi(\mathbf{x}) - \mathbf{x}] < 0$$

Hence, assumption **b** holds for function $\psi(\mathbf{x})$.

[This proof is from Cantarella, 1997, who slightly adapted the Feasible Flow Space Convergence Corollary in Daganzo, 1983.]

Weak form of the Corollary of Blum's theorem over compact sets.

Let $\boldsymbol{\varphi}(\mathbf{x})$ be a vector function from a non-empty compact convex set S to set $\boldsymbol{\varphi}(S) \subseteq S$, with a unique fixed-point $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*)$ in set S,

IF there exists

– a non negative real value function $\psi(\mathbf{x})$ defined over set S, with continuous first $\nabla\psi(\mathbf{x})$ and second $\nabla^2\psi(\mathbf{x})$ derivatives, and
– a subset $S^\bullet \subseteq S$, with $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^\bullet)\quad \forall\, \mathbf{x}^\bullet \in S^\bullet$, $\mathbf{x}^* \in S^\bullet$

such that

a. $|\psi(\mathbf{x}) - \psi(\mathbf{x}^\bullet)| > 0 \quad \forall\, \mathbf{x} \in S, \mathbf{x} \notin S^\bullet$

with $|\varphi(\mathbf{x}^\bullet) - \varphi(\mathbf{x}^*)| = 0 \quad \forall\, \mathbf{x}^\bullet \in S^\bullet$

b. $\nabla\psi(\mathbf{x})^{\mathrm{T}}[\boldsymbol{\varphi}(\mathbf{x}) - \mathbf{x}] < 0 \quad \forall\, \mathbf{x} \in S, \mathbf{x} \notin S$

with $\nabla\psi(\mathbf{x}^\bullet)^{\mathrm{T}}[\boldsymbol{\varphi}(\mathbf{x}^\bullet) - \mathbf{x}^\bullet] = 0 \quad \forall\, \mathbf{x}^\bullet \in S$

THEN the sequence $\mathbf{x}^k = \mathbf{x}^{k\text{-}1} + \alpha^k(\psi(\mathbf{x}^{k\text{-}1}) - \mathbf{x}^{k\text{-}1})$, with $\mathbf{x}^0 = \mathbf{x_o} \in S$ and $\alpha^k \in]0,1]\ \forall k$ such that $\Sigma_{k>0}\,\alpha^k = \infty$ and $\Sigma_{k>0}\,(\alpha^k)^2 < \infty$, converges to a point $\mathbf{x}^\bullet$ in set $S^\bullet$, thus called the convergence set. The fixed-point $\mathbf{x}^*$ can be computed as $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^\bullet)$.

Remark. Assessing convergence only requires that function $\psi(\mathbf{x})$ and subset $S^\bullet$ exist, but the sequence $\mathbf{x}^k$ is actually obtained without computing $\psi(\mathbf{x})$ (or subset $S^\bullet$).

Remark. From any function bounded below a non negative function $\psi(\mathbf{x})$ can be obtained.

Remark. The most natural choice for function $\psi(\mathbf{x})$ is a convex function that attains its minimum at any point $\mathbf{x}^\bullet \in S^\bullet$, in this case set $S^\bullet$ is convex.

All other remarks for the standard form of the corollary still applies.

[See details in Cantarella, 1997.]

As already noted, the most natural choice for function $\psi(\mathbf{x})$ is a strictly convex one with its minimum at the fixed-point $\mathbf{x}^*$. When only a convex, but not strictly convex, function $\psi(\mathbf{x})$ can be found, a weaker form of Blum's theorem can be adopted.

Convergence of the MSA-CA algorithm

As noted in the main text (p. 12) applying the Method of Successive Averages to model (3.18) the MSA-CA algorithm is obtained based on the recursive equation:

$$\mathbf{c}^k = \mathbf{c}^{k-1} + (1/k)\left(\mathbf{c}\left(\mathbf{f}\left(\mathbf{c}^{k-1}\right) - \mathbf{c}^{k-1}\right) \in S_c \quad \text{with}\ \mathbf{c}^0 = \mathbf{c}(\mathbf{f_0})\ \text{and}\ \mathbf{f_0} \in S_f (3.20)$$

where

- S_f is the feasible arc flow set non empty (if each o-d pair is connected by at least one route), compact (since closed and bounded), and convex; it also is the domain of the arc cost function $\mathbf{c}(\cdot)$;
- $\mathbf{c}(S_f)$ is the co-domain of the arc cost function $\mathbf{c}(\cdot)$; it is non empty and compact if the arc cost function $\mathbf{c}(\cdot)$is continuous, but generally not convex (unless the arc cost function were linear);
- S_c is the convex hull of set $\mathbf{c}(S_f)$; it is non empty and compact, as set $\mathbf{c}(S_f)$, and convex by definition.

If the existence and uniqueness /arc conditions hold and the Jacobian of the arc flow function is symmetric, the MSA-CA algorithm can be proved asymptotically converging to an arc cost vector $\mathbf{c}^\bullet \in S_c$ such that $\mathbf{c}^* = \mathbf{c}(\mathbf{f}(\mathbf{c}^\bullet;\ \mathbf{d});\ \boldsymbol{\kappa}) \in \mathbf{c}(S_f)$ and $\mathbf{f}^* = \mathbf{f}(\mathbf{c}^*;\ \mathbf{d}) \in S_f$ by applying the weak form of the corollary of Blum's theorem over compact sets as shown below. The arc cost vector $\mathbf{c}^\bullet$ does not need to be unique, and generally is not. Let.

$S^\bullet \triangleq \{\mathbf{c}^\bullet \in S_c : \mathbf{c}^* = \mathbf{c}(\mathbf{f}(\mathbf{c}^\bullet;\ \mathbf{d});\ \boldsymbol{\kappa})\}$ be the convergence set, with $\mathbf{c}^* \in S^\bullet$ and $\mathbf{f}^* = \mathbf{f}(\mathbf{c}^\bullet)$, since $\mathbf{c}^* = \mathbf{c}(\mathbf{f}^*)$.

The proof shows that the assumptions of the weak form of the corollary of Blum's theorem over compact sets hold, with reference to the MSA, If

- for existence conditions:

E1. each o-d pair i is connected by at least one route,

thus the feasible arc flow set S_f is non-empty, beside being compact and convex, therefore S_c is non-empty, compact and convex (see also E2),

E2. the arc cost function $\mathbf{c}(\mathbf{f})$ is continuous with respect to the arc flows $\mathbf{f}$,
E3. the arc flow function $\mathbf{f}(\mathbf{c})$ is continuous with respect to the arc costs $\mathbf{c}$;

- for uniquenss conditions /arc:

U1. the arc cost function $\mathbf{c}(\mathbf{f})$ is s. monotone increasing with respect to the arc flows $\mathbf{f}$,

---

Continued

U2. the arc flow function $\mathbf{f}(\mathbf{c})$ is monotone decreasing with respect to the arc costs $\mathbf{c}$ (as it occurs for invariant RUMs);

- moreover:

S2. the arc flow function $\mathbf{f}(\mathbf{c})$ is continuously differentiable with symmetric (as it occurs for invariant RUMs) negative semi-definite for real vectors Jacobian matrix, $-\nabla\mathbf{f}(\mathbf{c}) \succeq 0$.

Let $\boldsymbol{\varphi}(\mathbf{x}) = \mathbf{c}(\mathbf{f}(\mathbf{x}))$ as in the weak form of the corollary of Blum's theorem over a compact set; this fuction has exactly one fixed-point $\mathbf{c}^* \in \mathbf{c}(S_f) \subseteq S_c$, a non-empty, compact and convex set for existence and uniqueness conditions; let $\mathbf{f}^* = \mathbf{f}(\mathbf{c}^*)$. Moreover, let $S^\bullet$ be defined as above.

The vector valued function $\mathbf{f}^* - \mathbf{f}(\mathbf{x})$ is continuous over set S_c (according to existence conditions E3) with cont. Symmetric Jacobian $\nabla[\mathbf{f}^* - \mathbf{f}(\mathbf{x})] = -\nabla\mathbf{f}(\mathbf{x})$ (for hypothesis S2). Thus, there exists a real valued function $\psi(\mathbf{x})$, defined over set S_c, for which gradient and Hessian matrix are $\nabla\psi(\mathbf{x}) = \mathbf{f}^* - \mathbf{f}(\mathbf{x})$ and $\nabla^2\psi(\mathbf{x}) = \nabla[\mathbf{f}^* - \mathbf{f}(\mathbf{x})] = -\nabla\mathbf{f}(\mathbf{x})$, respectively. Hence, function $\psi(\mathbf{x})$ is twice differentiable.

Since the arc flow function $\mathbf{f}(\mathbf{x})$ is monotone decreasing (according to uniqueness conditions U2), $\nabla\psi(\mathbf{x}) = \mathbf{f}^*$ - $\mathbf{f}(\mathbf{x})$ is a vector-valued monotone increasing function over set S_c.. Thus, function $\psi(\mathbf{x})$ is convex over set S_c. Since $\nabla\psi(\mathbf{c}^*) = \mathbf{f}^* - \mathbf{f}(\mathbf{c}^*) = \mathbf{0}$, function $\psi(\mathbf{x})$ has a minimum at $\mathbf{c}^* \in \mathbf{c}(S_f) \subseteq S_c$, or $\psi(\mathbf{x}) \geq \psi(\mathbf{c}^*)$, $\forall \mathbf{x} \in S_c$, and $\psi(\mathbf{x})$ is bounded from below. Moreover, since $\nabla\psi(\mathbf{c}^\bullet) = \mathbf{f}^* - \mathbf{f}(\mathbf{c}^\bullet) = \mathbf{0}\ \forall\ \mathbf{c}^\bullet \in S^\bullet$, function $\psi(\mathbf{x})$ has a minimum at any point $\mathbf{c}^\bullet \in S^\bullet$ (but not at any other point) hence $\psi(\mathbf{x}) > \psi(\mathbf{c}^\bullet) = \psi(\mathbf{c}^*)$, $\forall \mathbf{x} \notin S^\bullet$ (and set $S^\bullet$ is convex). Thus assumption **a** holds for function $\psi(\mathbf{x})$.

- for $\mathbf{c}^\bullet \in S^\bullet \subseteq S_c$, that is $\mathbf{f}(\mathbf{c}^\bullet) = \mathbf{f}^* = \mathbf{f}(\mathbf{c}^*)$

$$\nabla\psi(\mathbf{c}^\bullet)^{\mathrm{T}}[\varphi(\mathbf{c}^\bullet) - \mathbf{c}^\bullet] = [\mathbf{f}* - \mathbf{f}(\mathbf{c}^\bullet)]^{\mathrm{T}}[\mathbf{c}(\mathbf{f}(\mathbf{c}^\bullet)) - \mathbf{c}^\bullet] = 0$$

- for $\mathbf{x} \in S_c$, $\mathbf{c} \notin S^\bullet \subseteq S_c$, that is $\mathbf{f}(\mathbf{x}) \neq \mathbf{f}^* = \mathbf{f}(\mathbf{c}^*)$

$$\nabla\psi(\mathbf{x})^{\mathrm{T}}[\varphi(\mathbf{x}) - \mathbf{x}] = [\mathbf{f}* - \mathbf{f}(\mathbf{x})]^{\mathrm{T}}[\mathbf{c}(\mathbf{f}(\mathbf{x})) - \mathbf{x}] =$$

$$= [\mathbf{f}(\mathbf{c}*) - \mathbf{f}(\mathbf{x})]^{\mathrm{T}}[\mathbf{c}(\mathbf{f}(\mathbf{x})) - \mathbf{c}(\mathbf{f}(\mathbf{c}^*)) + \mathbf{c}^* - \mathbf{x}] =$$

$$= [\mathbf{f}(\mathbf{c}*) - \mathbf{f}(\mathbf{x})]^{\mathrm{T}}[\mathbf{c}(\mathbf{f}(\mathbf{x})) - \mathbf{c}(\mathbf{f}(\mathbf{c}^*))] + [\mathbf{f}(\mathbf{c}*) - \mathbf{f}(\mathbf{x})]^{\mathrm{T}}[\mathbf{c}^* - \mathbf{x}] =$$

$$= -[\mathbf{c}(\mathbf{f}(\mathbf{x})) - \mathbf{c}(\mathbf{f}(\mathbf{c}*))]^{\mathrm{T}}[\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{c}^*)] + [\mathbf{f}(\mathbf{c}*) - \mathbf{f}(\mathbf{x})]^{\mathrm{T}}[\mathbf{c}^* - \mathbf{x}]$$

The arc cost function is strictly monotone increasing (uniqueness condition U1) and the arc flow function is monotone decreasing (uniqueness conditions U2), therefore

$$-[\mathbf{c}(\mathbf{f}(\mathbf{x})) - \mathbf{c}(\mathbf{f}(\mathbf{c}*))]^{\mathrm{T}}[\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{c}^*)] < 0 \text{ and } [\mathbf{f}(\mathbf{c}*) - \mathbf{f}(\mathbf{x})]^{\mathrm{T}}[\mathbf{c}^* - \mathbf{x}] \leq 0$$

thus $\nabla\psi(\mathbf{x})^{\mathrm{T}}[\varphi(\mathbf{x}) - \mathbf{x}] < 0$

Hence, assumption **b** holds for function $\psi(\mathbf{f})$.

[This proof has been adapted from Cantarella, 1997.]

References

Bifulco, G.N., 1993. A stochastic user equilibrium assignment model for the evaluation of parking policies. Eur. J. Oper. Res. 71, 269–287.

Cantarella, G.E., 1997. A general fixed-point approach to multimode multi-user equilibrium assignment with elastic demand. Transp. Sci. 31, 107–128.

Cantarella, G.E., Binetti, M.G., 1998. Stochastic equilibrium traffic assignment with value-of-time distributed among users. Int. Trans. Oper. Res. 5 (6), 541–553.

Cascetta, E., 2009. Transportation Systems Analysis: Models and Applications. Springer.

Daganzo, C., 1983. Stochastic network equilibrium with multiple vehicle types and asymmetric, indefinite link cost Jacobians. Transp. Sci. 17, 282–300.

Daganzo, C., Sheffi, Y., 1977. On stochastic models of traffic assignment. Transp. Sci. 11, 253–274.

Gentile, G., Nökel, K., 2016. Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems: COST Action TU1004 (TransITS). Springer.

Ghatee, M., Hashemi, S.M., 2009. Traffic assignment model with fuzzy level of travel demand: an efficient algorithm based on quasi-Logit formulas. EURO 194, 432–451.

Henn, V., 2000. Fuzzy route choice model for traffic assignment. Fuzzy Sets Syst. 116, 77–101.

Henn, V., Ottomanelli, M., 2006. Handling uncertainty in route choice models: from probabilistic to possibilistic approaches. EURO 175, 1526–1538.

Maher, M.J., Stewart, K., Rosa, A., 2005. Stochastic social optimum traffic assignment. Transp. Res. B 39B (8), 756–767.

Powell, W.B., Sheffi, Y., 1982. The convergence of equilibrium algorithms with predetermined step sizes. Transp. Sci. 16, 45–55.

Rasmussen, T.K., Watling, D.P., Prato, C.G., Nielsen, O.A., 2015. Stochastic user equilibrium with equilibrated choice sets: part II—solving the restricted SUE for the logit family. Transp. Res. B: Methodol. 77, 146–165.

Ridwan, M., 2004. Fuzzy preference based traffic assignment problem. Transp. Res. C: Emerg. Technol. 12, 209–233.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods. Prentice-Hall, Inc, Englewood Cliffs.

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proc. Inst. Civ. Eng. 2 (1), 325–378.

Watling, D.P., Rasmussen, T.K., Prato, C.G., Nielsen, O.A., 2015. Stochastic user equilibrium with equilibrated choice sets: part I—model formulations under alternative distributions and restrictions. Transp. Res. B: Methodol. 77, 166–181.

Watling, D.P., Rasmussen, T.K., Prato, C.G., Nielsen, O.A., 2018. Stochastic user equilibrium with a bounded choice model. Transp. Res. B: Methodol. 114, 254–280.

CHAPTER

Assignment to congested networks: Day-to-day dynamics—Deterministic processes

4

Giulio Erberto Cantarella
University of Salerno, Salerno, Italy

There is another danger which
you can scarcely hope to escape.
It is the weight of the past.
Vita Sackville-West

Outline. This chapter describes a comprehensive modelling approach to day-to-day dynamic assignment to congested networks through discrete-time Markovian deterministic process (DP) models; presented models are consistent with the SEAM modelling framework presented in Chapter 1; first simple models for cost and choice update are introduced and discussed, then deterministic process models based on them with respect to flows and costs are introduced and discussed. At last more general models encompassing most existing models are discussed.

Methods for day-to-day (or inter-periodic) dynamic assignment play a central role in advanced transportation system analysis, since they allow to analyse and forecast equilibrium stability and fluctuations around it, as a result of past events.

This chapter discusses deterministic process (DP) models for day-to-day (or intra-periodic) dynamic assignment to congested transportation networks, a kind of assignment still at research level and not yet fully implemented in commercial software. They will turn out closely related to stochastic process (SP) models, described in the next Chapter 5.

Even though exactly one user equilibrium flow and cost patterns exist (existence and uniqueness conditions hold), the equilibrium analysis discussed in the previous chapter does not allow to analyse whether (and under which conditions) the system state, flow and cost patterns, evolves towards the user equilibrium, and, if not whether, it evolves towards some kind of attractor. This stability analysis can be addressed through DP models, discussed in this chapter.

Dynamics and Stochasticity in Transportation Systems. https://doi.org/10.1016/B978-0-12-814353-7.00004-2
© 2020 Elsevier Inc. All rights reserved.

Moreover, the equilibrium analysis does not allow to analyse transients after demand and/or supply changes, nor to obtain a statistical description of the system state evolution over time, i.e. means, modes, moments and, more generally, frequency distributions, this kind of analysis requiring SP models discussed in the next chapter.

It should be stressed that day-to-day dynamics ontologically occur over discrete time, while within-day dynamics, discussed in Chapter 6, occur over continuous time, say any instant of time within a day (cfr Introduction). Thus in this chapter we will only discussed discrete-time deterministic process models, say models derived from the discrete-time non-linear dynamic system theory; apart from being the only consistent approach to day-to-day dynamics, it also allow us to easily bridge to (discrete) time-driven SP models, described in next chapter.

Some comments on DP models over continuous time are reported in the Summary at the end of this chapter, since some authors have proposed such a kind of models in the past. The special case of all users following a maximum utility or minimum cost routes will not be discussed since rather unrealistic under day-to-day dynamics, and not consistent with SP models; moreover, many models based on these behavioural assumptions are also based on continuous time modelling.

It is worth noting that in his seminal paper, Wardrop (1952) alluded to the role of dynamic adaptation in transportation networks. When providing a justification for the introduction of the equilibrium concept, by stating that 'it may be assumed that traffic will tend to settle down into an equilibrium situation'.

Horowitz (1984) was the first to explicitly describe day-to-day dynamics, proposing DP models for a two-link transportation network derived from discrete-time non-linear dynamic system theory. Cascetta (1987, 1989) was the first to propose SP models (more details in the next chapter) for analyse day-to-day dynamics in transportation systems. He also stressed that equilibrium models should be considered a special cases of day-to-day dynamic ones. Cantarella and Cascetta (1995) were the first to propose a unifying general theory, based on RUM, encompassing FP models for UE assignment and DP and SP models for DD assignment to general transportation networks. Since then several papers have been proposed, with an increasing interest in the last decade.

Even though day-to-day is the term commonly used in literature a much better term would be epoch, which, as stated in Cascetta (1989) 'can have either a "chronological" interpretation as successive reference periods of similar characteristics (e.g. the a.m. peak period of successive working days) or they can be defined as "fictitious" moments in which users acquire awareness of path attributes and make their choices'.

The specification of a DP models for day-to-day dynamic assignment requires an extension of both the supply and the demand models 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.

In this chapter, for pedagogical purpose we first introduce and discuss some simple DP models for day-to-day dynamic assignment to congested networks, including fixed-point models as a special case; one of these approaches allows also to analytically derive several interesting results about fixed-point stability. Then general DP models are discussed. They may include most route choice modelling approaches (cfr Appendix to the book). They are described under steady-state conditions, but they also apply to any TAN used for within-day dynamics, as discussed in Chapter 6. Presented DP models are consistent with the SEAM modelling framework presented in Chapter 1.

Section 4.1 introduces basic equations for simple deterministic process models discussed in Section 4.2; in Sections 4.3 and 4.4 long-term evolution over time is discussed and fixed-point stability and bifurcations are analysed respectively for those simple models; general models are discussed in Sections 4.5 and 4.6.

4.1 Basic equations for simple DP models

This section presents the basic equations for (within-day static) day-to-day dynamic assignment through DP models adding the cost updating and the choice updating filters to the six equations introduced in the previous Chapter 3; the presented approach can straightforwardly be applied to i-route variables instead of route ones and/or to multi-class assignment as well. All definitions and assumptions introduced in the previous chapters still hold, unless otherwise stated. Main vector notations from Chapters 2 and 3 as well few new ones used in the following are enlisted below in alphabetical order for reader's convenience (sets come first, then Roman letters, at last Greek letters). Variables that may change over the day have a superscript, usually k. When any ambiguity might occur with power exponent, the argument is between round brackets.

A is the set of arcs, with $m = |A|$ elements;
$\mathbb{E}^m$ is the set of real $m \times 1$ (column) vectors with Euclidean distance;
m is the number of arcs;
$\mathbb{N}$ is the set of natural numbers, that is positive integers;
n_i is the number of routes connecting o-d pair i;
$n = \sum_i n_i$ is the number of routes connecting all o-d pairs;
R_i is the set of routes for o-d pair i, with $n_i = |R_i|$ elements;
$R = \cup_i R_i$ is the set of routes for all o-d pairs, with $n = |R| = \sum_i n_i$ elements;
$S_c \subseteq \mathbb{E}^m$ is the arc cost set, given by the convex hull of set $\mathbf{c}(S_f)$;
$S_f \subseteq \mathbb{E}^m$ is the feasible arc flow set;
$S_{h_i} \subseteq \mathbb{E}^{ni}$ is the feasible route flow set for o-d pair i;
$S_h \subseteq \mathbb{E}^i$ is the feasible route flow set;

$S_{w_i} \subseteq \mathbb{E}^{ni}$ is the route cost set for o-d pair i, an affine transformation of the arc cost set S_c;

$\mathbf{B}_i$ is the $(m \times n_i)$ i-th block of the *ARGIM* for o-d pair i;
$\mathbf{B}$ is the $(m \times n)$ (row block) *ARGIM*;
$\mathbf{c}^k \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *actual arc costs* on day k;
$\mathbf{c}(\cdot)$ is the $m \times 1$ (column) *arc cost function*;
$d_i \geq 0$ is the *demand flow* for o-d pair i;
$\mathbf{d} \geq \mathbf{0}$ is the *demand flow* vector with entries d_i;
$\mathbf{f}^k \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *total arc flow*s on day k;
$\mathbf{f}_\mathbf{Z} \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *other arc flows*;
$\mathbf{f}(\cdot)$ is the $m \times 1$ (column) *arc flow function*;
$\mathbf{h}^k \geq \mathbf{0}$ is the $n \times 1$ (column) vector of *route flow*s for all o-d pairs on day k;
$\mathbf{h}^k;_i \geq \mathbf{0}$ is the $n_i \times 1$ i-th block of the (column) vector of *route flow*s for o-d pair i on day k;
$\mathbf{h}_i(\cdot)$ is the $n_i \times 1$ (column) vector *route flow function* for o-d pair i;
$\mathbf{h}(\cdot)$ is the $n \times 1$ (column) *route flow function* for all o-d pairs;
$\mathbf{p}^k;_i \geq \mathbf{0}$ is the $n_i \times 1$ i-th block of the (column) vector of *route choice proportions* for o-d pair i, with $\mathbf{1}^\mathbf{T}\mathbf{P}^k;_i = 1$, on day k;
$\mathbf{p}_i(\cdot)$ is the $n_i \times 1$ (column) vector *route choice function* for o-d pair i;
$\mathbf{v}^k;_i$ is the $n_i \times 1$ i-th block of the (column) vector of *route systematic utility* for o-d pair i on day k;
$\mathbf{w}^k \geq \mathbf{0}$ is the $n \times 1$ (column) vector of *actual route costs* on day k;
$\mathbf{w}^k;_i \geq \mathbf{0}$ is the $n_i \times 1$ i-th block of the (column) vector of *actual route costs* for o-d pair i on day k;
$\mathbf{w}_{\mathbf{Z}i} \geq \mathbf{0}$ is the $n_i \times 1$ i-th block of the (column) vector of *other route costs* for o-d pair i;
$\mathbf{w}_\mathbf{Z} \geq \mathbf{0}$ is the $n_i \times 1$ block of the (column) vector of *other route costs* for o-d pair i;
$\mathbf{w}(\cdot)$ is the $n \times 1$ (column) *actual route cost function* for all o-d pairs;
$\mathbf{w}_i(\cdot)$ is the $n_i \times 1$ block the (column) vector *actual route cost function* for o-d pair i;
$\mathbf{x}^k \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *forecasted arc costs* for day k;
$\mathbf{y}^k;_i \geq \mathbf{0}$ is the $n_i \times 1$ i-th block of the (column) vector of *forecasted route costs* for o-d pair i on day k;
$\mathbf{y}^k \geq \mathbf{0}$ is the $n \times 1$ (column) vector of *total route forecasted costs* on day k;

$\alpha \in]0, 1[$ is the *choice updating parameter*;
$\beta \in]0, 1[$ is the *cost updating parameter*;
ζ_j is the weight given to the actual cost occurred in any of the μ previous days, in a Moving Average filter
$\boldsymbol{\theta}_i > 0$ is the vector of the *route choice function parameters* for o-d pair i;
$\kappa_a > 0$ is the *capacity* of arc a;
$\boldsymbol{\kappa} > 0$ is the $m \times 1$ (column) vector of the *arc capacities*, with entries κ_a;
$\mu > 1$ is the integer *memory depth*, in a Moving Average filter;
$\psi_i > 0$ is the *utility scale parameter* in the route choice model, for o-d pair i.

4.1.1 Supply models for simple DP models

Transportation supply models express how user behaviour affects network performances. This section describes the three equations that according to the SEAM framework specify the transportation supply model for the day-to-day dynamics of a transportation system, needing an extension of EQN3 (or EQN2).

- Arc-route flow consistency relation

Under the steady-state assumption the yesterday arc flows due to all o-d pairs can be obtained from the yesterday route flows through an affine transformation from the route space to the arc space defined by the arc-route generalised incidence matrix (cfr Eq. 3.1), given $\mathbf{h}_i^0 \forall i$, blocks of $\mathbf{h}^0$:

$$\mathbf{f}^{k-1} = \sum\nolimits_i \mathrm{B}_i \cdot \mathbf{h}^{k-1}{;}_i + \mathbf{f_Z} \in S_f \quad \forall k \in \mathbb{N}$$
$$\text{or} \quad \mathbf{f}^{k-1} = \mathrm{B} \cdot \mathbf{h}^{k-1} + \mathbf{f_Z} \quad \forall k \in \mathbb{N} \tag{4.1}$$

omitting other arc flows $\mathbf{f_Z}$ in the following for simplifying notation.

- Arc cost function

Due to congestion, say driving user behaviour, yesterday actual arc costs depend on yesterday arc flows (cfr Eq. 3.2):

$$\mathbf{c}^{k-1} = \mathbf{c}\left(\mathbf{f}^{k-1}; \boldsymbol{\kappa}\right) \geq \mathbf{0} \in S_c \quad \forall \mathbf{f}^{k-1} \in S_f \forall k \in \mathcal{N} \tag{4.2}$$

where $\boldsymbol{\kappa} > 0$ is the vector of the *arc capacities*. The arc cost flow function as well its parameters are assumed day-invariant.

- Route-arc cost updating function

When modelling day-to-day dynamics the actual costs, resulting from congestion, are usually different from the forecasted costs affecting user choice behaviour, namely the route utility values in the demand models (see next sub-section).

The yesterday actual route costs for o-d pair i can be obtained from the yesterday actual arc costs through a transformation from the arc space to the route space defined by the transpose of arc-route generalised incidence matrix (cfr Eq. 3.3):

$$\mathbf{w}^{k-1}{;}_i = \mathbf{B}_i^{\mathbf{T}} \cdot \mathbf{c}^{k-1} + \mathbf{w}_{\mathbf{Z}i} \in S_c \quad \forall i \ \forall k \in \mathcal{N}$$

omitting other route costs $\mathbf{w}_{\mathbf{Z}i}$ and $\mathbf{w}_{\mathbf{Z}}$ in the following for simplifying notation.

After knowing yesterday actual costs, users, supported by words of mouth and possibly informative systems, elaborate them to get forecast costs. Let $\mathbf{y}^k{;}_i \geq \mathbf{0}$ be the $n_i \times 1$ block of the (column) vector of *total route forecasted costs* for o-d pair i on day k.

The *route cost updating filter* models how today forecasted costs are affected by yesterday actual cost as well as further past actual costs, that is user memory and learning process. Some simple models are described below, general ones in Section 4.5.

– Only yesterday filter

In the most straightforward approach to cost updating, the today forecasted costs are assumed equal the yesterday actual costs without being affected by further past costs; the so-called only yesterday (OY) filter is specified by the following recursive equation:

$$\mathbf{y}^k{}_{;i} = \mathbf{w}^{k-1}{}_{;i} \quad \forall i\ \forall k \in \mathbb{N}$$

Given $\mathbf{c}^0$, combining the route-arc consistency relation with the OY filter leads to:

$$\mathbf{y}^k{}_{;i} = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}^{k-1} \quad \forall i\ \forall k \in \mathbb{N} \tag{4.3.1}$$

– Exponential smoothing filter

According to an exponential smoothing, ES(β), filter they are defined by a strict convex combination of yesterday forecasted and yesterday actual costs, as specified by the following recursive equation, given

$$\mathbf{y}^k{}_{;i} = \beta \mathbf{w}^{k-1}{}_{;i} + (1-\beta)\mathbf{y}^{k-1}{}_{;i} \quad \forall i\ \forall k \in \mathbb{N}$$

where $\beta \in]0, 1[$ is the *cost updating parameter*, that is the weight given to yesterday actual costs; in the following the cost updating parameter β is assumed time invariant and common to all users.

Given $\mathbf{c}^0$, and $\mathbf{y}^0{}_{;i} = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}^0 \forall i$ combining the arc flow function with the ES(β) filter leads to:

$$\mathbf{y}^k{}_{;i} = \beta\, \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}^{k-1} + (1-\beta)\, \mathbf{y}^{k-1}{}_{;i} \quad \forall k \in \mathbb{N} \tag{4.3.2}$$

The OY filter (4.3.1) is a limit case obtained putting $\beta = 1$ in Eq. (4.3.2).

The ES(β) filter tries to model how each user make forecasts mixing own experience, experience shared with other users, as well as any other source of information such as ITS. It is worth noting that, the resulting forecasted costs are a convex combination of costs occurred today or on each previous day until first day, with weights β, $\beta\,(1-\beta)$, $\beta\,(1-\beta)^2$, …, respectively, the farther the day in the past the smaller the weight. [Still if today is day k the first day $k=0$ gets a weight $(1-\beta)^k$ that is greater than the weight $\beta\,(1-\beta)^{k-1}$ given to the first day $k=1$ if the updating parameter is less than one half, $\beta < 0.5$.] Even though according to an ES filter an infinite memory is assumed, the weight given to any of the past days becomes rather small after few days, for instance with $\beta = 0.5$, it is less that 0.1% after 9 days, and with $\beta = 0.6$, after 7 days; thus the ES(β) filter may be considered an effective approximation of a finite memory (Fig. 4.1).

If the ES(β) filter models user experience the value of the cost updating parameter β should be calibrated against real data, a still open issue; values of the cost updating parameter in the range [0.5, 0.8] seem likely. Extensive experimentations are surely needed to provide sound estimates. On the other hand, if the ES(β) filter models the forecasting filter of an ATIS the cost updating parameter β becomes a design variable.

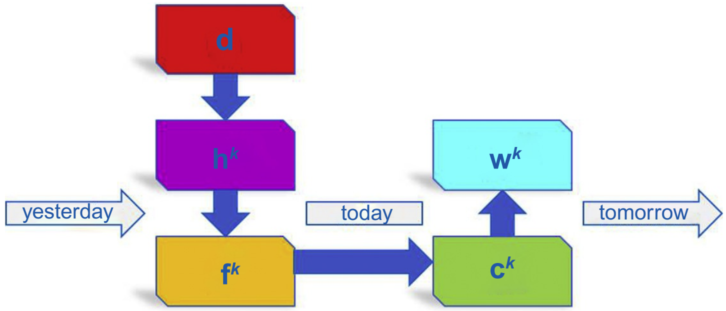


FIG. 4.1

Data-flow diagram of the DP models for day-to-day dynamic assignment.

– Moving average filter

Sometimes a model of user memory and learning process with explicitly finite memory depth μ may be useful, as in Chapter 5 for analysing some stochastic process models. Thus, a moving average, MA(β, μ), filter with one parameter β and memory depth μ is introduced. At this aim, first the weighs of the ES filter are only applied to μ previous days; thus at day t, the summation starts at day $t-\mu+1$; then applying a scaling factor ensures that these weights ζ_k sum to 1, leading to a (strict) convex moving average MA(β, μ) filter with one parameter; (normalised) decreasing weights ζ_j for last μ days are given by:

$$\zeta_j = \beta\,(1-\beta)^{j-1}/(1-(1-\beta)^{\mu}) \geq 0 \quad \forall j = 1, 2, \ldots, \mu$$

where

$\mu > 1$, $\mu \in \mathbb{N}$, is the *memory depth*, say how many previous day actual costs affect today forecasted costs;

$\beta \in]0, 1[$ is the *cost updating parameter*, that used to compute the weights given to previous actual costs to compute today forecasted costs;

ζ_j is the weight given to the actual cost occurred in any of the μ previous days, $\sum_j \zeta_j = 1$.

The weights of the MA filter may also be defined the following recursive equation, useful for computation:

$$\zeta_1 = \beta/(1-(1-\beta)^{\mu}) \quad j = 1$$

$$\zeta_j = \zeta_{j-1}(1-\beta) \quad \forall j = 2, \ldots, \mu$$

Using the above specified weights after the initialization step (see below), the *route cost updating function*, which expresses the relation between today forecasted route costs and previous day route costs, is defined as:

$$\mathbf{y}^k{}_{;i} = \sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{w}^{k-j}{}_{;i} \quad \forall i\ \forall k \in \mathbb{N},\ k > \mu$$

or $\mathbf{y}^k{}_{;i} = \mathbf{CM}^{k-1}{}_{;i} \cdot \zeta$

where

ζ is the $\mu \times 1$ vector with entries ζ_j;

$\mathbf{CM}^{k-1}{}_{;i}$ is the $\mu \times n$ memory matrix of costs with μ columns given by the actual costs in the μ previous days, $\mathbf{w}^{k-j}{}_{;i}$, $\forall j = 1, 2, \ldots, \mu$ for o-d pair i at the beginning of day k; at the end of day k the current cost memory matrix $\mathbf{CM}^{k-1}{}_{;i}$ is updated by dropping last column, moving all others columns rightwards and putting $\mathbf{w}^{\boldsymbol{k}}$ as first column to get $\mathbf{CM}^{k}{}_{;i}$.

Initialization of $\mathbf{CM}^{k}{}_{;i}$, say specification of $\mathbf{CM}_i{}^{\mu}$, may be carried out assuming that:

- all the μ columns of $\mathbf{CM}_i{}^{\mu}$ are equal to $\mathbf{B}_i{}^{\mathbf{T}}\ \mathbf{c}^0$;
- the ES filter (4.3.2) is applied for μ days to fill the μ columns of matrix $\mathbf{CM}_i{}^{\mu}$.

Combining the arc flow function with the MA(β, μ) filter leads to:

$$\mathbf{y}^k{}_{;i} = \sum_{j=1,\ldots,\mu} \zeta_j \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}^{k-j} \quad \forall i\ \forall k \in \mathbb{N},\ k > \mu \tag{4.3.3}$$

For $\mu = 1$ and/or $\beta = 1$ it is assumed $\zeta_1 = 1$, say today forecasted costs are equal to yesterday actual costs, as in a OY filter (4.3.1).

The weights given to actual travel experiences in the past depend only on the relative distance in time they are away from the present, i.e. the model MA(β, μ) is time-homogeneous.

With either of the above initialization approaches, as μ goes to infinite the first and last weights got to β and 0, respectively:

$$\zeta_1 = \beta/(1-(1-\beta)^{\mu}) > \beta \text{ goes to } \beta,$$

$$\zeta_\mu = \beta\,(1-\beta)^{\mu-1}/(1-(1-\beta)^{\mu}) > 0 \text{ goes to } 0.$$

Moreover as μ goes to infinite the MA(β, μ) filter (4.3.3) tends to the ES(β) filter (4.3.2). Fig. 4.1 shows a comparison between weights of the ES and the MS filters with same cost updating parameter β against different values of memory depth μ. Results show that the ES(β) filter is a very good approximation of a MA(β, μ), but for very low values of memory depth, $\mu \leq 3$ which seems rather unrealistic. Hence, from the practical point-of-view the preference between an ES filter or a MS filter is more a matter of mathematical convenience rather than a modelling issue. On the other hand, they have different theoretical features as shown in sub-section.

Equivalent i-route formulations can also be defined for all the above filters; they are omitted for brevity's sake (cfr Section 2.3).

- Route cost updating function

Eqs (4.1), (4.2), (4.3.#) describing the supply model can be combined to define the *route cost updating function*, which expresses the relation between today forecasted route costs and yesterday route flows and possibly other flows (cfr Eq. 3.4), for given $\mathbf{h}^{O}{}_{;i}\,\forall i$, blocks of $\mathbf{h}^{O}$ and $\mathbf{y}^{O}{}_{;i}\,\forall i$, blocks of $\mathbf{y}^{O}$. Let $\mathbf{y}^k \geq \mathbf{0}$ be the $n \times 1$ (column) vector of *total route forecasted costs* on day k.

– With only yesterday filter

$$\mathbf{y}^k{}_{;i} = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}\left(\sum\nolimits_i \mathbf{B}_i \cdot \mathbf{h}^{k-1}{}_{;i}; \boldsymbol{\kappa}\right) \quad \forall i \; \forall k \in \mathbb{N}$$

that are blocks of the block vector recursive function:

$$\mathbf{y}^k = \mathbf{B}^{\mathbf{T}} \cdot \mathbf{c}\left(\mathbf{B} \cdot \mathbf{h}^{k-1}; \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N}$$

or, remembering from Section 3.1.1 the *route cost function* (3.4) $\mathbf{w} = \mathbf{w}(\mathbf{h}; \boldsymbol{\kappa})$:

$$\mathbf{y}^k = \mathbf{w}\left(\mathbf{h}^{k-1}; \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N} \tag{4.4.1}$$

– With exponential smoothing filter

$$\mathbf{y}^k{}_{;i} = \beta\, \mathrm{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}\left(\sum\nolimits_i \mathbf{B}_i \cdot \mathbf{h}^{k-1}{}_{;i}; \boldsymbol{\kappa}\right) + (1-\beta)\, \mathbf{y}^{k-1}{}_{;i} \quad \forall i \; \forall k \in \mathbb{N}$$

that are blocks of the block vector recursive function:

$$\mathbf{y}^k = \beta\, \mathrm{B}^{\mathbf{T}} \cdot \mathbf{c}\left(\mathbf{B}_i \cdot \mathbf{h}^{k-1}; \boldsymbol{\kappa}\right) + (1-\beta)\, \mathbf{y}^{k-1} \quad \forall k \in \mathbb{N}$$

or remembering from Section 3.1.1 the *route cost function* (3.4) $\mathbf{w} = \mathbf{w}(\mathbf{h}; \boldsymbol{\kappa})$:

$$\mathbf{y}^k = \beta\, \mathbf{w}\left(\mathbf{h}^{k-1}; \boldsymbol{\kappa}\right) + (1-\beta)\, \mathbf{y}^{k-1} \quad \forall k \in \mathbb{N} \tag{4.4.2}$$

– With moving average filter

$$\mathbf{y}^k{}_{;i} = \sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathrm{c}\left(\sum\nolimits_i \mathbf{B}_i \cdot \mathbf{h}^{k-j}{}_{;i}; \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N}, \quad k > \mu$$

that are blocks of the block vector recursive function:

$$\mathbf{y}^k = \sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{B}^{\mathbf{T}} \cdot \mathbf{c}\left(\mathbf{B} \cdot \mathbf{h}^{k-j}; \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N}, \quad k > \mu$$

or remembering from Section 3.1.1 the *route cost function* (3.4) $\mathbf{w} = \mathbf{w}(\mathbf{h}; \boldsymbol{\kappa})$:

$$\mathbf{y}^k = \sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{w}\left(\mathbf{h}^{k-j}; \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N}, \; k > \mu \tag{4.4.3}$$

In multi-user assignment users may be grouped into classes (cfr Section 2.4), each with a different value of the cost updating parameter β and memory depth μ (if the case), this way commuters and non-commuters, ATIS equipped and non-equipped users and/or human driven vs. automated vehicles, different commodities might be differentiated, at the expense of increasing the number of parameters; other approaches to modelling user memory and learning may require several parameters as well. These general modelling approaches are not suitable for the discussion of the long-term evolution over time and the analysis of fixed-point stability and bifurcations in Sections 4.3 and 4.4.

In fully disaggregate approaches, each class is made up by a single user, thus the cost updating parameter β may be defined for each single user, and memory may only refer to personal experience. These models are better suited for disaggregate assignment through stochastic process models, as described in Chapter 5.

- Arc cost updating function

The very same route cost updating function (4.4.#) is obtained by first computing forecasted arc costs by applying any of the above cost updating filters to arc costs. Let $\mathbf{x}^k \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *arc forecasted costs* for day k.

– Only yesterday filter

$$\mathbf{x}^k = \mathbf{c}\left(\mathbf{f}^{k-1}; \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N} \tag{4.5.1}$$

for given $\mathbf{f}^0$.

– Exponential smoothing filter

$$\mathbf{x}^k = \beta\, \mathbf{c}\left(\mathbf{f}^{k-1}; \boldsymbol{\kappa}\right) + (1-\beta)\, \mathbf{x}^{k-1} \quad \forall k \in \mathbb{N} \tag{4.5.2}$$

for given $\mathbf{f}^0$, and $\mathbf{x}^0 = \mathbf{c}(\mathbf{f}^0)$
where

$\beta \in]0, 1[$ is the *cost updating parameter*.

– Moving average filter

$$\mathbf{x}^k = \sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{c}\left(\mathbf{f}^{k-j}; \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N},\ k > \mu \tag{4.5.3}$$

where

$\mu > 1,\ \mu \in \mathbb{N},$ is the *memory depth*;
$\beta \in\ ]0, 1[$ is the *cost updating parameter*;

$$\zeta_j = \beta\,(1-\beta)^{j-1} / (1-(1-\beta)^{\mu}) \geq 0 \quad \forall j = 1, 2, \ldots, \mu.$$

In this case the above equations are used instead of (4.2), with forecasted route costs defined by the following relation (cfr Eq. 3.3):

- Route-arc cost consistency relation

$$\mathbf{y}^k{}_{;i} = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{x}^k \quad \forall i\ \forall k \in \mathbb{N}$$

to be used instead of (4.3.#).

4.1.2 Demand models for simple DP models

Travel demand models express how network performances affect user choice behaviour. This section describes the three equations that according to the SEAM framework specify the travel demand model for the day-to-day dynamics of a transportation system, needing an extension of EQN5, as shown below.

- Route utility function

The utility function for o-d pair i is assumed specified through a linear transformation of today route forecasted costs, almost always in research analysis as well as in practical applications (cfr Eq. 3.6):

$$\mathbf{v}^k{}_{;i} = -\psi_i\, \mathbf{y}^k{}_{;i} \quad \forall i\ \forall k \in \mathbb{N} \tag{4.6}$$

where $\psi_i > 0$ is the utility scale parameter, such that the term $\psi_i\, \mathbf{y}^k;_i$ is dimensionless to be consistent with utility unit, it is assumed day-invariant.

[The features of whole model are not affected if first route utility values are computed from actual route costs, then any of the above filters is applied to past actual and/or forecasted utility values. Still this is not a general condition as discussed in Section 4.5.]

- Route choice updating function

When modelling day-to-day dynamics today route choice behaviour is generally affected by user habit effect and inertia to change yesterday choice. Very simple models are described below, more general ones are in Section 4.5.

– No inertia filter

In the most straightforward approach to choice updating both habit and inertia are neglected, thus for each o-d pair *i* the today route choice proportions depend on today route systematic utility values through any discrete choice models theory (see Appendix to the book) (cfr Eq. 3.7):

$$\mathbf{p}^k;_i = \mathbf{p}_i\left(\mathbf{v}^k;_i; \boldsymbol{\theta}_i\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{4.7.1}$$

where $\boldsymbol{\theta}_i$ is the choice function parameter vector, whose meaning depends on the choice model specification. If a utility scale parameter is present, it is considered included in the utility parameter ψ_i (or vice versa). The route choice function as well its parameters are assumed day-invariant.

– Exponential smoothing filter

In a more realistic, but still very simple, approach to choice updating, only some users reconsider yesterday choice (but not necessarily change them), and their route choice behaviour is modelled through any choice function as above; the others simple repeat their yesterday choices actual costs, that is their choice behaviour is modelled by yesterday route proportions. Following this approach for each o-d pair *i* today choice proportions are given by an Exponential Smoothing, ES(α), filter given $\mathbf{p}^0;_i$:

$$\mathbf{p}^k;_i = \alpha\, \mathbf{p}_i\left(\mathbf{v}^k;_i; \boldsymbol{\theta}_i\right) + (1-\alpha)\, \mathbf{p}^{k-1};_i \quad \forall i\ \forall k \in \mathbb{N} \tag{4.7.2}$$

where

$\alpha \in]0, 1[$ is the *choice updating parameter*, that is the proportion of users reconsidering yesterday choice; in the following the choice updating parameter α is assumed day-invariant and common to all users.

Remark. If $\mathbf{p}^{k-1};_i \geq \mathbf{0}$, $\mathbf{1}^{\mathbf{T}}\mathbf{p}^{k-1};_i = 1$, $\mathbf{p}_i(\mathbf{v}^k;_i; \boldsymbol{\theta}_i) \geq \mathbf{0}$, and $\mathbf{1}^{\mathbf{T}}\mathbf{p}_i(\mathbf{v}^k;_i; \boldsymbol{\theta}_i) = 1$, condition $\alpha \in]0, 1[$ assures that $\mathbf{p}^k;_i \geq \mathbf{0}$, $\mathbf{1}^{\mathbf{T}}\mathbf{p}^k;_i = 1$.

Remark. If the choice proportions, $\mathbf{p}_i(\mathbf{v}^k;_i; \boldsymbol{\theta}_i)$, are defined through choice functions derived from RUT, they are given the meaning of choice probability conditional to the event of reconsidering yesterday choice, and the choice updating parameter, α, is given the meaning of the reconsidering probability. [In this case Eq. (4.7.2) is an application of a basic lemma of Theory of Probability: P(B) = P(A) P(B/A) + (1 − P(A)) P(B/A^c).]

In the following the choice updating parameter α is assumed time invariant and common to all users. Comments made above for the cost updating parameter β about calibration, as well as on numerical interpretation, apply to the flow updating parameter α too; in this case values in the range [0.4, 0.6] seem likely.

Eq. (4.7.2) tries to model in simple but effective way user inertia to change and how much users are prone to review their habit.

– Moving average filter

Moving average filters for modelling habit and inertia have not been proposed so far. Their interpretation does not seem straightforward, thus they will not be discussed below.

• Route-demand flow consistency relation

Flow conservation for o-d pair i can be expressed as:

$$\mathbf{h}^k{}_{;i} = d_i \mathbf{p}^k{}_{;i} \quad \forall i\ \forall k \in \mathbb{N} \tag{4.8}$$

It assures that flows of all routes connecting the o-d pair i sum up to demand flow, that is $\mathbf{1}^{\mathbf{T}}\mathbf{h}^k{}_{;i} = d_i$, since $\mathbf{1}^{\mathbf{T}}\mathbf{p}^k{}_{;i} = 1$, and non-negative, $\mathbf{h}^k{}_{;i} \geq \mathbf{0}$, since $d_i \geq 0$ and $\mathbf{p}^k{}_{;i} \geq \mathbf{0}$.

• Route choice updating function

Eqs (4.6), (4.7.#), (4.8) describing the demand model can be combined to define the *route flow updating function*, which expresses the relation between today route flows and yesterday route flows (cfr Eq. 3.9), for given $\mathbf{h}^0{}_{;i} \forall i$, blocks of $\mathbf{h}^0$:

– With no inertia filter

$$\mathbf{h}^k{}_{;i} = d_i \mathbf{p}_i\left(-\psi_i \mathbf{y}^k{}_{;i};\, \boldsymbol{\theta}_i\right) \quad \forall i\ \forall k \in \mathbb{N}$$

or, remembering from Section 3.1.2 the *route flow function* (3.9) $\mathbf{h} = \mathbf{h}(\mathbf{w}; \mathbf{d})$:

$$\mathbf{h}^k = \mathbf{h}\left(\mathbf{y}^k; \mathbf{d}\right) \quad \forall k \in \mathbb{N} \tag{4.9.1}$$

– With exponential smoothing filter

$$\mathbf{h}^k{}_{;i} = \alpha\, d_i \mathbf{p}_i\left(-\psi_i\, \mathbf{y}^k{}_{;i};\, \boldsymbol{\theta}_i\right) + (1-\alpha)\, \mathbf{h}^{k-1}{}_{;i} \quad \forall i\ \forall k \in \mathbb{N}$$

or, remembering from Section 3.1.2 the *route flow function* (3.9) $\mathbf{h} = \mathbf{h}(\mathbf{w}; \mathbf{d})$:

$$\mathbf{h}^k = \alpha\, \mathbf{h}\left(\mathbf{y}^k; \mathbf{d}\right) + (1-\alpha)\, \mathbf{h}^{k-1} \quad \forall k \in \mathbb{N} \tag{4.9.2}$$

An equivalent i-route formulation can also be defined; it is omitted for brevity's sake (cfr Section 2.3).

In multi-user assignment users may be grouped into classes (cfr Section 2.4), each with a different value of the choice updating parameter α, thus increasing the number of parameters; other approaches to modelling user memory and learning may require several parameters as well. These general modelling approaches are not suitable for the discussion of the long-term evolution over time and the analysis of fixed-point stability and bifurcations in Sections 4.3 and 4.4.

In fully disaggregate approaches, each class is made up by a single user, thus the choice updating parameter α may be defined for each single user. These models are better suited for disaggregate assignment through stochastic process models, as described in Chapter 5.

4.1.3 Arc flow updating function

In this section the *arc flow updating function* is introduced and discussed. It can be obtained by combining together equations (4.1, 4.3.#, 4.6, 4.7.#, 4.8), or equations (4.1, 4.3.#, 4.9.#), and remembering the arc flow function (2.33) introduced in Section 2.5. Thus, given $\mathbf{f}^0$:

– With no inertia filter (cfr Eq. 2.34 in Section 2.5)

$$\mathbf{f}^k = \mathbf{f}(\mathbf{x}^k; \mathbf{d}) \in S_f \subseteq \mathbb{E}^m \qquad \forall k \in \mathbb{N} \tag{4.10.1}$$

– With exponential smoothing filter

$$\mathbf{f}^k = \alpha\, \mathbf{f}(\mathbf{x}^k; \mathbf{d}) + (1-\alpha)\, \mathbf{f}^{k-1} \in S_f \subseteq \mathbb{E}^m \qquad \forall k \in \mathbb{N} \tag{4.10.2}$$

Eq. (4.10.2) too can be considered an extension of the arc flow function (2.34) introduced in Section 2.5. Indeed, $\mathbf{f}(\mathbf{x}^k; \alpha\, \mathbf{d})$ are the today arc flows due to the ($\alpha\, \mathbf{d}$) users who have reconsider their yesterday choice, and $(1-\alpha)\, \mathbf{f}^{k-1}$ are the today arc flows due to the ($(1-\alpha)\, \mathbf{d}$) users who have not, thus summing up the today arc flows are:

$$\mathbf{f}^k = \mathbf{f}\left(\mathbf{x}^k; \alpha\, \mathbf{d}\right) + (1-\alpha)\, \mathbf{f}^{k-1} \quad \forall k \in \mathbb{N}$$

As noted in Section 2.5 the arc flow function is homogenous of degree 1 with respect to demand flows: $\mathbf{f}(\mathbf{c}; \alpha\, \mathbf{d}) = \alpha\, \mathbf{f}(\mathbf{c}; \mathbf{d})$, $\forall \alpha > 0$, (cfr 2.35), then Eq. (4.10.2).

4.2 Simple DP models

The set of six equations (4.1)–(4.3.#) and (4.6)–(4.8) defines a discrete time Markovian deterministic process (DP) model with respect to all the six basic variables, describing the evolution over time of them.

Discrete time Markovian deterministic processes

Let $\boldsymbol{\varphi}(\mathbf{x};\boldsymbol{\theta})$ be a vector function from set S to set $\boldsymbol{\varphi}(S)$, the recursive equation $\mathbf{x}^k = \boldsymbol{\varphi}(\mathbf{x}^{k-1}, \ldots; \boldsymbol{\theta}) \in S$ with $\mathbf{x}^0 \in S$ defines a discrete time deterministic process (DP), useful to describe the evolution over time of a system where the state at time (day) k is described by $\mathbf{x}^k$, S is the state space, $\boldsymbol{\varphi}(\cdot)$ is the transition function, and $\boldsymbol{\theta}$ are its parameters. The sequence $\mathbf{x}^0, \mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \ldots, \mathbf{x}^{k-1}, \mathbf{x}^k, \mathbf{x}^{k+1}, \ldots \in S$, called a *trajectory*, depends on the initial state $\mathbf{x}^0$ and the values of the parameters $\boldsymbol{\theta}$.

The process is called *Markovian* if today state $\mathbf{x}^k$ depends on yesterday state $\mathbf{x}^{k-1}$ only.

Remark. This condition can be obtained also if today state depends on finite number of previous day states by duly specifying an equivalent process with further state variables.

Remark. If today state also depends on itself, the DP can be put in a Markovian form through the approaches described in Appendix C.

Fig. 4.1 shows a data-flow diagram of the DP model (4.1–4.3) and (4.6–4.8), highlighting the roles of the main variables.

Deterministic processes as discrete-time non-linear dynamic systems

Discrete time deterministic processes are native discrete-time non-linear dynamic systems, for which time is integer and increased by 1 at each iteration. They should not be confused with discrete-time non-linear dynamic systems used to approximate continuous-time deterministic processes, the so-called Poincaré maps obtained through stroboscopic technique, for which time may be real and increased by a real value.

To further analyse the model it is better to reduce the number of equations and variables, as shown in the following (cfr Section 3.2). Given a specification of the cost updating filter and of the choice updating filter the resulting model can be specified with respect to route or i-route variables (not explicitly reported for brevity's sake) as well as to arc variables, leading to equivalent models.

4.2.1 Two equation assignment models

Given an ES cost updating filter and an ES choice updating filter, the resulting DP models are made by two equations with respect to two vectors, a flow vector and a cost vector, say a two equation assignment models (TEAMs). They can be specified with respect to route (i-route) or arc variables as show below.

- Route costs and flows—ES/ES

DP models based on route costs and flows may be specified by the ES route cost updating Eq. (4.4.2) describing the supply model and the ES route flow updating Eq. (4.9.2) the demand model, from a given initial state $(\mathbf{y}^0, \mathbf{h}^0)$:

$$\mathbf{y}^k = \beta\, \mathbf{w}\left(\mathbf{h}^{k-1}; \boldsymbol{\kappa}\right) + (1-\beta)\, \mathbf{y}^{k-1} \quad \forall k \in \mathbb{N} \tag{4.11}$$

$$\mathbf{h}^k = \alpha\, \mathbf{h}\left(\mathbf{y}^k; \mathbf{d}\right) + (1-\alpha)\, \mathbf{h}^{k-1} \quad \forall k \in \mathbb{N} \tag{4.12}$$

The DP model (4.11, 4.12) can easily be rewritten as a proper Markovian DP, today state only depends on yesterday one by putting Eq. (4.11) into (4.12), but still keeping Eq. (4.11). The state variables of DP model (4.11, 4.12) are ($\mathbf{y}^k$, $\mathbf{h}^k$); the updating parameters are α and β; other parameters are demand flow, **d**, and any other parameter in choice functions and in the arc cost function. This model is useful when explicit path enumeration can be carried out; it is worth noting that this is hardly the case if routes are hyperpaths. This model is also useful as a base for developing stochastic process models described in Chapter 5.

A DP model providing an equivalent evolution over time can be specified with respect to arc variables as shown below.

- Arc costs and flows—ES/ES

DP models based on arc costs and flows may be specified from given initial state by the ES arc cost updating Eq. (4.5.2) and the ES arc flow updating Eq. (4.10.2), from a given initial state ($\mathbf{x}^0 = \mathrm{c}(\mathbf{f}^0), \mathbf{f}^0 \in S_f$):

$$\mathbf{x}^k = \beta\, \mathbf{c}(\mathbf{f}^{k-1}; \boldsymbol{\kappa}) + (1-\beta)\, \mathbf{x}^{k-1} \in S_c \subseteq \mathbb{E}^m \qquad \forall k \in \mathbb{N} \tag{4.13}$$

$$\mathbf{f}^k = \alpha\, \mathbf{f}(\mathbf{x}^k; \mathbf{d}) + (1-\alpha)\, \mathbf{f}^{k-1} \in S_f \subseteq \mathbb{E}^m \qquad \forall k \in \mathbb{N} \tag{4.14}$$

where

S_f is the feasible arc flow set non empty;
S_c is the convex hull of set $\mathbf{c}(S_f)$, $\mathbf{c}(S_f)$ being the co-domain of the arc cost function $\mathbf{c}(\cdot)$.

The DP model (4.13, 4.14) can easily be rewritten as a proper Markovian DP by putting Eq. (4.13) into (4.14), but still keeping Eq. (4.13). The state variables of DP model (4.13, 4.14) are ($\mathbf{x}^k$, $\mathbf{f}^k$); the state space is $S_c \times S_f$; the updating parameters are α and β; other parameters are demand flow, **d**, and any other parameter in choice functions and in the arc cost function. This model is useful when routes (being them paths or hyperpaths) cannot be explicitly enumerated. This model also allows a complete stability analysis as shown in Sections 4.3 and 4.4.

Given an MA cost updating filter and an ES choice updating filter, the resulting DP models can be specified with respect to route (i-route) or arc variables as show below.

- Route costs and flows—MA/ES

DP models based on route costs and flows may also be specified from given initial state by the MA route cost updating Eq. (4.4.3) and the ES route flow updating Eq. (4.9.2) the demand model, from a given initial state defined route flows **h** in the first μ days (see the above Section 4.1.1 for details):

$$\mathbf{y}^k = \sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{w}\left(\mathbf{h}^{k-j}; \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N}, k > \mu \tag{4.15a}$$

$$\mathbf{h}^k = \alpha\, \mathbf{h}\left(\mathbf{y}^k; \mathbf{d}\right) + (1-\alpha)\, \mathbf{h}^{k-1} \quad \forall k \in \mathbb{N} \tag{4.15b}$$

with

$$\zeta_j = \beta\,(1-\beta)^{j-1}/(1-(1-\beta)^{\mu}) \geq 0 \quad \forall j = 1,2,\ldots,\mu \tag{4.16}$$

The DP model (4.15a, 4.15b) can easily be rewritten so that today state only depends on previous day ones by putting Eq. (4.15a) into (4.15b) actually leading to a OEAM:

$$\mathbf{h}^k = \alpha\,\mathbf{h}\left(\sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{w}\left(\mathbf{h}^{k-j};\boldsymbol{\kappa}\right);\mathbf{d}\right) + (1-\alpha)\,\mathbf{h}^{k-1} \tag{4.17}$$

Still the resulting DP model (4.17) is not properly Markovian, an equivalent Markovian DP can be specified as described below through $\mu+1$ equations.

The system state at day $k-1$ is described by a vector ${}^{|}\mathbf{h}^{k-1}$ made up by μ blocks, one block for yesterday flows ${}^{1}\mathbf{h}^{k-1}$, and one block ${}^{j}\mathbf{h}^{k-1}$ $j=2, \ldots, \mu$ for each of the $\mu-1$ previous days kept in memory. Therefore, on each day k today's flows, contained in the first block ${}^{1}\mathbf{h}^{k}$, are updated according to Eq. (4.15b), and each of all the other blocks are used to keep a memory of the $\mu-1$ previous days flows, whilst the μ-th previous day flows are no longer recorded. According to this state definition today state only depends on yesterday, leading to the following Markovian DP:

$${}^{1}\mathbf{h}^k = \alpha\,\mathbf{h}\left(\sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{w}\left({}^{j}\mathbf{h}^{k-1};\boldsymbol{\kappa}\right);\mathbf{d}\right) + (1-\alpha)\,{}^{1}\mathbf{h}^{k-1} \quad \forall k \in \mathbb{N}, k > \mu \tag{4.18}$$

$${}^{j}\mathbf{h}^k = {}^{j-1}\mathbf{h}^{k-1} \quad j = 2,\ldots,\mu \tag{4.19}$$

The above model is actually a multi equation assignment model, but it is still considered a TEAM since it is derived from a TEAM, and contains two types of equations only.

The state variables of DP model (4.18, 4.19) are the route flow blocks ${}^{j}\mathbf{h}^{k} j=1, \ldots,$ μ; the updating parameters are α and β and μ in Eq. (4.16) that defines memory weights ζ_j; other parameters are demand flow, $\mathbf{d}$, and any other parameter in choice functions and in the arc cost function.

This model is useful when explicit path enumeration can be carried out; it is worth noting that this is hardly the case if routes are hyperpaths. This model is also useful as a base for developing stochastic process models described in Chapter 5.

A DP model providing an equivalent evolution over time can be specified with respect to arc variables by the MA arc cost updating Eq. (4.5.3) and the ES arc flow updating Eq. (4.10.2) as shown below.

- Route costs and flows—MA/ES

$${}^{1}\mathbf{f}^k = \alpha\,\mathbf{f}\left(\sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{c}\left({}^{j}\mathbf{f}^{k-1};\boldsymbol{\kappa}\right);\mathbf{d}\right) + (1-\alpha)\,{}^{1}\mathbf{f}^{k-1} \quad \forall k \in \mathbb{N}, k > \mu \tag{4.20}$$

$${}^{j}\mathbf{f}^k = {}^{j-1}\mathbf{f}^{k-1} \quad j = 2,\ldots,\mu \tag{4.21}$$

The state variables of DP model (4.20, 4.21) are the arc flow blocks ${}^{j}\mathbf{f}^{k} j=1, \ldots, \mu+1$; the state space is the union of μ copies of set S_f; the updating parameters are α and β and μ in Eq. (4.16) that defines memory weights; other parameters are demand flow, $\mathbf{d}$, and any other parameter in choice functions and in the arc cost function. This model is useful when explicit route enumeration cannot be carried out.

4.2.2 One equation assignment models

Given an OY cost updating filter and/or an NI choice updating filter, the resulting DP models are made by one equation with respect to one vector, a flow vector or a cost vector, say one two equation assignment models (OEAMs). They can be specified with respect to arc or route variables, as well as i-route variables. Arc formulations only are reported below for brevity's sake.

- Arc costs and flows—OY/ES

$$\mathbf{f}^k = \alpha\, \mathbf{f}(\mathbf{c}(\mathbf{f}^{k-1}; \boldsymbol{\kappa}); \mathbf{d}) + (1-\alpha)\, \mathbf{f}^{k-1} \in S_f \subseteq \mathbb{E}^m \quad \forall k \in \mathbb{N} \tag{4.22}$$

This model can be obtained from (4.13, 4.14) with $\beta=1$ or from (4.20, 4.21) with $\beta=1$ and $\mu=1$.

- Arc costs and flows—ES/NI

$$\mathbf{x}^k = \beta\, \mathbf{c}\left(\mathbf{f}\left(\mathbf{x}^{k-1}; \mathbf{d}\right); \boldsymbol{\kappa}\right) + (1-\beta)\, \mathbf{x}^{k-1} \quad \forall k \in \mathbb{N} \tag{4.23}$$

This model can be obtained from (4.13, 4.14) with $\alpha=1$.

- Arc costs and flows—MA/NI

$$\mathbf{x}^k = \sum\nolimits_{j=1,\ldots,\mu} \zeta_j \mathbf{c}\left(\mathbf{f}\left(\mathbf{x}^{k-1}; \mathbf{d}\right); \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N}, k > \mu \tag{4.24}$$

This model can be obtained from (4.20, 4.21) with $\alpha=1$.

- Arc costs and flows—OY/NI

$$\mathbf{f}^k = \mathbf{f}(\mathbf{c}(\mathbf{f}^{k-1}; \boldsymbol{\kappa}); \mathbf{d}) \in S_f \subseteq \mathbb{E}^m \qquad \forall k \in \mathbb{N} \tag{4.25}$$

$$\text{or} \quad \mathbf{x}^k = \mathbf{c}\left(\mathbf{f}\left(\mathbf{x}^{k-1}; \mathbf{d}\right); \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N} \tag{4.26}$$

Either of these models can be obtained from (4.13, 4.14) with $\alpha=1$ and $\beta=1$ or from (4.20, 4.21) with $\alpha=1$, $\beta=1$ and $\mu=1$.

Even though all the above OEAMs can be obtained as limit cases of previous described TEAMs they may not share all their features as discussed in Sections 4.3 and 4.4.

4.2.3 Fixed point states

All the above discussed DP models have the same fixed-point states, which are consistent with CUE, as described by the fixed-point models presented in the previous Chapter 3. An explicit proof is given below for DP (4.13, 4.14) and (4.20, 4.21) only for brevity's sake. This result implies that fixed-point states of all the above described DP models do not depend on updating parameters. Thus, existence and uniqueness conditions discussed in Chapter 3 still hold. But, this condition may not hold for general DP models described in Sections 4.5 and 4.6. Anyway, fixed-point states depend on other parameters, such as demand flows, dispersion parameters, arc capacity, see Section 4.4.1 for further comments.

Fixed-point states of a deterministic process

Let $\mathbf{x}^k = \boldsymbol{\varphi}(\mathbf{x}^{k-1}; \boldsymbol{\theta}) \in S$ with $\mathbf{x}^0 \in S$ be a discrete-time Markovian deterministic process, any state $\mathbf{x}^* \in S$ such that

$$\mathbf{x}^* = \mathbf{x}^k = \mathbf{x}^{k-1} \text{ or } \mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta})$$

is a fixed-point state of it, generally depending on the parameters θ. It is a parametric fixed-point of the transition function $\boldsymbol{\varphi}(\cdot)$ as well (cfr the implicit function theorem).

Given $\mathbf{x}^* = \mathbf{x}^k = \mathbf{x}^{k-1}$ and $\mathbf{f}^* = \mathbf{f}^k = \mathbf{f}^{k-1}$, Eqs (4.13), (4.14) lead to

$$\mathbf{x}^* = \beta\, \mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa}) + (1-\beta)\, \mathbf{x}^*$$

$$\mathbf{f}^* = \alpha\, \mathbf{f}(\mathbf{x}^*; \mathbf{d}) + (1-\alpha)\, \mathbf{f}^*$$

or

$$\mathbf{0} = \beta\, \mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa}) - \beta\, \mathbf{x}^*$$

$$\mathbf{0} = \alpha\, \mathbf{f}(\mathbf{x}^*; \mathbf{d}) - \alpha\, \mathbf{f}^*$$

Since both updating parameters α and β are different from zero, it yields

$$\mathbf{x}^* = \mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa})$$

$$\mathbf{f}^* = \mathbf{f}(\mathbf{x}^*; \mathbf{d})$$

to be compared with fixed-point model (3.15, 3.16) for CUE.
Given $\mathbf{f}^* = {}^j\mathbf{f}^k = {}^j\mathbf{f}^{k-1}\ \forall j = 1, \ldots, \mu+1$, Eqs (4.20), (4.21) lead to

$$\mathbf{f}^* = \alpha\, \mathbf{f}\left(\Sigma_{j=1,\ldots,\mu}\, \zeta_j\, \mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa}); \mathbf{d}\right) + (1-\alpha)\, \mathbf{f}^*$$

$$\mathbf{f}^* = \mathbf{f}^*$$

or

$$\mathbf{f}^* = \alpha\, \mathbf{f}\left(\mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa})\, \Sigma_{j=1,\ldots,\mu}\, \zeta_j; \mathbf{d}\right) + (1-\alpha)\, \mathbf{f}^*$$

Since $\Sigma_{j=1,\ \ldots,\ \mu}\, \zeta_j = 1$, it yields:

$$\mathbf{f}^* = \alpha\, \mathbf{f}(\mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa}); \mathbf{d}) + (1-\alpha)\, \mathbf{f}^*$$

or

$$\mathbf{0} = \alpha\, \mathbf{f}(\mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa}); \mathbf{d}) - \alpha\, \mathbf{f}^*.$$

Moreover, since the choice updating parameter α is different from zero, it yields:

$$\mathbf{f}^* = \mathbf{f}(\mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa}); \mathbf{d})$$

to be compared with fixed-point model (3.17) for CUE.
Similar proofs exists for OEAMs (4.22), (4.23), (4.24), (4.25), (4.26).

A special case occurs when there is a unique fixed-point state and from any initial state belonging to the state space the system converges towards it. In this case the (unique) fixed point is called **globally stable**.

Sufficient conditions for existence, uniqueness and global stability are given by the Banach theorem and its corollaries, requiring that the state space is non empty and compact and the transition function is strictly non-expansive.

Sufficient conditions for fixed-point global stability—Corollary to Banach's theorem

Let $\mathbf{x}^k = \boldsymbol{\varphi}(\mathbf{x}^{k-1}; \boldsymbol{\theta}) \in S$ with $\mathbf{x}^0 \in S$ be a deterministic process, with compact state space S, and $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta}) \in S$ be a fixed-point state of it.

IF the transition function $\boldsymbol{\varphi}(\mathbf{x})$ is strictly non-expansive:

$$\|\boldsymbol{\varphi}(\mathbf{x}') - \boldsymbol{\varphi}(\mathbf{x}'')\| < \|\mathbf{x}' - \mathbf{x}''\| \forall \mathbf{x}_1 \neq \mathbf{x}_2 \in S$$

where $\|\cdot\|$ is any vector norm/distance.

THEN the fixed-point state $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta}) \in S$ is unique and globally stable, that is any trajectory $\mathbf{x}^k = \boldsymbol{\varphi}(\mathbf{x}^{k-1})$ converges to the fixed-point state $\mathbf{x}^*$ from any initial state $\mathbf{x}^0 \in S$ in the state space:

$$\lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}^* \quad \forall \mathbf{x}^0 \in S$$

Remark. If the transition function $\boldsymbol{\varphi}(\mathbf{x})$ is continuously differentiable with Jacobian $\nabla\boldsymbol{\varphi}(\mathbf{x})$, a sufficient conditions for being strictly non expansive is that the a matrix norm (induced by a vector norm) of its Jacobian is less than one, $|\|\nabla\boldsymbol{\varphi}(\mathbf{x})|\| < 1 \forall \mathbf{x} \in S$.

Still, strictly non-expansiveness of the transition function is a strong condition that do not hold for any of the DP models described in the previous section. A weaker condition for fixed-point global stability requiring that the state space is non empty and compact and the transition function is continuous can be stated by finding a Lyapunov function.

Sufficient conditions for fixed-point global stability—Lyapunov functions

Let $\mathbf{x}^k = \boldsymbol{\varphi}(\mathbf{x}^{k-1}; \boldsymbol{\theta}) \in S$ with $\mathbf{x}^0 \in S$ be a deterministic process, with compact state space S, and $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta}) \in S$ be a fixed-point state of it.

IF the transition function $\boldsymbol{\varphi}(\mathbf{x})$ is continuous and there exist a continuous non negative scalar function $\ell() \geq 0$, called a Lyapunov function, defined over the state space S such that:

$$\ell(\mathbf{x}) > 0 \quad \ell(\boldsymbol{\varphi}(\mathbf{x})) < \ell(\mathbf{x}) \quad \forall \mathbf{x} \in S, \mathbf{x} \neq \boldsymbol{\varphi}(\mathbf{x}; \boldsymbol{\theta})$$

$$\ell(\mathbf{x}^*) = 0 \quad \ell(\boldsymbol{\varphi}(\mathbf{x}^*)) = \ell(\mathbf{x}^*) \quad \forall \mathbf{x}^* \in S,\ \mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta})$$

THEN the fixed-point state $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta}) \in S$ is unique and globally stable, that is any trajectory $\mathbf{x}^k = \boldsymbol{\varphi}(\mathbf{x}^{k-1})$ converges to the fixed-point state $\mathbf{x}^*$ from any initial state $\mathbf{x}^0 \in S$ in the state space:

$$\lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}^* \quad \forall \mathbf{x}^0 \in S$$

Remark. If the transition function is strictly non-expansive a Lyapunov function exists given by $\ell(\mathbf{x}) = \|\boldsymbol{\varphi}(\mathbf{x}) - \mathbf{x}\|$, thus the existence of a Lyapunov function is a weaker condition than the non-expansiveness of the transition function.

Remark. Sometimes the Lyapunov function is defined non positive.

Remark. Sometimes a Lyapunov function can only be defined locally, say over a subset of the state space.

To author' knowledge, no condition based on existence of a Lyapunov function has been proposed assuring global stability of a unique fixed-point of a DP of the kind discussed in this book, such as DP-ES/ES and DP-MA/ES. Thus, this is still an open issue.

When conditions for globally stability cannot be stated, even if a fixed-point state exists and is unique the system may not evolve towards it, as discussed below. Local stability analysis can be carried out based on the features of the Jacobian matrix of the recursive equations specifying the DP, as shown Sections 4.3 and 4.4.

4.2.4 Solution issues and convergence analysis

Any of the above described discrete-time Markovian deterministic process models can be solved by repeatedly applying the recursive equations that specify it, given an initial state and values of all parameters, and is suitable for large scale applications.

> **Solution of deterministic processes**
>
> Given an initial state $\mathbf{x}^0 \in S$, the recursive equation $\mathbf{x}^k = \boldsymbol{\varphi}(\mathbf{x}^{k-1}; \boldsymbol{\theta}) \in S$ is repeatedly applied for a pre-fixed number of iterations or until a sort of convergence is reached (see below). The resulting trajectory generally depends on the initial state $\mathbf{x}^0$.
>
> Remark. A fixed-point state is not suitable initial state since in this case the state will always reproduces itself.

With reference to the examples already discussed in Chapters 2 and 3, Fig. 4.2 shows the trajectories of flow on route 1 from day 105 to day 120 obtained by applying DP-ES/ES with $\alpha = 0.50$, $\beta = 0.60$, dispersion parameter $\theta = 7$, and increasing demand flow d from 3600 to 3900, 4200. The trajectories with small values of the demand flow d reach the unique fixed-point state (consistent with SUE, cfr Table 3.2 and Fig. 3.3 for $d = 3600$), but for larger values the trajectories keep oscillating between two states. Similar results can be observed decreasing the Logit dispersion parameter θ and/or the arc capacities.

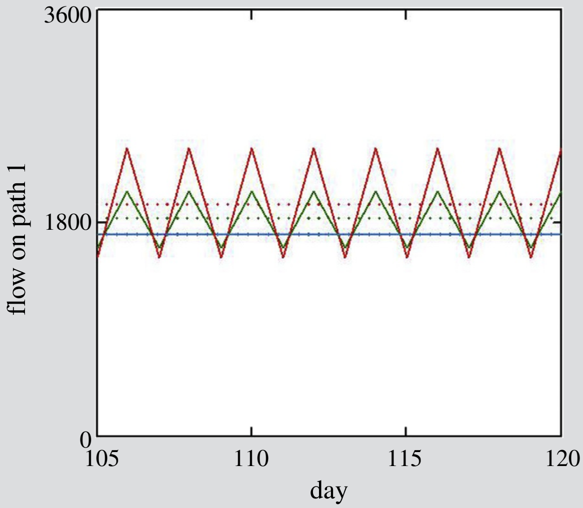


FIG. 4.2

A route flow against day given by DP-ES/ES (dotted line shows SUE) with $d = 3600$ (*flat line*), 3900 (*spiky line*), 4200 (*largely spiky line*).

Fig. 4.3 shows the trajectories obtained increasing β from 0.60 to 0.65, 0.70. The trajectories of DP-ES/ES with small values of the cost updating parameter β reach the unique fixed-point state (consistent with SUE, cfr Table 3.2 and Fig. 3.3), but for larger values the trajectories keep oscillating between two states. As shown above, the fixed-point state is not affected by the values of β. Similar results can be observed increasing the choice updating parameter α.

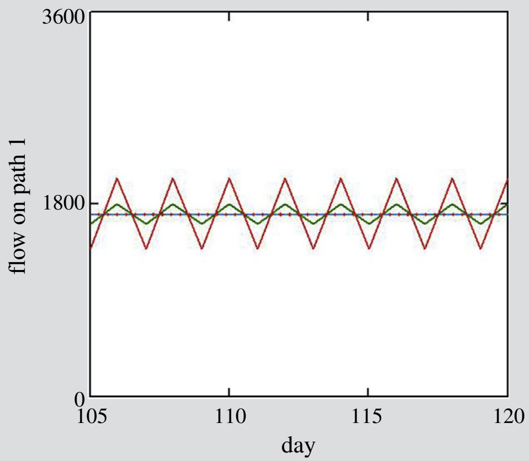


FIG. 4.3

A route flow against day given by DP-ES/ES (dotted line shows SUE) with $\beta = 0.60$ (*flat line*), 0.65 (*spiky line*), 0.70 (*largely spiky line*).

Fig. 4.4 shows a comparison between the trajectories obtained by applying DP-ES/ES with $\beta = 0.60$ and those of DP-MA/ES with same value of β and memory depth $\mu = 3, 4$. The trajectories from DP-ES/ES and DP-MA/ES with $\mu = 4$ reach the unique SUE (and cannot be distinguished), but with $\mu = 4$ the trajectory from DP-MA/ES keep oscillating between two states.

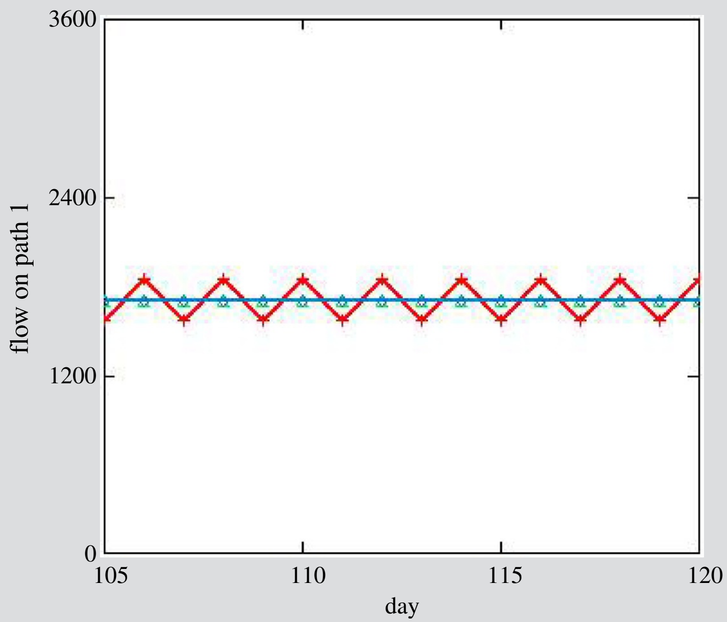


FIG. 4.4

A route flow against day given by DP with MA $\mu = 3$ (*largely spiky line*), 4 (*flat line*), and with ES (*hidden flat line*).

Results of simple examples show that the trajectory of a DP may not converge to a fixed-point state (consistent with SUE) even if it is unique, as stated below where some notions about the convergence of a deterministic process are briefly reviewed to introduce notations and definitions (and to support the unfamiliar reader).

Convergence sets of a deterministic process

A *convergence set*, is a proper subset of the state space, having the following properties:

– it has a dimension strictly less than the dimension of the state space,
– the system cannot evolve towards a state out of the convergence starting from its interior;
– the convergence set is minimal, that is it does not strictly include any other convergence set.

An example of convergent set is a fixed-point state.

Attractors of a deterministic process

An *attractor* is a convergence set that

– has an *attraction domain* (also called basin of attraction), which is a proper super-set of the convergence set such that from any initial state belonging to the attraction domain the system converges towards the attractor; the attraction domain may be a proper sub-set of the state space or the whole state space.

Convergent deterministic processes

A DP is called *convergent* from an initial state if its state tend to a convergence set. The state space may be considered partitioned into *convergence regions*, say attraction domains, and *non-convergence regions*, from which the DP does not tend to any attractor, say it is not convergent. Usually borders among regions can only be identified through brute force.

Types of attractors of a deterministic process

There are four main **types of attractors**:

– *fixed-point attractors*: once reached the system always takes up the same point, their dimension is zero;
– *periodic attractors*: once reached the system periodically moves among a finite number of points, their dimension is zero;
– *quasi-periodic attractors*: once reached the system moves on a torus, that is a toroidal surface containing infinite many points, with non zero integer dimension;
– *a-periodic attractor*s: once reached the system moves within a fractal set containing infinite many points with non zero non-integer dimension.

If there is no attraction domain any of the sets described above is just a convergence set. For example a fixed-point state is a *repulsor*, if from any other initial state the system diverges from the fixed-point state, or a *saddle*, if from some initial states the system converges to the fixed-point state and from others diverges from it.

Chaotic attractors of a deterministic process

An attractor is called *chaotic* if two trajectories starting however close will greatly diverge after some time (but they still are in the attractor in any case). This is often called the *butterfly effect* (usually with reference to continuous-time NDSs). Fixed-point, periodic and quasi-periodic attractors are non-chaotic. Almost always an a-periodic attractor is *chaotic*; in this case the system state evolves over time through successive contractions on some directions and stretching and folding on others.

According to the shadow theorem although a numerically computed chaotic trajectory diverges exponentially from the true trajectory with the same initial coordinates, there exists an errorless trajectory with a slightly different initial condition that stays near ('shadows') the numerically computed one. Therefore, the fractal structure of chaotic trajectories seen in computer maps is real. (Cfr Ott, E. (1993). Chaos in Dynamical Systems. New York: Cambridge Univ. Press, pp. 18–19.)

As shown above a qualitative analysis of convergence and types of attractors (if any) can be carried-out by repeatedly applying the recursive equations that specify a DP. Analytical tools to address these issues are described in the next Sections 4.3 and 4.4.

4.3 Dissipativeness analysis

This section presents analytical conditions for differentiable DP models being convergent to some attractors, possibly depending on the initial state and the model parameters, through a dissipativeness analysis.

Differentiable deterministic processes

A deterministic process $\mathbf{x}^k = \boldsymbol{\varphi}(\mathbf{x}^{k-1}; \boldsymbol{\theta}) \in S$ is called *differentiable* (DDP) if its transition function $\boldsymbol{\varphi}(\cdot)$ s differentiable over the state space S (or subset of it), with $\mathbf{J}(\mathbf{x}) = \nabla\boldsymbol{\varphi}(\mathbf{x})$ being the Jacobian matrix of the transition function.

Dissipative deterministic processes

A differentiable deterministic process is called *dissipative* if a (small enough) ball round any initial state will shrink as the DP evolves, and the state space tends to reduce to a null measure set, that is a set with a dimension smaller than the state space, say an *attractor*, or a convergence set. If the ball shrinks after stretching and folding a chaotic fractal attractor will likely be observed. Let $\delta(\mathbf{J}(\mathbf{x})) = |\det(\mathbf{J}(\mathbf{x}))|$ be the absolute value of the determinant (*absdet*) of matrix $\mathbf{J}(\mathbf{x})$. A sufficient condition for dissipativeness is:

$$\delta(\mathbf{J}(\mathbf{x})) < 1 \forall \mathbf{x} \in S \text{ with } \delta \neq 0.$$

If $\delta = 0$, the analysis should be moved to a space with reduced dimensions where $\delta \neq 0$, through proper linear transformations; otherwise dissipativeness cannot be assessed.

A DP may be dissipative, say convergent, from any initial state in a sub-set (locally dissipative) of the state space only or in the whole state space (globally dissipative).

From the Jacobian matrix of a DDP model the Lyapunov multipliers can be defined to formally identify the type of an attractor (instead of the qualitative analysis of trajectories).

Lyapunov multipliers of a differentiable deterministic processes

Given a differentiable DP specified by transition function $\varphi(\cdot)$, let $\boldsymbol{\varphi}^k(\cdot)$ be the transition function applied k times with Jacobian matrix $\nabla\boldsymbol{\varphi}^k(\cdot)$ that can be computed through the chain rule. Given an initial state $\mathbf{x}_0$ and values of all parameters let be $\lambda^k{}_{;j}$ be the j-th eigenvalue of matrix $\nabla\boldsymbol{\varphi}^k(\mathbf{x}_0)$, the contracting or expanding factor for each direction, called $\mu_j \geq 0$ (not to be confused with the memory depth of a MA filter), is given by:

$$\mu_j = \lim_{k\to\infty}\left(|\lambda^k{}_{;j}|\right)^{1/k}$$

If the product of the n multipliers is less than one, $\mu_1\,\mu_2\ldots\mu_n < 1$, the DP is converging to an attractor; $\mu_j < 1$ defines a shrink direction, while $\mu_j > 1$ a stretching direction followed by a folding, as noted above. Assuming the multipliers sorted in a ascent order, $0 \leq \mu_n \leq \ldots \leq \mu_j \leq \ldots \leq \mu_1$, the type of attractor can be identified by the following conditions.

fixed-point or periodic attractor	$0 \leq \mu_n \leq \ldots \leq \mu_1 < 1$
quasi-periodic attractor	$0 \leq \mu_n \leq \ldots \leq \mu_{j+1} < \mu_j = \ldots = \mu_1 = 1$ the number of $\mu_j = 1$ is the dimension of the torus
a-periodic attractor	$0 \leq \mu_n \leq \ldots \leq \mu_{j+1} \leq 1 < \mu_j = \ldots = \mu_1$ the number of $\mu_j > 1$ is the dimension of the fractal

Remark. Condition $\mu_1\,\mu_2\ldots\mu_n < 1$, which holds for all types of attractors, is clearly implied by the reported conditions for the first three types.

Remark. If the DP specified by transition function $\varphi(\cdot)$ is dissipative, $\delta(\nabla\boldsymbol{\varphi}(\cdot)) < 1$, then $\delta(\nabla\boldsymbol{\varphi}^k(\mathbf{x}_0)) < 1$, thus confirming condition $\mu_1\,\mu_2\ldots\mu_n < 1$.

Remark. Sometimes Lyapunov multipliers are named after Poincarè or Floquet.

The dissipativeness of the DP TEAMs described in Section 4.2.1 is discussed below; some comments are also provided about DP OEAMs from Section 4.2.2.

4.3.1 Dissipativeness of DP-ES/ES

To analyse whether the DP-ES/ES model (4.13, 4.14), specified with respect to arc variables and repeated below for reader's convenience:

$$\mathbf{x}^k = \beta\,\mathbf{c}\left(\mathbf{f}^{k-1};\boldsymbol{\kappa}\right) + (1-\beta)\,\mathbf{x}^{k-1} \quad \forall k \in \mathbb{N} \tag{4.13}$$

$$\mathbf{f}^k = \alpha\,\mathbf{f}(\mathbf{x}^{k-1};\mathbf{d}) + (1-\alpha)\,\mathbf{f}^{k-1} \in S_f \subseteq \mathbb{E}^m \qquad \forall k \in \mathbb{N} \tag{4.14}$$

is dissipative is necessary to first define its Jacobian matrix $\mathbf{J}(\mathbf{x},\mathbf{f})$, which is given by:

$$\mathbf{J}(\mathbf{x},\mathbf{f}) = \left[\begin{array}{c|c} (1-\beta)\,\mathbf{I_m} & \beta\,\mathbf{J}_c(\mathbf{f}) \\ \hline \alpha(1-\beta)\,\mathbf{J}_f(\mathbf{x}) & (1-\alpha)\,\mathbf{I_m} + \alpha\beta\,\mathbf{J}_f(\mathbf{x})\cdot\mathbf{J}_c(\mathbf{f}) \end{array}\right] \tag{4.27}$$

where m is the number of arcs, $\mathbf{J}_f(\mathbf{x}) = \nabla\,\mathbf{f}(\mathbf{x})$ is the Jacobian of the arc flow function, and $\mathbf{J}_c(\mathbf{f}) = \nabla\,\mathbf{c}(\mathbf{f})$ is the Jacobian of the arc cost function (cfr Chapter 3).

The absdet of Jacobian matrix $\mathbf{J}(\mathbf{x},\mathbf{f})$ can be computed from its block structure giving:

$$\delta(\mathbf{J}(\mathbf{x},\mathbf{f})) = (1-\beta)^m\,(1-\alpha)^m \quad \forall(\mathbf{x},\mathbf{f}) \tag{4.28}$$

Since $\alpha\in]0,1[$ and $\beta\in]0,1[$, then $\delta(\mathbf{J}(\mathbf{x}))\in]0,1[$, thus DP-ES/ES is dissipative from any initial state in the whole state space. It is worth noting that the absdet $\delta(\mathbf{J}(\mathbf{x},\mathbf{f}))$ only depends on the updating parameters α and β, and does not depend on the point where is computed, nor on any others parameters such demand flows, capacities, …

An equivalent results is obtained with respect to route (or i-route) variables, with reference to DP (4.11, 4.12):

$$\delta(\mathbf{J}(\mathbf{y},\mathbf{h}))=(1-\beta)^n(1-\alpha)^n \quad \forall(\mathbf{y},\mathbf{h})$$

where n is the number of routes

> Determinant of block matrices 1 (Theorem 3 in Sylvester (2000))
>
> Let $\mathbf{M}$ be a 2×2 block matrix, with first row $[\mathbf{M}_{11}\,|\,\mathbf{M}_{12}]$ and second row $[\mathbf{M}_{21}\,|\,\mathbf{M}_{22}]$, and $\mathbf{M}_{11}$, $\mathbf{M}_{12}$, $\mathbf{M}_{21}$ and $\mathbf{M}_{22}$ square matrices of the same size. If $\mathbf{M}_{11}\cdot\mathbf{M}_{21}=\mathbf{M}_{21}\cdot\mathbf{M}_{11}$, then the determinant of matrix $\mathbf{M}$ is given by: $\det(\mathbf{M})=\det(\mathbf{M}_{11}\cdot\mathbf{M}_{22}-\mathbf{M}_{21}\,\mathbf{M}_{12})$

Since $(1-\beta)\,\mathbf{I_m}\cdot\alpha(1-\beta)\,\mathbf{J}_f(\mathbf{x})=\alpha(1-\beta)\,\mathbf{J}_f(\mathbf{x})\cdot(1-\beta)\,\mathbf{I_m}$,

$$\begin{aligned}\det(\mathbf{J}(\mathbf{x},\mathbf{f}))&=\det((1-\beta)\,\mathbf{I_m}\cdot((1-\alpha)\,\mathbf{I_m}+\alpha\beta\,\mathbf{J}_f(\mathbf{x})\cdot\mathbf{J}_c(\mathbf{f}))-\alpha(1-\beta)\,\mathbf{J}_f(\mathbf{x})\cdot\beta\,\mathbf{J}_c(\mathbf{f})))\\&=\det((1-\beta)(1-\alpha)\,\mathbf{I_m}+\alpha\beta(1-\beta)\,\mathbf{J}_f(\mathbf{x})\cdot\mathbf{J}_c(\mathbf{f}))-\alpha\beta(1-\beta)\,\mathbf{J}_f(\mathbf{x})\cdot\mathbf{J}_c(\mathbf{f})))\\&=(1-\beta)^m(1-\alpha)^m\end{aligned}$$

4.3.2 Dissipativeness of DP-MA/ES

To analyse whether the DP-MA/ES model (4.18, 4.19), specified with respect to route variables and repeated below for reader's convenience:

$$^{1}\mathbf{h}k=\alpha\,\mathbf{h}\left(\Sigma_{j=1,\dots,\mu}\,\zeta_j\,\mathbf{w}\left(^{j}\mathbf{h}k-1;\boldsymbol{\kappa}\right);\mathbf{d}\right)+(1-\alpha)^{1}\mathbf{h}^{k-1}\forall k\in\mathbb{N},k>\mu \tag{4.18}$$

$$^{j}\mathbf{h}^{k}={}^{j-1}\mathbf{h}^{k-1}\quad j=2,\dots,\mu \tag{4.19}$$

is dissipative is necessary to first define its Jacobian matrix $\mathbf{J}(^{|}\mathbf{h})$, with respect to the block vector of route flows $^{|}\mathbf{h}={}^{j}\mathbf{h}$ $j=1,\dots,\mu$. The Jacobian matrix $\mathbf{J}(^{|}\mathbf{h})$ has a special structure [indeed it is a Frobenius block matrix], as shown below for $\mu=4$:

$$\mathbf{J}(|\,\mathbf{h})=\begin{array}{|c|c|c|c|}\alpha\zeta_1\,\mathbf{G}_1+(1-\alpha)\,\mathbf{I_n} & \alpha\zeta_2\,\mathbf{G}_2 & \alpha\zeta_3\,\mathbf{G}_3 & \alpha\zeta_4\,\mathbf{G}_4\\\hline \mathbf{I_n} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\\hline \mathbf{0} & \mathbf{I_n} & \mathbf{0} & \mathbf{0}\\\hline \mathbf{0} & \mathbf{0} & \mathbf{I_n} & \mathbf{0}\end{array} \tag{4.29}$$

where $\mathbf{y}=\mathbf{y}(|\,\mathbf{h})$ is a compact way to express Eq. (4.4.3), $\mathbf{J_h}(|\,\mathbf{h})=\nabla\,\mathbf{h}(\mathbf{y}=\mathbf{y}(|\,\mathbf{h}))$, $\mathbf{J_w}(^{j}\mathbf{h})=\nabla\,\mathbf{w}(^{j}\mathbf{h})$, $\mathbf{G}_j=\mathbf{J_h}(\mathbf{y}(|\,\mathbf{h}))\cdot\mathbf{J_w}(^{j}\mathbf{h})$ with entries depending on system parameters also. Moreover, cfr (4.26), $\zeta_j=\beta\,(1-\beta)^{j-1}/(1-(1-\beta)^{\mu})\geq0\,\forall j=1,2,\dots,\mu$.

The absdet of Jacobian matrix $\mathbf{J}(|\mathbf{h})$ can be computed from its block structure giving:

$$\delta(\mathbf{J}(|\,\mathbf{h})) = \alpha\,\zeta_\mu\,\delta(\mathbf{G}_\mu) \tag{4.30}$$

Determinant of block matrices 2 (cfr Acknowledgements in Sylvester (2000))

Let $\mathbf{M}$ be a 2×2 block matrix, with first row $[\mathbf{M}_{11} \mid \mathbf{M}_{12}]$ and second row $[\mathbf{M}_{21} \mid \mathbf{M}_{22}]$, and $\mathbf{M}_{11}$ and $\mathbf{M}_{22}$ square matrices, $\mathbf{M}_{12}$ and $\mathbf{M}_{21}$ not necessarily square matrices, $\mathbf{M}_{22}$ non singular/invertible.

The determinant of matrix $\mathbf{M}$ is given by: $\det(\mathbf{M}) = \det\left(\mathbf{M}_{11}\,\mathbf{M}_{22} - \mathbf{M}_{12}\,\mathbf{M}_{22}{}^{-1}\,\mathbf{M}_{21}\,\mathbf{M}_{22}\right)$

From matrix algebra the absdet $\delta(\mathbf{J})$ of the above Jacobian matrix $\mathbf{J}$ is equal to absdet of the following matrix $\mathbf{J}'$ obtained through properly interchanging some columns:

$\mathbf{J}'(|\,\mathbf{h}) =$

$\alpha\,\zeta_4\,\mathbf{G}_4$	$\alpha\,\zeta_1\,\mathbf{G}_1 + (1-\alpha)\,\mathbf{I_n}$	$\alpha\,\zeta_2\,\mathbf{G}_2$	$\alpha\,\zeta_3\,\mathbf{G}_3$
$\mathbf{0}$	$\mathbf{I_n}$	$\mathbf{0}$	$\mathbf{0}$
$\mathbf{0}$	$\mathbf{0}$	$\mathbf{I_n}$	$\mathbf{0}$
$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{I_n}$

thus $\delta(\mathbf{J}) = \delta(\mathbf{J}') = \alpha\;\zeta_4\;\delta(\mathbf{G}_4)$

Omitting arguments for simplicity's sake, matrix $\mathbf{G}_\mu = \mathbf{J_h} \cdot \mathbf{J_w}$ is singular since $\mathbf{J_w}$ is singular due to the demand-route flow consistency equation; thus the above result, however elegant, is not useful to assess dissipativeness.

As shown in Appendix A, dissipativeness analysis can be further advanced by specify a DP model with respect to i-route flows equivalent DP-MA/ES model (4.18, 4.19), such that the equivalent of matrix $\mathbf{G}_\mu$ is non-singular.

It turns out that the DP-MA/ES may not be dissipative, that is may not converge to any kind of attractor (from some initial states at least) as day goes to infinite, especially for high values of β and low values of μ. On the other hand, whichever is the value of β, there always exists a large enough memory depth μ^* such that for any memory larger than this value μ^* the DP-MA/ES is *dissipative* from any initial state (further details in Appendix A to this chapter).

4.3.3 Dissipativeness of DP-OEAMs

As in the previous cases, to analyse whether each of the DP-OEAMs described in Section 4.2.2 is dissipative is necessary to first define its Jacobian matrix $\mathbf{J}$ (omitting argument). In all cases the absolute value of the determinant of $\mathbf{J}$ may be out of the range]0, 1[, each of these DP models may be not *dissipative*, that is it may not converge to any kind of attractor, details are omitted for brevity's sake. Therefore, even if there is a unique fixed-point $\mathbf{x}^*$, it may be an attractor from some initial states only, but not from all of them, or not an attractor at all.

Jacobian matrices for models (4.22) and (4.23) are given below for reference.

$$\mathbf{J}(\mathbf{f}) = (1-\alpha)\,\mathbf{I_m} + \alpha\,\mathbf{J}_\mathrm{f}\,(\mathbf{x} = \mathbf{c}(\mathbf{f}))\cdot\mathbf{J}_\mathrm{c}(\mathbf{f})$$

$$\mathbf{J}(\mathbf{x}) = (1-\beta)\,\mathbf{I_m} + \beta\,\mathbf{J}_\mathrm{c}(\mathbf{f} = \mathbf{f}(\mathbf{x}))\cdot\mathbf{J}_\mathrm{f}(\mathbf{x})$$

In both cases the determinant is a function of the entries of matrices $\mathbf{J}_\mathrm{f}\,(\mathbf{x})$ and $\mathbf{J}_\mathrm{c}(\mathbf{f})$ too, besides the updating parameters.

4.4 Fixed-point state local stability and bifurcation analysis

As already noted a deterministic process may evolve towards fixed-points as well as other kind of attractors, such periodic, quasi-periodic, and a-periodic attractors, or may not converge at all, depending on values of parameters and/or initial states. This section describes methods to analyse whether a fixed-point is an attractor without explicitly running the underlying deterministic process model.

A special case of convergent DP occurs when a fixed-point state is an attractor and its attraction domain is the whole state space: from any initial state belonging to the state space the system converges towards it, the (unique) fixed point is globally stable as discussed in Section 4.2.3. When conditions for globally stability cannot be stated, local analysis can be carried out based on the features of the Jacobian matrix of the recursive equations specifying the DP, as shown below.

4.4.1 Local stability analysis

This section presents sufficient conditions for local stability of a fixed-point state of DP-ES/ES, that is a local attraction domain exists thus the fixed-point state is an attractor. These conditions are based on a spectral analysis of the Jacobian matrix of the transition function. As already noted, fixed-points of DP-ES/ES are consistent with CUE.

Some results from matrix analysis

Given a $n \times n$ real matrix $\mathbf{J}$, let

λ_j be one of the n eigenvalues of matrix $\mathbf{J}$, real or component of a complex conjugate pair: $\lambda_j = \lambda_{\mathrm{R}j} \pm i\,\lambda_{\mathrm{I}j}$, with no loss of generality $\lambda_{\mathrm{I}j} > 0$;
$\nu_\#(\mathbf{J}) \geq 0$ be any matrix norm of matrix, where subscript # highlights that there exist several different norms.

the following relations hold:

$$(\lambda_{\mathrm{R}j} + i\,\lambda_{\mathrm{I}j})\,(\lambda_{\mathrm{R}j} - i\,\lambda_{\mathrm{I}j}) = (\lambda_{\mathrm{R}j})^2 + (\lambda_{\mathrm{I}j})^2 = |(\lambda_{\mathrm{R}j} + i\,\lambda_{\mathrm{I}j})|^2 = |(\lambda_{\mathrm{R}j} - i\,\lambda_{\mathrm{I}j})|^2$$

$\det(\mathbf{J}) = \Pi_j\,\lambda_j$, for the determinant of matrix $\mathbf{J}$,
$\delta(\mathbf{J}) = |\det(\mathbf{J})| = \Pi_j|\lambda_j| \geq 0$, for the absolute value determinant of matrix $\mathbf{J}$,
$\rho(\mathbf{J}) = \max_j|\lambda_j| \geq 0$, for the spectral radius of matrix $\mathbf{J}$ by definition,
$\nu_\#(\mathbf{J}) \geq \rho(\mathbf{J})$, for any matrix norm.

Moreover $\nu_\#(\mathbf{J}) < 1 \Rightarrow \rho(\mathbf{J}) < 1 \Rightarrow \delta(\mathbf{J}) < 1$.

Local stability of fixed-point states of a differentiable deterministic process

Let $\mathbf{x}^k = \varphi(\mathbf{x}^{k\text{-}1}; \theta) \in S \subseteq \mathbb{E}^n$ with $\mathbf{x}^0 \in S$ be a differentiable deterministic process with $n \times n$ Jacobian matrix $\nabla \boldsymbol{\varphi}(\mathbf{x}; \boldsymbol{\theta})$, and $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta}) \in S$ be a fixed-point state of it.

Let $\lambda_j{}^*$ be one of the n eigenvalues of the real matrix $\nabla \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta})$, real or component of a complex conjugate pair: $\lambda_j{}^* = \lambda_{Rj}{}^* \pm i\,\lambda_{Ij}{}^*$, with no loss of generality $\lambda_{Ij}{}^* > 0$;

$\rho^* = \text{MAX}_j |\lambda_j{}^*| \geq 0$ be the spectral radius of matrix $\nabla \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta})$.

IF $\rho^* < 1$

THEN the fixed-point state $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta}) \in S$ is locally stable, that is there exists a neighbourhood $S_{\mathbf{x}*} \subseteq S$ of $\mathbf{x}^*$, such that any trajectory converges to the fixed-point state $\mathbf{x}^*$ from any initial state in the neighbourhood:

$$\lim_{k \to \infty} \mathbf{x}^k = \mathbf{x}^* \quad \forall \mathbf{x}^0 \in S_{\mathbf{x}*}$$

In other words, $\mathbf{x}^*$ is a fixed-point attractor and $S_{\mathbf{x}*}$ is its attraction domain.

Remark. This results can be proved by showing that under condition $\rho^* < 1$ a local Lyapunov function exists in a neighbourhood of the fixed-point state. Thus global stability implies local stability.

Remark. On the Argand real—imaginary plan, condition $\rho^* < 1$ means that all the n eigenvalues are within the unitary circle round the origin.

Remark. A sufficient condition for $\rho^* < 1$ is that there exists a matrix norm strictly less than one, $\nu_\#(\nabla \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta})) < 1$, since for any matrix $\mathbf{J}$ $\nu_\#(\mathbf{J}) \geq \rho(\mathbf{J})$ for any matrix norm.

Relationship between dissipativeness and local stability conditions

Let $\mathbf{x}^k = \varphi(\mathbf{x}^{k\text{-}1}; \theta) \in S \subseteq \mathbb{E}^n$ with $\mathbf{x}^0 \in S$ be a differentiable deterministic process with $n \times n$ Jacobian matrix $\nabla \boldsymbol{\varphi}(\mathbf{x}; \boldsymbol{\theta})$, whose spectral radius is $\rho(\nabla \boldsymbol{\varphi}(\mathbf{x}; \boldsymbol{\theta}))$.

IF $\rho\,(\nabla \boldsymbol{\varphi}(\mathbf{x}; \boldsymbol{\theta})) < 1\, \forall \mathbf{x} \in S$

THEN any fixed-point state $\mathbf{x}^* = \boldsymbol{\varphi}(\mathbf{x}^*; \boldsymbol{\theta}) \in S$ is locally stable, moreover the DP is dissipative, since for any matrix $\mathbf{J}$, $\rho(\mathbf{J}) < 1 \Rightarrow \delta(\mathbf{J}) < 1$.

Remark. If the transition function $\boldsymbol{\varphi}(\mathbf{x})$ is continuously differentiable with Jacobian $\boldsymbol{\nabla\varphi}(\mathbf{x})$, a sufficient conditions for being strictly non expansive, and the fixed-point globally stable, is that the a matrix norm (induced by a vector norm) of its Jacobian is less than one, $\nu_\#(\nabla \boldsymbol{\varphi}(\mathbf{x})) < 1\, \forall \mathbf{x} \in S$, For any matrix $\mathbf{J}$ $\nu_\#(\mathbf{J}) \geq \rho(\mathbf{J})$, so $\nu_\#(\mathbf{J}) < 1 \Rightarrow \rho(\mathbf{J}) < 1$, thus local stability conditions hold.

To analyse whether a fixed-point state ($\mathbf{x}^*$, $\mathbf{f}^*$) of the DP-ES/ES model (4.13, 4.14), specified with respect to arc variables, is local stable is necessary to first define the $2m \times 2m$ Jacobian matrix $\mathbf{J}(\mathbf{x}, \mathbf{f})$ (cfr Eq. 4.27, given below for reader's convenience) then its eigenvalues

$$\mathbf{J}(\mathbf{x}, \mathbf{f}) = \left[\begin{array}{c|c} (1-\beta)\,\mathbf{I_m} & \beta\,\mathbf{J_c}(\mathbf{f}) \\ \hline \alpha\,(1-\beta)\,\mathbf{J_f}(\mathbf{x}) & (1-\alpha)\,\mathbf{I_m} + \alpha\,\beta\,\mathbf{J_f}(\mathbf{x}) \cdot \mathbf{J_c}(\mathbf{x}) \end{array}\right] \tag{4.27}$$

where m is the number of arcs, $\mathbf{J_f}(\mathbf{x}) = \nabla\,\mathbf{f}(\mathbf{x})$ is the Jacobian of the arc flow function, and $\mathbf{J_c}(\mathbf{f}) = \nabla\,\mathbf{c}(\mathbf{f})$ is the Jacobian of the arc cost function. In order to highlight the role of the updating parameters as well as of the other parameters and input data, let

$\lambda_j(\mathbf{f}, \mathbf{x})$ be one of the $2m$ eigenvalues of the $2m \times 2m$ matrix $\mathbf{J}(\mathbf{x}, \mathbf{f})$, depending on demand flows, arc capacities, dispersion parameters, updating parameters, etc.; they may be real or may occur in complex conjugate pairs, $\lambda_j = \lambda_{Rj} \pm i\,\lambda_{Ij}$;

$\omega_j(\mathbf{f}, \mathbf{x})$Z be one of the m eigenvalues of the $m \times m$ matrix $\mathbf{G}(\mathbf{x}, \mathbf{f}) = \mathbf{J_f}(\mathbf{x}) \cdot \mathbf{J_c}(\mathbf{f})$, depending on demand flows, arc capacities, dispersion parameters, etc., but not on the updating parameters α and β; these eigenvalues are a-dimensional after the entries of matrix $\mathbf{G}(\mathbf{x}, \mathbf{f})$; they may be real or may occur in complex conjugate pairs, $\omega_j = \omega_{Rj} \pm i\,\omega_{Ij}$.

Omitting arguments to simplify notations, for each eigenvalue ω_j of matrix $\mathbf{G}$ two eigenvalues λ_j and λ_{m+j} of matrix $\mathbf{J}$ can be obtained as solutions of the following (reduced) quadratic equation:

$$\lambda^2 - ((1-\alpha) + (1-\beta) + \alpha\,\beta\,\omega)\,\lambda + ((1-\alpha)\,(1-\beta)) = 0 \tag{4.31}$$

Determinant of block matrices 1 (Theorem 3 in Sylvester (2000))

Let $\mathbf{M}$ be a 2×2 block matrix, with first row $[\mathbf{M}_{11} \mid \mathbf{M}_{12}]$ and second row $[\mathbf{M}_{21} \mid \mathbf{M}_{22}]$, and $\mathbf{M}_{11}$, $\mathbf{M}_{12}$, $\mathbf{M}_{21}$ and $\mathbf{M}_{22}$ square matrices of the same size. If $\mathbf{M}_{11} \cdot \mathbf{M}_{21} = \mathbf{M}_{21} \cdot \mathbf{M}_{11}$, then the determinant of matrix $\mathbf{M}$ is given by: $\det(\mathbf{M}) = \det(\mathbf{M}_{11} \cdot \mathbf{M}_{22} - \mathbf{M}_{21}\,\mathbf{M}_{12})$

Eigenvalues λ of matrix $\mathbf{J}$ are the solutions of the polynomial equation $\det(\mathbf{J} - \lambda\,\mathbf{I_{2m}}) = 0$, that is $\delta(\mathbf{J} - \lambda\,\mathbf{I_{2m}}) = 0$. From at the 2×2 block structure of matrix $\mathbf{J} - \lambda\,\mathbf{I_{2m}}$

$$\mathbf{J} - \lambda\,\mathbf{I_{2m}} = \left[\begin{array}{c|c} (1-\beta)\,\mathbf{I_m} - \lambda\,\mathbf{I_m} & \beta\,\mathbf{J_c} \\ \hline \alpha\,(1-\beta)\,\mathbf{J_f} & (1-\alpha)\,\mathbf{I_m} + \alpha\,\beta\,\mathbf{J_f} \cdot \mathbf{J_c} - \lambda\,\mathbf{I_m} \end{array}\right]$$

it yields:

$$\begin{aligned}\delta(\mathbf{J} - \lambda\,\mathbf{I_{2m}}) &= \delta((1-\beta-\lambda)\,\mathbf{I_m} \cdot ((1-\alpha-\lambda)\,\mathbf{I_m} + \alpha\beta\,\mathbf{J_f}\,\mathbf{J_c})) - \alpha\,(1-\beta)\,\mathbf{J_f} \cdot \beta\,\mathbf{J_c}) \\ &= \delta((1-\beta-\lambda)\,((1-\alpha-\lambda)\,\mathbf{I_m} + \alpha\,\beta\,(1-\beta-\lambda)\,\mathbf{J_f}\,\mathbf{J_c} - \alpha\,\beta\,(1-\beta)\,\mathbf{J_f} \cdot \mathbf{J_c})) \\ &= \delta((1-\beta-\lambda)\,((1-\alpha-\lambda)\,\mathbf{I_m} - \alpha\,\beta\,\lambda\,\mathbf{J_f} \cdot \mathbf{J_c}))\end{aligned}$$

From $\delta((1-\beta-\lambda)\,((1-\alpha-\lambda)\,\mathbf{I} - \alpha\,\beta\,\lambda\,\mathbf{J_f} \cdot \mathbf{J_c})) = 0$, $(1-\beta-\lambda)\,(1-\alpha-\lambda)$ are the eigenvalues of $(\alpha\,\beta\,\lambda)\,\mathbf{G} = (\alpha\,\beta\,\lambda)\,\mathbf{J_f} \cdot \mathbf{J_c}$; moreover, the eigenvalues of $(\alpha\,\beta\,\lambda)\,\mathbf{G}$ are $(\alpha\,\beta\,\lambda)\,\omega$, therefore: $((1-\beta)-\lambda)\,((1-\alpha)-\lambda) = (\alpha\,\beta\,\lambda)\,\omega$, that is Eq. (4.31).

Remark. Since $\lambda_j + \lambda_{m+j} = ((1-\alpha) + (1-\beta) + \alpha\,\beta\,\omega)$ and $\lambda_j\,\lambda_{m+j} = (1-\alpha)(1-\beta)$, $\omega_j = 0$ implies $\lambda_j = (1-\alpha)$ and $\lambda_{m+j} = (1-\beta)$; moreover $\lambda_j = (1-\alpha)$ implies $\lambda_{m+j} = (1-\beta)$, then $\omega_j = 0$, as well as $\lambda_j = (1-\beta)$ implies $\lambda_{m+j} = (1-\alpha)$, then $\omega_j = 0$; therefore $\omega_j = 0$ is equivalent to $\lambda_j = (1-\alpha)$ and $\lambda_{m+j} = (1-\beta)$.

Remark. With $(1-\alpha) = 0$ or $(1-\beta) = 0$ Eq. (4.31) becomes:

$$\lambda^2 - ((1-\alpha) + \alpha\,\omega)\,\lambda = 0 \text{ or } \lambda^2 - ((1-\beta) + \beta\,\omega)\,\lambda = 0$$

with solutions $\lambda_{m+j}=0$, and $\lambda_j=(1-\alpha)+\alpha\ \omega$ or $\lambda_j=(1-\beta)+\beta\ \omega$ respectively; as expected λ_j are the eigenvalues of the Jacobian matrices for models (4.22) and (4.23).

Conditions for local stability of a fixed-point $(\mathbf{x}^*, \mathbf{f}^*)$ requires that the spectral radius, ρ^*, of the Jacobian matrix computed at the fixed-point, $\mathbf{J}(\mathbf{x}^*, \mathbf{f}^*)$, is less than one:

$$\rho^* = \mathrm{MAX}_j|\lambda_j^*| < 1 \tag{4.32}$$

Remark. Since $\omega_j=0$ is equivalent to $\lambda_j=(1-\alpha)\in]0, 1[$ and $\lambda_{m+j}=(1-\beta)\in]0, 1[$, zero eigenvalues $\omega_j=0$ have no effect of fixed-point local stability. Thus, the very same stability condition can be stated with respect to route (or i-route), since in the case ω_j are the eigenvalues of matrix $\nabla\mathbf{h}(\mathbf{w})\cdot\nabla\mathbf{w}(\mathbf{h})$ that has the same non zero eigenvalues of matrix $\mathbf{G}$ (cfr uniqueness condition C/arc-D. Remark 2, in Chapter 2).

Remark. The above considerations also explain the difference between the absdet with respect to arc variables, $(1-\beta)^m\,(1-\alpha)^m$ and the one with respect to route variables $(1-\beta)^n\,(1-\alpha)^n$, without any effect on local stability as well as on dissipativeness.

Remark. Since $\lambda_j\,\lambda_{m+j}=(1-\alpha)\,(1-\beta)<1$, at least one of the two eigenvalues λ_j and λ_{m+j} has absolute value smaller than 1.

Local stability condition

By combining solutions of Eq. (4.31) with Eq. (4.32) stability condition (4.32) can be stated for the DP-ES/ES with respect to the eigenvalues ω^* of matrix $\mathbf{G}(\mathbf{x}^*, \mathbf{f}^*)$ computed at the fixed-point (as shown in Appendix B).

$$\left(\omega_{Rj}{}^* - 1 + e_R\right)^2/e_R{}^2 + \left(\omega_{Ij}{}^*\right)^2/e_I{}^2 < 1 \quad \forall j = 1,\ldots,m \tag{4.33a}$$

where

$$e_R = (1 + (1-\alpha)\,(1-\beta))/(\alpha\,\beta) \geq 1 \tag{4.33b}$$

$$e_I = (1 - (1-\alpha)\,(1-\beta))/(\alpha\,\beta) \leq 1 \tag{4.33c}$$

In other words, all the eigenvalues $\omega_j{}^*$ should be within the interior of a stability region defined by an ellipse on the Argand plan, with shape depending on e_R and e_I being the real and the imaginary semi-axis respectively (Fig. 4.5). It is worth noting that exchanging each other the values of α and β has no effect on the stability region.

The stability region is located on real axis between the values $(1-2\ e_R) = -\ \omega_o(\alpha,\beta)$ and 1, with

$$\omega_o(\alpha, \beta) = 1 + 2\,((1-\alpha) + (1-\beta))/(\alpha\,\beta) \geq 1 \tag{4.34}$$

Updating parameters have a symmetrical effect on the *stability bound function* $\omega_o(\cdot) : \omega_o(\alpha,\beta) = \omega_o(\beta,\alpha)$. It always gets values greater than or equal to 1 and goes to infinity as either of the updating parameters, α or β, goes to zero (Fig. 4.6), thus the lower the values of updating parameters, the greater the area of stability region is. The value of this function is to be considered an input data, since it depends on updating parameters α and β, which are input data resulting from the calibration of the model or design scenario. Several values of parameters α and β map into the same value of function $\omega_o(\alpha,\beta)$, all of them having the same effect on the stability bound on the real axis.

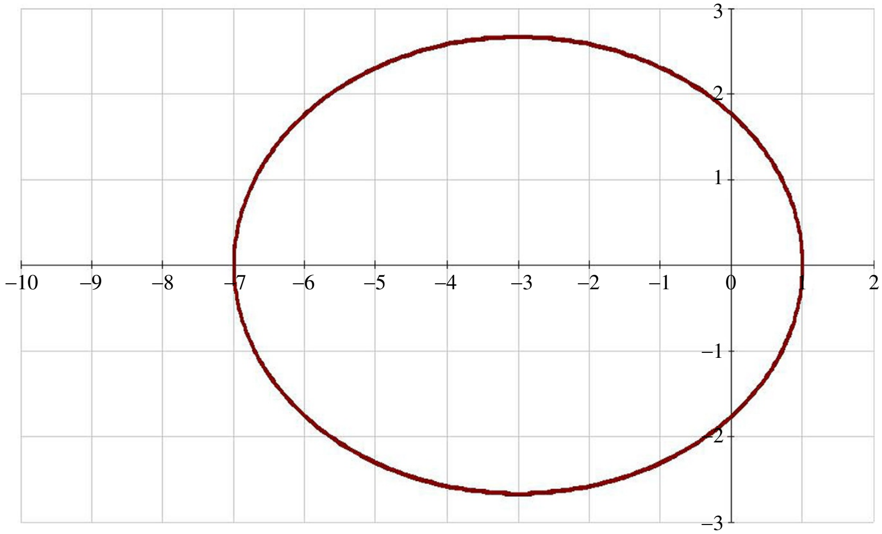


FIG. 4.5

Stability region with $\alpha = 0.50$ and $\beta = 0.60$ (or with $\alpha = 0.60$ and $\beta = 0.50$).

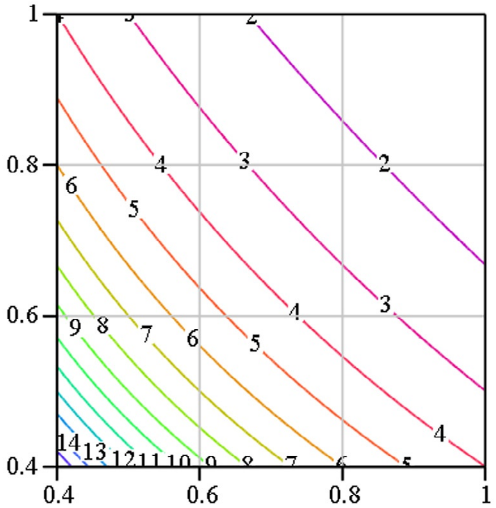


FIG. 4.6

The stability bound function $\omega_o(\alpha, \beta)$ against α and β.

On the imaginary axes ($\omega_R = 0$) the stability region is located between the lower bound $-(1+(1-\alpha)(1-\beta))/(\alpha\beta) \leq -1$ and the upper bound $(1+(1-\alpha)(1-\beta))/(\alpha\beta) \geq +1$; both bounds tend to (minus/plus) infinity as either of the updating parameters, α or β, goes to zero, thus confirming that low values of α and/or β have a stabilisation effect.

When either of the updating parameters is equal to one, $\alpha = 1$ or $\beta = 1$, cfr DP-OY/ES or DP-ES/NI, the stability region is a circle and $\lambda_{n+j} = 0\,\forall j$, when both are equal to one, cfr DP-OY/NI, the stability region is the unitary circle, in this case indeed $\lambda_j = \omega_j\,\forall j$.

Since the stability region is located on the Argand plan on the left of the point (1, 0), if all the eigenvalues ω_j^* have real part less than one, $\omega_{Rj}^* < 1$, there always are values of updating parameters α and β small enough to ensure stability. Otherwise, if there exist at least an eigenvalue ω_j^* that have real part greater than one, $\omega_{Rj}^* > 1$, the fixed-point is always non-stable whatever the values of updating parameters α and β; in this case multiple fixed-points can be found (cfr Uniqueness condition C/arc-D. Remark 2), further comments on this issue in the following Section 4.4.2.

Local stability condition (4.33) allows us to clearly distinguish the role of updating parameters α and β, which only affects the size of the ellipse, from the roles of other parameters, such demand flows, arc capacities, parameters of choice functions and cost function, …, which only affects the eigenvalues ω_j^*.

Hence, the effect of any change of updating parameters α and β can be analysed without re-computing the eigenvalues ω_j, which do not depends on them. Moreover, since eigenvalues ω_j^* have no dimension fixed-point local stability is not influenced by the units chosen for flows and costs.

With reference to the trajectories of flow on route 1 shown in Fig. 4.2, Fig. 4.7 shows the stability region and the eigenvalues ω_j^*, the later depending on the demand flow d, results confirm the qualitative analysis in Fig. 4.2. Similar results can be observed decreasing the Logit dispersion parameter θ and/or the arc capacities.

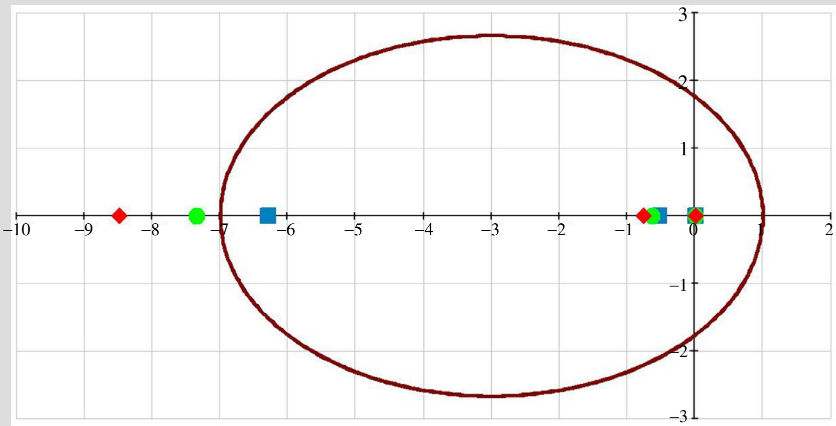


FIG. 4.7

Stability region and eigenvalues for trajectories in Fig. 4.2.

Eigenvalues of a matrix positive semi-definite for real vectors

Let $\mathbf{A}$ be a $m \times m$ square real matrix positive semi-definite for real vectors, $\mathbf{A} \succeq 0$, that is $\mathbf{x}^{\mathbf{T}} \cdot \mathbf{A} \cdot \mathbf{x} \geq 0 \forall \mathbf{x} \neq \mathbf{0}$, $\mathbf{x} \in \mathbb{R}^m$, and $\lambda = \lambda_R + i\lambda_I$ be one of its eigenvalues,

it has non-negative real part, $\lambda_R \geq 0$.

Let be $\mathbf{y} = \mathbf{y}_R + i\,\mathbf{y}_I \neq \mathbf{0}$ be one of the corresponding eigenvectors, so that $\mathbf{A} \cdot \mathbf{y} = \lambda\,\mathbf{y}$,

$\mathbf{A} \cdot \mathbf{y}_R = \lambda_R\,\mathbf{y}_R - \lambda_I\,\mathbf{y}_I$ and $\mathbf{A} \cdot \mathbf{y}_I = \lambda_R\,\mathbf{y}_I + \lambda_I\,\mathbf{y}_R$,

$\mathbf{y}_R^{\mathbf{T}} \cdot \mathbf{A} \cdot \mathbf{y}_R = \lambda_R\,\mathbf{y}_R^{\mathbf{T}} \cdot \mathbf{y}_R - \lambda_I\,\mathbf{y}_R^{\mathbf{T}} \cdot \mathbf{y}_I$ and $\mathbf{y}_I^{\mathbf{T}} \cdot \mathbf{A} \cdot \mathbf{y}_I = \lambda_R\,\mathbf{y}_I^{\mathbf{T}} \cdot \mathbf{y}_I + \lambda_I\,\mathbf{y}_I^{\mathbf{T}} \cdot \mathbf{y}_R$

$\mathbf{y}_R^{\mathbf{T}} \cdot \mathbf{A} \cdot \mathbf{y}_R + \mathbf{y}_I^{\mathbf{T}} \cdot \mathbf{A} \cdot \mathbf{y}_I = \lambda_R\,\mathbf{y}_R^{\mathbf{T}} \cdot \mathbf{y}_R + \lambda_R\,\mathbf{y}_I^{\mathbf{T}} \cdot \mathbf{y}_I = \lambda_R\left(\mathbf{y}_R^{\mathbf{T}} \cdot \mathbf{y}_R + \mathbf{y}_I^{\mathbf{T}} \cdot \mathbf{y}_I\right)$.

Since $\mathbf{y} \neq \mathbf{0}$, $\mathbf{y}_R \neq \mathbf{0}$ and/or $\mathbf{y}_I \neq \mathbf{0}$, thus $\mathbf{y}_R^{\mathbf{T}} \cdot \mathbf{y}_R + \mathbf{y}_I^{\mathbf{T}} \cdot \mathbf{y}_I > 0$; since $\mathbf{x}^{\mathbf{T}} \cdot \mathbf{A} \cdot \mathbf{x} \geq \mathbf{0} \forall \mathbf{x} \neq \mathbf{0}$, $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y}_R^{\mathbf{T}} \cdot \mathbf{A} \cdot \mathbf{y}_R + \mathbf{y}_I^{\mathbf{T}} \cdot \mathbf{A} \cdot \mathbf{y}_I \geq 0$; therefore $\lambda_R \geq 0$.

Eigenvalues of product of definite matrices

Let $\mathbf{A}$ and $\mathbf{B}$ be two $m \times m$ square matrices, $\mathbf{B}$ symmetric positive semi-definite, $\mathbf{B} \succeq 0$.

Remark. Since $\mathbf{B}$ is symmetric positive semi-definite, there exist a matrix $\mathbf{C}$, such that $\mathbf{B} = \mathbf{C} \cdot \mathbf{C}^{\mathrm{T}}$, thus $\mathbf{A} \cdot \mathbf{B} = \mathbf{A} \cdot \mathbf{C} \cdot \mathbf{C}^{\mathrm{T}}$ with the same non zero eigenvalues of matrix $\mathbf{C}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{C}$ (cfr Uniqueness condition C/arc-D. Remark 2).

- IF $\mathbf{A}$ is symmetric THEN each eigenvalue of $\mathbf{A} \cdot \mathbf{B}$ is real ($\mathbf{A} \cdot \mathbf{B}$ may not be symmetric). Indeed in this case matrix $\mathbf{C}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{C}$ is symmetric, thus all its eigenvalues are real.
- IF $\mathbf{A}$ is positive semi-definite for real vectors, $\mathbf{A} \succeq 0$, THEN each eigenvalue of $\mathbf{A} \cdot \mathbf{B}$ has non-negative real part. Indeed, in this case matrix $\mathbf{C}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{C}$ is positive semi-definite for real vectors, thus its eigenvalues have non-negative real part.

Particular conditions can be stated if the arc flow function has a symmetric negative semi-definite Jacobian (as it occurs for invariant choice functions, derived from RUT).

Remark for arc cost functions with positive semi-definite Jacobian

Assuming that the arc flow function has a symmetric negative semi-definite Jacobian, $-\mathbf{J_f}(\mathbf{x}) \succeq 0$, if the Jacobian of arc cost function is *positive semi-definite for real vectors*, $\mathbf{J_c}(\mathbf{f}) \succeq 0$, as it occurs for increasing monotone arc cost functions, each eigenvalue ω_j^* of matrix $-\mathbf{G}(\mathbf{x},\mathbf{f}) = -\mathbf{J_f}(\mathbf{x}) \cdot \mathbf{J_c}(\mathbf{f})$ has non-negative real part, $-\omega_{Rj}^* \geq 0 \forall j$, or $\omega_{Rj}^* \leq 0 \forall j$, thus there always are values of α and β small enough to ensure stability.

Local stability condition for arc cost functions with symmetric Jacobian

Assuming that the arc flow function has a symmetric negative semi-definite Jacobian, $-\mathbf{J_f}(\mathbf{x}) \succeq 0$, if the Jacobian of arc cost function is *symmetric*, as it occurs for separable arc cost functions, each eigenvalue ω_j of matrix $-\mathbf{G}(\mathbf{x},\mathbf{f}) = -\mathbf{J_f}(\mathbf{x}) \cdot \mathbf{J_c}(\mathbf{f})$ is real, $\omega_j = \omega_{Rj} \forall j$, thus the stability conditions (4.33) becomes:

$$\omega_j^* \in]-\omega_o(\alpha,\beta), 1[\quad \forall j=1,\ldots,m \tag{4.35}$$

Conditions (4.35) still holds in any case where all eigenvalues ω_j^* are real.

Local stability condition for arc cost functions with symm. positive semi-definite Jacobian

Assuming that the arc flow function has a symmetric negative semi-definite Jacobian, $-\mathbf{J_f}(\mathbf{x}) \succeq 0$, if the Jacobian of arc cost function is *symmetric positive semi-definite*, as it occurs for separable increasing monotone arc cost functions, each eigenvalue ω_j of matrix $-\mathbf{G}(\mathbf{x},\mathbf{f}) = -\mathbf{J_f}(\mathbf{x}) \cdot \mathbf{J_c}(\mathbf{f})$ is non-positive real, $\omega_j = \omega_{Rj} \le 0 \,\forall j$, thus the stability conditions (4.35) becomes:

$$|\omega_j^*| \le \omega_o(\alpha,\beta)\quad \forall j=1,\ldots,m \tag{4.36}$$

or

$$\rho_G{}^* = \mathrm{MAX}_j|\omega_j^*| \le \omega_o(\alpha,\beta).$$

According to stability conditions (4.36), after the computation of the spectral radius $\rho_G{}^* < 1$ for a fixed-point state, if the choice updating parameter α is known, the maximum value β_{MAX} of the cost updating parameter to guarantee stability, may easily be obtained from condition (4.36) leading to:

$$\beta \le \beta_{MAX} = (4-2\alpha)/(2-\alpha(1-\rho_G{}^*)) \tag{4.37}$$

As expected, the greater $\rho_G{}^*Z$ the smaller the upper bound β_{MAX} is.

Remark. Condition (4.37) may be useful to support ES filter design for information systems.

Similar considerations hold for choice updating parameter α, but they mainly have a theoretical use, since parameter α has a behavioural meaning only and cannot be designed, but has to be calibrated against real data.

Even though the DP-ES/ES model is suitable for large applications, as already noted, the large scale application of stability conditions (4.33), (4.35), (4.36) as such seems quite hard requiring the computation of the eigenvalues of large matrices. In this case it is useful to remember that any matrix norm is an upper bound for the spectral radius, $\nu_\#(\mathbf{J}) \ge \rho(\mathbf{J})$; thus, the spectral radius in stability conditions may be approximated by any matrix norm, such the vector-induced 2-norm $\nu_2(\mathbf{J}) = \|\mathbf{J}\|_2$; this norm usually provides a tight approximation of the spectral radius, but it is rather hard to compute. The Frobenius norm, ν_F, even though not a vector-induced norm, provides values very close to the 2-norm and can easily be computed, since given by the square root of the sum of the square of the entries of the Jacobian matrix; at this aim entries can be generated when needed avoiding the explicit definition of the whole Jacobian matrix.

This approach is very effective for approximating stability condition (4.36) leading to

Approximated local stability condition

$$\nu_F(\mathbf{J}(\mathbf{x}^*, \mathbf{f}^*)) < \omega_o(\alpha, \beta) \tag{4.38}$$

All the above stability conditions are independent of the reference variables, say arc, route, or i-route variables, moreover they hold with $\alpha = 1$ and/or $\beta = 1$, for DP-OY/ES, DP-ES/NI, DP-OY/NI. Similar results have not been fully exploited yet for DP-MA/ES.

4.4.2 Local bifurcation analysis

If a DP is dissipative, but one of its fixed-point states is not an attractor, it is useful to verify towards which type of attractor it converges. Effects of changes of parameters on the type of attractor can be analysed through a bifurcation analysis to further deepen the local stability analysis so far.

Bifurcations of a dissipative DP

A stable fixed-point of a dissipative DP may move towards instability after the change of an eigenvalue induced by a parameter change according to three type of bifurcations.

Flip bifurcation. A real eigenvalue decreases along the real axis below the value -1; the fixed-point becomes unstable and the DP evolves toward a periodic attractor, after a period-doubling an a-periodic one may be reached.

Neumark bifurcation. The absolute value of a pair of complex conjugate eigenvalues increase above the value 1, so that the eigenvalues traverse the border of the unitary circle; the fixed-point becomes unstable and the DP evolves toward a quasi-periodic attractor.

Pitchfork bifurcation. A real eigenvalue increases along the real axis above the value 1; the fixed-point becomes unstable and two new stable fixed-points are attractors, their attraction domains being separated by the unstable fixed-point state. This kind of bifurcation is closely related with fixed-point uniqueness.

Further bifurcations may be observed for non-dissipative systems, which may exhibit a non-converging evolution over time (at least from some initial states).

DP-ES/ES is always dissipative for any state only three types of bifurcations may occur, as discussed below with respect to each eigenvalue ω_j^* (Fig. 4.8).

A real eigenvalue ω_j may lead to a complex conjugate pair of eigenvalues λ_j and λ_{n+j} with $|\lambda_j| = |\lambda_{n+j}|$. In this case $\lambda_j\,\lambda_{n+j} = |\lambda_j|^2 = |\lambda_{n+j}|^2 = (1-\alpha)^2(1-\beta)^2 < 1$ thus the two eigenvalues for a fixed-point state are always within the stability region, and these eigenvalues may not lead to any kind of a bifurcation.

A real eigenvalue ω_j may also lead to two real eigenvalues λ_j and λ_{n+j} such that $\lambda_j\lambda_{n+j} = (1-\alpha)(1-\beta) < 1$. Therefore a real eigenvalue ω_j^* may lead to

a *Flip bifurcation* if ω_j^* is negative and decreases below the value $-\omega_o(\alpha, \beta)$, leading to a periodic attractor; or

a *Pitchfork bifurcation* if ω_j^* is positive and increases beyond the value 1 leading to multiple fixed-points.

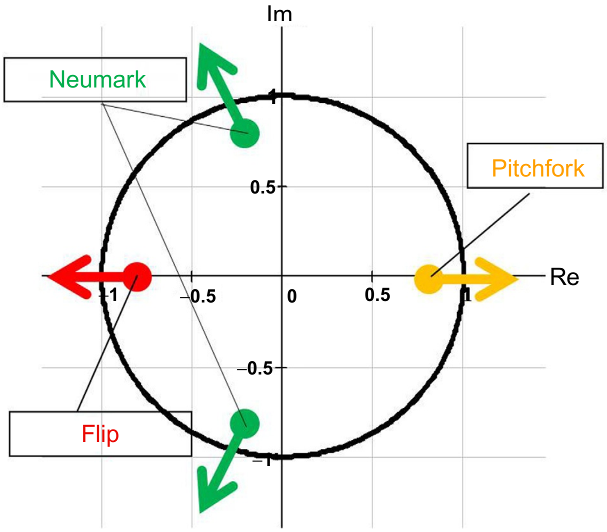


FIG. 4.8

Stability region and bifurcations.

A complex conjugate pair of eigenvalues $\omega_j = \omega_R + i\,\omega_I$ and $\omega_{j+1} = \omega_R - i\,\omega_I$, with $\omega_I > 0$, leads to four two complex eigenvalues λ_j, λ_{m+j}, λ_{j+1}, and λ_{m+j+1} that must occur in two conjugate pairs since they are complex eigenvalues λ of a real matrix**:**

λ_j and λ_{j+1}, with $|\lambda_j| = |\lambda_{j+1}|$

λ_{m+j} and λ_{m+j+1} with $|\lambda_{m+j}| = |\lambda_{m+j+1}|$

Since $|\lambda_j|^2\,|\lambda_{m+j}|^2 = (1-\alpha)^4(1-\beta)^4$, $|\lambda_j|\,|\lambda_{m+j}| = (1-\alpha)^2(1-\beta)^2 < 1$, thus only one pair may be out the unitary circle. Therefore a complex conjugate pair of eigenvalues ω_j^* and ω_{j+1}^* may lead to

a *Neumark bifurcation* if their absolute value $|\omega_j^*| = |\omega_{j+1}^*|$ increases outside the border of the stability region leading to a quasi-periodic attractor.

Assuming that the arc flow function has a symmetric negative semi-definite Jacobian, $-\mathbf{J_f}(\mathbf{x}) \succeq 0$, some further considerations can be drawn.

If the Jacobian of arc cost function is *symmetric*, as it occurs for separable arc cost functions, each eigenvalue ω_j is real, as in 'Local stability condition for arc cost functions with symmetric Jacobian' (Eq. 4.35); thus Flip and Pitchfork bifurcations only may be observed. Hence quasi-periodic attractors may only occur with arc cost function with asymmetric Jacobian; furthermore at least two independent variables are needed to have a pair of complex conjugate eigenvalues, thus a quasi-periodic attractor.

If the Jacobian of arc cost function is *positive semi-definite for real vectors*, $\mathbf{J_c}(\mathbf{f}) \succeq 0$, as it occurs for increasing monotone arc cost functions, each eigenvalue ω_j has non-negative real part, as in 'Remark for arc cost functions with positive semi-definite Jacobian'; thus Flip and Neumark bifurcations only may be observed. Hence, multiple fixed-points may not occur in this case.

If the Jacobian of arc cost function is *symmetric positive semi-definite for real vectors*, $\mathbf{J_c}(\mathbf{f}) \succeq 0$, as it occurs for separable increasing monotone arc cost functions, each eigenvalue ω_j is non-positive real, as in 'Local stability condition for arc cost functions with symm. positive semi-definite Jacobian' (Eq. 4.36); thus Flip bifurcations only may be observed.

If the Jacobian of arc cost function is not *positive semi-definite*, pitchfork bifurcation may be observed only if any eigenvalue $\omega_a{}^*$ has a real part greater than 1 (cfr Uniqueness conditions C/arc-D); in this case the fixed-point state actually reached depend on the start.

Fixed-point bifurcation analysis with respect to a parameter is generally represented by a bifurcation diagram showing the point belonging to the attractor against the values of the parameter, the fixed-points are shown in any case even if they are not attractors.

With reference to the input data of the examples in Figs 4.2 and 4.7, Fig. 4.9 shows the bifurcation diagram of flow on route 1 against the demand flow *d*. Similar results can be observed decreasing the Logit dispersion parameter θ and/or the arc capacities.

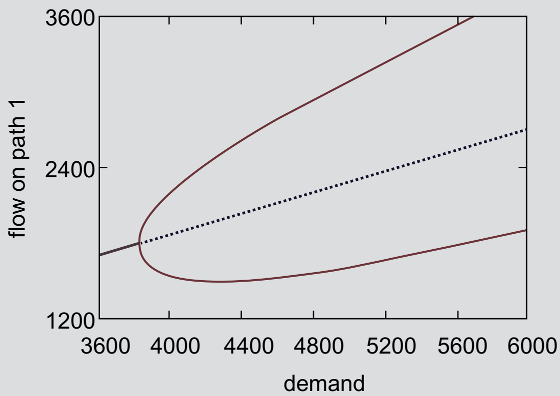


FIG. 4.9

Bifurcation diagram against the demand flow d.

4.5 Basic equations for general models

This section present more general models for cost and choice updating equations EQN3 and EQN5, respectively. DP models based on them are discussed in Section 4.6.

4.5.1 Supply models for general DP models

The arc-route flow consistency relation (4.1) and the arc cost function (4.2) still hold, thus general models for specifying the route-arc cost consistency and updating equation (4.3.#) only are discussed below.

- Route-arc cost consistency and updating function

As already state (cfr Section 4.1.1), yesterday actual route costs depend on yesterday arc costs (cfr Eq. 3.3) through the *route-arc cost consistency relation*:

$$\mathbf{w}^{k-1};_i = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}^{k-1} + \mathbf{w}_{\mathbf{Z}i} \quad \forall i \ \forall k \in \mathbb{N} \tag{4.39a}$$

omitting other route costs $\mathbf{w}_{\mathbf{Z}i}$ and $\mathbf{w}_{\mathbf{Z}}$ in the following for simplifying notation.

Generally, today forecasted costs depend on costs incurred on previous days. Hence, learning and forecasting processes can be modelled through filters applied to costs incurred on previous days. The route-arc cost updating function is made up by combining the *route-arc cost consistency relation* with a *route cost updating filter*.

For simplify notations, forecasted costs $y^k;_i$ are expressed as a function of yesterday actual $\mathbf{w}^{k-1};_i$ and forecasted costs $\mathbf{y}^{k-1};_i$ only, but they may depends on past days:

$$\mathbf{y}^k;_i = \mathbf{y}_i\left(\mathbf{w}^{k-1};_i, \mathbf{y}^{k-1};_i, \ldots\right) \quad \forall i \ \forall k \in \mathbb{N} \tag{4.39b}$$

thus

$$\mathbf{y}^k;_i = \mathbf{y}_i\left(\mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}^{k-1}, \mathbf{y}^{k-1};_i, \ldots\right) \quad \forall i \ \forall k \in \mathbb{N} \tag{4.39}$$

Forecasting filters are assumed day-invariant, that is their functional form and parameters are independent of day k. A filter is *time-homogeneous* if

$$\mathbf{w}^{k-1};_i \neq \mathbf{y}^{k-1};_i \Rightarrow \mathbf{y}^k;_i \neq \mathbf{y}^{k-1};_i \text{ or } \mathbf{y}^k;_i = \mathbf{y}^{k-1};_i \Rightarrow \mathbf{w}^{k-1};_i = \mathbf{y}^{k-1};_i \quad \forall i \ \forall k \in \mathbb{N}$$

that is, if yesterday users experienced costs different from their forecasted costs, today they change their forecasts; vice versa if today they do not change their forecasts, today forecasts are equal to yesterday experience. The above conditions implies that $\mathbf{y}^* = \mathbf{y}(\mathbf{w}, \mathbf{y}^*) \Rightarrow \mathbf{y}^* = \mathbf{w}$: the user forecasting process can start from any initial guess about attributes making up the transportation costs, provided that it can be modified if not confirmed by experience. On the other hand, non modifiable prejudices on congestion dependent attributes are excluded; this assumption does not refer to cost congestion independent attributes, such as fares or scenic quality. ES(β) filter, or its generalisation based on a matrix rather than a single parameter β, is an instance of time-homogenous filter, MA (β, μ) as well.

If the general filter (4.39) is linear with respect to yesterday actual and forecasted costs, the same results are obtained by applying it to arc costs, otherwise two different models are obtained.

An arc cost filter may be time-homogenous as well:

$\mathbf{c}^{k-1} \neq \mathbf{x}^{k-1} \Rightarrow \mathbf{x}^{k} \neq \mathbf{x}^{k-1}$ or $\mathbf{x}^{k} = \mathbf{x}^{k-1} \Rightarrow \mathbf{x}^{k} = \mathbf{c}^{k-1}$

- Route cost updating function

The equations (4.1, 4.2, 4.39) can be combined to define the general *route cost updating function* (G) for given $\mathbf{h}^{0};_{i}\ \forall i$, blocks of $\mathbf{h}^{0}$ and $\mathbf{y}^{0};_{i}\ \forall i$, blocks of $\mathbf{y}^{0}$:

$$\mathbf{y}^{k};_{i} = \mathbf{y}_{i}\left(\mathbf{B}_{i}^{\mathbf{T}} \cdot \mathbf{c}\left(\sum\nolimits_{i} \mathbf{B}_{i} \cdot \mathbf{h}^{k-1};_{i}; \boldsymbol{\kappa}\right), \mathbf{y}^{k-1};_{i}\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{4.40}$$

4.5.2 Demand models for general DP models

The route utility function (4.6) and the demand-route flow consistency relation (4.8) still hold, thus general models for specifying the route choice updating equation (4.7.#) only are discussed below.

If first route utility values are computed from actual route costs, then a general filter is applied to past actual and/or forecasted utility values features of whole model may be affected if it is not linear.

- Route choice updating function

As already noted, when modelling day-to-day dynamics today route choice behaviour is generally affected by user habit effect and inertia to change yesterday choice. In the most general approach, the proportion of users moving between o-d pair i (and belonging to a user class) along route r at day k depends on the route j chosen the previous day due either to habit and conservative behaviour, desire for variety, and/or available information. With reference to day k and o-d pair i (and a user class), let:

$p^{k};_{i\ r} \in [0,1]$ be the route choice proportion that a user chooses route r on, with $\sum_{r} p^{k};_{i\ r} = 1$;

$\mathbf{p}^{k};_{i} \geq \mathbf{0}$ be the $n_i \times 1$ (column) vector of *route choice proportions* with entries $p^{k};_{i\ r}$ such that $\mathbf{1}^{\mathbf{T}}\mathbf{p}^{k};_{i} = 1$;

$p^{k};_{i\ r/j} \in [0,1]$ be the route transition proportion that a user chooses route r, given that j is the route chosen the previous day, such that $\sum_{r} p^{k};_{i\ r/j} = 1$;

$\mathbf{S}^{k};_{i}$ be the $n_i \times n_i$ route transition matrix with entries $p^{k};_{i\ r/j}$; all its entries are non-negative, $\mathbf{S}^{k};_{i} \geq \mathbf{0}$, with column sum equal to 1, $\mathbf{1}^{\mathbf{T}}\mathbf{S}^{k};_{i} = \mathbf{1}^{\mathbf{T}}$, thus $\mathbf{S}^{k};_{i}$ is a column stochastic matrix;

$\mathbf{S}_{i}(\cdot)$ be the $n_i \times n_i$ *route transition matrix function* assumed time-independent.

A *route proportion consistency equation* holds between the today and yesterday route choice proportions and the route transition proportions:

$$p^{k};_{i\,r} = \sum\nolimits_{j} p^{k};_{i\,r/j}\, p^{k-1};_{i\,j} \quad \forall r \in R_i \forall i \forall k \in \mathbb{N}$$

$$\text{or} \quad \mathbf{p}^{k};_{i} = \mathbf{S}^{k};_{i} \cdot \mathbf{p}^{k-1};_{i} \quad \forall i\ \forall k \in \mathbb{N} \tag{4.41a}$$

Generally, the route transition matrix depends on the today route utilities, as well as on yesterday day actual and forecasted route costs (and possibly further past costs, not indicated to simplify notations), through the *route choice updating* filter:

$$\mathbf{S}^k{}_{;i} = \mathbf{S}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}, \ldots\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{4.41b}$$

The route choice updating function is made up by combining the *route proportion consistency equation* with the *route transition function*, leading to the *S filters*:

$$\mathbf{p}^k{}_{;i} = \mathbf{S}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \cdot \mathbf{p}^{k-1}{}_{;i} \quad \forall i\ \forall k \in \mathbb{N} \tag{4.42}$$

Very often conditional route choice models are formulated to explicitly simulate the choice switching process leading to two-stage choice models. Let

$p^k{}_{;i\ C/j} \in]0, 1]$ be the switching choice proportion that a user reconsiders the route j chosen the previous day;

$\mathbf{Q}^k{}_{;i}$ be the $n_i \times n_i$ diagonal *route switching choice matrix* with entries on main diagonal given by $p^k{}_{;i\ C/j}$ and null off-diagonal entries; since the entries on the main diagonal are strictly positive this matrix is not singular (otherwise some users would never reconsider previous day choice);

$\mathbf{Q}_i(\cdot)$ be the $n_i \times n_i$ *route switching choice matrix function* assumed time-independent;

$p^k{}_{;i\ r/Cj} \in [0,1]$ be the active route choice proportion that today a user chooses route r after reconsidering the route j chosen the previous day, with $\sum_r p^k{}_{;i\ r/Cj} = 1$;

$\mathbf{Z}^k{}_{;i}$ be the $n_i \times n_i$ square *route active choice matrix* on day k for each o-d pair i, with entries $p^k{}_{;i\ r/Cj}$; since all its entries are non-negative, $\mathbf{Z}^k{}_{;i} \geq 0$, and its column sum is equal to 1, $\mathbf{1}^{\mathbf{T}}\mathbf{Z}^k{}_{;i} = \mathbf{1}^{\mathbf{T}}$, $\mathbf{Z}^k{}_{;i}$ is a column stochastic matrix;

$\mathbf{Z}_i(\cdot)$ be the $n_i \times n_i$ *route active choice matrix function* assumed time-independent.

As for the route transition matrix, both the route choice switching and route active choice matrices depend on the route utilities as well as other route costs; a *consistency equation* holds between the transition matrix function and these two matrices:

$$\begin{aligned}\mathbf{S}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) = {} & \mathbf{Z}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \cdot \mathbf{Q}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \\ & + \left(\mathbf{I}_{n_i} - \mathbf{Q}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right)\right) \quad \forall i \forall k \in \mathbb{N}\end{aligned} \tag{4.43}$$

Thus Eq. (4.42) can re-written as follows, leading to *ZQ filters*:

$$\begin{aligned}\mathbf{p}^k{}_{;i} = {} & \mathbf{Z}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \cdot \mathbf{Q}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \cdot \mathbf{p}^{k-1}{}_{;i} \\ & + \left(\mathbf{I}_{n_i} - \mathbf{Q}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right)\right) \cdot \mathbf{p}^{k-1}{}_{;i} \quad \forall i \forall k \in \mathbb{N}\end{aligned} \tag{4.44}$$

Eq. (4.43) shows that S filters (4.42) and ZQ filters (4.44) are equivalent; S filters are preferred for theoretical analysis due their compact notations (see also Chapter 5), while ZQ filters are preferred for specification.

The Logit choice function derived from RUT (cfr Chapters 2 and 5; for details see Appendix A to the book) can be used to specify a S or a ZQ filter, as shown below for a user having travelled yesterday along route j, omitting indices for day and o-d pair. Proportions are defined by probabilities.

The switching choice probability $p_{C/j}$ that a user reconsiders the route j chosen the yesterday, an entry of the diagonal matrix $\mathbf{Q}$, can be defined as a function of yesterday forecasted and actual route costs by a Binomial Logit choice function. The choice set contains only two alternatives: reconsider (R), non-reconsider (N). The utility of alternative R is given by the difference $w_j - y_j$ between the yesterday actual w_j and forecasted route costs y_j, thus the greater the difference, the greater the propensity to reconsidering is. The utility of alternative N, say the inertia to change utility, may be assumed proportional to the opposite of the actual cost $-\theta_1 w_j$ with $\theta_1 \geq 0$; thus the smaller the actual cost the greater the propensity to not reconsider. Therefore:

$$p_{C/j} = \exp\left(\left(w_j - y_j\right)/\theta_0\right) / \left(\left(\exp\left(-\theta_1 w_j/\theta_0\right) + \exp\left(\left(w_j - y_j\right)/\theta_0\right)\right)\right)$$

where $\theta_0 > 0$ is the dispersion parameter including the scale factor. This is an instance of the logistic function, often used to described stochastic thresholds.

The j-th column of matrix $\mathbf{Z}$ gives the today active route choice probabilities $p_{r/Cj}$ and can be defined as a function of today forecasted route costs by a Multinomial Logit choice function. The choice set contains all available routes (for the o-d pair i). The utility of each route r different from j is given by the opposite of the route forecasted cost $-y_r$; but for route j chosen yesterday an extra-utility is added; the greater the extra-utility the greater the propensity to not change the route chosen yesterday.

S filters (4.42) and ZQ filters (4.44) are very general and includes as special cases simple models described in Section 4.1.2, and others described below, as well as most existing models. To simplify notation, below the today forecasted route costs only are explicit included as argument of any function related to user choice behaviour.

A simpler modelling approach is obtained by assuming that the switching choice proportion do not depend on the route j chosen made the previous day, and possibly depends on the o-d pair i only: $p^k{}_{;i\ C/j} = p^k{}_{;i\ C}$. Let

$p^k{}_{;i\ C} \in]0, 1]$ be the route independent switching choice proportion; so that:

$\mathbf{A}^k{}_{;i} = p^k{}_{;i\ C}\,\mathbf{I}_{n_i}$ be the $n_i \times n_i$ diagonal *route generic switching choice matrix*, with entries on main diagonal given by $p^k{}_{;i\ C}$ and null off-diagonal entries; since the entries on the main diagonal are strictly positive this matrix is not singular;

$p_{Ci}(\cdot)$ be the *route generic switching choice function* assumed time-independent;

$\mathbf{A}_i(\cdot) = p_{Ci}(\cdot)\mathbf{I}_{n_i}$ be the $n_i \times n_i$ *route generic switching choice matrix function* assumed time-independent.

The generic switching choice matrix generally depends on the difference between actual and forecasted route costs $\mathbf{A}^k{}_{;i} = \mathbf{A}(\mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}) = p_{Ci}(\mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i})\mathbf{I}_{n_i}$, if $\mathbf{Q}^k{}_{;i} = \mathbf{A}^k{}_{;i}$,then $\mathbf{Z}^k{}_{;i} \cdot \mathbf{Q}^k{}_{;i} = \mathbf{A}^k{}_{;i} \cdot \mathbf{Z}^k{}_{;i}$ thus, a simpler version of ZQ filters (4.44) is obtained, leading to *AZ filters*:

$$\begin{aligned}\mathbf{p}^k{}_{;i} &= p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right)\mathbf{Z}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \cdot \mathbf{p}^{k-1}{}_{;i} \\ &\quad + \left(1 - p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right)\right)\mathbf{p}^{k-1}{}_{;i} \quad \forall i \forall k \in \mathbb{N}\end{aligned} \tag{4.45}$$

An even simpler modelling approach is obtained by assuming that all the columns of each matrix $\mathbf{Z}_i(\cdot)$ of are equal to the vector function $\mathbf{p}_i(\mathbf{v}^k{}_{;i};\theta_i)$. If $\mathbf{Z}^k{}_{;i}=\mathbf{p}^k{}_{;i}\cdot\mathbf{1}^{\mathbf{T}}$, then $\mathbf{Z}_i(\mathbf{v}^k{}_{;i})\cdot\mathbf{p}^{k-1}{}_{;i}=\mathbf{p}^k{}_{;i}\cdot\mathbf{1}^{\mathbf{T}}\cdot\mathbf{p}^{k-1}{}_{;i}=\mathbf{p}^k{}_{;i}$, thus a simpler version of AZ filters (4.45) is obtained, leading to *specific Ap filters*:

$$\mathbf{p}^k{}_{;i}=p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\mathbf{p}_i\left(\mathbf{v}^k{}_{;i};\theta_i\right)+\left(1-p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\right)\mathbf{p}^{k-1}{}_{;i}\quad\forall i\forall k\in\mathcal{N}\tag{4.46}$$

If the switching choice proportion function does not depend on the o-d pair (nor on the user class) $p_{ci}(\cdot)=p_c(\cdot)$ a simpler model is obtained, leading to *generic Ap filters*:

$$\mathbf{p}^k{}_{;i}=p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\mathbf{p}_i\left(\mathbf{v}^k{}_{;i};\boldsymbol{\theta}_i\right)+\left(1-p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\right)\mathbf{p}^{k-1}{}_{;i}\quad\forall i\;\forall k\in\mathcal{N}\tag{4.47}$$

If the switching choice matrix does not depend on costs, thus it is constant over time, model (4.47) becomes the ES(α) filter (4.7.2) with $\alpha=p_C(\mathbf{v}^k{}_{;i})$.

- Route flow updating function

Either of the above choice updating model (4.42) or (4.44) can be combined with the utility function (4.6) and the demand-route consistency equation (4.8) to define the general *route flow updating function* for given $\mathbf{h}^0{}_{;i}\forall i$, blocks of $\mathbf{h}^0$:

$$\mathbf{h}^k{}_{;i}=\mathbf{S}_i\left(-\psi_i\,\mathbf{y}^k{}_{;i},\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\cdot\mathbf{h}^{k-1}{}_{;i}\quad\forall i\;\forall k\in\mathcal{N}\tag{4.48}$$

$$\begin{aligned}\mathbf{h}^k{}_{;i}=&\,\mathbf{Z}_i\left(\mathbf{v}^k{}_{;i},\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\cdot\mathbf{Q}_i\left(\mathbf{v}^k{}_{;i},\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\cdot\mathbf{h}^{k-1}{}_{;i}\\&+\left(\mathbf{I}_{\mathbf{n_i}}-\mathbf{Q}_i\left(\mathbf{v}^k{}_{;i},\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\right)\cdot\mathbf{h}^{k-1}{}_{;i}\quad\forall i\;\forall k\in\mathcal{N}\end{aligned}\tag{4.49}$$

with $\mathbf{v}^k{}_{;i}=-\psi_i\,\mathbf{y}^k{}_{;i}$

Similar are results are obtained with simpler choice updating models (4.45), (4.46), and (4.47):

$$\begin{aligned}\mathbf{h}^k{}_{;i}=&\,p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\mathbf{Z}_i\left(-\psi_i\,\mathbf{y}^k{}_{;i}\right)\cdot\mathbf{h}^{k-1}{}_{;i}\\&+\left(1-p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\right)\mathbf{h}^{k-1}{}_{;i}\quad\forall i\;\forall k\in\mathcal{N}\end{aligned}\tag{4.50}$$

$$\mathbf{h}^k{}_{;i}=d_ip_{Ci}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\mathbf{p}_i\left(-\psi_i\,\mathbf{y}^k{}_{;i};\boldsymbol{\theta}_i\right)+\left(1-p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\right)\mathbf{h}^{k-1}{}_{;i}\quad\forall i\;\forall k\in\mathcal{N}\tag{4.51}$$

$$\mathbf{h}^k{}_{;i}=d_ip_{C}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\mathbf{p}_i\left(-\psi_i\,\mathbf{y}^k{}_{;i};\boldsymbol{\theta}_i\right)+\left(1-p_{Ci}\left(\mathbf{w}^{k-1}{}_{;i},\mathbf{y}^{k-1}{}_{;i}\right)\right)\mathbf{h}^{k-1}{}_{;i}\quad\forall i\;\forall k\in\mathcal{N}\tag{4.52}$$

Each of the functions $\mathbf{S}_i(\cdot)$, $\mathbf{Z}_i(\cdot)$, $\mathbf{Q}_i(\cdot)$, $p_{Ci}(\cdot)$, $p_C(\cdot)$ can be specified by applying any choice modelling theory (see Appendix A to the book). Some existing approaches to modelling user inertia to change and habit are briefly described below within the general framework proposed above.

In aggregate approaches to model the effect of reliability of several information sources the switching choice behaviour depends on the aggregate reliability, thus may change over time. Examples of this kind models, sometimes called *bounded-rationality behavioural models*, are mostly based on probabilistic (or deterministic) threshold filters with respect to differences between actual and forecasted costs. Modelling effects of an ATIS reliability is addressed by Bifulco et al. (2016) through a modelling approach consistent with (4.52). Other aggregate approaches are based on extra utility inertia: the route chosen the previous day is given an extra utility, expressing the so-called transition cost to a different alternative.

In route-disaggregate approaches, a choice updating parameter is defined for each route separately depending on the difference between experienced and forecasted arc costs. The use of probabilistic thresholds leads to route choice switching models. This approach is rather effective when only two routes are available between each O-D pair, since there is no need of any route choice function. Indeed, when more than two routes are available, a conditional route choice function should be applied to model route choice behaviour of users who decide to reconsider their yesterday choice.

All above models can be applied in a multi-user framework, as already noted for the simple models discussed in Section 4.1.2. In fully disaggregate approaches, each class is made up by a single user, thus the choice updating parameters may be defined for each single user. These models are better suited for disaggregate assignment through stochastic process models, as described in Chapter 5.

4.5.3 Arc flow updating function

An explicit arc flow updating function extension of (4.10.2) cannot be obtained from any of the above route choice updating functions (4.48), (4.49), (4.50), (4.51), (4.52).

A special case occurs if the switching choice proportion α does not depend on the o-d pair, nor on route forecasted costs, but only depends on arc forecasted costs, $\alpha = \alpha(\mathbf{x})$; in this case the arc flow updating function can properly be defined as:

$$\mathbf{f}^k = \alpha\left(\mathbf{x}^k\right)\mathbf{f}\left(\mathbf{x}^k; \mathbf{d}\right) + \left(1 - \alpha\left(\mathbf{x}^k\right)\right)\mathbf{f}^{k-1} \quad \forall k \in \mathbb{N} \tag{4.53}$$

which results an extension of Eq. (4.10.2).

4.6 General DP models

The set of six equations (4.1) (4.2), (4.40), (4.6), any choice updating equation, and (4.8) defines a Markovian deterministic process (DP) model with respect to all the six basic variables, describing the evolution over time of them.

As in Section 4.2, to further analyse the resulting model it is better to reduce the number of equations and variables. None of the models discussed below is suitable for the explicit stability analysis carried out in Sections 4.3 and 4.4. They can be specified with respect to route (i-route) vectors, and in some particular cases only with arc variables.

4.6.1 Two equation assignment models

Given a cost updating filter and any choice updating filter, the resulting DP models are made by two equations with respect to two vectors, a flow vector and a cost vector, say a two equation assignment models (TEAMs). Let

- Route costs and flows—G/S OR ZQ

$$\mathbf{w}^{k-1};_i = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}\left(\sum\nolimits_i \mathbf{B}_i \cdot \mathbf{h}^{k-1};_i; \boldsymbol{\kappa}\right) \quad \forall i \; \forall k \in \mathbb{N} \tag{4.54}$$

$$\mathbf{y}^k{}_{;i} = \mathbf{y}_i\left(\mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \quad \forall i\ \forall k \in \mathcal{N} \tag{4.55}$$

$$\mathbf{h}^k{}_{;i} = \mathbf{S}_i\left(-\psi_i\,\mathbf{y}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \cdot \mathbf{h}^{k-1}{}_{;i} \quad \forall i\ \forall k \in \mathcal{N} \tag{4.56}$$

Eq. (4.40) has been split into equations (4.54)—made up by (4.1), (4.2) and (4.39a)—and (4.55)—say (4.39b)—for better readability; indeed Eq. (4.54) is an auxiliary static equation (yesterday actual costs depend on yesterday flows); thus this model is still considered a TEAM. A similar model is obtained by applying the ZQ filter instead of the S filter:

$$\begin{aligned}\mathbf{h}^k{}_{;i} = {} & \mathbf{Z}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \cdot \mathbf{Q}_i\left(\mathbf{v}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right) \cdot \mathbf{h}^{k-1}{}_{;i} \\ & + \left(\mathbf{I}_{\mathbf{n_i}} - \mathbf{Q}_i\left(\mathrm{v}\mathbf{y}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}\right)\right) \cdot \mathbf{h}^{k-1}{}_{;i} \quad \forall i\ \forall k \in \mathcal{N}\end{aligned} \tag{4.57}$$

with $\mathbf{v}^k{}_{;i} = -\psi_i\,\mathbf{y}^k{}_{;i}$

The DP model (4.54), (4.55) and (4.56 or 4.57) can easily be rewritten as a proper Markovian DP, today state only depends on yesterday one, by putting Eq. (4.54) into (4.55) and both Eqs (4.54) and (4.55) into (4.56 or 4.57). The main state vectors of DP model (4.54), (4.55) and (4.56 or 4.57) are $(\mathbf{y}^k{}_{;i}\ \mathbf{h}^k{}_{;i})\forall i$, $\mathbf{w}^{k-1}{}_{;i}\forall i$ being auxiliary vectors introduced for readability only; the updating parameters are those in the updating filters; other parameters are demand flows, and any other parameter in choice functions and in the arc cost function. Model (4.54), (4.55) and (4.56) is useful to highlight the non-linearity of the relationship between today and yesterday flows, while (4.54), (4.55) and (4.57) is useful to clearly distinguish the user route choice behaviour and the switching one. Both the above models are useful when explicit path enumeration can be carried out; it is worth noting that this is hardly the case if routes are hyperpaths. These models are also useful as a base for developing stochastic process models described in Chapter 5. No specification of this general model has been analysed in literature, still some theoretical results can be drawn as shown in the next sub-sections.

Any simple cost updating equation presented in Section 4.1.1 may be used instead of (4.54) and (4.55) as well as any simplified model (4.50), (4.51), (4.52) can be used instead of (4.56) or (4.57). With reference to the route cost updating Eq. (4.52), Eqs (4.40) and (4.53) defines an equivalent model based on arc vectors $(\mathbf{x}^k, \mathbf{c}^k)$:

- Arc costs and flows—G/A

$$\mathbf{x}^k = \mathbf{x}\left(\mathbf{c}\left(\mathbf{f}^{k-1}; \boldsymbol{\kappa}\right), \mathbf{x}^{k-1}\right) \geq \mathbf{0} \quad \forall \mathbf{f}^{k-1} \in S_f\ \forall k \in \mathcal{N} \tag{4.40}$$

$$\mathbf{f}^k = \alpha\left(\mathbf{x}^k\right)\mathbf{f}\left(\mathbf{x}^k; \mathbf{d}\right) + \left(1 - \alpha\left(\mathbf{x}^k\right)\right) \quad \forall k \in \mathcal{N} \tag{4.53}$$

4.6.2 OEAMs

One equation assignment models (OEAMs) may be based on the route cost updating Eq. (4.4.1) with the Only Yesterday (OY) cost updating filter combined with any of the above route flow updating equation. In this case indeed Eq. (4.55) becomes: $\mathbf{y}^k{}_{;i} = \mathbf{w}^{k-1}{}_{;i},\ \forall i\ \ \forall k \in N$. Another kind of OEAMs are based on the route flow updating Eq. (4.9.1) with the no inertia (NI) flow updating filter combined with the general route cost updating Eq. (4.40). These models have been quoted out of sense of completeness, since they combine a very detailed specification of an updating equation with a very simple specification of the other.

4.6.3 Fixed-point states of general deterministic processes

Generally fixed-point states of model (4.54), (4.55) and (4.56) or (4.55) and (4.57) are not consistent with CUE, as described by the fixed-point models presented in Chapter 3.

If the cost updating filter is not time-homogenous (cfr Section 4.5.1), the fixed-point states depend on the route cost updating filter, and the route actual costs $\mathbf{w}^*$ at a fixed-point state may be different from route forecasted costs $\mathbf{y}^*$ Indeed the fixed-point state conditions

$$\mathbf{y}^k;_i = \mathbf{y}^{k-1};_i = \mathbf{y}_i{}^* \text{ and } \mathbf{h}^k;_i = \mathbf{h}^{k-1};_i = \mathbf{h}_i{}^*, \text{with } \mathbf{w}_i{}^* = \mathbf{w}_i(\mathbf{h}^*; \boldsymbol{\kappa}) \quad \forall i$$

combined with Eqs (4.54), (4.55) and (4.56), gives for each o-d pair i:

$$\mathbf{w}_i{}^* = \mathbf{B}_i^{\mathbf{T}} \cdot \mathbf{c}\left(\sum\nolimits_i \mathbf{B}_i \cdot \mathbf{h}^{k-1};_i; \boldsymbol{\kappa}\right) \quad \forall i \tag{4.58}$$

$$\mathbf{y}_i{}^* = \mathbf{y}_i(\mathbf{w}_i{}^*, \mathbf{y}_i{}^*) \quad \forall i \tag{4.59}$$

$$\mathbf{h}_i{}^* = \mathbf{S}_i(-\psi_i\, \mathbf{y}_i{}^*, \mathbf{w}_i{}^*, \mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^* \quad \forall i \tag{4.60}$$

On the other hand, if the cost updating filter is time-homogenous:

$$\mathbf{y}^k;_i = \mathbf{y}^{k-1};_i \Rightarrow \ \mathbf{y}^k;_i = \mathbf{w}^{k-1};_i$$

Eq. (4.59) becomes: $\mathbf{y}_i{}^* = \mathbf{w}_i{}^*$ that combined with Eq. (4.58) gives:

$$\mathbf{y}_i{}^* = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}\left(\sum\nolimits_i \mathbf{B}_i \cdot \mathbf{h}^{k-1};_i; \boldsymbol{\kappa}\right) \quad \forall i \tag{4.61}$$

Let $\overline{\mathbf{S}}_i(\mathbf{y}_i{}^*) = \mathbf{S}_i(-\psi_i \mathbf{y}_i{}^*, \mathbf{y}_i{}^*, \mathbf{y}_i{}^*)$ simplifying notations of arguments for readability, Eq. (4.60) can be re-written:

$$\overline{\mathbf{S}}_i(\mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^* = \mathbf{h}_i{}^* \quad \forall i \tag{4.62}$$

Thus, for time-homogenous cost updating filters fixed-point states ($\mathbf{y}^*$,$\mathbf{h}^*$) are the solutions of the two equations (4.61) and (4.62). The analysis of these equations shows that *the fixed-point states do not depend on any time-homogenous cost updating filter*.

Eigenvalues and eigenvectors of stochastic matrices

Let $\mathbf{S}$ be a $n \times n$ (non-negative) column (or left) stochastic matrix, that is all its entries are non-negative, $\mathbf{S} \geq \mathbf{0}$,with column sum equal to 1, $\mathbf{1}^{\mathbf{T}} \cdot \mathbf{S} = \mathbf{1}^{\mathbf{T}}$, it has

- at least an eigenvalue equal to 1, $\mathbf{S} \cdot \boldsymbol{\pi} = \boldsymbol{\pi}$, with a non-negative real eigenvector $\boldsymbol{\pi} \geq \mathbf{0}$, the eigenvector $\boldsymbol{\pi}$ with sum equal to 1 is called a *Perron vector* $\boldsymbol{\pi}_{\mathbf{P}} \geq \mathbf{0}$ with $\mathbf{1}^{\mathbf{T}} \boldsymbol{\pi}_{\mathbf{P}} = 1$;
  if $\mathbf{S}$ is also *irreducible*, a necessary and sufficient condition being $(\mathbf{I_n} + \mathbf{S})^{n-1} > 0$, it has
- one eigenvalue equal to 1 with a positive real eigenvector $\boldsymbol{\pi} > \mathbf{0}$ (Frobenius theorem), and the Perron vector $\boldsymbol{\pi}_{\mathbf{P}}$ is unique;
  if $\mathbf{S}$ is also positive, $\mathbf{S} > \mathbf{0}$, it is irreducible and
- each of the other eigenvalues has absolute value less than 1 (Perron theorem).

Since $\overline{\mathbf{S}}_i(\mathbf{y}_i{}^*)$ is a column stochastic matrix it has at least an eigenvalue equal to 1, thus for given $\mathbf{y}_i{}^*$ each route flow vector $\mathbf{h}_i{}^*$ solution of (4.62) is equal to the product between the demand flow and a Perron vector, $\mathbf{h}_i{}^* = d_i \boldsymbol{\pi}_\mathbf{P}$;if $\overline{\mathbf{S}}_i(\mathbf{y}_i{}^*)$ is also irreducible it has exactly one eigenvalue equal to 1, and the route flow vector $\mathbf{h}_i{}^* = d_i \boldsymbol{\pi}_\mathbf{P}$ is unique.

Similar considerations hold with the choice updating filter ZQ leading to the route flow updating Eq. (4.57); and the following equation is used instead of (4.60):

$$\mathbf{h}_i{}^* = \mathbf{Z}_i(\mathbf{v}_i{}^*, \mathbf{w}_i{}^*, \mathbf{y}_i{}^*) \cdot \mathbf{Q}_i(\mathbf{v}_i{}^*, \mathbf{w}_i{}^*, \mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^* + (\mathbf{I}_{\mathbf{n_i}} - \mathbf{Q}_i(\mathbf{I}_{\mathbf{n_i}} - \mathbf{Q}_i(\mathbf{v}_i{}^*, \mathbf{w}_i{}^*, \mathbf{y}_i{}^*)) \cdot \mathbf{h}_i{}^* \quad \forall i \tag{4.63}$$

with $\mathbf{v}_i{}^* = -\psi_i\, \mathbf{y}_i{}^*$

As above, if the cost updating filter is time-homogenous, Eq. (4.59) combined with Eq. (4.58) gives Eq. (4.61). In this case, let $\overline{\mathbf{Z}}_i(\mathbf{y}_i{}^*) = \mathbf{Z}_i(-\psi_i \mathbf{y}_i{}^*, \mathbf{y}_i{}^*, \mathbf{y}_i{}^*)$ and $\overline{\mathbf{Q}}_i(\mathbf{y}_i{}^*) = \mathbf{Q}_i(-\psi_i \mathbf{y}_i{}^*, \mathbf{y}_i{}^*, \mathbf{y}_i{}^*)$, simplifying notations of arguments for readability, Eq. (4.63) can be re-written:

$$\overline{\mathbf{Z}}_i(\mathbf{y}_i{}^*) \cdot \left(\overline{\mathbf{Q}}_i(\mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^*\right) = \left(\overline{\mathbf{Q}}_i(\mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^*\right) \quad \forall i \tag{4.64}$$

Fixed-point states ($\mathbf{y}^*$, $\mathbf{h}^*$) are the solutions of the two equations (4.61) and (4.64). The analysis of these equations shows that the fixed-point states do not depend on any time-homogenous cost updating filter.

Since $\overline{\mathbf{Z}}_i(\mathbf{y}_i{}^*)$ is a column stochastic matrix it has at least an eigenvalue equal to 1, thus $\overline{\mathbf{Q}}_i(\mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^* = \delta_i \boldsymbol{\pi}_\mathbf{P}$ where δ_i is a scaling factor. Remembering that $\overline{\mathbf{Q}}_i(\mathbf{y}_i{}^*)$ is non-singular, for given $\mathbf{y}_i{}^*$ each route flow vector $\mathbf{h}_i{}^*$ solution of (4.65) can be obtained from relation $\mathbf{h}_i{}^* = \delta_i \overline{\mathbf{Q}}_i(\mathbf{y}_i{}^*)^{-1} \cdot \boldsymbol{\pi}_\mathbf{P}$ with scaling factor $\delta_i = d_i / \left(\mathbf{1}^\mathbf{T} \cdot \overline{\mathbf{Q}}_i(\mathbf{y}_i{}^*)^{-1} \cdot \boldsymbol{\pi}_\mathbf{P}\right)$ so that $\mathbf{1}^\mathbf{T}\mathbf{h}_i{}^* = d_i$;if $\overline{\mathbf{Z}}_i(\mathbf{y}_i{}^*)$ is also irreducible it has exactly one eigenvalue equal to 1, and the route flow vector $\mathbf{h}_i{}^* = d_i \overline{\mathbf{Q}}_i(\mathbf{y}_i{}^*)^1 \cdot \boldsymbol{\pi}_\mathbf{P} / \left(\mathbf{1}^\mathbf{T} \cdot \overline{\mathbf{Q}}_i(\mathbf{y}_i{}^*)^{-1} \cdot \boldsymbol{\pi}_\mathbf{P}\right)$ is unique.

With reference to the choice updating filter AZ leading to the route flow updating Eq. (4.50); in this case the following equation is used instead of (4.62) or (4.64):

$$\mathbf{h}_i{}^* = p_{Ci}(-\psi_i\, \mathbf{y}_i{}^*) \overline{\mathbf{Z}}_i(\mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^* + (1 - p_{Ci}(-\psi_i\, \mathbf{y}_i{}^*)) \mathbf{h}_i{}^* \quad \forall i$$

$$\text{or} \quad \mathbf{0} = p_{Ci}(-\psi_i\, \mathbf{y}_i{}^*) \overline{\mathbf{Z}}_i(\mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^* - p_{Ci}(-\psi_i\, \mathbf{y}_i{}^*) \mathbf{h}_i{}^* \quad \forall i$$

$$\text{or} \quad \mathbf{0} = \overline{\mathbf{Z}}_i(\mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^* - \mathbf{h}_i{}^* \quad \forall i$$

since $p_{Ci}(-\psi_i\, \mathbf{y}_i{}^*) > 0$. Therefore

$$\overline{\mathbf{Z}}_i(\mathbf{y}_i{}^*) \cdot \mathbf{h}_i{}^* = \mathbf{h}_i{}^* \quad \forall i \tag{4.65}$$

The analysis of Eqs (4.58), (4.59) and (4.65) shows that *the fixed-point states do not depend on switching behaviour for any ZQ filters*. Eq. (4.65) looks like (4.62) with matrix $\overline{\mathbf{Z}}_i(\mathbf{y}_i{}^*)$ instead of $\overline{\mathbf{S}}_i(\mathbf{y}_i{}^*)$, thus all above comments still applies.

If the switching choice proportion only depends on arc forecasted costs, $\alpha = \alpha(\mathbf{x}) > 0$,the arc flow updating function can properly be defined by Eq. (4.53) to be coupled with the arc cost updating Eq. (4.40). Assuming a time-homogenous cost updating filter, $\mathbf{x}^k = \mathbf{x}^{k-1} = \mathbf{x}^* \Rightarrow \mathbf{x}^* = \mathbf{c}^*$, the fixed-point states ($\mathbf{x}^*$,$\mathbf{f}^*$) of this DP are given by:

$$\mathbf{x}^* = \mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa})$$

$$\mathbf{f}^* = \alpha(\mathrm{x})\mathbf{f}(\mathbf{x}^*; \mathbf{d}) + (1 - \alpha(\mathrm{x}^*))\mathbf{f}^*$$

or

$$\mathbf{x}^* = \mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa}) \tag{4.66}$$

$$\mathbf{f}^* = \mathbf{f}(\mathbf{x}^*; \mathbf{d}) \tag{4.67}$$

They are independent of the switching behaviour and consistent with fixed-point states of simple DP models discussed in Section 4.2.3 and with CUE, as in Chapter 3.

General fixed-point existence conditions

Sufficient conditions for the existence of fixed-points of general deterministic process models with time-homogeneous cost updating filters are given in the following. These conditions are a generalisation of those already discussed for CUE in Section 3.2.3. For brevity's sake they are discussed with S filters for choice updating only, Eq. (4.43) allowing to apply the so-obtained results to ZQ filters, and their simpler versions.

It is worth to express Eq. (4.61) in a different way. At this aim, remembering the arc-route flow consistency equation (3.1), the arc cost function (3.2), and the route-arc cost consistency equation (3.3), let

- $\mathbf{f}^* = \sum_i \mathbf{B}_i \mathbf{h}_i^*$ be the arc flows corresponding to the fixed-point route flows $\mathbf{h}_i^* \forall i$ i;
- $\mathbf{c}^* = \mathbf{c}(\mathbf{f}^*; \mathbf{k})$ be the corresponding arc costs;
- $\mathbf{y}_i^* = \mathbf{B}_i^{\mathbf{T}} \cdot \mathbf{c}^*$ be the route costs corresponding to the fixed-point route flows $\mathbf{h}_i^* \forall i$.

Thus, Eq. (4.61) is equivalent to the following equations:

$$\mathbf{f}^* = \sum_i \mathbf{B}_i \mathbf{h}_i^* + \mathrm{f}_{\mathrm{Z}} \tag{4.68}$$

$$\mathbf{c}^* = \mathbf{c}(\mathbf{f}^*; \mathbf{k}) \tag{4.69}$$

$$\mathbf{y}_i^* = \mathbf{B}_i^{\mathbf{T}} \cdot \mathbf{c}^* + \mathbf{w}_{\mathbf{Z}i} \quad \forall i \tag{4.70}$$

Moreover, Eq. (4.62) implicitly defines a map between the route flow vector $\mathbf{h}_i^*$ and the route cost vector $\mathbf{y}_i^*$; if matrix $\mathbf{S}_i(\mathbf{y}_i^*)$ is irreducible there is one route flow vector $\mathbf{h}_i^*$ for each route cost vector $\mathbf{y}_i^*$, thus this map is a function $\boldsymbol{\varphi}_{\mathbf{S}i}(\mathbf{y}_i^*)$, called the *S route flow function* for o-d pair i, that can be used to express equation(4.62):

$$\mathbf{h}_i^* = \boldsymbol{\varphi}_{\mathbf{S}i}(\mathbf{y}_i^*) \quad \forall i \tag{4.71}$$

Since the eigenvectors are continuous function of the entries of a matrix, if function $\mathbf{S}_i(\cdot)$ is continuous, the S route flow function $\boldsymbol{\varphi}_{\mathbf{S}i}(\cdot)$ is continuous too.

Therefore, fixed-point states $(\mathbf{h}_i{}^*, \mathbf{y}_i{}^*)$ defined by Eqs (4.61) and (4.62) can also be described by (4.68), (4.69), (4.70), and (4.71). Furthermore, the *generalised arc flow function* can be defined combining together Eqs (4.68), (4.70), and (4.71):

$$\boldsymbol{\varphi}_{\mathbf{ARC}}(\mathbf{c}) = \sum\nolimits_i \mathbf{B}_i \boldsymbol{\varphi}_{\mathbf{S}i}\left(\mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}\right) \tag{4.72}$$

It is continuous if the S route flow function is continuous. It has values in the feasible arc flow set $S_f \subseteq \mathbb{E}^m$, which

- has a finite dimension if the number of arcs is finite,
- is non empty if each o-d pair is connected by at least one route,
- is compact, since closed and bounded,
- is convex.

Thus, fixed-point states $(\mathbf{h}_i{}^*, \mathbf{y}_i{}^*)$ $\forall i$ are equivalent to the solutions ($\mathbf{f}^*$, $\mathbf{c}^*$) of the following two equations:

$$\mathbf{c}^* = \mathbf{c}(\mathbf{f}^*; \mathbf{k}) \tag{4.73}$$

$$\mathbf{f}^* = \boldsymbol{\varphi}_{\mathbf{ARC}}(\mathbf{c}^*) \tag{4.74}$$

These equations are a generalisation of the TEAM (3.15), (3.16) presented in Chapter 3 for CUE. Combining them together the following fixed-point model with respect to arc flows can be defined as a generalisation of the OEAM (3.17):

$$\mathbf{f}^* = \boldsymbol{\varphi}_{\mathbf{ARC}}(\mathbf{c}(\mathbf{f}^*; \boldsymbol{\kappa})) \in S_f \tag{4.75}$$

Solutions $\mathbf{f}^*$ of (4.75) allow to define the solutions ($\mathbf{f}^*$, $\mathbf{c}^*$) of (4.73) and (4.74) that are equivalent to the solutions $(\mathbf{h}_i{}^*, \mathbf{y}_i{}^*)$ $\forall i$ of (4.68–4.72), say of (4.61) and (4.62). Hence existence conditions can be stated with respect to the OEAM (4.75).

Thus, sufficient conditions for existence of solutions $\mathbf{f}^*$ of (4.75) can be derived with reference to the composed function $\boldsymbol{\varphi}_{\mathbf{G}}(\mathbf{c}(\cdot))$ defined over set S_f with values in the same set, as:

1. each o-d pair i is connected by at least one route,
2. the arc cost function $\mathbf{c}(\mathbf{f}; \boldsymbol{\kappa})$ is continuous with respect to the arc flows $\mathbf{f}$,
3. the generalised arc flow function $\boldsymbol{\varphi}_{\mathbf{ARC}}(\mathbf{c})$ is continuous with respect to the arc costs $\mathbf{c}$, that is functions $\boldsymbol{\varphi}_{\mathbf{S}i}(\cdot)\,\forall i$ are continuous, say functions $\mathbf{S}_i(\cdot)$ are continuous and matrices $\mathbf{S}_i(\cdot)$ are irreducible (otherwise $\boldsymbol{\varphi}_{\mathbf{S}i}(\cdot)$ is not a function).

Proof, based on the Brouwer's theorem, is similar to the proof of existence conditions in Chapter 3, with reference to the composed function $\boldsymbol{\varphi}_{\mathbf{ARC}}(\mathbf{c}(\cdot))$ instead of $\mathbf{f}(\mathbf{c}(\cdot))$.

General fixed-point uniqueness conditions

Sufficient conditions for the uniqueness of fixed-points of general deterministic process models with time-homogeneous cost updating filters can stated by applying the uniqueness conditions discussed in Chapter 3 to the two equations (4.61) and (4.71) with respect to route vectors, a generalisation of equations (3.11) and (3.12), or to two equations (4.73) and (4.74) with respect to arc variables, a generalisation of equations (3.15) and (3.16). But it should be remarked that generally it is not easy to check whether functions $\boldsymbol{\varphi}_{\mathbf{S}}(\cdot)$ or $\boldsymbol{\varphi}_{\mathbf{ARC}}(\cdot)$ have the required features.

On the other hand, with route choice updating Eq. (4.51) based on Ap choice updating filters (4.46), since $d_i p_{Ci}(\mathbf{y}_i{}^*) > 0\ \forall i$, Eq. (4.62) becomes:

$$\mathbf{h}_i{}^* = d_i \mathbf{p}_i(\mathbf{v}_i{}^* = -\psi_i\, \mathbf{y}_i{}^*; \theta_i)\ \ \forall i$$
$$\text{or}\ \ \mathbf{h}_i{}^* = \mathbf{h}_i(\mathbf{y}_i{}^*; d_i)\ \ \forall i \tag{4.76}$$

with $\boldsymbol{\varphi}_{\mathbf{S}i}(\mathbf{y}_i{}^*) = d_i \mathbf{p}_i(\mathbf{v}_i{}^* = -\psi_i\, \mathbf{y}_i{}^*; \boldsymbol{\theta}_i)$. Eqs (4.61) and (4.76) are exactly equations (3.11) and (3.12) in Chapter 3.

4.6.4 Solution issues and convergence analysis

Any of the above described discrete-time Markovian deterministic process models can be solved by repeatedly applying the recursive equations that specify it, given an initial state and values of all parameters.

At each iteration of general models based on route costs and flows, arc costs and flows can be computed through equations (4.3.#) and (4.1) respectively. Still these models require explicit route enumeration, thus they may be unsuitable for very large scale applications.

In any case, the trajectory of a DP may not converge to a fixed-point state and other kinds of attractors may be reached possibly depending on the initial state, as already noted for simple DP models.

4.7 Summary

4.7.1 Major findings

This chapter presented several simple and some general discrete-time Markovian deterministic process models, casted within the general SEAM framework, for the day-to-day dynamic assignment to congested transportation networks. They are based on a model of user memory and learning and a model of user habit and inertia to change. The presented DP models may include route choice functions from any choice modelling theory (cfr Appendix A to the book). They have been developed under steady-state conditions, but they can be applied to any transportation system with supply modelled by a TAN. Relationship with fixed-point models for CUE, discussed in Chapter 3 has also been analysed.

This chapter also presented a general methodology to analyse discrete-time Markovian deterministic process models (see for instance Stokey and Lucas, 1989 for more theoretical details). First, fixed-point states are defined and existence, uniqueness, and global stability are investigated; then, if global stability cannot be assessed, the Jacobian matrix is defined and dissipativeness analysis is carried out; furthermore, the eigenvalues of the Jacobian matrix at a fixed-point state are defined and the fixed-point local stability and bifurcation analyses can be carried out.

This methodology has been applied to the presented DP models as far as possible, a full analysis being available for some simple models only. The simple DP models based on MA cost updating filters deserve further research effort to get complete fixed-point local stability and bifurcation analyses. General conditions for global stability are still an open issue. Whether the general fixed-point existence conditions could be extended to non irreducible stochastic matrices through the Kakutani's theorem is still an open issue as well.

The today state may also depends on itself when the system state is the result of aggregation and/or averaging over sub-periods of the day k, or for idealised systems used as 'benchmarks', such as traveller information systems where the future can perfectly be forecasted, see for instance Bifulco et al. (2016). As already noted, the resulting DP models can be put in Markovian form by applying either of the approaches described in Appendix C.

As already noted for CUE models, the proposed modelling approach can rather easily be extended to assignment with demand flows variable with respect to costs, and/or multi-type or multi-mode assignment, where the choice behaviour among vehicle types or transportation modes is explicitly described by choice models. These extensions are out of the scope of this book (and will possibly be described in a future book on advanced topics). All parameters introduced above are to be calibrated against real/simulated data (Cheng et al. (2019) recently addressed this issue); this relevant issues as well as implementation and application issues are out the scope of this book, mainly focusing on mathematical features.

4.7.2 Further readings

Very few applications exist based on real input data. Surely worth of mention are the several papers dedicated to modal split in freight transportation by Ferrari (2009, 2011, 2014, 2015, 2016, 2018).

Since the seminal paper by Horowitz (1984), the basic framework of DP models has been developed in some papers published during 90s. After a break in the sequent decade, many papers have been published from 2010, too many to be mentioned. Some papers published by the authors are enlisted below. An analysis of attraction basins or domains is in (Bie and Lo, 2010). An application for congestion toll design is in (Han et al., 2017). An application to a simple multi-modal systems is in Cantarella et al. (2019).

The approximated stability conditions presented in Section 4.4 may be included as a constraint in optimization models for Transportation Supply Design with equilibrium assignment (see Cantarella et al., 2012, for an example of large scale applications). The relevance of fixed-point stability analysis for transportation supply design and project appraisal is discussed in Cantarella (2013).

4.7.3 Remarks

Some papers follow a continuous time approach to day-to-day dynamics modelling. They (do not address this issue or) state that it can often be convenient for obtaining analytical-theoretical results, which might be easier to establish in the continuous-time case; still this is not case for relevant modelling approaches. On the other hand, for numerical solution time discretisation is necessary; indeed even if the model were originally specified in continuous-time, it needs to be discretised for computational purposes at least. Furthermore, the properties of the resulting DP models can be rather different in the two cases, some qualitative phenomena evident in one and not in another (see Cantarella and Watling, 2016). From a general point of view, continuous time DP models cannot be considered an effective approximation of discrete time ones as much as like a circle generally is not a good approximation of a square or a rectangle.

Quite often, but not always, continuous time DP models are based on Wardrop (1952) or Deterministic Utility route choice behaviour. This approach to route choice modelling could be rather effective when only two routes are available between each O-D pair, and the choice updating filter is specified by a switching model, since there is no need of any route choice function. Indeed, when more than two routes are available, a conditional route choice function should be applied to model route choice behaviour of users who decide to reconsider their yesterday choice. As noted in the previous chapters Deterministic Utility choice modelling approach shows several drawbacks, such as several path flow patterns may correspond to the one link flow pattern, …; indeed, models based on this approach are classified as link-based or path-based depending on the variables used to specify the models; this distinction is meaningless if applied to more general choice modelling approaches.

DP models should not be confused with CUE solution algorithms, however similar they may look. Indeed, in a DP model parameters are to be calibrated against real data and convergence to a fixed-point state equivalent to CUE may not be guaranteed, as shown above through some simple examples. On the other hand, CUE solution algorithms are specified in order to converge to CUE anyway.

Appendix A: Dissipativeness of DP-MA/ES (adapted from Cantarella and Watling, 2016)

The dissipativeness analysis of the DP-MA/ES model (4.18, 4.19) can be carried out by specify an equivalent DP model with respect to i-route flows, see Section 2.3.

The i-route cost (block vector) function can be specified by combining together equations (2.23) and (2.24) with arc cost function (4.2):

$$\widetilde{\mathbf{w}} = \widetilde{\mathbf{w}}\left(\widetilde{\mathbf{h}}\right) \tag{A1.1}$$

where the i-th block is given by $\widetilde{\mathbf{w}}_i\left(\widetilde{\mathbf{h}}_i\right) = \mathbf{B}_i'\mathbf{T}\cdot\mathbf{c}\left(\mathbf{B}_i'\cdot\widetilde{\mathbf{h}}_i + \mathbf{f}_Z'; \boldsymbol{\kappa}\right) + \widetilde{\mathbf{w}}_{\mathbf{Z}i}$.

The i-route flow (block vector) function can be specified by combining together equations (2.25), (2.26) and (2.27):

$$\widetilde{\mathbf{h}} = \widetilde{\mathbf{h}}(\widetilde{\mathbf{w}}) \tag{A1.2}$$

where the i-th block is given by $\widetilde{\mathbf{h}}_i(\widetilde{\mathbf{w}}) = d_i\widetilde{\mathbf{p}}_i(-\psi_i\widetilde{\mathbf{w}}_i; \boldsymbol{\theta}_i)$.

Thus an equivalent formulation of the the DP-MA/ES model (4.18, 4.19) with respect to iro flows is given by:

$$^{1}\widetilde{\mathbf{h}}^{k} = \alpha\widetilde{\mathbf{h}}\left(\sum\nolimits_{j=1,\ldots,\mu}\zeta_j\widetilde{\mathbf{w}}\left({}^{j}\widetilde{\mathbf{h}}^{k-1};\boldsymbol{\kappa}\right);\mathbf{d}\right) + (1-\alpha)\,^{1}\widetilde{\mathbf{h}}^{k-1} \quad \forall k \in \mathbb{N}, k > \mu \tag{A1.3}$$

$$^{j}\widetilde{\mathbf{h}}k = {}^{j}\text{-}1\widetilde{\mathbf{h}}k - 1 \quad j = 2,\ldots,\mu \tag{A1.4}$$

The DP (A1.3, A1.4) is formally similar to the DP (4.18, 4.19) and shares with it the Jacobian matrix structure with reference to the μ matrices $\widetilde{\mathbf{G}}_j = \widetilde{\mathbf{J}}_{\mathbf{h}} \cdot \widetilde{\mathbf{J}}_{\mathbf{w}}$ generally different since computed in different points, with $\widetilde{\mathbf{J}}_{\mathbf{h}} = \nabla\widetilde{\mathbf{h}}(\widetilde{\mathbf{w}})$ and $\widetilde{\mathbf{J}}_{\mathbf{w}} = \nabla\widetilde{\mathbf{w}}\left(\widetilde{\mathbf{h}}\right)$, thus:

$$\delta\left(\widetilde{\mathbf{G}}_{\mu}\right) = \delta\left(\widetilde{\mathbf{J}}_{\mathbf{h}}\right)\delta\left(\widetilde{\mathbf{J}}_{\mathbf{w}}\right) \quad j = 1,\ldots,\mu \tag{A1.5}$$

If for each o-d pair i, the choice function is strictly (positive) regular (see Definition 2 in Section 2.3), such as invariant strictly positive probabilistic choice functions derived from the RUT, the Jacobian of the i-route flow function is non-singular $\det\left(\widetilde{\mathbf{J}}_{\mathbf{h}}\right) \neq 0$. Moreover, the Jacobian of the i-route cost function is given by: $\widetilde{\mathbf{J}}_{\mathbf{w}} = \mathbf{B}'^{\mathbf{T}}\,\mathbf{J}_{\mathbf{c}}\,\mathbf{B}'$ with $\mathbf{J}_{\mathbf{c}} = \nabla\,\mathbf{c}(\mathbf{f})$, thus:

$$\delta\left(\widetilde{\mathbf{G}}_{\mu}\right) = \delta\left(\widetilde{\mathbf{J}}_{\mathbf{h}}\right)\,\delta\left(\mathbf{B}'^{\mathbf{T}}\mathbf{J}_{\mathbf{c}}\,\mathbf{B}'\right) \tag{A1.6}$$

Assuming that $\mathbf{J}_{\mathbf{c}}$ is a (not necessarily symmetric) positive definite matrix with respect to real vectors, thus it is non singular, $\det(\mathbf{J}_{\mathbf{c}}) \neq 0$, two cases may occur, as discussed below, about the rank $\text{rank}(\mathbf{B}')$ of the $m \times \widetilde{n}$ arc—i-route incidence matrix $\mathbf{B}'$, with blocks $\mathbf{B}'_i$.

- The rank is equal to the number of i-routes, $\text{rank}(\mathbf{B}') = \widetilde{n} \leq m$, and $\widetilde{\mathbf{B}}$ is full rank. In this case, matrix $\widetilde{\mathbf{B}}^{\mathbf{T}}\,\mathbf{J}_{\mathbf{c}}\,\widetilde{\mathbf{B}}$ is a (not necessarily symmetric) positive definite matrix with respect to real vectors, thus it is non singular. Thus, in this case $\delta\left(\widetilde{\mathbf{G}}_{\mu}\right) \neq 0$.

Indeed if $\mathbf{Q}$ is a $n \times n$ (not necessarily symmetric) positive definite matrix with respect to real vectors (but necessarily with respect to complex vectors too):

$$\mathbf{x}^{\mathbf{T}}\,\mathbf{Q}\,\mathbf{x} > 0 \quad \forall \mathbf{x} \neq 0, \mathbf{x} \in \mathbb{R}^n$$

and if $\mathbf{M}$ is a $m \times n$ full rank matrix with $n \leq m$,
then the $n \times n$ matrix $\mathbf{M}^{\mathrm{T}}\,\mathbf{Q}\,\mathbf{M}$ is positive definite matrix with respect to real vectors:

$$\mathbf{y}^{\mathrm{T}}\,\mathbf{M}^{\mathrm{T}}\,\mathbf{Q}\,\mathbf{M}\,\mathbf{y} > 0 \quad \forall \mathbf{y} \neq 0, \mathbf{y} \in \mathbb{R}^n$$

since $\mathbf{M}\;\mathbf{y} > 0 \forall \mathbf{y} \neq 0, \mathbf{y} \in \mathbb{R}^n$.

- The rank is less than the number of i-routes, $\text{rank}(\mathbf{B}') < \widetilde{n}$; in this case, matrix $\mathbf{B}'$ may be expressed as the product of two full rank matrices both with rank $\text{rank}(\mathbf{B}')$ through a rank factorization: $\mathbf{B}' = \mathbf{B_1} \times \mathbf{R}$.

Indeed a $m \times n$ matrix $\mathbf{Q}$ with rank $r \leq \min(m,n)$ contains r linearly independent columns making up the $(m \times r)$ full rank sub-matrix $\mathbf{Q_1}$, hence after some re-arrangements of the column $\mathbf{Q} = [\mathbf{Q_1} | \mathbf{Q_2}]$, with the $(m \times (n-r))$ matrix $\mathbf{Q_2}$ containing the $(n-r)$ linearly dependent columns, thus for a suitable $(r \times (n-r))$ full rank matrix $\mathbf{A}$, $\mathbf{Q_2} = \mathbf{Q_1} \times \mathbf{A}$.

Therefore, $\mathbf{Q} = [\mathbf{Q_1} \mid \mathbf{Q_2}] = \mathbf{Q_1} \times [\mathbf{I_r} \mid \mathbf{A}]$, where the $(r \times n)$ matrix $\mathbf{R} = [\mathbf{I_r} | \mathbf{A}]$ is full rank.

In this case the DP (A1.3, A1.4) can be reformulated in the space of the i-routes corresponding to the r linearly independent columns of $\mathbf{B}'$ through a linear transformation defined by matrix $\mathbf{R}$. In this space for the reformulated DP it occurs that $\delta\left(\widetilde{\mathbf{G}}_\mu\right) \neq 0$, properly redefining matrix $\widetilde{\mathbf{G}}_\mu$.

Some assumptions about the arc-route incidence matrix are useful to reduce the number of linearly dependent columns (or rows):

1. each arc belongs to at least a route, thus no row is null,
2. each route contains at least one arc, thus no column is null,
3. no pair of routes are equal, thus no pair of columns are equal,
4. no route is properly contained in another route.

All the above assumptions are quite mild and/or reasonable and can easily be accepted. On the other hand two arcs may well have equal rows if the share all routes, as it occurs for instance for two arc in series or in parallel.

Once in a properly defined space the matrix $\widetilde{\mathbf{G}}_\mu$ is non-singular the dissipativeness analysis can be carried out. If the partial derivatives in matrix $\widetilde{\mathbf{G}}_\mu$ are well-defined, say finite and continuous, the absolute value of the determinant of matrix $\widetilde{\mathbf{G}}_\mu, \delta\left(\widetilde{\mathbf{G}}_\mu\right)$, is a continuous function defined over a compact set, thus $\delta\left(\widetilde{\mathbf{G}}_\mu\right)$ has an upper bound δ_{MAX} (and a lower bound, as well), so that $\alpha\, \zeta_\mu\, \delta\left(\widetilde{\mathbf{G}}_\mu\right) \leq \alpha\, \zeta_\mu\, \delta_{MAX}$, thus $\alpha\, \zeta_\mu\, \delta_{MAX} < 1$ implies $\delta(\mathbf{J}') < 1$. Value of δ_{MAX} cannot easily computed, an approximation may be obtained through matrix norms. It is worth noting that the entries of matrix $\widetilde{\mathbf{G}}_\mu$ as well as δ_{MAX} have no dimension, thus the above condition is not affected by the units used to measure flows or costs.

It can easily demonstrated that ζ_μ is decreasing with μ, and $\lim_{\mu \to \infty} \zeta_\mu = 0$ whichever the value of $\beta \in]\ 0.5,\ 1.0\ [$ is, thus there always exists a large enough memory depth μ^* such that for any memory deeper than this value μ^* the DP is *dissipative* from any initial state. The minimum memory depth value μ^* is defined by:

$$\mu^* = \min\left\{\mu : \zeta_\mu < 1/(\alpha\, \delta_{MAX})\right\}$$

Appendix B: Local stability conditions for of DP-ES/ES

Below some general considerations are reported about the combination of the general stability conditions $\text{MAX}_j\ |\lambda_j^*| < 1$ or $|\lambda_j^*|^2 < 1, \forall j = 1,\ldots,2m$ with either of the two solutions (roots) λ_1 and λ_2 of a reduced quadratic equation:

$$\lambda^2 - p\lambda + q = 0 \quad \text{(A2.1)}$$

$$\lambda_1 = p/2 + \Delta^{1/2} \quad \text{(A2.1a)}$$

$$\lambda_2 = p/2 - \Delta^{1/2} \quad \text{(A2.1b)}$$

$$\Delta = p^2/4 - q \quad \text{(A2.1c)}$$

where Δ is the discriminant of Eq. (A2.1).

Remark. The linear coefficient, p, is equal to the sum of the two solutions, $\lambda_1 + \lambda_2 = p$, and the constant coefficient, q, is equal to the product of the two solutions, $\lambda_1\ \lambda_2 = q$.

Remark. If the linear and/or the constant coefficient is null, $p = 0$ and/or $q = 0$, the Eq. (A2.1) can easily be solved without using Eqs (A2.1a), (A2.1b), (A2.1c), thus in the following these coefficients are assumed non null, $p \neq 0$ and $q \neq 0$.

- If both p and q are real the discriminant Δ is real, and

Δ is real non-negative, $\Delta \geq 0$, then $\Delta^{1/2}$ is real and non-negative, $\Delta^{1/2} \geq 0$;

in this case λ_1 and λ_2 are both real with $\lambda_1 \geq \lambda_2$; thus the stability conditions becomes

$$1 > \lambda_1 \geq \lambda_2 > -1, \text{thus:}$$

$$1 > p/2 + \Delta^{1/2} \geq p/2 - \Delta^{1/2} > -1$$

or

$$\Delta^{1/2} < (1 - p/2)$$

$$\Delta^{1/2} < (1 + p/2)$$

If both right sides are non-negative, $(1 - p/2) > 0$ and $(1 + p/2) > 0$, or

$$-2 < p < 2 \quad \text{(A2.2a)}$$

squaring both of the above conditions do not affect the inequality thus the stability conditions become:

$$\Delta < (1 - p/2)^2$$

$$\Delta < (1 + p/2)^2$$

Therefore the stability conditions are equivalent to (A2.2a) together with:

$$-(1+q) < p < 1+q \tag{A2.2b}$$

Depending on the value of q either of (A2.2a) and (A2.2b) dominates the other.

- If both p and q are real the discriminant Δ is real, and

Δ is real negative, $\Delta < 0$, then $\Delta^{½}$ is imaginary $\Delta^{½} = i\,(-\Delta)^{½}$;
in this case λ_1 and λ_2 occur in a complex conjugate pair such that $\lambda_1 = p/2 + i\,(-\Delta)^{½}$
and $\lambda_2 = \lambda_1 = p/2 - i\,(-\Delta)^{½}$; thus the stability conditions becomes $|\lambda_1|^2 = |\lambda_2|^2 < 1$,
thus:

$$p^2/4 - \Delta = q < 1 \tag{A2.3}$$

with $p^2/4 - \Delta = q \geq 0$ in any case.

- If either p or q or both are complex the discriminant Δ is complex, $\Delta = (a + i\,b)$

in this case λ_1 and λ_2 are both complex, but do not occur in a conjugate pair; thus the stability conditions becomes:

$$|\lambda_1|^2 < 1 \tag{A2.4a}$$

$$|\lambda_2|^2 < 1 \tag{A2.4b}$$

The above stability conditions (A2.2) or (A2.3) turn out special cases of (A2.4). To apply (A2.4). It is useful to remember that the square root of the discriminant is given by:

$$\Delta^{½} = (a + ib)^{½} = \left(\left((a^2+b^2)^{½} + a\right)/2\right)^{½} + i\,(|b|/b)\left(\left((a^2+b^2)^{½} - a\right)/2\right)^{½}$$

where $(a^2+b^2)^{½}$, $((a^2+b^2)^{½}+a)/2)^{½}$, $(((a^2+b^2)^{½}-a)^{½}$ are *positive* square roots, and $(|b|/b)$ is the sign of b.

Local stability conditions for of DP-ES/ES

In the following the solutions of Eq. (4.31) relative to DP-ES/ES are combined with the stability conditions to get the stability conditions (4.33) for a DP-ES/ES model.

As proved in the main text, for each eigenvalue ω_j of matrix **G** two eigenvalues λ_j and λ_{n+j} of matrix **J** can be obtained from each eigenvalue ω_j of matrix **G**, as solutions of the reduced quadratic equation (4.31). It is reported below omitting superscript to simplify notations:

$$\lambda^2 - ((1-\alpha) + (1-\beta) + \alpha\beta\omega)\,\lambda + ((1-\alpha)\,(1-\beta)) = 0 \tag{4.31}$$

with $p = (1-\alpha) + (1-\beta) + \alpha\ \beta\ \omega$, $q = (1-\alpha)(1-\beta) \in\ ]\ 0,\ 1\ [$

$$\Delta^{½} = ((1-\alpha) + (1-\beta) + \alpha\ \beta\ \omega)^2/4 - ((1-\alpha)\,(1-\beta))$$

If the linear coefficient is zero, $p=0$, that is $\omega = -((1-\alpha)+(1-\beta))/(\alpha\,\beta)$, the two solutions are pure imaginary, $\lambda_j = i\,((1-\alpha)\ (1-\beta))^{½}$ and $\lambda_{n+j} = -i\ ((1-\alpha)\ (1-\beta))^{½}$, in this case the stability condition $|\lambda_j|^2 = |\lambda_{n+j}|^2 < 1$ become $((1-\alpha)\ \ (1-\beta)) < 1$, thus the stability conditions is always satisfied. Therefore, below the linear coefficient is assumed different from zero, $p \neq 0$.

A **real eigenvalue** ω_j may lead to two real eigenvalues $\lambda_j \geq \lambda_{n+j}$; in this case the two eigenvalues λ_j and λ_{n+j} have the same sign since $\lambda_j\,\lambda_{m+j} = (1-\alpha)\,(1-\beta) > 0$, and at least one satisfies the stability condition $\mathrm{MAX}_j|\ \lambda_j| < 1$ since $\lambda_j\ \lambda_{m+j} = (1-\alpha)$ $(1-\beta) < 1$; moreover they are both positive if $p > 0$, otherwise they are both negative.

Since $q = (1-\alpha)\,(1-\beta) < 1$, $1+q < 2$, thus conditions (A2.2b) dominates (A2.2a) and describe the stability conditions:

$$-(1+q) < p < 1+q \tag{A2.2b}$$

After some algebra we get the stability conditions below:

$$-\omega_o(\alpha,\beta) < \omega < 1 \quad \forall j = 1,\ldots,m$$

where: $\omega_o(\alpha,\beta) = 1+2\ \ ((1-\alpha)+(1-\beta))/(\alpha\ \ \beta) \geq 1$, as in (4.34) in the main text.

A real eigenvalue ω_j may also lead to a complex conjugate pair of eigenvalues λ_j and λ_{m+j} with $|\ \lambda_j| = |\ \lambda_{m+j}|$; in this case both the eigenvalues λ_j and λ_{n+j} always satisfy the stability condition (A2.4) since $\lambda_j\ \ \lambda_{m+j} = |\ \ \lambda_j|^2 = |\ \ \lambda_{m+j}|^2 = (1-\alpha)^2$ $(1-\beta)^2 < 1$. Therefore, above stability conditions are equivalent to 'Local stability condition for arc cost functions with symmetric Jacobian' (4.35).

A **complex conjugate pair of eigenvalues** $\omega_j = \omega_R + i\,\omega_I$ and $\omega_{j+1} = \omega_R - i\,\omega_I$, with $\omega_I > 0$, leads to four two complex eigenvalues λ_j, λ_{m+j}, λ_{j+1}, and λ_{m+j+1} that must occur in two conjugate pairs since they are complex eigenvalues λ of a real matrix**:**

λ_j and λ_{j+1}, with $|\lambda_j| = |\lambda_{j+1}|$

λ_{m+j} and λ_{m+j+1} with $|\lambda_{m+j}| = |\lambda_{m+j+1}|$

The general stability conditions become:

$$(\lambda_{Rj})^2 + (\lambda_{Ij})^2 < 1 \tag{A2.5a}$$

$$(\lambda_{Rm+j})^2 + (\lambda_{Im+j})^2 < 1 \tag{A2.5b}$$

Since $|\ \lambda_j| = |\lambda_{j+1}|$ and $|\ \lambda_{n+j}| = |\lambda_{n+j+1}|$ with no loss of generality in the following $\omega_j = \omega_R + i\,\omega_I$ with $\omega_I > 0$ is considered only. At least one of two eigenvalues λ_j and $\lambda_{n+j'}$ satisfies the stability conditions (A2.4) since $|\lambda_j|^2\ \ |\lambda_{n+j}|^2 =$ $(1-\alpha)^4\ \ (1-\beta)^4 < 1$.

Generally the discriminant is complex, $\Delta = \mathrm{a} + i\,\mathrm{b}$, after the eigenvalue ω, let

$\mathrm{a} = \tau^2 - (\alpha\ \beta\ \omega_I)^2 - 4\ (1-\alpha)\ (1-\beta)$ be the real part of it,
$\mathrm{b} = 2\ \ \alpha\,\beta\,\omega_I \geq 0$ be the imaginary part of it, non-negative as commented above.

Thus, from the expression of the square root of a complex number $(\mathrm{a} + i\,\mathrm{b})$ with $\mathrm{b} \geq 0$, the *positive* square root of the discriminant is given by:

$$\Delta^{1/2} = (a + i\,b)^{1/2} = \left(\left((a^2+b^2)^{1/2}+a\right)/2\right)^{1/2} + i\left(\left((a^2+b^2)^{1/2}-a\right)/2\right)^{1/2}.$$

Thus:

$$\lambda_{1,2} = \left(\tau + i\,\alpha\,\beta\,\omega_I \pm \left(\left((a^2+b^2)^{1/2}+a\right)/2\right)^{1/2} + i\left(\left((a^2+b^2)^{1/2}-a\right)/2\right)^{1/2}\right)/2$$

$$= \left(\tau \pm \left(\left((a^2+b^2)^{1/2}+a\right)/2\right)^{1/2}\right)/2 + i\left(\alpha\,\beta\,\omega_I \pm \left(\left((a^2+b^2)^{1/2}-a\right)/2\right)^{1/2}\right)/2$$

and

$$|\lambda_{1,2}|^2 = \left(\tau \pm \left(\left((a^2+b^2)^{1/2}+a\right)/2\right)^{1/2}\right)^2/4 + \left(\alpha\,\beta\,\omega_I \pm \left(\left((a^2+b^2)^{1/2}-a\right)/2\right)^{1/2}\right)^2/4$$

$$= 0.25\left(\tau^2 \pm 2\,\tau\left(\left((a^2+b^2)^{1/2}+a\right)/2\right)^{1/2} + \left(\left((a^2+b^2)^{1/2}+a\right)/2\right)\right)$$

$$+ 0.25\left((\alpha\,\beta\,\omega_I)^2 \pm 2\,\alpha\,\beta\,\omega_I\left(\left((a^2+b^2)^{1/2}-a\right)/2\right)^{1/2} + \left(\left((a^2+b^2)^{1/2}-a\right)/2\right)\right)$$

$$= 0.25\,(\tau^2 + (\alpha\,\beta\,\omega_I)^2 + (a^2+b^2)^{1/2}$$

$$\pm 0.50\left(\tau\left(\left((a^2+b^2)^{1/2}+a\right)/2\right)^{1/2} + \alpha\,\beta\,\omega_I\left(\left((a^2+b^2)^{1/2}-a\right)/2\right)^{1/2}\right))$$

Therefore the stability conditions (A2.4), that is $|\lambda_{1,\,2}|^2 < 1$, are equivalent to the following two conditions, depending on the sign of the left-side term:

$$\pm 2\left(\tau\left(\left((a^2+b^2)^{1/2}+a\right)/2\right)^{1/2} + \alpha\,\beta\,\omega_I\left(\left((a^2+b^2)^{1/2}-a\right)/2\right)^{1/2}\right) < 4 - \left(\tau^2 + (\alpha\,\beta\,\omega_I)^2 + (a^2+b^2)^{1/2}\right) \quad \text{(A2.6)}$$

Conditions (A2.6) with positive left-side term can only be satisfied if the right-hand side is positive:

$$(a^2+b^2)^{1/2} < 4 - \left(\tau^2 + (\alpha\,\beta\,\omega_I)^2\right) \quad \text{(A2.7)}$$

On the other hand, if condition (A2.7) is satisfied, condition (A2.6) with a negative left-side term becomes redundant, thus conditions (A2.7) can be restated as:

$$2\left|\tau\left(\left((a^2+b^2)^{1/2}+a\right)/2\right)^{1/2} + \alpha\,\beta\,\omega_I\left(\left((a^2+b^2)^{1/2}-a\right)/2\right)^{1/2}\right| < 4 - \left(\tau^2 + (\alpha\,\beta\,\omega_I)^2 + (a^2+b^2)^{1/2}\right) \quad \text{(A2.8)}$$

Since both the left-side and the right-side terms are positive condition (A2.8) can be restated squaring both sides:

$$4\left(\tau\left(\left((a^2+b^2)^{1/2}+a\right)/2\right)^{1/2} + \alpha\,\beta\,\omega_I\left(\left((a^2+b^2)^{1/2}-a\right)/2\right)^{1/2}\right)^2 < \left(4 - \left(\tau^2 + (\alpha\,\beta\,\omega_I)^2 + (a^2+b^2)^{1/2}\right)\right)^2 \quad \text{(A2.9)}$$

therefore the stability conditions (A2.4) are equivalent to (A2.7) and (A2.9).

Since $(((a^2+b^2)^{½})^2-a^2)^{½}=(a^2+b^2-a^2)^{½}=(b^2)^{½}=b=2\,\tau\,\alpha\,\beta\,\omega_I$ (remembering that $b\geq 0$), after some algebra on either term, condition (A.2.9) becomes:

$$2\left(\tau^2-(\alpha\beta\omega_I)^2\right)a+2\left(\tau^2+(\alpha\beta\omega_I)^2\right)\left(a^2+b^2\right)^{½}+8\,\tau^2(\alpha\beta\omega_I)^2$$
$$<16+\tau^4+(\alpha\beta\omega_I)^4+\left(a^2+b^2\right)-8\tau^2-8\,(\alpha\beta\omega_I)^2-8\left(a^2+b^2\right)^{½}$$
$$+2\,\tau^2(\alpha\beta\omega_I)^2+2\,(\alpha\beta\omega_I)^2\left(a^2+b^2\right)^{½}+2\,\tau^2\left(a^2+b^2\right)^{½}$$

or

$$8\,\tau^2(\alpha\beta\omega_I)^2+2\,\tau^2\,a-2\,a\,(\alpha\beta\omega_I)^2a$$
$$<16+\tau^4+(\alpha\beta\omega_I)^4-8\,\tau^2-8\,(\alpha\beta\omega_I)^2+2\,\tau^2(\alpha\beta\omega_I)^2-8\left(a^2+b^2\right)^{½}+\left(a^2+b^2\right) \tag{A2.10}$$

Since $\left(a^2+b^2\right)=\tau^4+(\alpha\beta\omega_I)^4+16\,(1-\alpha)^2(1-\beta)^2+2\,\tau^2(\alpha\beta\omega_I)^2$

$$-8\,\tau^2\,(1-\alpha)\,(1-\beta)+8\,(\alpha\beta\omega_I)^2\,(1-\alpha)\,(1-\beta)$$

The above condition becomes:

$$0<16-8\tau^2-8\,(\alpha\beta\omega_I)^2-8\left(a^2+b^2\right)^{½}+16\,(1-\alpha)^2\,(1-\beta)^2$$

Thus, after reordering and dividing by 8 (A.19) becomes:

$$\left(a^2+b^2\right)^{½}<2\left(1+(1-\alpha)^2(1-\beta)^2\right)-\left(\tau^2+(\alpha\beta\omega_I)^2\right) \tag{A2.11}$$

Since $0<\alpha\leq 1$ and $0<\beta\leq 1$ it results that $0\leq 1-\alpha<1$ and $0<1-\beta<1$, thus $0\leq(1-\alpha)^2(1-\beta)^2<1$ and then $1\leq(1+(1-\alpha)^2(1-\beta)^2)<2$, therefore the right-hand side of (A2.11) is less than the right-hand side of (A2.7), and condition (A2.11) dominates condition (A2.7).

Before squaring both sides of (A2.11) in order to eliminate the square root, it is necessary to assure that the right-hand side is non-negative, thus condition (A2.11) is equivalent to the two following conditions:

$$0\leq 2\left(1+(1-\alpha)^2(1-\beta)^2\right)-\left(\tau^2+(\alpha\beta\omega_I)^2\right) \tag{A2.12}$$

$$\left(a^2+b^2\right)<\left(2\left(1+(1-\alpha)^2(1-\beta)^2\right)-\left(\tau^2+(\alpha\beta\omega_I)^2\right)\right)^2 \tag{A2.13}$$

Condition (A.2.12) can be restated as:

$$\tau^2+(\alpha\beta\omega_I)^2\leq(1+(1-\alpha)\,(1-\beta))^2+(1-(1-\alpha)\,(1-\beta))^2 \tag{A2.14}$$

Remembering the expression of (a^2+b^2) condition (A2.13) becomes:

$$-8\tau^2(1-\alpha)\,(1-\beta)+4\tau^2\left(1+(1-\alpha)^2(1-\beta)^2\right)$$
$$+8\,(\alpha\beta\omega_I)^2(1-\alpha)(1-\beta)+4(\alpha\beta\omega_I)^2\left(1+(1-\alpha)^2(1-\beta)^2\right) \tag{A2.13}$$
$$<4\left(1+(1-\alpha)^2(1-\beta)^2\right)^2-16\,(1-\alpha)^2(1-\beta)^2$$

Therefore dividing by 4 condition (A2.13) can be re-stated as:

$$\tau^2(1-(1-\alpha)(1-\beta))^2+(\alpha\beta\omega_I)^2(1+(1-\alpha)(1-\beta))^2 < (1-(1-\alpha)(1-\beta))^2(1+(1-\alpha)(1-\beta))^2 \quad \text{(A2.15)}$$

As already noted, assumption $0<\alpha\leq 1$ and $0<\beta\leq 1$ implies $0\leq 1-\alpha<1$ and $0<1-\beta<1$, thus: $0\leq(1-\alpha)(1-\beta)<1$ and then:

$$0<(1-(1-\alpha)(1-\beta))\leq 1 \text{ and } 1\leq(1+(1-\alpha)(1-\beta))<2$$

$$\text{or } 0<(1-(1-\alpha)(1-\beta))^2\leq 1 \text{ and } 1\leq(1+(1-\alpha)(1-\beta))^2<4$$

Thus, dividing condition (A.2.15) by $(1-(1-\alpha)(1-\beta))^2>0$ leads to:

$$\tau^2+(\alpha\beta\omega_I)^2((1+(1-\alpha)(1-\beta))^2/(1-(1-\alpha)(1-\beta))^2<(1+(1-\alpha)(1-\beta))^2 \quad \text{(A2.16)}$$

Since $((1+(1-\alpha)(1-\beta))^2/(1-(1-\alpha)(1-\beta))^2)\geq 1$, the left-hand side of (A2.16) is not less than the left-hand side of (A2.14), moreover since $(1-(1-\alpha)(1-\beta))^2>0$ the right-hand side of (A2.16) is less than the right-hand side of (A2.14), thus condition (A2.16) dominates condition (A2.14).

As a conclusion, stability condition (A2.2) is equivalent to condition (A2.16), which can be restated as:

$$(\tau/(\alpha\beta))^2/((1+(1-\alpha)(1-\beta))/(\alpha\beta))^2+(\omega_I)^2/((1-(1-\alpha)(1-\beta))/(\alpha\beta))^2<1 \quad \text{(A2.17)}$$

Remembering the expression of τ it results that:

$$\tau/(\alpha\beta)=\omega_R+((1-\alpha)+(1-\beta))/(\alpha\beta)=\omega_R-(1-(1+(1-\alpha)(1-\beta))/(\alpha\beta))$$

Thus stability condition (A2.17) becomes (4.33):

$$\left(\omega_{Rj}*-(1-e_R)\right)^2/{e_R}^2+\left(\omega_{Ij}*\right)^2/{e_I}^2<1 \quad \forall j=1,\ldots,m \quad \text{(4.33a)}$$

where

$$e_R=(1+(1-\alpha)(1-\beta))/(\alpha\beta)\geq 1 \quad \text{(4.33b)}$$

$$e_I=(1-(1-\alpha)(1-\beta))/(\alpha\beta)\leq 1 \quad \text{(4.33c)}$$

Appendix C: DP with today states depending on itself (adapted from Cantarella and Watling, 2016)

In some case it may occurs that today state also depends on itself. This condition may occur for instance when the system state is the result of aggregation and/or averaging over sub-periods of the day k, or at idealised systems which cannot exist in the real-world but act as 'benchmarks', such as idealised traveller information systems where the ITS and/or travellers can see perfectly into the future

$$\mathbf{x}^k = \boldsymbol{\varphi}\left(\mathbf{x}^{k-1}, \mathbf{x}^k\right) \in S \quad \forall k \in \mathbb{N} \tag{A3.1}$$

Implicit function theorem

briefly: if $\mathbf{f}(\mathbf{x}, \mathbf{y})$ is continuously differentiable and the Jacobian matrix $\nabla_{\mathbf{y}}\mathbf{f}(\mathbf{x},\mathbf{y}) = \mathbf{I} - \nabla_{\mathbf{y}}\boldsymbol{\varphi}\ (\mathbf{x},\mathbf{y})$ is invertible, then there exists a unique continuously differentiable function $\mathbf{y} = \mathbf{g}(\mathbf{x})$, from an open set X to an open set Y, such that for any given $\mathbf{x} \in X$, $\mathbf{y} = \mathbf{g}(\mathbf{x}) \in Y$ is a solution in $\mathbf{y}$ to $\mathbf{f}(\mathbf{x},\mathbf{y}) = \mathbf{0}$.

Quite often the following approach may be followed to express the above DP model as a proper Markovian model $\mathbf{x}^k = \boldsymbol{\varphi}(\mathbf{x}^{k-1})$.

The *implicit function method*. Let us express Eq. (A3.1) as $\mathbf{y} = \boldsymbol{\varphi}(\mathbf{x},\mathbf{y})$, assuming $\mathbf{f}(\mathbf{x},\mathbf{y}) = \mathbf{y} - \boldsymbol{\varphi}(\mathbf{x},\mathbf{y})$ it yields $\mathbf{f}(\mathbf{x},\mathbf{y}) = \mathbf{0}$. Thus, if the hypotheses of the implicit function theorem hold then the following equation can be obtained for a properly defined function $\mathbf{g}(\cdot)$:

$$\mathbf{x}^k = \mathbf{g}\left(\mathbf{x}^{k-1}\right) \in S \quad \forall k \in \mathbb{N}$$

where $\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}) = -\left(\nabla_{\mathbf{y}}\mathbf{f}(\mathbf{x},\mathbf{y})|_{\mathbf{y}=\mathbf{g}(\mathbf{x})}\right)^{-1} \cdot \left(\nabla_{\mathbf{x}}\mathbf{f}(\mathbf{x},\mathbf{y})|_{\mathbf{y}=\mathbf{g}(\mathbf{x})}\right)$. This expression of Jacobian is remarkably useful when analysing the evolution over time close to a fixed-point state $\mathbf{x}^* = \mathbf{x}^k = \mathbf{x}^{k-1}$, that is $\mathbf{x}^* = \boldsymbol{\varphi}\ (\mathbf{x}^*,\mathbf{x}^*)$, since it does not require to know function $\mathbf{g}(\cdot)$.

If the transition function in Eq. (A3.1) is separable with respect the two arguments another approach is also available as described below. This approach, applied in Bifulco et al. (2016), can be proved a particular instance of the previous one. Anyhow it is outlined below for comparison's purpose.

Global inverse function theorem

briefly: if $\mathbf{f}(\mathbf{y})$ is continuously differentiable and the Jacobian matrix $\nabla_{\mathbf{y}}\ \mathbf{f}(\mathbf{y}) = \mathbf{I} - \nabla_{\mathbf{y}}\ \boldsymbol{\varphi}_2(\mathbf{y})$ is invertible in an open set Y, then there exists a unique continuously differentiable inverse function $\mathbf{h}(\mathbf{z}) = \mathbf{f}^{-1}(\mathbf{z})$ for $\mathbf{z} \in \mathbf{f}(Y)$, where $\mathbf{f}(Y)$ denotes the image of the set Y.

The *inverse function method*. If the transition function $\boldsymbol{\varphi}\ (\cdot,\cdot)$ in Eq. (A3.1) is separable with respect the two arguments: $\boldsymbol{\varphi}\ (\mathbf{x}^{k-1},\mathbf{x}^k) = \boldsymbol{\varphi}_1(\mathbf{x}^{k-1}) + \boldsymbol{\varphi}_2(\mathbf{x}^k)$, Eq. (A3.1) may be rewritten as:

$$\mathbf{x}^k - \boldsymbol{\varphi}_2\left(\mathbf{x}^k\right) = \boldsymbol{\varphi}_1\left(\mathbf{x}^{k-1}\right) \tag{A3.2}$$

Let us express Eq. (A3.2) as $\mathbf{y} - \boldsymbol{\varphi}_2(\mathbf{y}) = \boldsymbol{\varphi}_1(\mathbf{x})$, assuming $\mathbf{f}(\mathbf{y}) = \mathbf{y} - \boldsymbol{\varphi}_2(\mathbf{y})$ it yields $\mathbf{f}(\mathbf{y}) = \boldsymbol{\varphi}_1(\mathbf{x})$. Thus, if the hypotheses of the global inverse function theorem hold, then the following equation can be obtained for a properly defined function $\mathbf{h}\ (\cdot)$:

$$\mathbf{x}^k = \mathbf{h}\left(\boldsymbol{\varphi}_1\left(\mathbf{x}^{k-1}\right)\right)$$

where, since the Jacobian matrix of function $\mathbf{h}(\mathbf{z})$ is $\nabla_{\mathbf{z}}\mathbf{h}(\mathbf{z}) = (\mathbf{I} - \nabla_{\mathbf{y}}\boldsymbol{\varphi}_2(\mathbf{y})|_{\mathbf{y}=\mathbf{h}(\mathbf{z})})^{-1}$, the Jacobian matrix of function $\mathbf{h}(\boldsymbol{\varphi}_1(\mathbf{x}))$ is $\nabla_{\mathbf{x}}\mathbf{h}(\boldsymbol{\varphi}_1(\mathbf{x})) = (\mathbf{I} - \nabla_{\mathbf{y}}\boldsymbol{\varphi}_2(\mathbf{y})|_{\mathbf{y}=\boldsymbol{\varphi}_1(\mathbf{x})})^{-1} \cdot (\nabla_{\mathbf{x}}\boldsymbol{\varphi}_1(\mathbf{x}))$. [The same result may be obtained by applying the implicit function method expressing Eq. (A3.2) as: $-(\mathbf{y} - \boldsymbol{\varphi}_2(\mathbf{y})) + \boldsymbol{\varphi}_1(\mathbf{x}) = 0$.] This expression of Jacobian is remarkably useful when analysing the evolution over time close to a fixed-point state $\mathbf{x}^* = \mathbf{x}^k = \mathbf{x}^{k-1}$, that is $\mathbf{x}^* - \boldsymbol{\varphi}_2(\mathbf{x}^*) = \boldsymbol{\varphi}_1(\mathbf{x}^*)$, since it does not require to know the inverse function $\mathbf{h}(\cdot)$.

References

Bie, J., Lo, H.K., 2010. Stability and attraction domains of traffic equilibria in a day-to-day dynamical system formulation. Transp. Res. Part B 44 (1), 90–107.

Bifulco, G.N., Cantarella, G.E., Simonelli, F., Velonà, P., 2016. Advanced traveller information systems under recurrent traffic conditions: network equilibrium and stability. Transport. Res. B Method. 73–87.

Cantarella, G.E., 2013. Day-to-day dynamic models for intelligent transportation systems design and appraisal. Transp. Res. C Emerg. Technol. 29, 117–130.

Cantarella, G.E., Cascetta, E., 1995. Dynamic process and equilibrium in transportation networks: towards a unifying theory. Transp. Sci. 29, 305–329.

Cantarella, G.E., Watling, D.P., 2016. Modelling road traffic assignment as a day-to-day dynamic, deterministic process: a unified approach to discrete- and continuous-time models. EURO J. Transp. Logist. 5, 69–98.

Cantarella, G.E., Velonà, P., Vitetta, A., 2012. Signal setting with demand assignment: global optimization with day-to-day dynamic stability constraints. J. Adv. Transp. 46, 254–268.

Cantarella, G.E., Velonà, P., Watling, D.P., 2019. Day-to-day dynamics & equilibrium stability in a two-mode transport system with responsive bus operator strategies. Netw. Spat. Econ. 15 (3), 485–506.

Cascetta, E., 1987. Static and dynamic models of stochastic assignment to transportation networks. In: Szaego, G., Bianco, L., Odoni, A. (Eds.), Flow Control of Congested Networks. Springer Verlag, Berlin.

Cascetta, E., 1989. A stochastic process approach to the analysis of temporal dynamics in transportation networks. Transp. Res. B 23, 1–17.

Cheng, Q., Wang, S., Liu, Z., Yuan, Y., 2019. Surrogate-based simulation optimization approach for day-to-day dynamics model calibration with real data. Transp. Res. C 105, 422–438.

Ferrari, P., 2009. The effect of the competition between cars and trucks on the evolution of the motorway transport system. Transp. Res. C Emerg. Technol. 17 (6), 558–570.

Ferrari, P., 2011. The dynamics of the competition between cars and trucks on motorways. Transp. Res. C Emerg. Technol. 19 (4), 579–592.

Ferrari, P., 2014. The dynamics of modal split for freight transport. Transp. Res. E Logist. Transport. Rev. 70 (1), 163–176.

Ferrari, P., 2015. Dynamic cost functions and freight transport modal split evolution. Transp. Res. E Logist. Transport. Rev. 77, 115–134.

Ferrari, P., 2016. Instability and dynamic cost elasticities in freight transport systems. Transp. Policy 49, 226–233.

Ferrari, P., 2018. Some necessary conditions for the success of innovations in rail freight transport. Transp. Res. A Policy Pract. 118, 747–758.

Han, L., Wang, D.Z.W., Lo, H.K., Zhu, C., Cai, X., 2017. Discrete-time day-to-day dynamic congestion pricing scheme considering multiple equilibria. Transp. Res. Part B 44 (1), 1–16.

Horowitz, J.L., 1984. The stability of stochastic equilibrium in a two-link transportation network. Transport. Res. B 18, 13–28.

Stokey, N., Lucas, R., 1989. Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge, MA.

Sylvester, J.R., 2000. Determinants of block matrices. Math. Gazette 84 (501), 460–467.

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proc. Inst. Civ. Eng. 2 (1), 325–378.

Further reading

Cantarella, G.E., Velonà, P., 2003. Stability analysis of equilibrium patterns in a transportation network. In: Proceedings of the European Transport Conference 2003.

Linghui, H., Wang David, Z.W., Lo, H.K., Zhu, C., Xingju, C., 2017. Discrete-time day-to-day dynamic congestion pricing scheme considering multiple equilibria. Transp. Res. Part B 44 (1), 1–16.

CHAPTER

5

Assignment to congested networks: Day-to-day dynamics—Stochastic processes

Giulio Erberto Cantarella[a], **David Paul Watling**[b]

[a]*University of Salerno, Salerno, Italy*

[b]*Institute for Transport Studies, University of Leeds, Leeds, United Kingdom*

All my days I have longed equally
to travel the right road and
to take my own errant path.
Sigrid Undset

The problem with putting two and two together is that
sometimes you get four, and sometimes you get twenty-two.
Dashiell Hammett, The Thin Man

Outline. This chapter describes a comprehensive modelling approach to day-to-day dynamic assignment to congested networks through discrete-time Markovian stochastic process (SP) models; presented models are consistent with the SEAM modelling framework presented in Chapter 1; contents of this chapter mirrors those of Sections 4.5 and 4.6, in the previous Chapter 4, on general deterministic process models.

As stated in the previous Chapter 4, methods for day-to-day (or inter-periodic) dynamic assignment play a central role in advanced transportation system analysis, since they allow to analyse and forecast equilibrium stability and fluctuations around it, as a result of past events.

This chapter discusses stochastic process (SP) models for day-to-day dynamic assignment to congested transportation networks, a kind of assignment still at research level and not yet implemented in commercial software. They can be considered a sort of generalisation of the deterministic process (DP) models, described in the previous Chapter 4. Indeed in a DP model the system state is described by

Dynamics and Stochasticity in Transportation Systems. https://doi.org/10.1016/B978-0-12-814353-7.00005-4
© 2020 Elsevier Inc. All rights reserved.

deterministic variables, while in a SP one it is described by random variables. Still a SP model for day-to-day dynamic assignment is meaningful under the assumptions of probability theory only, whilst any uncertainty theory can be applied to specify the demand model in a DP model.

Even though exactly one user equilibrium flow and cost patterns exist (cfr Chapter 3) and conditions for its local stability can be stated by embedding it within a deterministic process model (cfr Chapter 4), a full description of day-to-day dynamics can only be obtained through a stochastic process analysis. Indeed, the DP analysis does not allow to analyse transients after demand and/or supply changes, nor to obtain a statistical description of the system state evolution over time, i.e. means, modes, moments and, more generally, frequency distributions, this kind of analysis requiring SP models.

Indeed, a complete day-to-day dynamic analysis cannot be based on DP models only, it also needs the specification and application of stochastic process (DP) models, the only models that can provide full statistical description of the evolution over time of the system state for asymptotic behaviour through invariant probability distribution(s) as well as for transient from an initial state (distribution). On the other hand, a proper day-to-day dynamic analysis we need both kinds of models: DP models mainly provide the basis for fixed-point stability conditions including bifurcation analysis useful to support sensitivity analysis with respect to model parameters, while SP model provide a complete statistical characterization of the limit state distribution as well as tools for transient analysis. … .

As already noted in the previous chapter, day-to-day dynamics ontologically occur over discrete time, while within-day dynamics, discussed in Chapter 6, occur over continuous time, say any instant of time within a day (cfr Introduction). Thus in this chapter we will only discussed discrete-time stochastic process models.

Just few years after Horowitz (1984) first proposed to analyse day-to-day dynamics in transportation networks through DP models, Cascetta (1987, 1989) was the first to apply models derived from the theory of stochastic processes to analyse day-to-day dynamics in transportation systems. Davis and Nihan (1993) discussed some issues of stochastic modelling of assignment. Then, Cantarella and Cascetta (1995) were the first to propose a unifying general theory, based on RUM, encompassing FP models for UE assignment and DP and SP models for day-to-day dynamic assignment to general transportation networks. Since then few other papers have been proposed, such as Hazelton and Watling (2004). Recently Watling and Cantarella (2013) proposed a general framework and Watling and Cantarella (2015) discussed some examples.

In this chapter, we introduce and discuss general SP models for day-to-day dynamic assignment to congested networks, consistent with general DP models discussed in Sections 4.5 and 4.6. They may include most route choice modelling approaches based on the theory of probability (cfr Appendix A to the book). They described under steady-state conditions, but they also apply to any TAN used for within-day dynamics, as discussed in Chapter 6. Presented SP models are consistent with the SEAM modelling framework presented in Chapter 1. Section 5.1 introduces basic equations for general stochastic process models discussed in Section 5.2.

5.1 Basic equations for SP models

This section presents the basic equations for day-to-day dynamic assignment through SP models; again the presented approach can straightforwardly be applied to i-route variables instead of route ones and/or to multi-class assignment as well. All definitions and assumptions introduced in the previous chapters still hold, unless otherwise stated.

As noted above in SP models the state of the system is described by random variables. Each of the main vectors appearing in the SEAM are modelled by a random vector so that the state of the system at day k is described by a realisation of each of them.

For clear notation a random variable is denoted by an italic upper case letter as usually, even though an ambiguity may occur with set notation; a random vector (r.v.) or matrix (r.m.) is denoted by an italic bold lower or upper case, respectively; a realisation of a random variable or vector or matrix is denoted as a deterministic variable or vector or matrix; a discrete or a continuous r.v. or r.m. is described by the joint probability mass (pmf) or the probability density function (pdf), respectively.

The equations introduced for general DP models in Sections 4.5 and 4.6 are used to specify relationships among the following random vectors:

- $\boldsymbol{c}^k \in S_c$ is the $m \times 1$ (column) random vector of *actual arc costs* on day k;
- $\boldsymbol{f}^k \in S_f$ is the $m \times 1$ (column) random vector of *arc flows* on day k, *it has discrete entries* as explained in Section 5.1.1;
- $\boldsymbol{h}^k;_i \in S_{h_i}$ is the $n_i \times 1$ (column) random vector of *route flows* for o-d pair i on day k, *it has integer entries* as explained in Section 5.1.2;
- $\boldsymbol{p}^k;_i \geq \boldsymbol{0}$ is the $n_i \times 1$ (column) random vector of *route choice probabilities* for o-d pair i on day k;
- $\boldsymbol{v}^k;_i$ is the $n_i \times 1$ (column) random vector of *route systematic utilities* for o-d pair i on day k;
- $\boldsymbol{w}^k;_i \in S_{w_i}$ is the $n_i \times 1$ (column) random vector of *actual route costs* for o-d pair i on day k;
- $\boldsymbol{y}^k;_i \in S_{y_i}$ is the $n_i \times 1$ (column) random vector of *forecasted route costs* for o-d pair i on day k.
- $d_i \geq 0$ is the *demand flow* for o-d pair i, *it is assumed integer* as explained in Section 5.1.2.

If any input data, such as demand flows, are to be modelled through random vectors, they have to be included among the state vectors, even if they do not depend on day.

Main vector notations from Chapter 4 as well few new ones used in the following are enlisted below in alphabetical order for reader's convenience (sets come first, then Roman letters, at last Greek letters). Variables, vectors, or matrices that may change over the day have a superscript, usually k.

A is the set of arcs, with $m = |A|$ elements;
$\mathbb{E}^m$ is the set of real $m \times 1$ (column) vectors with Euclidean distance;
m is the number of arcs;
$\mathbb{N}$ is the set of natural numbers, that is positive integers;
$\mathbb{N}_0$ is the set of non-negative integers;
n_i is the number of routes connecting o-d pair i;
R_i is the set of routes for o-d pair i, with $n_i = |R_i|$ elements;
$S_c \subseteq \mathbb{E}^m$ is the arc cost set, given by the convex hull of set $\mathbf{c}(S_f)$, non-empty, compact and convex, since $\mathbf{c}(S_f)$, the co-domain of the arc cost function $\mathbf{c}(\cdot)$, is non-empty and compact for a continuous arc cost function;
$S_f \subseteq \mathbb{E}^m$ is the feasible arc flow set, non-empty, compact since finite, see Section 5.1.1;
$S_{h_i} \subseteq \mathbb{N}^{ni};_0$ is the feasible route flow set for o-d pair i, non-empty, compact since finite see Section 5.1.2;
$S_{w_i} \subseteq \mathbb{E}^{ni}$ is the route cost set for o-d pair i, non-empty, compact and convex since an affine transformation of the arc cost set S_c, see Eq. (5.3a);

$\mathbf{B}_i$ is the $(m \times n_i)$ i-th block of the *ARGIM* for o-d pair i;
$\mathbf{c}^k \in S_c$ is the $m \times 1$ (column) vector of *arc actual costs* on day k;
$\mathbf{c}(\cdot)$ is the $m \times 1$ (column) *arc cost function*;
$\mathbf{f}^k \in S_f$ is the $m \times 1$ (column) vector of *total arc flows* on day k;
$\mathbf{f_Z} \geq \mathbf{0}$ is the $m \times 1$ (column) vector of *other arc flows*;
$\mathbf{f}(\cdot)$ is the $m \times 1$ (column) *arc flow function*;
$\mathbf{h}^k;_i \in S_{h_i}$ is the $n_i \times 1$ (column) vector of *route flows* for o-d pair i on day k;
$\mathbf{h}^k \in S_{h_i}$ is the $n \times 1$ (column) vector of *route flows* for all o-d pairs on day k;
$\mathbf{h}_i(\cdot)$ is the $n_i \times 1$ (column) vector *route flow function* for o-d pair i;
$\mathbf{p}^k;_i \geq \mathbf{0}$ is the $n_i \times 1$ (column) vector of *route choice proportions* with entries $p^k;_{i\ r}$ such that $\mathbf{1}^\mathbf{T}\mathbf{p}^k;_i = 1$;
$\mathbf{p}_i(\cdot)$ is the $n_i \times 1$ (column) vector *route choice function* for o-d pair i;
$\mathbf{S}^k;_i$ be the $n_i \times n_i$ *route transition matrix* with entries $p^k;_{i\ r/j}$; all its entries are non-negative, $\mathbf{S}^k;_i \geq \mathbf{0}$, with column sum equal to 1, $\mathbf{1}^\mathbf{T}\mathbf{S}^k;_i = \mathbf{1}^\mathbf{T}$, thus $\mathbf{S}^k;_i$ is a column stochastic matrix;
$\mathbf{S}_i(\cdot)$ be the $n_i \times n_i$ *route transition matrix function* assumed time-independent;
$\mathbf{v}^k;_i$ is the $n_i \times 1$ i-th block of the (column) vector of *route systematic utility* for o-d pair i on day k;
$\mathbf{w}^k;_i \in S_{w_i}$ is the $n_i \times 1$ i-th block of the (column) vector of *actual route costs* for o-d pair i on day k;
$\mathbf{w}_{\mathbf{Z}i} \geq \mathbf{0}$ is the $n_i \times 1$ (column) vector of *other route costs* for o-d pair i;
$\mathbf{w}_i(\cdot)$ is the $n_i \times 1$ block the (column) vector *route cost function* for o-d pair i;

$\mathbf{x}^k \in S_c$ is the $m \times 1$ (column) vector of *forecasted arc costs* for day k;
$\mathbf{y}^k{}_{;i} \in S_{y_i}$ is the $n_i \times 1$ (column) vector of *forecasted route costs* for o-d pair i on day k;
$\mathbf{y}^k \in S_y$ is the $n \times 1$ (column) vector of *forecasted route costs* for all o-d pairs on day k;

$\alpha \in]0, 1[$ is the *choice updating parameter*;
$\beta \in]0, 1[$ is the *cost updating parameter*;
ζ_j is the weight given to the actual cost occurred in any of the μ previous days, in a moving average filter;
$\boldsymbol{\theta}_i > 0$ is the vector of the *route choice function parameters* for o-d pair i;
$\boldsymbol{\kappa} > 0$ is the $m \times 1$ (column) vector of the *arc capacities*, with entries κ_a;
$\mu > 1$ is the integer *memory depth*, in a moving average filter;
$\psi_i > 0$ is the *utility scale parameter* in the route choice model, for o-d pair i.

Relationships between two random vectors or matrices

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two random vectors with values in sets S_X and S_Y respectively, and $\boldsymbol{\varphi}(\cdot)$ be a vector function from sets S_X and S_Y.

- Relationship in variables (RiV) occurs if $\boldsymbol{y} = \boldsymbol{\varphi}(\boldsymbol{x})$, often referred to as function of a r.v.;
  - if function $\boldsymbol{\varphi}(\cdot)$ is linear, moments—such means, variances, and co-variances—of r.v. $\boldsymbol{y}$ may easily be obtained from those of r.v. $\boldsymbol{y}$;
  - if function $\boldsymbol{\varphi}(\cdot)$ is invertible, or generally if for any given $\mathbf{y} \in S_Y$, equation $\mathbf{y} = \boldsymbol{\varphi}(\mathbf{x})$ has at most a countable number of roots, the pmf/pdf of $\boldsymbol{y}$ may be defined from that of $\boldsymbol{x}$, then moments of r. v. $\boldsymbol{y}$ may easily be defined.
- Relationship in all parameters (RiAP) occurs if all parameters of r.v. $\boldsymbol{y}$ are a function of r.v. $\boldsymbol{x}$; since moments of r.v. $\boldsymbol{y}$ depend on its parameters they may easily be defined. The same relationship can be obtained if (a large enough number of) moments are expressed as functions of r.v. $\boldsymbol{x}$.

Remark. In both cases the uncertainty modelled by r.v. $\boldsymbol{y}$ is the same of that modelled by r.v. $\boldsymbol{x}$, even though their dispersion indices, such as variances, take different values.

- Relationship in some parameters (RiSP) occurs if some but not all parameters of r.v. $\boldsymbol{y}$ are function of r.v. $\boldsymbol{x}$ (this case may only occurs if r.v. $\boldsymbol{y}$ has at least two parameters). A similar relationship is obtained if some moments are expressed as a function of r.v. $\boldsymbol{x}$.

The most common case occurs if the mean E[$\boldsymbol{y}$] only of r.v. $\boldsymbol{y}$ is defined as a function of r.v. $\boldsymbol{x}$, $E[\boldsymbol{y}] = \boldsymbol{\varphi}(\boldsymbol{x})$, and the r.v. $\boldsymbol{y}$ is defined by the sum of its mean and an additional r.v. $\boldsymbol{y}_A$ with null mean, $E[\boldsymbol{y}_A] = \mathbf{0}$, say $\boldsymbol{y} = \boldsymbol{\varphi}(\boldsymbol{x}) + \boldsymbol{y}_A$.

Remark. If the two r.v.'s $\boldsymbol{\varphi}(\boldsymbol{x})$ and $\boldsymbol{y}_A$ are independently distributed, they are uncorrelated thus the uncertainty, as measured by variances for instance, modelled by $\boldsymbol{y}$ is always greater than that of $\boldsymbol{\varphi}(\boldsymbol{x})$ (if $\mathrm{Var}[\boldsymbol{y}_A] > 0$). On the other hand, if the two r.v.'s $\boldsymbol{\varphi}(\boldsymbol{x})$ and $\boldsymbol{y}_A$ are correlated the effect depends on the sign of correlation.

Similar considerations apply for a matrix function of random vectors and/or matrices.

5.1.1 Supply models for SP

Transportation supply models express how user behaviour affects network performances. This section describes the three equations that according to the SEAM framework specify the transportation supply model for the day-to-day dynamics of a transportation system within a stochastic framework. The arc-route flow consistency relation (4.1), the arc cost function (4.2) and the route-arc cost consistency and updating function (4.39) still apply with some extensions to random vectors.

- Stochastic arc-route flow consistency relation

The yesterday arc flows due to all o-d pairs are assumed modelled by a random vector. It can be obtained through an affine transformation from the route space to the arc space defined by the arc-route generalised incidence matrix (cfr Eq. 4.1),

$$\boldsymbol{f}^{k-1} = \sum_i \mathbf{B}_i \cdot \boldsymbol{h}^{k-1};_i + \mathbf{f_Z} \in S_f \quad \forall k \in \mathbb{N} \tag{5.1.1}$$

with $\boldsymbol{f}^0 = \sum_i \mathbf{B}_i \cdot \boldsymbol{h}^0;_i + \mathbf{f_Z} \in S_f$, given $\boldsymbol{h}^0;_\mathrm{i} \in S_{h_i} \forall i$

Thus its stochastic characterisation can easily be obtained. Moreover, the uncertainty modelled by r.v. $\boldsymbol{f}^{k-1}$ is the same of that modelled by r.v.'s $\boldsymbol{h}^{k-1};_i \forall i$. Since all random vectors $\boldsymbol{h}^{k-1};_i$ are integer, $\boldsymbol{f}^{k-1}$ is a discrete random vector as well, and S_f is finite.

According to the more general RiSP modelling approach, some but not all parameters (or moments) of r.v. $\boldsymbol{f}^{k-1}$ may depend on r.v.'s $\boldsymbol{h}^{k-1};_i \forall i$. In a simple instance of this approach the mean $\mathrm{E}[\boldsymbol{f}^{k-1}]$ is defined as a linear function of r.v.'s $\boldsymbol{h}^{k-1};_i \forall i$ through (5.1.1) leading to:

$$\boldsymbol{f}^{k-1} = \sum_i \mathbf{B}_i \cdot \boldsymbol{h}^{k-1};_i + \mathbf{f_Z} + \boldsymbol{f_A} \quad \forall k \in \mathbb{N} \tag{5.1.2}$$

where $\boldsymbol{f_A}$ is the additional arc flow r.v. with $\mathrm{E}[\boldsymbol{f_A}] = \mathbf{0}$, it is assumed independent of r.v.'s $\boldsymbol{h}^{k-1};_i \forall i$, and day-invariant; the r.v. $\boldsymbol{f_A}$ tries to model uncertainty about arc flows due to missing arcs after zoning, route definition, flow composition, lack of information about other flows, … .

- Stochastic arc cost function

Due to congestion, say driving user behaviour, yesterday actual arc costs depend on yesterday arc flows (cfr Eq. 4.2):

$$\boldsymbol{c}^{k-1} = \mathbf{c}(\boldsymbol{f}^{k-1}; \boldsymbol{\kappa}) \in S_c \quad \forall k \in \mathbb{N} \tag{5.2.1}$$

with $\boldsymbol{c}^0 = \mathbf{c}(\boldsymbol{f}^0; \boldsymbol{\kappa}) \in S_c$

Arc cost function are assumed day-invariant. If the arc cost flow function may be assumed strictly monotone, thus invertible, the stochastic characterisation of r.v. $\boldsymbol{c}^{k-1}$ can be obtained from that of $\boldsymbol{f}^{k-1}$. Moreover, the uncertainty modelled by r. v. $\boldsymbol{c}^{k-1}$ is the same of that modelled by r.v. $\boldsymbol{f}^{k-1}$.

According to the more general RiSP modelling approach, some but not all parameters (or moments) of r.v. $\boldsymbol{c}^{k-1}$ may depend on r.v. $\boldsymbol{f}^{k-1}$. In a simple instance of this approach the mean E[$\boldsymbol{c}^{k-1}$] is defined as a function of r.v. $\boldsymbol{f}^{k-1}$ through Eq. (5.2.1) leading to:

$$\boldsymbol{c}^{k-1} = \mathbf{c}\left(\boldsymbol{f}^{k-1}; \boldsymbol{\kappa}\right) + \boldsymbol{c}_{\boldsymbol{A}} \quad \forall k \in \mathbb{N} \tag{5.2.2}$$

where $\boldsymbol{c}_{\boldsymbol{A}}$ is the additional arc cost r.v. with E[$\boldsymbol{c}_{\boldsymbol{A}}$]$=\mathbf{0}$, it is assumed independent of r.v. $\boldsymbol{f}^{k-1}$, and day-invariant; the r.v. $\boldsymbol{c}_{\boldsymbol{A}}$ tries to model uncertainty about arc costs due to weather and lighting conditions and traffic control that may affect capacity, monetary cost heterogeneity and dispersion, … .

- Stochastic route-arc cost consistency and updating function

The yesterday actual route cost r.v. for o-d pair i can be obtained from the yesterday actual arc cost r.v. through an affine transformation (cfr Eq. 4.39a):

$$\boldsymbol{w}^{k-1};_i = \mathbf{B}_i{}^{\mathbf{T}} \cdot \boldsymbol{c}^{k-1} + \mathbf{w}_{\mathbf{Z}i} \in S_{w_i} \quad \forall i \; \forall k \in \mathcal{N} \tag{5.3a}$$

with $\boldsymbol{w}^0;_i = \mathbf{B}_i{}^{\mathbf{T}} \cdot \boldsymbol{c}^0 + \mathbf{w}_{\mathbf{Z}i} \in S_{w_i} \quad \forall i$

Thus its stochastic characterisation can easily be obtained. Moreover, the uncertainty modelled by r.v.'s $\boldsymbol{w}^{k-1};_i \;\; \forall i$. is the same of that modelled by r.v. $\boldsymbol{c}^{k-1}$. (Usually the additional actual route cost r.v.'s are not explicitly considered, since they are defined by the affine transformation of the additional arc cost r.v. if present.)

Today forecasted costs depend on actual or forecasted costs incurred on a finite number of previous days (cfr Eq. 4.39b) through a forecasting filter assumed day-invariant and continuous:

$$\boldsymbol{y}^k;_i = \mathbf{y}_i\left(\boldsymbol{w}^{k-1};_i, \boldsymbol{y}^{k-1};_i, \ldots\right) \in S_{y_i} \quad \forall i \; \forall k \in \mathcal{N} \tag{5.3b}$$

with $\boldsymbol{y}^0;_i = \boldsymbol{w}^0;_i \in S_{w_i} \quad \forall i$

where

S_{y_i} is the set of forecasted route costs for o-d pair i, it assumed compact and independent of the day k, as explained below.

Indeed, the domain of the function specifying the forecasting filter is the product of a finite number of sets, each containing the actual or the forecasting route costs relative to one previous day. If each of these sets is compact their product is a compact set. In this case for a *continuous filter* the function is continuous and its co-domain (or image) $S_{yik}\, \forall i$ is a compact set as well, generally depending on the day k. By induction if all the costs at day 0 belong to a compact set, such as $S_{w_i} \forall i$, the forecasted route set at any day k $S^k;_{yi}\forall i$ is a compact set; moreover the union of the first $k+1$ sets S_{yij} from day $j=0$ up to day $j=k$ is a compact set. Still, the limit as k goes to infinity of the sequence of the unions up to day k may not be compact. In the following it is assumed that this limit set $S_{y_i}\forall i$ exists and is compact. This surely is the case for a *convex filter*, based on a convex combination of past actual or forecasted costs, such as the ES(β) and the MA(β,μ), if all the costs at day 0 belong to a convex set, such as $S_{w_i}\forall i$, thus the co-domain is a convex set.

The route-arc cost consistency and updating function is defined combining together Eqs (5.3a) and (5.3b).

$$\boldsymbol{y}^k;_i = \mathbf{y}_i\left(\mathbf{B}_i{}^{\mathbf{T}} \cdot \boldsymbol{c}^{k-1} + \mathbf{w}_{\mathbf{Z}i}, \boldsymbol{y}^{k-1};_i, \ldots\right) \in S_{y_i} \quad \forall i\ \forall k \in \mathbb{N} \tag{5.3.1}$$

The uncertainty modelled by r.v.'s $\boldsymbol{y}^k;_i$ is the same of that modelled by r.v. $\boldsymbol{c}^{k-1}$ and r.v.'s $\boldsymbol{y}^{k-1};_i\forall i$. The stochastic characterisation of r.v.'s $\boldsymbol{y}^k;_i$ can easily be obtained with linear filters; examples are the ES(β) filter or the MA(β,μ) filter described in Section 4.1.1:

$$\boldsymbol{y}^k;_i = \beta\,\mathbf{B}_i{}^{\mathbf{T}} \cdot \boldsymbol{c}^{k-1} + (1-\beta)\boldsymbol{y}^{k-1};_i \quad \forall k \in \mathbb{N}$$

$$\boldsymbol{y}^k;_i = \sum_{j=1,\ldots,\mu} \zeta_j \mathbf{B}_i{}^{\mathbf{T}} \cdot \boldsymbol{c}^{k-j} \quad \forall i\ \forall k \in \mathbb{N}, k > \mu$$

with $\zeta_j = \beta\,(1-\beta)^{j-1}/(1-(1-\beta)^{\mu}) \geq 0 \quad \forall j = 1,\ 2,\ \ldots,\ \mu$

According to the more general RiSP modelling approach, some but not all parameters (or moments) of r.v.'s $\boldsymbol{y}^k;_i$ may depend on other r.v.'s. In a simple instance of this approach the mean $E[\boldsymbol{y}^k;_i]$ is defined as a function of the other r.v.'s through Eq. (5.3.1) leading to:

$$\boldsymbol{y}^k;_i = \mathbf{y}_i\left(\mathbf{B}_i{}^{\mathbf{T}} \cdot \boldsymbol{c}^{k-1} + \mathbf{w}_{\mathbf{Z}i}, \boldsymbol{y}^{k-1};_i, \ldots\right) + \boldsymbol{y}_{\boldsymbol{A}} \quad \forall i\ \forall k \in \mathbb{N} \tag{5.3.2}$$

where $\boldsymbol{y}_{\boldsymbol{A}}$ is the additional forecasted route cost r.v. with $E[\boldsymbol{y}_{\boldsymbol{A}}] = \mathbf{0}$, it is assumed independent of the other r.v.'s, and day-invariant; the r.v. $\boldsymbol{y}_{\boldsymbol{A}}$ tries to model uncertainty about forecasted route costs due to user heterogeneity, variations of attitude, … .

- Stochastic route cost updating function

Eqs (5.1.1), (5.2.1), and (5.3.1) can be combined to define the general *route cost updating function* (cfr Eq. 4.40):

$$\boldsymbol{y}^k;_i = \mathbf{y}_i\left(\mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}\left(\sum_i \mathbf{B}_i \cdot \boldsymbol{h}^{k-1};_i + \mathbf{f}_{\mathbf{Z}};\boldsymbol{\kappa}\right) + \mathbf{w}_{\mathbf{Z}i}, \boldsymbol{y}^{k-1};_i, \ldots\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{5.4}$$

Eq. (5.4) can be specified through the ES(β) filter or the MA(β,μ) filter:

$$\boldsymbol{y}^k;_i = \beta\,\mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}\left(\sum_i \mathbf{B}_i \cdot \boldsymbol{h}^{k-1};_i + \mathbf{f}_{\mathbf{Z}};\boldsymbol{\kappa}\right) + \mathbf{w}_{\mathbf{Z}} + (1-\beta)\boldsymbol{y}^{k-1};_i \quad \forall i\ \forall k \in \mathbb{N}$$

$$\boldsymbol{y}^k;_i = \sum_{j=1,\ldots,\mu} \zeta_j \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}\left(\sum_i \mathbf{B}_i \cdot \boldsymbol{h}^{k-j};_i + \mathbf{f}_{\mathbf{Z}};\boldsymbol{\kappa}\right) + \mathbf{w}_{\mathbf{Z}} \quad \forall i\ \forall k \in \mathbb{N}, k > \mu$$

with $\zeta_j = \beta(1-\beta)^{j-1}/(1-(1-\beta)^{\mu}) \geq 0 \quad \forall j = 1, 2, \ldots, \mu$

- Stochastic arc cost updating function

The very same route cost updating function (5.4) is obtained by first computing forecasted arc costs by applying any of the above linear cost updating filters to arc costs:

$$\boldsymbol{x}^k = \beta\, \mathbf{c}\left(\boldsymbol{f}^{k-1}; \boldsymbol{\kappa}\right) + (1-\beta)\boldsymbol{x}^{k-1} \quad \forall k \in \mathbb{N}$$

$$\boldsymbol{x}^k = \sum_{j=1,\ldots,\mu} \zeta_j \mathbf{c}\left(\boldsymbol{f}^{k-j}; \boldsymbol{\kappa}\right) \quad \forall k \in \mathbb{N}, k > \mu$$

with $\zeta_j = \beta(1-\beta)^{j-1}/(1-(1-\beta)^{\mu}) \geq 0 \quad \forall j = 1, 2, \ldots, \mu$

In this case either of the above equations are used instead of Eq. (5.2.1) or (5.2.2), with forecasted route cost r.v.'s defined by the following relation (cfr Eq. 4.3.#) to be used instead of Eq. (5.3.#):

$$\boldsymbol{y}^{k-1};_i = \mathbf{B}_i^{\mathbf{T}} \cdot \boldsymbol{x}^{k-1} + \mathbf{w}_{\mathbf{Z}i} \quad \forall i\ \forall k \in \mathbb{N} \tag{5.5}$$

5.1.2 Demand models for SP models

Travel demand models express how network performances affect user choice behaviour. This section describes the three equations that according to the SEAM framework specify the travel demand model for the day-to-day dynamics of a transportation system within a stochastic framework.

The route utility function (4.6) and the general route choice updating equation (4.42) as well as the route-demand flow consistency relation (4.8) still apply with some extensions to random vectors and matrices. In SP models route choice proportions have to be considered choice probabilities obtained by applying any choice modelling theory based on probability, such as the Random Utility Theory referred to in this chapter (as briefly reviewed above, for details see Appendix A to the book).

Random utility theory (RUT) is based on three main choice modelling hypotheses.

A. Perfect rationality hypothesis: each decision-maker, belonging to a group of homogeneous individuals, for instance all users travelling between the same o-d pair i (and belonging to same class):

   A.1 considers all the alternatives in a set of relevant alternatives, called *choice set*, for instance R_i the set of routes available for travelling between o-d pair i;

   A.2 gives each alternative r, for instance a route, a value of *perceived utility*, $U_{i,r}$;

   A.3 chooses an alternative r^* with maximum value of perceived utility, $U_{i,\, r*} \geq U_{i,\, r} \forall r \in R_i$.

   Remark. Hypothesis A is rather unrealistic, but it is mitigated by the following one.

B. Uncertainty hypothesis: the perceived utility of each alternative is modelled taking into account uncertainty regarding the non-complete information available to each user as well as to the modeller.

Continued

Remark. Hypothesis B greatly weaken the above hypothesis A, modelling user errors and heterogeneity as well as unavoidable modelling approximations about availability of alternatives and perceived utility (such as dividing the study area into zones, where a journey starts or ends, and singling out main infrastructures and services to support journeys between any pair of them, cfr Preface).

C. Randomness hypothesis: the perceived utility of each alternative is modelled through a continuous random variable; the mean of the perceived utility is called the systematic utility, $v_{i,r} = E[U_{i,r}]$.

Thus, as for any random variable with finite mean, the random residual can be defined as $\xi_{i,r} = U_{i,r} - E[U_{i,r}] = U_{,r} - v_{i,r}$, with $E[\xi_{i,r}] = 0$ and $Var[\xi_{i,r}] = Var[U_{i,r}]$; the distribution of the random residual can be obtained from that of the perceived utility.

Remark. This hypothesis rules out modelling other kinds of uncertainty not included in Theory of Probability, such as vagueness in Theory of Possibility.

Therefore, the probability $p_{i,r}$ that the decision-maker chooses alternative r is given by the probability that the perceived utility of this alternative is greater than or equal to the perceived utility of any other alternative:

$$p_{i,r} = \Pr\left[U_{i,r} \geq U_{i,j}, \forall j \in R_i\right] = \Pr\left[v_{i,r} - v_{i,j} \geq \xi_{i,j} - \xi_{i,r}, \forall j \in R_i\right]$$

Remark. Since the perceived utility is assumed a continuous random variable the two conditions $U_{i,r*} \geq U_{i,r}$ and $U_{i,r*} > U_{i,r} \forall r \in R_i$ are equivalent and the same results is obtained by the following equation (often found in literature):

$$p_{i,r} = \Pr\left[U_{i,r} > U_{i,j}, \forall j \neq r \in R_i\right] = \Pr\left[v_{i,r} - v_{i,j} > \xi_{i,j} - \xi_{i,r}, \forall j \neq r \in R_i\right]$$

According to RUT it is assumed that each user travelling between o-d pair i associate to each route r in the set of the available routes R_i a perceived utility $U_{i,r}$ modelled as a random variable, let

$\boldsymbol{u}^k_{;i}$ be the $n_i \times 1$ (column) random vector of *perceived utilities* for o-d pair i on day k, with entries U_{irk}; its mean $E[\boldsymbol{u}^k_{;i}]$ is given by the systematic utility vector.

The r.v. $\boldsymbol{u}^k_{;i}$ tries to model several sources of uncertainty about perceived utility from the point of view of the users as well as the modeller (see Section 1.3). Some of them are enlisted below.

- User perception errors: users may take wrong decisions since they wrongly perceived or weight attributes such as travel time or money affecting the set and the utility values of the available options.
- User heterogeneity: aggregation is necessary to keep any model at a manageable level of complexity but it introduces some unavoidable modelling errors:
  - over space, for example during study area delimitation and zoning;
  - over time, for example neglecting difference among days of the week;
  - over type, for example grouping users with respect to class of income, age, education degree.

- Missing attributes: due lack of data modeller may decide to exclude some attributes who affect users' behaviour, for example weather conditions, or may ignore them.
- Attribute measurement errors: attribute measurements may be affected by errors due to for example data collection procedures, different conditions during collection.

In a simplified approach the perceived utility r.v. also model any additional uncertainty about arc flows or costs, or forecasted route costs as in Eqs (5.1.2), (5.2.2), or (5.3.2).

> A **random utility model** (RUM), derived from RUT, is fully specified by two functions:
>
> - the *utility function* between the systematic utility and attributes that can be measured in the current scenario or assumed in a design scenario, for instance the route cost;
> - the *choice function* between the choice probabilities and the systematic utilities and attributes that can be measured in the current scenario or assumed in a design scenario, for instance the route cost; its expression depends on the joint distribution of the perceived utilities.
>
> Parameters of the systematic utility function as well as those of the distribution of the perceived utility can be calibrated through statistical inference applied to a sample of observed choices (disaggregate calibration) and/or on data about user flows (aggregate calibration).
>
> A random utility choice model is defined PROBABILISTIC if the perceived utilities have non-null (finite) variance. (If all the variances are null the deterministic utility choice model is obtained.)
>
> A probabilistic choice model is defined STRICTLY POSITIVE if the choice probability of any alternative is strictly positive, whichever are the systematic utility values.
>
> A probabilistic choice model is defined INVARIANT if the random residual distribution (and the choice set) is independent of the systematic utility. In this case the choice probability vector actually depends only on the differences between the values of systematic utility and any reference value; the choice function is increasing monotone, and if differentiable has a symmetric positive semi-definite (with respect to real vectors) Jacobian matrix.

- Stochastic route utility function

Under the assumptions of the RUT, the systematic utility vector $\boldsymbol{v}^k{}_{;i}$ is the mean of the random vector of perceived utilities $\boldsymbol{u}^k{}_{;i}: \boldsymbol{v}^k{}_{;i} = \mathrm{E}[\boldsymbol{u}^k{}_{;i}], \forall i \quad \forall k \in \mathbb{N}$. It is a r.v. given by a linear transformation of the r.v. of today route forecasted costs (cfr Eq. 4.6):

$$\boldsymbol{v}^k{}_{;i} = -\psi_i \boldsymbol{y}^k{}_{;i} \quad \forall i \ \forall k \in \mathbb{N} \tag{5.6}$$

where $\psi_i > 0$ is the utility scale parameter.

- Stochastic route choice updating function

A general approach to choice updating can be defined through an S filter (cfr Eq. 4.42) or the equivalent ZQ filter described in Section 4.5.2:

$$\boldsymbol{p}^k{}_{;i} = \mathbf{S}_i\left(\boldsymbol{v}^k{}_{;i}, \boldsymbol{w}^{k-1}{}_{;i}, \boldsymbol{y}^{k-1}{}_{;i}, \ldots\right) \cdot \boldsymbol{p}^{k-1}{}_{;i} \quad \forall i \ \forall k \in \mathbb{N} \tag{5.7}$$

where $\boldsymbol{S}^k{}_{;i} = \mathbf{S}_i(\boldsymbol{v}^k{}_{;i}, \boldsymbol{w}^{k-1}{}_{;i}, \boldsymbol{y}^{k-1}{}_{;i}, \ldots)$ is a random matrix since function of r.v.'s.

In the most straightforward approach to choice updating for each o-d pair i the today route choice probabilities depend on today route systematic utility values through any RUM (cfr Eq. 4.7.1):

$$\boldsymbol{p}^k{}_{;i} = \mathbf{p}_i\left(\boldsymbol{v}^k{}_{;i}; \boldsymbol{\theta}_i\right) \quad \forall i \;\; \forall k \in \mathbb{N}$$

In a very simple, approach to choice updating, only some users reconsider yesterday choice (but not necessarily change them), and their route choice behaviour is modelled through any RUM as in Eq. (5.7) while the other users' choice behaviour is modelled by yesterday route probabilities. For each o-d pair i today choice proportions are given by an exponential smoothing filter (cfr Eq. 4.7.2 or 4.47):

$$\boldsymbol{p}^k{}_{;i} = \alpha\, \mathbf{p}_i\left(\boldsymbol{v}^k{}_{;i}; \boldsymbol{\theta}_i\right) + (1-\alpha)\boldsymbol{p}^{k-1}{}_{;i} \quad \forall i \; \forall k \in \mathbb{N}$$

where

$\alpha \in]0, 1[$ is the *choice updating parameter*, that is the probability that a user reconsiders yesterday choice; it is assumed day-invariant, independent of route chosen yesterday, and common to all users (anyhow dispersion among users is somehow modelled through the randomness of perceived utilities).

- Stochastic route-demand flow consistency relation

For each o-d pair i the demand flows $d_i \forall i$ and the route flows $\boldsymbol{h}_i \forall i$ are assumed integer; hence the feasible route flow set S_{h_i} for o-d pair i is discrete and finite (since the route flows are upper bounded by the demand flows), thus is compact (but not convex); it also is non-empty if at least one route connect each o-d pair.

For each o-d pair i there are d_i (indistinct) users who may choose among $n_i = |R_i|$ distinct routes in set R_i, thus the number of feasible route flow vectors $|S_{h_i}|$, say the number of elements in set S_{h_i} is equal to the number of the d_i-multisubsets of $R_i : |S_{h_i}| = (n_i + d_i - 1)!/(d_i)!$ (cfr enumerative combinatorics: case 2.1 of the 12-fold way, or case 7.1 of the 20-fold way).

If users choose routes independently from each other, for each o-d pair i the route flow r.v. can be assumed distributed as a sum of d_i identically and independently distributed (i.i.d.) Categorical r.v.'s (a generalisation of the Bernoulli random variable) with category probabilities given by the choice probabilities $\boldsymbol{p}^k{}_{;i}$ (however specified) that is a multinomial (MN) r.v. (a generalisation of the binomial random variable), with number of categories given by the demand flow d_i and category probabilities given by the route choice probabilities $\boldsymbol{p}^k{}_{;i}$ (RiAP modelling approach):

$$\boldsymbol{h}^k{}_{;i} \sim \mathrm{MN}\left(d_i, \boldsymbol{p}^k{}_{;i}\right) \quad \forall i \; \forall k \in \mathbb{N} \tag{5.8.1}$$

with mean $\mathrm{E}[\boldsymbol{h}^k{}_{;i}] = d_i, \boldsymbol{p}^k{}_{;i}$ (cfr Eq. 4.8) expressing flow conservation.

If for each o-d pair i the route choice updating function is specified through a S filter, as in Eq. (5.7) another specification of the route flow r.v. can be adopted. Let

$H^{k-1}{;_{ij}}$ be an entry of r.v. $\boldsymbol{h}^k{;_i}$, that is the yesterday integer flow of route j, say the number of users who followed it, connecting o-d pair i;

$\boldsymbol{s}^k{;_{ij}} = \mathbf{s}_{\mathrm{ij}}(\boldsymbol{v}^k{;_i}, \boldsymbol{w}^{k-1}{;_i}, \boldsymbol{y}^{k-1}{;_i}, \ldots)$ be the j-th column of the r.m. $\boldsymbol{S}^k{;_{\mathrm{i}}} = \mathbf{S}_{\mathrm{i}}(\boldsymbol{v}^k{;_{\mathrm{i}}}, \boldsymbol{w}^{k-1}{;_{\mathrm{i}}}, \boldsymbol{y}^{k-1}{;_{\mathrm{i}}}, \ldots)$, say the today choice probability (random) vector for users who yesterday chose route j.

For each o-d pair i the route flow r.v. is assumed distributed as a sum of n_i independently distributed MN r.v.'s, one for each route j, with number of categories given by $H^{k-1}{;_{ij}}$ and category probabilities given by $\boldsymbol{s}^k{;_{ij}}$:

$$\boldsymbol{h}^k{;_i} \sim \Sigma_j \mathrm{MN}\left(H^{k-1}{;_{ij}}, \boldsymbol{s}^k{;_{ij}}\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{5.8.2}$$

If the switching choice behaviour is explicitly simulated through a ZQ filter, or any of its particular instance such as the ES(α) filter (5.7), the route flows can be modelled as the sum of the combination of a Bernoulli random variable, for switching choice behaviour, and a categorical random vector for (conditional) route choice behaviour.

- Stochastic route flow updating function

Combining together Eqs (5.6), (5.7) and (5.8.1) leads to the general *route flow updating function*

$$\boldsymbol{h}^k{;_i} \sim \mathrm{MN}\left(d_i, \mathbf{S}_i\left(-\psi_i \boldsymbol{y}^k{;_i}, \boldsymbol{w}^{k-1}{;_i}, \boldsymbol{y}^{k-1}{;_i}, \ldots\right) \cdot \boldsymbol{p}^{k-1}{;_i}\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{5.9}$$

or using Eq. (5.8.2) instead of Eq. (5.8.1) to

$$\boldsymbol{h}^k{;_i} \sim \Sigma_j \mathrm{MN}\left(H^{k-1}{;_{ij}}, \mathbf{s}_{ij}\left(-\psi_i \boldsymbol{y}^k{;_i}, \boldsymbol{w}^{k-1}{;_i}, \boldsymbol{y}^{k-1}{;_i}, \ldots\right)\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{5.10}$$

5.1.3 Arc flow updating function

Within SP modelling approach an explicit arc flow updating function, extension of Eq. (4.10.2), that does not requiring explicit route enumeration cannot be obtained but for very simple cases; thus this topic is not discussed.

5.2 General SP models

The set of six equations (5.1.#), (5.2.#), (5.3.#) and (5.6), (5.7), (5.8.#) defines a discrete time Markovian stochastic process (SP) model with respect to all the six basic variables, describing the stochastic evolution over time of them.

Discrete time Markovian stochastic processes

Let $\boldsymbol{\varphi}(\mathbf{x};\boldsymbol{\theta})$ be a vector function from set S to set $\boldsymbol{\varphi}(S)$, and $\boldsymbol{x}^0, \boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^{k-1}, \boldsymbol{x}^k, \boldsymbol{x}^{k+1}, \ldots \in S$ be a sequence of random vectors defined by any kind of relationship between consecutive r.v.'s $\boldsymbol{x}^k$ and $\boldsymbol{x}^{k-1}, \ldots$ specified through function $\boldsymbol{\varphi}(\mathbf{x},\ldots;\boldsymbol{\theta})$, the sequence is a discrete time Stochastic—or random—process (SP), useful to describe the stochastic evolution over time of a system where the state at time (day) k is described by $\mathbf{x}^k \leftsquigarrow \boldsymbol{x}^k$ a realisation of the random vector $\boldsymbol{x}^k$ with joint probability function (jpf) $\boldsymbol{\phi}^k(\mathbf{x}) = \boldsymbol{\phi}_{\boldsymbol{x}^k}(\mathbf{x})$ on day k, S is the state space, $\boldsymbol{\varphi}(\cdot)$ is the transition function, and $\boldsymbol{\theta}$ are its parameters; the state space S may be discrete or continuous or mixed.

The process is called *Markovian* if today state depends on yesterday state only.

Remark. This condition can be obtained even if today state depends on finite number of previous day states by duly specifying an equivalent process with further state variables.

Remark. A SP may also be specified by an explicit relationship between the jpf $\boldsymbol{\phi}_{\boldsymbol{x}^k}(\mathbf{x})$ of the random vector $\boldsymbol{x}^k$ with respect the jpf $\boldsymbol{\phi}_{\boldsymbol{x}^{k-1}}(\mathbf{x})$ of $\boldsymbol{x}^{k-1}$.

Remark. A stochastic process can be interpreted as a deterministic process in the space (with infinite many dimensions) of the jpf $\boldsymbol{\phi}_{\boldsymbol{x}}(\mathbf{x})$.

Stochastic processes and deterministic processes

To each SP model defined by the state space S and the transition function $\boldsymbol{\varphi}(\cdot)$ a DP model with same state space and transition function can be associated.

Remark. If the transition function is linear the sequence $\mathbf{x}^k\, \forall k \in \mathbb{N}$ given by the associated DP is equal to the sequence of the means $\mathrm{E}[\boldsymbol{x}^k]\, \forall k \in \mathbb{N}$ of the SP model, otherwise these sequences are generally different.

Remark. If the r.v. $\boldsymbol{x}^k$ has the same jpf each day k so that only the parameters of jpf change each day, the SP can be described by a DP over the space of parameters of jpf.

As in Sections 4.2 and 4.6, to further analyse the resulting model it is better to reduce the number of equations and variables. The resulting model can be specified with respect to i-route variables as well (not explicitly reported for brevity's sake) but not to arc variables, apart very particular cases.

5.2.1 General two equation assignment models

Given a cost updating filter specified as a RiV and a choice updating filter specified as a RiAP, the resulting SP models are made by two equations with respect to two vectors, a flow vector and a cost vector, say a two equation assignment models (TEAMs). Let

$$\boldsymbol{w}^{k-1}{}_{;i} = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}\left(\sum_i \mathbf{B}_i \cdot \boldsymbol{h}^{k-1}{}_{;i}; \boldsymbol{\kappa}\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{5.11}$$

$$\boldsymbol{y}^k{;}_i = \mathbf{y}_i\left(\boldsymbol{w}^{k-1}{;}_i, \boldsymbol{y}^{k-1}{;}_i\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{5.12}$$

$$\boldsymbol{h}^k{;}_i \sim \mathrm{MN}\left(d_i, \mathbf{S}_i\left(-\psi_i \boldsymbol{y}^k{;}_i, \boldsymbol{w}^{k-1}{;}_i, \boldsymbol{y}^{k-1}{;}_i, \ldots\right) \cdot \boldsymbol{p}^{k-1}{;}_i\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{5.13}$$

$$\text{or}\ \boldsymbol{h}^k{;}_i \sim \Sigma_j \mathrm{MN}\left(H^{k-1}{;}_{ij}, \mathbf{s}_{ij}\left(-\psi_i \boldsymbol{y}^k{;}_i, \boldsymbol{w}^{k-1}{;}_i, \boldsymbol{y}^{k-1}{;}_i, \ldots\right)\right) \quad \forall i\ \forall k \in \mathbb{N} \tag{5.14}$$

Eq. (5.3.#) has been split into Eqs (5.11)—made up by Eqs (5.1.#) and (5.2.#) and (5.3a)—and (5.12)—say (5.3b)—for better readability; indeed Eq. (5.11) is an auxiliary static equation (yesterday actual costs depend on yesterday flows); thus this model is still considered a TEAM, with two vectors defining the system state. Eq. (5.13) can be specified by Eq. (5.9) or (5.10).

The SP model (5.11), (5.12) and (5.13 or 5.14) can easily be rewritten as a proper Markovian SP, today state only depends on yesterday one, by putting Eq. (5.11) into Eq. (5.12) and both Eqs (5.11) and (5.12) into Eq. (5.13 or 5.14). Let

- $S_y = \cup_i S_{y_i}$ be the set of route cost vectors for all o-d pairs, assumed compact;
- $\boldsymbol{y}^k \in S_y$ be the $n \times 1$ (column) block random vector of *forecasted route costs* for all o-d pairs, with i-th block given by $\boldsymbol{y}^k{;}_i$;
- $S_h = \cup_i S_{h_i}$ be the set of feasible route flow vectors for all o-d pairs, a finite (compact) set with $\Pi_i |S_{h_i}|$ integer elements;
- $\boldsymbol{h}^k \in S_h$ be the $n \times 1$ (column) block random vector of *route flow*s for all o-d pairs on day k, with the i-th block given by $\boldsymbol{h}^k{;}_i$, it has integer entries.

The main state vectors of SP model (5.11), (5.12) and (5.13 or 5.14) are $(\boldsymbol{y}^k, \boldsymbol{h}^k)$, $\boldsymbol{w}^{k-1}{;}_i \forall i$ being auxiliary vectors introduced for readability only; the state space is $S_y \times S_h$, assumed compact; the updating parameters are those in the updating filters; other parameters are demand flows, and any other parameter in choice functions and in the arc cost function.

The SP model (5.11), (5.12) and (5.13 or 5.14) can be applied to define the jpf $\mathbf{\Phi}^k(\mathbf{y}, \mathbf{h}) = \mathbf{\Phi}_{(\boldsymbol{y}^k, \boldsymbol{h}^k)}(\mathbf{y}, \mathbf{h})$ of the today state $(\boldsymbol{y}^k, \boldsymbol{h}^k)$ conditional to the yesterday state space $(\boldsymbol{y}^{k-1}, \boldsymbol{h}^{k-1})$ starting from an initial state $(\boldsymbol{y}^{\mathrm{o}}, \boldsymbol{h}^{\mathrm{o}})$, which might also be defined by two deterministic vectors $(\mathbf{y}^{\mathrm{o}}, \mathbf{h}^{\mathrm{o}})$. Since the forecasted route costs are assumed as a continuous r.v., while the route flows are assumed an integer r.v., the joint probability function $\boldsymbol{\phi}^k(\mathbf{y}, \mathbf{h})$ is a joint probability density function with respect to $\boldsymbol{y}$ and a joint probability mass function with respect to $\boldsymbol{h}$. The marginal probability function of route flow r.v. is given by: $\boldsymbol{\phi}^k(\mathbf{h}) = \int_{Sy} \boldsymbol{\phi}^k(\mathbf{y}, \mathbf{h})\, d\mathbf{y}$.

The SP model (5.11), (5.12) and (5.13 or 5.14) can be generalised to include an additional arc flow r.v., and/or additional arc cost r.v., as well as additional forecasted route costs r.v.'s.

The ES(β) filters are an example of cost updating filters that fit well in the above general SP model. Further considerations are useful for MA(β, μ) filters, as reported in the next subsection.

5.2.2 Simple two equation assignment models with MA cost updating filters

SP models with MA(β, μ) cost updating filters can be put in a Markovian form as shown in the Section 4.2 (see Eqs 4.18 and 4.19); the state is defined by a random vector including the route flows of the last μ days; the state space is the union of μ copies of set S_h. In this case the SP modes is a non-linear Markov chain since the state space is finite with integer elements, as noted above.

Markov chains

A *Markov chain* is a discrete time Markov stochastic process with a n-dimensional finite state space S. The joint probability mass function (pmf) of today state $\boldsymbol{\pi}^k(\boldsymbol{x}) = \boldsymbol{\pi}_{(\boldsymbol{x}^k)}(\boldsymbol{x})$ is defined through a linear transformation of the yesterday joint pmf $\boldsymbol{\pi}^{k-1}(\boldsymbol{x}) = \boldsymbol{\pi}_{(\boldsymbol{x}^{k-1})}(\boldsymbol{x})$: $\boldsymbol{\pi}^k(\boldsymbol{x}) = \mathbf{P} \cdot \boldsymbol{\pi}^{k-1}(\boldsymbol{x})$, where $\mathbf{P}$ is called a Markov matrix, an entry p_{ij} being the probability of jumping from state j to state i.

Each Markov matrix is a $n \times n$ square column stochastic matrix, that is all its entries are non-negative, $\mathbf{P} \geq \mathbf{0}$, with column sum equal to 1, $\mathbf{1}^{\mathbf{T}} \cdot \mathbf{P} = \mathbf{1}^{\mathbf{T}}$; n, the number rows and columns, is equal to the number of distinct states in the finite state space S.

If the matrix $\mathbf{P}$ does not depend on the system state the Markov chain is called linear; otherwise, if it depends on the state $\mathbf{P} = \mathbf{P}(\boldsymbol{x})$ the Markov chain is called non-linear (for some authors the latter is not a Markov chain).

Remark. Different definitions of a Markov chain may be found in literature.

5.2.3 Ergodic sets of stochastic processes

Below some notions about the convergence of a stochastic process are briefly reviewed below to introduce notations and definitions (and to support the unfamiliar reader).

Ergodic sets and stationary joint probability functions of a stochastic process

An *ergodic set* of a SP is a minimal subset of the state space such that there is a null probability of leaving it from a state inside it; *minimal* means that it does not contain any proper subset with this property. (Cfr convergence sets and attractors of a DP.)

An SP may have several ergodic sets. (Cfr convergent DP's.)

A *stationary joint probability function* is associated to each ergodic set, it expresses the probability that the system state belongs to the ergodic set as $k \to \infty$:

$$\boldsymbol{\phi}^*(\boldsymbol{x}) = \lim_{k\to\infty} \boldsymbol{\phi}^k(\boldsymbol{x})$$

Each stationary joint probability function completely defines stationary moments, such as the mean vector and the co-variance matrix.

Remark. If the stochastic process is interpreted as a deterministic process in the space of the jpf $\boldsymbol{\phi}_{\boldsymbol{x}}(\mathbf{x})$, a stationary probability function is to be compared with the definition of fixed-point attractor.

Stationary, ergodic, regular or strongly converging stochastic processes

A stochastic process is called *stationary* if its state space contains at least one ergodic set, thus there exists at least on stationary probability function $\boldsymbol{\phi}^*(\boldsymbol{x})$.

A stationary stochastic process is called *ergodic* if its state space contains only one ergodic set, that is there exists at only one stationary probability function $\boldsymbol{\phi}^*(\boldsymbol{x})$.

An ergodic stochastic process is called *regular* if its probability converges towards the unique stationary probability function, regardless of the initial state (or its distribution).

In this case, the moments of the jpf $\boldsymbol{\phi}^k(\boldsymbol{x})$ converge to the stationary moments of $\boldsymbol{\phi}^*(\boldsymbol{x})$.

Remark. If the stochastic process is interpreted as a deterministic process in the space of the jpf $\boldsymbol{\phi}_{\boldsymbol{x}}(\mathbf{x})$, stationarity, ergodicity and regularity are to be compared with the concepts of existence, uniqueness and global stability of a fixed-point states.

Regularity conditions—Invariant distribution existence and uniqueness conditions

Regularity conditions for a general S stochastic process model (5.11), (5.12) and (5.13 or 5.14) mainly require all involved functions are continuous, and that each matrix $\mathbf{S}_i$ is positive.

Necessary and sufficient regularity conditions for a Markov stochastic process

A necessary and sufficient condition for the regularity of a discrete time Markov SP requires that there exists a finite number of days such that the probability of a transition from any feasible state to any subset of the state space or to its complement is significantly greater than zero. More formally, for a given Markov process, let

$\Pr^m(\mathbf{x}, E)$ be the probability of a transition in m days from state $\boldsymbol{x} \in S$ to the subset $E \subseteq S$.

The Markov process is regular if and only if

$$\exists \varepsilon > 0, m \geq 1: \ \forall E \subseteq S, \ \ \Pr^m(\mathbf{x}, E) \geq \varepsilon \forall \boldsymbol{x} \in S \text{ OR } \ \Pr^m(\mathbf{x}, S - E) \geq \varepsilon \ \forall \mathbf{x} \in S$$

where $S - E$ is the complement of E with respect to S (see Stokey and Lucas, 1989).

Assuming that each O-D pair is connected by at least one path, IF

- each matrix $\mathbf{S}_i = \mathbf{S}_i(-\psi_i\ \mathbf{y}_i, \mathbf{y}_i, \mathbf{y}_i)$ is positive whatever the value of $\mathbf{y}$ in set S_y,
- each function $\mathbf{S}_i(\cdot, \cdot, \cdot)$ is continuous over the compact set $S_y \times S_y \times S_y$,
- the forecasting filter $\mathbf{y}(\cdot, \cdot)$, made up by blocks $\mathbf{y}_i(\cdot, \cdot)$, is continuous over $S_y \times S_y$,
- arc cost-flow functions $\mathbf{c}(\mathbf{f})$ are continuous over set S_f,

THEN the resulting stochastic process is regular.

Indeed, since each matrix $\mathbf{S}_i$ is positive the transition probability to any path flow vector from a given previous day state is positive:

$$\forall \left(\mathbf{y}^{k-1}, \mathbf{h}^{k-1}\right) \in S_y \times S_h \Rightarrow \Pr\left[\mathbf{h}^k \middle| \left(\mathbf{y}^{k-1}, \mathbf{h}^{k-1}\right)\right] > 0 \ \ \forall \mathbf{h}^k \in S_h$$

Since the transition probability function is positive over a compact set, it is positively lower bounded, that is:

$$\exists \varepsilon > 0, \varepsilon \leq 1 : \forall \left(\mathbf{y}^{k-1}, \mathbf{h}^{k-1}\right) \in S_y \times S_h \Rightarrow \Pr\left[\mathbf{h}^k \mid \left(\mathbf{y}^{k-1}, \mathbf{h}^{k-1}\right)\right] \geq \varepsilon \ \forall \mathbf{h}^k \in S_h$$

Moreover, given any previous day state, the corresponding forecasted path cost vector belongs to the set of feasible forecasted path cost vectors:

$$\forall \left(\mathbf{y}^{k-1}, \mathbf{h}^{k-1}\right) \in S_y \times S_h \Rightarrow \ \mathbf{y}^{k-1} = \mathbf{y}\left(\mathbf{y}^{k-1}, \mathbf{h}^{k-1}\right) \in S_y$$

Therefore, for any initial state and any subset of the state space, only two cases are possible corresponding to the required condition:

$$\forall \left(\mathbf{y}^{k-1}, \mathbf{h}^{k-1}\right) \in S_y \times S_h, \quad \forall E \subseteq S_y \times S_h,$$

- $\exists \mathbf{h}^k \in S_h : (\mathbf{y}^k, \mathbf{h}^k) \in E \subseteq S_y \times S_h$

then $\Pr\left[\left(\mathbf{y}^k, \mathbf{h}^k\right) \in E \mid \left(\mathbf{y}^{k-1}, \mathbf{h}^{k-1}\right)\right] \geq \varepsilon$

- $\forall \mathbf{h}^k \in S_h : (\mathbf{y}^k, \mathbf{h}^k) \notin E \Leftrightarrow$

$\Leftrightarrow \forall \mathbf{h}^k \in S_h : \left(\mathbf{y}^k, \mathbf{h}^k\right) \in S_y \times S_h - E \subseteq S_y \times S_h$

then $\Pr\left[\left(\mathbf{y}^k, \mathbf{h}^k\right) \in S_y \times S_h - E \mid \left(\mathbf{y}^{k-1}, \mathbf{h}^{k-1}\right)\right] = 1 \geq \varepsilon$

Remark. The above conditions are adapted from Cantarella and Cascetta (1995) who extended to a general S formulation the conditions proposed by Davis and Nihan (1993) for a simpler formulation.

Remark. The assumption of positive matrices $\mathbf{S}_i \, \forall i$ is satisfied by any strictly positive random utility model, as those usually adopted. This assumption also implies that each matrix $\mathbf{S}_i$ is irreducible, thus map $\mathbf{S}_i(\cdot, \cdot, \cdot)$ is a function.

Remark. The extension to irreducible (but not positive) matrix $\mathbf{S}_i \, \forall i$ is still an open issue.

Remark. The assumption that function $\mathbf{S}_i(\cdot, \cdot, \cdot)$ is defined over a compact, that is closed and bounded, set $S_y \times S_y \times S_y$ requires that forecasted path costs, and thus actual path and link costs cannot tend to infinity even if any link flow is over the link capacity (as already noted for deterministic processes, and for fixed-point models for equilibrium).

Remark. If switching choice behaviour is explicitly simulated following a ZQ formulation the regularity conditions require that the diagonal matrices $\mathbf{Q}_i \forall i$ are non-singular, that is all entries on the main diagonal are strictly positive (as already assumed in Chapter 4), and that matrices $\mathbf{Z}_\mathrm{i} \forall i$ are positive, a condition which again occurs for any strictly positive random utility model.

Remark. The above regularity conditions can almost straightforwardly be applied to other filters depending on several past costs.

Remark. The above regularity conditions can almost straightforwardly be applied to SP models including additional an arc flow r.v., and/or an additional arc cost r.v.

Remark. Regularity of a stochastic process is a weaker property than the existence, uniqueness and stability of a fixed-point of the associated deterministic process; indeed no assumptions about monotonicity of demand and/or supply functions or the structure of their Jacobian matrices are needed to assure regularity of the stochastic process. Moreover the forecasting filter does not need to be homogeneous.

Remark. The above regularity conditions implies convergence with respect to moments of the state space jpf, such as mean vector, co-variance matrix, … too.

Strongly and weakly converging stochastic processes

A SP is called *weakly converging*, if convergence occurs with respect to moments of the state space jpf, such as mean vector, co-variance matrix, … .

From this point of view a regular SP is called *strongly converging*.

A strongly converging SP is also weakly converging, but general the converse is not true. But if the r.v. $\boldsymbol{x}^k$ has the same jpf each day k so that only the jpf parameters change each day, an SP weakly converging with respect to a large enough number of moments to completely define the parameters of the jpf is also strongly converging.

5.2.4 Solution issues and convergence analysis

Any of the above described discrete-time Markovian stochastic process models can be numerically solved, say the probability of each state on each day k can be estimated, by applying Monte Carlo techniques, given an initial state and values of all parameters.

Monte Carlo techniques are used to compute (an unbiased estimate of) the probability that a random variable X gets value x when direct computation is unavailable, such as the probability of each state on each day k of a stochastic process.

Let $\Phi_x(x) = \Pr[X \leq x]$ be the distribution function of a random variable X (if this function cannot be expressed in a closed form, approximation functions can be used).

If X is a continuous random, for each value x that X may assumed, variable $\Phi_x(x)$ is strictly increasing monotone, thus its inverse function $\Phi_{X^{-1}}(\cdot)$ exists. Given a random number p, uniformly distributed over interval [0, 1], a realisation $x \twoheadleftarrow X$ of the random variable X is $\Phi_X{}^{-1}(p)$. Value p can actually be generated with a *Pseudo-Random Number Generator* (PRNG). In literature various procedures have been proposed for PRNG.

If X is a discrete random variable, for each value x_i that X may assumed $\Phi_x(x_i)$ has a jump equal to $p_i = \Pr[X = x_i]$, and a similar procedure may be applied, dividing the interval [0,1] in to as many segments, each long p_i, as the number of values x_i that X may assume.

Solution of stochastic processes

The jpf of the state at each day k can be estimated by averaging several realisations of the process, obtained through Monte Carlo techniques, given an initial state $\mathbf{x}^0 \in S$. Each realisation, say each sequence $\mathbf{x}^0$, $\mathbf{x}^1$, $\mathbf{x}^2$, … , generated this way is often called a *trajectory*.

From the jpf moments, such as means, variances, co-variances, can be computed.

Solution of regular stochastic processes

The stationary jpf $\boldsymbol{\phi}^*(\boldsymbol{x})$ of a regular SP can be estimated from a single realisation of the process, obtained through Monte Carlo techniques, given an initial state $\mathbf{x}^0 \in S$, collecting results after the transient long enough to reach stationarity, say from a day k_0 far enough from the first day $k=1$.

Remark. The end of transient can be checked through duly statistical tests, or through checking the trajectory of the associated DP.

Remark. The statistical characterisation of the transient requires averaging several realisations of the process anyway.

A realisation of any of the a general S stochastic process model (5.11), (5.12) and (5.13 or 5.14) can be computed recursively applying the following steps given initial route flows $\mathbf{h}^0;_\text{i} \in \ S_{h_i} \forall i$, and a proper initialisation of the forecasted route costs $\mathbf{y}^0;_\text{i} \in \ S_{y_i} \forall i$.

- Arc flows

A realisation of the yesterday arc flows is computed from route flows through a deterministic version of Eq. (5.1.1):

$$\mathbf{f}^{k-1} = \sum_i \mathbf{B}_i \cdot \mathbf{h}^{k-1};_i + \mathbf{f_Z} \in S_f \tag{5.15}$$

If additional arc flows $\boldsymbol{f_A}$ are considered as in Eq. (5.1.2), a realisation $\mathbf{f_A} \leftsquigarrow \boldsymbol{f_A}$ of them is computed applying Monte Carlo techniques and added to $\mathbf{f}^{k-1}$.

- Actual arc costs

A realisation of the yesterday actual arc costs is computed from yesterday arc flows through a deterministic version of Eq. (5.2.1):

$$\mathbf{c}^{k-1} = \mathbf{c}\left(\mathbf{f}^{k-1}; \boldsymbol{\kappa}\right) \in S_c \tag{5.16}$$

If additional arc costs $\boldsymbol{c_A}$ are considered as in Eq. (5.2.2), a realisation $\mathbf{c_A} \leftsquigarrow \boldsymbol{c_A}$ of them is computed applying Monte Carlo techniques and added to $\mathbf{c}^{k-1}$.

- Actual route costs

A realisation of the yesterday actual route costs is computed from yesterday arc costs through a deterministic version of Eq. (5.3a):

$$\mathbf{w}^{k-1};_i = \mathbf{B}_i{}^{\mathbf{T}} \cdot \mathbf{c}^{k-1} + \mathbf{w}_{\mathbf{Z}i} \in S_{w_i} \quad \forall i \tag{5.17}$$

- Forecasted route costs

A realisation of the today forecasted route costs is computed from yesterday actual and forecasted costs (and possibly further past costs) through any cost updating filter, say a deterministic version of Eq. (5.3b):

$$\mathbf{y}^k;_i = \mathbf{y}_i\left(\mathbf{w}^{k-1};_i, \mathbf{y}^{k-1};_i, \ldots\right) \in \ S_{y_i} \quad \forall i \tag{5.18}$$

Instances of Eq. (5.17) are the ES(β) and the MA(β, μ) filters:

$$\mathbf{y}^k{}_{;i} = \beta\,\mathbf{w}^{k-1}{}_{;i} + (1-\beta)\,\mathbf{y}^{k-1}{}_{;i} \quad \forall i \tag{5.18.1}$$

$$\mathbf{y}^k{}_{;i} = \sum_{j=1,\ldots,\mu} \zeta_j \mathbf{w}^{k-1}{}_{;i} \quad \forall i, k > \mu \tag{5.18.2}$$

Remark. In fully *aggregate* approaches to memory and learning behaviour users are assumed sharing information to generate a common memory, thus all users have the same forecasted route costs. Any heterogeneity and/or dispersion is modelled through additional forecasted route costs $\boldsymbol{y}_{\boldsymbol{A}}$ or in the randomness of the perceived utility. In fully *disaggregate* approaches to memory and learning behaviour users are assumed to have information about own experience only, thus each user has specific forecasted route costs. In this case the memory of each single user is separately modelled. Real world behaviour is in between the above two limit cases depending on availability of information systems.

- Route choice probabilities

Applying the S choice updating filter, a realisation of the today choice probabilities is computed as a function of today forecasted route costs and possibly other past route costs through the following deterministic version of Eq. (5.7) equation including a deterministic version of the utility function (5.6):

$$\mathbf{p}^k{}_{;i} = \mathbf{S}_i\left(-\psi_i\,\mathbf{y}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}, \ldots\right) \cdot \mathbf{p}^{k-1}{}_{;i} \quad \forall i \tag{5.19}$$

where $\mathbf{p}^{k-1}{}_{;i} = (1/d_i)\,\mathbf{h}^{k-1}{}_{;i}$

An instance of Eq. (5.19) is the ES(α) filter:

$$\mathbf{p}^k{}_{;i} = \alpha\,\mathbf{p}_i\left(-\psi_i\,\mathbf{y}^k{}_{;i};\,\boldsymbol{\theta}_i\right) + (1-\alpha)\,\mathbf{p}^{k-1}{}_{;i} \quad \forall i$$

- Route flows—general method

Once choice probabilities have been computed a realisation of the route flows is computed applying Monte Carlo techniques to Eq. (5.8.1)

$$\mathbf{h}^k{}_{;i}\mathrm{MN}\left(d_i,\ \mathbf{p}^k{}_{;i}\right) \quad \forall i \tag{5.20}$$

A multinomial random vector is obtained by independently repeating several times a Categorical random vector (in the very same way that a Binomial random variable is obtained by independently repeating several times a Bernoulli random variable). Thus for each o-d pair i and for each of the d_i users travelling between o-d pair i a route is associated as a realisation of a Categorial r.v. with category probabilities $\mathbf{p}^k{}_{;i}$.

- Route flows—alternative method for s choice updating filters

According to an alternative method for SP models based on S choice updating filters, first a realisation of the route transition matrix is computed as:

$$\mathbf{S}^k{}_{;i} = \mathbf{S}_{\mathrm{i}}\left(-\psi_i\,\mathbf{y}^k{}_{;i}, \mathbf{w}^{k-1}{}_{;i}, \mathbf{y}^{k-1}{}_{;i}, \ldots\right) \quad \forall i \tag{5.21}$$

then route flows are computed as a realisation of the sum of Multinomial r.v. applying Eq. (5.8.2)

$$\mathbf{h}^{k};_{i} \sum_{j} \mathrm{MN}\left(h^{k-1};_{ij},\ \mathbf{s}^{k};_{ij}\right) \quad \forall i \tag{5.22}$$

- Route flows—alternative method for ES choice updating filters

According to an alternative method for SP models based on ES(α) choice updating filters, first for each o-d pair i and for each of the d_i users travelling between o-d pair i the choice of reconsidering or not reconsidering yesterday choice is defined as a realisation of a Bernoulli random variable with probability α. Then, a route is associated as a realisation of a Categorial r.v. with category probabilities $\mathbf{p}^{k};_{i}$ or $\mathbf{p}^{k-1};_{i} = (1/d_i)\ \mathbf{h}^{k-1};_{i}$ depending of the realisation of the Bernoulli random variable.

Remark. In the above described method Monte Carlo techniques have been applied to route flows only, but these techniques may be applied to other variables.

From a single trajectory after the transient of a regular SP estimates of the jpf function through frequencies can be obtained, as well as of mean, variance, covariances, and the autocorrelogram, say the correlation index between any pair of states in two different days.

With reference to the examples already discussed in Chapters 2–4, Fig. 5.1 shows the trajectories of flow on route 1 from day 105 to day 120 obtained by applying SP and DP-ES/ES with $\alpha = 0.50$, $\beta = 0.60$, dispersion parameter $\theta = 7$, and demand flow $d = 3600$. The trajectory of DP reaches the unique fixed-point state.

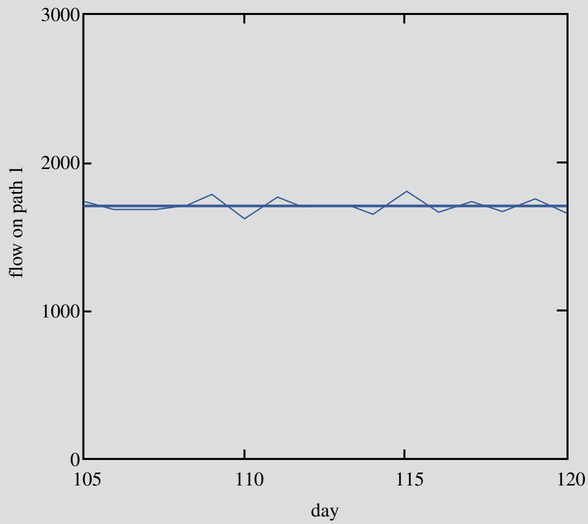


FIG. 5.1

A route flow against day given by SP-ES/ES or DP-ES/ES with $d = 3600$.

Fig. 5.2 shows the trajectories of flow on route 1 from day 75 to day 90 obtained by applying SP and DP-ES/ES with $\alpha = 0.50$, $\beta = 0.60$, dispersion parameter $\theta = 7$, and demand flow $d = 3900$. The trajectory of DP reaches a 2-periodic attractor, since the unique fixed-point state is not stable.

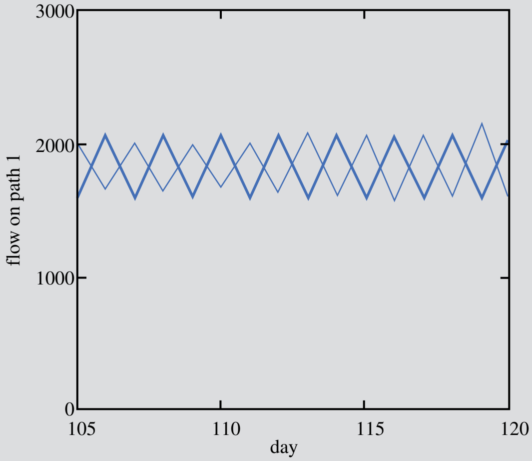


FIG. 5.2

A route flow against day given by SP-ES/ES or DP-ES/ES with $d = 3900$.

In both case SP and DP trajectories are not very different since the number of users, say the demand flow, is rather great.

5.3 Summary

5.3.1 Major findings

This chapter presented several simple and some general discrete-time Markovian stochastic process models, casted within the general SEAM framework, for the day-to-day dynamic assignment to congested transportation networks. As the DP models presented in the previous Chapter 4, SP models are based on a model of user memory and learning and a model of user habit and inertia to change. Presented SP models have been developed under steady-state conditions, but they can be applied to any transportation system with supply modelled by a TAN.

A complete day-to-day dynamic analysis of a transportation systems cannot be based on deterministic process models only, it also needs the specification and application of stochastic process models, the only models that can provide full statistical description of the evolution over time of the system state for asymptotic behaviour through invariant probability distribution(s) as well as for transient from an initial state. Moreover SP models allow disaggregate approaches to users memory and learning behaviour. On the other hand, DP models provide the basis for fixed-point stability and bifurcation analysis useful to support sensitivity analysis with respect to model parameters; DP models may also be used to provide indications about transient length. Therefore a proper day-to-day dynamic analysis needs both kinds of dynamic models. (Some papers seem suggesting that one kind only of dynamic process models is the right tool for day-to-day dynamic analysis, in some cases in favour of SP models in others of DP models, thus missing the useful contribution of the other kind of models.)

The SP models may include route choice functions from any choice modelling theory based on the theory of probability only, such the random utility theory (cfr Appendix A to the book), while DP models can be applied with any choice modelling theory. Moreover DP can be used to specify stability equilibrium constraint for transportation supply design methods.

Generally for non-linear dynamic process models, considering a regular SP model and the associated DP model, if the DP has one fixed-point attractor, it may be considered an approximation of the mean, or the mode, of the stationary jpf of the SP; if two (or more) stable fixed-point attractors exist, they may be considered an approximation of the modes of stationary jpf. Similar considerations hold for periodic and quasi-periodic as well as aperiodic attractors.

Applying a regular SP model there always is a positive probability for the system state jumping from the neighbour of one fixed-point to another one, in other words SP models provide (more realistic) stochastic boundaries between the attraction basins. On the other hand DP models provide only deterministic boundaries between the attraction basin (often called domain) of each fixed-point attractor.

Main open issues about SP models regard extension to random-fuzzy vectors, possibly defined over a fuzzy set.

As already noted for CUE and DP models, the proposed modelling approach can rather easily be extended to assignment with demand flows variable with respect to costs, and/or multi-type or multi-mode assignment, where the choice behaviour among vehicle types or transportation modes is explicitly described by choice models. These extensions are out of the scope of this book (and will possibly be described in a future book on advanced topics). All parameters introduced above are to be calibrated against real data; this relevant issues as well as implementation and application issues are out the scope of this book, mainly focusing on mathematical features.

5.3.2 Further readings

The application of SP models to uncongested vs congested transportation networks is described by Watling and Cantarella (2013). The application of SP model to schedule-based assignment of transit systems is described by Nuzzolo et al. (1999).

For more considerations about DP models as approximation of SP for assignment see Cascetta (1989), Davis and Nihan (1993), Hazelton and Watling (2004), Watling and Cantarella (2015).

5.3.3 Remarks

SP models based on Wardrop or deterministic utility route choice behaviour are inconsistent.

References

Cantarella, G.E., Cascetta, E., 1995. Dynamic process and equilibrium in transportation networks: towards a unifying theory. Transp. Sci. 29, 305–329.

Cascetta, E., 1987. Static and dynamic models of stochastic assignment to transportation networks. In: Szaego, G., Bianco, L., Odoni, A. (Eds.), Flow Control of Congested Networks. Springer Verlag, Berlin.

Cascetta, E., 1989. A stochastic process approach to the analysis of temporal dynamics in transportation networks. Transp. Res. B 23, 1–17.

Davis, G., Nihan, N., 1993. Large population approximations of a general stochastic traffic assignment model. Oper. Res. 41, 169–178.

Hazelton, M.L., Watling, D.P., 2004. Computation of equilibrium distributions of Markov traffic assignment models. Transp. Sci. 38, 331–342.

Horowitz, J.L., 1984. The stability of stochastic equilibrium in a two-link transportation network. Transp. Res. B 18, 13–28.

Nuzzolo, A., Russo, F., Crisalli, U., 1999. A doubly dynamic assignment model for congested urban transit networks. In: Proceedings of 27th European Transportation Forum, Cambridge, England, Seminar F, pp. 185–196.

Stokey, N., Lucas, R., 1989. Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge, MA.

Watling, D.P., Cantarella, G.E., 2013. Modelling sources of variation in transportation systems: theoretical foundations of day-to-day dynamic models. Transportmetrica B 1, 3–32.

Watling, D.P., Cantarella, G.E., 2015. Model representation & decision-making in an ever-changing world: the role of stochastic process models of transportation systems. Netw. Spatial Econ. 15 (3), 843–882.

Further reading

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proc. Inst. Civ. Eng. 2 (1), 325–378.

CHAPTER

6 Assignment to transportation networks: Within-day dynamics

Giulio Erberto Cantarella
University of Salerno, Salerno, Italy

Nothing, of course, begins at the time you think it did.
Lillian Hellman—An Unfinished Woman (1969)

Outline. This chapter proposes a modelling approach to within-day dynamic assignment to transportation networks through models derived from the Traffic Flow Theory.

Methods for within-day (or intra-periodic) dynamic assignment play a central role in advanced transportation system analysis, since they allow to analyse the effects of variability over time of capacities and demand flows, queue formation and dissipation at bottlenecks, real time traffic management systems as well as user while-trip re-routing due to information (or indications) provided by an ATIS or to unexpected traffic conditions.

This chapter discusses deterministic and stochastic process models for within-day dynamic travel demand assignment to a transportation network, a kind of assignment still worth of further research efforts, but some software is available for real case applications under some simplifying assumptions. DP and SP models used for within-day dynamic analysis are properly specified in continuous time (see Chapter 1), even though they need to be discretised over time and possibly space for solution, as in the modelling approach followed in this chapter.

Macroscopic continuous time DP models are based on sets of differential equations derived from macroscopic models of the Traffic Flow Theory (see Appendix B); discretisation over time leads to sets of finite difference equations. As already noted in a mathematical note in Chapter 4, in the resulting discrete-time DP models time might well take real values (for example one tenth of second); these models should not be confused with native discrete time DP models, for which time is integer and increased by 1 at each iteration, as those used for day-to-day dynamic analysis.

Dynamics and Stochasticity in Transportation Systems. https://doi.org/10.1016/B978-0-12-814353-7.00006-6
© 2020 Elsevier Inc. All rights reserved.

Continuous time or event-driven SP models, often called discrete event simulation (DES) methods when coupled with Montecarlo solution techniques, are based on microscopic models of the Traffic Flow Theory (see Appendix B); discretisation over time leads to discrete time or time-driven SP models, in which time might well take real values (for example one tenth of second). These models should not be confused with discrete time SP models for which time is integer and increased by 1 at each iteration, as those used for day-to-day dynamic analysis.

Continuous time or event-driven stochastic process models

In a continuous time or event-driven stochastic process model the system state is updated each time an event (in a discrete set) occurs; the solution approach based on Montecarlo techniques is often called next-event time progression.

Discrete-time or time-driven stochastic process models

In a discrete time or time-driven stochastic process model time is divided into intervals (of same duration) and the system state is updated at the end of each of such intervals, cumulating all changes of state occurred within the interval; the solution approach based on Montecarlo techniques is often called fixed-increment time progression.

Under within-day dynamic conditions, travel time should explicitly be modelled, if different from transportation cost, in order to specify the consistency equations in the supply model and the utility function in the demand models, as shown in Section 6.1.

Indeed, the relation between arc and route flows is highly non-linear since the flow entering an arc at a given time depends on travel time to reach the arc, generally through different routes; the travel time of each of these routes depends on the travel time of each arcs previously traversed, which in turn depend on the flow that has traversed them. Hence, within-day dynamic models for transportation supply analysis are highly non-linear including several feedbacks.

Moreover, demand modelling requires to include departure time choice behaviour through pre-fixed proportions or explicit choice models derived from a Choice Modelling Theory (see Appendix A). Utility functions for modelling departure time choice behaviour usually include a disutility attribute for early/late arrival with respect to the desired time; this attribute depends on the route travel time generally depending on the departure time.

Under within-day dynamic conditions different modelling approaches are usually followed to describe transportation supply with discrete (scheduled) or continuous service systems (see Section 6.1.2). In the former case, indeed, a diachronic discrete TAN can effectively be used to model both space and time (see for instance Nuzzolo, 2009), each node representing a point in space and an instant of time. Thus models for steady-state conditions can almost straightforwardly be applied, as already noted in previous chapters (see also Gentile and Nökel, 2016). On the other hand, generally diachronic networks are not suitable for properly modelling Within-day Dynamics for continuous service systems to avoid a huge number of nodes and arcs, apart some very particular cases (see the recent Watling et al., 2019).

This chapter proposes a general approach for within-day dynamic assignment to continuous service systems consistent with the SEAM modelling framework. Since flows and costs depend on time and on the position within the arc, most modelling approaches are based on continuous networks (see Section 6.1.1) and continuous time. However, space and time discretisation is assumed as in any solution method suitable for real size applications.

Section 6.1 introduces basic equations for within-day dynamic supply and demand models after space and time discretisation; in Section 6.2 within-day assignment models are briefly discussed first for uncongested networks, then for congested network under the assumptions of equilibrium or day-to-day dynamics.

6.1 Basic equations

This section presents the basic equations for within-day dynamic assignment mirroring the approaches already followed under steady-state conditions in previous chapters; all definitions and assumptions introduced in the previous chapters still hold, unless otherwise stated. The proposed modelling approach is consistent with the SEAM framework even if in some cases a basic equation is split into more equations for better readability.

User modelling. In aggregate modelling approaches users are grouped into o-d pairs (and user classes), and the results of their routing behaviour is described by flows and related variables, while in disaggregate modelling approaches each single user is distinguished from the others and the routing behaviour is described by the user trajectory, say the user's position over time.

Space modelling. Space is assumed modelled through a continuous network, so that flows and related variables as well as costs can be attached to any point along an arc, and a vehicle trajectories may be traced by its position over time. However, for solution space discretisation is needed, in this case each link may be described by a single arc, or by several arcs (called sub-arcs, segments, cells, etc.) and variables are attached to arcs only. A variable may be defined with respect the beginning or the end of an arc or to the whole arc. Let

Δx_a be the length of arc a; if a links is described by several arcs, usually each of them has the same length, the same notation is used in any case.

Time modelling. Time is assumed continuous for model specification. However, for solution, time discretisation is needed; thus time is assumed divided in small intervals of same (real) duration, thus all time variables are integer, actually defining the number of intervals. Therefore time appears as a subscript for flow and cost variables. The analysis time period, such as the morning peak period, is made by n_T

intervals. A variable may be defined with respect the beginning or the end of an interval or to the whole interval. Let

Δt be the duration of each time interval.

Main vector notations from previous chapters should be reviewed to take into account within-day dynamics. Route variables common to both the supply and demand models are introduced below. Let

A is the set of arcs, with $m = |A|$ elements;
m is the number of arcs;
n_i is the number of routes available to o-d pair/user i;
n_T is the number of time intervals within the analysis time period;
R_i is the set of routes available to o-d pair/user i, with $n_i = |R_i|$ elements;
i be an o-d pair—that is all the users travelling between the o-d—or a single user for aggregate or disaggregate modelling, respectively;
$h_{isr} \geq 0$ be the route flow moving on route r and departing at the beginning of interval s for o-d pair i/user i; in the latter case $h_{isr} = 1$ if route r is the chosen route $h_{isr} = 1$ otherwise;
$\mathbf{h}_{is} \geq \mathbf{0}$ be the vector of the route flows departing at the beginning of interval s for o-d pair i/user i, with entries h_{isr} $\forall r \in R_i$;
$y_{isr} \geq 0$ be the route travel time along route r departing at the beginning of interval s for o-d pair i/user i;
$\mathbf{y}_{is} \geq \mathbf{0}$ be the vector of the route travel times departing at the beginning of interval s for o-d pair i/user i, with entries y_{isr} $\forall r \in R_i$;
$w_{isr} \geq 0$ be the route transportation cost along route r departing at the beginning of interval s for o-d pair i/user i; this cost may include other disutility attributes beside the travel time, such monetary costs, on-board crowding, etc.;
$\mathbf{w}_{is} \geq \mathbf{0}$ be the vector of the route transportation costs departing at the beginning of interval s for o-d pair i/user i, with entries w_{isr} $\forall r \in R_i$.

6.1.1 Supply models

Transportation supply models express how user behaviour as described by route flows per departure interval affects network performances as described by route travel times and route transportation costs per departure interval. After the introduction of some preliminary assumptions, this section describes the three main relations that according to the SEAM framework specify the transportation supply model for a within day-dynamic transportation system, assuming space and time discretisation.

The proposed modelling approach is rather formal trying to encompassing most of the main approaches to within-day dynamics in a transportation networks available from the Traffic Flow Theory (see Appendix B).

Arc flow modelling approaches

- *aggregate approaches* through mono-dimensional fluid (MDF) approximation; in this case conditions are needed to assure *mono-dimensional fluid* assumption: a particle may never reach (or overtake) a particle who has entered the link before, otherwise the fluid is no longer mono-dimensional; this condition is often referred to in literature as the (MDF) FIFO rule;
- *disaggregate approaches*, through direct representation by tracing the position of each vehicle; in this case FIFO rule is irrelevant.

Arc speed modelling approaches

- *aggregate approaches*, speed and travel time are function of the arc density;
- *disaggregate approaches*, the speed of each single vehicle is modelled as a results of the interaction with the surrounding vehicles.

In *macroscopic models* flow is represented through aggregate variables, so far single vehicles are not explicitly traced, and aggregate speed-density (or other LoS attributes) relations are used, derived from stationary models. Macroscopic models can be:

- continuous in time and space;
- continuous in time and discrete in space;
- discrete in time and space.

For solution models of the first two types are formulated as models of the third type, after space and time discretisation. FIFO rule is a relevant issue for proper specification of macroscopic models; space discretisation may lead to FIFO rule violation. The less coarse the space discretisation, the less significant the FIFO rule violation is.

In *mesoscopic models*: flow is represented through disaggregate variables, by explicitly tracing each single vehicle but aggregate speed-flow relations are used, derived from stationary models. In *microscopic models* flow is represented through disaggregate variables, by explicitly tracing single vehicles, and disaggregate speed modelling is adopted based on explicit modelling of driver behaviour of speed adjustment (through well established models of car following, lane changing, overtaking, gap-acceptance, etc.). FIFO rule is not an issue for mesoscopic or microscopic models.

As already noted, in macroscopic and mesoscopic models, the flow entering an arc at a given time depends on travel time to reach the arc, generally through different routes, the travel time of each of these routes depends on the travel time of each arcs previously traversed, which in turn depend on the flow that has traversed them. This condition leads to a fixed-point between flow and density on one hand and speed and travel time on the other. Time discretisation breaks down this feedback providing an approximated solution to this fixed-point. The less coarse the time discretisation, the less significant the fixed-point violation is.

Main notations for describing within-day dynamics in an arc and between adjacent arcs are introduced below together with main relationships existing among them.

- Arc flow models

- Aggregate modelling is applied for macroscopic models. Let
  m_{INat} be the number of users entering arc a during interval t;
  m_{OUTat} be the number of users exiting arc a during interval t;
  $n_{ONa(t-1)}$ be the number of users on arc a at the end of interval $t-1$, that is at the beginning of interval t;
  n_{ONat} be the number of users on arc a at the end of interval t;
  $k_{a(t-1)} = n_{ONa(t-1)}/\Delta x_a$ be the density on arc a at the beginning of interval t;
  $k_{at} = n_{ONat}/\Delta x_a$ be the density on arc a at the end of interval t;
  $f_{INat} = m_{INat}/\Delta t$ be the flow entering arc a during interval t;
  $f_{OUTat} = m_{OUTat}/\Delta t$ be the flow exiting arc a during interval t.

The following general conservation equation holds:

$$n_{ONat} = n_{ONa(t-1)} + m_{INat} - m_{OUTat} \tag{6.1}$$

or

$$\mathrm{n_{ONat}} = \mathrm{n_{ONa(t-1)}} + (\mathrm{f_{INat}} - \mathrm{f_{OUTat}})\Delta t \tag{6.2}$$

thus

$$k_{at} = (f_{INat} - f_{OUTat})\,\Delta t/\Delta x_a + k_{a(t-1)} \tag{6.3}$$

Remark. Initial density k_{a0}, say density at the beginning of the analysis period can be assumed equal to zero, or given by initial traffic conditions (see pre-load in Section 6.2).

- Disaggregate modelling is applied for mesoscopic and microscopic. Let

  x_{iat} be the position of user i on arc a time at time t.

From the position of each vehicle the above variables m_{INat}, m_{OUTat}, $n_{ONa(t-1)}$, n_{ONat}, can easily be computed, thus flows and densities.

- Arc speed and travel time models
- Aggregate modelling is applied for macroscopic and mesoscopic models. Let

  vv_{at} be the average (space) speed entering arc a during interval t;
  tt_{at} be the average travel time entering arc a during interval t.

The average speed is assumed a function of density at the beginning of interval t after time discretisation:

$$vv_{at} = vv_a\left(k_{a(t-1)}\right) \tag{6.4}$$

and the travel time is given by:

$$tt_{at} = \Delta x_a / vv_{at} \tag{6.5}$$

- Disaggregate modelling is applied for microscopic models. Let

  vv_{iat} be the speed of user i on arc a at time t.

From the speed of each vehicle on arc a at time t the average speed can be computed.

- Arc entering and exiting models

According to aggregate modelling used for macroscopic models after time discretisation, the flow of users f_{OUTat} exiting arc a during interval t is function of the entering flow f_{INat} and the average travel time tt_{at} during interval t:

$$f_{OUTat} = f_{OUTat}(f_{INat}, tt_{at}) \tag{6.6}$$

The detailed expression of this function depends on the macroscopic model actually used. Generally, exit flow also depend on the exiting capacity of arc a.

Remark. If time discretisation is not carried out some feedbacks may occur leading to fixed-point conditions.

- Arc queuing models

Queue formation and dissipation at bottlenecks, such as the approaches of a junction, can explicitly be modelling by dividing the number of users n_{ONat} on a arc a into

- the number of moving users at the speed above defined, and
- the number of queuing users exiting the queue at a rate given by the reciprocal of the service time, depending on the control strategy.

In this case a conservation equation hold separately for moving and for queuing users. Moreover, the travel time is the sum of the running time and the waiting time.

The queue length greatly affects the entering capacity of an arc up to spillback, occurring when the whole arc is occupied by queuing users, and no entering capacity is available. This way queues may spread backward through all the network.

- Network flow propagation models

If arc a precedes one arc a' only the exit flow f_{OUTat} is the entering flow of arc a' during interval t. If a diversion occurs after arc a the exit flow is distributed among several arcs to contribute to the entering flows of these arcs. On the other hand, if a merging occurs before an arc the entering flow of this arc depends on the exit flows of all the merging arcs; all route flows departing during interval time t from arc a have also to be considered. Thus a relationship holds between arc and route flows, defined by the so-called network flow propagation (NFP) models. The specification of the diversion or the merging flow models and of the NFP model depends on the macroscopic model actually used.

Remark. In advanced models the exit flow from an arc a is also constrained by the available entering capacity on the arcs beyond arc a, with respect to the number of users on these arcs. This condition is more relevant if queuing is explicitly modelled.

- Main arc variables and functions

After time discretisation with reference to arc a and interval $(t-1)$, given the density at the beginning of the interval $k_{a(t-2)}$, the set of four equations (Eqs 6.3–6.6) contains five variables: the entering flow $f_{INa(t-1)}$, the exiting flow $f_{OUTa(t-1)}$, the density $k_{a(t-1)}$, the average speed $vv_{a(t-1)}$, and the travel time $tt_{a(t-1)}$. Thus 4 variables may be considered dependent variables, and the entering flow during interval $(t-1) f_{INa(t-1)}$ is the only independent variable, therefore it id the flow associated to each arc a and interval $(t-1)$:

$$f_{a(t-1)} = f_{INa(t-1)} \tag{6.7}$$

For aggregate modelling approaches the average speed vv_{at}, the average travel time tt_{at} and the average transportation cost c_{at} entering arc a during interval t are assumed function of this flow:

$$vv_{at} = vv_a\left(f_{a(t-1)}\right) \tag{6.8}$$

$$tt_{at} = tt_a\left(f_{a(t-1)}\right) \tag{6.9}$$

$$c_{at} = c_a\left(f_{a(t-1)}\right) \tag{6.10}$$

Remark. If time discretisation is not carried out the average travel time tt_{at} and the average transportation cost c_{at} would be function of the arc flow during the same interval f_{at}. In this case fixed-point conditions may occur due to feedbacks among the variables.

The three main equations for supply modelling consistent with the SEAM framework are discussed below; a main equation may be split into more equations for readability.

- Arc-route flow consistency relation

Under the within-day dynamic assumption, after time discretisation, the arc flows during time interval t can be obtained from the route flows departed in any interval up to interval t through an affine transformation from the route space to the arc space, defined by the flow dynamic arc-route generalised incidence matrix:

$$\mathbf{f}_t = \sum\nolimits_{s \le t} \sum\nolimits_i \mathbf{B}_{\mathbf{F}its} \cdot \mathbf{h}_{is} + \mathbf{f}_{\mathbf{Z}t} \tag{6.11}$$

where

i stays for an o-d pair i or a single user i, as defined at the beginning of the section;
$\mathbf{f}_{\mathbf{Z}t} \ge \mathbf{0}$ is the $m \times 1$ (column) vector of arc other flows during interval t;
$\mathbf{h}_{is} \ge \mathbf{0}$ is the $n_i \times 1$ (column) vector of the route flows departing at the beginning of interval s, as defined at the beginning of the section;

$\mathbf{B}_{\mathbf{F}its}$ is the $m \times n_i$ flow dynamic *ARGIM* for intervals t and $s \le t$; each entry $b_{Fitsar} \in [0,1]$ describes how much the flow on route r departed in interval s contributes to the flow on an arc a during interval t; accordingly $\mathbf{B}_{\mathbf{F}its} = \mathbf{0}$ for $s > t$;

$\mathbf{f}_t \ge \mathbf{0}$ is the $m \times 1$ (column) vector of arc flows during interval t, with entries defined according to Eq. (6.7).

The flow dynamic *ARGIM* $\mathbf{B}_{\mathbf{F}its}$ for intervals t and $s \le t$ is a non-linear function of the travel times up to interval t:

$$\mathbf{B}_{\mathbf{F}its} = \mathbf{B}_{\mathbf{F}i}\left(\mathbf{tt}_j\, \mathrm{j} = 1, \ldots, t\right) \tag{6.12}$$

where

$\mathbf{tt}_j$ is the $m \times 1$ (column) vector of the average arc travel times during interval j.

$\mathbf{B}_{\mathbf{F}i}(\cdot)$ is the flow dynamic *ARGIM* function assumed independent of the time intervals.

Remark. Any model expressing the relation between arc and route flows obtained combining together Eqs (6.11), (6.12) is called a NFP model as said above.

- Arc travel time and transportation cost functions

As already stated in Eqs (6.9), (6.10) for a single arc, like in the previous Chapters 3–5, the arc average travel times and transportation costs depend on the arc flows due to congestion, say driving user behaviour:

$$\mathbf{tt}_t = \mathbf{tt}\left(\mathbf{f}_{(t-1)}\right) \ge \mathbf{0} \tag{6.13}$$

$$\mathbf{c}_t = \mathbf{c}\left(\mathbf{f}_{(t-1)}\right) \ge \mathbf{0} \tag{6.14}$$

where

$\mathbf{c}_t$ is the $m \times 1$ (column) vector of the average arc transportation costs during interval t, possibly different from the travel times since they may include other cost attributes;

$\mathbf{tt}(\cdot)$ is the arc travel time function assumed independent of the time intervals;

$\mathbf{c}(\cdot)$ is the arc transportation cost function assumed independent of the time intervals.

Remark. The combination of a NFP model and some travel time functions, Eqs (6.11)–(6.13), is often called a dynamic network loading (DNL) model.

Remark. If time discretisation is not carried out, travel time would be function of the arc flows during the same interval, and the DNL model would be specified by a fixed-point.

- Route-arc cost consistency relation

Under the within-day dynamic assumption, after time discretisation, the travel time $y_{isra} \geq 0$ along route r up to arc $a \in r$ departing at interval s can be obtained through a recursive equation. Let a' and a'' be two consecutive arcs along route r, it yields:

$$y_{isra''} = y_{isra'} + tt_{a't} \text{ where } t = y_{isra'} \tag{6.15a}$$

with $y_{israo} = 0$, a_o being the first arc of route r.

According to the above equation the route travel time y_{isr} along route r departing at interval $s \geq 0$ is given by the travel time up to the last arc of route r:

$$y_{isr} = y_{isra*} \tag{6.15b}$$

$a*$ being the last arc of route r.

From the above Eqs (6.15a), (6.15b) the route travel times departing at interval s can be obtained from the arc travel times during any interval equal to or after s through an affine transformation from the arc space to the route space, defined by the cost dynamic arc-route generalised incidence matrix:

$$\mathbf{y}_{is} = \sum\nolimits_{t \geq s} \mathbf{B}_{\mathbf{C}ist} \cdot \mathbf{tt}_t \forall i \tag{6.16}$$

where

i stays for an o-d pair i or a single user i, as defined at the beginning of the section;
$\mathbf{B}_{\mathbf{C}ist}$ is the $n_i \times m$ cost dynamic *ARGIM* for intervals s and $t \geq s$; each entry $b_{Cistar} \in \{0,1\}$ describes whether the travel time on an arc a during interval t affects the flow on route r departed in interval s or not; accordingly $\mathbf{B}_{\mathbf{C}ist} = \mathbf{0}$ for $t < s$;
$\mathbf{y}_{is} \geq \mathbf{0}$ is the $n_i \times 1$ (column) vector of the route travel times departing at the beginning of interval s, as defined at the beginning of the section.

REMARK. Summation for $t \geq s$ in Eq. (6.13) ends as the last arc of a route is reached or the analysis period ends, whichever occurs first (see post-load in Section 6.2).

Remark. Matrix $\mathbf{B}_{\mathbf{C}ist}$ is not the transpose of matrix $\mathbf{B}_{\mathbf{F}its}$; thus the above equations do not define a TAN (see Chapter 1).

The cost dynamic *ARGIM* $\mathbf{B}_{\mathbf{C}ist}$ for intervals s and $t \geq s$ is a non-linear function of the travel times from interval s:

$$\mathbf{B}_{\mathbf{c}ist} = \mathbf{B}_{\mathbf{c}i}\left(\mathbf{tt}_j \, \mathrm{j} = s, \ldots\right) \tag{6.17}$$

The very same matrix $\mathbf{B}_{\mathbf{c}ist}$ defines the relation that holds between arc and route transportation costs:

$$\mathbf{w}_{is} = \sum\nolimits_{t \geq s} \mathbf{B}_{\mathbf{C}ist} \cdot \mathbf{c}_t + \mathbf{w}_{\mathbf{Z}i} \forall i \tag{6.18}$$

where

$\mathbf{w}_{is} \geq \mathbf{0}$ is the $n_i \times 1$ (column) vector of the route transportation costs departing at the beginning of interval s;

$\mathbf{w}_{\mathbf{Z}is} \geq \mathbf{0}$ is the $n_i \times 1$ (column) vector of the route other transportation costs departing at the beginning of interval s.

- Route travel time and transportation cost functions

Eqs (6.11)–(6.13), (6.16), (6.17) describing the supply model can be combined to define the *route travel time function*:

$$\mathbf{y} = \mathbf{y}(\mathbf{h}) \tag{6.19}$$

where

$h \geq \mathbf{0}$ is the whole vector of the route flows made by blocks $\mathbf{h}_{is}$;
$\mathbf{y} \geq \mathbf{0}$ is the whole vector of the route travel times made by blocks $\mathbf{y}_{is}$.

Adding Eqs (6.14), (6.18) that describe the effects of route flows on route costs, if different from travel times, the *route transportation cost function* is defined:

$$\mathbf{w} = \mathbf{w}(\mathbf{h}) \tag{6.20}$$

where

$\mathbf{w} \geq \mathbf{0}$ is the whole vector of the route transportation costs made by blocks $\mathbf{w}_{is}$.

Remark. The specification of the route transportation cost function requires that the arc travel time functions have already been defined, since they are needed to assure consistency between arc and route variables through matrices $\mathbf{B}_{\mathbf{F}its} = \mathbf{B}_{\mathbf{F}i}(\mathbf{tt}_j \text{ j} = 1, \ldots, t)$ and $\mathbf{B}_{\mathbf{c}ist} = \mathbf{B}_{\mathbf{c}i}(\mathbf{tt}_j \text{ j} = s, \ldots)$.

6.1.2 Demand model

Travel demand models express how network performances as described by route travel times and route transportation costs per departure interval affect user choice behaviour as described by route flows per departure interval. This section describes the three main equations that according to SEAM framework specify the travel demand model for a within day-dynamic transportation system, assuming space and time discretisation; a main equation may be split into more equations for readability.

- Route utility function

The utility function is assumed specified through a linear combination of the route transportation costs and the disutility for early/late arrival with respect a desired arrival time:

$$\mathbf{v}_{is} = -\psi_{1i}\,\mathbf{w}_{is} - \psi_{2i}\,\mathbf{z}_{is} \tag{6.21}$$

where

i stays for an o-d pair i or a single user i, as defined at the beginning of the section;
$z_{isr} \geq 0$ is the early/later arrival disutility departing at interval s and following route r;

$\mathbf{z}_{is} \geq \mathbf{0}$ is the $n_i \times 1$ (column) vector of the early/later arrival disutility departing at interval s, with entries z_{isr};
$\psi_{2i} > 0$ is the utility scale parameter for early/late arrival disutility;
$\mathbf{w}_{is} \geq \mathbf{0}$ is the $n_i \times 1$ (column) vector of the route transportation costs departing at interval s, as already introduced;
$\psi_{1i} > 0$ is the utility scale parameter for transportation costs;
$\mathbf{v}_{is} \leq \mathbf{0}$ is the $n_i \times 1$ (column) vector of the route utilities departing at interval s.

The disutility for early/late arrival with respect a desired arrival time departing at is a function of the difference between the desired and the actual arrival time:

$$z_{isr} = z_i(s\,\Delta t + y_{isr} - dat_i) \geq 0 \tag{6.22}$$

where

dat_i is the desired arrival time with respect to the start of the analysis time period;
$s\,\Delta t$ is the departure time with respect to the start of the analysis time period;
$s\,\Delta t + y_{isr}$ is the arrival time with respect to the start of the analysis time period.

All the above time variables as well as the early/late arrival disutility are real numbers. Some specifications of function (Eq. 6.22) are described in Appendix A.4.

- Route choice function

The departure time and route choice behaviour can be described by applying any discrete choice modelling theory (see Appendix A), thus choice proportions depend on systematic utilities.

In a general modelling approach the choice behaviour occurs over two levels, in the bottom one the choice options are the routes r conditional to a departure interval s, in the top one the choice options are the departure intervals s:

$$\mathbf{p}_{is} = p_{Dis}\,\mathbf{p}_{\mathbf{R}is}(\mathbf{v}_{is};\boldsymbol{\theta}_i) \tag{6.23}$$

where

$\boldsymbol{\theta}_i$ is the route choice function parameter vector;
$\mathbf{p}_{\mathbf{R}is}$ is the $n_i \times 1$ (column) vector of the route proportions conditional to the departure interval s;
p_{Dis} is the choice proportion of departing at interval s;
$\mathbf{p}_{\mathbf{R}is}(\cdot)$ is the route choice function conditional to the departure interval s;
$\mathbf{p}_{is}$ is the $n_i \times 1$ (column) vector of route proportions departing at interval s.

The choice proportion of departing at interval s, p_{Dis}, can be an input data or the result of a choice model. In this case the utility function is a linear transformation of an aggregate value of the route utilities departing at interval s, such as the Expected Maximum Perceived Utility for choice models derived from Random Utility Theory (see Appendix A):

$$v_{Dis} = v_{Di}(\mathbf{v}_{is}) \tag{6.24}$$

where

v_{Dis} is the utility of departing at interval s;
$v_{Di}(\cdot)$ is the departure interval utility function.

Any choice function can be applied to model the departure choice behaviour.

$$\mathbf{p}_{\mathbf{D}i} = \mathbf{p}_{\mathbf{D}i}(\mathbf{v}_{\mathbf{D}i}; \boldsymbol{\theta}_{\mathbf{D}i}) \tag{6.25}$$

where

$\boldsymbol{\theta}_{\mathbf{D}i}$ is the choice function parameter vector;
$\mathbf{v}_{\mathbf{D}i}$ is the $n_T \times 1$ (column) vector of the departure interval utilities, with entries v_{Dis};
$\mathbf{p}_{\mathbf{D}i}$ is the $n_T \times 1$ (column) vector of the departure interval proportions, with entries p_{Dis};
$\mathbf{p}_{\mathbf{D}i}(\cdot)$ is the departure interval choice function.

Remark. In simpler modelling approaches the choice options are the departure time and route pairs (s, r).

Remark. In advanced route choice modelling while-trip re-routing after en-route diversions due to the availability of real-time information can also be considered.

- Route-demand flow consistency relation

Flow conservation for each departure interval s can be expressed as:

$$\mathbf{h}_{is} = d_i\, p_{Dis}\, \mathbf{p}_{\mathbf{R}is} \tag{6.26}$$

It assures that flows of all routes departing at interval s sum up to demand flow $d_i\, p_{Dis}$.

- Route flow function

Eqs (6.21)–(6.26) describing the demand model can be combined to define the *route flow function*:

$$\mathbf{h} = \mathbf{h}(\mathbf{w}, \mathbf{y}) \tag{6.27}$$

As in the steady-case (see Section 6.2) since demand flows are non-negative the route flow function has the same features of the route choice proportion functions.

6.2 Assignment

This section briefly describes how assignment models presented in Chapters 2–5 can be extended to within-day dynamics. Many features of the involved functions do not hold in this case, thus general conditions for fixed-point existence, uniqueness and stability as well as for stationary joint probability functions cannot be stated.

Whichever is the adopted assignment model a so-called pre-load of the network is needed for effective solutions, that is some users are already loaded on the network at the beginning of the analysis time period, otherwise speed values would be unrealistic high. At the end of the analysis time period some users may still be on the network without having reached their destination yet, thus a so-called post-load is generated. The assignment algorithm may further be applied until all users have reached their destination using arc speeds constant over time such the values at the last interval.

6.2.1 Pseudo-dynamics

Pseudo-dynamics is a heuristic approach to within-day dynamic user equilibrium assignment. The whole analysis time period is subdivided into rather large time intervals, for instance 2 h are subdivided into six 20 min sub-intervals. Then demand flows are distributed among these sub-intervals and a user equilibrium assignment is carried out for reach sub-interval; queue lengths, parking use, ... at the end of each sub-interval are used as initial conditions for the next sub-interval. This approach is rather crude, but can be used to provide coarse solutions as indications for more sophisticated approaches.

6.2.2 Uncongested networks

If the network is uncongested (cfr Chapter 2) arc travel times and transportation costs do not depend of arc flows, thus Eqs (6.13), (6.14) are not considered. Therefore, the flow and the cost dynamic *ARGIM's* are fixed, since Eqs (6.12), (6.17) can be used once for all before the application of the model. The resulting network models is both a FAN and a CAN (see Chapter 1); still, as already remarked, either matrix is not the transpose of the other, thus the resulting network is not a TAN (see Chapter 1).

The *within-day dynamic arc flow function* can be defined combining together Eqs (6.11), (6.16), (6.18) from the supply model with the route flow function Eq. (6.27) describing the demand model:

$$\mathbf{f}(\mathbf{c}, \mathbf{tt}) \triangleq \mathbf{B_F} \cdot \mathbf{h}(-\mathbf{B_T} \cdot \mathbf{c}, -\mathbf{B_T} \cdot \mathbf{tt}) + \mathbf{f_Z} \tag{6.28}$$

where vectors $\mathbf{f_Z}$, $\mathbf{tt}$, $\mathbf{c}$, as well as $\mathbf{f}$ and matrices $\mathbf{B_T}$ and $\mathbf{B_F}$ are made up by blocks given by the corresponding vectors and matrices already introduced. Even though Eq. (6.28) is formally similar to Eq. (2.33) in Chapter 2, the within-day dynamic arc flow function is much more complex and hardly shows features similar to the steady-state counterpart.

Applying the within-day dynamic arc flow function, the arc flows over all intervals can be expressed as a function of the arc travel times and transportation costs over all intervals:

$$\mathbf{f} = \mathbf{f_{WD}}(\mathbf{c}, \mathbf{tt}; \mathbf{d}) \tag{6.29}$$

Eq. (6.29) formally describes the within-day dynamic travel demand assignment to an uncongested network. In some cases it can be computed avoiding explicit route enumeration.

6.2.3 Congested networks: User equilibrium—Fixed point models

User Equilibrium assignment searches for mutually consistent arc flows and costs. It can effectively be addressed through fixed-point models as discussed in Chapter 3 under steady state conditions. The set of equations describing the supply and the demand models defines a fixed-point model with respect to all the basic variables.

To further analyse the model it is better to reduce the number of equations and variables with references to Eqs (6.9), (6.10), (6.27) for route variables, and to Eqs (6.13), (6.14), (6.29) for arc variables. Explicit formulations is not reported for brevity's sake.

Quite effective instances of macroscopic models based on arc variables are the link transmission model (LTM) and the cell transmission models (CTM); an instance of mesoscopic model is the TRAFFMED. All these models are described in Appendix B to the book. Many microscopic models exist allowing very detailed modelling of arc flow and speed as well as of other traffic phenomena, some of them are also available as open software; still most of them are specifications of the arc route consistency equation, since route flows are input data, thus they do not provide a complete assignment model, not including a model of the departure interval and route choice behaviour.

6.2.4 Congested networks: Day-to-day dynamics—Dynamic process models

Day-to-day dynamic assignment tries to describe the evolution over days of the costs and the flows in a transportation system. Under steady state conditions it can be addressed by the discrete time deterministic process (DP) models discussed in Chapter 4 under steady state conditions or through discrete time stochastic process (SP) models discussed in Chapter 5.

The supply and demand models described in Section 6.1 can be combined with any of the general cost updating filters and flow updating filters described in Section 4.5. Explicit formulations are not reported since some notations used in Chapters 4 and 5 have different meanings in this chapter due to the limited numbers of available letters.

Resulting models are often referred to in literature as double dynamic assignment models. Cascetta and Cantarella (1991) was one of the first papers addressing this topic proposing a general formal framework for both macroscopic DP and SP models with reference to choice models from Random Utility Theory. However, operative doubly dynamic models are still open research issue.

More recently, operative macroscopic modelling approaches based on DP models have been proposed under some simplifications assumptions; Guo et al. (2018)

review several approaches of this kind based on the Wardrop choice behaviour (see remark at the end of the chapter), and some extensions of it. Some extensions of the mesoscopic TRAFFMED model to DP model have also been proposed.

Microscopic models are better suited to be combined with discrete time SP models, in particular with disaggregate modelling approaches to user memory and habit. No significant microscopic models have been proposed to author's knowledge, perhaps due to the limitation of operative microscopic models already discussed in the previous sub-section on user equilibrium assignment.

6.3 Summary

6.3.1 Major findings

This chapter presented a general modelling framework for within-day dynamic assignment including most of the models from Traffic Flow Theory presented in Appendix B. It is not surprising that no fully consistent unifying general theory is available yet; indeed, conditions have not been stated yet for consistency among the different approaches and with steady state conditions as limit cases. Effective modelling of entering exiting flows from a junction under consistent behavioural hypotheses is another relevant still open research issue.

6.3.2 Further readings

Several references are reported at the end of Appendix B to the book. Some papers by Malachy Carey and other authors proposed requirements for effective macroscopic modelling focusing on the so-called FIFO rule (Bar-Gera and Carey, 2017; Carey et al., 2014a, b).

6.3.3 Remarks

Several macroscopic within-day dynamic assignment modelling approaches are based on Wardrop choice behaviour and extensions of it, leading to variational inequality models of the kind briefly discussed in the remarks at the end of Chapters 2 and 3. The several drawbacks of this choice modelling approach have already been discussed in the remarks at the end of the previous chapters. These drawbacks seem even more relevant in within-day dynamic assignment, since it is really unrealistic assuming that users behave in such a way the used departure time and route pairs have the same disutility.

References

Bar-Gera, H., Carey, M., 2017. Representation requirements for perfect first-in first-out verification in continuous flow dynamic models. Transp. Res. B Methodol. 100, 284–301.

Carey, M., Bar-Gera, H., Watling, D., Balijepalli, C., 2014b. Implementing first-in first-out in the cell transmission model. Transp. Res. B Methodol. 65, 105–118.
Carey, M., Humphreys, P., McHugh, M., McIvor, R., 2014a. Extending travel-time based models for dynamic network loading and assignment, to achieve adherence to first-in-first-out and link capacities. Transp. Res. B Methodol. 65, 90–104.
Cascetta, E., Cantarella, G.E., 1991. A day-to-day and within-day dynamic stochastic assignment model. Transp. Res. A: Gen. 25 (5), 277–291.
Gentile, G., Nökel, K., 2016. Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems: COST Action TU1004 (TransITS). Springer.
Guo, R., Yang, H., Huang, H., 2018. Are we really solving the dynamic traffic equilibrium problem with a departure time choice? Transp. Sci. 52 (3), 603–620.
Nuzzolo, A., 2009. 7.6. Supply models for transport systems with scheduled services. In: Cascetta, E. (Ed.), Transportation Systems Analysis: Models and Applications. Springer.
Watling, D.P., Connors, R., Chen, H., 2019. Sensitivity analysis of optimal routes, departure times and speeds for fuel-efficient truck journeys. In: Proceedings of the MT-ITS 2019.

Further reading

Gentile, G., Papola, N., 2009. 7.5. Dynamic traffic assignment with non-separable link cost functions and queue spillover. In: Cascetta, E. (Ed.), Transportation Systems Analysis: Models and Applications. Springer.

CHAPTER

Conclusion

7

Giulio Erberto Cantarella
University of Salerno, Salerno, Italy

Praise a book when 'tis read.
Cfr. the Hávamál

Outline. This chapter resumes the general framework presented in the previous chapter and outlines some of the major research in progress and perspectives.

A quite long title better reflecting the contents of this book would be "Graph, Dynamic and Stochastic models for analysing distribution over Space, evolution over Time and Uncertainty (due to lack of information) about flows and costs in a Transportation System: the modern theory of travel demand assignment to a transportation network."

Indeed, this book presented a general theory encompassing most already existing or devisable assignment models for transportation systems analysis. It focused on the mathematical issues needed to fully understand transportation systems analysis, rather than implementation or practical issues (possibly discussed in a future book).

The proposed general modelling theory of demand assignment has been presented in a progressive way from simpler to increasingly more general and complex modelling approaches. All of them have been described within the powerful framework of the Six Equation Assignment Modelling (SEAM) approach useful to classify and specify assignment models. In many case the six equations can be reduced to two equations only, one about costs and one about flows to ease the analysis, leading to the Two Equation Assignment Modelling (TEAM) approach.

The possibly too ambitious purpose is that this theory will help researchers to share a common language about the core elements of demand assignment and to propose advancements in a clear and consistent way. This book will hopefully be useful to teachers and students too for supporting exchange of knowledge within a common framework.

At this aim a consistent set of notations has been developed as described at the end of most chapters and proposed to the scientific community; in some cases traditional notations have been changed to clearly distinguish scalar variables from vectors and matrices, deterministic variables from random ones, main variables (Roman letters) from parameters (Greek letters).

Dynamics and Stochasticity in Transportation Systems. https://doi.org/10.1016/B978-0-12-814353-7.00007-8
© 2020 Elsevier Inc. All rights reserved.

This book will also provide a tool box for practitioners supporting the understanding and the specification of the right modelling approach for applications such as project assessment and evaluation, feasibility studies, urban planning, evacuation plan design, transportation systems with ITS and/or autonomous-connected vehicles.

The presented assignment models can be embedded within transportation supply design methods regarding for instance traffic lights, urban lane allocation, fares and tolls, transit frequencies and stops, … . However, a full integration is still an open issue, in particular with reference to dynamic models.

All the presented models have been described for transportation systems with a single transportation mode, without explicit modelling of mode choice behaviour as well as of demand variability with costs (elastic demand). Indications for the extension to multi-type multi-mode transportation systems with elastic demand have been provided at the end of Chapter 3, but this is still an open issue for dynamic assignment (possibly discussed in a future book on advanced topics).

- *Additional materials*

  A mathematical companion is currently in progress to support the unfamiliar reader with the mathematical background behind the mathematical notes within the main text. Materials is also planned about detailed discussions of presented examples as well as of others about cost functions with asymmetric Jacobian or with indefinite Jacobian.
- *Research in progress*

  The stability and bifurcation analysis of Deterministic Processes with Moving Average cost updating filters with μ-day memory requires the specification of a μ-th degree polynomial equation. The analysis of this model with $\mu = 2$, mirroring results reported in Sections 4.3 and 4.4 for Deterministic Processes with Exponential Smoothing cost updating filters is currently in progress and will be the main topic of an add-on. Main results will possibly be anticipated in a paper.

  Day-to-day Dynamic Process models for assignment with daily updated traffic control strategies are still to be specified and analysed requiring a further updating equation describing how control variables are defined each day. Preliminary results indicates that the kind of stability and bifurcation analysis carried out in Sections 4.3 and 4.4 requires the specification of a 3-rd degree polynomial equation.
- *Technical details*

  All numerical results in the reported examples have been computed through an-hoc Mathcad 15 code. Figures showing numerical results have been provided by this code, then further edited in PowerPoint, the others have directly been drawn in PowerPoint. Tables showing numerical results provided by this code have been designed in Word over a grid with 5×10 mm cells, then further edited in PowerPoint.

7.1 Research perspectives

Several issues are worth of further research efforts, beside those already mentioned above, such as detailed specifications of the discussed assignment models for transit systems and for freight transportation, use of other choice modelling theories, relationship with (stochastic) evolutionary game theory.

As noted in Chapters 4 and 5, very few results are already available about the parameter calibration of Dynamic Process models, another topic surely worth of further research efforts.

Looking into the future of demand assignment, the most promising approach seems based on the integration of multi-agent systems within disaggregate stochastic process models, allowing to deal with both day-to-day and within-day dynamics.

7.2 Remarks

He is no true friend who only says pleasant things.
from the Hávamál.

The many drawbacks of assignment models based on the assumption Wardrop route choice behaviour have already been highlighted in several chapters; main ones are briefly remembered below.

– Behavioural issues: it is assumed that users have a perfect and complete knowledge of the route costs, and the modeller has a perfect and complete knowledge of the costs affecting user choice behaviour; even unavoidable uncertainty about costs due to modelling simplifications are neglected.
– Mathematical issues: resulting flow maps are not functions, uniqueness of arc flows does not guarantee uniqueness of route flows; more than that, this behavioural approach is arguable for specifying Deterministic Process models, and cannot be applied at all to support Stochastic Process modelling.

In a general perspective, the wide set of available choice modelling theories include as a special case utility-based theories, which in turn include uncertainty-based theories, which include theories based on belief theory including as special case probability and possibility theories leading to random variables and fuzzy numbers, respectively. In comparison to this richness of modelling approaches, still to be fully exploited, the Wardrop choice behaviour modelling approach is just a limit case of the last two, say Random and Fuzzy Utility Theories. These considerations are even more relevant remembering that Random Utility Theory has been proposed forty five years ago.

Indeed often the simplest approach is not the most effective one, as shown by an analogy. Five centuries ago Nicolaus Copernicus presented his well known heliocentric model, the orbit of each planet supposed being a circle; after one century Johannes Kepler showed that the orbit of each planet is an ellipse. Afterwards nobody used circles for modelling planet orbits even if a circle a simpler curve than an ellipse, and a circle is the limit of a succession of ellipses as eccentric goes to zero.

Needless to say, all researchers belonging to the scientific community working on travel demand assignment to a transportation network will forever be in debt to J.D. Wardrop who gave birth to this topic and more generally to transportation system analysis with his 1952 seminal paper, worth to be quoted again as end of this chapter.

Further reading

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proc. Inst. Civ. Eng. 2 (1), 325–378.

Postface

No book can ever be finished. While working on it we learn just enough to find it immature the moment we turn away from it.
Karl Popper

Outline. This chapter presents some final philosophical considerations about the content of this book.

As noted in the Preface modelling is an attitude intrinsic to the mankind nature and to its approach to the knowledge of the real word; still each models, whichever the kind, is an opinion and as such should not be confused with subject of the modelling activity: the real world. A mathematical model can be considered a metaphor from the real world to the mathematical one [interestingly the Greek word μεταφορά also means transportation].

In this book the *epistemologems* (cfr *mythologems* introduced by Károly Kerényi) of the general complex system modelling and analysis have been discussed and used to propose *epistemologems* for the transportation system analysis as founding elements of a consistent theory of travel demand assignment to a transportation network. Indeed, this book describes a general approach to develop an effective mathematical theory for analysing any consistent class of highly dimensional real systems with non-linear feedbacks, and proposes an application of this approach to develop such a theory to demand assignment.

On the other hand, Karl Popper warned against any ultimate theory stating that *whenever a theory appears to you as the only possible one, take this as a sign that you have neither understood the theory nor the problem which it was intended to solve*. Thus the presented theory should mainly be considered just a step towards a more general theory of demand assignment, in much as the same way as a traveller stopping in an inn at evening before re-starting the trip next day towards the final destination; hence readers are expected to argue about the contents of the book, to further advance them, to propose modification, … .

A short history of this book

I began working on travel demand assignment to a transportation network in the early 1990s and rather soon I established a fruitful cooperation with Ennio Cascetta, as well as a still warm deep friendship, born when the two of us were at the University of Reggio Calabria. After having published some papers on dynamic assignment we decided to prepare some more extended documents, and we involved Maria Nadia

FIG. 1

An old fashion diskette containing the very first version of this book.

Postorino in what now can be considered an early version of this book and of the mathematical companion. Fig. 1 shows the diskette containing the two files.

In the next decade I moved to the University of Salerno, close to my hometown, but I still had the chance to working on this topic as a tutor of a PhD student, Pietro Velonà, at the University of Reggio Calabria. Together we kept working on Deterministic Process models and on tools for analysing them for several years, before he became a college teacher in a small town close to Reggio Calabria.

In 2010 David Paul Watling and I met at a congress in London (we were already in acquaintance since the jotter initiative in 2004) and decided to work together and to focus our research efforts on both Deterministic and Stochastic Process models and their relationships. We developed a special approach to working, based on thinking, walking, talking together during meeting in Leeds or in Salerno, before starting writing separately. I really enjoyed the time spent together speaking about several fundamental topics of transportation analysis as well as any other topic about life.

In 2017 David and I decided to answer to a call for books from Elsevier, and after some months our proposal was accepted. Unfortunately afterwards David had to withdraw from this project due to new unforeseeable compelling academic commitments. Therefore the original contents of the book was redesigned and Stefano de Luca and Roberta Di Pace both members of the Transportation Systems Analysis and Design Team at the University of Salerno, and good friends of mine since long were involved.

A final comment

There are two kinds of truth: the truth that lights the way and the truth that warms the heart. The first of these is science, and the second is art. Neither is independent of the other or more important than the other. Without art science would be as useless as a pair of high forceps in the hands of a plumber. Without science art would become a crude mess of folklore and emotional quackery. The truth of art keeps science from becoming inhuman, and the truth of science keeps art from becoming ridiculous. (Great Thought, February 19, 1938)

Raymond Chandler (from The Notebooks of Raymond Chandler)

In my opinion mathematics is a sort of language for speaking of science rather than a science in itself, a kind of poetry, a form of art. Thus, to get an effective mathematical theory you have to look at the same time for a beautiful elegant theory. Fig. 2 below tries to suggest the reader which is the right perspective needed to catch the deep meaning of this book. Moreover it is an artistic portrayal of the state space of a dynamic system with several attractors and circling trajectories to reach each of them.

Giulio Erberto Cantarella

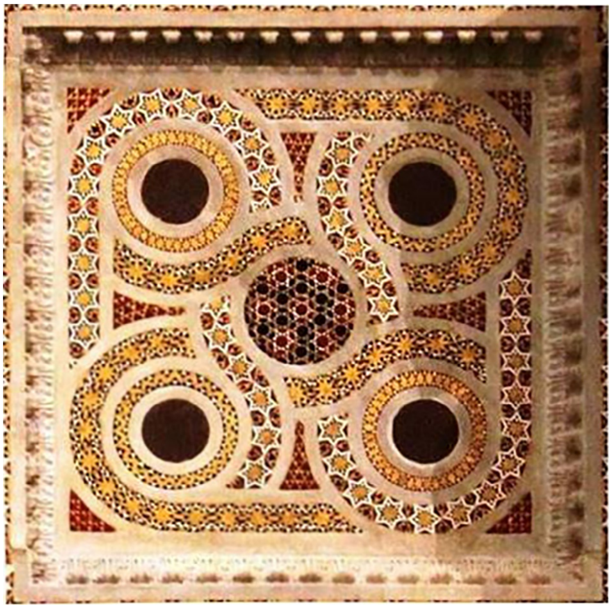


FIG. 2

A particular of the inside of the Salerno Cathedral.

Photo shoot by Roberta Di Pace.

Discrete choice modelling with application to route and departure time choice

A

Stefano de Luca
University of Salerno, Salerno, Italy

Not to be absolutely certain is, I think, one of the essential things in rationality.
Bertrand Russell

Knowledge is an unending adventure at the edge of uncertainty.
Jacob Bronowski

Outline. Travel demand analysis and simulation is one of the main issues of Transportation Engineering and represents one of the most investigated topics in the last 40 years.

Generally a travel demand model could be interpreted as a tool to estimate origin-destination demand flows, to simulate travel behaviour, but also a tool to understand and quantify the determinants of travel choices.

In general, a travel/demand choice model tries to simulate the real choice phenomenon by describing input factors and output effects through numerical variables and existing relationships through a function

$$\mathbf{y} = g(\mathbf{x}; \boldsymbol{\theta})$$

where $\mathbf{y}$ is the vector of the output variables, $\mathbf{x}$ is the vector of the input variables, $g(\bullet)$ is the modelling function, and $\boldsymbol{\theta}$ is the vector of the model parameters.

The meaning of input and output variables come out from the description of the real phenomenon itself (assuming that they can be observed and measured). For instance, when dealing with discrete choice analysis, choice attributes play the role of input variables and observed choice fractions (equal to 0 or 1) of output ones.

Provided that a sample of observations concerning the real phenomenon is available, a model can be completely defined by a trial-and-error process which consists in three main stages: specification, calibration and validation.

The *specification* (the definition of a specific mathematical expression) of the modelling function, $g(\mathbf{x}; \boldsymbol{\theta})$ is a matter of the analyst's judgement based on observations, except in the very simplest cases, since several options are often available. It should be stressed that, as is well-known from calculus, given a finite number of points, infinite many functions can be found that fit them perfectly (provided that the functions include a large enough number of parameters).

Three different modelling approaches can be distinguished to specify modelling functions:

- White Box: the function is derived from theoretical considerations which try to explain the phenomenon itself. This is often the case for models emanating from the direct application of physics or the like (e.g. models describing the trajectory of a car moving on a road) or through some analogy (e.g. models describing car flow along a highway). In these cases the model parameters may often be given a clear interpretation.
- Black Box: the function is derived from empirical considerations. This is often the case when no satisfactory theoretical paradigm is available. In this case parameters may be interpreted with difficulty, although a larger set of functions may be taken into consideration; this is the case of ANNs.
- Grey Box: any mix of the two types above and usually adopted when only a partially satisfactory theoretical paradigm is available. This is generally the case when human decision-making is concerned, e.g. econometric choice models such as Random Utility Models.

Once a type of modelling function has been specified, the *calibration* of its parameters should be performed in an attempt to best reproduce a sample of observations (data-set).

Usually, a (scalar) calibration function $\kappa(\boldsymbol{\theta}; S)$ (not to be confused with the modelling function) is defined as an indicator expressing how well the observations in the sample S are reproduced by a given set of parameters, $\boldsymbol{\theta}$ (the modelling function having already been defined).

The values of parameters corresponding to the best value of the calibration function are considered the most sensible values. When a statistical technique may be employed, the obtained values of the parameters can be considered statistical estimates of those values. An example of this approach is the maximum likelihood estimation, well-known within statistical inference.

More generally, any distance function (d) over the sample S, between the ith observed inputs ($\mathbf{x}_i$) and predicted outputs ($\mathbf{y}_i$) can be adopted, $\kappa(\boldsymbol{\theta}; S) = \sum_i d(\mathbf{y}_i, g(\mathbf{x}_i; \boldsymbol{\theta}))$, such as the minimum of the sum of the squares of differences being the most commonly adopted choices (least square estimators within a statistical framework). Special care should be devoted to the possibility of multiple local optimal points. It is common practice to deal with this condition by applying the optimisation algorithm to several (randomly generated) starting conditions, when an in-depth study of the calibration function is not easy.

It should be noted that even though the calibrated values of the model parameters are optimal (in the sense that they optimise the adopted calibration function), the resulting model may yet perform poorly, since the modelling function might have been badly specified (apart from the rather irresolvable case of the availability of poor data). Thus, the effectiveness of the model resulting from the specification and calibration stages must be explicitly analysed, that is, the *validation* stage should be carried out, before the model can be effectively used to generalise. Besides consistency with modelling assumptions, at least two features should be considered: quality of *approximation*, in other words, how closely the real phenomenon is reproduced, and *robustness*, that is, the stability of the model with respect to small variations in calibration data or input variables, the last feature being closely linked to generalisation. Several indices have been developed with this in mind, which are easily available within a statistical inference framework. Moreover, it is good practice to use another sample of observations, hold-out or validation sample, in order to test the generalisation capability of the obtained model on it (despite the fact that this stage is often skipped to reduce the costs of model building). Once a model has been validated, it can be compared with others through the same indices adopted for validation. The final selection, among all the available specified and calibrated models, is made by also considering their efficiency, or rather, computational speed and memory requirements.

Within the before mentioned general context, different theoretical paradigms have been developed in the last decades: behavioural and not-behavioural.

Behavioural models are mainly founded on the Utility theory, and may refer to perfect rationality and/or bounded rationality paradigms. Non-behavioural are mainly nonlinear regressive approach that may rely on traditional methods or on Neural Networks..

The aim of the chapter is threefold, it introduces

(i) Consolidated approaches for modelling disaggregate travel behaviour.
(ii) The most commonly adopted choice models for route choice and departure choice.
(iii) Alternative approaches to disaggregate travel behaviour that can be interpreted as an alternative paradigm and/or as a benchmark in terms of interpretation of the uncertainty and/or in terms of capability to reproduce users' choice.

It should be clarified that the aim of this chapter is not to cover any possible approach for travel behaviour analysis and modelling, but to give a synthetic and compact overview of the possible solutions that may be implemented within the theoretical framework proposed in the book.

The chapter is organised as follows. Section A.1 introduces the random utility theory (RUT) and the main random utility models (RUMs), moreover the issues related to the calibration and validation of RUMs are discussed. Section A.2 introduces a general framework for modelling the route choice process, and the modelling solutions founded on random utility theory. Section A.3 introduces the modelling solutions for departure time choice issues. Section A.4 introduces an alternative paradigm to interpret uncertainty within utility theory: fuzzy utility theory and fuzzy utility models are discussed. Section A.5 introduces the non-behavioural modelling approach founded on artificial neural network model.

A.1 Random utility theory for modelling traveller's choice

Transportation Systems users' choices are commonly analysed through models derived from the random utility theory (Domencich and McFadden, 1975; reviews in Ben-Akiva and Lerman, 1985; Train, 2009; see also Daganzo, 1979, Hensher and Button, 2000; and Cascetta, 2009), where:

a. Each user (of a class of homogenous users) of a transportation system.

- **a1.** considers a set of alternatives;
- **a2.** gives each alternative a value of perceived utility;
- **a3.** chooses the alternative with the maximum value of perceived utility (homo-economicus assumption).

b. The perceived utility of a mode is modelled through a (continuous) random variable due to several sources of uncertainty, such as unobserved attributes, unobserved taste variations, measurements errors and imperfect information, instrumental variables (Ben-Akiva and Lerman, 1985).

To apply models derived from this theory three main elements must be specified:

R the choice set, which is the set of available alternatives, mutually exclusive, it is assumed non-empty and finite, with $m = |R|$;

U_r the perceived utility, or better its distribution ϕ_{Ur};

$v_r = \mathrm{E}[U_r]$ the systematic utility, or the expected value of the perceived utility, specified as a function, $v_r = \mathrm{v}(\mathbf{x}_r; \boldsymbol{\psi})$ of a vector of attributes ($\mathbf{x}_r$) which can be measured for the current state or assumed for a design scenario and of a vector of parameters that should be estimated ($\boldsymbol{\psi}$).

Hypotheses on perceived utility pdf, ϕ_{Ur}, allow to define the probability that the perceived utility of alternative r be maximum, that is the probability of choosing alternative r conditional to the choice-set R:

$$\mathrm{p}[r/R] = \Pr[U_r > U_{M^*}]$$

where $U_{M^*} = \max_{h \neq r} [U_h]$. Thus, the choice probability vector, $\mathbf{p}$, results a function of the systematic utility vector, $\mathbf{v}$, and (possibly) other parameters, $\boldsymbol{\theta}$, of the perceived utility distribution:

$$\mathbf{p} = \mathbf{p}(\mathbf{v}; \boldsymbol{\theta})$$

Parameters of the utility function, $\boldsymbol{\psi}$, as well as those of the perceived utility pdf, $\boldsymbol{\theta}$, can be estimated through statistical inference from a sample of observed choices (disaggregate) and/or data from user flows (aggregate).

In the following sections the following three main issues will be addressed:

(i) the definition of the choice set (R),
(ii) the specification of systematic utility function (v_r; $\mathbf{v}$),
(iii) the hypotheses on perceived utility pdf, ϕ_{Ur}.

The discussion will adopt the traditional formulation which expresses the perceived utility of the generic alternative U_r as the sum of the systematic utility, v_r, e and of a random residual ξ_r which represent the dispersion between systematic and perceived utilities (uncertainty). Moreover, all the specification that will be introduced refers to a specific user class, thus no specific identifying symbol will be introduced.

A.1.1 Choice set definition

Users may differ with respect to available alternatives, say choice set, R. Alternative availability can be simulated through several approaches, as for any other socio-economic characteristic, such as income or sex:

- availability attributes within the choice model (implicit approach) and
- segmentation of demand with a discrete distribution (explicit approach), which may be prefixed or explicitly modelled.

Availability attributes, may take binary values, say 0 or 1, to simulate non-availability or availability of the alternative; availability can be stated by the user during the demand survey, or checked by the modeller. These attributes may be used jointly with or instead of other socio-economic ones, such as the ratio between the number of cars and the number of licensed drivers in the user household.

More generally, availability attributes in the range [0,1] can be considered as degree of possibilities, say the choice set is modelled as a fuzzy set (e.g. implicit availability perception approach, introduced by Cascetta and Papola, 1997).

A simple, but effective, approach to model alternatives availability is represented by the Dogit model (Gaudry and Dagenais, 1979a; Gaudry, 1981). For each alternative r a non-negative parameter, q_r, is introduced proportional to the share of users who have available that alternative only ($q_r = 0$ meaning no captivity on alternative r):

$$p_r = 1/\left(1+\sum\nolimits_{\mathrm{h}} q_h\right) \cdot p_{r/nc} + q_r/\left(1+\sum\nolimits_{\mathrm{h}} q_h\right)\ \mathrm{h} \in \mathrm{R}$$

where $q_r/(1+\sum_h q_h)$ is the probability of being captive to alternative r; $1/(1+\sum_h q_h)$ is the probability of not being captive to any alternative; $p_{r/nc}$ is the probability of choosing alternative r conditional to not being captive to any alternative, which may be specified by any random utility model.

Parameters to be calibrated are those within the utility specification and the Dogit captivity parameters (q_r).

Another approach, also known as implicit availability perception (IAP) method, considers that an alternative may have intermediate levels of availability/perception to a decision-maker (Cascetta and Papola, 1997). The decision-maker's choice set is then viewed as a "fuzzy set"; it is no longer represented as a set of [0/1] Boolean variables, but as a set of continuous variables $\tau_R(r)$ defined on the interval [0,1] and with respect to choice-set R.

The model accounts for different levels of availability and perception of an alternative by directly introducing an appropriate functional transformation of $\tau_R(r)$ into the alternative's utility function:

$$U_r = v_r + \ln \tau_R(r) + \xi_r$$

where U_r is the perceived utility of alternative r; v_r is the systematic utility of alternative r; ξ_r is the random residual of alternative r; $\tau_R(r)$ is the level of membership of alternative r in the choice set R $(0 \le \tau_R \le 1)$.

In this way, all the alternatives can be considered as theoretically available, but if alternative r is not available ($\tau_R(r) = 0$), the term $\ln(\tau_R(r))$ forces its perceived utility U_r to minus infinity and the probability of choosing it to zero, regardless of the value of v_r.

The main issue of this approach relies on the specification of the term $\ln(\tau_R(r))$ which may expressed as random variable with mean valued expressed as a function of the availability and perception attributes (e.g. a Binomial Logit model may be used). However practical formulations may be found in Cascetta (2009).

A different approach may consist in explicitly modelling the composition of the generic decision-maker's choice set. In particular, in the explicit approach, the choice probability of an alternative r may be expressed through a two-stage choice model:

$$\mathrm{p}[r] = \sum_{R \in G} \mathrm{p}[r, R] = \sum_{R \in G} \mathrm{p}[r/R] \cdot \mathrm{p}[R]$$

where R is the generic choice set; G is the set made up of all possible non-empty choice sets (non-empty subsets of the set of all the possible alternatives); $p[r,R]$ is the joint probability that decision-maker will choose alternative r and that R is his/her choice set; $p[r/R]$ is the probability that decision-maker will choose alternative r, his/her choice set being R; $p[R]$ is the probability that R is the choice set.

The choice probability $p[r/R]$ conditional on set R can be represented with any of choice models that will be introduced in the next sections, whilst the probability that R is the choice set can be formulated in terms of probability that each single alternative belongs to the choice set, which, in turns, may be expressed through regressive or logistic functions properly calibrated from a sample of users (details in Cascetta, 2009; Mansky, 1977; Morikawa, 1996; Swait and Ben-Akiva, 1987a, 1987b).

The explicit approach, although very interesting and consistent from a theoretical point of view, presents significant computational problems for large number of alternatives.

A.1.2 Specification of the systematic utility

For each alternative r, the systematic utility, v_r is generally specified as a function of several attributes, x_{rj}, and parameters, ψ_j. For instance a linear in parameters and in attributes utility specification is usually adopted:

$$v_r = \sum_j \psi_j x_{rj}$$

The effectiveness of a model can be greatly improved by considering user *socio-economic* attributes (*SE*—such as income, age, etc.) and *land-use* ones (*LU*—such as population of zones, etc.) a part from those describing *level-of-service* (*LoS*—such as travel time and monetary cost). Moreover, for every alternative, but one, an *alternative specific* attribute (*ASA*) is usually introduced to consider (quantify) those determinants that may significantly affect choice behaviour, but known to the analyst and/or not easily measurable.

Attributes can be introduced in the above systematic utility specification as such (absolute attributes), or as a ratio with respect to an alternative used as reference (relative attributes). Moreover, but depending on the choice context, *LoS* attributes may be weighted by continuous/discrete attributes, usually socio-economic, in order to explicitly represent (and weight) the different sensitivity to the Los of different *SE* characteristics.

Attributes effectiveness may further be improved by considering non-linearity with respect to continuous attributes; a part from a better estimation of the effect of such attributes, non-linearity allows to simulate asymmetry of choice probability elasticity. Several approaches can be devised as described below.

- *Dummy variables.*

ASA's can further be differentiated by the range of values. Then, users are grouped into demand segments with common value of *ASA's*.

- *Threshold variables.*

The parameter of a (*LoS*) attribute varies according to the some (pre-fixed) ranges of values. Ranges and parameters can be specified in order to get a (continuous) piecewise linear relation between the values of the systematic utility and of the attribute (continuous thresholds). Else it may assumed that the parameter increases when (given) thresholds about the value of an attribute are overcome (non-continuous thresholds).

- *Box-Cox transformations.*

A more general approach is based on Box-Cox transformations (Box and Cox, 1964; Gaudry, 1981; Gaudry and Dagenais, 1979b). Such a (monotone strictly increasing) transformation y_{rj} of attribute x_{rj}, that may be applied to a strictly positive variable only, is defined by a non-negative shape parameter, λ_j:

$$y_{rj}(\lambda_j) = \left(x_{rj}^{\lambda j} - 1\right)/\lambda_j \text{ if } \lambda_j \neq 0$$

$$y_{rj}(\lambda_j) \rightarrow \ln\left(x_{rj}\right) \text{ if } \lambda_j \rightarrow 0$$

Box-Cox transformation can be considered as a (continuous) non-linear generalisation of the continuous threshold introduced above. The partial first derivative of systematic utility, v_r, with respect to the attribute, x_{rj}, depends on the value of the attribute, according to the value of the shape parameter, λ_j; however, it will be linear with respect to the corresponding beta coefficient.

In principle, a different Box-Cox transformation can be implemented to each attribute, but it is usually more effective to apply it to few attributes only. The introduction of Box-Cox parameters slightly increases the computational effort needed for calibration, but the specification of the resulting model is more flexible, allows representing not symmetric behaviour and does not require any prefixed values with respect to threshold variables.

All the previous attributes measure characteristics easily identifiable of the alternatives and/or characteristics of the decision makers, but they are not able to grasp the psychological factors that may significantly affect users' behaviour, such as perceptions, beliefs, attitudes.

Within this context, an increasing attention towards specifications able to embed in random utility model such attributes has been observed.

In particular, random utility choice models with latent variables (often defined hybrid models) have been increasingly adopted for simulating choice contexts in which psychological factors may play a significant role.

The hybrid choice model (HCM) based on random utility theory is a discrete choice model which integrates and simultaneously estimates different types of sub-models into a unique structure. If the HCM includes a latent variable model, it is possible to take into account the effects of users' latent attitudes, perceptions and concerns (i.e. Integrated Choice and Latent Variable model, ICLV).

Fig. A.1 introduces the general structure of a ICLV and allow to comprehend the different sub-models that define a ICLV: the latent variable model and the discrete choice model. In particular, the ellipses represent the unobservable (latent) variables, the rectangles represent the observable variables, and the circles represent the error variance or disturbance terms.

Since the latent variables (attitudes, perceptions and concerns) cannot be directly observed and measured from a revealed choice or a stated preference experiment, they have to be modelled and then indirectly identified starting from a set of indicators. The latent variable model allows to identify and measure these unobservable variables as a function of the indicators, in order to include them in a choice model.

Mathematically, a latent variable is treated as a random variable; the latent variable is specified through a *structural equation* formalising it as a function of several parameters and a random error term. With regard to the relationship between indicators and latent variables, it can be formalised through a *measurement equation*, in

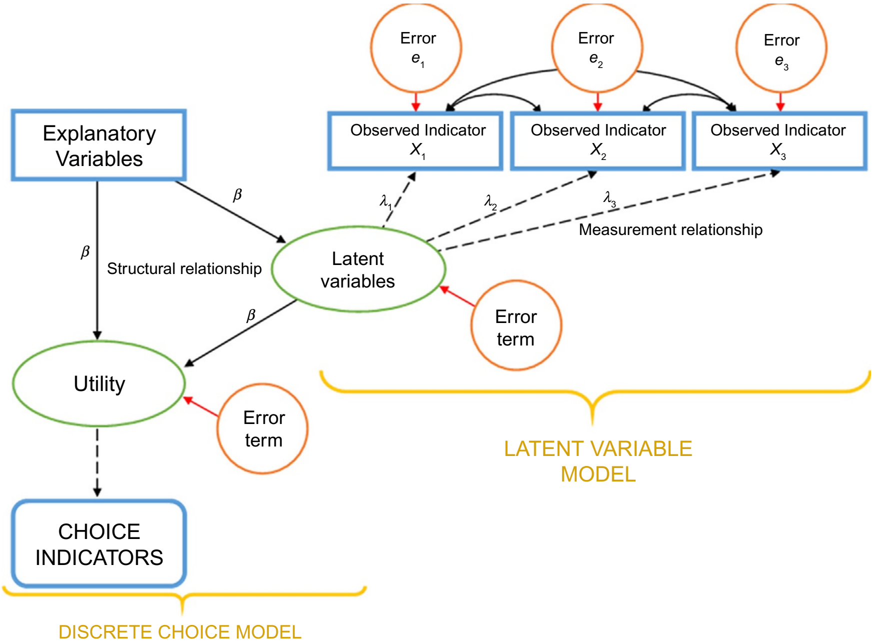


FIG. A.1

Scheme of integrated choice and latent variable model (ICLV).

which each observed psychological indicator is a function of a latent variable and a random error term. In general, each latent variable may be part of more than one measurement equation.

Finally, in accordance with the random utility theory, the latent variables are included in the utility function of the alternatives as explanatory variables. Indeed, ICLV models may have any of the previously introduced mathematical formulations, but they differ from traditional models for the systematic utilities specification. Indeed, the systematic utility functions may be expressed as a function of instrumental attributes of alternative r (x_{rj}), users' specific attributes (x_s) and latent variables, LV_l.

$$v_r = \sum_j \psi_j x_{rj} + \sum_s \psi_h x_s + \sum_l \psi_l LV_l$$

With reference to the LV_l two equations should be specified: the structural and the measurement equations.

The structural equations are introduced in order to specify the latent variables, whilst the measurement equations are introduced in order to specify the perception indicators.

In particular, if l is the generic latent variable, the structural equation for each latent variable may be expressed as follows:

$$LV_l = \gamma_l + \sum_p \phi_p X_{SE,p} + \omega_l$$

where γ_l is the intersect; $X_{SE,p}$ is the pth users' socioeconomic attribute which concurs in the definition of the LV_l; ϕ_p is the pth coefficient associated with the pth users' socioeconomic attribute; ω_l be the error term which is usually normally distributed with zero mean and standard deviation to be estimated.

The specification/estimation of each LV_l requires the specification/calibration of psychometric indicators. Indeed, different psychometric indicators ($I_{l,n}$) reveal the latent variable LV_l and, vice versa, each psychometric indicator concurs to the definition of the LV_l. $I_{l,n}$ which may be expressed through a measurement equation as follows (measurement equation):

$$I_{l,n} = \alpha_{l,n} + \lambda_{l,n} LV_l + \nu_{l,n}$$

where $I_{l,n}$ be the nth perception; $\alpha_{l,k}$ is the intersect; $\lambda_{l,k}$ is the coefficient associated with the latent variable (to be estimated); $\nu_{l,k}$ is the error terms usually assumed normally distributed with zero mean and standard deviation to be estimated.

The psychometric indicators may be observed through ad hoc surveys and are, usually, coded using a Likert scale. These indicators can be considered to be a linear continuous expression of the *LV's* or an ordered discrete variable. The first approach has been historically chosen because simpler and more practical with lower computational cost.

In recent years, several studies have treated them as discrete variables, but with a higher computational cost. In particular, if the measurement is represented by an ordered discrete variable Φ having m elements $\{\phi_1, \phi_2, \ldots, \phi_m\}$, one for each response in the survey, we may have:

$$\Phi = [\phi_1 \text{ if } I_{l,n} < \tau_1; \phi_2 \text{ if } \tau_1 \le I_{l,n} < \tau_2; \ldots; \phi_m \text{ if } \tau_{m-1} \le I_{l,n} < \tau_m]$$

where and $\tau_1 \le \tau_2 \le \cdots \le \tau_{m-1}$ are parameters to be estimated, but due to symmetry constraints, only $(m-1)/2$ parameters can be independently be estimated. Usually, the ordered Probit model is used for estimating the probability of Φ elements.

Hybrid choice models with latent variables can be an effective modelling solution to embed psychological factors in traditional random utility models. However, their operative implementation in travel demand estimation and/or traffic assignment problems could be particularly demanding. In this sense, the approach is mainly finalised to better understand travel behaviour and/or to interpret the role that not-instrumental attributes play in the choice process and within the "utilitarian" behaviour.

A.1.3 Distribution of perceived utility and choice functions

Several assumptions can be adopted about the perceived utility random distribution.

The probability of selecting alternative r conditional on his/her choice set R, can be formally expressed as the probability that the perceived utility of alternative r is greater than that of all the other available alternatives:

$$p[r/R] = \Pr[U_r > U_h; \forall r \neq h, h \in R]$$

That, in turn, may be expressed into the following formulation which depend on the vector of perceived utilities (**U**) and on their joint density function *f*(**U**).

$$p[r/R] = \int_{U1<Ur} \int_{Uj=-\infty\ldots}^{+\infty} \int_{\mathrm{Um<Uj}} f(\mathbf{U}) \cdot \mathrm{d}U_1 \ldots \mathrm{d}U_{\mathrm{m}}$$

In conclusion, the estimation of choice probabilities of a random utility model involves multi-dimensional integral computation and its complexity mainly depend on the assumption that can be made on *f*(**U**).

The general, and most important, assumption concerns the independence of joint density function *f*(**U**) from the perceived utilities mean values and from the choice-set *R* composition. Such an hypothesis defines a class of models, also known as *invariant random utility models*, that allows to simplify the expression of the probability of choosing the generic alternative *r*.

Indeed, expressing the *r*th perceived utility U_r as the sum of the systematic utility v_r and the *random residual* ξ_r, the choice probability may be formulated as

$$\mathrm{p}[r/R] = \Pr[v_r - v_h > \xi_h - \xi_r; \forall r \neq h, h \in R]$$

and can be calculated assuming the difference, $\xi_h - \xi_r$, independently distributed from the difference between systematic utilities $v_r - v_h \ \forall r \neq h$.

Under the assumptions made above, choice probabilities depend on the differences among the utilities, and not on the absolute values of the utilities. Furthermore, *invariant random utility models* not only allow an easier computation of choice probabilities, but they also have some important properties that are useful for the specification od RUM and are worth of interest for the general mathematical framework of the book (see mathematical notes 1 for proofs and further details):

(1) Choice probabilities do not change if the utility values are multiplied for a scale factor. In other words. The scale does not matter in RUMs.

(2) The only parameters that can be estimated are those concurring to define utilities differences.

(3) If the joint density function of the random residuals is continuous with continuous first derivatives, the choice probabilities are also continuous functions of the systematic utilities with continuous first derivatives.

(4) The Jacobian of choice probabilities is (symmetric and) positive semidefinite.

(5) The choice probabilities are monotonic increasing functions of the systematic utilities. Indeed, the choice probability of a generic alternative does not decrease as its systematic utility increases, if all the other systematic utilities remain unchanged. Using an analogous argument, it can be demonstrated that, as v_r tends to minus infinity, the choice probability of alternative *r* tends to zero.

Before introducing the possible mathematical formulations, it is useful to introduce a classification based on the possible assumptions that can be made on the random

Table A.1 Classification of type of random utility models

			Parameters of systematic utility functions	
			Not distributed	**Distributed**
Random residuals	*Identically distributed*	Not jointly	Homoscedastic with no correlation among perceived utilities	Heteroscedastic models and perceived utilities potentially correlated
		Jointly	Homoscedastic and correlated	Heteroscedastic and perceived utilities correlated
	Not identically distributed	Not jointly	Heteroscedastic and not correlated	Heteroscedastic and perceived utilities potentially correlated (also over time)
		Jointly	Heteroscedastic and correlated	Heteroscedastic and perceived utilities correlated (also over time)

distribution of the vector of random residuals (ξ) and/or on the vector of parameters of systematic utility functions (ψ). (See Table A.1.)
Indeed, random residuals can be.

a1 identically distributed among alternatives and
a2 not identically distributed among alternatives.

Random residuals can be.

b1 not jointly distributed among alternative and
b2 jointly distributed.

Parameters of systematic utility functions can be.

c1 deterministic,
c2 randomly distributed across the users,
c3 randomly distributed across the alternatives, and
c4 randomly distributed over time.

A family of models may be obtained by the combination of the above hypotheses and, in general, three main issues require particular attention in practical applications:

(i) modelling the correlation across perceived utilities (usually related to alternatives that are similar and/or that share same characteristics);

(ii) modelling the correlation over time of user choices;
(iii) modelling the heterogeneity across users and/or over time.

With respect the above-mentioned classification, the mathematical complexity of the corresponding choice models can be significantly different. In particular all the model highlighted in grey may have a closed-form solution, whereas all the others need simulation for estimating the choice probabilities.

In the following, the most adopted random utility models are briefly introduced for each of the classes introduced in the table.

Homoscedastic and not correlated perceived utilities

The *multinomial logit* (MNL) model is the simplest random utility model. It is based on the assumption that the random residuals ξ_r are independently and identically distributed (i.i.d.) according to a Gumbel random variable of zero mean and parameter θ. Under the assumptions made, the probability of choosing alternative r among those available belonging to choice-set R can be expressed in closed form as:

$$\mathrm{p}(r) = \exp(v_r/\theta)/\sum\nolimits_{h\in R} \exp(v_h/\theta)$$

The MNL model is characterised by three main properties:

- dependence on the differences among systematic utilities,
- influence of residual variance,
- independence from irrelevant alternatives (I.I.A.).

In particular, the last property, can be an advantage, but sometimes may lead to unrealistic results if the choice set is not composed by clearly distinct choice alternatives.

As a consequence, the variance–covariance matrix of perceived utilities has a block diagonal structure and the elements of the main diagonal are all equal to $(\pi^2\theta^2/6)$.

MNL model can be easily adopted for modelling any travel choice dimensions, but with perceived utilities not correlated.

Homoscedastic and correlated perceived utilities

A closed form, but effective, alternative to the MNL is the single-level *Nested Logit* model (Williams, 1977), which allows to partially overcome the assumption of independent random residuals underlying the MNL model, retaining, at the same time, a closed analytical expression.

The *Nested Logit* (NL) assumes that alternatives are hierarchically nested in a finite number of nests (G) and the random residuals of each alternative are decomposed into two independently distributed Gumbel random variables: one for each alternative and one for each nest.

The mathematical formulation continues to hold a closed form, and can be expressed as the product of two MNL formulations, the former estimating the probability to choose the nest g among the nests G, the latter the probability to choose the generic alternative r belonging to the nest g. Each alternative is characterised by its systematic utility function, whilst each nest is characterised by the "nest utility" measured as maximum expected utility associated to the nest g.

$$\mathrm{p}(r)=\mathrm{p}(g)\,\mathrm{p}(r/g)=\left(\exp\left(\delta Y_g\right)/\sum\nolimits_{n\in G}\exp\left(\delta Y_n\right)\right)\cdot\left(\exp\left(v_r/\theta\right)/\sum\nolimits_{h\in g}\exp\left(v_h/\theta\right)\right)$$

where $p(g)$ is the probability to choose the nest (group of alternatives) g; $p(r/g)$ is the probability to choose the alternative r, belonging to the nest g, once chosen the nest g; θ is the parameter associated to the choice within the group g; θ_o is the parameter associated to the choice between the nests (root); δ is a scale parameter obtained as the ratio of the scale parameters: θ_o/θ; Y_g is the maximum expected utility associated to the nest g (satisfaction). In the case of single level NL, Y_g, also known as logsum variable, may be expressed as follows: $Y_g = \ln\,[\Sigma_h \exp\,(v_h)]$, with h the alternatives belonging to nest g.[a]

NL allows a positive co-variance between each pair of alternatives belonging to a same nest, as described by parameter δ (the ratio θ_o/θ) which may have value $\in[0,1]$ with $\delta = 1$ meaning no correlation, that is MNL.

From the previous results, the variance–covariance matrix of random residuals has a block diagonal structure. The elements of the main diagonal are all equal to $(\pi\theta_o)^2/6$, the covariance between each pair of alternatives belonging to the same group is constant and equal to $(\pi^2\theta_o^2 - \pi^2\theta^2)/6$, while the covariance between alternatives belonging to different groups is null. Therefore, if the alternatives of each group are ordered sequentially, the resulting variance–covariance matrix has a block diagonal structure.

The single level NL structure may be relaxed by assuming different θ parameters for each group and/or by introducing more levels, thus more root parameters.

In order to overcome the limitation of a block-diagonal structure of the variance–covariance matrix, a generalisation of the NL model can be introduced assuming that an alternative may belong to more than one nest (g), with different degrees of membership. The degree of membership of an alternative r to a nest g is denoted by the parameter α_{rg} which is included in the [0,1] interval. Degrees of membership have to satisfy the normalising equation: $\Sigma_g\ \alpha_{rg} = 1\ \forall\ r$.

This model, also known as *Cross-Nested Logit*—CNL (Vovsha, 1997), keeps the same formulation of the single-level NL model with the only difference that the summation is extended over all the nests (G). Indeed, the choice probability of the generic alternative r, continues to hold a closed-form expression and it can be expressed as:

$$\mathrm{p}(r)=\sum\nolimits_{g\in G}\mathrm{p}(g)\,\mathrm{p}(r/g)$$

[a] Nested Logit formulation can also be specified assuming, for each choice level, random residuals not Gumbel distributed. In this case the satisfaction variabile could not be expressed in a closed form, as well as the resulting choice probabilities.

$$p(g) = \left(\sum\nolimits_{n\in Ig} \alpha_{ng}{}^{1/\delta g} \cdot \exp\left(v_n/\theta_g\right)\right)^{\delta g} / \sum\nolimits_{g'\in G} \left(\sum\nolimits_{n\in Ig'} \alpha_{ng'}{}^{1/\delta g'} \cdot \exp\left(v_n/\theta_{g'}\right)\right)^{\delta g'}$$

$$p(r/g) = \alpha_{rg}{}^{1/\delta k} \cdot \left(\exp\left(v_r/\theta_g\right) / \sum\nolimits_{n\in \mathrm{Ig}} \alpha_{ng}{}^{1/\delta g} \cdot \left(\exp\left(v_n/\theta_g\right)\right.\right.$$

where I_g $(I_{g'})$ is the generic set of alternatives belonging to nest g (g'); θ_g ($\theta_{g'}$) is the parameter associated to the choice between alternatives belonging to nest g (g'); θ_o is the parameter associated to the root; and δ_g ($\delta_{g'}$) the ratio θ_g/θ_o ($\theta_{g'}/\theta_o$).

In this case, the variance–covariance matrix of random residuals may have covariances different from zero but with the elements of the main diagonal all equal.

Although the CNL allows a greater flexibility in substitution across alternatives, it should be noted that it is still an homoscedastic model and that the covariances of the CNL model cannot be computed with a closed-form expression. Moreover, it has been demonstrated that the CNL is not able to reproduce the whole. Homoscedastic covariance matrices domain (Marzano and Papola, 2008).

Heteroscedastic and correlated perceived utilities

Two classes of model allow to take into account both of heteroskedasticity and correlation among perceived utilities: the *Probit* model and the *Mixed*-models.

The *Probit* model (Daganzo, 1979) overcomes most of the drawbacks of the previous models and its generalisations, though at the cost of analytical tractability. It is based on the hypothesis that residuals ξ_r are distributed according to a multivariate normal (MVN) random variable with zero mean and general variances and covariances.

The choice probability of alternative r can be formally expressed as the joint probability that utility U_r will assume a value within an infinitesimal interval and that the utilities of the other alternatives will have lower values. The choice probability can be expressed as multi-dimensional integral, and this probability must be integrated over all possible values of Ur.

$$p(r/R) = \int_{U1<Ur} \int_{Uj=-\infty}^{+\infty} \int_{Um<Uj} \exp\left(-1/2\,(\mathbf{U}-\mathbf{v})^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\,(\mathbf{U}-\mathbf{v})/((2\pi)^{\mathrm{m}}\det(\boldsymbol{\Sigma}))^{1/2} \cdot \mathbf{dU}_1 \ldots \mathbf{dU}_{\mathrm{m}}\right.$$

The flexibility of the *Probit* model is paid at the cost of a computational complexity greater than the other previous models.

Indeed, the *Probit* model does not allow to express the choice probabilities in closed form, therefore approximation methods based on simulation should be adopted. A general overview on simulation procedure for the *Probit* model may be found in Hajivassiliou et al. (1996), the most adopted are: Monte-Carlo, Kernel-Smoothed Frequency (McFadden, 1989), Stern Decomposition (Stern, 1992), GHK (Geweke, 1991; Börsch-Supan and Hajivassiliou, 1993; Keane, 1994), Acceptance/Rejection or Gibbs Sampler (Hajivassilioù and McFadden, 1990), Sequentially Unbiased and Approximately Unbiased.

With regard to the variance–covariance matrix, although there are at most $(m(m+1))/2$ different values, with m the number of choice alternatives, the estimation of all the possible values can be problematic if the number of alternatives is large. To this aim, different methods might be adopted to reduce the number of unknown elements by assuming some structure underlying to the random residuals. One of the most adopted is the Factor Analytic Probit which expresses the vector of random residuals as a linear function of a vector of independent standard normal variable.

An effective alternative to the Probit model may be presented by mixed models, in which it is assumed that random residuals are distributed as sum of two terms: one distributed anyway and the second one identically, independently distributed as an extreme value random variable:

$$\xi_r = \omega_r + \gamma_r$$

As a consequence, the choice probability becomes an integral in joint density $f(\boldsymbol{\omega},\gamma_r)$, that, if assumed to be independent from each other, allows formally expressing the choice probabilities the following integral:

$$p(r/R) = \int_{\boldsymbol{\omega}} p[r/\boldsymbol{\omega}] \cdot f(\boldsymbol{\omega}) \cdot \mathrm{d}\boldsymbol{\omega}$$

Although any kind of joint probability distribution with any variance–covariance matrix may be adopted, generally simplified, but realistic, assumptions can be made in order to simplify the final mathematical formulation of the model.

One of the most used Mixed models is the *Mixed-Logit* which assumes that the second term, γ_r, is identically and independently as a Gumbel variable among all the alternatives. It allows to express the probability of choosing the alternatives r, as the integral over $\boldsymbol{\omega}$ of a MNL function conditional on values of $\boldsymbol{\omega}$.

$$\mathrm{p}(r/R) = \int_{\boldsymbol{\omega}} (\exp(v_r + \omega_r)/\theta) / \sum_{h \in R} \exp(v_h + \omega_h/\theta)) \cdot f(\boldsymbol{\omega}) \cdot \mathrm{d}\boldsymbol{\omega}$$

Moreover, if random residuals ω_r are expressed through independent, but non identical random variables, the final mathematical formulation, and the consequent computational efforts, can be significantly simplified.

Mixed Logit model allows to approximate any random utility model with a satisfying degree of closeness (McFadden and Train, 2000) and may be expressed into two different specification, but formally equivalent:

1. the error component formulation and
2. the random parameters formulation.

In the error component formulation, the γ_r residuals are i.i.d. *Gumbel* random variable and independent randomly distributed error components are introduced among the alternatives. It allows to represent heteroscedasticity and cross-correlation among perceived utilities that share the same error components. However, it should be pointed out that identification issues (Walker et al., 2007) may arise when the number of alternatives and the error components raise. To avoid such a drawback, it is advisable to preventively hypothesise parsimonious and realistic correlation structures.

The random parameters formulation assumes the systematic utility parameters, ψ_j, randomly distributed, whilst the error components remain distributed as i.i.d. *Gumbel* random variables. Random parameters formulation allows to explicitly take into account taste heterogeneity among users, but in this case the taste heterogeneity and the cross correlation between perceived utilities are expressed (and interpreted) in terms of specific attributes (e.g. travel time). Therefore, correlations among perceived utilities depend on the systematic utility attributes and not on specific error components; moreover, correlation magnitudes depend on the values assumed by the attributes whose parameters are assumed randomly distributed.

However, it should be noted that also for Mixed-models, the dimensionality of the multifold integrals requires the use of unbiased efficient estimators of choice probabilities as, for instance, the MonteCarlo method.

In conclusion, *Mixed Logit* models allow more flexible choice models without significant computational efforts, and also allows an interpretation of taste heterogeneity and of correlation. Indeed, MNP model, although allowing complete flexibility in the variance–covariance matrix, on the other hand it does not permit an easy interpretation of correlations and variances, except for specific choice contexts, such as route choice, in which the variance–covariance structure is predefined. Furthermore, the parameters identification issue should be carefully addressed.

A.1.4 Calibration and validation of a choice model

Parameters of the utility function, $\boldsymbol{\psi}$, as well as those of the choice function, $\boldsymbol{\theta}$, can be estimated by trying to reproduce a sample of observations of user choices (disaggregate) and/or demand flows (aggregate).

The parameters estimation is a calibration-validation procedure, not easily automatable, that should be performed until the most effective modelling solution is achieved.

Random Utility models are usually calibrated through the *Maximum Likelihood Method* (*ML*).

The method, as well known, maximises the probability L of observing the whole sample S, indeed the probabilities that each user (i) chooses alternative actually chosen by him/her, $r(i)$. The probabilities $p^i[r(i)](\mathbf{X}^i; \boldsymbol{\psi}, \boldsymbol{\theta})$ are computed by the random utility model and therefore depend on the coefficient vectors. Thus, the probability L of observing the whole sample is a function (the likelihood function) of the unknown parameters:

$$L = \mathrm{L}(\boldsymbol{\psi}, \boldsymbol{\theta})$$

The *Maximum Likelihood* estimates the vectors of parameters $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ by maximising the previous equation or, more conveniently, its natural logarithm (the log-likelihood function):

$$[\boldsymbol{\psi}, \boldsymbol{\theta}]_{ML} = argmax \, \ln L(\boldsymbol{\psi}, \boldsymbol{\theta})$$

The formulation of the likelihood function depends on the type of sample strategy (e.g. stratified or simple), on the type of the survey (revealed preferences or stated preferences) and on the type collected observations (cross-sectional or panel). For more details the reader may refer to Cascetta, 2009).

Once estimated the parameters, the effectiveness of the model must then be explicitly analysed, that is, the validation stage should be carried out before the model can effectively accepted and/or be used.

Indeed, a choice model should be able to address the following issues:

(1) *Description* and *interpretation* of the phenomenon through its parameters.

Choice model parameters should allow an interpretation of the phenomenon, giving a clear meaning of the variables and of their relationships. The parameters interpretation allows a first validation of the choice model goodness and gives first insights to the analyst and to the policy makers.

(2) *Reproduction* of observations about choice behaviour used to calibrate model parameters (calibration sample).

A choice model at least should reproduce the users' choices (or choice fractions) that have been used to calibrate its parameters (calibration sample). Such an analysis may be carried out through some benchmarking indicators.

(3) *Generalisation* to choice behaviour in the same transportation scenario (observations not used to calibrate model parameters—hold-out sample) and/or in different transportation scenarios (model sensitivity to level of service attributes).

Even though reproduction capabilities are usually used to evaluate and choose the best model or approach, the most performing model may show very poor generalisation capabilities because of over-fitting phenomena. The resulting choice model might not be able to reproduce users' behaviours even if they belong to the same transportation scenario, and it might not be able to simulate model sensitivity.

In the following sections, the analyses that should be carried out are summarised. All the proposed indicators can be calculated for RUMs, but most of them can be easily used for different approaches, such those introduced in Sections A.4 and A.5 (see also de Luca and Cantarella, 2009).

Analysis on utility parameters

First, an informal stage consistency between the meaning of the jth attribute and the sign of the relative parameter (ψ_j) is checked. Unexpected signs of parameters likely indicate errors in available data and/or model misspecification, moreover the ratios between the parameters of some pairs of attributes may be expected to be greater than one, such as waiting time and on-board time for transit. In addition, the ratio between the parameters of monetary cost and travel time attributes can be considered an estimate of the value-of-time (VoT), which can also be estimated through other approaches.

Once, passed the first informal validation, some formal tests are usually adopted.

For RUMs, if utility parameters are estimated through maximum likelihood estimators, the *t*-test may be applied to verify statistical significance of parameters, for large samples of available data, through statistics:

$$t_j = |\psi_j^{ML}| / \mathrm{Var}\left[\psi_j^{ML}\right]^{0.5}$$

where an approximate estimate of the variances (and covariances) of parameters ψ_j^{ML} can be easily obtained since equal (minus) the inverse of the log-likelihood function Hessian. Usually the size of the available sample is large enough to approximate the T random variable through a Normal r.v.; if the value of t_j is >1.96, the value ψ_j^{ML} may be considered (statistically) different from zero at a confidence level 0.95.

Analysis through aggregate indicators

The indicators described below when applied to calibration (or hold-out) sample help us to understand how the model reproduces (generalises) the observed choice scenario (to other choice scenarios). All the indicators are to be compared with a reference value obtained from the simplest choice model or the benchmark modelling approach.

Simulated vs. observed shares for each choice alternative

Since differences take a null value over the calibration sample for any Logit model calibrated through maximum likelihood and with only alternative specific constants, such a comparison is useful only when referred to the hold-out sample and/or to other choice models.

Test and indicators based on Log-Likelihood value

This value is always less than or equal to zero, zero meaning that all choices in the calibration sample are simulated with probability equal to one. Usually, the comparisons carried out are based on the goodness of fit statistic (*pseudo*-ρ^2) and the Likelihood Ratio test; less widely used is the comparison test (it will be called *adjusted* ρ^2).

- *Goodness of fit statistic*

The goodness of fit statistic checks the null hypothesis that the maximum value of log-Likelihood, $\ln L(\boldsymbol{\beta}^{ML})$, is equal to the value corresponding to null vector of coefficients, $\ln L(\mathbf{0})$. Thus their difference is due to sampling errors. This test is based on the so-called (pseudo-)*rho-square* statistic:

$$\rho^2 = \left(1 - \ln \mathrm{L}\left(\boldsymbol{\beta}^{ML}\right)/\ln\mathrm{L}(\mathbf{0})\right) \in [0, 1]$$

If equal to zero, $L(\boldsymbol{\beta}^{ML}) = L(\mathbf{0})$, the model has no explanatory capability; if equal to one, the model gives, for each user in the sample, a probability equal to one to the alternative actually chosen, thus the model perfectly reproduces observed choices. Usually, the model with a better *rho-square* statistic is preferred, even if the differences are small this criterion does not seem very effective.

- *Adjusted* ρ^2

The adjusted ρ^2 test is based on an enhanced value of rho-square statistic (sometimes called rho-square bar):

$$^{-}\rho^2 = \left(1 - \left(\ln \mathrm{L}\left(\boldsymbol{\beta}^{ML}\right) - N_\beta\right) / \ln \mathrm{L}(\mathbf{0})\right)$$

It attempts to eliminate the effect of the number of parameters included in the model (N_β) to allow the comparison of models with different numbers of parameters.

- *Adjusted* ρ^2 *test*

Starting from the proposed statistics, a specific test may be carried out. It assumes as null hypothesis that the statistic $\overline{\rho}_1^2$ for model *1* is not greater than the statistic $\overline{\rho}_2^2$ (assuming $\overline{\rho}_2^2 \leq \overline{\rho}_1^2$) for model 2, their actual difference being due to sampling errors. It is based on the relation:

$$\Pr\left(|^{-}\rho_1{}^2 - {}^{-}\rho_2{}^2 | > z\right) \leq \Phi(^{-}z)$$

where $\Phi(\cdot)$ is the distribution function of standard normal, $\overline{z} = -[-2\ z \ln L(\mathbf{0}) + (N_1 - N_2)]^{1/2}$; N_1 and N_2 are the number of parameters in model 1 and 2 respectively.

This test can be used to ascertain whether model 2 can be considered (statistically) better than model 1.

- *Likelihood ratio test*

The likelihood ratio test, $\mathrm{LR}(\boldsymbol{\psi}^{\mathrm{ML}})$, checks the null hypothesis that the maximum value of log-likelihood, $\ln L(\boldsymbol{\psi}^{\mathrm{ML}})$, is equal to the maximum value corresponding to a reference model with null utility parameters or to a simpler one. Hence thus the difference is due to sampling errors. This test is based on the *likelihood ratio* statistic:

$$\mathrm{LR}\left(\boldsymbol{\psi}^{\mathrm{ML}}\right) = -2\left[\ \ln \mathrm{L}(\boldsymbol{\psi}^{\circ}) - \ \ln \mathrm{L}\left(\psi^{\mathrm{ML}}\right)\right.$$

which, on the null hypothesis, is asymptotically distributed according to a chi-square variable with a number of degrees of freedom equal to the number of constraints imposed (the number of parameters). This test can be used to check whether utility parameters ($\boldsymbol{\psi}^{\mathrm{ML}}$) are (statistically) different from zero or from those of a reference model ($\boldsymbol{\psi}^{\circ}$).

As stated for the ρ^2 test, the likelihood ratio and the adjusted ρ^2 tests might not be very significant, since small differences are sufficient to verify such tests.

Indicators based on Simulated vs. observed choice fractions for each user in the sample

- *Mean square error—MSE*

$MSE = \Sigma_i \Sigma_r \left(p^{sim}{}_{r,i} - p^{obs}{}_{r,i}\right)^2 / N_{users} \geq 0$

Mean square error between the user observed choice fractions and the simulated ones of alternative r belonging choice-set R, over the number of users (i) in the sample (N_{users}) Apart of MSE indicators, the corresponding standard deviation (SD) may be computed, representing how the predictions are dispersed, if compared with the choices observed. If different models have similar MSE errors, the one with smaller SD value is preferable.

- *Fitting factor—FF*

$FF = \Sigma_{\mathrm{i}}\, p^{sim}(r_{chosen,i}) / \mathrm{N_{users}} \in [0, 1]$

It is the ratio between the sum over the users in the sample of the simulated choice probability for the mode actually chosen, $p^{sim}(r_{chosen,i}) \in [0,1]$, and the number of users in the sample, N_{users}. $FF = 1$ means that the model perfectly simulates the choice actually made by each user (say with $p^{sim}(r_{chosen,i}) = 1$).

- *Level of service impact—*LoS_{impact}

The level of service impact may be calculated by the sum of the square differences between choice probabilities computed with or without ASA's, hence it describes the effect of ASA's: the less this index, the better the corresponding model is.

Analysis of clearness of predictions

It is common practice that this analysis is carried out through the *%right* indicator, that is the percentage of users in the calibration sample whose observed choices are given the maximum probability by the model. This index, very often reported, is somewhat meaningless if the number of alternatives is greater than two.

Really effective analysis can be carried out through the indicators proposed below:

- $\%_{clearly\ right}(t)$ percentage of users in the sample whose observed choices are given a probability greater than threshold t by the model;
- $\%_{clearly\ wrong}(t)$ percentage of users in the sample for whom the model gives a probability greater than threshold t to a choice alternative differing from the observed one;
- $\%_{unclear}(\mathrm{t}) = 100 - (\%_{clearly\ right} + \%_{clearly\ wrong})$ percentage of users for whom the model does not give a probability greater than threshold t to any choice.

The three indicators may be defined for each alternative and/or for all the alternatives available to users in the sample. Moreover, they can be plotted as the threshold changes, thus allowing the comparison between different models by comparing the diagrams themselves, or by measuring the difference between the areas below the diagrams. Indeed, integrating the areas under both curves, it is possible to estimate different indicators able to measure model effectiveness, and to appreciate differences between models and/or approaches.

As for aggregate indicators, the indicators applied to the calibration (hold-out) sample allow us to understand how the model reproduces (generalises) the observed choice scenario (to other choice scenarios).

Elasticity analysis

Elasticity analysis is a further approach, well-consolidated in the literature, helps understand how a model generalises to other choice scenarios. This is called elasticity analysis and is based on indicators obtained from partial derivatives (point elasticity) or finite differences (arc elasticity) of choice probabilities w.r.t. specific attributes.

The point direct elasticity of the choice probability for alternative r with respect to an infinitesimal variation in the jth attribute (x_{jr}) of its own utility function is defined as:

$$E_{r/j} = \left(\partial\, \mathrm{p}[r](\mathbf{X})/\partial x_{j,r}[r]\right) \cdot \left(x_{j,r}/\mathrm{p}[r]\right)$$

where $\mathbf{X}$ includes the vectors of attributes for all alternatives.

The point cross elasticity of the choice probability of alternative r with respect to an infinitesimal variation of the kth attribute, $x_{k,h}$, of the utility function of alternative h is defined as:

$$E_{r/kh} = (\partial\, \mathrm{p}[r](\mathbf{X})/\partial x_{k,h}) \cdot (x_{k,h}/\mathrm{p}[r])$$

The "arc" elasticity is calculated as the ratio of incremental ratios over an "arc" of the demand curve. In general, *direct elasticity, $E_{j,r}$*, may be defined as the percentage variation in choice probability of an alternative r divided by the percentage variation in the value of an attribute j of the systematic utility, $x_{j,r}$:

$$E_{r/j} = \Delta p[r]/\mathrm{p}[r] \cdot \Delta x_{j.r}/x_{j,r}$$

Analogously, *cross elasticity $E_{r/kh}$* is defined as the percentage variation in choice probability of an alternative r divided by the percentage variation in the value of an attribute k of another alternative h, $x_{k,h}$:

$$E_{r/kh} = \Delta p[r]/\mathrm{p}[r] \cdot \Delta x_{k,h}/x_{k,h}$$

A.2 Random utility models for route choice

Route choice represents one of the most complex characteristics of travel demand to observe, understand and simulate. Elements of complexity in other choice dimensions stem from the larger number of *degrees of freedom* characterising route choice:

– *Updating between one trip and the next: high day-to-day elasticity*.

Route choice is very elastic to the functional characteristics of the transport supply system and the experience of previous trips; hence it is an easily modifiable choice from one day to the next (between homogeneous reference periods).

– *Updating within the same trip: high within-day elasticity.*

  Within the same period the user may decide whether to depart or postpone his/her departure, and once the trip has begun, the user may decide to modify his/her (preventively chosen) route one or more times in the presence of changed traffic conditions that make a route deviation worth taking.

– *Effect of user heterogeneity.*

  Compared with other demand characteristics, in route choice there may be more significant user heterogeneity by virtue of the larger number of alternatives, the non-clear identification of the set of alternatives (choice-set), the variety of variables affecting choice, and the greater weight that psychological and cognitive factors assume in choice behaviour.

The above issues are assuming particular importance alongside the increasing development of new technologies (e.g. intelligent information systems) which require the simulation of increasingly realistic and disaggregate route choice behaviour.

The aim of this section is to schematise route choice behaviour by formalising an interpretative paradigm (a set of rules capable of describing the phenomenon) and a theoretical paradigm (formulating a theory or general principles to simulate individual phenomena) for simulating route choice issues in dynamic contexts.

A.2.1 Formalisation of an interpretative framework

In studying route choice behaviour, it is important to make the choice process explicit and how it is expressed in the time between homogeneous time periods and within each time period.

In general, the route choice process may be expressed as the process of the possible choice behaviours that a generic user may have:

(i) at the trip origin node (o),
(ii) at a diversion nodes (q),
(iii) at an information nodes (q_{inf}).

In this context, simulation of route choice behaviour may be studied in a day-to-day time context (between homogeneous reference periods) and/or in a within-day space–time context.

In the day-to-day context we are interested in the evolution of the choice process between one trip and another in the next day, i.e. whether and how the user modifies his/her choice process with the variation in experience accumulated on previous trips and/or in the presence of an information system.

In the within-day context we are interested in the evolution of the choice process within the same trip, i.e. whether and how the user modifies his/her choices once a trip has begun.

If we suppose a subdivision of the phenomenon into homogeneous periods (days), we may distinguish:

- the day on which the phenomenon begins (g^{0}),
- subsequent days (g^{+1}, g^{+2}, etc…), and
- possibly the day (g^{*}) on which the phenomenon reaches a steady-state condition.

To each of the above-mentioned contexts, different interpretative paradigms may be associated (e.g. holding or switching choices, etc….) and it is necessary to define theoretical paradigms and effective/realistic mathematical models.

By combining the day-to-day time dimension and the within-day dimension we may distinguish the following operational contexts and modelling issues (Table A.2).

In the following each of the afore mentioned issues are addressed.

A.2.2 Trip behaviour and alternatives in route choice modelling

As regards the trip behaviour we may distinguish:

(a) a completely *preventive* behaviour: route choice is carried out at the trip origin and the sequence of nodes to be crossed by the user is completely known and not

Table A.2 Operational contexts and modelling issues

When	Where		
	Origin node o	Diversion node q	Information node q_{inf}
g^{O}, g^{*}	(g^{O}, g^{*}, o) - type of trip behaviour - type of alternative - choice set - reaction to pre-trip information - holding choice	(g^{O},, g^{*}, q) or (g^{+}, q) - reaction to en-route information - switching choice in a "reference route" approach - holding choice in a "strategy approach"	(g^{O},g^{*}, q_{inf}) or (g^{+}, q_{inf}) - reaction to en-route information
g^{+}	(g^{+}, o) - reaction to pre-trip information - costs and choice updating		

modifiable by events or information which the user might encounter along the way;

(b) a mixed *preventive/adaptive* behaviour: the user extends his/her decisional process to the whole trip duration and the chosen route cannot be defined a priori.

The interpretative paradigm of the phenomenon for this choice dimension may be (i) not-behavioural or (ii) behavioural.

Following a non-behavioural approach, it is assumed that the user does not make a choice every time he/she travels, and the most reasonable theoretical paradigm must be sought within regressive-classification methods.

Under the behavioural approach, the user makes a choice which is sensitive to system functionality and can be modified from one day to the next if boundary conditions vary and/or if an information system exists. In this context, the user chooses from the alternatives available and can change his/her choice as the boundary conditions vary (e.g. existence of an information system and type of information supplied). In such a context it is reasonable to imagine a theoretical paradigm based on a binary choice model founded on random utility theory.

Once the interpretative paradigm on the type of behaviour has been chosen, the type of alternative should be defined.

In a completely *preventive* context the choice alternatives consist of routes connecting the origin–destination, whilst in a mixed *preventive/adaptive* context the user has the possibility to make route diversions. In this regard, two possible behaviours can be distinguished:

(a) *reference route* and
(b) *strategy*.

In the *reference route* approach, the user chooses a route preventively, then at specific nodes (diversion nodes) he/she considers the possibility of modifying his/her choice, taking account of the most efficient changes. Each of these changes will begin at the diversion node and will be characterised by network conditions which have either been directly experienced or exogenously supplied (information).

In the *strategy* approach, the user starts by choosing a set of routes and a strategy to adopt along the way on the basis of the functional characteristics encountered and/or the information supplied. The approaches are significantly different if threshold and non-compensatory behaviours are present (in particular cases they may be equivalent).

The type of alternative which users consider in their mind should be previously investigated and modelled. It may depend on the user characteristics, on the trip purpose, on the type of network, on the existing information systems and/or on the type of supplied services (e.g. transit, services of sharing, etc.). To this aim, classification methods may be adopted such as cluster analyses and or regressive classification methods based on Neural Networks models.

A.2.3 Choice set in route choice modelling

Modelling choice behaviour requires to defines the set of alternatives from which each user will choose (Prato and Bekhor, 2007; Bekhor et al., 2006; Cascetta et al., 2002).

The choice set depends on the user's experience and the information available. The information system, in particular, may either extend the perceived choice set, or reduce it to a few more efficient alternatives.

However, it may be hypothesized that the user considers all the elementary alternatives available, or he/she takes only part of them into account.

In the first case, the user considers all available alternatives (routes, reference routes or strategies), although the topological complexity of the network may generate alternatives which are scarcely realistic or consistent with a behavioural interpretation of the phenomenon.

In a deeper analysis three types of approaches may be adopted:

(a) Exhaustive, e.g. all the elementary routes without circuits are considered admissible. This approach may generate a significant computational complexity due to the overall large number of routes and due to the number of routes that may share same links, generating correlation among their perceived utilities. However, these issues may be addressed through *implicit route enumeration* algorithms and by adopting proper random utility models.

(b) Selective, e.g. applies heuristic behavioural rules to identify only a subset of the elementary routes (examples in Cascetta, 2009). In general the selective approach requires *explicit route* enumeration between each *O-D* pair, and usually applies a combination of criteria.

(c) Modelling, e.g. if the set of admissible alternatives is defined by a further behavioural model which simulates perception/availability of the alternatives and supplies the probability of each route belonging to the set of alternatives perceived by a generic user. In the literature two types of modelling approaches are consolidated: *explicit* and *mixed*. The *explicit* solution entails specification of a random utility model which allows us to assign to each non-empty sub-set of the alternatives choice set a probability of being the actual choice set of the generic user (see Section A.1.1).

The *mixed* solution tackles simulation of the choice set and alternative choice at the same time. The problem may be formalised in the framework of random utility theory by introducing an attribute that measures the availability/perception of the within the utility function (see Section A.1.1).

A.2.4 Holding choice in route choice modelling

Depending on his/her previous decisions the user chooses the alternative may be a route, a reference route or a strategy. Holding choices may occur before the trip, but also at a diversion node.

The most adopted paradigm in holding choices relies on the random utility theory. In this case, the specification of a route choice model requires, as usual, the definition of the attributes in the systematic utility function and of the joint probability distribution of random residuals. It is usually assumed that the variables influencing route choice are disutility attributes that enter in the utility function, and may be expressed as a linear combination of user's socio-economic characteristics, of performance attributes which may be *additive* (the sum of the corresponding arc performance attributes) or might include some attributes that cannot be obtained as the sum of arc variables (*non-additive attributes*).

The simplest and first proposed route choice model is the *deterministic utility* model, which can be seen as a special case of a random utility model in which the variance of the residuals ξ_r is assumed to be equal to zero, thus $U_r = v_r$. In this case, a route r can be used only if, from among the set of alternative routes, its utility v_r is max:

$$\mathrm{p}[r] > 0 \Rightarrow v_r \geq v_h \forall h \neq r\, r, h \in R$$

In particular, the deterministic utility model does not provide a unique route choice probability vector, except when there is a unique minimum cost route. Although deterministic choice models are less realistic than probabilistic models, for computational reasons they are often applied to very large networks with implicit route enumeration. See remarks at the end of the chapter for further details.

In the following, the most adopted route choice RUMs are introduced.

Multinomial Logit and Logit-based model

The easiest probabilistic choice models used to calculate route choice probability is the multinomial Logit model (MNL), however the assumption of Independence of irrelevant alternatives (I.I.A.) property may be unrealistic when the routes in the choice set overlap (shared links). At the same time, the hypothesis of homoscedasticity may be quite unrealistic in some choice contexts. The choice probability expression of the resulting model has been introduced in Section A.1.3.

To reduce the effects of the I.I.A. property, the MNL model should be used with an explicit route enumeration method that eliminates highly overlapping routes.

The MNL is usually implemented by means of the Dial's algorithm (Dial, 1971). However, Dial's algorithm has strong limitations, other than the limitations of the MNL itself, indeed, it restricts the choice set to efficient routes only and, if single-step formulation is used, it introduces the unrealistic assumption of equal variance parameter for each O-D pair.

To overcome the problems deriving from the Logit I.I.A. property, a modification to the MNL route choice model is widely adopted in practical application and several commercial software.

The modification consists in introducing a negative term (*commonality factor*, CF_r) in the systematic utility specification of each alternative. It reduces the systematic utility of a route according to its degree of overlap with other routes. Indeed, CF_r

is inversely proportional to route r's degree of independence from other routes, and is equal to zero if no other route shares links with route r.

The resulting model, usually called *C-Logit* model (Cascetta et al., 1996; Zhou et al., 2012), reduces the probability of choosing overlapping routes and increases the probability of choosing non-overlapping routes. It has the following specification:

$$p(r) = \exp[(v_r/\theta) - CF_r]/\sum_{h\in R} \exp[(v_h/\theta) - CF_r]$$

The *commonality factor* may be specified in various ways.

The most used expression reported in the following allows using implicit route enumeration algorithms similar to Dial's.

$$CF_r = \ln\left(1 + \sum_{h\neq r} z_{hr}/(z_h \cdot z_r)^{1/2}\right) \mathrm{h} \in \mathrm{R}$$

where z_r is an additive route cost attribute of route r. It is *different* from the actual route cost w_r in order to satisfy the random utility model's property of additivity; z_h, is the additive route cost attribute of route h belonging to the choice-set R; and z_{hr} is the cumulative value of the cost attribute over the links belonging the two routes h and r.

The previous specification does not take into account the relative weight of shared links in the overall route cost. Therefore, different specifications of the communality factor may be adopted.

One of the most adopted defines a communality factor larger for a route whose shared links contribute a larger fraction to its total length or cost:

$$CF_r = \sum_{a\in r} \mathrm{q}_{ar} \cdot \ln(N_a)$$

where a is the generic arc belonging to route r, q_{ar} is the weight of arc a in route r which may be expressed as the ratio (s_a/z_r) with s_a being the arc, a, cost attribute coherent with the used route cost attribute, and N_a is the number of routes of choice-set R pair using arc a.

Another possible formulation may be obtained by expliciting the dependence of the communality factor on the cost of its non-shared links. In this case the commonality factors of two routes increases as the percentage of common attributes cost, with respect to the total one, increases:

$$CF_r = \ln\left[1 + \sum_{h\neq r} \left(z_{hr}/(z_h \cdot z_r)^2\right) \cdot (z_r - z_{hr}/z_h - z_{hr})\right]$$

Multinomial Weibit model

The Multinomial Weibit model (MNW) has been applied to route choice issues by Castillo et al. (2008) and Kitthamkesorn and Chen (2014), and it adopts the multiplicative random utility model formulation with the Weibull distribution as the random error term.

The Weibull distribution, as the Gumbel distribution, is a particular case of the more general Generalised Extreme Value distribution. The latter, in general, depends on three parameters: the systematic utility (v_r), the scale parameter representing the variance (θ) and a parameter γ. Whilst the Gumbel distribution is obtained for $\gamma \to 0$, the Weibull distribution is obtained when the parameter $\gamma < 0$.

Gammit formulation has been predominantly implemented to route choice models and with systematic utility functions that coincide with route path cost, w_r.

The MNW utility function in usually written in terms of route costs as follows:

$$u_r = \left(w_r - \underline{w_r}\right)^{\theta} \cdot \xi_r \forall r \in R$$

where w_r is the route r cost; ξ_r is the random residual independently Weibull distributed; θ is the scale parameter and is related to the route perception variance with $\theta > 0$; λ is the location parameter which identifies the lower bound of route perceived travel cost between the considered origin–destination pair, with $0 < w_r \leq w_r$;

The choice probability expression of the resulting model, is

$$\mathrm{p}(r/R) = (w_r - \lambda)^{-\theta} / \sum_{h \in R} (w_h - \lambda)^{-\theta}$$

The resulting model is a heteroscedastic model, but not able to reproduce any cross-correlation among alternatives perceived utility.

Multinomial Probit model

An effective alternative to the multinomial Logit model is represented by the multinomial Probit model (MNP).

Although the general formulation of the MNP model may be adopted, the most widely used specification is usually obtained by applying the Factor Analytic approach to the route choice context.

The factor analytic approach assumes a specific structure underlying the random residuals which allows reducing the number of unknown variance–covariance matrix elements, and it is particularly suitable for modelling route choice under reasonable assumptions.

Within this framework, alternatives random residual (the vector of random residuals) may be expressed as a linear function of independent standard normal variables (of a vector $\boldsymbol{\zeta}$), $\boldsymbol{\xi} = \mathbf{F}\,\boldsymbol{\zeta}$, which in scalar form may be expressed as:

$$\xi_r = \sum_{t \in T} f_{rt}\, \zeta_t$$

where ζ_t is the tth identical and independent standard normal random variables which composes the vector $\boldsymbol{\zeta} \sim \mathrm{MVN}(\mathbf{0}, \mathbf{I})$ vector of identical and independent standard normal random variables; f_{rt} is the tth loading factor belonging to the matrix ($\mathbf{F}$) of factor "loadings" which maps the vector of standard normal random variables ($\boldsymbol{\zeta}$) to the vector of random residuals ($\boldsymbol{\xi}$).

The extension of the factor analytic approach to the route choice problem assumes that the variance of the random residuals is proportional to an additive route cost attribute z_r (*different* from the actual route cost in order to satisfy the random utility model's property of additivity), and that the covariance of the residuals of two routes is proportional to the cumulative value of the cost attribute over the links that are shared by the two routes (z_{rh}):

$$\mathrm{var}[\xi_r] = \gamma\, z_r \; r \in R$$

$$\mathrm{cov}\left[\xi_{r,}\xi_h\right] = \gamma\, z_{rh} \; h, r \in R$$

Since each route perceived utility may be expressed in terms of the perceived utility of the arcs belonging to the route, the random residual of each route may be expressed as the sum of arc random residuals.

Assuming each arc random residual independently normally distributed with variance proportional (through the parameter γ) to the value of the arc performance variable of which z_r is the sum, each route random residual may be expressed as a sum of univariate normal random residuals that may associated to each arc belonging to the route (η_a).

In other words, the formulation of the Probit model can be obtained by applying the Factor Analytic approach to the route choice context assuming a matrix of factor "loadings" given by the arc-route incidence matrix and the parameter γ.

$$\xi_r = U_r - v_r = \sum\nolimits_a \delta_{ak}\,(U_a - v_a) = \sum\nolimits_a \delta_{ak}\,\eta a$$

Which may be expressed in vectorial form as.

$$\boldsymbol{\xi} = \mathbf{F}\boldsymbol{\zeta} = \mathbf{B}^{\mathrm{T}}\boldsymbol{\eta}$$

where $\boldsymbol{\xi}$ is the vector of multivariate normal distributed route random residuals, $\boldsymbol{\xi} \sim$ MVN($\mathbf{0}, \boldsymbol{\Sigma}$); $\mathbf{B}$ is the arc-route incidence matrix; $\boldsymbol{\eta}$ is the vector of independent normal distributed arc random residuals, $\boldsymbol{\eta} \sim$ MVN($\mathbf{0}$, $\boldsymbol{\Sigma}_\eta$); $\boldsymbol{\zeta}$ is the vector of i.i.d. standard normal random variables, $\boldsymbol{\zeta} \sim$ MVN ($\mathbf{0}, \mathbf{I}$); $\mathbf{F}$ is the matrix that maps the random vector $\boldsymbol{\zeta}$ into route choice random residuals $\boldsymbol{\xi}$.

The Probit model allows handling perceived utility correlation (route overlapping) and particularly suitable for applications with exhaustive route generation (implicit enumeration). Obviously, the choice probabilities cannot be expressed in closed form, thus simulation method should be adopted, but Factor analytic Probit formulation allow the use of algorithms that are based on Monte Carlo simulation.

Finally, if the hypothesis of a unique variance parameter (γ) may be considered a strong assumption, on the other hand no effective alternative solutions have been proposed in literature.

Multinomial Gammit model

The Gammit model is obtained assuming that perceived utilities are jointly distributed as a non-negative *shifted* MultiVariate Gamma random variable, with mean equal to the systematic utilities and variance–covariance matrix **Σ**.

Gammit formulation has been predominantly implemented to route choice models (Cantarella and Binetti, 2002) and with systematic utility functions that coincide with route cost, w_r.

The Gammit formulation to route choice, once defined the systematic utility for each alternative, can be specified as follows.

Let

c_a be the arc cost of arc a.

$\underline{c_a}$ the reference cost on arc a (for instance zero-flow cost), assumed not greater than the arc cost: $0 < \underline{c_a} \leq c_a$.

The perceived utility of arc a, U_a, is assumed distributed (independently of the perceived disutility of any other arc) as a non-negative shifted Gamma variable with mean given by the arc cost (c_a), variance proportional to the reference arc cost ($\sigma_a = \theta \underline{c_a}$) and shifting factor given by the difference between the arc cost and the reference cost (c_a - $\underline{c_a}$):

$$U_a \sim \left(c_a - \underline{c_a}\right) + Gamma\left(\alpha_a = \underline{c_a}/\theta, \beta = \theta\right)$$

In other words, the arc perceived utility is the sum of a non-negative deterministic term depending on arc flows and non-negative stochastic term independent from arc flows. The assumption on link reference cost yields that the corresponding reference cost on route r, $\underline{w_r}$, is strictly positive and not greater than the systematic utility cost, w_r. Thus, the perceived utility on route r is marginally distributed as a non-negative shifted Gamma variable.

$$U_r \sim \left(w_r - \underline{w_r}\right) + Gamma\left(\alpha_r = \underline{w_r}/\theta, \underline{\beta} = \theta\right)$$

The described specification leads to an additive choice model if the variance parameter θ and the reference arc costs do not vary with arc costs.

The resulting model presents a route perceived utility vector distributed as a non-negative shifted MultiVariate Gamma with variance–covariance matrix proportional to route reference cost through the parameter θ.

In conclusion, the Gammit model allows simulating heteroscedasticity and taking into account the covariance between overlapping routes. Indeed, the variance for route r is proportional to the reference cost $\underline{w_r}$, whilst the covariance between two routes is proportional to the reference costs of arc shared by the two routes.

Gammit path choice probabilities cannot be expressed in a closed form, and their computation requires simulation techniques such as MonteCarlo method.

A.2.5 Updating choices in route choice modelling

On the generic day $g \neq g\text{-}1$ the user elaborates his/her experience from the previous day, past experience and any information which is supplied exogenously, and redefines his/her own trip behaviours. In general an *indifferent* behaviour and an *active* behaviour may be distinguished.

Under the assumption of *indifferent* behaviour, the user does not question the choice made the previous day in the same context (purpose, OD, time slice) and hence does not update the previous day's behaviour or the characteristic variables, such as level of service attribute, on which the previous day's choice behaviour was based. Under this hypothesis, the user repeats his/her holding choice according to the paradigm proposed for day g^0; no evolution occurs between one period and the next unless there is an information system.

In the case of *active* behaviour, the user analyses and elaborates experiences during the trip on previous days (from g^0 to $g\text{-}1$), receives information on the current day and evaluates whether to behaviour in the same way as day $g\text{-}1$.

In general, a choice updating process can be defined, and in which we distinguish:

(a) the analysis phase and
(b) the choice phase.

In the analysis phase the user analyses the choice already made the day before by comparing experience and prediction and the choice by comparing personal experience with the possible available alternatives perceived (or supplied by an information system). Then, in the choice phase, the user decides whether to reconsider all the behaviours which led to the final choice, or only some them and, finally, takes the choice for the next period.

In specifying a theoretical paradigm, there are two major issues:

(i) simulating cost updating, in other words, simulating how experience and information on costs for the previous and current days affect current choices and
(ii) simulating the phenomenon of choice updating, that is, how choices in a day are affected by choices made on previous days.

Costs and choices may be updated using models that simulate choice behaviour in day g (in a certain cost configuration), cost updating models and choice updating models (for further details, see Cascetta, 2009).

The cost updating model simulates the way the predicted utilities vector are affected by the costs on previous days. Costs relative to the previous days may be the result of direct experience or they may be representative of information obtained before undertaking the trip.

A simple example of *cost updating model* is defined by means of an exponential filter in which utility predicted at day g is expressed by a convex combination of predicted utility on the previous day *g-1*, v^{g-1}, and of the opposite of route costs incurred on day *g*-1, $-w_r^{g-1}$:

$$v_r{}^g = -\beta\, w_r{}^{g-1} + (1-\beta)\, u_r{}^{g-1} \forall od$$

where $\beta \in]0,1]$ is the weight attributed by users to costs that are occurred on day *g*-1; with $\beta = 1$ the predicted cost coincides with the costs incurred on day *g*-*1*, and the costs of the days even before do not affect user behaviour.

As regards the specification of the *choice updating model*, it is necessary to simulate how daily choices are influenced by previous days' choices. The most general approach may be expressed by a matrix, known as the matrix of conditional choices, with as many rows and columns as there are routes and whose elements define the percentage of conditional route choice, that is, the percentage of users choosing route *r* on day *g*, having chosen route *r* on the previous day, *g*-1.

As with cost updating, an efficient model of efficient choice updating may be defined by means of an exponential filter. It hypothesizes that every day a fraction of users $(1-\alpha)$ repeats the choices of the previous day, while the remaining part (α) chooses independently of the choice actually made the previous day. At day *g* let $p_r^g \in]0,1]$ be the probability of a user choosing route *r* having reconsidered the choice made the previous day, $g-1$, and q_r^g be the probability of the user choosing route *r* on day *g*, it yields:

$$q_r{}^g = \alpha\, p_r{}^g + (1-\alpha) \cdot q_r{}^{g-1}$$

In this approach $\alpha \in]0,1]$ is the probability of a user reconsidering the choice made the day before, *g*-*1*. It may be estimated using regressive non-behavioural models or behavioural models. Probabilities p_r^g may be simulated with a model derived from random utility theory as a function of utilities predicted for day *g*.

A.2.6 Switching choices in route choice modelling

The choice behaviours are termed *switching* (Ben-Akiva and Morikawa, 1990) when the choice occurs over an alternative which can be considered as a reference alternative and treated differently from others (e.g. the reference route, previous day's choices, etc…).

A possible modelling solution consists of the implementation of random utility models, possibly combined with a bounded rationality approach that leads to stochastic thresholds.

Switching choices may occur at the origin of the trip (node *o*) or at a diversion node (node *q*). In the former case the user may switch from the pre-trip choice made in the previous day (*g*-*1*), in the latter case the user may switch from his/her pre-trip choice to different solution.

In defining the utility function we consider as an attribute the difference between preventive route travel time (from origin *o*—or node *q*—to destination *d*) and the travel time of the best route connecting *o* (*q*) and *d*. The problem consists in estimating the probability of users changing his/her pre-trip choice. This probability, below indicated as the probability of switching, may be expressed as the probability that the time saved exceeds the maximum of the two stochastic thresholds (INT_j):

- *threshold 1,* which is the threshold below which the user does not modify his/her choice and
- *threshold 2*, which is the of total travel time beyond which the user is believed to consider the hypothesis of changing his/her preventive choice.

User makes his/her route choice preventively to travel from origin *o*, or from diversion node *q*, to destination *d*. Once the choice has been made, a route *r* (in a *reference route* approach) or a strategy I_R (in a *strategy* approach) is identified (each with non-zero probability).

Hence it is possible to define with regard to a time interval *t* (subsequently neglected):

$UTTP_{r,od}$ is the utility associated to route *r* (or to strategy I_R) relative to origin *o* (ore node *q*) and destination *d*;
$UTTM_{s,od}$ is the utility associated to the shortest route *s* between node origin *o* (or node *q)* and destination *d*.

The difference between the two costs gives the saving obtainable from adopting the minimum route instead of route *r* or strategy I_R: $UTTS_{r,od} = UTTP_{r,od} - UTTM_{s,od}$.

The user changes his/her route if, and only if, the saving in terms of perceived utility exceeds an indifference threshold (ΔINT_{od}) which may be deterministic or stochastic.

Following a deterministic approach the phenomenon of switching may be schematised by a binary variable, $\phi_{r,od}$, which is equal to 1 if the user changes his/her preventive choice, 0 otherwise:

$$\phi_{r,od} = 0 \text{ if } 0 \leq UTTS_{r,od} \leq \Delta INT_{od}$$

$$\phi_{r,od} = 1 \text{ otherwise}$$

Assuming a stochastic approach, diversion probabilities are thus defined:

$$\Pr\left[\phi_{r,od} = 0\right] = \Pr\left[0 \leq \mathrm{UTTS_{r,od}} \leq \Delta\ \mathrm{INT_{od}}\right] \Rightarrow \mathrm{p}\left[\phi_{\mathrm{r,od}} = 1\right] = 1 - \mathrm{p}\left[\phi_{\mathrm{r,od}} = 0\right]$$

In both cases it is necessary to make both $UTTS_{r,j\grave{i}od}$ and ΔINT_{od} explicit.

As regards the estimate of $UTTS_{r,od}$ the problem basically entails the estimate of the maximum utility (or minimum route cost), $UTTM_{s,od}$, from origin *o* (or node *q*) to destination *d*. $UTTM_{od}$ in turn may be the maximum utility (minimum route cost) perceived by the user at node *o* (or node *q*), possibly different from the preventive one.

In the first case, the user observes and elaborates the traffic flow and re-elaborates his/her trip costs. Specification of such an attribute requires an easy implementation of a model of travel costs. The main difficulties concern the need to have a significant number of test observations and reliable measurements of vehicle flow characteristics at the time of the interview. Unlike many studies developed in simulation laboratories, in this case it is hard to reproduce in a virtual environment the perturbations and sensations that can be perceived in a real context.

Under the assumption of exogenously supplied route costs, the problem may be seen as one of minimum route cost estimation, which may be expressed by one of two approaches:

(a) *instantaneous*: according to the transportation system characteristics at the moment of the switching choice and
(b) *predictive*: based on forecasts of network future conditions so that it is congruent with the cost that the user will encounter along the way.

As already specified above, *instantaneous information* is based on the network flow configuration at the moment when the user moves from origin o, or arrives at diversion node q. In this case the user is supplied with instantaneous information which does not predict system evolution over time. The benefit lies in the greater simplicity in calculating level of service attributes; the main drawback lies in the incongruence of information against the costs that the user will actually encounter along the way. Non-congruent information may cause indifference to information, or may render the day-to-day evolution of the system highly unstable, thereby distancing the system from equilibrium. In addition, this kind of information may cause the user wander through the network without reaching intended destination. For all these considerations the instantaneous approach is not much considered in literature.

Predictive information may be obtained by simulating: (a) the evolution of the current system without route diversions, (b) evolution of the system allowing for possible diversions. The first hypothesis, undoubtedly less computationally burdensome, leads to costs which are not necessarily congruent with the costs that the user will encounter up to destination. The second hypothesis, at least in theory, guarantees congruence between the costs supplied to the diversion nodes and costs which the user will encounter on the network. The computational burden of such a solution is not banal, especially in the presence of many diversion possibilities, above all in light of the need to ensure information in real time.

As regards the estimate of the interval of indifference (ΔINT_{od}), it may be represented as a deterministic variable or as a random variable. In the former case, the upper limit of the above interval may be estimated through test observations and consolidated statistical techniques. In the latter case ΔINT_{od} is assumed characterised by a known distribution function with parameters that can be estimated using the usual statistical inference techniques. The latter approach, which is definitely preferable, allows for the heterogeneity among the perceptions of various users (cross-sectional heteroscedasticity), the heterogeneity among the various perceptions of the same user due to explicitly not easy to simulate factors (longitudinal heteroscedasticity) and some estimation errors of level-of-service attributes which combine to define travel costs. Generally, the band of indifference may be expressed as the maximum of two deterministic or random variables:

$$\Delta INT_{od} = \max(\eta_{od}\,TTP_{od}, \pi_{od})$$

where η_{od} is the percentage of the chosen travel cost TVP_{od} of the route which the user believes significant to change his/her choice. π_{od} indicates the minimum time that user expects to save following a change in his/her preventively chosen route.

In practice the user, to be able to change his/her preventive choice, wishes to perceive a time saving greater than, or equal to, part of the current travel time and, at any rate, no less than a minimum threshold.

In the case of the random interval of indifference, it is reasonable to assume η_{od} and π_{od} as two random variables, characterised by a systematic component and by a random component:

$$\eta_{od} = \phi(\mathbf{SE}, \mathbf{LoS}_{\mathrm{od}}, \boldsymbol{\theta}) + \xi_{\mathrm{od}}$$

$$\pi_{od} = \varphi(\mathbf{SE}, \mathbf{LoS}_{\mathrm{od}}, \boldsymbol{\theta}) + \zeta_{od}$$

where $\boldsymbol{SE}$ is a vector of socioeconomic attributes; $\boldsymbol{LoS}_{od}$ is the vector of level-of-service attributes concerning the pre-trip choice; $\boldsymbol{\theta}$ is the vector of parameters to be estimated of functions ϕ and φ; and ξ_{od} and ζ_{od} are the random components.

Also in this case, it may be assumed that the probabilities of diversion are independent or dependent among themselves. In the latter case the probabilities of diversion may be expressed as the probability of diversion conditional upon have made a diversion at the previous node, that is, have already affected a diversion during the current trip.

The problem may be tackled by a modelling approach, specifying models that take into account the correlation between choices made at different times (e.g. Probit or Mixed-Logit).

A.2.7 Diversion choices in route choice modelling

The Analysis of behaviour at the diversion node is important within a mixed preventive/adaptive approach.

On the basis of the trip behaviour and the type of alternative (*route*, *strategy or reference route*), the choice set (*exhaustive* vs. *selective*) and the choice made (*route* or *strategy chosen*), the user begins his/her trip and starts to acquire information from experience or from the information system if any.

Users compare what they know from past experience with what they observe (in real time) and with what they receive, and begin to build expectations regarding the network functional characteristics and, especially, those of the reference route or strategy. Downstream of this process, once a possible diversion node has been reached, users decide whether to change their own reference route, or choose how to continue the trip within the preventively chosen strategy.

In the *strategy approach* it is supposed that the user, at each diversion node belonging to the strategy, chooses which arc (or sequence of arcs up to the next diversion node) to follow in continuing his/her trip. Such decisional behaviour may be schematised as a holding choice between a set of alternatives known a priori whose costs may differ from those perceived at the beginning of the trip both in terms of values and quality.

In the *reference route* approach, there are two choice dimensions to be revisited:

(a) the definition of the choice set and
(b) the choice of a new reference route.

The diversion node becomes the new origin and a behaviour similar to that described in the previous sections may be assumed. If information systems exist, choice behaviour may be strongly affected as the type of information varies.

In the presence of *descriptive information* (e.g. travel time, congestion rates, etc.), users may choose among the available route from the diversion node to the final destination. In this case a holding model can be adopted, with the only difference that each user will have a better knowledge of the current network conditions and will choose combining past experience and personal forecast of route costs. It can be expected a less dispersed behaviour, since the users may have a reduced (or more crisp) choice-set and more accurate information on network conditions. In this case, a holding choice model should be specifically calibrated, since different dispersion parameters are expected.

In the presence of *prescriptive* (or mixed) *information*, users may choose to comply or not comply with the proposed solutions, if alternative to the *reference route*. To this aim two approaches may be pursued: (i) with complete rationality, (ii) with bounded rationality. The complete rationality approach is founded on the utilitarian paradigm, the same adopted for the preventive choice behaviour. The bounded rationality approach assumes a behavioural paradigm for which the user does not change his/her preventively chosen alternative as long as the variation in perceived utility stays within an interval of indifference.

Both types of behaviour are significantly affected by the part of the trip already undertaken and according to the diversion behaviours adopted hitherto.

A.2.8 Reaction to information in route choice modelling

Simulation of the degree of compliance to the supplied information is important for realistic simulation of route choice behaviour in the presence of an information system. As the type of information varies, we may distinguish three approaches to modelling compliance to information as described in the following.

Explicit model of adaptation to information

In the presence of prescriptive information, the approach simulates the user's compliance to the information (that is, the probability of complying), as binomial choice behaviour, based on random utility theory and in which the utility of the alternative of not adapting is assumed zero. Under this assumption the problem may be formulated as follows:

$$\Pr[\delta=1]=\Pr[U_c>0]=\Pr[\xi_c>v_c]$$

where δ is a binary variable with the value of 1 if the user follows the indication supplied; $\Pr[\delta=1]$ is the probability to comply to the information, also called compliance rate; U_c is the utility perceived to comply to the information; $v_c = v_c(SE, LoS)$: is the systematic utility associated to the alternative, SE being socioeconomic variables and LoS the level of service variables; and ξ_c is the associated random residual.

Several modelling solutions may be adopted, which must in any case take into account that the phenomenon is a succession of correlated choices over time (day-to-day) and in space (with respect to possible diversion nodes), and that there may be relevant behavioural dispersion.

If user choice perceived utilities can be assumed to be independent of one another over time and space, the most effective modelling solution is a binomial Logit model.

To take into account the correlation among different day and/or different diversion (decision) nodes, it is necessary to interpret the phenomenon according to a process of choice from a discrete number of alternatives which is no longer binary. Let us suppose that each diversion (origin) node is a decision node q and that each node has a time index representing the generic day g. At each node (q), at day g, the user associates a perceived utility $u_c^{q,g}$ and makes a binary choice. The choice probability is a probability conditional upon the choices already made at the other decision nodes and/or on previous days. The problem may be also approached by introducing some error terms (Probit or Mixed-Logit model) which allow simulation of correlations between utilities of the same node but referring to different days, or between utilities of different nodes but crossed on the same day.

An alternative approach may be based on the theoretical paradigm of bounded rationality and on switching models.

Implicit model of adaptation to information

In the presence of prescriptive user information, reaction to information may be simulated as a holding choice introducing information within systematic utility functions.

Let:

- q be the decision node (origin or diversion node);
- R_q be the set of possible alternatives to reach destination d starting from q (the choice set may be exogenously supplied and/or defined by the user);
- ch_q be the pre-trip chosen alternative;
- b_q be the best alternative (possibly supplied by an information system), it may be equal to ch_q; and
- h_q be one of the remaining choice alternatives ($h_q \neq b_{q,}\ ch_q \in R_q$).

The user's behaviour can be schematised as a choice between: alternative b_q, alternative ch_q and the remaining alternatives h_q.

Following an approach based on random utility theory, to each alternative a perceived utility function may be associated:

$$U_{hq} = v_{hq} + \xi_{hq} \forall h_q \neq b_q, ch_q$$

$$U_{bq} = v_{bq} + \xi_{bq}$$

$$U_{chq} = v_{chq} + \xi_{chq}$$

Although different modelling solutions could be hypothesized, the most effective is the Mixed-logit model. Since alternative b_q is supplied by an information system, it is considered different by the user and it should be modelled differently. This can be schematised by introducing a component of utility (latent variable), U_{bq2}, in formulating the perceived utility function of alternative b_q. Such component measures the degree of user adaptation to the information, if zero there is no reason why the information supplied should increase the utility of the alternative indicated. Hence:

$$U_{hq} = v_{hq} + \xi_{hq} \forall h_q \neq b_q, ch_q$$

$$U_{bq} = U_{1,bq} + U_{2,bq} = v_{1,bq} + \xi_{bq} + v_{2,bq} + \eta_{bq}$$

$$U_{chq} = v_{chq} + \xi_{cq}$$

where $U_{2,bq}$ is sum of systematic utility, $v_{2,bq}$, and a random residual, η_{bq}, which in turn permits us to simulate the heteroscedasticity of utility U_{bq} in any case. While the systematic utilities ($v_{1,bq}$, v_{hq}, v_{chq}) express the typical characteristics that affect the route choice (e.g. socio-economic, level of service variables). The systematic utility $v_{2,bq}$, in turn, may be expressed as a linear combination of variables representing the degree of *adaptation* to information.

The degree of user adaptation to the information should be calibrated, four types of variables may be reasonably used:

- **(i)** variables representing past user experience (e.g. observed delay on previous days following information or otherwise);
- **(ii)** variables representing the quality/reliability of information (e.g. dispersion of travel time supplied);
- **(iii)** variables representing the network congestion level (usually represented by aggregate measures for crossed links or by symbolic variables calibrated ad hoc);
- **(iv)** variables representing the benefits to be gained by adaptation to information (e.g. travel time saving supplied together with the information but also total travel time);
- **(v)** socio-economic variables.

As regards the probability distribution functions of the random residuals (ξ_{bq}, ξ_{chq}, ξ_{hq}), they could be differently distributed in order to simulate the heteroscedasticities among the alternative's utilities.

At this point it may be worth studying the two possible situations separately:
($b_q = ch_q$), if the best route coincides with the route currently chosen;
($b_q \neq ch_q$), if the two routes do not coincide.

In the former case ($b_q = ch_q$) it can be useful to introduce an error term which simulates heteroscedasticity between the two groups of alternatives (b_q and h_q). In the latter case ($b_q \neq ch_q$), by the same token, it could be worth introducing an error term which simulates the correlation between the utilities of alternatives ch_q and h_q, in that they are not affected by the adaptation phenomenon.

Together with the adaptation phenomenon, the *inertia* phenomenon can be simulated implicitly, i.e. what induces the user not to change his/her preventive choice.

Since *inertia* affects the choice, it is correct to introduce into the perceived utility of the preventively chosen alternative a latent variable that simulates the dis/utility that the user attributes to abandoning his/her choice. As with the above case, a Mixed-Logit model seems to be the best option, and the perceived utilities have similar formulation.

If ($b_q \neq ch_q$), it is hypothesized that the utilities of alternatives $h_q \neq (b_q, ch_q)$ and ch_q, by means of an error term τ_q, are correlated as they are not characterised by information and lastly, that the utilities of alternatives $h_q \neq (b_q, ch_q)$ and b_q are correlated by means of the error term ω_q since they are not characterised by inertia:

$$U_{hq} = v_{hq} + \xi_{hq} + \omega_q + \tau_q \forall h_q \neq b_q, ch_q$$

$$U_{bq} = v_{bq} + \xi_{bq} + \omega_q$$

$$U_{chq} = U_{1,chq} + U_{2,chq} = v_{1,chq} + \xi_{chq} + v_{2,chq} + \eta_{chq} + \tau_q$$

Also in this case, it is possible to assume the same four types of variables described above for compliance models and make the same considerations on the residuals and the possible correlations existing.

Explicit model of the cognitive process to acquire and use the information

The process underlying route choice in the presence of an information systems may be expressed as follows:

(a) the user acquires the information,
(b) the user decides to use it, and
(c) the user chooses on the basis of the acquisition process and the use of information.

While issue (c) may be addressed through the modelling approaches introduced before, issues (a) and (b) are worth of interest in the presence of different information systems.

Let be:

φ is the generic source of information (e.g. variable message panels) and that I_φ is the set of all the available sources of information;
$U_{a\varphi}$ is the perceived utility associated to acquiring information of type φ;
$U_{r\varphi}$ is the perceived utility of using the information of type φ;
a_φ is a binary variable which has the value of 1 if the user obtains information from source φ, 0 otherwise;
r_φ is a binary variable which has the value of 1 if the user refers (thus use) to the information received from source φ in the process of choosing his/her route, 0 otherwise.

The user's decisional process can be schematised assuming a hierarchical structure in which the user first chooses to acquire information from the source φ, then decides to refer to it (or use). In this case, the phenomenon may be modelled as the probability of acquiring information multiplied by the probability of using it, conditional upon having acquired it.

$$\Pr[a_\varphi, r_\varphi] = \mathrm{p}[a_\varphi]\cdot \mathrm{p}[r_\varphi/a_\varphi]$$

Adopting the Utility Theory, the user uses information φ that he/she receives ($a_\varphi = 1$) and uses it ($r_\varphi = 1$), if perceived utilities $U_{a\varphi}$ and $U_{r\varphi}$ exceed two threshold values $T_{a\varphi}$ and $T_{r\varphi}$:

$\rightarrow$ acquire if $U_{a\varphi} \geq T_{a\varphi}$

$\rightarrow$ refer to if $U_{r\varphi} \geq T_{r\varphi}$

where $U_{a\varphi}$ is the perceived utility to acquire information from source φ; $U_{r\varphi}$ is the perceived utility to refer to information from source φ; $T_{a\varphi}$ is the threshold relative to the acquisition process and may be expressed as a random variable consisting of a systematic part $\tau_{a\varphi}$ and a random residual $\eta_{a\varphi}$; and $T_{r\varphi}$ is the threshold relative to the process of information use, and may be expressed as a random variable consisting of a systematic part $\tau_{r\varphi}$ and a random residual $\zeta_{r\varphi}$.

Thus $\Pr[a_\varphi, r_\varphi]$ may be written as follows:

$$\Pr[a_\varphi] = \Pr[U_{a\varphi} \geq T_{a\varphi}] = \Pr[V_{a\varphi} + \xi_{a\varphi} \geq \tau_{a\varphi} + \eta_{a\varphi}] = \Pr[V_{a\varphi} - \tau_{a\varphi} \geq \eta_{a\varphi} - \xi_{a\varphi}$$

$$\Pr[r_\varphi/a_\varphi] = \Pr[U_{r\varphi} \geq T_{r\varphi}] = \Pr[V_{r\varphi} + \xi_{r\varphi} \geq \tau_{r\varphi} + \zeta_{r\varphi}] = \Pr[V_{r\varphi} - \tau_{r\varphi} \geq \zeta_r - \xi_{r\varphi}]$$

The perceived utilities may be expressed as follows:

$$U_{a\varphi} = \Sigma_{\mathrm{m}}\left(\psi_{m\varphi}\cdot\lambda_m\right) + \Sigma_{\mathrm{p}}\left(\gamma_{p\varphi}\cdot trip_{p\varphi}\right) + \xi_{a\varphi}$$

$$U_{r\varphi} = \Sigma_{\mathrm{n}}\left(\gamma_{n\varphi}\cdot trip_{n\varphi}\right) + \Sigma_{\mathrm{q}}\left(\delta_{q\varphi}\cdot m_{q\varphi}\right) + \xi_{r\varphi}$$

where λ_m is mth latent variable that measure cognitive involvement and the capacity to process information (see Section A.1.2); $trip_{n(or\ p)}$ is the nth (pth) variable characterising the trip made up to that moment or the user's past experience on the trip he/she is about to undertake: e.g. trip purpose, route characteristics, level of congestion encountered along the route or expected in the past days. In general, the choice of *acquiring* and *referring* to information may be expressed as a function of benefits that the user may obtain (or have obtained in the past) from the same information (e.g. travel time saving). $m_{q\varphi}$ is the qth variable characterising the source of information φ: e.g. type, accuracy. $\psi_{m\varphi}$, $\gamma_{n\varphi}$, $\delta_{q\varphi}$ are the systematic utility parameters; and $\xi_{a\varphi}$ and $\xi_{r\varphi}$ are the random residuals of the perceived utilities.

Assuming that the random residuals ($\zeta_{r\varphi}$, $\xi_{r\varphi}$, $\eta_{a\varphi}$, $\xi_{a\varphi}$) are Gumbel i.i.d. random variables, the probabilities may be formulated as the product of two MNL models, but different choice models can also be adopted.

The hierarchical structure specified may also be extended to the choice of the alternative. In this case a mathematical formulation which is internally consistent with a behavioural approach can be obtained.

A.3 Random utility models for departure time and route choice modelling

Route choice may be strictly correlated to departure time choice process, and this issue is particularly relevant in dynamic traffic assignment frameworks.

Usually elastic demand and rigid demand profiles can be distinguished.

In rigid demand profile models, it is assumed that the distribution of demand flows over departure times is known and independent from variations in travel times. It follows that path is the only choice dimension considered given a departure time.

In *elastic demand*, the flow of users following a route r connecting a generic od pair and starting at time τ can be estimated simulating in addition to path, departure time choice given the desired arrival time at destination, τ_d, or the desired departure time from the origin τ_o.

Let:

$p_{od,k}(\tau/\tau_d)$ be the choice probability of time τ and route r, given the O-D pair od and the desired arrival time τ_d;

$v_r(\tau/\tau_d)$ be the systematic utility of route r and departure time τ, given the desired arrival time τ_d;

$v_r(\tau/\tau_o)$ be the systematic utility of route r and departure time τ, given the desired arrival time τ_d;

T be the simulation interval in which the departure time choice may occur.

$\mathbf{v}_{od}(\tau/\tau_d)$ be the vector of systematic utilities relative to all the paths connecting the pair od, $r \in R$, for a given departure time τ and desired arrival time τ_d.

Choice probabilities of departure time τ and route r are usually expressed with random utility models as a function of the systematic utilities of available path-departure time alternatives:

$$p_{od,r}(\tau/\tau_d) = p_{od,r}(\mathbf{v}_{od}(\tau'/\tau_d), \forall\tau')$$

Such models are usually "single-level" random utility models with mixed continuous (departure time)/discrete (path) alternatives, or partial share models.

Usually different sequences may be adopted.

(1) $p_{od,r}(\tau/\tau_d) = p_{od}(\tau/\tau_d)\ \mathrm{p_{od}}[r/\tau,\tau_d]$

where $p_{od,r}$ (τ/τ_d) is the product of the probability to choose route r given the departure time τ, and the probability to depart at time τ.

(2) $p_{od,r}(\tau/\tau d) = pod[r]\ p_{od}(\tau/r,\tau d)$

where $p_{od,r}$ (τ/τ_d) is the product of the probability to choose route r, and the probability to depart at time τ given the route r.

With regard to the specification of departure time choice model $p_{od,r}(\tau/\tau_d)$, continuous or discrete approaches may be adopted.

In the continuous approach, a simultaneous Multinomial Logit model may be adopted:

$$p_{od,r}(\tau/\tau_d) = \exp(v_r(\tau/\tau_d))/\Sigma_{\mathrm{r'}\in\mathrm{R}}\int_0^T \exp(v_{r'}(\tau'/\tau_d))\,d\tau'$$

In the discrete approach, users do not choose among an infinite number of departure instant (τ_j), but rather among a finite number of times intervals (e.g. 5 min long). Discrete departure time models can be adopted for the continuous flows assuming that that actual departure times are uniformly distributed within the chosen interval. In this case the probability of leaving at time $\tau(j)$, within interval j, and following path r computed with a Multinomial Logit model would be:

$$p_{od,r}(\tau_j/\tau_d) = (\exp(v_r(\tau_j/\tau_d))/\Sigma_{j'}\Sigma_{r'\in R}(\exp(v_{r'}(j'/\tau_d))/\mathrm{T}$$

However, departure time and path choice probabilities can be expressed through different discrete choice models depending on the type of choice set. For instance Multinomial Logit specification as introduced before, but also through ordered choice models (Small, 1987) or hierarchical specifications able to introduce a correlation structure among adjacent departure intervals (e.g. with a Cross-Nested Logit).

Finally, it can be useful to highlight that some dynamic assignment models proposed in the literature assume deterministic utility departure time and route models. In this case, as for static systems, choice probabilities cannot be expressed in closed form as may exist several departure time/route alternatives with equal systematic disutilities.

With regard to the systematic utility functions, together with route attributes (usually Level of Service), attributes representing the penalty for arriving early or late with respect to the desired arrival time. In particular:

- the penalty, E_r, related to early arrival with respect to τ_d, departing in τ and following route r. This penalty is considered only if the early arrival is above a minimum threshold Δ_e, and it is usually expressed as function of the travel time on route r at time τ, $TT_r(\tau)$:

$$E_r(\tau, \tau_d, TT_r(\tau)) = \tau_d - \Delta_e - (\tau + TT_r(\tau)) \text{ if } \tau_d - \Delta_e - (\tau + TT_r(\tau) > 0$$

$$= 0 \text{ otherwise}$$

- The penalty, L_r, related to a delay with respect to τ_d, departing in τ and following route t. This penalty is usually considered only if the delay is above a minimum threshold Δ_l, and it is usually expressed as function of the travel time of route r at time τ

$$L_r[\tau, \tau_d, TT_r(\tau)] = \tau + TT_r(\tau) - \tau_d - \Delta_l \text{ if } \tau + TT_r(\tau) - \tau_d - \Delta_l > 0$$

$$= 0 \text{ otherwise}$$

For both penalties, deterministic or stochastic approaches may be pursued; moreover perfect rationality o bounded rationality approaches may pursued, however to address such an issue the reader ay refer to the considerations and approaches discussed in Section A.2.6.

The problem can be also expressed in terms of desired departure time τ_o from the origin. In this case the probability $p_{od,r}$ (τ/τ_o) can be expressed in a similar manner of $p_{od,r}$ (τ/τ_d), but with different specification of the systematic utility functions.

Indeed, the desired departure time from the origin τ_o is function of a scheduled delay, but in this case the scheduled delay does not depend on the route travel time $TT_r(\tau)$. Two kind of penalty may be considered:

- the penalty, E, related to early departure with respect to τ_o, departing in τ, usually considered only if the early departure is above a minimum threshold Δ_e.

$$E(\tau, \tau_o) = \tau_o - \Delta_e - \tau \text{ if } \tau_o - \Delta_e - \tau > 0$$

$$= 0 \text{ otherwise}$$

- the penalty, L, related to a delay with respect to τ_o, departing in τ, usually considered only if the delay is above a minimum threshold Δ_l:

$$L(\tau, \tau_o) = \tau - \tau_o - \Delta_l \text{ if } \tau - \tau_o - \Delta_l > 0$$

$$= 0 \text{ otherwise}$$

A.4 Fuzzy utility models for modelling traveller's choice

Fuzzy utility models share the same hypothesis of Random Utility models except for the assumption on uncertainty paradigm. Instead of the probability theory, perceived utility as a random variable, the uncertainty is modelled within the possibility theory, perceived utility as a fuzzy number (see mathematical notes 2 for further details).

Possibility theory allows taking into account two types of uncertainty: *dispersion* (due to contrasting alternatives) and *imprecision of information* (or data) available to the users (or to the modeller).

Most existing approaches to simulate uncertainty in choice behaviour through fuzzy numbers are based on ranking indices: given a set of fuzzy numbers, a crisp number (ranking index) is associated to each of them, so that the fuzzy numbers can linearly be ordered. The most effective indices seem those proposed by Dubois and Prade (1983). Once the value of the ranking index has been computed for the fuzzy number describing the perceived utility of each alternative, an estimate of the choice share for each alternative is obtained from the values of the ranking index. Henn (2002) presents a review on this approach (see also Henn, 2000; Binetti and De Mitri, 2002). Approaches based on ranking indices may lead to some counter intuitive results, as argued also by Henn (2002).

To overcome these limitations, a different approach, coherent with the Utility Theory may be pursued assuming that the perceived utility is modelled through a fuzzy number, leading to the fuzzy utility theory and to fuzzy utility models (FUM). Fuzzy utility models may be specified assuming that.

a. Each user (of a class of homogenous users) of a transportation system.
- **a1** considers a set of alternatives,
- **a2** gives each alternative a value of perceived utility, and
- **a3** chooses the alternative with the maximum value of perceived utility (homo-economicus assumption);

b. The perceived utility of a mode is modelled through a fuzzy variable due to several sources of the above-mentioned uncertainties. Let.

R be the choice set, containing all the available alternatives, it is assumed non-empty and finite, with $m = |R|$;
U_r be the perceived utility associated to alternative $r \in R = \{1, \ldots, m\}$, considered a fuzzy number with fuzzy distribution function (fdf) $\mu_{Ur}(U_r)$;
v_r is the core value of the fuzzy distribution function;
$q_r \geq 0$ be the choice possibility associated to alternative r, with $\max_r q_r = 1$;
$z_r \geq 0$ be an estimate of the choice fraction associated to alternative r, with $\sum_r z_r = 1$, hence, they be considered as a probability distribution.

FUM can be formulated by defining the maximum perceived utility fuzzy distribution (defined from the perceived utility fuzzy distribution of each alternative as for RUT) and defining the possibility that the perceived utility of an alternative be equal to the maximum value. Then, choice fractions are deducted from choice possibilities, which model both non-specificity and discord, through a transformation which preserves the total amount of uncertainty considering choice fractions as a probability distribution.

FUMs allow estimating choice fractions as a function of the core values of the alternatives perceived utility ($\mathbf{v}$) and of vector of parameters $\boldsymbol{\theta}$.

$\mathbf{z} = z(\mathbf{v}; \boldsymbol{\theta})$

The specification of a FUM requires to address 3 main issues:

(i) the definition of the choice set,
(ii) the specification of systematic utility function, and
(iii) the identification of the most effective choice function (fuzzy distribution function).

A.4.1 Choice set definition

As regards the definition of the choice set, the same considerations introduced in Section A.1.1 hold for a FUM, thus no details are reported. In particular, it could be useful to point out how the perception/availability approach is consistent with the overall fuzzy utility theory.

A.4.2 Specification of the systematic utility

As for RUM's, the application of a FUM requires that the core value, v_r, of each choice alternative r be specified as a function, $v_r = \mathrm{v}(\mathbf{x}_\mathrm{r}; \boldsymbol{\psi})$, of attributes, $\mathbf{x}_\mathrm{r}$, measured or assumed for a design scenario. A linear expression is commonly used: $v_r = \sum_\mathrm{j} \psi_\mathrm{j}\, \mathrm{x}_\mathrm{rj}$.

With regard to the attributes to use, the same consideration introduced in Section A.1.2 hold.

A.4.3 Distribution of perceived utility and choice functions

As for random utility models, FUM can be formulated by defining the maximum perceived utility fuzzy distribution and defining the possibility that the perceived utility of an alternative be equal to the maximum value.

Choice possibilities, q_r, can be obtained from the perceived utility fdf, $\mu_{\mathrm{Ur}}(\mathrm{u_r})$, of each alternative, r, through a two-step procedure.

The (possibilistic) maximum perceived utility, U_{MAX}, can be defined by repeatedly applying the fuzzy MAX operator (it can equivalently be defined by the general expression of the maximum of several fuzzy numbers):

$$U_{MAX} = U_1$$

$$U_{MAX} = \mathrm{MAX}\ (U_{MAX}, U_r) \forall r \geq 2$$

The possibility that any alternatives is chosen, say it is a maximum perceived utility alternative, is given by:

$$q_r = \mathrm{Pos}(U_r = U_{MAX}) = \max_\mathrm{x} \min\{\mu_{Ur}(\mathrm{x}), \mu_{UMAX}(\mathrm{x})\} \in [0, 1] \forall r \geq 1$$

It should be noted that at least one alternative will have a choice possibility equal to one, say $\max_\mathrm{r} q_r = 1$, thus a n-valued discrete possibility distribution is obtained.

Choice fractions, s_r, can be obtained from choice possibilities (q_r) by applying a general possibility-probability (P-P) transformation, $\mathbf{p} = \mathbf{p}(\mathbf{q})$, between a possibility distribution, **q**, and a probability one, **p**.

Such a transformation is called *consistent* if the corresponding components of the two distributions are ordered the same way, $p_r \geq p_{r+1} \Leftrightarrow q_r \geq q_{r+1}$. Generally, infinite many consistent transformations may be devised, such as the power transformation which assures existence and uniqueness of transformation (in both directions). The following equation, which assures both normalisation constraints, is usually adopted:

$$z_r = q_r^{1/\alpha} / \sum_{h \in R} q_h^{1/\alpha} \forall r \geq 1$$

where $0 < \alpha < 1$ assures a consistent transformation.

For example, for two alternatives, probability and possibility definitions yield: p_1, $p_2 \in [0,1]$, $p_1 + p_2 = 1$ or $p_1 = 1 - p_2$, and q_1, $q2 \in [0,1]$, $\max\{q_1, q_2\} = 1$. Due to normalisation, a consistent P-P transformation is expressed by any relation $p_2 = p_2(q_2)$ for which both $p_2 \leq 0.5$(since $p_2 \leq p_1 \Leftrightarrow q_2 \leq q_1 = 1$)and $p_2 \leq q_2$ (besides $p_1 \leq q_1 = 1$) hold.

The successive application of the above presented equations lead to a relation between the choice share vector (**z**), the central value (core) vector **v** and the vector of parameter **θ**.

$$\mathbf{s} = \boldsymbol{s}(\mathbf{v}; \boldsymbol{\theta})$$

The explicit expression of this relation depends on the assumptions adopted about the perceived utility fuzzy distribution (fdf) as well as the values of its parameters (such as values of the variability index).

Also for FUMs, different formulations may been derived and it is possible to introduce a classification based on the possible assumptions that can be made on the fdf of the perceived utilities.

Usually a fdf can be described by a specific shape (e.g. triangular), by a core value of v_r, by width and/or shape parameters. In general, perceived utilities can be identically distributed among alternatives or not identically distributed among alternatives. In this case, differently from RUMs, two families of models may be obtained by the combination of the above hypotheses: homoscedastic or heteroscedastic models.

To this aim the following fuzzy distribution functions have been adopted in practical application:

- triangular symmetric–generic: characterised by the core value, and by equal width for the right and the left sides;
- triangular symmetric–specific: characterised by the core value, and by the same widths for the right and the left sides;
- triangular asymmetric–generic: characterised by the core value, and by different widths for the right and the left sides.

The last allowing to take into account heteroscedasticity between alternative perceived utilities.

The parameters, $\boldsymbol{\psi}$, of the core value function, as well as other parameters for the fdf (such as the ratio between left and right widths, w_L/w_R, for asymmetric triangular fuzzy distribution function), are calibrated trying to reproduce a sample of observations about user choices (disaggregate) and/or demand flows (aggregate).

Any distance function between observed and modelled values, such as the most widely used Euclidean one, can express how well the observations are reproduced by a set of parameters; the minimum of the distance function leads to the estimates of parameters. (this approach can also be adopted for RUM's, which are usually calibrated through maximisation of log-Likelihood.)

It should be noted that if all the choice possibilities are independent of any linear transformation of the core value utility, a scale parameter cannot be identified.

As it occurs for models derived from random utility theory, if all the parameters of the descriptive functions, but the position index, do not depend on the position index itself, the choice possibilities are independent of any (crisp) linear transformation of perceived utility, say $U_r' = a + b \cdot U_r$, in other word any change of utility origin or scale does not affect choice shares, as usually assumed in economic-based choice behaviour theories.

In conclusion, it should be highlighted that FUM have been implemented to mode choice contexts, but several issues are worth of further research work before a fully consistent fuzzy utility theory could be considered acquired, namely:

- models resulting from different fuzzy distributions of perceived utility, such as the beta fdf, which also includes a shape parameter (besides the core, the left and the right widths), but requires a numerical computation of U_{MAX};
- fuzzy vector distributions to model similarity between perceived utility of pairs of alternatives,
- analysis of mathematical features, such as continuity and monotonicity, of choice fractions from FUM against core values;
- calibration method;
- algorithms for application.

Finally, specific solutions should be formulated for modelling the correlation across perceived utilities (usually related to alternatives that are similar and/or that share same characteristics) and/or the correlation over time of user choices.

A.5 Neural network for modelling traveller's choice

A different approach to travel demand and to travel choice analysis based on artificial neural network models has been proposed by several researchers (a review in Cantarella and de Luca, 2005a; Karlaftis and Vlahogianni, 2011) and has showed excellent capability to simulate choice behaviour.

Artificial neural networks (ANNs) were first developed as a simplified model of the neural tissue which made up animal neural systems. According to this biological metaphor an ANN may be considered a modelling function made up of sub-modelling functions, namely the called neurons, and in this sense it may be considered a parallel distributed processing (PDP) model. As such they are capable to generalise a phenomenon starting from experimental input–output pairs of data (supervised learning).

Formally, a parallel distributed processing (PDP) model is a function made up of several sub-functions or basic processing units (PUs) working as follows. At each *PU*, *k*, first a linear transformation of the upstream inputs is performed: namely each received input, x_j, is weighted, $w_{jk} \cdot x_j$, all weighted inputs are summed up, $\sum_{j \text{ UP } k} w_{jk} \cdot x_j$ and added to a constant (usually called bias), b_k, leading to $z_k = \sum_j w_{jk}\, x_j + b_k$. Then, an activation function is valued, $y_k = \varphi(z_k)$. Finally, the output value y_k is forwarded to downstream *PUs*, (Fig. A.2).

Output processing units supply their outputs to the model-user, whilst *input processing units* just receive their inputs from the model-user during the model application and forward them to the downstream *PUs* without any transformation (so far, they actually do not perform any process). All the other *PUs* are called *hidden processing units*, since the model-user does not see them (of course the model-builder is well aware of them). The layouts can be described by a digraph, with a node for each *PU*, and an arc for each pair of connected *PUs*.

A multilayered feedforward network (MLFFN) is obtained when the layout can be described by a multi-level graph, where all *PUs* are grouped into layers, and the layers are sequentially ordered from the input one to the output one so that each *PU* is only connected with all those in the upstream layer (if any) and in the downstream layer (if any) but not with those in the same layer, Fig. A.3. This architecture is the most used ANN models for classification and non-linear regression analysis as well as choice analysis.

According to these assumptions, a MLFFN should be considered a black box model of the relationship $\boldsymbol{y} = \psi(\boldsymbol{X}; \boldsymbol{w}, \boldsymbol{b})$, between attributes $\boldsymbol{X}$ and users' choices $\boldsymbol{y}$, and not an explicit model of the cognitive processes underlying user choices.

Traditional MLFFN does not allow any interpretation, but a possible solution may be achieved by specifying a utility based lay-out (Cantarella and de Luca, 2005b).

Indeed, the proposed lay-out is made-up by two intermediate layers besides the input and output ones (Fig. A.4), expressing the combination of two functions,

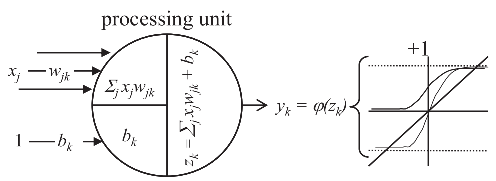


FIG. A.2

A processing unit (PU) and its activation function.

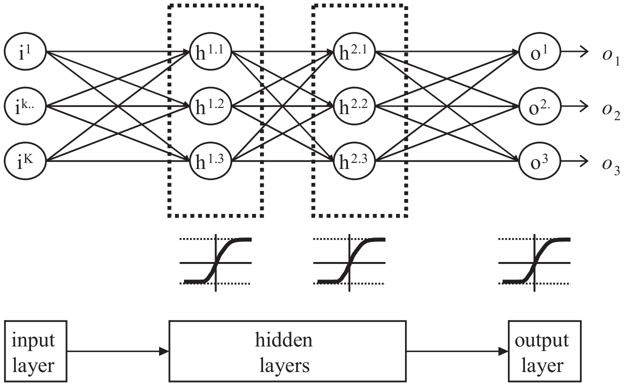


FIG. A.3

A multilayered feedforward network ($K \times 3 \times 3 \times 3$).

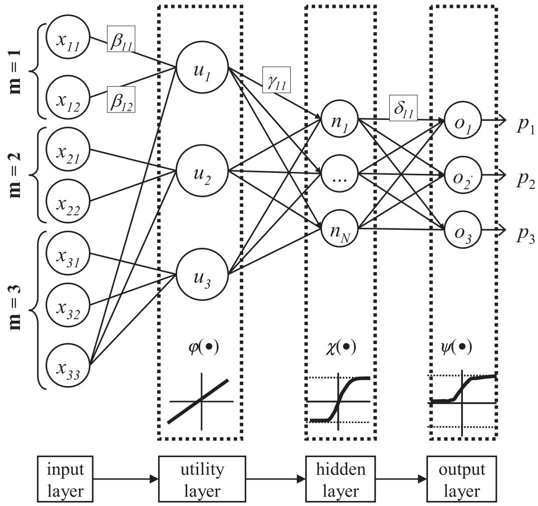


FIG. A.4

Utility-based MLFFN architecture.

according to the utility-based approach introduced above, with the combination of two functions:

$\boldsymbol{v} = v(\boldsymbol{x})$ utility function: from layer 0 to layer 1.
$\boldsymbol{p} = p(\boldsymbol{v})$ choice function: from layer 1 to 3 through 2.

[0] *Input layer*. This layer contains one *PU mj* for each attribute *j* (for instance: travel time, monetary cost, …) and each alternative *m*, it simply forwards the input value x_{mj} (from the data-set given by the model user) to downstream *PUs*, without processing it.

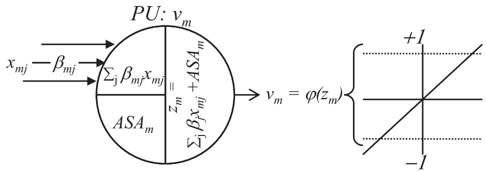


FIG. A.5

PU of the utility layer.

[1] *Utility layer*. This layer contains one *PU m* for each mode *m*, which only receives input values from the upstream input *PUs mj* corresponding to the same alternative *m* (Fig. A.5). So far the input and utility layers are not fully connected.

Assuming the identity function as activation function, $\varphi(\bullet)$, the output, v_m, is given by a linear combination of attributes:

$$v_m = \sum_j \beta_{mj} x_{mj} + ASA_m$$

where β_{mj} is the weight associated to the connection between input PU *mj* (attribute *j* for mode *m*) and utility PU *m* and ASA_m is the bias (constant) associated to utility PU *m* (named after Alternative Specific Attribute from econometrics).

The provided output v_m is formally analogous to commonly adopted utility function, quoted in the previous sub-section. Once parameters such as weights and biases have been calibrated they may be given an interpretation.

[2] *Hidden layer*. The number of *PUs* in this layer is defined during the model-building stage. Each *PU n* in this layer is connected to (receives an input from) each *PU m* in the utility layer (Fig. A.6). Assuming an activation function $\chi(\bullet)$, common to all *PUs* in this layer, the output, y_n, is given by:

$$y_n = \chi\left(\sum_m \gamma_{mn} v_m + c_n\right)$$

where γ_{mn} is the weight associated to the connection between utility *PU m* (for mode *m*) and hidden *PU n* and c_n is the bias (constant) associated to hidden *PU n*.

It should be noted that the calibrated weights and biases in this layer should hardly be given a clear interpretation.

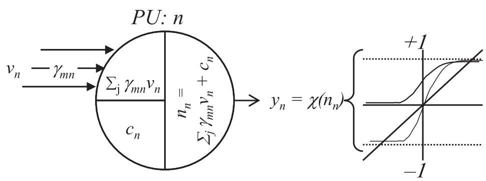


FIG. A.6

PU of the hidden layer.

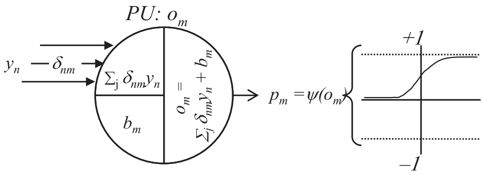


FIG. A.7

PU of the output layer.

[3] *Output layer*. This layer contains one *PU m* for each alternative *m*, which is connected to (receives an input from) each *PU n* in the hidden layer. Assuming an activation function $\psi(\bullet)$ with values in the range [0,1] and common to all *PUs* in this layer, the output, p_m, is given by (Fig. A.7):

$$p_m = \psi_m\left(\sum\nolimits_n \delta_{nm}\, y_n + b_m\right)$$

where δ_{nm} is the weight associated to the connection between hidden PU *n* and output *PU m* (for mode *m*) and b_m is the bias (constant) associated to hidden *PU n*.

According to the proposed lay-out allows both the utility function and the choice function are explicitly and separately specified, as in random utility models (RUMs) within an econometric framework.

The specification of a ANN for discrete choice analysis requires to address 3 main issues:

- **(i)** the definition of the choice set,
- **(ii)** the specification of systematic utility function,
- **(iii)** the identification of the most effective choice function (fuzzy distribution function).

A.5.1 Choice set definition

As regards the definition of the choice set, the same considerations introduced in Section A.1.1 hold for a the ANN, thus no details are reported. However, it should be noted that availability may be considered as a PU unit and availability may be implicitly calibrated into the overall calibration procedure of the ANN model.

A.5.2 Specification and calibration of a neural network model

Given the choice set, *R*, in order to define a MLFFN model for user choice analysis, one input processing unit is introduced for each attribute of each alternative, and one output processing unit, giving a value in the range [0,1], is introduced for each alternative. In order to compensate for numerical errors, a normalisation stage is usually carried out to ensure that the sum of output values (non-negative in any case) over all the alternatives is equal to one. The input and output layers are connected by one or more hidden layers.

The specification of a MLFFN model requires to address the following main operational issues:

(1) input attributes definition,
(2) selection of MLFFN architecture: general or utility-based,

Besides assigning the meaning of input and output *PUs*, which depends on the particular problem at hand, three items should be defined for a complete specification of a MLFFN:

(a) number of layers, including the input and output ones, as well as some hidden layers, conventionally numbered from 0 (the input layer) to N (the output layer);
(b) number of *PUs* for each layer k, say n_k;
(c) activation function, usually common to all *PUs* in the same layer k, say $\varphi_k(\bullet)$, clearly no activation function is associated to *PUs* in the input layer; commonly adopted activation functions are:

- linear, say $\varphi(z_k) = \alpha + \beta\ z_k$,
- logistic, say $\varphi(z_k) = 1/(1 + e^{-\beta\ zk})$
- hyperbolic tangent, say $\varphi(z_k) = (e^{\beta\ zk} - e^{-\beta\ zk})/(e^{\beta\ zk} + e^{-\beta\ zk})$

Usually, the parameters of the activation functions, say α and β in the examples above, are chosen by the model-builder. The MLFFN parameters to be calibrated include one weight for each connection and one bias for each *PU* except the input ones, hence their number is $\sum_{k=1,N} (n_{k-1} \times n_k) + \sum_{k=1,N} n_k$. Due to the high number of parameters the possibility of over-fitting should be carefully checked.

Once specified the ANN architecture, the calibration can be schematised in the following main operational issues:

(1) calibration and validation data-sets definition,
(2) error function to minimise,
(3) parameters (weight and biases) initialisation technique,
(4) number of epochs and of starting conditions (initialisations),
(5) selection of MLFFN architecture,
(6) computation of the parameters for the selected MLFFN architecture.

First of all, the available sample of observations (data-set) has to be split into the calibration data-set, and the hold-out or validation data-set, which is set aside to carry out the validation stage, some items should be defined to carry out a calibration. This stage is very important, since the calibration of a ANN model simultaneously carry out the validation stage to avoid over-training and over-fitting phenomena.

Secondly, the error function (calibration function) to minimise should be defined. The calibration function is the mean sum of the squares of differences (MSE) between the observed and predicted outputs; a (local) minimum can be found through the back-propagation algorithm, a specific implementation of the gradient algorithm that duly exploits the structure of the MLFFN by backwards updating weights and biases from the output layer to the input layer (different algorithm may be also used).

As many non-linear optimisation problems, MLFFN calibration has to deal with many local minima due to several reasons, the obtained solution can be greatly affected by the starting values of weights and biases, thus requires different initialisations of parameters. Since the convergence is not guaranteed when all the weights are set to the same value, an appropriate weights initialisation can improve both rate of convergence and precision. Two techniques may be used: random or Nguyen–Widrow initialisation. The latter distributes the initial weights and bias values for a layer so that the active region of the layer's processing units is distributed roughly evenly over the input space.

Once defined the desired input attributes, then the maximum number of epochs must be defined, that is the number of copies of the calibration data-set which are actually used for calibration. An excessively high number of epochs may lead to over-training, that is, a very close reproduction of calibration data-set with poor generalisation of validation data-set. Over-training can be prevented by heuristic techniques or by the Early Stopping criterion that stops the calibration when the error, computed on a limited data set, increases. Then, the number of repetitions of the back-propagation algorithm with different starting conditions should also be defined. Clearly the higher the number of repetitions the more likely a good approximation of the global minimum is obtained. This approach leads to several values for calibration parameters, the values corresponding to the best value of the calibration function are usually adopted during the applications.

Generally, it can be useful to adopt techniques to deal with over-fitting (such as the well-known heuristic criteria, weight decay technique or pruning techniques), that is, the MLFFN is able to reproduce well the calibration sample but unable to generalise new samples. This depends on the whole number of parameters (depending on the numbers of PUs and hidden layers) which should be much less than the number of observations in the calibration data-set. Over-fitting may also be affected by factors causing over-training.

So far different architectures should be calibrated with different number of epochs (usually from 50 to 5000) and different number of starting conditions (e.g. from 10 to 200). It should be noted that each starting condition leads to a different set of parameters, still the values of the calibration function (MSE) can be quite similar, in other words different sets of parameters may show similar reproduction capability.

The MLFFN architecture is selected among all instances obtained by varying the number of hidden layers (e.g. 1, 2, 3), the number of processing units per layer (e.g. from 10 to 50), and the activation function per processing unit (e.g. linear, hyperbolic tangent and sigmoid functions, usually common to all processing units in the same layer). It should be recalled that to each instance is associated a different set of parameters for each starting condition.

Generally, an architecture should be preferred when it shows:

(a) Good reproducibility, as measured by the error between observations and simulated values for the calibration data-set,
(b) Good generalisation, as measured by the error between observations and simulated values for the validation data-set (using parameters obtained from calibration).
(c) Low dependency on starting conditions, as measured by the dispersion of error between observations and simulated values for the calibration data-set over the starting conditions. Several indices may be adopted to describe the above criteria, for instance: mean square error between the user observed mode choice fractions and the simulated ones (over the calibration or the validation data-sets), measures of reproducibility, measures of generalisation, etc.

Usually, a multi-criteria technique can be applied based on the three criteria above introduced. First, non-dominated (Pareto-optimal) architectures are devised, then criteria are normalised over the range [0,1], finally an architecture is selected through the least-distance-from-origin criterion. More sophisticated selection criteria could be implemented.

Once a MLFFN architecture has been selected during the previous step, for each starting condition a different set of parameters is obtained by applying the calibration algorithm, as already pointed out. A set of parameters must be selected to apply the calibrated and validated model to any scenario. The most adopted approach is also the simplest one: just select the set of parameters with the best value of error function. Other approaches may consist in selecting a combination of sets of the obtained *parameters*.

A.5.3 Validation

Due to the high number of parameters of an MLFFN as well as its flexible structure, the validation stage plays a very relevant role in checking whether generalisation vs. reproduction of observations is obtained. To this end, an in-depth validation analysis, as previously introduced, should be carried out. An effective approach is the bootstrap technique that re-samples data from the original calibration sample many times in order to generate an empirical distribution of the calibration error. However, the validation of an ANN model outputs can be carried out through some of the indicators proposed in section 2.4.

A.6 Summary

A.6.1 Major findings

This chapter presented a general overview of the possible approaches for modelling disaggregate travel behaviour with special attention to route choice modelling. In particular, three theoretical paradigms are discussed: the consolidated Random

Utility behavioural paradigm, the Fuzzy Utility behavioural paradigm and the not-behavioural paradigm based on artificial neural network models.

First, random utility theory (RUT) is introduced and all the issues regarding the specification, calibration and validation of a random utility model (RUM) are addressed. Two relevant choice dimensions are discussed in details: route and departure time choice. Route choice process are analysed by distinguishing the different types of choice that a user may be called to take; indeed, the following issues are addressed: pre-trip choices, en-route choices, reaction to information, switching behaviour. For each of the above-mentioned choice problems, the most adopted RUMs are formalised and discussed.

Then, fuzzy utility theory, though not yet widely applied in practice, represents an alternative paradigm that explicitly allows to interpret and quantify a different type of uncertainty. Fuzzy utility models represent a potential alternative to RUMs for simulating choice behaviour, but also a reference approach to understand if and how a different type of uncertainly may lead to most performing models and/or to different interpretation of the choice process determinants.

ANN models are a powerful tool to predict users' behaviour and represent a benchmark modelling solution for the above-mentioned utilitarian paradigms. Their use may be rather effective in those choice contexts in which choice behaviour may rapidly change and the choice model parameter's may require to be continuously, or frequently, updated. Finally, a general protocol for validating any choice model is presented.

A.6.2 Further readings

The chapter does not aim to cover all the possible paradigms and all the possible route choice models, but those paradigms that can be easily embedded into the mathematical framework proposed in the book. Indeed, the literature of the past decade is rich of several and different interpretative and theoretical paradigms, also fruitfully applied to disaggregate (discrete) choice modelling.

For a general and comprehensive framework on choice modelling and on all the related issues, the reader may refer to the contributions by Luce (1959), Domencich and McFadden (1975), Daganzo (1979), Ben-Akiva and Lerman (1985), Klir and Wierman (1999), Hensher and Button (2000), Louviere et al. (2000), McFadden (2001), Washington et al. (2003), Train (2009), Cascetta (2009), and Rasouli and Timmermans (2015).

With regard to route choice modelling, different models, not introduced in this chapther, have been proposed within the Random Utility paradigms and can be worth of interest: the paired combinatorial logit model (Chu, 1989), the link-Nested Logit (Vovsha and Bekhor, 1998), the path-size Logit (Ben-Akiva and Bierlaire, 1999), the implicit availability perception Logit (Cascetta and Papola, 2001; Cascetta et al., 2002),the quantum utility models (Vitetta, 2016), the CoRum (Papola, 2016). However an interesting state of art can be found in Prashker and Bekhor (2004).

Recently different theoretical paradigms have been investigated and applied to discrete choice issues: the Prospect theory (Katsikopoulos et al., 2000;

Avineri and Prashker, 2004; Gao et al., 2010; de Luca and Di Pace, 2015); Elimination by Aspects theory (Tversky, 1972a,b; Batley and Daly, 2003; the Random Regret minimisation framework (Chorus et al., 2008; Prato, 2013; Mai et al., 2017).

A.6.3 Remarks (G. E. Cantarella)

Some choice modelling approaches assume that all sources of uncertainty are negligible, thus all users travelling between o–p pair i follow maximum utility routes, and do not use at all any of the other routes. In this case, a route r can be used only if, from among the set of alternative routes, its utility v_r is the max. This approach does not provide a unique route choice probability vector, except when there is a unique maximum utility route.

This user choice behaviour modelling approach (cfr Wardrop, 1952) may be obtained from Deterministic Utility Theory, or as a limit of random/fuzzy utility theory when dispersion goes to zero, as well as from the expected utility theory, where perceived utility is modelled by a random variables but users are assumed choosing according the expected values of the perceived utility. Although these choice models are less realistic than probabilistic models, for computational reasons they are often applied to very large networks with implicit route enumeration.

In this case, as noted above, the route choice function $\mathbf{p} = \mathbf{p}(\mathbf{v}; \boldsymbol{\theta})$ is actually a multi-valued function (also called a one-to-many or a point-to-set function or a map), since it may well be the case that the systematic utility values of two or more routes are equal to the maximum. A different approach is often followed to avoid this kind of functions. Let.

$\mathbf{p_D}(\mathbf{v}) \geq \mathbf{0}$ be any of the route deterministic choice proportion vectors corresponding to systematic utility vector $\mathbf{v}$ with $\mathbf{1}^{\mathrm{T}}\ \mathbf{p_D}(\mathbf{v}) = 1$; for any route r

$$\text{if } v_r < v_{max} \text{ then } p_r = 0 \Leftrightarrow \text{if } p_r > 0 \text{ then } v_r = v_{max}$$

[The case $p_r = 0$ with $v_r = v_{max}$ it is not ruled out by this condition.] From the above condition, for any systematic utility vector $\mathbf{v}$, $\mathbf{p_D} = \mathbf{p_D}(\mathbf{v})$ is equivalent to the following condition (see Appendix A for more details):

$$\mathbf{v}^{\mathrm{T}} \cdot (\mathbf{p_D} - \mathbf{q}) \geq 0 \forall \mathbf{q} \geq \mathbf{0} \text{ with } \mathbf{1}^{\mathrm{T}}\ \mathbf{q} = 1$$

A.7 Mathematical notes (*G. E. Cantarella*)

A.7.1 Mathematical properties of random or deterministic utility models

Several useful mathematical properties of random or deterministic utility models can be proved as shown below (reported proofs are adapted from Cantarella, 1997, see also Daganzo, 1979). The expected maximum perceived utility (EMPU) variable, s, may associated to any random utility models, it is defined as the expected value of perceived utility over the alternatives available in the choice set:

$$s = s(\mathbf{U}) = E[\max_r(\mathbf{U})] = E[\max(\mathbf{v}+\boldsymbol{\xi})] = \int_{\dots}\int_{\dots}\int \max(\mathbf{v}+\boldsymbol{\xi})\, f(\boldsymbol{\xi})\, d\boldsymbol{\xi}$$

The EMPU is a function of the systematic utilities of all the alternatives, vector $\mathbf{v}$, and that it depends on the joint probability density function of the random residuals, $f(\boldsymbol{\xi})$, as well as on the composition of the choice set R.

The EMPU is always greater than or equal to the maximum systematic utility:

$$s(\mathbf{v}) \geq \max(\mathbf{v})$$

Indeed, by definition,

$$s(\mathbf{v}) = \int_{\varepsilon 1 \dots}\int_{\varepsilon m} \max(\mathbf{v}+\boldsymbol{\xi})\, f(\boldsymbol{\xi})\, d\boldsymbol{\xi}$$

and since $f(\boldsymbol{\xi}) \geq 0$ and $\max(\mathbf{v} + \boldsymbol{\varepsilon}) \geq V_r + \varepsilon_r \ \forall r \in R$, it follows that:

$$s(\mathbf{v}) = \int_{\varepsilon 1 \dots}\int_{\varepsilon m} \max(\mathbf{v}+\boldsymbol{\xi})\, f(\boldsymbol{\xi})\, d\boldsymbol{\xi}$$

$$\geq \int_{\varepsilon 1 \dots}\int_{\varepsilon m} v_r\, f(\boldsymbol{\xi})\, d\boldsymbol{\xi} + \int_{\varepsilon 1 \dots}\int_{\varepsilon m} \xi_r\, f(\boldsymbol{\xi})\, d\boldsymbol{\xi}$$

$$= v_r \int_{\varepsilon 1 \dots}\int_{\varepsilon m} f(\boldsymbol{\xi})\, d\boldsymbol{\xi} + \int_{\varepsilon 1 \dots}\int_{\varepsilon m} \xi_r\, f(\boldsymbol{\xi})\, d\boldsymbol{\xi}$$

$$= v_r + E[\xi_r] = v_r \forall h \in R$$

Therefore $s(\mathbf{v})$ is greater than or equal to the largest systematic utility, $s(\mathbf{v}) \geq V_h \ \forall h \in R$.

In addition, the mean systematic utility, calculated by weighing the systematic utility of each alternative r by its respective choice probability $p_r(\mathbf{v})$, is less than or equal to the EMPU variable. From the previous expression, it follows that:

$$p(\mathbf{v})^T \mathbf{v} = \Sigma_r\, p_r(\mathbf{v})\, v_r \leq \Sigma_r\, p_r(\mathbf{v}) \max(v) = \max(\mathbf{v}) \leq s(\mathbf{v})$$

Writing the EMPU variable for a MNL model with constant parameter θ, other properties can be derived. Indeed, since the EMPU variable for MNL can be written in the following closed-form formulation (also known as logsum variable)

$$s(\mathbf{v}) = \theta\ \ln_{\Sigma r} \exp(v_r/\theta)$$

It is important to note that the EMPU increases if the systematic utility of one or more alternatives increases since the functions $ln(\cdot)$ and $exp(\cdot)$ are both monotonic increasing. Furthermore, because of the non-negativity of the exponential function, the EMPU increases with the number of available alternatives. In fact, the addition of a new alternative to the choice set results in an increase in the EMPU even if the new alternative has a systematic utility less than that of the alternatives already available. This is because of the randomness of perceived utilities: there is a positive probability that the new alternative will be perceived as having a utility greater than that of any other alternative.

These properties of EMPU, directly derived here for the MNL model, also apply to the larger class of *invariant random utility models*. Recall that, for these models, the density function of the random residuals does not depend on $\boldsymbol{v}$.

Moreover, if the joint density function of the random residuals $f(\boldsymbol{\xi})$ is continuous with continuous first derivatives, the choice probabilities $p(\mathbf{v})$ and the EMPU $s(\mathbf{v})$ are also continuous functions of $\mathbf{v}$ with continuous first derivatives. All random utility models described in the previous sections satisfy these continuity requirements. Under these assumptions, invariant random utility models share a number of general mathematical properties that are connected with the Expected Maximum Perceived Utility. In particular, the following properties are worth of interest for the aims of the book.

(1) The *partial derivative* of the EMPU with respect to the systematic utility v_r is equal to the choice probability of alternative r:

$$\partial s(\mathbf{v})/\partial v_r = p_r(\mathbf{v})$$

The gradient of the EMPU is thus equal to the vector of choice probabilities:

$$\nabla s(\mathbf{v}) = \mathbf{p}(\mathbf{v})$$

and its Hessian is equal to the Jacobian of choice probabilities:

$$\mathbf{Hess}[s(\mathbf{v})] = \mathbf{Jac}[p(\mathbf{v})]$$

Indeed, for a continuous function with continuous first derivatives, the integration and differentiation operators can be exchanged:

$$\partial s(\mathbf{v})/\partial v_r = \partial \left(\int_{\varepsilon 1} \ldots \int_{\varepsilon m} \max(\mathbf{v}+\boldsymbol{\xi})\, f(\boldsymbol{\xi}) d\boldsymbol{\xi} \right) / \partial v_r = \int_{\varepsilon 1} \ldots \int \varepsilon m\, (\partial \max(\mathbf{v}+\boldsymbol{\xi})/\partial v_r)\, f(\boldsymbol{\xi}) d\boldsymbol{\xi})$$

Since $\partial \max(\mathbf{v}+\boldsymbol{\xi})/\partial v_r = 1$ for r such that $(v_r+\xi_r) = \max(\mathbf{v}+\boldsymbol{\xi})$, $= 0$ otherwise, the integral is equal to the probability that the perceived utility of alternative r, $v_r+\xi_k$, is the largest among all the m alternatives available. This result can be checked immediately for the Multinomial Logit model, for which the EMPU can be differentiated analytically.

Furthermore, since the choice probability p_r is always greater than or equal to zero, it can be easily demonstrated that the derivative of the EMPU with respect to the systematic utility is always non-negative: the EMPU increases (or does not decrease) as the systematic utility of each alternative increases and, by extension, as the number of available alternatives increases.

(2) The EMPU function is convex with respect to $\mathbf{v}$, the vector of systematic utilities.

Indeed, for each $\boldsymbol{\xi}$, $f(\boldsymbol{\xi}) \geq 0$ and $\max(\mathbf{v}+\boldsymbol{\xi})$ is a convex function of $\mathbf{v}$; it follows that the Expected Maximum Perceived Utility function $s(\mathbf{v})$ is a linear combination with non-negative coefficients of convex functions, and therefore is convex too. By virtue of this property, the EMPU function has a Hessian matrix, $\mathbf{Hess}(s(\mathbf{v}))$, which is (symmetric and) positive semidefinite. Consequently, the Jacobian of choice probabilities, $\mathbf{Jac}(\mathbf{p}(\mathbf{v}))$, is (symmetric and) positive semidefinite.

(3) If the EMPU function is continuous and differentiable then:

$$s(\mathbf{v}') \geq s(\mathbf{v}'') + p(\mathbf{v}'')^T (\mathbf{v}' - \mathbf{v}'') \forall \mathbf{v}', \mathbf{v}'' \tag{a}$$

and the choice probabilities are monotonic increasing functions of the systematic utilities.

$$(p(\mathbf{v}') - p(\mathbf{v}''))^T (\mathbf{v}' - \mathbf{v}'') \geq 0 \forall \mathbf{v}', \mathbf{v}'' \tag{b}$$

Indeed, because the EMPU function is convex and differentiable, it follows that:

$$s(\mathbf{v}') \geq s(\mathbf{v}'') + \nabla s(\mathbf{v}'')^T (\mathbf{v}' - \mathbf{v}'') \forall \mathbf{v}', \mathbf{v}''$$

and its gradient must be an increasing monotonic function:

$$(\nabla s(\mathbf{v}') - \nabla s(\mathbf{v}''))^T (\mathbf{v}' - \mathbf{v}'') \geq 0 \forall \mathbf{v}', \mathbf{v}''$$

The two preceding expressions can be also formulated in terms of the vector of choice probabilities:

$$s(\mathbf{v}') - s(\mathbf{v}'') \geq p(\mathbf{v}'')^T (\mathbf{v}' - \mathbf{v}'') \forall \mathbf{v}', \mathbf{v}''$$

$$s(\mathbf{v}'') - s(\mathbf{v}') \geq p(\mathbf{v}')^T (\mathbf{v}'' - \mathbf{v}'') \forall \mathbf{v}', \mathbf{v}''$$

Summing the last two inequalities yields:

$$0 \geq p(\mathbf{v}'')^T (\mathbf{v}' - \mathbf{v}'') + p(\mathbf{v}')^T (\mathbf{v}'' - \mathbf{v}'') \forall \mathbf{v}', \mathbf{v}''$$

from which Eq. (a) is easily obtained.

Eq. (b) can be expressed for a single alternative, assuming that the systematic utilities of all other choice alternatives are constant:

$$p_r(v_r') \geq p_r(v_r'') \text{ if } v_r' \geq v_r''$$

In other words, the choice probability of a generic alternative does not decrease as its systematic utility increases, if all the other systematic utilities remain unchanged. Using an analogous argument it can be demonstrated that, as v_r tends to minus infinity, the choice probability of alternative r tends to zero.

The *deterministic choice model* satisfies condition invariant models. If there are two or more alternatives with (equal) maximum systematic utility, there are infinitely many choice probability vectors satisfying the above conditions. In this case, the relation $\mathbf{p}(\mathbf{v})$ is not a function, but a one-to-many map. Let $\mathbf{p_D}(\mathbf{v})$ be one of the possible choice probability vectors corresponding to vector $\mathbf{v}$ through the deterministic choice map.

The following necessary and sufficient condition guarantees that a probability vector $\mathbf{p}^*$ (with $\mathbf{p}^* \geq \mathbf{0}$ and $\mathbf{1}^T\mathbf{p}^* = 1$) is a deterministic choice probability vector:

$$\mathbf{p}^* = \mathbf{p_D}(\mathbf{v}) \Leftrightarrow \mathbf{v}^T\mathbf{p}^* = \max(\mathbf{v})\mathbf{1}^T\mathbf{p}^* = \max(\mathbf{v})$$

Given a vector of deterministic probabilities $\mathbf{p}^* = \mathbf{p_D}(\mathbf{v})$, it follows that $\mathbf{v}^T\mathbf{p}^* = \max(\mathbf{v})$ since p_r^* can be positive only for an alternative r having maximum systematic utility, and conversely. Furthermore, the condition $\mathbf{1}^T\mathbf{p}^* = 1$ implies that $\max(\mathbf{v})\mathbf{1}^T\mathbf{p}^* = \max(\mathbf{v})$.

In general, for any vector of choice probabilities $\mathbf{p}$, since $\mathbf{1}^T\mathbf{p} = 1$ then, as observed earlier:

$$\mathbf{v}^T\mathbf{p} \leq \max(\mathbf{v})\mathbf{1}^T\mathbf{p} = \max(\mathbf{v}) \forall \mathbf{p} : \mathbf{p} \geq \mathbf{0}, \mathbf{1}^T\mathbf{p} = 1$$

Therefore, equality holds in the above relationship only for a vector of deterministic probabilities. Combining the two above relationships, the following basic relationship can be obtained:

$$(\mathbf{v} - \max(\mathbf{v})\mathbf{1})^T(\mathbf{p} - \mathbf{p_D}(\mathbf{v})) \leq 0 \forall \mathbf{p} : \mathbf{p} \geq \mathbf{0}, \mathbf{1}^T\mathbf{p} = 1$$

The deterministic utility model has properties (2) and (3) described above for probabilistic and invariant models. Regarding property (2), the Expected Maximum Perceived Utility of a deterministic model is a convex function of systematic utilities and is equal to the maximum systematic utility.

$$s(\mathbf{v}) = \max(\mathbf{v}) = \mathbf{p_D}(\mathbf{v})^T \mathbf{v}$$

This condition implies that, for a given vector of systematic utilities $\mathbf{v}$, the EMPU of a deterministic choice model is less than or equal to that of any probabilistic choice model involving the same systematic utility. A behavioural interpretation of this result suggests that the presence of random residuals makes the perceived utility for the chosen alternative, on average, larger than the alternative's systematic utility, which is the perceived utility in a deterministic choice model.

Regarding property (3), the deterministic choice map is monotone non-decreasing with respect to systematic utilities, just as are invariant probabilistic choice functions.

$$s(\mathbf{v}') - s(\mathbf{v}'') \geq \mathbf{p_D}(\mathbf{v}'')^T (\mathbf{v}' - \mathbf{v}'') \forall \mathbf{v}', \mathbf{v}''$$

$$s(\mathbf{v}'') - s(\mathbf{v}') \geq \mathbf{p_D}(\mathbf{v}')^T (\mathbf{v}'' - \mathbf{v}'') \forall \mathbf{v}', \mathbf{v}''$$

or

$$(\mathbf{p_D}(\mathbf{v}') - \mathbf{p_D}(\mathbf{v}''))^T (\mathbf{v}' - \mathbf{v}'') \geq 0 \forall \mathbf{v}', \mathbf{v}''$$

Indeed, $\max(\mathbf{v}') = (\mathbf{v}')^T \mathbf{p_D}(\mathbf{v}')$ and $\max(\mathbf{v}'') = (\mathbf{v})^T \mathrm{p_D}(\mathbf{v}'')$. Subtracting the last two equations term by term gives:

$$\max(\mathbf{v}') - \max(\mathbf{v}'') = (\mathbf{v}')^T \mathbf{p_D}(\mathbf{v}') - (\mathbf{v}'')^T \mathbf{p_D}(\mathbf{v}'') \tag{c}$$

Since:

$$(\mathbf{v}')^T \mathbf{p_D}(\mathbf{v}') = \max(\mathbf{v}') \geq (\mathbf{v}')^T \mathbf{p} \forall \mathbf{p}$$

for $\mathbf{p} = \mathbf{p_D}(\mathbf{v}'')$ it follows that:

$$(\mathbf{v}')^T \mathbf{p_D}(\mathbf{v}') \geq (\mathbf{v}')^T \mathbf{p_D}(\mathbf{v}'')$$

from which:

$$(\mathbf{v}')^T \mathbf{p_D}(\mathbf{v}') - (\mathbf{v}'')^T \mathbf{p}_\mathrm{D}(\mathbf{v}'') \geq (\mathbf{v}')^T \mathbf{p}_\mathrm{D}(\mathbf{v}'') - (\mathbf{v}'')^T \mathbf{p}_\mathrm{D}(\mathbf{v}'') \tag{d}$$

Therefore, combining Eqs. (c), (d) yields:

$$\max(\mathbf{v}') - \max(\mathbf{v}'') \geq (\mathbf{v}' - \mathbf{v}'')^T \mathbf{p_D}(\mathbf{v}'')$$

as above for random utility models.

A.7.2 Modelling uncertainty

Some approaches to uncertainty modelling are briefly reviewed below, for more details see for instance:

Klir, G.J., Yuan, B., 1995. Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, Upper Saddle River, NJ.

Klir, G.J., Wierman, M.J., 1999. Uncertainty-Based Information. Physica-Verlag, Heidelberg, New York.

Fuzzy and crisp sets

A *set* is a collection of elements, taken from a *universe class*, that show a relevant feature, with reference to the problem at the hand. Let.

- X be the universe class, containing all relevant values, for instance the set of real numbers R, or any n-dimensional extension R^n;
- $\varnothing$ be the empty set, assumed included within the universe class, as denoted by $\varnothing \subseteq X$;
- x, y be elements of the universe class, as denoted by $x \in X$, $y \in X$;
- A, B be sets containing elements of the universe class, as denoted by $A \subseteq X$, $B \subseteq X$.

Each *fuzzy* set within the universe class, $A \subseteq X$, may be described by a *membership function*: $\mu_A(x)$: $X \rightarrow [0, 1]$, which gives the membership grade of element x within set A. If the membership function may take only binary values 0 / 1, it is also called a *characteristic function*, $\chi_A(x)$: $X \rightarrow \{0, 1\}$, and the corresponding set is called *crisp*. A set A is said a *subset* of set B, as denoted by $A \subseteq B$, if $A = A \cap B$ or $\mu_A(x) \leq \mu_B(x)\ \forall x$.

Three basic operations can be defined on sets (all the other set operations can be obtained by duly combining them):

$C = {}^{-}A$ or $\neg A$ *complement* of set A (with respect to the universe class X), in this case
$\mu_{\neg A}$ (x) = 1 - μ_A (x);
$C = A \cup B$ *union* of sets A e B, with μ_C (x) = max{μ_A (x), μ_B (x)};
$C = A \cap B$ *intersection* of sets A e B, with μ_C (x) = min{μ_A (x), μ_B (x)}.

Moreover, the Zadeh extension principle makes possible to extend any function on real numbers to a function on fuzzy quantities.

Monotone set measures

The degree of uncertainty about how much an uncertainly defined object satisfied a feature expressed by a crisp set can be described by a *monotone set measure*. [The use of term fuzzy set measures use sometimes may be misleading since no fuzziness is modelled by these set measures.] Two well-known examples of such a measure are *possibility-necessity* and *probability measures*, special case of measures defined within the (Dempster-Shafer) *evidence theory*. Their main features are briefly reviewed below (for simplicity's sake max denotes both usual max or sup operators).

Possibility-necessity measures	**Probability measures**
Pos(A), Nec(A) (uncertainty as imprecision)	Pro(A) (uncertainty as dispersion)
Pos(A) $\in$ [0, 1], with Pos(X) = 1 and Pos($\emptyset$) = 0	Pro(A) $\in$ [0, 1], with Pro(X) = 1 and Pro($\emptyset$) = 0
$A \subseteq B \Rightarrow$ Pos(A) $\leq$ Pos(B)	$A \subseteq B \Rightarrow$ Pro(A) $\leq$ Pro(B)
Nec(A) = 1 – Pos(($\neg A$)	Pro(A) = 1 – Pro($\neg A$)
Pos($A \cup B$) = max{Pos(A), Pos(B)}	Pro($A \cup B$) + Pro($A \cap B$) = Pro(A) + Pro(B)
Nec($A \cap B$) = min{Nec(A), Nec(B)}	
Normalisation:	Normalisation:
max{Pos(A), Pos($\neg A$)} = 1	Pro(A) + Pro($\neg A$) = 1
With Pos(A) + Pos($\neg A$) $\geq$ 1	With max{Pro(A), Pro($\neg A$)} $\leq$ 1

Monotone set measures can be applied to fuzzy set too.

Uncertain numbers

According to above introduced monotone set measures, a real *uncertain number* U can be described by two types of variables; main features of their descriptive functions are briefly presented below.

Fuzzy numbers (f.n.) U	Continuous random variables (r.v.) U
Described by a fuzzy distribution function $\mu_U(x) \in [0,1]$, with	Described by a probability density function $\phi_U(x) \in [0,+\infty)$
$\{x \in X \mid \mu_U(x) > 0\}$ bounded	
$\{x \in X \mid \mu_U(x) = \alpha\}$ closed and convex $\forall \alpha \in (0,1]$	
Normalisation:	*Normalisation*:
$\max_{x \in X} \mu_U(x) = 1$	$\int_X \phi_U(x)\, dx = 1$
Relationship with possibility measures:	*Relationship with probability measures*:
$Pos(U = x) = \mu_U(x)$	$dPro(U = x) = \phi_U(x)\, dx$
$Pos(U \in [a,b]) = \max_{x \in [a,b]} \mu_U(x)$	$Pro(U \in [a,b]) = \int_{[a,b]} \phi_U(x)\, dx$

It is usually assumed that functions $\mu_U(x)$ and $\phi_U(x)$ are continuous over the *support set*, say $\{x \in X \mid \mu_U(x) > 0\}$ or $\{x \in X \mid \phi_U(x) > 0\}$ respectively. A special case is obtained when the object can precisely be defined, say $U = v$, with

Crisp numbers:	Deterministic variables:
$\mu_U(v) = 1,\ \mu_U(x) = 0\ \forall\ x \neq v\ x \in X$	$\phi_U(v) = 1,\ \rho\phi_U(x) = 0\ \forall\ x \neq v\ x \in X$

For random variables it is common practice to use the *mean* (value):

$$v = \int_X x\, \phi_U(x)\, dx$$

as a position index (less used are the mode and the median), as well as the *variance:*

$$\sigma^2 = \int_X (x - v)^2\, \phi_U(x)\, dx$$

(or derived indices) as a variability index.

For fuzzy numbers a *central value* v can be defined by the *barycentre* of $\mu_U(x)$ or by the *core value*, equal to middle point of the core set of $\mu_U(x)$

$$\{y : \mu_U(y) = \max_{x \in X} \mu_U(x) = 1\} = [u_1, u_2]$$

Clearly, if $\mu_U(x)$ is symmetrical, the central value and the core value are equal. (Sometimes a fuzzy number with a non-singleton core set is called a flat fuzzy number or better a fuzzy interval). The area of below $\mu_U(x)$,

$$\omega = \chi\delta$$

can be used as *variability index*, it is given by the product of the highness, equal to 1 due to normalisation, say the extension of the support set δ, and a factor, $\chi \in [0,1]$, depending on the shape of $\mu_U(x)$, (for instance $\chi = ½$ for a triangular fuzzy numbers).

The possibility that two fuzzy numbers U and T are equal is given by:

$$Pos(U = T) = \max_x \min\{\mu_U(x), \mu_T(x)\}$$

The possibility that the (fuzzy) maximum, MAX(U,T), of two fuzzy numbers, U and T, gets a value equal to z is given by:

$$Pos(MAX(U,T) = z) = \mu_{MAX(U,T)}(z) = \max_{z=\max(x,y)} \min\{\mu_U(x), \mu_T(y)\}$$

Several properties hold for MAX operator, idempotence, MAX(U, U) = U, commutativity, MAX(U, T) = MAX (T, U), and associativity, MAX(U, MAX(T, S)) = MAX(MAX(U, T), S). Associativity allow to extend the MAX operator to more than two numbers (the general expression is not reported for brevity's sake). It is also worth noting that the features of the MAX operator assure that the results is a fuzzy number (say $\max_{x \in X} \mu_{UMAX}(x) = 1$).

Possibility values given by the fuzzy distribution function of a fuzzy number can be transformed into choice proportions through uncertainty preserving transformation, see above quoted references for details.

References

A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, Econometrica 57 (5) (1989), 995–1026.

Avineri, E., Prashker, J.N., 2004. Violations of expected utility theory in route-choice stated preferences. Transp. Res. Rec. J. Transp. Res. Board 1894, 222–229.

Batley, R.P., Daly, A., 2003. Establishing equivalence between nested logit and hierarchical elimination-by-aspects choice models, paper presented at the European Transport Conference, Strasbourg, France.

Ben-Akiva, M., Lerman, S., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA.

Ben-Akiva, M., Morikawa, T., 1990. Estimation of switching models from revealed preferences and stated intentions. Transp. Res. A 24A (6), 485–496.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice models with applications to departure time and route choice. Handb. Transp. Sci. 26, 5–34.

Bekhor, S., Ben-Akiva, M., Ramming, M.S., 2006. Evaluation of choice-set generation algorithms for route choice models. Ann. Oper. Res. 144, 235–247.

Binetti, M., De Mitri, M.ì., 2002. Traffic assignment model with fuzzy travel cost. In: Proceedings of the 13th Mini_EURO Conference and the 9th Meeting of the EURO Working Group on Transportation, Bari (Italy) 10–13 June. Vol. 2002. Edited by the Local Organising Committee, Polytechnic of Bari, pp. 805–812.

Börsch-Supan, A., Hajivassiliou, V., 1993. Smooth unbiased multivariate probability simulators for limited dependent variable models. J. Econ. 58 (1993), 347–368.

Box, G.E., Cox, D.R., 1964. An analysis of transformations Journal of the Royal Statistical Society, Series B, 26, 211–243.

Cantarella, G.E., 1997. A general fixed-point approach to multimode multi-user equilibrium assignment with elastic demand. Transp. Sci. 31, 107–128.

Cantarella G.E. and Binetti M. G. (2002). Stochastic assignment with gammit path choice models. In Transportation Planning, M. Patriksson (ed.). Kluwer Academic Publisher. Printed in the Netherlands, 53–67.

Cantarella, G.E., de Luca, S., 2005a. Multilayer feedforward networks for transportation mode choice analysis: an analysis and a comparison with random utility models. Transp. Res. C 13, 121–155.

Cantarella, G.E., de Luca, S., 2005b. Modeling transportation choice through utility-based multi-layer feedforward networks. In: Attoh-Okine, N.O., Ayyub, B.M. (Eds.), Applied Research in Uncertainty Modeling and Analysis. International Series in Intelligent Technologies. In: vol. 20. Springer, Boston, MA.

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified Logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. In: Lesort, J.-B. (Ed.), Proceedings of the 13th International Symposium on Transportation and Traffic Theory. Pergamon, pp. 696–711.

Cascetta, E., Russo, F., Viola, F.A., Vitetta, A., 2002. A model of route perception in urban road networks. Transp. Res. B 36 (7), 577–592.

Cascetta, E., 2009. Transportation Systems Analysis: Models and Applications. Springer.

Cascetta, E., Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. Transp. Res. C 9 (6), 249–263.

Castillo, E., Menéndez, J.M., Jiménez, P., Rivas, A., 2008. Closed form expression for choice probabilities in the Weibull case. Transp. Res. B 42 (4), 373–380.

Chorus, C.G., Arentze, T.A., Timmermans, H.J.P., 2008. A random regret minimization model of travel choice. Transp. Res. B 42 (1), 1–18.

Chu, C., 1989. A paired combinatorial Logit model for travel demand analysis. In: Proceedings of 5th World Conference on Transportation Research, Ventura, CA.

Daganzo, C.F., 1979. Multinomial Probit: The Theory and its Application to Demand Forecasting. Academic Press, New York.

de Luca, S., Cantarella, G.E., 2009. Validation and comparison of choice models. In: Saleh, W., Sammer, G., (Eds.), Success and Failure of Travel Demand Management Measures. UK: Ashgate publications, 37–58.

de Luca, S., Di Pace, R., 2015. Evaluation of risk perception approach in route choice experiments: an application of the cumulative prospect theory. In: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. vol. 2015. pp. 309–315.

Dial, R.B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. Transp. Res. 5, 83–111.

Domencich, T.A., McFadden, D., 1975. Urban Travel Demand —A Behavioral Analysis. North Holland Publishing Company/Oxford American Elsevier Publishing Company Inc, Amsterdam/New York.

Dubois, D., Prade, H., 1983. Ranking fuzzy numbers in the setting of possibility theory. Inf. Sci. 30, 183–224.

Gao, S., Frejinger, E., Ben-Akiva, M., 2010. Adaptive route choice in risky traffic networks: A prospect theory approach. Transp. Res. C: Emerg. Technol. 18 (5), 727–740.

Gaudry, M., Dagenais, M., 1979a. The Dogit model. Transp. Res. B 13, 105–112.

Gaudry, M., Dagenais, M., 1979b. Heteroscedaticity and the use of Box-Cox transformations. Econo. Lett. 2–3, 225–229.

Gaudry, M., 1981. The inverse power transformation Logit and Dogit mode choice models. Transp. Res. B 15, 97–103.

Hajivassiliouì, V., McFadden, D., 1990. The Method of Simulated Scores for the Estimation of LDV Models. Cowles Foundation Discussion Paper No. 967. Yale University, New Haven, CT.

Hajivassiliou, V., McFadden, D., Ruud, P., 1996. Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. J. Econ. 72 (2), 85–134.

Henn, V., 2000. Fuzzy route choice model for traffic assignment. Fuzzy Sets Syst. 116 (1), 77–101.

Henn, V., 2002. What is the meaning of fuzzy costs in fuzzy traffic assignment models? In: Proceedings of the 13th Mini_EURO conference and the 9th Meeting of the EURO Working Group on Transportation, Bari (Italy) 10–13 June 2002. Local Organising Committee, Polytechnic of Bari, pp. 805–812.

Geweke, J., 1991. Efficient simulation from the Multivariate Normal and Student-T distributions subject to linear constraints and the evaluation of constraints probabilities. In: Meeting Computing Science and Statistics: The Twenty-Third Symposium on the Interface, Seattle.

Hensher, D.A., Button, K.J. (Eds.), 2000. Handbook of Transport Modelling. Pergamon Press.

Katsikopoulos, K.V., Duse-Anthony, Y., Fisher, D.L., Duffy, S.A., 2000. The framing of drivers' route choices when travel time information is provided under varying degrees of congnitive load. Human Factor 42 (3), 470–481.

Karlaftis, M.G., Vlahogianni, E.I., 2011. (2011). Statistics versus neural networks in transportation research: differences, similarities and some insights. Transp. Res. C Emerg. Technol. 19 (3), 387–399.

Kitthamkesorn, S., Chen, A., 2014. Unconstrained weibit stochastic user equilibrium model with extensions. Transp. Res. B 59, 1–21.

Keane, M.P., 1994. A computationally practical simulation estimator for panel data. Econometrica 62, 95–116.

Louviere, J.J., Hensher, D.A., Swaitì, J.F., 2000. Stated Choice Methods and Analysis. Cambridge University Press, Cambridge.

Luce, R.D., 1959. Individual Choice Behavior. John Wiley & Sons, Inc, New York, NY.

Mai, T., Bastin, F., Frejinger, E., 2017. On the similarities between random regret minimization and mother logit: The case of recursive route choice models. J. Choice Model. 23, 21–33.

Mansky, C., 1977. The structure of random utility models. Theor. Decis. 8, 229–254.

Marzano, V., Papola, A., 2008. On the covariance structure of the cross-nested Logit model. Transp. Res. B 42 (2), 83–98.

Morikawa, T., 1996. A hybrid probabilistic choice set model with compensatory and non compensatory rules. In: Proceedings of 7th WCTR, Sydney, Australia.

Klir, G.J., Wierman, M.J., 1999. Uncertainty-Based Information. Physica-Verlag, Heidelberg, New York.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econ. 15, 447–470.
McFadden, D., 2001. Economic choices. Am. Econ. Rev. 91 (3), 351–378.
Papola, A., 2016. A new random utility model with flexible correlation pattern and closed-form covariance expression: the CoRUM. Transp. Res. B 94, 80–96.
Prashker, J.N., Bekhor, S., 2004. Route choice models used in the stochastic user equilibrium problem: a review. Transp. Rev. 24 (6), 437–463.
Prato, C.G., 2013. Expanding the applicability of random regret minimization for route choice analysis Transportation, pp. 1–25.
Prato, C.G., Bekhor, S., 2007. Modeling route choice behavior: how relevant is the composition of choice set? Transp. Res. Rec. Iss. 2003, 64–71.
Rasouli, S., Timmermans, H.J.P. (Eds.), 2015. Bounded Rational Choice Behaviour: Applications in Transport. Emerald, Bingley.
Small, K.A., 1987. A discrete choice model for ordered alternatives. Econometrica 55 (2), 409–424.
Stern, S., 1992. A method for smoothing simulated moments of discrete probabilities in multinomial probit models. Econometrica 60, 943–952.
Swait, J., Ben-Akiva, M., 1987a. Incorporating random constraints in discrete choice models of choice set generation. Transp. Res. B 21, 91–102.
Swait, J., Ben-Akiva, M., 1987b. Empirical test of a constrained choice discrete model: mode choice in Sao Paulo Brazil. Transp. Res. B 21, 103–115.
Train, K., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.
Tversky, A., 1972a. Elimination by aspects: a theory of choice. Psych. Rev. 79 (4), 281–299.
Tversky, A., 1972b. Choice by elimination. J. Math. Psych. 9 (4), 341–367.
Vitetta, A., 2016. A quantum utility model for route choice in transport systems. Travel Behav. Soc. 3, 29–37.
Vovsha, P., 1997. The cross-nested Logit model: application to mode choice in the Tel-Aviv metropolitan area. In: Proceedings of the 76th TRB Meeting.
Vovsha, P., Bekhor, S., 1998. Link-nested Logit model of route choice. Overcoming route overlapping problem. Transp. Res. Rec. 1645, 133–142.
Walker, J., Ben-Akiva, M., Bolduc, D., 2007. Identification of parameters in normal error component Logit-mixture (NECLM) models. J. Appl. Econ. 22, 1095–1125.
Wardop, J.G., 1952. Some Theoretical Aspects of Road Traffic Research. In Proc. Inst. Civ. Eng. 2 (1), 325–378.
Washington, S., Karlaftis, M., Mannering, F., 2003. Statistical and Econometricmethods for Transportation Data Analysis. CRC Press, Boca Raton, FL.
Williams, H.C.W.L., 1977. On the formation of travel demand models and economic evaluation measures of user benefit. Environ. Plan. A. 285–344.
Zhou, Z., Chen, A., Behkor, S., 2012. C-Logit stochastic user equilibrium model: formulations and solution algorithm. Transportmetrica 8 (1), 17–41.

Further reading

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–291.

Kitthamkesorn, S., Chen, A., 2013. A path-size weibit stochastic user equilibrium model. Transp. Res. B 57, 378–397.

Koppelman, F.S., Wen, C.H., 2000. The paired combinatorial Logit model: properties, estimation and application. Transp. Res. B 34 (2), 75–89.

Klir, G.J., Yuan, B., 1995. Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, Upper Saddle River, NJ.

Wen, C.H., Koppelman, F.S., 2001. The generalized nested Logit model. Transp. Res. B 35 (7), 627–641.

APPENDIX

Traffic flow theory

B

Roberta Di Pace
University of Salerno, Salerno, Italy

The enjoyment of scientific research also means coming up against obstacles to overcome, coming up with even better investigation tools and even more complex theories while endeavouring to always move forward despite knowing that we will likely get closer to comprehending reality, without ever fully being able to understand it.
Margherita Hack

Outline. Preliminary studies on traffic flow theory may be found in literature since 1930s (Greenshields, 1935) as the congestion phenomena increased due to the impact of vehicles interactions.

In the first part of the appendix the observable variables will be initially defined and then the relationships between them and models are introduced and analysed. In particular phenomena along links (*running links*) are discussed in accordance with the *uninterrupted flow theory* whilst queuing phenomena are analysed in accordance with the *interrupted flow theory* (*queuing links*). Depending on the nature of the input variables, *stationary* and *non-stationary* conditions may be identified then phenomena respectively related to the running and queuing links are analysed with respect to each one of two conditions.

The second part of the appendix focuses on non-stationary models. The approaches classification is based on the level of aggregation of traffic flow variables; in particular, as discussed in more detail in Section B.2, users and supply variables are distinguished in aggregate or disaggregate.

Firstly macroscopic models may be identified in which users' behaviour variables are aggregate (arc density or entry flows can be obtained from the vehicle position on the arc) as well as level of service variables (space mean speed, arc performance functions are derived from fundamental diagram). Furthermore they may be classified in accordance with time and space. Then there are the mesoscopic models in which users' behaviour variables are disaggregate (packets of users or single users are considered; arc density or entry flows can be obtained from packets/users position on the arc) and the level of service variables are aggregate (such as space mean speed; arc performance functions are derived from the fundamental diagram).

Finally, microscopic models able to describe traffic flow dynamics in terms of a single vehicle may be identified. Indeed, in this models users' behaviour variables are disaggregate (single users are considered; link density or entry flows can be obtained from the users' position on the link) as well as the level of service variables (time speed and link performance functions are derived from the drivers' behaviour models such as car-following models).

In the perspective of extended applications, each model is also analysed with respect to the network equations that are introduced in order to define the outflow rates from incoming links at each node within the network.

The chapter is organised as follows. First we introduce basic notations and definitions in Section B.1, by distinguishing stationary from non-stationary models then we discuss each class of non-stationary models; in particular in Sections from B.2 to B.4 macroscopic models are analysed, whilst mesoscopic and microscopic models are respectively displayed in Sections B.5 and B.6.

Finally it must also be highlighted that some of the symbols adopted in this appendix will appear with a different meaning from the rest of the book chapters.

B.1 Basic TFT

The first aim of the section is to provide an overview of the main observable variables describing the running and queuing phenomena. Then the variables classification will be related to the running and queuing links. Starting from these observable variables and depending on nature, stationary or non-stationary, of the input variables, two classes of models are presented: *steady state and non-steady state models*.

Main variables are enlisted below in alphabetical order for reader's convenience (notations come first, then Roman letters, at last Greek letters):

a is an index denoting a arc;
Cap is the capacity of the road measured in vehicles per unit of time;
f is the flow measured in vehicles per unit of time;
f_K is the equilibrium flow as a function of density speed;
f_{IN} is the entering flow;
f_{OUT} is the exiting flow during;
f_V is the equilibrium flow as a function of density speed;
h is an index denoting the headway between successive users measured in time per unit of vehicle;
i is an index denoting an observed vehicle/user;
k is the density measured in vehicles per unit of length;
l is the length of road segment;
l_a is the length of road segment corresponding to arc a;

m is the number of vehicles traversing a point in a time interval;
n is the number of vehicles between two points at a given time;
p is the pressure term reflecting the speed variance and then the effect of different vehicles;
s is the spacing between vehicles at a given time;
t is the time at which the system/traffic is observed;
t_{si} is the service time of user i;
t_{wi} is the total waiting time of user i;
$u(t)$ is the function representing the vehicles cumulative *arrival function*;
v is the speed measured in space per unit of time;
v_0 is the free flow speed measured in space per unit of time;
v_i is the speed of vehicle i measured in space per unit of time;
v_k is the equilibrium speed as a function of density k;
v_S is the space mean speed, among all vehicles between two points at a given time;
v_T is the time mean speed, among all vehicles crossing a point during time interval.
$w(t)$ is the function representing the vehicles cumulative *departure function*.
x is a point along an arc, or rather, its abscissa increasing (from a given origin, usually located at the beginning of the arc) along the traffic direction ($s \in [0, l_a]$);
τ is the relaxation time representing the aggressiveness of drivers;
Δt is the time variation;
Δx is a position variation;
φ diffusion term in Payne's model.

B.1.1 Fundamental variables

Running links

Running links are introduced in order to describe the vehicles interactions along links; indeed vehicles using the same link may interact with each other and the level of interaction depends on the demand. In particular, if the demand is great enough that the interaction may affect the link performances in terms of mean speed and travel time the congestion phenomenon may occur.

In general, modelling may be based on deterministic or stochastic approaches depending; it is very often sufficient to adopt the aggregate deterministic models described below in case of running links whilst stochastic models may also be used in case of queuing representation in order to characterise an interaction event that causes a delay in a probabilistic sense.

The observable variables will be initially defined and then the relationship between some observed variables and uninterrupted flow models in stationary and non-stationary conditions will be introduced.

Let us consider a road infrastructure which may be represented by a segment, several variables may be referred to the arc and representing the vehicles moving along a road segment. In particular, for traffic observed at time t in a road segment $[x, x+\Delta x]$ the variables to be defined are:

$s_i(t)$ the spacing between vehicles i and $i-1$ at time t; that is the front spacing vehicle to vehicle at time t;
$n(t; x, x+\Delta x)$ the number of vehicles at time t between points x and $x+\Delta x$;
$s(t)=\Sigma_{i=1, \ldots, n}\, s_i(t)/n(t; x, x+\Delta x)$ the mean spacing, among all vehicles between points x and $x+\Delta x$ at time t; and
$v_i(s, t)$ the speed of vehicle i at time t while traversing point (abscissa) x;

For traffic observed at point s during time interval $[t, t+\Delta t]$, several variables can be defined:

$h_i(x)$ the headway (temporal spacing) between vehicles i and $i-1$ crossing point x;
$m(x; t, t+\Delta t)$ the number of vehicles traversing point x during time interval $[t, t+\Delta t]$;
$h(x) = \Sigma_{i=1,\ldots,m}\, h_i(x)/m(x; t, t+\Delta t)$ the mean headway, among all vehicles crossing point x during time interval $[t, t+\Delta t]$;
$\mathrm{v_S}(t) = \Sigma_{i=1,\ldots,n}\, v_i/n(t; x, x+\Delta x)$ the space mean speed, among all vehicles between points x and $x+\Delta x$ at time t.
$v_T(x) = \Sigma_{i=1, \ldots, m}\, v_i\,(x)/m(x; t, t+\Delta t)$ the time mean speed, among all vehicles crossing point y during time interval $[\tau, \tau+\Delta t]$.

Some other relationships may be defined:

$f(x; t, t+\Delta t) = m(x; t, t+\Delta t)/\Delta t$ is the flow of vehicles crossing point x during time interval $[t, t+\Delta t]$, measured in vehicles per unit of time;
$k(t; x, x+\Delta x) = n(t; x, x+\Delta x)/\Delta x$ is the density between points x and $x+\Delta x$ at time t, measured in vehicles per unit of length.

Regarding the variation in the number of vehicles between points x and x + Δx during Δt and the variation in the number of vehicles during time interval [t, t + Δt] over space Δx they may be respectively represented as in the following:

$$\Delta n(x, x+\Delta x; t, t+\Delta t) = n(t+\Delta t; x, x+\Delta x) - n(t\,; x, x+\Delta x);$$
$$\Delta m(x, x+\Delta x; t, t+\Delta t) = m(x+\Delta x\,; t, t+\Delta t) - m(x\,; t, t+\Delta t);$$

thus the general the flow conservation may be formulated in accordance with the *flow conservation equation* as follows:

$$n(t; x, x+\Delta x) + m(x; t, t+\Delta t) = m(x+\Delta x; t, t+\Delta t) + n(t+\Delta t; x, x+\Delta x) \tag{B.1}$$

Furthermore let.

$\Delta z(x, x + \Delta x, t, t + \Delta t)$ be the number of entering minus exiting vehicles (if any) during time interval $[t, t + \Delta t]$, due to entry/exit points (e.g. on/off ramps), between points x and $x + \Delta x$ the flow conservation equation may be generalised as

$$\Delta n(x, x+\Delta x, t, t+\Delta t) + \Delta m(x, x+\Delta x, t, t+\Delta t) = \Delta z(x, x+\Delta x, t, t+\Delta t) \tag{B.2}$$

Furthermore, let

$\Delta f(x, x + \Delta x, t, t + \Delta t) = \Delta m(x, x + \Delta x, t, t + \Delta t)/\Delta t$ be the variation of the flow over space;

$\Delta k(x, x + \Delta x, t, t + \Delta t) = \Delta n(x, x + \Delta x, t, t + \Delta t)/\Delta s$ be the variation of the density over time.

$\Delta e(x, x + \Delta x, t, t + \Delta t) = \Delta z(y, x + \Delta x, t, t + \Delta t)/\Delta t$ be the (net) entering/exiting flow.

if the *general flow conservation* Eq. (B.2) is divided by Δt, it may also be rewritten as:

$$\Delta n/\Delta t + \Delta f = \Delta e \tag{B.3}$$

and finally if the equation is divided again by Δs it may be rewritten as

$$\Delta k/\Delta t + \Delta f/\Delta x = \Delta e/\Delta x \tag{B.4}$$

Queuing links

The average delay experienced by vehicles that queue to cross a flow interruption point (intersections, toll barriers, merging sections, etc.) is affected by the number of vehicles waiting. This phenomenon may be analysed with models derived from queuing theory, developed to simulate any waiting or user queue formation at a server (administrative counter, bank counter, etc.). Below the subject is treated with reference to generic users, at the same time highlighting the similarities with uninterrupted flow.

The main variables that describe the queuing phenomena are:

$h_i = t_i - t_{i-1}$ the headway between successive users i and $i - 1$ joining the queue at times t_i and t_{i-1};

$m_{IN}(t, t+\Delta t)$ number of users joining the queue during $[t, t+\Delta t]$;

$m_{OUT}(t, t+\Delta t)$ number of users leaving the queue during $[t, t+\Delta t]$;

$h(t, t + \Delta t) = \Sigma_{i=1, \ldots, m}\, h_i/m_{IN}(t, t+\Delta t)$ mean headway between all vehicles joining the queue in the time interval $[t, t+\Delta t]$;

$n(t)$ number of users waiting to exit (*queue length*) at time t;

With reference to observable quantities, flow variables may be introduced:

$f_{IN}(t, t + \Delta t) = m_{IN}(t, t + \Delta t)/\Delta t$ *arrival* (entering) flow during $[t, t + \Delta t]$;

$f_{OUT}(t, t+\Delta t) = m_{OUT}(t, t+\Delta t)/\Delta t$ *exiting* flow during $[t, t + \Delta t]$;

Note that the main difference with the basic variables of running arcs is that space $(x, \Delta x)$ is no longer explicitly referred to since it is irrelevant. Some of the above variables are shown in Fig. B.1.

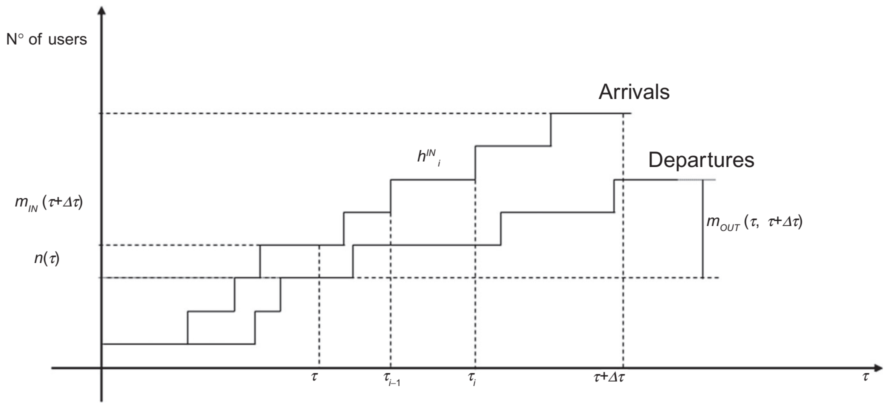


FIG. B.1

Fundamental variables for queuing systems.

B.1.2 Steady-state models

In this section we describe several deterministic models developed under the assumption of stationarity, running phenomena along links are analysed in accordance with the *uninterrupted flow theory* whilst queuing phenomena at network nodes are analysed in accordance with the *interrupted flow theory*. All models are formally presented in each section about running links and queuing links.

Running links

In formulating such steady state models for running links models it is assumed that a traffic stream (a discrete sequence of vehicles) is represented as a continuous (one-dimensional) fluid.

Traffic flow is called *stationary* during a time interval $[t, t + \Delta t]$ between points x and $x + \Delta x$ under the following conditions.

- flow is (on average) independent of point s, hence $f(x; t, t+\Delta t) = f$
- density is independent of time t, hence $k(\text{t}; x, x + \Delta x) = k$
- time mean speed is independent of location and the space mean speed is independent of time:

$$\bar{v}_T(s) = \bar{v}_T \text{ and } \bar{v}_S(t) = \bar{v}_S$$

In the case of stationarity, both terms in the left side of the conservation equation are identically null, anyhow other flow conservation conditions may be formulated.

In particular in the case of stationarity

- the number of vehicles crossing each cross-section during time Δt is equal to $f\,\Delta t$;
- the number of vehicles, time-independent due to the assumption of stationarity, on the stretch of road between cross-sections x and $x + \Delta x$, equals $k\,\Delta x$;

v_s, the space mean speed of these vehicles on the stretch of road equals to $k\,v_s\,\Delta t$.

Thus the number of vehicles in the stretch of road will be made up of vehicles entering the section during the time interval Δt and $f\,\Delta t = k\,v_S\,\Delta t$. Hence, under stationary conditions, if the previous equation is divided by Δt, flow, density and space mean speed must satisfy the *stationary flow conservation equation:*

$$f = k\,v \tag{B.5}$$

where $v = \bar{v}_S$ is the space mean speed, simply called speed for further analysis of stationary conditions.[a]

Flow and density are related to mean headway and mean spacing through the following relations:

$$f(x;\,t,\,t+\Delta t) \cong 1/\bar{h}(s) \text{ and } k\,(t;\,x,\,x+\Delta x) \cong 1/\bar{s}(t)$$

thus the stationary conservation equation may be rewritten as follows:

$$f \approx (1/\bar{s})\,v \text{ then } (1/\bar{h}) \approx k\,v \text{ and } \bar{s} \approx \bar{h}v.$$

In stationary conditions, empirical relationships can be observed between each pair of variables: flow, density, speed. In general, observations are rather scattered (see Fig. B.2 for an example of a speed-flow empirical relationship) and various models may be adopted to describe such empirical relationships.

These models are generally given the name *fundamental diagram (of traffic flow)* (see Fig. B.3) and are specified by the following relations:

$$v_K = v(k) \tag{B.6}$$

$$f_K = f(k) \tag{B.7}$$

$$f_V = f(v) \tag{B.8}$$

Though only a model representation of empirical observations, this diagram permits some useful considerations to be made. It shows that flow may be zero under two conditions: when density is zero (no vehicles on the road) or when speed is zero (vehicles are not moving). The latter corresponds in reality to a stop-and-go condition.

In the first case the speed assumes the theoretical maximum value, *free - flow speed*, v_0, while in the second the density assumes the theoretical maximum value,

[a] It is worth noting that the time mean speed is not less than the space mean speed, as can be shown since the two speeds are related by the equation $\bar{v}_T = \bar{v}_s + \sigma_2/\,\bar{v}_S$, where σ^2 is the variance of speed among vehicles. $\sigma^2 = 0$, hence $\bar{v}_\tau = \bar{v}_s \ldots$ var.(v) $= \Sigma_{i=1..n}\,(v_i - v)^2/n$.

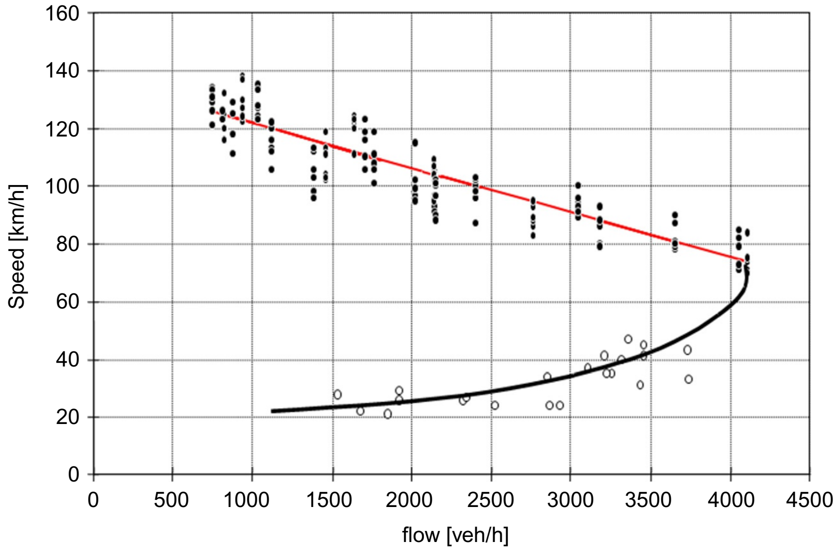


FIG. B.2

Relationship between speed and flow.

Source: Cascetta, E., 2009. Transportation Systems Analysis: Models and Applications. Springer, pp. 448–477.

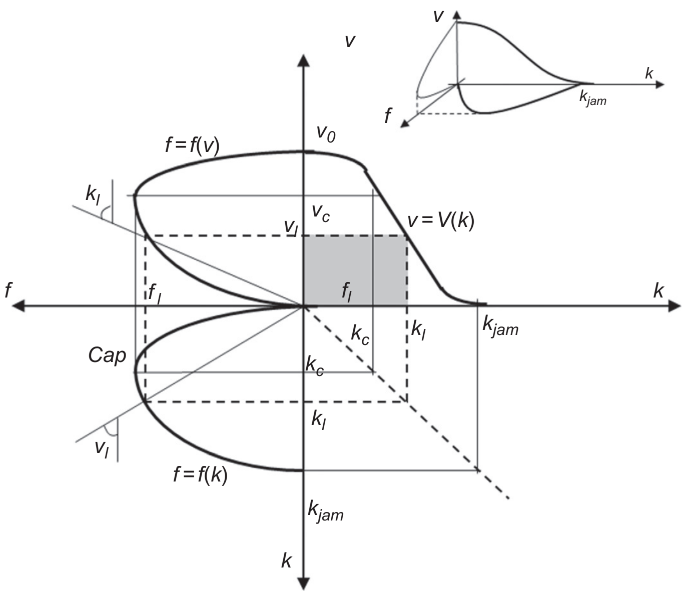


FIG. B.3

Fundamental diagram of traffic flow.

Source: Cascetta, E., 2009. Transportation Systems Analysis: Models and Applications. Springer, pp. 448–477.

jam density, k_{jam}. Therefore, a traffic stream may be modelled through a *partially compressible fluid*, i.e. a fluid that can be compressed up to a maximum value.

The peak of the *speed - flow* (and *density - flow*) curve occurs at the theoretical maximum flow, *capacity*, Cap, of the facility; the corresponding speed v_c and density k_c are referred to as the *critical speed* and the *critical density*. Thus any value of flow (except the capacity) may occur under two different conditions: low speed and high density and high speed and low density. The first condition represents an *unstable* state for the traffic stream, where any increase in density will cause a decrease in speed and thus in flow. This action produces another increase in density and so on until traffic becomes jammed. Conversely, the second condition is a *stable* state since any increase in density will cause a decrease in speed and an increase in flow. At capacity (or at critical speed or density) the stream is *non-stable*, this being a boundary condition between the other two.

These results show that flow cannot be used as the unique parameter describing the state of a traffic stream; speed and density, instead, can univocally identify the prevailing traffic condition. For this reason the relation $v_K = v(k)$ is preferred to study traffic stream characteristics.

Mathematical formulations have been widely proposed for the fundamental diagram, based on *single regime* or *multi-regime* functions.

An example of a single regime function is Greenshields' linear model:

$$v(k) = v_0 \left(1 - k/k_{jam}\right)$$

or Underwood's exponential model (useful for low densities):

$$v(k) = v_0 \, e^{-k/kc}.$$

An example of a multi-regime function is Greenberg's model:

$$v(k) = a_1 \ \ln\left(a_2/k\right) \text{ for } k > k_{min}$$
$$v(k) = a_1 \ \ln\left(a_2/k_{min}\right) \text{ for } k \le k_{min}$$

where a_1, a_2 and $k_{min} \le k_{jam}$ are constants to be calibrated.

Starting from the speed-density relationship, the flow-density relationship, $f_K = f(k)$, may be easily derived by using the flow conservation equation under stationary conditions, or fundamental conservation equation:

$$f(k) = v(k)\, k$$

Greenshields' linear model yields:

$$f(k) = v_0 \left(k - k^2/k_{jam}\right)$$

In this case the capacity is given by:

$$Cap = v_0 \, k_{jam}/4$$

Moreover the flow-speed relationship can be obtained by introducing the inverse speed-density relationship: $k = v^{-1}(v)$, thus

$$f(v) = v\left(k = v^{-1}(v)\right) v^{-1}(v) = v\, v^{-1}(v).$$

For example, Greenshields' linear model yields: $v^{-1}(v) = k_{jam}\,(1 - v/v_0)$ thus.

$$f(v) = k_{jam}\left(v - v^2/v_0\right)$$

Continued

In general, the flow-speed relationship may be inverted by only considering two different relationships, one in a stable regime, $v \in [v_c, v_o]$, and the other in an unstable regime, $v \in [0, v_c]$.

Greenshield's linear model leads to:

$$v_{stable}(q) = (v_0/2)\left(1 + \left(1 - 4f/\left(v_0\, k_{jam}\right)\right)\right)^{(1/2)} = (v_0/2)(1 + (1 - f/Cap))^{(1/2)}$$

$$v_{unstable}(q) = (v_0/2)(1 - (1 - f/Cap))^{(1/2)}$$

In the particular case that one can assume the flow regime is always stable, with reference to relation $v = v_{stable}(f)$ the corresponding relationship between travel time, t, and flow may be defined as

$$t = t(f) = l/v_{stable}(f)$$

Other models have been proposed in the literature; these models are listed below with respect to the speed–density relationship:

$$\text{Drew}: V(k) = v_0 \left[1 - \left(k/k_{jam}\right)^{0.5}\right]$$

$$\text{Greenberg}: V(k) = -v^*_0 \;\ln\left(k/k_{jam}\right)$$

$$\text{Underwood}: V(k) = v_0 \exp\left[-k/k_0\right]$$

$$\text{Drake}: V(k) = v_0 \exp\left[-(k/k_0)^2/2\right]$$

The first two models are very similar to Greenshields's approach and still fail in simulating the low or critical densities situations. In particular Greenberg's model is very limited in simulating the high speed scenarios corresponding to the low density situations. Regarding the Underwood and Drake models, these consider the density k_o in reference to the road capacity.

Greenberg's model tends towards an infinite value in case of density which tends towards zero value thus $v_0/3$ (v_0*) is usually adopted in place of v_0. In general Underwood and Drake provide a non-null speed value for $k_{jam.}$

For all models the flow density relationships may be easily derived by using the flow conservation equation under stationary conditions, or fundamental conservation equation: $f(k) = v(k)k$, as in the previous case of Greenshields's models.

Queuing links—Deterministic models

In this subsection we describe several deterministic models developed under the assumption that the arrival flow and the service time are represented by deterministic variables. The following definitions may be introduced:

$t_s(t, t + \Delta t)$ average service time among all users joining the queue in time interval $[t, t + \Delta t]$;

$t_w(t, t + \Delta t)$ average total waiting time among all users joining the queue in time interval $[t, t + \Delta t]$;

$f(t, t + \Delta t) = 1/t_s(t, t + \Delta t)$ the (trasversal[b]) capacity or maximum exit flow, i.e. the maximum number of users that may be served in the time unit, assumed constant during $[t, t + \Delta t]$ for simplicity's sake (otherwise Δt can be redefined).

[b] In some cases it is also necessary to introduce longitudinal capacity, i.e. the maximum number of users that may form the queue.

Regarding the capacity constraint on exiting flow it is expressed by: $f_{OUT} \leq f$.

A general conservation equation, introduced for uninterrupted flow, holds in this case:

$$n(t) - n(t+\Delta t) = m_{OUT}(t, t+\Delta t) - m_{IN}(t, t+\Delta t) \tag{B.9}$$

Moreover, dividing by Δt we obtain:

$$\Delta n/\Delta t + [f_{OUT}(t, t+\Delta t) - f_{IN}(t, t+\Delta t)] = 0 \tag{B.10}$$

Deterministic queuing systems can also be analysed through the cumulative number of users that have arrived at the *server* by time t, and the cumulative number of users that have departed from the *server* (leaving the queue) at time t, as expressed by two functions termed *cumulative arrival function*, $u(t)$, that is the cumulative number of users that arrive and *cumulative departure function*, $w(t)$ that is the cumulative number of users that leave constrained to the cumulative arrival function as in the following:

$$u(t) \leq w(t)$$

Queue length $n(t)$ at any time t is given by:

$$n(t) = u(t) - w(t) \tag{B.11}$$

provided that the queue at time 0 is given by $n(0) = u(0) \geq 0$ with $w(0) = 0$. The arrival and departure functions are linked to entering and exiting users by the following relationships:

$$m_{IN}(t, t+\Delta t) = u(t+\Delta t) - w(t) \tag{B.12}$$

$$m_{OUT}(t, t+\Delta t) = u(t+\Delta t) - w(t) \tag{B.13}$$

If during time interval $[t_0, t_0 + \Delta t]$ the entering flow is constant over time, $f_{IN}(t) = -f_{IN}$, then the queuing system is named (*flow*) *stationary* and the arrival function $u(t)$ is linear with slope given by $-u$:

$$u(t) = u(t_0) + -f_{IN}(t - t_0)\, t \in [t_0, t_0 + \Delta t]$$

The exit flow may be equal to the entering flow, $\bar{f}_{IN}$, or to the capacity, *Cap*, as described below in more detail.

When the arrival flow is less than capacity, $\bar{f}_{IN} < Cap$, the system is *under-saturated* and two conditions may occur: (i) the queue length at the beginning of period is zero or (ii) there is a queue at time t_0.

If the queue length at the beginning of period is zero.

$$n(t_0) = 0$$

$$n(t) = 0$$

$$n(t) = n(t_0) - (Cap - \bar{f}_{IN})\,\Delta t \text{ and } \bar{f}_{OUT} = \bar{f}_{IN}$$

If there is a queue at time t_0, its length decreases with time and vanishes after a time $\Delta t = t - t_0$.

Before time $t_0 + \Delta t$, the queue length is linearly decreasing with t and the exiting flow $\bar{f}_{OUT}$ is equal to capacity Cap:

$$\begin{aligned} &n(t) = n(t_0) - (Cap - \bar{f}_{IN})\,\Delta t \\ &f_{OUT} = Cap \\ &w(t) = w(t_0) + Cap\,\Delta t \end{aligned} \tag{B.14a}$$

After time $t_0 + \Delta t$ the queue length is zero and the exiting flow $\bar{f}_{OUT}$ is equal to the arrival flow $\bar{f}_{IN}$:

$$\begin{aligned} &n(t_0 + \Delta t) = 0 \\ &\bar{f}_{OUT} = \bar{f}_{IN} \\ &\Delta t = n(t_0)/(\text{Cap} - f_{IN}) \\ &w(t) = u(t) = u(t_0) + \bar{f}_{IN}\,\Delta t \end{aligned} \tag{B.14b}$$

When the arrival flow rate is larger than capacity, $\bar{f}_{IN} \geq Cap$, the system is *over-saturated*. As in case of under-saturated conditions two situations may occur: (i) the queue length at the beginning of period is zero or (ii) there is a queue at time t_0.

If the queue length at the beginning of period is zero

$$\begin{aligned} &n(t) = (\bar{f}_{IN} - Cap)\,\Delta t_0 \\ &\bar{f}_{OUT} = Cap \\ &w(t) = w(t_0) + Cap\,\Delta t_0 \end{aligned} \tag{B.15a}$$

If there is a queue at time t_0, the queue length linearly increases with time t and the exiting flow is equal to the capacity Cap

$$\begin{aligned} &n(t) = n(t_0) + (\bar{f}_{IN} - Cap)\,\Delta t_0 \\ &\bar{f}_{OUT} = Cap \\ &w(t) = w(t_0) + Cap\,\Delta t_0 \end{aligned} \tag{B.15b}$$

By comparing Eqs (B.14a), (B.14b), (B.15a), (B.15b) it is possible to formulate this general equation for calculating the queue length at generic time instant t:

$$n(t) = \text{MAX}\left\{0, (n(t_0) + (\bar{f}_{IN} - Cap)\,\Delta t_0) \geq 0\right\} \tag{B.16}$$

With the above results, any general case can be analysed by modelling a sequence of periods during which arrival flow and capacity are constant.

Finally the delay can be defined as the time needed for a user to leave the system (passing the server), accounting for the time spent queuing (pure waiting). Thus the delay is the sum of two terms:

$$t_w = t_s + t_{wq} \tag{B.17}$$

where t_w is the total delay; $t_s = 1/Cap$ is the average service time (time spent at the server); and t_{wq} is the queuing delay (time spent in the queue).

In under-saturated conditions ($\bar{f}_{IN} < Cap$) if the queue length at the beginning of period is zero, the queuing delay is equal to zero, $t_{wq}(\bar{f}_{IN}) = 0$, and the total delay is equal to the average service time: $t_w = t_s$ otherwise $t_{wq}(\bar{f}_{IN}) = n(t_0)\,\Delta t_0/2$ and then t_w is derived.

In over-saturated conditions ($\bar{f}_{IN} \geq Cap$), the queue length, and respective delay, would tend towards infinity in the theoretical case of a stationary phenomenon lasting for an infinite time. In practice, however, over-saturated conditions last only for a finite period, T.

If the queue length is equal to zero at the beginning of the period, it will reach a value $(\bar{f}_{IN} - Cap)\,T$ at the end of the period. Thus, the average queue over the whole period T is:

$$\bar{n} = (\bar{f}_{IN} - \mathrm{Cap})\,\mathrm{T}/2 \tag{B.18}$$

In this case the average queuing delay is $\bar{n}/Cap$, and the average total delay is

$$t_w(\bar{f}_{IN}) = t_s + (\bar{f}_{IN} - Cap)\,T/2\,Cap \tag{B.19}$$

Queuing links—Stochastic models

In order to properly apply the deterministic models the simulation interval is usually discretized in sub-intervals during which it is supposed that the input variables (flows and capacity) are stationary. However if the flow fluctuations are observed and then these cannot be modelled through deterministic models, stochastic models are required.

If the system is under-saturated, it can be analysed through (stochastic) queuing theory which includes the particular case of the deterministic models already discussed.

It is particularly necessary to specify the stochastic process describing the sequence of user arrivals, the sequence of service times and the queue discipline.

Main variables are enlisted below in alphabetical order for reader's convenience (notations come first, then Roman letters, at last Greek letters).

f_{IN}, is the arrival rate or the expected value of the arrival flow;
$Cap = 1/t_s$, is the service rate (or capacity) of the system, the inverse of the expected service time;
$f_{IN}\,/Cap$, is the traffic intensity ratio or utilisation factor;
n is a value of the random variable N, number of users present in the system, consisting of the number of users queuing plus the user present at the server, if any (the significance of the symbol n is thus slightly different);
tw is a value of the random variable TW, the time spent in the system or overall delay, consisting of queuing time plus service time.

The queuing discipline:

FIFO = First In – First Out (i.e. service in order of arrival);
LIFO = Last In – First Out (i.e. the last user is the first served);
SIRO = Service In Random Order;
HIFO = High In – First Out (i.e. the user with the maximum value of an *indicator* is the first served).

The probability density function describing the intervals between two successive arrivals/departures:

D = deterministic variable.
M = negative exponential random variable.
E = Erlang random variable.
G = general distribution random variable.

The probability density function describing the intervals between two successive departures.

The characteristics of a queuing phenomenon can be redefined in the following concise notation:

$$a/b/c\,(d, e)$$

where a denotes the type of arrival pattern represented as already described above;
b denotes the type of departure represented as already described above; c is the number of service channels: {1, 2, …}; d is the queue storage limit: $\{\infty, n_{max}\}$ or longitudinal capacity; and e denotes the queuing discipline represented as already described above.

Fields d and e, if defined respectively by ∞ (no constraint on maximum queue length) and by *FIFO*, are generally omitted.

In the following we will report the main results for the $M/M/1$ and the $M/G/1$ queue systems, which are commonly used for simulating transportation facilities, such as signalised intersections.

Some definitions or notation differ from those traditionally adopted in dealing with queuing theory (the relative symbols are in brackets) so as to be consistent with those adopted above. The parameters defining the phenomenon are as follows:

$M/M/1$ systems

In *under-saturated* conditions ($f_{IN}/Cap < 1$):

The expected number of users in the system may be generally defined as

$$E[N] = (f_{IN}/Cap)/(1 - f_{IN}/Cap) = f_{IN}/(Cap - f_{IN})$$

Let

$$\mathrm{E[TW]} = 1/(1 - f_{IN}/\mathrm{Cap})$$

then in accordance with Little's formula.

$$E[N] = f_{IN}\,E[TW]$$
$$VAR[N] = (f_{IN}/Cap)/(1 - f_{IN}/Cap)^2$$

The expected time spent in the queue, $E[tw_q]$, (or queuing delay) is given by the difference between the expected delay, $E[tw]$, and the average service time $t_s = 1/Cap$:

$$E\left[TW_q\right] = 1/(Cap - f_{IN}) - 1/Cap = f_{IN}/Cap\,(Cap - f_{IN})$$

According to Little's second formula, the expected value of the number of users in the queue, $E[N_q]$, is the product of the expected queuing delay, $E[TW_q]$, multiplied by the arrival rate, u:

$$E[N_q] = f_{IN}\, E[TW_q]$$

and then:

$$E[N_q] = f_{IN}{}^2 / Cap(Cap - f_{IN})$$

$M/G/1$ (∞, FIFO) systems

In this case the main results are the following:

$$E[N] = (f_{IN}/Cap)(1 + f_{IN}/2(Cap - f_{IN}$$
$$E[TW] = 1/Cap\,(1 + f_{IN}/2(Cap - f_{IN}))$$
$$E[TW_q] = f_{IN}/2Cap(Cap - f_{IN})$$

Non-steady state models

Running links

In the case of non-stationary models the main variable describing the uninterrupted flow conditions, will be considered as a function of space and time. In particular flow speed and density will be represented as a function of space and time. In this model, also called *first order* model, the macroscopic variables will be represented as.

$$f = f(x, t)$$
$$v = v(x, t)$$
$$k = k(x, t)$$

The observed variables, m(x; t, t+Δt), the number of vehicles traversing point x during time interval $[t, t + \Delta t]$ and $n(t; x, x + \Delta x)$, the number of vehicles at time t between points x and $x + \Delta x$, can be averaged (flow with respect to space and density with respect to time) hence density and flow will be consistently defined.

Thus flow is related to m through following equation

$$m(x, t, t + \Delta T) = \int_t^{t+\Delta T} f(x, t)\, dt \tag{B.20}$$

density is related to n through following equation

$$n(t, x, x + \Delta x) = \int_y^{y+\Delta x} k(x, t)\, dx \tag{B.21}$$

In this context the continuity equation is a partial differential equation for the macroscopic quantities' density, flow and speed. It is able to describe the density variations in terms of gradients (or differences) of the flow. The continuity equation may be completed through the general equation of flow ($f = k\ v$), which can be usually applied in case of stationary conditions; to this aim the time intervals may be fixed with a sufficiently small size in order to guarantee the flow stationarity condition. However this approach is not suitable because it is unrealistic.

Furthermore the $f_{IN}(t)$ may be introduced and then the cumulative in-flow may be derived and the $f_{OUT}(t)$ from which the cumulative outflow may be derived; in particular let.

$f_{IN}(t) = f(x_{1;}t)$ then the cumulative in (entering)—flow is given by

$$\mathrm{u(t)} = \int_0^{\mathrm{t}} \mathrm{f}\left(\mathrm{x}_1;\mathrm{z}\right) \mathrm{dz}$$

and let

$f_{OUT}(t) = f(x_{2;}t)$ then the cumulative out (exiting)—flow is given by

$$w(t) = \int_0^t f\left(x_2;z\right) dz$$

With respect to the speed

$$(1/(x_2 - x_1)) \int_{x1}^{x2} v\left(x;t\right) dx = v_s\ (t;x_1,x_2)$$

All these variables may be represented in aggregate or disaggregate ways (in particular position and speed are the most common aggregate or disaggregate variables) therefore non-stationary traffic flow models may be classified on the base of their representation.

In macroscopic models users' behaviour variables are aggregate (arc density or entry flows can be obtained from the vehicle position on the arc) as well as level of service variables (space mean speed, arc performance functions are derived from the fundamental diagram). They may be classified in accordance with time and space. Time-continuous and space-continuous macroscopic models are formulated by differential equations in time and space dimensions; a solution approach in discrete time and space is adopted using the finite difference method. The first class of model is also called *point based models* whereas the second class is the *finite difference class of models*. The first class of models is discussed in this section whilst the finite difference class of models is discussed in Section B.2.

In mesoscopic models users' behaviour variables are disaggregate (packets of users or single users are considered; arc density or entry flows can be obtained from packets/users position on the arc) and the level of service variables are aggregate (such as space mean speed; arc performance functions are derived from the fundamental diagram).

Microscopic models describe traffic flow dynamics in terms of a single vehicle. In particular users' behaviour variables are disaggregate (single users, i, are considered; arc density or entry flows can be obtained from the users' position, x_i, on the arc) as well as the level of service variables (time speed and arc performance functions are derived from the drivers' behaviour models such as the car-following models) (Fig. B.4).

speed variables / user variables	aggregate $v(t)$ $v(x,t)$	disaggregate $v_i(t)$
$f_{IN}(t)$ $f_{OUT}(t)$ $k(t)$ aggregate $f(x,t)$ $k(x,t)$	MACRO scopic	-
disaggregate $x_i(t)$	MESO scopic	MICRO scopic

FIG. B.4

Overview of TFM.

Queuing links

Unlike stationary models, in the case of non-stationary models the input variables are not fixed.

With reference to the following figure, the first diagram displays the arrivals and departures trajectories, whilst the second figure shows the cumulative value of arrivals and departures. The figures may be analysed with respect to four successive steps and in particular with respect to the time window between t_1 and t_3 during which a capacity variation (reduction) occurs:

- at time t_1 the beginning of the queue propagation may be observed
- at time t_2 the link capacity decreases until a minimum value
- at time t_3 the link capacity increases until the f_{IN}
- at time t_5 the queue discharging appears
- at time t_3 the queue length achieves the maximum value

and the following equations may be obtained (Fig. B.5):

$$dn(t)/dt + (f_{OUT}(t) - f_{OUT}(t)) = 0$$

$$n \geq 0$$

$$f_{OUT}(t) \leq Cap(t)$$

or equivalently

$$n(t) = (w(t) - u(t))$$

Then in terms of the general expressions of the average unitary delay, the total delay may be derived.

In particular with respect to the unitary delay it may be computed as.

$$d(t) = n(t)/Cap$$

whilst the total delay is given by

$$D(t) = \int_0^T n(t)\,dt$$

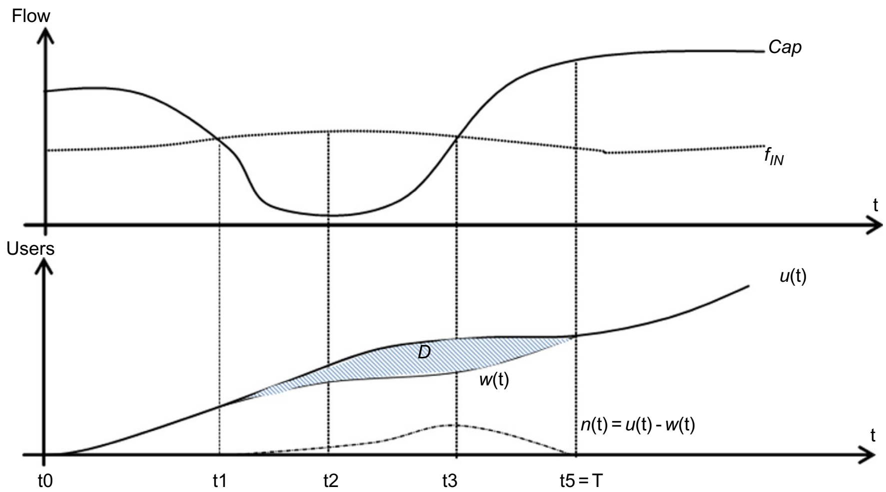


FIG. B.5

(A) arrivals and departures flows and (B) arrivals and departures cumulative values.

and finally the average unitary delay during a time interval [0, T] is equal to.

$d(0,T) = D(t)/u(t)$

B.2 Continuous time continuous space macroscopic models

Two types of kinds are identified in the class of continuous time continuous space (CC) models: the *first order models* also called *point based models* and the *second order models*. The main difference between them is in the introduction of the acceleration equation in the second order model in order to properly reproduce the traffic inhomogeneity with respect to different vehicles desired speed.

The main variables are:

f is the flow measured in vehicles per unit of time;
k is the density measured in vehicles per unit of length;
v is the speed measured in space per unit of time;
v_k is the equilibrium speed as a function of density k.

B.2.1 Point based models

An approach in which the flow is given as a function of density was introduced by Lighthill and Whitham (1955) and Richards (1956) thus this class of models are also called LWR models and differ only for the functional form of the fundamental diagram; the corresponding equations are listed below.

$f(x,t) = f\,(k(x,t))$ which represents the relationship between flow and density;
$v(x,t) = v_K(k(x,t))$ which represents the relationship between speed and density.

Thus the conservation equation may be rewritten as follows:

$$(\partial f/\partial k)\,(\partial k/\partial x) + (\partial k/\partial t) = 0 \qquad \text{(B.23)}$$

which is also called the LWR model.

The previous equation may be rewritten considering the fundamental diagram

$$(v_k + k\,\partial v_k/\partial k)\,(\partial k/\partial x) + (\partial k/\partial t) = 0 \qquad \text{(B.24)}$$

Since the functional form of the fundamental diagram is not specified these models are generally classified as LWR models. Furthermore, it must be observed that this class of model has only one dynamic equation which is represented by the continuity equation thus they are also called first order models.

Propagation of density variations

Let d be $\partial f/\partial k$, the conservation equation may be rewritten as

$$d\,(\partial k/\partial x) + (\partial k/\partial t) = 0 \qquad \text{(B.25)}$$

thus the solution of the differential equation is a function φ (differentiable) and the argument of the function is x – d t thus the solution equals

$$k = \varphi(x - d\,t) \qquad \text{(B.26)}$$

Furthermore

$k_0\ (x) = k(x,0)$ defines the initial density which according to Eq. (B.26) uniformly moves with speed d and let k'_0 be the derivative with respect to its argument, then

$$(\partial k/\partial t) = -dk'_0\,(x - d\,t)$$
$$(\partial k/\partial x) = k'_0\,(x - d\,t)$$

The conservation equation may be rewritten as

$$(\partial f/\partial k)\,k'_0\,(x - d\,t) - dk'_0\,(x - d\,t) = 0 \qquad \text{(B.27)}$$

Some considerations may be made about d which is assumed equal to $\partial f/\partial k$ (see Fig. B.6); indeed the propagation speed d depends on the density consistently with the steady state flow density relationship (fundamental diagram). In particular the density variations may propagate in driving direction (with positive derivative) in the right part of the diagram (free flow condition, stable flow) and against the driving

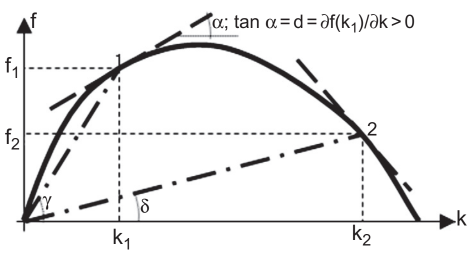


FIG. B.6

Flow-density diagram.

direction (with negative derivative) in the left part of the diagram (congested condition, unstable flow).

Shock waves

Let us consider a road segment (see Fig. B.7) in which at time t* two different flow conditions may be observed which are respectively stable in the first section (before section A-A) and unstable in the second section (after section A-A); a space- time diagram may be introduced to support the description of the phenomenon. Phenomena in the first section propagate with speed tan(α) and may be represented in the space diagram through parallel segments with angular coefficient equals to α; the same approach may be applied for section 2. Therefore space–time segments related to section 1 and space–time segments related to section 2, will meet in the wave front (continuous line) which propagates against driving direction and the relative speed equals the slope of the secant of two different states. A further example may be made if the state 2 is represented by point A in Fig. B.8, which is close to the capacity; in this case the wave front propagates in driving directions.

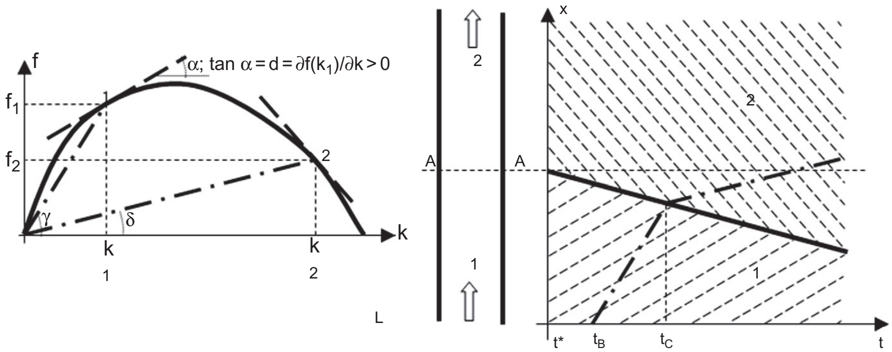


FIG. B.7

(A) Flow-density diagram. (B) Waves trajectories in a space–time diagram.

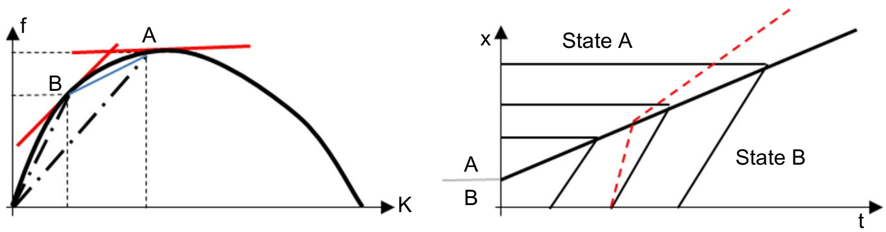


FIG. B.8

(A) Flow-density diagram. (B) Waves trajectories in a space–time diagram.

This is the case of the density discontinuity which is described though the shock wave propagation; in particular regarding the previous example, the shock wave speed propagation is computed as $z = (f_2 - f_1)/(k_2 - k_1)$.

Another example may be represented with the capacity variation (this is the case of the bottleneck; see Fig. B.9) indeed in this case it is necessary to consider a different fundamental diagram; in particular due to the effect of the lower capacity the internal fundamental diagram must be considered. If the flow is lower than the bottleneck capacity, the shock wave propagates stationary in a horizontal direction, when flow is higher than capacity the flow propagates under capacity restriction thus the shock wave propagates backwards.

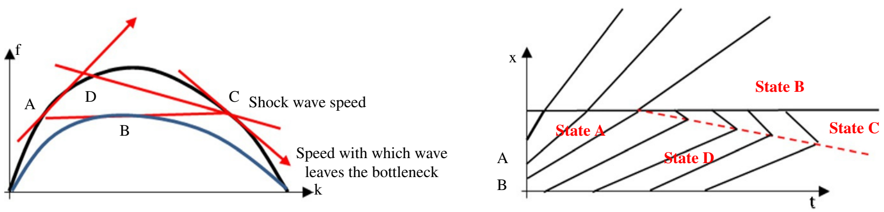


FIG. B.9

(A) Flow-density diagram. (B) Waves trajectories in a space–time diagram.

B.2.2 Second order models

Two main limitations may be identified in the LWR model: the first one refers to the model assumption of instantaneous adaptation of the vehicles speed which is certainly an idealisation; the second one is about non-homogeneous traffic indeed different desired speed for different vehicle class may be expected. Therefore the acceleration equation is introduced and these models are classified in a second order model

$$dv/dt = \partial v/\partial t + v\,\partial v/\partial x = 1/\tau\,[v_K - v] \tag{B.28}$$

Payne model

A generalised expression for the acceleration equation is considered in the Payne-Whitham model (Payne, 1979; Whitham, 1974)

$$dv/dt = \partial v/\partial t + v\,\partial v/\partial x + (1/k)\,\partial p/\partial x = 1/\tau\,[v_K - v] \tag{B.29}$$

which has a formal relationship with the Euler and Navier-Stockes equation from hydrodynamics. In particular

- the term $v\ \partial v/\partial x$ represents the *transport or convection* term and describes the speed profile of vehicles;
- the term $(1/k)\ \partial p/\partial x$ represents the *pressure term* reflecting the speed variance and then the effect of different vehicles;

- the term $1/\tau\ [v_K - v]$ represents the *relaxation term* and delineates the adaptation of the speed to the equilibrium speed based on speed – density relationship and τ is the relaxation time representing the aggressiveness of drivers.

Models differences are based on the specification of the traffic pressure, p, the relaxation time, τ, and finally the equilibrium speed, v_K. For instance supposing that

$$p(k) = (1/2\tau)\,[v_0 - v_K]$$

In which v_0 is the free flow speed, if we combine this equation with the acceleration equation Payne's model is derived:

$$\partial v/\partial t + v\,\partial v/\partial x - 1/\tau\,[v_K - v] + (1/2k)\,(\partial v_K\,k)/\partial k)(\partial k/\partial x) = 0$$

which may be generalised for $\tau \to 0$ in Payne's simplified model:

$$v = v_K - (\tau/k)\,(\partial p/\partial x) = v_K + (1/2\mathrm{k})\,(\partial v_K\,k)(\partial k/\partial x)$$

In accordance with Payne's simplified model the conservation equation may be rewritten as:

$$\partial k/\partial t + \partial(k\,v_K + (1/2)\,(\partial v_K/\partial k)\,(\partial k/\partial x))/\partial x = 0 \qquad \text{(B.30)}$$

that is the LWR model including the *diffusion* term:

$$\partial k/\partial t + \partial(k\,v_K)/\partial x = \partial(\varphi(k)\,\partial k/\partial x)/\partial x \qquad \text{(B.31)}$$

Where the diffusion term depends on density and it is equal to

$$\varphi\,(k) = (-1/2k)\,(\partial v_K/\partial k) \geq 0$$

Ross model

Regarding the Ross model this originates from the consideration that the vehicle speed is influenced (limited) by congestion even if each driver would like to travel at free flow speed, v_0 which is independent of density and density is limited by density k_{jam} thus let T be the time interval necessary for the flow equilibrium the acceleration equation may be rewritten as follows:

$$\partial v/\partial t + v\partial v/\partial x = (1/T)[v_0 - v], k \leq kj_{am}$$

Kerner and Konhäuser model [KK model]

Regarding the pressure term in the KK model it is equal to

$$P = k\theta_0 - \gamma_0 \partial v/\partial x$$

where $\theta_0 > 0$ is a constant term and $\gamma_0 > 0$ is the viscosity coefficient thus the acceleration equation will be rewritten as below

$$\partial v/\partial t + v\partial v/\partial x = -(\theta_0/k)(\partial k/\partial x) + (\gamma_0/k)\partial^2 v/\partial^2 x + (1/T)[(\partial\,v_K - v]$$

In general the space and time continuous second order model may be formulated through the following equations[c]

$$(\partial f/\partial k)\,(\partial k/\partial x) + (\partial k/\partial t) = 0 \quad \text{(B.32)}$$

$$f(x,t) = k(x,t)\,v(x,t) \quad \text{(B.33)}$$

and a third additional equation representing the acceleration.

B.2.3 Network equations

In order to properly apply these models at network level it is necessary to define the outflow rates from incoming links at each node within the network. To this aim the network equations are required which are able to define the inflows into and outflows out of each node taking into account the inflow capacities into outgoing links.

However in practical applications the network equations for CC models are obtained through a discretized approach then a proper and more detailed discussion is proposed in Section B.4.

B.3 Continuous time discrete space macroscopic models

In terms of macroscopic models other approaches to flow propagation are based on *time-continuous* and *space-discrete* representation (CD), also referred to as *link-based* models. This class of models is specified through differential equations with respect to the time, for each arc. Main notations used in the following are enlisted below in alphabetical order (notations come first) for reader's convenience(notations come first, then Roman letters, after Greek letters).

$n_L(t)$ is the number of vehicles on a arc l, at time t;
$f_{INL}(t)$ is the inflow on a arc l, at time t;
$f_{OUTL}(t)$ is the outflow on a arc l, at time t;
$tt_L(t)$ is the travel time on a arc l, at time t;
v_F be the free flow speed.

These models can be subdivided into *whole-link* and *wave models*. Each one of them is discussed in more detail in the following sections.

[c] The expression may be generalised to the case of on—ramps, with in—flows equals to E*(y,t) thus the conservation equation may be rewritten as $(\partial q/\partial y) + (\partial k/\partial t) = E^*(y,t)$.

B.3.1 Whole link models

Running links

The time-continuous and arc based models may be classified into two main approaches: the *exit function* formulation and the *travel time* formulation. The first class of model is based on the exit function relationship which governs the outflow from a arc; whilst the second class of model is based on the arc travel time. The second class of model has been introduced in order to overcome the theoretical limitations of exit function approaches. Indeed these models do not satisfy the following requirements: (1) the First in First out (FIFO[d]) rule able to ensure that no overtaking can occur between two users who have entered an arc at different times and (2) the consistency between flow propagation and speed which has to be lower than free flow speed.

Starting from the equation.

$$n_L(t+1) = n_L(t) + (f_{INL}(t) - f_{OUTL}(t))\,\Delta t \tag{B.34}$$

then the conservation equation may be rewritten as:

$$dn_L/dt = f_{INL}(t) - f_{OUTL}(t) \tag{B.35}$$

In the case of the exit function model, the exit arc function governing the outflow from the arc (the inflow which is supposed to be known and independent from preceding arcs) is introduced. Depending on the number of users n(t), the exit function arc may be formalised as in the following:

$$f_{OUT}\,(t) = w(n(t)) \tag{B.36}$$

then the arc is modelled through the system of differential equations made up of Eqs. (B.35), (B.36).

An example of formulation for the exit function might be the following:

$$f_{OUT}\,(t) = Cap\left(1 - e^{-n(t)/tt}\right)$$

where Cap is the arc capacity.

In the case of travel time formulation the arc travel time function tt of the user who arrives at the time t at the beginning of the arc is introduced as $tt(t)$. A travel time value $tt(t)$ depends on the number of users on the arc thus it may be expressed through a travel time function $T(n)$ as

$$tt(t) = T(n(t)) \tag{B.37}$$

Regarding the arc model it is formalised through the system of differential equations made up of Eqs (B.34), (B.36).

A further discussion for the FIFO necessary and sufficient condition is required; in particular the rule behaviour is:

$$dtt(t)/dt > -1$$

[d] Under the assumption of one-dimensional fluid flow, the first vehicle entering the arc will be the first vehicle exiting the arc.

which guarantees that the exit flows are positive in the case of positive entry flows, then the condition is equivalent to.

$$dtt(t)/dt = -1 + f_{IN}(t)/f_{OUT}\,(t + tt(t))$$

which implies that $f_{IN}(t)$ has to be >0 if some users are entering the arc and $f_{OUT}(t) \geq 0$ is not moving in the wrong direction and one of two of the above conditions ensures that no FIFO violation occurs. The condition may be also rewritten as:

$$f_{OUT}(t + tt(t)) = f_{IN}\,(t)/(1 + dtt(t)/dt) \tag{B.38}$$

furthermore it is still guaranteed that the denominator in the previous expression is positive.

Finally the equation states that, if the flow on arc at time t is decelerating ($dt/dtt > 0$), the flow exiting the arc after the travel time required to cross it will be less than the flow entering at t. Vice versa, if the flow on the arc is accelerating ($dt/dtt < 0$), the flow exiting the arc will exceed that entering. However, when the arc travel time at time t is constant ($dt/dtt = 0$) (in non-congested networks with constant supply or in steady state conditions in the system), the flow exiting from an arc is simply translated in time compared with the entry flow.

Therefore the arc model it is formalised through the system of differential equations composed by Eqs (B.35), (B.36), (B.38).

In general different expressions are proposed for travel time representation (see Eq. B.37) however the FIFO condition is not necessarily guaranteed. For instance, an example of travel time expression for running links constrained to the FIFO condition is given by the following linear function

$$T(n(t)) = tt^0 + n(t)/Cap$$

where tt^0 represents the free flow travel time and implies that the exiting flow is constrained to the capacity and then guarantees that the FIFO condition is never violated.

A similar equation is proposed for the queuing links but the estimated time is the queuing time (the time spent in the waiting link; T_w)

$$T_w(n(t)) = 1/Cap + n(t)/Cap$$

Network equations

In order to define the network loading model a further flow conservation for each node within the network is required. However the solution requires the time discretization of Eq. (B.34) and the link discretization in segments. The final result is the time-discrete and space-discrete models discussed in more details in following Section B.4.

B.3.2 Wave models

In order to discuss the wave models and in particular the Link Transmission Model (LTM), first of all the Newell simplified theory of kinematic waves (NSTKW) must be introduced; indeed the LTM uses the NSTKW to propagate traffic flows on arcs.

Newell model

Main notations used in the following are enlisted in Section B.1.1 in alphabetical order (notations come first) for reader's convenience(notations come first, then Roman letters, after Greek letters) in addition.

n(x, t) is the number of the last vehicle to pass section x before time t;

A curve of cumulative flow versus time is also known as a Newell-curve or simply N-curve (Daganzo, 1994), since Newell 1993 developed a simplified version of the LWR kinematic wave theory based on this concept. This theory is founded on the conservation of vehicles.

The cumulative function N(x, t) is the number of the last vehicle to pass section x before time t; the space–time diagram shows the vehicles position as a function of time (trajectories). The vehicles have been numbered in increasing order in the direction of increasing time (Fig. B.10).

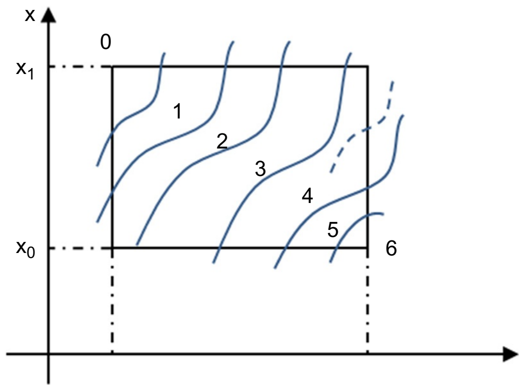


FIG. B.10

vehicles trajectories and values of cumulative vehicle function N(x,t).

In particular if one draws the curves of the cumulative number of vehicles to pass some locations x_1 and x_2 by time t, the vertical distance at time t between two curves represents the number of vehicles between x_1 and x_2, the horizontal distance between two curves at height j, represents the trip time of the j_{th} vehicle from x_1 to x_2; the area between curves represents the total travel time of all vehicles. Therefore the evolution of the traffic is represented by the cumulative number of vehicles that pass the up and downstream ends x_1 and x_2 of each arc by time t.

Let

- $n(x, t_1) - n(x, t_0)$ be the number of vehicles observed at location x during time interval $[t_0, t_1]$;
- $n(x_0, t) - n(x_1, t)$ be the number of vehicles in section $[x_0, x_1]$ at time t.

Starting from the relations already introduced in Section B.3 the following equations may be obtained:

the first one is between observed vehicles during time interval and flow

$$\partial n(x,t,t+\Delta t))/\partial t = f(x,t)$$

the second one is between observed vehicles on section and density

$$\partial n(t,x,x+\Delta x)/\partial x = -k(x,t)\,dx$$

where the negative sign arises because n decreases in the direction of increasing x. The function n(x, t) ensures vehicles are not created nor lost along the road segment; thus n(x,t) exists in a certain region (Δx, Δt) as well as first and second derivatives; hence

$$\partial^2 n(x,t,t+\Delta T))/\partial t\partial x = \partial^2 n(t,x,x+\Delta x)/\partial x\partial t$$

and the conservation equation may be rewritten as

$$\partial f(x,t)/\partial x + \partial k(x,t)/\partial t = 0 \qquad \text{(B.39)}$$

Finally the Green's theorem yields:

$$n(x_2,t_2) - n(x_1,t_1) = \int_t^{t+\Delta T} \partial m(x,t,t+\Delta T))/\partial t\,dt + \int_x^{x+\Delta x} \partial n(t,x,x+\Delta x)/\partial x\,dx$$

$$= \int_t^{t+\Delta T} q(x,t)\,dt - \int_x^{x+\Delta x} k(x,t)\,dx \qquad \text{(B.40)}$$

Instead of using the NSTKW to evaluate the flows and densities, and then determine the cumulative number of vehicles, Newell uses the NSTKW to directly evaluate the cumulative vehicle number n(x, t).

Newell uses the triangular fundamental diagram and four main parameters are identified: the free flow speed, the maximum capacity, the critical density and the traffic jam density; for density less than critical value, vehicles propagate with a free flow speed otherwise for density higher than the critical value, vehicles propagate in congested regime. Finally two speed values are identified: v that is a positive speed and is referred to the free flow conditions and z that is negative and is referred to the congested flow conditions.

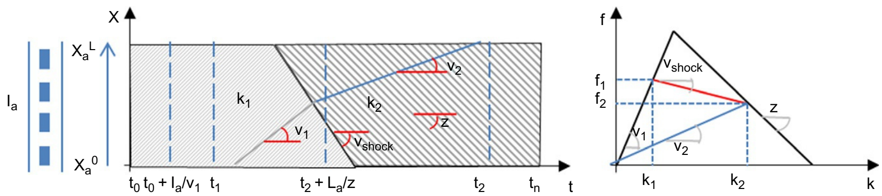


FIG. B.11

(A) homogeneous road segment with capacity restriction and (B) triangular fundamental diagram.

Assume an arc $[x_a^0; x_a^1]$ representing a homogeneous road segment of length l_a; vehicles travel belong arc from x_a^0 to x_a^1; suppose that at time t_1 a capacity restriction

(f_2) at downstream (x_a^l) occurs, traffic states may be described by considering a triangular fundamental diagram. The initial condition is $k(x, t_0) = 0$ and boundary conditions are

$$k\left(x^0{}_a, [t_0, t_n]\right) = k_1$$

$$k\left(x^l{}_a, [t_1, t_n]\right) = k_2$$

As displayed in Fig. B.11A the two traffic states (which are identified in accordance with figure B), respectively, free flow in case of traffic demand f_1, and congested in case of traffic demand f_2, and respectively identified with following variables, k_1, f_1 and v_1 and k_2, f_2 and v_2, intersect each other in a shock; the shock wave travels with a negative speed and back propagates against traffic direction. At time t_2 the congested states reach arc boundary x_0^a and from that time the road segment is congested. Below the cumulative vehicle curve at upstream and downstream are displayed (Fig. B.12).

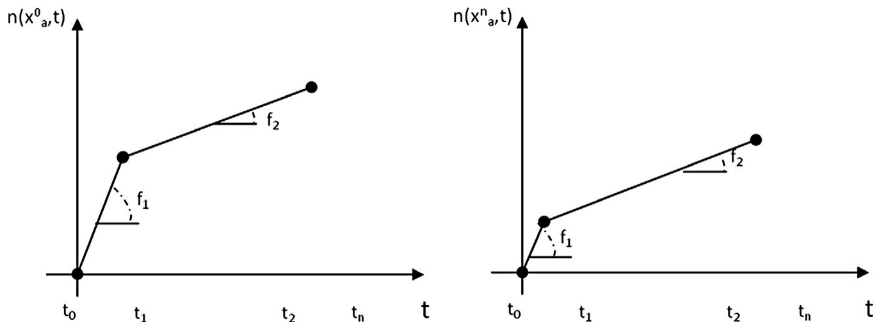


FIG. B.12

(A) Cumulative vehicle numbers at the upstream boundary x_a^O and (B) at the downstream boundary x_a^L.

Consider now the following example (Fig. B.13):

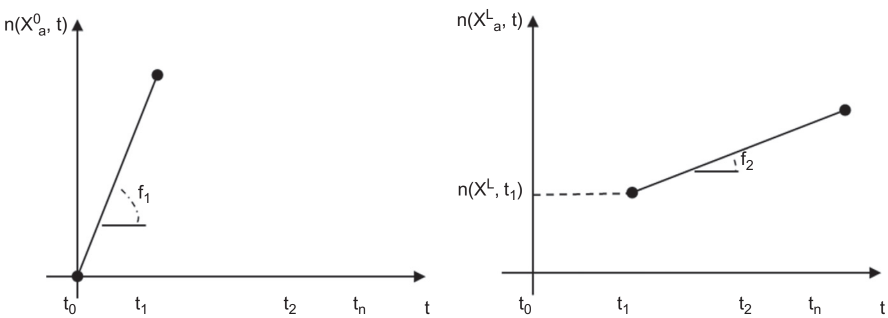


FIG. B.13

Upstream x_a^O (A) x_a^L boundary conditions on a arc a (B).

Free flow traffic state q_1 travels form upstream; if this state does not intersect with another state, then it will reach the downstream arc in l_a/v_1 time, and the initial traffic condition at t_0 will be identical to the traffic condition at time t_0+l_a/v_1 hence the Green's theorem may be rewritten as

$$n(x_0{}^a, t_0) - n\left(x^L{}_a, t_0 + l_a/v_1\right) = f_1(-l_a/v_1) - k_1(-l_a) = l_a\,(-f_1/v_1 + k_1) = 0$$

The congested traffic state q_2 travels from downstream to upstream; if it does not intersect any other state then (Fig. B.14)

$$n(x_0{}^a, t_2) - n\left(x^L{}_a, t_2 + l_a/w\right) = f_2(-l_a/w) - k_2(-l_a) = l_a\,(-f_2/w + k_2) = k_{jam} l_a$$

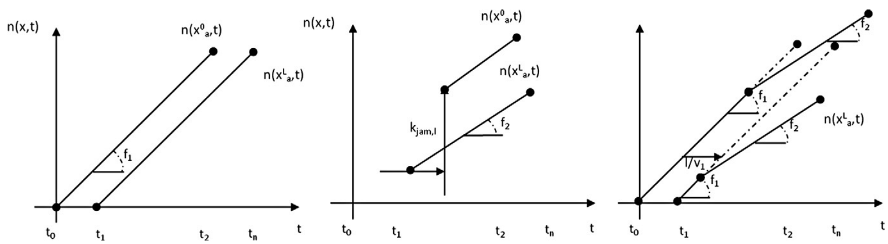


FIG. B.14

Translations of upstream(A) and downstream (B) boundary conditions and the lower envelope of the double-valued solution (C).

Based on the NSTKW a very accurate procedure to determine sending and receiving may be defined.

Link transmission model (LTM)

Running links

Regarding the LTM it uses Newell's simplified theory to propagate traffic flows on arcs. In LTM traffic states are updated in successive steps, hence the algorithm provides a discrete time solution of the KW model. At each time interval, Δt, the algorithm determines the sending flow at the downstream arc, and the receiving flow at the upstream arc.[e]

The *sending flow* is defined as the maximum number of vehicles that could leave the downstream end of this arc during time interval $[t, t+\Delta t]$.

With reference to the sending flow the NSTKW states that in case of free flow traffic condition (see Fig. B.15), the states have been emitted from upstream t_F free flow time units earlier (let v_F be the free flow speed, $t_F = l/v_F$).

[e] The general application of the LTM algorithm is based on three successive steps which may be summarised in: (i) estimation of sending and receiving flows, (ii) generation of transition flows at nodes, and (iii) update of cumulative vehicle number (based on transition flows estimation).

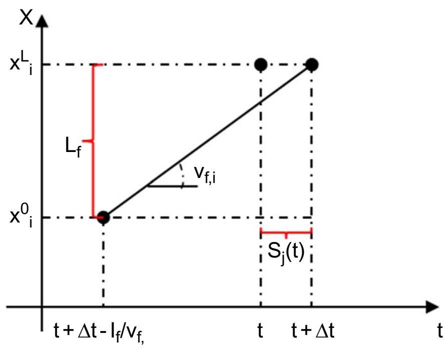


FIG. B.15

Free flow traffic state propagation.

Regarding the conservation equation it may be rewritten as in the following:

$$n\left(x_i{}^0, t+\Delta t-l_f/v_F\right)-n\left(x^L{}_i, t+\Delta t\right)=f_1\left(-l_i/v_F\right)-k_1\left(-l_i\right)=l_i\left(-f/v_F+k\right)=0 \tag{B.41}$$

thus

$$n\left(x_i{}^0, t+\Delta t-l_F/v_F\right)=n\left(x^L{}_i, t+\Delta t\right)$$

Furthermore in accordance with the sending flow definition, may be formulated as:

$$S_i(t)\leq n\left(x_i{}^l, t+\Delta t\right)-n\left(x^l{}_i, t\right)=n\left(x_i{}^0, t+\Delta t-l_F/v_F\right)-n\left(x^l{}_i, t\right)$$

and the sending flow is also constrained to the arc's capacity thus:

$$S_i(t)=\min\left(n\left(x_i{}^0, t+\Delta t-l_F/v_F\right); f\,\Delta t\right)$$

The *receiving flow* is defined as the maximum number of vehicles that could enter the upstream end of this arc during time interval [t, t+Δt]. The NSTKW states that in case of congested traffic flow condition (see Fig. B.16), the states have been emitted from upstream t_F free flow time units earlier (let v_F be the free flow speed, $t_F=-\,l/w_F$).

$$\begin{aligned} &n\left(x_i{}^0, t+\Delta t\right)-n\left(x^l{}_i, t+\Delta t+l_F/w_f\right)=f\left(-l_F/w_F\right)-k\left(-l_F\right)\\ &=l_F\left(-f/w_F+k\right)=k_{jam}l_F \\ &n\left(x_i{}^0, t+\Delta t\right)=n\left(x^l{}_i, t+\Delta t+l_F/w_F\right)+k_{jam}l_F \end{aligned} \tag{B.42}$$

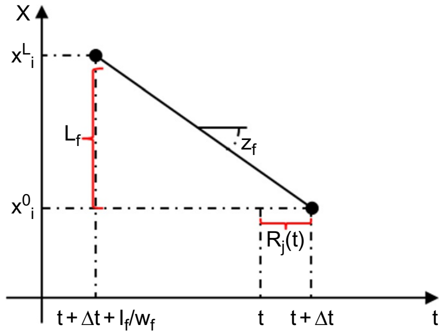


FIG. B.16

Congested traffic state propagation.

As in case of the sending flow, the receiving flow is still constrained by the arc's capacity, therefore the receiving flow will be given by:

$$R_j(t) = \min\left(n\left(x^l{}_i, t+\Delta t + l_F/w_F\right) + k_{jam}\, l_F - n\left(x_j{}^0, t\right); f\,\Delta t\right)$$

Network equations

In order to define the network loading model a further flow conservation for each node within network is required and in particular a node model needs to be introduced. This model is able to describe the rules governing the vehicles transferring from upstream to downstream. Three main examples are discussed: the ordinary node, the merge and diverge nodes (see Fig. B.17).

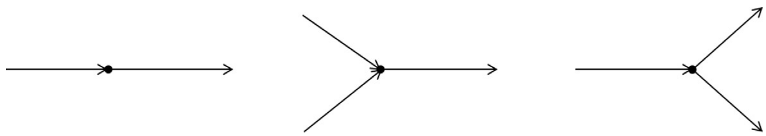


FIG. B.17

(A) ordinary node, (B) merge node, and (C) diverge node.

An ordinary node connects an incoming link, to an outgoing link, c. The flow through an ordinary node is the maximum that can be sent by the upstream link to the downstream link then the following expression may be applied:

$$n\left(x^l{}_a, t+\Delta t\right) = n\left(x^l{}_a, t\right) + \min\left\{Sa(t), R_c(t)\right\}$$

A merge node connects two incoming links, a and b, to only one outgoing link, c. The flow through a merge node is the sum of the Sending flows of the incoming links constrained to the Receiving flow of the outgoing link j. Indeed if the sending flow exceeds the Receiving flow, it will be assumed that the maximum amount that can be received by link j will be transferred according to p_{ij} which represents the fraction of the total amount of vehicles, coming from link i and $\Sigma_i\ p_{ij} = 1$.

Each merge model may be applied in order to determine the transition flows y_{ij} for the upstream links. Then the cumulative curves at the link boundaries are updated as follows:

$$n\left(x^L{}_a, t+\Delta t\right) = n\left(x^L{}_a, t\right) + y_{ac}(t)$$
$$n\left(x^L{}_b, t+\Delta t\right) = n\left(x^L{}_b, t\right) + y_{bc}(t)$$
$$n\left(x^0{}_c, t+\Delta t\right) = n\left(x^0{}_c, t\right) + y_{ac}(t) + y_{bc}(t)$$

where the transition flows may be modelled in accordance with Daganzo (1995) as in the following:

$$y_{ij} = median\ \left(S_{ij}, R_j, -\left(\left(\Sigma_i\, S_{ij-}\, S_{ij}\right), p_{ij}\, R_j\right)\right.$$

A diverge node connects one incoming link, a, to two outgoing links, c and d. Diverge models determine the transition flows that are used to update the cumulative curves as follows:

$$n\left(x_a{}^L, t+\Delta t\right) = n\left(x_a{}^L, t\right) + y_{ac}(t) + y_{ad}(t)$$
$$n\left(x_c{}^0, t+\Delta t\right) = n\left(x_c{}^0, t\right) + y_{ac}(t)$$
$$n\left(x_d{}^0, t+\Delta t\right) = n\left(x_d{}^0, t\right) + y_{ad}(t)$$

The flow through a diverge node is the maximum that can be sent by the incoming link, unless one of the outgoing links is unable to receive its allocated part of the Sending flow. We study two approaches for diverging modelling: the FIFO based on split factor (see Daganzo, 1995) and the parallel method with split factors (see Lebacque, 1996). Both methods are based on exogenously defined split factors, p_c and p_d representing the percentage of vehicles going on link c and link d (the sum equals 1). Two methods are described below:

[FIFO model]

$$G_{ij} = p_{ij}\ \min\left[S_i,\ \min_i\left(R_{ij}/p_{ij}\right)\right]$$

[Parallel method with split factors]

$$G_{ij} = \min\left[p_{ij}\, S_{i,}\ R_{ij}\right]$$

B.4 Discrete time discrete space macroscopic models

In this section the finite difference models as an example of the macroscopic discrete time and discrete space models are discussed. In particular as already stated in Section B.3, CC macroscopic models are formulated by differential equations in time and space dimensions and a solution approach in *discrete time* and *space* (this class of models is also called discrete time and discrete space, DD) is adopted using the finite difference method. In this models it is assumed that the road segment is divided into cells.

Main notations used in the following are enlisted in Section B.3 in alphabetical order (notations come first) for reader's convenience(notations come first, then Roman letters, after Greek letters). In addition.

k is a uniform density in a cell during time interval Δt;
n_i be the number of vehicles on cell i, equals to $k\ \Delta x$;
N_i is the maximum number of vehicles present in cell i.
Cap is the maximum flow rate in cell i;
v is the free flow speed coefficient;
z is the wave speed coefficient;
$y_i(t)$ is the inflow (to cell i) at time t;
$y_{i+1}(t)$ is the outflow (from cell i) at time t;
$\Delta = z/v$ with $z \leq v$.

B.4.1 Finite difference models

The LWR models need to be solved numerically by *finite-difference* methods. In particular the road segment is divided into cells of constant length and time in the index k increasing in the downstream direction. Each cell is characterised by the same density and speed (as a function of the speed–density relationship) as well as the flow between neighbouring cells which is constant during the time interval.

The most common integration method for LWR models solution is the *Godunov scheme*. This method is based on an exact solution of the continuity equation for a one-time step assuming stepwise initial conditions given by the actual densities of the cells.

In particular at the first step the road is divided into cells each one of width Δx. The cell length Δx is the distance a vehicle would travel in a free flow condition, in a one-time step hence it is equal to free flow speed times the length of the time step (also called clock tick), $\Delta x = v\Delta t$. In must be remarked that the relationship between the cell length and the time step corresponds to the *Courant-Friedrich-Lewy* condition and for stability of explicit solution methods that is $v\Delta t \leq \Delta x$.

Regarding the Godunov scheme, the densities are initially averaged for each cell (each cell has a constant value of density), and from one step, t, to the successive one, $t + \Delta t$, the solution evolution is averaged again in order to obtain a the piecewise constant solution.

The density is obtained as a function of flows at the cell boundaries (Fig. B.18).

$$k_j(t+\Delta t) = k_j(t) + (\Delta t/\Delta x)\left(f_{j-1/2} - f_{j+1/2}\right)$$

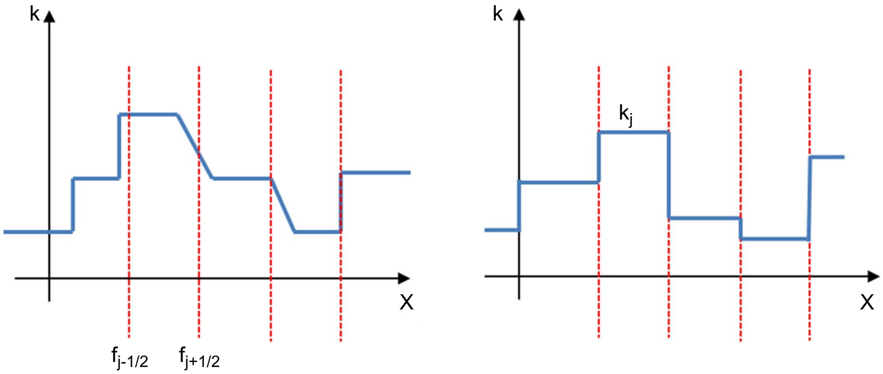


FIG. B.18

Solution evolved exactly (A) solution at time $t+\Delta t$.

The cell transmission model

Let $N_i(t)$ be the vehicle holding capacity of cell i and k_{jam} is the jam density (the maximum number of vehicles which can fit into cell i) thus (Fig. B.19)

$$N_i(t) = k_{jam}\Delta x$$

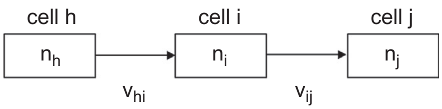


FIG. B.19

Junction represented trough cell transmission model.

The conservation equation may be rewritten as

$$n_{i+1}(t+1) = n_i(t) + y_i(t) - y_{i+1}(t)$$

The key quantities of the method can be introduced based on the (trapezoidal) fundamental diagram Fig. B.20.

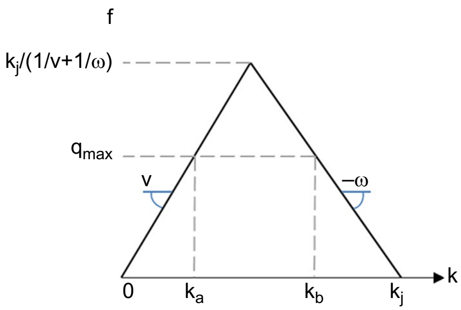


FIG. B.20

Trapezoidal fundamental diagram.

The number of vehicles moving from cell i to cell j is given by.

$$y_i(t) = \min\{n_i - 1(t), Cap, \delta[Ni(t) - n_i]\} = \min\left\{kv, Cap_, z\left(k_{jam} - k\right)\right\} \quad \text{(B.43)}$$

and this is the result of a comparison between the maximum number of vehicles that can be *sent* by the cell i directly upstream of the boundary $S_i(t) = min\{kv, Cap\}$ and those that can be *received* by the downstream cell i, $R_{i+1}(t) = min\ \{Cap, z(k_{jam}\text{-}k)\}$.

Summing up, the flow equation inherently accommodates different traffic conditions from low level flow to oversaturation. In low level traffic (*uncongested*), the flow is equal to the number of vehicles in cell i at time t, n_i; in bottleneck traffic (*flow capacity*), the flow is equal to the saturation flow rate f_i times Δt, and in oversaturated traffic (*congested case*), the flow is restricted by the jam density and depends on the available space in cell i at time t, $z_i/v[N_i - n_i]\}$. This allows us to simulate the propagation of blocking back phenomena by considering constraints on the cell outflow equation (receiving function).

Hence in accordance with the *Godunov scheme*, the flow $y_i(t)$ can be rewritten in accordance with the supply (*sending*)-demand (*receiving*) rule of the cell transmission model (Fig. B.21):

$$y_i(t) = \min\{S_i, R_{i+1}\}$$

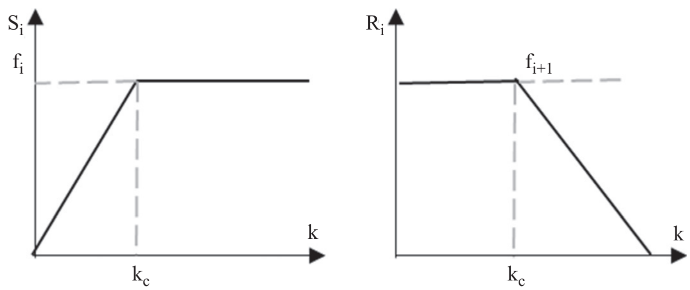


FIG. B.21

The sending (A) and receiving rule of cell transmission model.

Depending on the density of each cell, the cells flow propagates into the upstream cell (propagation speed <0) or downstream (propagation speed >0). Summing up all steps of the Godunov scheme for the triangular fundamental diagram leads to the supply–demand method constrained as in following: if supply is the limiting factor, information travels upstream; if demand limits the traffic flow, information travels downstream.

B.4.2 Network equations

In order to define the network loading model a further flow conservation for each node within the network is required and in particular a node model need to be introduced. This model is able to describe the rules governing the vehicles transferring from upstream to downstream. In accordance with the LTM model three main examples are discussed: the ordinary node, the merge and diverge nodes. In the specific case of CTM the node is represented by a cell then the classification is not referred to a node but to a cell (Fig. B.22).

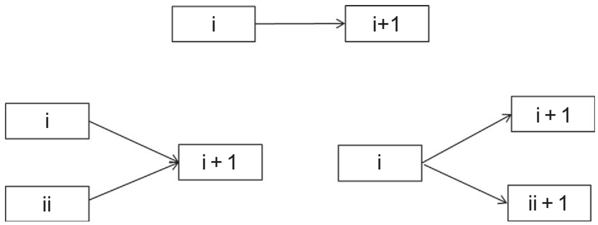


FIG. B.22

Ordinary, merging and diverging cell transmission model.

In an ordinary cell if one entering and one leaving are considered, the number of vehicles moving from cell i to cell j is given by

$$y_i(t) = \min\{n_i, \min[Cap_i, Cap_{i+1}], \delta[N_{i+1-} n_{i+1}]\}$$

however if we consider the maximum flows that can be sent and received by cell i in the interval between t and t+1 a further simplification may be introduced:

$$S_i(t) = \min\{n_i, Cap_i\}$$

$$R_i(t) = \min\{w(k_{jam} - k), Cap_i\} = \min\{\delta[N_i - n_i], Cap_i\}$$

Then the number of vehicles moving from i to j is given by

$$y_i(t) = \min\{S_i(t), R_{i+1}(t)\}$$

The cell occupancies may be updated in accordance with following expressions[f]:

$$n'_i(t+1) = n_i(t+1) - y_i(t)$$

$$n_{i+1}(t+1) = n'_{i+1}(t+1) + y_i(t)$$

In the case of merge there are two links that enter and one leaves. In general flows must satisfy the following conditions.

$$y_i(t) \le S_i; y_{ii}(t) \le S_{ii}$$

$$y_i(t) + y_{ii}(t) \le R_{i+1}$$

as in the case of ordinary links, however it will be assumed that cells i and ii send the maximum traffic possible if cell i+1 can receive it then if the condition is

$$R_{i+1} > S_i + S_{ii}$$

$$y_i(t) = S_i; y_{ii}(t) = S_{ii}$$

If this condition is not satisfied it will assume that the maximum number of possible vehicles R_{i+1} advances in i+1 and the fractions of the vehicles come from i and ii, respectively given by p_i and p_{ii} (and the sum equals 1) then if $R_{i+1} < (S_i + S_{ii})$.

[f] The occupancies $n'_k(t+1)$ are intermediate variables introduced only for mathematical notation purposes; they can be eliminated during computer implementation.

$y_i = \text{median}\ ((S_i, R_{i+1} - S_{ii}), p_i R_{i+1})$
$y_{ii} = median\ ((S_{ii}, R_{i+1} - S_i), p_{ii} R_{i+1})$

These flows are finally combined with the cell occupancies equations.

The last case is related to the diverge in which one link enters and two leave it. It must be specified that vehicles diverging is regulated on the base of the FIFO rule. Supposing that the proportions of $S_{i,}$ α_{i+1} and $\alpha_{ii+1,}$ are exogenously determined, then the number of vehicles exiting from i, y_i^j generates.

$y_i = \alpha_{i+1}\, y_i$

$y_{ii+1=}\ \alpha_{ii+1}\, y_i$

furthermore in order to constrain the number of vehicles that can be received by i + 1 and ii + 1, with respect to R_{i+1} and R_{ii}, the following conditions are introduced:

$\max\{y_i \leq S_{i,}\ \alpha_{i+1}\, y_i \leq R_{i+1}, \alpha_{ii+1}\, y_i \leq R_{ii+1}\}$

and the solutionof the linear problem is given by

$y_{i+1} = \min\{S_{i,}\ R_{i+1}/\alpha_{i+1}, R_{ii+1}/\alpha_{ii+1}\}$

as in case of ordinary links and merging, these flows are finally combined with the cell occupancies equations.

B.5 Mesoscopic models

These models may be classified in terms of flow representation. Two different approaches may be identified in the literature: the packet, say the group of users/ the single user representation and the single vehicle representation. A packet of vehicles acts as one entity and its speed on each road (arc) is derived from a speed-density function defined for that arc. Each packet is dealt with as a single entity which experiences the same traffic conditions. Several authors have proposed methods based on packets of vehicles to reduce the computed effort with respect to available computer resources. This feature is significant in order to classify papers proposed in the past since computing resources which are currently available make it possible to consider each packet made up of one vehicle only and, therefore, this distinction is no longer available.

A further classification of packet based models can be made in terms of a discrete packet and a continuous packet. In the first case, a discrete distribution of vehicles in the packet is considered. All users are grouped in a single point (for instance the head of the packet) and, therefore, are located contemporarily at the same position over the arc. In the continuous packet based approach, the vehicles are considered uniformly distributed (in time or space) in the packet, which is thus identified by two main points i.e. the head and the tail of the packet. Due to their inherent difficulties related to numerical problems of internal consistency when instantaneous density variations between adjacent simulation steps occur, only a few authors in literature have investigated continuous packet models and the most relevant contributions rely on discrete packet methods. It is worth noting that continuous packets are relevant only when packets are made up of more than one vehicle.

In general, these models do not allow detailed simulation of the behaviour of individual vehicles (overtaking, lane-changing, etc.)

Mesoscopic traffic flow models may be further classified in terms of queuing representation in arc based models, which are in turn grouped further depending on the performance function, and node based models, usually referring to the models which consider the flow splitting rates. In particular, arc based models are also divided in (i) travel time models and (ii) exit function models.

Travel time models may also be classified depending on the arc representation: by considering the whole arc so that the travel time is the sum of running travel time and queuing time or by considering the arc divided in a running part (free flow moving) and a queuing part.

B.5.1 Traffic analysis and flow forecasting mesoscopic dynamic [TRAFFMED]

In this section an overview of the TRAFFMED (traffic analysis and flow forecasting mesoscopic dynamic) is provided in particular with reference to the packet generation and the traffic dynamics.

This traffic model is based on discrete packet representation and each packet is made up of a single vehicle only. It is an arc based model, falling within the class of the travel time models, and makes it possible to explicitly represent: horizontal queues; proportional traffic leaving a node through an explicit path choice model; the dynamic generation of the path-flows incidence matrix.

Furthermore, a speed-density relationship is adopted to evaluate the speed of a packet on the running part only and it is not used as a cost function. In this way, double counting the delay experienced by the vehicles is avoided, firstly due to the traversal speed (which is computed as a function of the actual density) and then due to the queuing delay (which is obtained as a result of the simulation). It is considered a dynamic queue-running part representation by including arc capacity restrictions which are due to the signal timings variations over the time steps of simulation.

Main notations used in the following are enlisted below in alphabetical order (notations come first) for reader's convenience(notations come first, then Roman letters, after Greek letters).

a is the arc;

k_c^a is the critical density on the arc *a*, i.e. the density corresponding to the capacity in the fundamental diagram;

k_a^γ is the density of the running part of the arc *a* at beginning of sub-interval γ;
l_a is the length of arc *a*,
$l_a - x_a^S$ is the part of the arc occupied by the queue;
$n_e(P)$ is the number of elements of packet P, that is, $n_e(P) = 1$;
Cap_a is the capacity at the final section of the arc *a*;
t are the discrete simulation time intervals;
$u_a(t_1)$ is the number of packets which left the arc until time t_1;
v_a^γ is the speed on the running part of the arc *a*, updated at the beginning of each sub-interval γ, as a function of density k_a^γ;
x_a^S is the abscissa of a section S in arc *a* that divides the arc into two parts named respectively running part $[0, x_a^S)$ and queuing part $[x_a^S, l_a]$;
γ are the sub-intervals of the discrete simulation time intervals;
δ is the length of the generic sub-interval γ and let $t_1 \in [0, \delta]$;
δ/T is the ratio between the length of interval γ and the time unit T considered for the capacity.

The generic packet P, is characterised by a departure time η within the *departure interval*, h (which is the simulation interval during which the departure occurs) and an origin/destination pair rs; path choice is evaluated at the beginning of each departure interval, say at the origin and may be updated at the beginning of each simulation interval t, if en route rerouting occurs.

A packet P moves in the network and it is subject to queuing phenomena. The network is represented by a graph G(N,A), with N the set of nodes and A the set of arcs.

The position of section S is obtained by the dynamic network loading (a detailed description is provided in Fig. B.23) and is updated at each time interval t considering the number of packets in the queuing part of the arc at the previous interval t − 1; the speed on the running part of the arc at interval t depends on the outflow conditions of the previous interval t − 1.This approach is slightly inconsistent since the level of service in the interval is considerably affected by flows and queues in the previous interval and not in the current. This assumption makes it possible to greatly simplify the computation while the inconsistence can be limited by reducing the length for the sub-interval, t.

Outflow conditions on each arc are considered homogeneous and constant for the entire duration of a sub-interval, δ. They are estimated at the beginning of each sub-interval and maintained for its entire duration, thus avoiding the occurrence of the internal fixed point problem.

Summing up, the movement of a generic packet depends on its position on the arc (running or queuing) and it can change its outflow conditions by moving forward from the running part to the queuing part of the arc.

At each simulation time, the length of the queue on the arc (and, consequently, the abscissa of section S) is updated and it is possible to take into account the eventual occurrence of queue spillback.

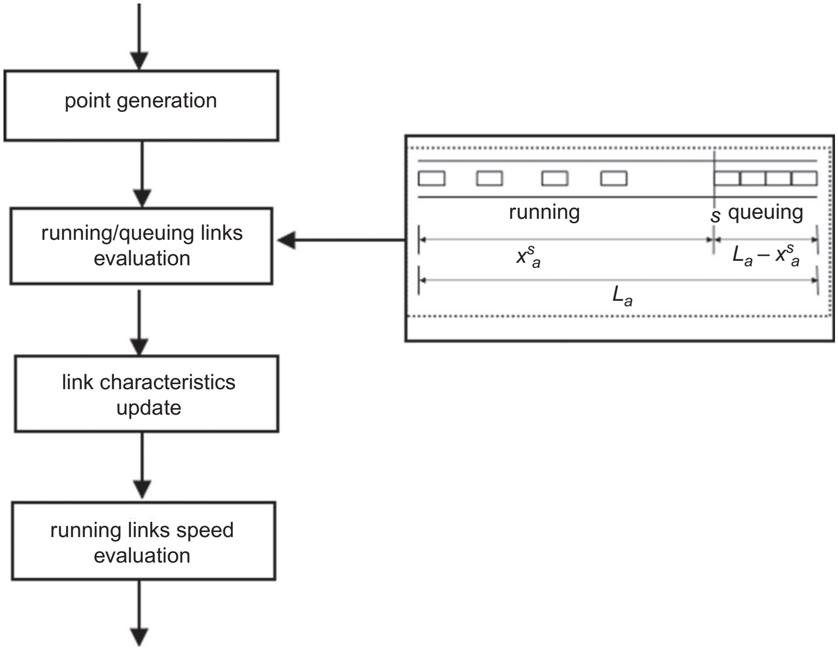


FIG. B.23

Packet generation.

Then the movement rules of packets within each part of the arc and the queue spillback simulation are described. These rules also make it possible to compute the travel time on the running part and on the queuing part.

Let $\Gamma_{od\eta}$ be a sub-graph composed of the set of arcs that belong to the feasible paths connecting the O/D pair computed at time t_0, and let x be the abscissa on arc a belonging to sub-graph $\Gamma_{od\eta}$ representing the position, at time t_1 of interval t_n, of packet P left at the time t_0 of the departure interval $h \leq t_1$ (if $h = t_1$ then $t_0 < t_1$).

With reference to the arc model described above, the following cases may occur:

if $x < x^S{}_{a,}$ packet P is located on the running part of the arc, it moves forward at a speed $v_a{}^{\gamma;}$ then within the interval t_n, packet P may reach a maximum abscissa $x^S{}_a$; consequently.

if $(x^S{}_a - x) < (\delta - t_1)\ v^{\gamma}_{a}$, packet P enters the queuing part of the arc a, before the end of the interval, at time t' computed as in follows:

$$t'_1 = t_1 + \left(x^S{}_a - x\right)/v^{\gamma}{}_a$$

otherwise, at the end of the interval it remains located on the running part;

if $x \geq x^S{}_a$, packet P moves on the queuing part of the arc a; the length of the queuing part travelled by packet P by end of the interval is given by:

$$\Delta = ((\delta - t_1)\,\mathrm{Cap_a}/k^a{}_c$$

if $x + \Delta > l_a$, packet P leaves the arc a during the interval t at time:

$$t_1'' = t_1 + ((l_a - x)\,k^a{}_c)/l_a$$

otherwise it remains on the queuing part of the arc a.

Network equations

When packet P reaches the end of arc a, it is necessary to identify the next arc of its followed path.

Let A^+ be the choice set of available arcs (made up by the arcs of Directed Acyclic Graph—DAG) whose initial node correspond to the final node of arc a, a choice weight is associated to each one of these arcs so that the arc-choice problem can be formulated. All arcs in the set, belong to at least one path connecting the rs pair, starting from r at time t, the arrival at destination of the packet is ensured. For each origin–destination pair, and departure time a DAG sub-graph is associated to the packet P. Such a sub-graph is composed of the set of arcs that belong to the feasible paths connecting the origin–destination pair rs computed at time τ.

A choice weight is associated to each arc; the sub-graph is generated by implicit paths enumeration (i.e. Dial's STOCH; see Dial, 1971) thus the arc-path incidence matrix is dynamically generated at the beginning of each interval t as well as the path flow patterns.

Before entering the next arc a+ it must be verified that (i) arc a+ can accept the incoming packet and (ii) the residual capacity of arc a allows packet P to move to the next arc a+.

When packet P reaches the end of the arc a, before moving forward to the next arc a+, it has to be verified that the length of the running part of the arc a+ is not null, that it is $x^S{}_a + > 0$.

If $x^S{}_a + > 0$, packet P may enter the arc a+ otherwise it means that the entire length of arc a+ is occupied by a queue and packet P remains on arc a until the queue length on arc a+ is smaller than the length of the arc a+, that is for the time until the condition $l_{a+} - x^S{}_{a+} < l_{a+}$ is satisfied.

Once verified that the length of the running part of a arc $a^+ \epsilon A^+$ is not null, that is $x^S_{a+} > 0$, it must be verified that the residual capacity of the arc a allows packet P to move to the next arc a^+. It must be remarked that consistently with the discrete time simulation x^S_a is defined at the beginning of each simulation interval and then it is assumed constant during the whole interval; thus the error strictly depends on the simulation interval size (a smaller interval size induces lower error).

In general two topological representations may be considered at node as displayed in following Fig. B.24; in particular Fig. B.24 guarantees a more realistic representation of spillback phenomena.

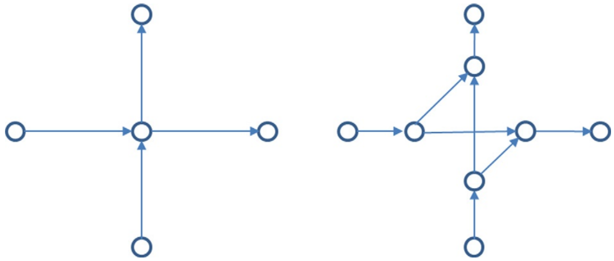


FIG. B.24

Junction topological representation.

Finally some further information needs to be introduced in terms of capacity constraint.

If $[u_a(t_1) + n_e(P)]/t \leq Cap\ \delta/T$ packet P may leave the arc a, otherwise the arc a is saturated and packet P remains on the arc as long as a residual capacity, which allows the packet to leave the arc, becomes available.

Finally it is possible to express:

the arc queue length, $(l_a - x^S{}_a)$.
the (arc) delay, $(l_a - x^S{}_a)\ k_c^a\ Cap_a$.
the network total delay (the total queuing time): $\Sigma_a\ (l_a - x^S{}_a)\ k^a{}_c\ Cap_a$.

B.6 Microscopic models

Models able to simulate the vehicles interactions can be classified in car following models for longitudinal interactions for vehicles along the same lane and lane changing models for vehicles travelling along different lanes. In this section some of the most relevant car following models are discussed.

Main notations used in the following are enlisted below in alphabetical order (notations come first) for reader's convenience(notations come first, then Roman letters, after Greek letters) (Fig. B.25).

a_n is the desired acceleration of vehicle n at time t.
$a_{n,\ max}$ be the upper bound of the vehicle acceleration.
$b^*{}_n$ is an estimate of the deceleration applied by the preceding vehicle.
b_n is the desired deceleration.
$d'{}_{n\text{-}1}$ be the leading vehicle deceleration.
d_n, is the vehicle acceleration.
$d_{n,max}$ is the max deceleration in the braking condition.
n, is the following vehicle.
n-1 is the leading vehicle.

$s_n(t) = x_n(t) - x_{n\text{-}1}(t)$, is the spacing between two successive vehicles, n and n-1, front to front;
$s_{n,\ max}$ is the lower bound of the free flow driving condition.
$s_{n,\ min}$ is the upper bound of the braking condition.
$s_{n,\ s}$ is the safety spacing representing the upper bound of the collision condition.
$s_{n\text{-}1}$ is the effective length.
$T = 1/w\ k_{jam}$ is the time shift between two consecutive trajectories with w wave speed and k_{jam} density;
u is the speed of vehicles travelling along the highway.
v is the vehicle speed.
$v'_n(t)$ is the vehicle free flow speed.
$v^{\beta}{}_n(t)$ is the vehicle speed due to the presence of the leading vehicle.
$v^{d}{}_n(t)$ is the desired vehicle speed.
v_K is the equilibrium speed as a function of density k.
x is the vehicle position.
$x_n(t+T)$ is the longitudinal position of vehicle n at time t+T.
$\Delta v_n\ (t\text{-}\tau_n)$ is the speed difference between subject vehicle and leading vehicle.
Δx is the relative position between leading and following vehicle.
$\delta = 1/k$ is the space shift.
λ is the sensitivity parameter which may assume different functional forms.
τ_n is the reaction time.

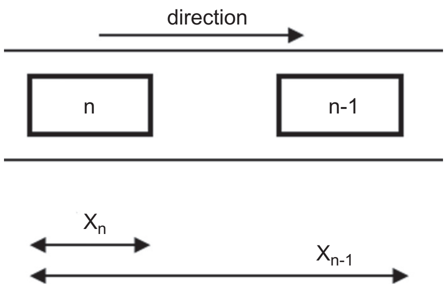


FIG. B.25

General overview of the road stretch with leading and following vehicles.

In particular in this section an overview of five main models is provided: the Stimulus- Response models, the Safety and Collision Avoidance models, the Lower Order models, the Psycho-Physical models and finally an alternative approach based on a Cellular Automata models is also briefly presented (see Fig. B.26).

- Engineering approach
  - GHR models
    - Chandler et al. (1958)
    - Helly (1959)
    - Edie (1960)
    - Gazis et al.(1961)
    - Yang and Koutsopoulus (1996)
  - Safe distance models
    - Gipps (1981)
    - Leutzbach (1988)
    - Jepsen (1998)
  - Lower order model
    - Newell (2002)
- Psyco - physical approaches
  - Action point model
    - Michaels (1963)
    - Wiedemann (1974)
    - Hoffman and Mortimer (1996)
    - Francher and Bareker (1998)
    - Krauss et al. (1999)
- Cellular automata model
  - Nagel and Schreckenberg (1992)
  - Barlovic et al. (1998)
  - Kerner, Klenov and Wolf (2002)

FIG. B.26

Microscopic models overview.

B.6.1 Stimulus–response models

Gazis-Herman-Rothery (GHR) model

The Gazis-Herman-Rothery (GHR) family of models is probably the most studied model class. The basic relationship between a leader and a follower vehicle is in this case a *stimulus-response* type of function that was first introduced by the General Motors research laboratories (Chandler et al., 1958; Gazis et al., 1961). The framework assumes that each driver responds to a given stimulus in accordance with the following relationship:

$$\text{response} = \text{sensitivity} \times \text{stimulus}$$

In general the following vehicle response (the acceleration) is strictly influenced by the speed difference between follower and leader, and the space headway.

The first developed model was by a Chandler et al. (1958); it was a mono-regime model based on a linear expression in which the vehicle acceleration is proportional to the relative speed between follower and leader (stimulus)

$$a_n(t) = \lambda \Delta v_n (t - \tau_n) \quad \text{(B.44)}$$

The main limitation of this model was on the independence of the stimulus from vehicles distance thus if the speed of the leading vehicles is higher than the speed of the following vehicle the acceleration is positive however the acceleration is still positive also in the case of small distances. Further developments were proposed in order to increase the model realism specifically the sensitivity term was modified in order to be proportional to the speed and inversely proportional to the vehicles distance.

The final expression is

$$a_n(t) = \alpha \cdot v^{\beta}{}_n (t) \cdot \Delta v_n (t - \tau_n) / \Delta x_n (t - \tau_n)^{\gamma} \quad \text{(B.45)}$$

where $\alpha > 0$, β and γ are model parameters that control the proportionalities.

However this model is unrealistic in representing the human ability to perceive small changes in driving conditions in particular any response is appreciated when the speed difference is null and in case of low density conditions.

In order to capture differences in driving behaviour Yang and Koutsopoulos (1996) proposed a multi-regime model in which depending on the spacing between vehicles three different driving conditions may occur:

- *emergency*: if the headway is lower than a fixed threshold (h_{lower});
- *free flow driving*: if the headway is higher than a fixed threshold (h_{upper});
- *car-following*: if the headway is between two thresholds above.

$$a_n(t) = \alpha_{acc/dec} \cdot v_n{}^{\beta}{}_{acc/dec} (t) \cdot \Delta v_n (t - \tau_n)^{\lambda}{}_{acc/dec} / \Delta x_n (t - \tau_n)^{\gamma}{}_{acc/dec} \quad \text{(B.46)}$$

where $\alpha_{acc/dec}$, $\beta_{acc/dec}$, $\gamma_{acc/dec}$ and $\lambda_{acc/dec}$ are parameters to be calibrated.

Stability in microscopic models

The equation proposed by Chandler is a differential equation thus it is necessary to study the stability and the stationarity of the equation and then of a vehicles stream. As already anticipated this model is realistic only in the case of high spatial density and it is not able to analyse vehicles travelling in platoons which are independent of each other.

A vehicles stream can be considered stable if a perturbation of the motion of one vehicle in the flow slightly affects the others vehicles without introducing any phenomenon amplification.

A vehicles stream can be considered stationary if all variables describing the flow are constant over time; in general an unstable vehicles stream never achieves the stationary condition.

Stability may be studied with reference to two successive vehicles and all analyses may be further extended to the whole vehicles stream.

In particular the local stability may be distinguished if the analyses are limited to a set of nearby vehicles whereas in the case of asymptotic stability the analyses may be extended to all vehicles in stream.

Usually stability is primarily studied at local level and then extended to the global conditions.

Local stability has been proposed by Herman in the case of $\beta=\gamma=0$ and the general car-following equation may be rewritten as:

$$a_n(t)=\alpha\cdot\Delta v_n\,(t-\tau_n) \quad \text{(B.47)}$$

stability depends on $\alpha\tau_n$; in particular four different conditions may be observed (Fig. B.27):

$\alpha\tau_n\leq 1/e$: non-harmonic motion;
$1/e<\alpha\tau_n<\pi/2$: damped harmonic motion;
$\alpha\tau_n=\pi/2$: simple harmonic motion;
$\alpha\tau_n>\pi/2$: increasing harmonic motion.

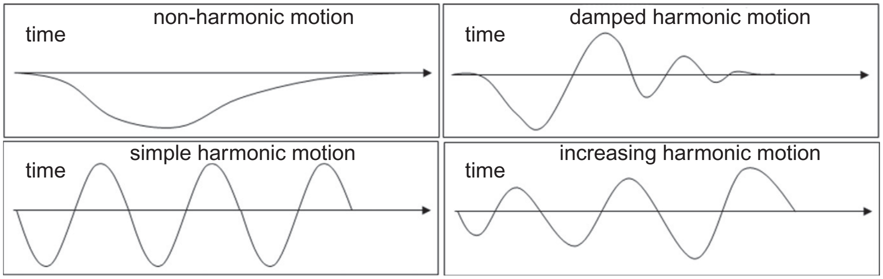


FIG. B.27

Local stability in a car-following model: vehicles relative speed.

The local stability may occur when drivers travelling in the same stream are characterised by the same reaction times and they rapidly react to the stimulus.

If for a specific driver the term $\alpha\tau_n$ is lower than 1/e (around 0.37) the asymptotic stability will be observed otherwise in the opposite condition local stability cannot exist.

Asymptotic stability may be observed in two conditions:

$\alpha\tau_n\leq 1/e$ stable;
$\alpha\tau_n > 1/e$ unstable.

The stationarity phenomenon may exist only when delay τ_n equals zero otherwise stability will never be observed.

When driver react with any delay, the following general equation may be applied:

$$a_n(t)=\alpha_{acc/dec}\cdot {v_n}^{\beta}{}_{acc/dec}\,(t)\cdot\Delta v_n\,(t-\tau_n)_{/}\Delta x_n\,(t-\tau_n)^{\gamma}{}_{acc/dec} \quad \text{(B.48)}$$

The solution for the equation with specific values of β and γ, provides some macroscopic models in particular:

$\beta=0$, $\gamma=1$: Greenberg's model;
$\beta=0$, $\gamma=2$: Greenshields's model;
$\beta=0$, $\gamma=3$: Drake's model.
$\beta=1$, $\gamma=2$: Underwood's model.
$\beta=1$, $\gamma=3/2$: Drew's model.

B.6.2 Safety and collision avoidance models

In this class of models drivers of the following vehicle try to completely preserve the safety distance with respect to the leading vehicle. In particular the speed is selected by the driver in order to ensure that the vehicle can be safetly stopped in the case that the preceding vehicle should suddenly brake.

The safety distance is computed on the base of the motion equations. Gipps (1981) proposed a multi-regime model in which two driving conditions are identified: the free flow driving and the car-following regime; the driver chooses the smaller speeds between them:

$v_n(t+\tau_n) = \min\{c; d\}$ where c and d are equal to.

$$c = v_n(t) + 2.5\, a_n\, \tau_n\, (1 - v_n(t))/v^d{}_n)\left(0.025 + v_n(t)/v^d{}_n\right)^{1/2};$$

$$d = b_n(t)\, \tau_n + b_n{}^2(t)\tau_n{}^2 - b^*{}_n\left[2(\Delta x_n(t) - s_{n-1}) - v_n\, \tau_n - v_{n-1}(t)^2/b^*\right] \quad \text{(B.49)}$$

B.6.3 Lower order models

These classes of models unlike models that operate on acceleration or speed directly operate on vehicles position thus are called *lower order models*.

A model developed by Newell (2002) was based on the assumption that the time space trajectory of vehicles on a homogeneous highway is identical to the preceding vehicles' trajectory except for space and time shifts then

$$x_n(t+T) = min\begin{cases} x_n(t) + vT, & infreeflowconditions \\ x_{n-1}(t) - \delta, & incongestionconditions \end{cases} \quad \text{(B.50)}$$

It must be observed that in this model a driver's reaction time is not considered.

Link between microscopic and macroscopic models

It may be verified that a macroscopic model, in particular Payne's model, is derived from a car following model.

$$v(x(t+\tau_n), t+\tau_n) = v_k(k(x+\Delta x), t)$$

In order to derive the partial differential equation of Payne's model, Taylor's expansion rule has to be applied respectively on the left and right term of the previous equation

$$v(x(t+\tau_n), t+\tau_n) \simeq v(x,t) + \tau_n v(x,t)\, \partial v/\partial x + \tau_n \partial v/\partial t$$

and

$$v_k(k(x+\Delta x), t) \simeq v_k(k(x,t) + \Delta x \partial k/\partial x\, \partial\, v_k(k(x,t)/\partial\, k$$

The traffic density k equals $1/\Delta x$ thus the first equation about the car following model may be reduced to

$$v(x,t)\, \partial v/\partial x + \partial v/\partial t = 1/\tau_n\, (v_k(k(x,t) - v(x,t)) - (1/\tau_n \partial v_k/\partial k)\, (1/k\, \partial k/\partial x)$$

which may be reduced to the following equation.

$$v(x,t)\,\partial v/\partial x + \partial v/\partial t = 1/\tau_n\left(v_k(k(x,t) - v(x,t)) - c^2(k)\,(1/k\,\partial k/\partial x)\right.$$

$$if\ (1/\tau_n\,\partial v_k/\partial k) = c^2(k)$$

B.6.4 Psycho-physical or action point models

This class of models was introduced in order to improve traffic safety and to better understand some of the most relevant traffic phenomena such as capacity drop, hysteresis phenomenon, stop and go oscillations. These models aim to reproduce the human abilities and the errors by explicitly introducing the human-factors in the representation of the driving process.

In general it is expected that each driver is different in his driving style therefore it is necessary to integrate the latest CF models from both engineering and psychological perspectives.

Some of the human factors affecting drivers' behaviour are

- The socioeconomic characteristics
- The reaction time
- The accuracy (the error) in estimating spacing and speeds
- The perception threshold
- The ability in predicting traffic situations
- The impact of different traffic situations (context sensitivity)
- The imperfect driving: the same driver may exhibit different behaviours even though the traffic conditions are the same
- The aggressiveness and risk propensity
- The distraction
- The desired speed/spacing/headway

The basics of the *psycho-physical* models are the introduction of perceptual thresholds aiming to define the minimum value of the stimulus affecting the driver's reaction avoiding the simulation of the driver's reaction in case of the small changes.

These thresholds or *action points* introduced by Michaels (1963) are expressed as a function of speed difference and spacing between two successive vehicles in a car following regime. In general thresholds are able to alert drivers or provide more freedom depending on the spacing if it is small or large. The key point is the introduction of driver's perception of vehicle distance by the effect of different relative speed perception due to the visual angle threshold.

Wiedemann's model

Wiedemann (1974) proposed a further method for thresholds computation in order to identify four driving regimes (Fig. B.28)

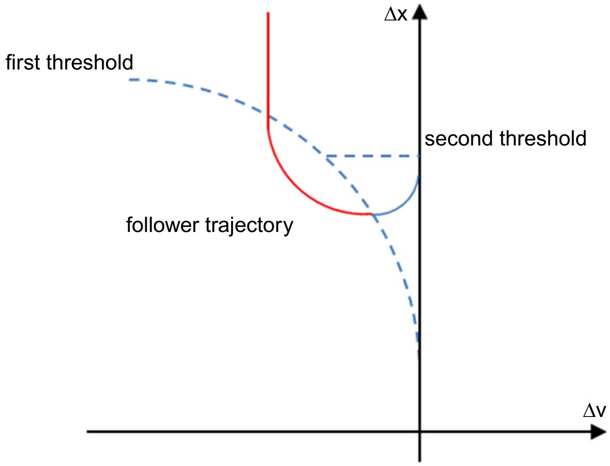


FIG. B.28

Thresholds of the action point model.

A more detailed example is provided in following figure:

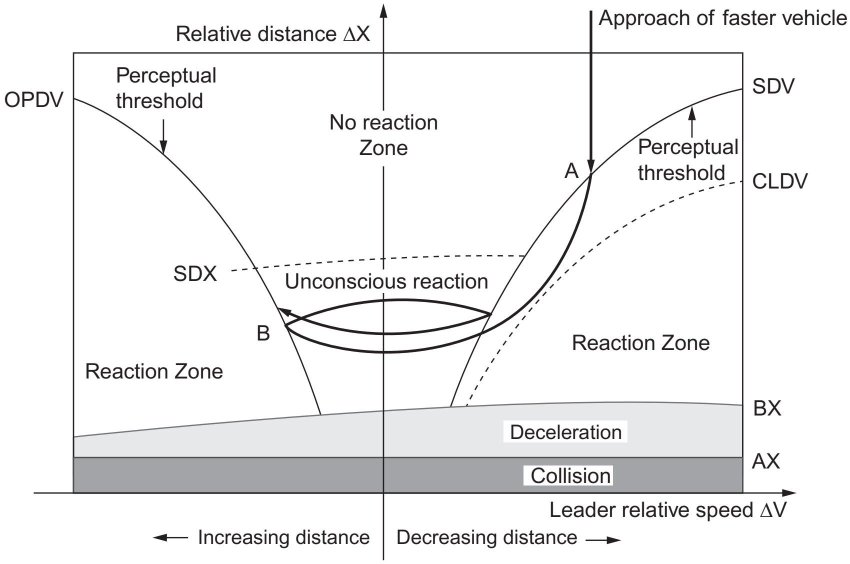


FIG. B.29

Wiedmann's CF model.

Source: Wiedemann, R., 1974. Simulation des Strassenverkehrsflusses. Schriftenreihe des Institutes für Verkehrswesen der Universität Karlsruhe.

AXE: the desired spacing between the front sides of two successive vehicles in a standing queue; $AXE = L_{n-1} + AX_{add}$ where L_{n-1} is the length of the leading vehicle.

ABX: the desired minimum following distance; it is a function of AXE, the safety distance and speed; $ABX(t) = AXE + BX\ [v_n(t)]^{(1/2)}$ where $v_n(t)$ is the minimum of the speed of the subject vehicle and the lead vehicle.

SDV[g]: the action point where a driver consciously observes that he/she is approaching a slower leading vehicle $SDV(t) = [(x_{n-1}(t) - x_n(t) - L_{n-1} - AXE)/CX]^2$.

CLDV: closing delta velocity is an additional threshold that accounts for additional deceleration by the application of brakes; $CLDV\ (t) = [(x_{n-1}(t) - x_n(t) - L_{n-1} - AXE)/CLDVCX]^2$.

OPDV (Opening Difference in Velocity) curve is primarily a boundary to the *unconscious* reaction region. It represents the point where the driver notices that the distance between his or her vehicle and the lead vehicle is increasing over time. When this realisation is made the driver will accelerate in order to maintain the desired space headway thus this is the action point where a driver notices that he/she is slower than a leading vehicle; $OPDV = CLDV \cdot k$.

SDX: A perception threshold to model the maximum following distance; it is 1.5–2.5 times BX; $SDX(t) = AXE + EX \cdot BX\ [v_n(t)]^{(1/2)}$.

$AX_{add,}$ CLDVCX, k, EX.

The dark line in Fig. B.29 shows the decision path of an approaching vehicle. A vehicle travelling faster than the leader will get close to it until the deceleration perceptual threshold (SDV) is crossed (*at Point A*). The driver will then decelerate to match the leader's speed. However, as a human being, the driver is unable to accurately replicate the leader's speed, and spacing will increase until the acceleration perceptual threshold (OPDV) is reached (*at Point B*). The driver will again accelerate to match the leader's speed and the process continues, as shown in the unconscious reaction zone.

In particular in accordance with the identified action points four driving regimes are identified:

Free deriving regime: the driver applies the maximum value of acceleration, b_{max}, in order to achieve the desired speed;

Closely Approaching regime: the regime occurs when a vehicle in the Free Driving Regime passes the SDV Perception Threshold the driver of the following vehicle applies deceleration in order to preserve the distance value ABX

$$a_n(t) = (1/2)\,(\Delta v)(t-\tau_n)^2/[ABX\,(t-\tau) - \Delta x(t-\tau_n)] + a_{n-1}(t-\tau_n) \quad \text{(B.51)}$$

Car-following process is defined in accordance with following constraints:

$$ABX(t) \le \Delta x \le SDX(t)$$
$$CLDV(t) \le \Delta v \le OPDV(t)$$

[g] A further similar threshold can be applied when the subject vehicle is already engaged in following the lead vehicle; it is still the point where the driver notices that the distance between his or her vehicle and the lead vehicle is decreasing over time.

and further described through the two following regimes:

Deceleration following regime: The deceleration following regime occurs as a result of a vehicle in the approaching or closely approaching regime passes the perception threshold or a vehicle in the acceleration following regime passes the second perception threshold.

Acceleration following regime: The acceleration following regime occurs when a vehicle in the deceleration following regime passes the opening difference in velocity threshold or a vehicle in the emergency regime passes the minimum following distance threshold.

Emergency regime: regime occurs any time that the space headway is below the minimum following distance threshold and may be influenced by the leading vehicle's behaviour that suddenly decelerates:

$$a_n(t) = (1/2)\,(\Delta v)(t-\tau_n)^2/AX \cdot \Delta x\,(t-\tau_n) + a_{n-1}(t-\tau_n) + b_{\text{min}}.\,[ABX(t-\tau_n) - \Delta x\,(t-\tau_n)]/BX \qquad \text{(B.52)}$$

where b_{min} is the maximum value of deceleration.

B.6.5 Cellular automata models

Another approach that may be proposed for microscopic traffic flow modelling, is the cellular automata. In this kind of model the space is discretized in an array of cells and each cell may be empty or not depending on the vehicle presence. The state of each cell is described in discrete time steps and is affected by the state of neighbouring cells.; the length of each cell is related to the vehicles length (around two/three vehicles) and the array dimension depends on the number of lanes. Vehicles move to cells ahead based on specified rules.

Nagel-Schreckenberg model

A simplified approach has been proposed by the Nagel-Schreckenberg model in which a single lane is considered and each cell size corresponds to the length of one vehicle. Speed is unidimensional and is given by the ratio between the cell length and the time step; furthermore speed is usually constrained by a maximum value.

Vehicle speed and position at the next time steps are updated in accordance with the following three equations.

Deterministic acceleration:

v(t) is the vehicle speed.

The vehicle speed $v^*(t+1)$ may be updated considering the minimum value among three terms:

g_α is the number of empty cells between the vehicle and the next vehicle and the corresponding speed is $g_\alpha - 1$.

the desired speed, v_0.

the speed obtained when accelerating and given by $v+1$ in free flow conditions.

the safety speed.

Dawdling:

depending on a random term p, the speed may not be accelerating or decelerating thus

$$v(t+1) = \max \begin{cases} (v^*(t+1) - 1, 0) & \textit{with probability } p \\ v^*(t+1) & \textit{otherwise} \end{cases}$$

Driving:

$$\mathrm{x}(\mathrm{t}+1) = \mathrm{x}(\mathrm{t}) + \mathrm{v}(\mathrm{t}+1)$$

B.7 Summary

B.7.1 Major findings

This chapter aims to provide an overview of the traffic flow theory. In particular starting from the main definitions and assumptions, the focus in the first part is on the stationary models whilst in the second part on the non-stationary models. In both cases all analyses are referred to the running and queuing links. Additionally in non-stationary models the network equations are also described in order to properly support the applications at a network level.

Some relevant considerations may be made specifically with respect to each of dynamic models discussed in the chapter.

Macroscopic models are based on a compact analytical formulation however it is not easy to support the extension at network level; furthermore, as will be discussed in the following section about main remarks, another limitation is on queues modelling. Finally these models require, as properly discussed in Section B.4, a finite difference of solution methods.

Some relevant difference may be revealed in mesoscopic models. Indeed these models are not based on a compact formulation however the extension at network level may be easily pursued. A proper queue modelling and spillback simulation is also formalised in mesoscopic models. Finally these models may be easily solved through a simulation approach. Both models, macroscopic and mesoscopic should be considered for large scale analyses. Regarding the microscopic models these provide a very detailed representation including queuing phenomena, junction control etc. however these models are based on several parameters thus the calibration procedure is very complicated unlike macroscopic and mesoscopic models.

In general, as already specified in the book, several parameters in the chapter need to be calibrated however this relevant issue is out the scope of this book.

B.7.2 Further readings

The most elementary continuous traffic flow model is the first order model developed concurrently by Lighthill and Whitham (1955) and Richards (1956), based around the assumption that the number of vehicles is conserved between any two points if there are no entrances (sources) or exits (sinks). They finally proposed a continuous model known as the Lighthill-Whitham-Richards (LWR). This particular model suffers from several limitations. The model does not contain any inertial effects, which implies that vehicles adjust their speeds instantaneously, nor does it contain any diffusive terms, which would model the ability of drivers to look ahead and adjust to changes in traffic conditions, such as shocks, before they arrive at the vehicle itself. In order to address these limitations Payne (1971), Ross (1988), and Kerner and Konhäuser (1994) proposed a second order continuous model governing traffic flow.

Daganzo (1995) demonstrates that the Payne model, as well as several other second-order models available in the literature, produces unrealistic behaviour for some traffic conditions. Specifically, it is noted that traffic arriving at the end of a densely-packed queue would result in vehicles travelling backwards in space, which is physically unacceptable. This is due to the vehicles behaviour that is influenced by vehicles behind them due to diffusive effects.

As the differential equations used in the LWR model are difficult to solve, especially in situations of high density variations like bottlenecking (in these cases the LWR calls for a shock wave), different approximate techniques have been proposed to solve those equations. Newell (1993), introduces a simplified theory of kinematic waves in which, by using cumulative inflow/outflow curves, the state of flow at an extreme, according to the traffic conditions of another one, can be predicted without considering traffic conditions at intermediate sections. This theory provides a relation between traffic flow and density. Therefore the author proposes a space discrete model (link based) which provides link travel times complying with the simplified kinematic wave theory.

Consistently with simplified first order kinematic wave theory after Newell, Yperman et al. (2006) it presents the link transmission model (LTM) in which link volumes and link travel times are derived from cumulative vehicle numbers. Another way to solve the LWR space continuous problem is introduced by Daganzo (1994) through the "Cell Transmission model", developed as a discrete analogue of the LWR differential equations in the form of difference equations which are easy to solve and also take care of high density changes.

With reference to mesoscopic models classification in the literature at present concerns headway distribution, cluster and gas kinetic models. Regarding the first group, Hoogendoorn and Bovy (1998) developed a headway distribution model for multiclass traffic flow and multiclass multilane traffic flow, whilst an example of cluster models based on the homogeneous representation of the cluster according to the speed and size of the cluster itself was provided by Botma (1978). In contrast, some mesoscopic models are derived in analogy to gas-kinetic theory and are based on dynamic representation of speed distributions (Prigogine and Herman, 1971; Paveri-Fontana, 1975; Hoogendoorn and Bovy, 1998). Moreover, as with cluster-based models, a further classification proposed in the literature refers to packet-based models in which vehicles are assumed grouped into packets (Leonard et al., 1989; Cascetta et al., 1991; Dell'Orco, 2006; Celikoglu and Dell'Orco, 2007). In this case two main methods may be distinguished depending on the representation of packets, namely discrete or continuous (Di Gangi, 1992). These methods consider vehicles as packets and their travel times are calculated as a function of current flow on the link. A similar approach may be found in Mahut et al. (2002) who, on the basis of the space–time queue concept, estimate travel time in accordance with the flow-density relationship.

As highlighted in Dell'Orco (2006), a common problem in mesoscopic models is how to represent the anisotropic property of traffic flow models. In order to obtain a more realistic representation and thus overcome the limitation of all vehicles moving at the same speed, the author proposed a model in which packets are uniformly accelerated, thereby avoiding averaging speed.

Further model developments are proposed in the literature in terms of: acceleration behaviour description (Celikoglu and Dell'Orco, 2007; Celikoglu et al., 2009), link representation which may be considered instantaneous (Ran and Boyce, 1996) or traffic-responsive (see He, 1997; Di Gangi et al., 2016), outflow capacity which may be classified as fixed (He, 1997) or variable (Bliemer, 2006), and multicommodity modelling (Bliemer, 2006). Finally, path choice modelling has been treated by Ben-Akiva et al. (1996), Jayakrishnan et al. (1994), Celikoglu and Dell'Orco (2007), and Bliemer (2006).

Although mesoscopic traffic dynamics (queue lengths, delays, shockwaves etc.) have been widely investigated especially in terms of queue modelling (Ben-Akiva et al., 1998; Ben-Akiva, 2003a, b; Burghout, 2004; Di Gangi et al., 2016), less is known about traffic flow propagation phenomena in terms of dispersion and vehicle discharging.

Regarding microscopic models the main efforts are on the car-following theory able to model the interaction between the leading and the following vehicle; each vehicle reacts to the stimulus of the leading vehicle in terms of driving behaviour by accelerating or decelerating. The basic representation of the dynamic representing the interaction in the car-following models, is the stimulus–response approach in which the stimulus may be represented through the vehicle speed, the acceleration, the relative speed and the spacing between vehicles. The General Motors model was the most well-known stimulus–response model, which was first put forward by Chandler et al. (1958); in particular in this model stimulus is specified through the relative vehicles speed then each vehicle tends to move in accordance with the speed of the leading vehicle. Further developments may be found in Gazis et al. (1959) in order to overcome the main limitation of not explaining the traffic situation in higher density. The Collision avoidance (Gipps, 1981) models focus on a safe distance rather than describing a stimulus–response type function; in accordance with this model according, the collision would be unavoidable if the distance is shorter than the safe distance. However one of the model limitations is that the driver might take the behaviour of several preceding vehicles into account and predict to what extent the preceding vehicle might then react by decelerating. Another approach is the physiology-psychology model is also called the Action Point firstly introduced by Michaels (1963). The main idea is that driver reacts if he perceives that it is now approaching the vehicle then thresholds must be defined before the driver reacts. Further enhancements are proposed by Wiedemann (1974) aiming to define the different regimes in car following based on the driver's relative distance and velocity to the front vehicle. Therefore the model considers that larger headways driving behaviour is not influenced or alternatively small headways driving behaviour is influenced only if changes in relative speed and headways are large enough to be perceived. It is assumed that the driver behaves differently in each regime, and then, the acceleration is calculated differently. The considered regimes are the free driving, closing in, and emergency regimes, corresponding to a set of thresholds. Some of these thresholds use a speed parameter, but others rely solely upon the difference in speed between the subject vehicle and the lead vehicle. In the class of microscopic models the cellular automata model firstly introduced by Nagel and Schreckenberg (1992) for traffic flow simulation may also be considered. This model is based on a space discretization corresponding to the road representation by cells. The model was based on a one- dimensional array of cells with some boundary conditions; each cell may be occupied or empty depending if there is a car or not and each cell may be occupied by no more than one car. Regarding the speed it is represented by an integer variable in the range between zero and an upper bound corresponding to the maximum speed (equal to 5 in the original research); the number of cells that vehicles may progress depends on the speed (if speed equals 3 the vehicle may move forward 3 cells).

For further details on other models and some practical examples the book by Treiber and Kesting (2013) about data, models and simulation in traffic flow dynamics is suggested to the reader.

B.7.3 Remarks

In general it must be remarked that the appendix aim is to provide the reader with the main notations, assumptions and definitions with reference to the traffic flow theory. In particular the main idea is to propose to the reader an useful appendix supporting the chapter about within day traffic flow modelling. Indeed arc traffic performance models can be used to specify dynamic network loading (DNL) models, which provide arc flows and travel times consistent with path flows. Therefore DNL can be combined with a travel time model aiming at path travel times computing; the path travel times can be used to define transportation costs.

Another remark may be referred to the spillback and vehicles dispersion phenomena that significantly affect the realism in traffic flow simulation; however these phenomena are not reproduced by all models in literature.

In particular regarding vehicles dispersion during the last five decades several methods have been proposed to predict traffic flow profiles in order to derive the link's delay/offset relation (kinematic theory, diffusion theory, etc.); one of the most straightforward is undoubtedly the platoon dispersion model (PDM) presented by Robertson (1969), widely adopted in a number of practical applications. After discussing the queuing representation in macroscopic models, two different approaches are proposed below in macroscopic and mesoscopic models for vehicles dispersion reproduction.

Queuing in macroscopic models

Below some further considerations are provided about queue models in the case of macroscopic models. As already discussed in the section about major findings, one of the main limitations of macroscopic models is on queue simulation and spillback representation. In particular let

l be the arc length.
n_r be the vehicle running (the number of vehicles in the running link).
n_w be the vehicle waiting (the number of vehicles in the queuing link).

v be the free flow speed.
Cap be arc capacity.

two main queue representations may be identified in literature

- the vertical queue given by $l/v(n_r(t)/l)+0$
- the horizontal queue: $l/v(n_r(t)/l)+n_w(t)/Cap$

where the horizontal queue representation provides a more realistic simulation of spillback phenomena.

However a third approach may also be considered in which the length of the running link is variable and depends on the number of the waiting vehicles then the physical queue [spillback] is given by.

$$((l-n_w(t)l_{veh})/v(n_r(t)/l))+n_w(t)/Cap$$

Dispersion in macroscopic models

As shown above, in CTM equations, due to the assumption that all vehicles travel at the same speed (thereby retaining the same density) to reach the downstream section, there may be no detection of platoon dispersion. To overcome this shortcoming, a platoon dispersion volume function is taken into account for describing situations of low-density cells. In particular, by employing the well-known Underwood (or alternative) speed-density relationship we introduce a further equation:

$$X_i(t)=k_i(t)*v_{0*}\exp[-(k_i(t)/k_c)]$$

where $X_i(t)$ is the platoon dispersion volume function; $k_i(t)=[n_i(t)+n_{i+1}(t)]/2\,\mathrm{L}$ is the density of cell i and cell i+1 at time t, l being the length of the cell; v_0 is the free-flow speed; k_c is the traffic density at maximum flow.Then the flow $f_i(t)$ is given by[h]

$$f_i(t)=\min\{S_i(t),R_{i+1}(t),X_i(t)\}$$

Furthermore it may be verified that the model is still consistent with the first-order traffic flow theory. In particular, let.

k be the density,
f be the flow,
v be the average speed,
v_K be the equilibrium speed.

For completeness, the proposed model is based on the following three equations, consistent with first-order traffic flow theory:

$$(conservation\ \ equation)\,(\partial f/\partial x)+(\partial k/\partial t)=0$$
$$(flow-density\ \ relationship)\,f=kv$$
$$(speed-density\ \ relationship)\,v_K=v(k)$$

[h] The modified CTM differs from the traditional CTM only in the definition of an additional parameter which refers to traffic jam conditions. The definition of demand and supply, the flow function and the density updating equation remain the same as those in the traditional CTM. Thus the Godunov scheme (Lebacque, 1996; Daganzo, 1995) may still be applied.

Furthermore, it is a discretised version of the LWR thus let.

δx be the cell length.
δt be the time step.
$k_i(t)$ be the average density at time t.
$f_i(t)$ be the average exit flow in the interval $[t;\delta t]$

$$k_i\,(t+\delta t)=k_i\,(t)+(\partial t/\partial x)\,(f_{i-1}(t)-f_i(t))$$

Dispersion in mesoscopic models

In order to reproduce the vehicles dispersion in mesoscopic models (consider for instance the dispersion implementation in the proposed model TRAFFMED) the Robertson's model may be directly represented. In particular this model takes the following mathematical form. Let.

T be mean link travel time;
t be 0.8 T;
f_d (j) be the flow rate over a time step Δt arriving at the downstream signal at time interval j;
f_0 (i) be the discharging flow over time step Δt observed at the upstream signal at time interval i;
Δt be the time step duration, usually assumed as 1 s;
F be the smoothing factor;
α and β be dimensionless model parameters.

Let us consider a generic interval j of length Δt and a generic time instant $0 < t_0 < \Delta t$ within interval j. At time interval j the following are known: the density k(j), the link speed v(j) and the cruise time T(j), derived from the previous time interval j-1. The speed could be obtained by any speed density function (fundamental diagram – stable regime). Let us consider a packet p which reaches link a at time t_0, and the number of packets $n(t_0)$ which reach link a during interval j until time instant t_0. The entry flow at link a could thus be obtained as $f(t_0) = n(t_0)/t_0$.

Assuming that within the generic interval j steady state conditions hold, the previous equation could be rewritten as

$$k(j)\,v(j)=f_d\,(j)$$

References

Ben-Akiva, M., 2003a. Development of a deployable real-time dynamic traffic assignment system. In: Evaluation Report (Part A): Evaluation of Estimation and Prediction Capabilities. Massachusetts Institute of Technology.
Ben-Akiva, M., 2003b. Development of a deployable real-time dynamic traffic assignment system. In: Evaluation Report (Part B): Planning Functionality (DynaMIT-P) and Applications. Massachusetts Institute of Technology.

Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H., Mishalani, R., 1998. DynaMIT: A Simulation-Based System for Traffic Prediction. (Presented at DACCORD Short Term Forecasting Workshop).

Ben-Akiva, M.E., Bierlaire, M., Koutsopoulos, H.N., Mishalani, R.G., 1996. DynaMIT: dynamic network assignment for the management of information to travelers. In: Presented at the 4th Meeting of the EURO Working Group on Transportation, Newcastle.

Bliemer, B.C.J., 2006. Dynamic queuing and spillback in an analytical multiclass dynamic traffic assignment model. In: Proceedings of the 1st International Symposium on Dynamic Traffic Assignment, 21–23 June 2006. The University of Leeds, UK.

Botma, H., 1978. State of the Art Report Traffic Flow Models. Part viii: Microscopic Traffic Flow Models. Stichting Wetinschappiligk Undesgack Veskussveiligheid. No. R-78-42.

Burghout, W., 2004. Hybrid Microscopic-Mesoscopic Traffic Simulation. PhD Diss, Royal Institute of Technology.

Cascetta, E., Cantarella, G.E., Di Gangi, M., 1991. Evaluation of control strategies through a doubly dynamic assignment model. Transp. Res. Rec. N. Y., 1306.

Celikoglu, H.B., Gedizlioglu, E., Dell'Orco, M., 2009. A node-based modeling approach for the continuous dynamic network loading problem. IEEE Trans. Intell. Transp. Syst. 10 (1), 165–174.

Celikoglu, H.B., Dell'Orco, M., 2007. Mesoscopic simulation of a dynamic link loading process. Transp. Res. C: Emerg. Technol. 15 (5), 329–344.

Chandler, R.E., Herman, R., Montroll, E.W., 1958. Traffic dynamics: studies in car following. Oper. Res. 6, 165–184.

Daganzo, C.F., 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B: Methodological, 28 (4), 269–287.

Daganzo, C.F., 1995. Requiem for second-order fluid approximations of traffic flow. Transp. Res. B 29 (4), 277–286.

Dell'Orco, M., 2006. A dynamic network loading model for mesosimulation in transportation systems. Eur. J. Oper. Res. 175 (3), 1447–1454.

Di Gangi, M., 1992. Continuous-flow approach in dynamic network loading. In: Second International CAPRI Seminar on Urban Traffic Networks, Compendium, vol. I.

Di Gangi, M., Cantarella, G.E., Di Pace, R., Memoli, S., 2016. Network traffic control based on a mesoscopic dynamic flow model. Transp. Res. C: Emerg. Technol. 66, 3–26.

Dial, R.B., 1971. A probabilistic multi-path traffic assignment model which obviates path enumeration. Transp. Res. 5, 83–111.

Gazis, D.C., Herman, R., Potts, R.B., 1959. Car-following theory of steady-state traffic flow. Oper. Res. 7 (4), 499–505.

Gazis, D.C., Herman, R., Rothery, R.W., 1961. Nonlinear follow the leader models of traffic flow. Oper. Res. 9, 545–567.

Gipps, P.G., 1981. A behavioural car-following model for computer simulation. Transp. Res. B 15, 105–111.

Greenshields, B.D., 1935. A study of traffic capacity. In: Proceedings of the Highway Research Board. vol. 14. Highway Research Board, Washington, DC.

He, Y., 1997. A Flow-Based Approach to the Dynamic Traffic Assignment Problem: Formulations, Algorithms, and Computer Implementations. Doctoral Dissertation, Massachusetts Institute of Technology.

Hoogendoorn, S.P., Bovy, P.H., 1998. New estimation technique for vehicle-type-specific headway distributions. Transp. Res. Rec. 1646 (1), 18–28.

Jayakrishnan, R., Mahmassani, H.S., Hu, T.-Y., 1994. An evaluation tool for advanced traffic information and management systems in urban network. Transp. Res. C 2, 129–147.

Kerner, B.S., Konhäuser, P., 1994. Structure and parameters of clusters in traffic flow. Phys. Rev. E 50 (1), 54.

Lebacque, J.P., 1996. The Godunov scheme and what it means for first order macroscopic traffic flow models. In: Lesort, J.B. (Ed.), Proceedings of the 13th ISTTT. Pargamon, Oxford, pp. 647–677.

Leonard, D.R., Gower, P., Taylor, N.B., 1989. CONTRAM: structure of the model. In: TRRL Research Report 178. Crowtorne.

Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. In: Paper Presented at the Roy. Soc. London.

Mahut, M., Florian, M., Tremblay, N., 2002. Application of a simulation-based dynamic traffic assignment model. In: IEEE 5th International Conference on Intelligent Transportation Systems, Singapore.

Michaels, R.M., 1963. Perceptual factors in car following. In: Proceedings of the Second International Symposium on the Theory of Road Traffic Flow. OECD, Paris, pp. 44–59.

Nagel, K., Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. J. Phys. I France 2, 2221–2229.

Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic, part I: General theory. Transportation Research Part B: Methodological, 27 (4), 281–287.

Newell, G.F., 2002. A simplified car-following theory—a lower order model. Transp. Res. B 36, 195–205.

Paveri-Fontana, S.L., 1975. On Boltzmann-like treatments for traffic flow: a critical review of the basic model and an alternative proposal for dilute traffic analysis. Transp. Res. 9 (4), 225–235.

Payne, H.J., 1971. Models of freeway traffic and control. Mathematical Models of Public Systems. Simulation Council Proceedings Series, vol. 1, no 1, 51–61.

Payne, H.J., 1979. In: Levine, W.S., Lieberman, E., Fearnsides, J.J. (Eds.), Research Directions in Computer Control of Urban Traffic Systems. American Society of Civil Engineers, New York.

Prigogine, I., Herman, R., 1971. Kinetic Theory of Vehicular Traffic. 100 pp.

Ran, B., Boyce, D.E., 1996. Modeling Dynamic Transportation Networks: An Intelligent Transportation System Oriented Approach, second ed. Springer-Verlag, Berlin.

Richards, P.I., 1956. Shockwaves on the highway. Oper. Res. 4, 42–51.

Robertson, D.I., 1969. TRANSYT: A Traffic Network Study Tool. RL Report LR 253, Road Research Laboratory, England.

Ross, P., 1988. Traffic dynamics. Transp. Res. B 22 (6), 421–435.

Treiber, M., Kesting, A., 2013. Traffic flow dynamics. In: Traffic Flow Dynamics: Data, Models and Simulation. Springer-Verlag Berlin Heidelberg.

Whitham, G.B., 1974. Linear and Nonlinear Waves. Wiley, New York.

Wiedemann, R., 1974. Simulation des Strassenverkehrsflusses. Schriftenreihe des Institutes für Verkehrswesen der Universität Karlsruhe.

Yang, Q., Koutsopoulos, H.N., 1996. A microscopic traffic simulator for evaluation of dynamic traffic management systems. Transp. Res. C 4, 113–129.

Yperman, I., 2007. The link transmission model for dynamic network loading. Ph.D. thesis.

Further reading

Barlovic, R., Santen, L., Schadschneider, A., Schreckenberg, M., 1998. Metastable states in cellular automata for traffic flow. Eur. Phys. J. B Conden. Matter Complex Syst. 5 (3), 793–800.

Cascetta, E., 2009. Transportation Systems Analysis: Models and Applications. Springer, pp. 448–477.

Edie, L.C., 1960. Car following and steady state theory for non-congested traffic. Oper. Res. 9, 66–76.

Fancher, P.S., Bareket, Z., 1998. Evolving model for studying driver-vehicle system performance in longitudinal control of headway. In: Transportation Research Record. vol. 1631. TRB, National Research Council, Washington, DC, pp. 13–19.

Helly, W., 1959. Simulation of bottlenecks in single lane traffic flow. In: Proceedings of the Symposium on Theory of Traffic Flow, Research Laboratories, General Motors. Elsevier, New York, pp. 207–238.

Hoffman, E.R., Mortimer, R.G., 1996. Scaling of relative velocity between vehicles. Accid. Anal. Prevent. 28 (4), 415–421.

Jepsen, M., 1998. On the speed-flow relationships in road traffic: a model of driver behaviour. In: Proceedings of the Third International Symposium on Highway Capacity, Copenhagen, Denmark, pp. 297–319.

Kerner, B.S., Klenov, S.L., Wolf, D.E., 2002. Cellular automata approach to three-phase traffic theory. J. Phys. A Math. Gen. 35 (47), 9971.

Krauss, S., Nagel, K., Wagner, P., 1999. The mechanism of flow breakdown in traffic flow models. In: Proceedings of the 14th International Symposium of Transportation and Traffic Theory (Abbreviated Presentations).

Leutzbach, W., 1988. An Introduction to the Theory of Traffic Flow. Springer-Verlag, Berlin.

Payne, H.J., 1971. In: Bekey, G.A. (Ed.), Mathematical Models of Public Systems. In: vol. 1. Simulation Council, La Jolla, CA, pp. 51–61.

Yperman, I., 2007. The Link Transmission Model. Doctoral Dissertation, PhD Thesis, Department of Transport and Infrastructure. Katholieke Universiteit Leuven.

Index

Note: Page numbers followed by *f* indicate figures, *t* indicate tables, and *np* indicate footnotes.

D

Dynamics and Stochasticity in Transportation Systems

Tools for Transportation Network Modelling

Dynamic assignment modeling approaches are critical for the future of modeling transportation networks, providing peak-period solutions where static models fall short. Dynamics and Stochasticity in Transportation Systems: Tools for Transportation Network Modeling breaks new ground, providing consistent and comprehensive coverage of steady state equilibrium assignment, and dynamic assignment within a common framework.

The book details the most recent advances in network assignment, including day-to-day, and within-day dynamics, providing a solid foundation to help transportation planners solve transient overload, and other problems. It fills a gap, describing how to use, and employ, the latest dynamic network models for evaluation of traffic and transport demand interventions.

Dynamics and Stochasticity in Transportation Systems: Tools for Transportation Network Modeling demystifies the many different dynamic traffic assignment approaches, requiring no previous knowledge on the part of the reader. All results are fully described and proven, eliminating the need to seek out other references. The skills described will appeal to transportation professionals, researchers, and graduate students alike.

About the Authors

Giulio E. Cantarella is a full professor of Transportation Systems Analysis and Design in the Department of Civil Engineering at the University of Salerno, Italy, and currently coordinates the department's PhD program. His research focus includes travel demand assignment, transportation systems analysis and design, choice modeling, signal setting design, and urban network design. He has authored more than 80 book chapters and journal articles, including in Elsevier's Transportation Research Part B: Methodological, and Transportation Research Part C: Emerging Technologies.

Stefano de Luca is an associate professor of Transportation Planning and Transportation Systems Theory in the Department of Civil Engineering at the University of Salerno, Italy, and currently the director of the Transportation Planning and Modelling Laboratory. His research focus includes transportation planning techniques, choice modelling, freight/passenger terminal simulation and optimization. He has authored more than 60 book chapters and journal articles, including in Elsevier's Transportation Research Part A: Policy and Practice, Transportation Research Part B: Methodological, Transportation Research Part C: Emerging Technologies, Journal of Transport Geography and Transport Policy.

Roberta Di Pace is a tenure-track assistant professor of transportation systems in the Department of Civil Engineering at the University of Salerno, Italy, and currently she is a committee member of the department's research commission and PhD program. Her research focus includes the traffic flow modeling and simulation, and the analysis of Cooperative – Intelligent Transportation Systems. She has authored more than 50 book chapters and journal articles, including in Elsevier's Transportation Research Part C: Emerging Technologies, Transportation Research Part B: Methodological and Transportation Research Part A: Policy and Practice.

David P. Watling is a centenary professor of transport analysis for the Institute for Transport Studies, at the University of Leeds, Leeds, UK. His research focus includes the analysis, mathematical modelling, and optimization of transportation systems, as well as visioning and alternative / qualitative paradigms for long-range transport planning. He has authored more than 80 journal articles, including in Elsevier's Transportation Part Research B: Methodological, Transportation Research Part A: Policy and Practice, Transportation Research Part E: Logistics and Transportation Review, and Journal of Transport Geography.

ISBN 978-0-12-814353-7

9 780128 143537

ELSEVIER

elsevier.com/books-and-journals