

STRATEGY-BASED DYNAMIC ASSIGNMENT IN TRANSIT NETWORKS WITH PASSENGER QUEUES

Valentina Trozzi

A thesis submitted for the degree of Doctor of Philosophy of Imperial College London

Centre for Transport Studies

Department of Civil and Environmental Engineering

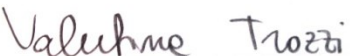
Imperial College London, London, UK

DECLARATION OF ORIGINALITY

At various stages during my PhD, collaboration has taken place with colleagues working on similar subjects. My supervisor, Prof Daniel Graham, has advised me during the last part of the PhD, while Prof Michael G.H. Bell of the University of Sidney advised me during the first part. Prof Guido Gentile of Rome University ‘La Sapienza’ and Dr Ioannis Kaparias of City University London have also advised me at different levels and are co-authors, together with Prof Bell, of some of my conference and published papers on the topic of my PhD.

The *demand model* for dynamic transit assignment, which is formulated in Chapter 3, is the main contribution of this research and includes an innovative *Stop Model* (SM) as well as *Route Choice Model* (RCM). On the other hand, the *supply model* for dynamic assignment includes an extension of the *Network Flow Propagation Model* (NFPM) and the *Bottleneck Queue Model* originally proposed for frequency-based dynamic transit assignment without hyperpaths (Meschini et al., 2007). The formulation of the supply model, together with the methodological implications brought about by the extension of the NFPM by Meschini et al. (2007), is detailed in Chapter 4. The solution algorithm (Chapter 5) is an extension to the context of interest of the *Decreasing Order of Time* method, presented by Chabini (1998) for many-to-one dynamic shortest paths.

Besides the collaboration mentioned above, the work described in this thesis has been carried out by me.


.....
Valentina Trozzi

COPYRIGHT DECLARATION

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

ABSTRACT

This thesis develops a mathematical framework to solve the problem of dynamic assignment in densely connected public transport (or *transit* – the two words are interchangeably used) networks where users do not time their arrival at a stop with the lines' timetable (if any is published).

In the literature there is a fairly broad agreement that, in such transport systems, passengers would not select *the single best* itinerary available, but would choose a *travel strategy*, namely a bundle of partially overlapping itineraries diverging at stops along different lines. Then, they would follow a specific path depending on what line arrives first at the stop. From a graph-theory point of view, this route-choice behaviour is modelled as the search for the shortest *hyperpath* (namely an acyclic sub-graph which includes partially overlapping single paths) to the destination in the *hypergraph* that describes the transit network.

In this thesis, the hyperpath paradigm is extended to model route choice in a dynamic context, where users might be prevented from boarding the lines of their choice because of capacity constraints. More specifically, if the supplied capacity is insufficient to accommodate the travel demand, it is assumed that passenger congestion leads to the formation of passenger First In, First Out (FIFO) queues at stops and that, in the context of commuting trips, passengers have a good estimate of the expected number of vehicle passages of the same line that they must let go before being able to board.

By embedding the proposed demand model in a fully dynamic assignment model for transit networks, this thesis also fills in the gap currently existing in the realm of strategy-based transit assignment, where – so far – models that employ the FIFO queuing mechanism have proved to be very complex, and a theoretical framework for reproducing the dynamic build-up and dissipation of queues is still missing.

TABLE OF CONTENTS

| | |
|---|----|
| 1. Introduction..... | 16 |
| 1.1. Background and objectives | 16 |
| 1.2. Contribution of the thesis | 19 |
| 1.3. Thesis structure | 23 |
| 2. Research background..... | 26 |
| 2.1. Approaches to transit assignment..... | 26 |
| 2.1.1. Basic modelling frameworks | 32 |
| 2.2. Demand-side phenomena | 36 |
| 2.2.1. Application of Random Utility Theory for route-choice modelling..... | 36 |
| 2.2.2. Travel strategies | 41 |
| 2.3. Supply-side phenomena | 44 |
| 2.3.1. Service information and regularity | 44 |
| 2.3.2. Passenger congestion and capacity constraints..... | 45 |
| 2.4. Discussion | 48 |
| 3. Demand model for strategy-based transit assignment with capacity constraints | 51 |
| 3.1. Introduction | 51 |
| 3.1.1. Network representation..... | 52 |

| | | |
|--------|---|-----|
| 3.1.2. | Demand models: basic nomenclature | 57 |
| 3.2. | Demand model for static strategy-based assignment without congestion effects | 59 |
| 3.2.1. | Stop Model: original formulation | 59 |
| 3.2.2. | Route Choice Model: original formulation | 62 |
| 3.2.3. | Stop Model extension: wayside information | 65 |
| 3.2.4. | Stop Model extension: service regularity | 66 |
| 3.3. | Demand model for strategy-based assignment with congestion effects | 68 |
| 3.3.1. | Models with mingling | 69 |
| 3.3.2. | Models with FIFO queues | 74 |
| 3.4. | The proposed dynamic demand model | 77 |
| 3.4.1. | Stop Model | 78 |
| 3.4.2. | Extensions of the dynamic Stop Model | 82 |
| 3.4.3. | Route Choice Model: dynamic hyperpath search | 89 |
| 3.5. | Discussion | 91 |
| 4. | Supply and demand–supply interaction models for strategy-based dynamic transit assignment | 94 |
| 4.1. | Introduction | 94 |
| 4.2. | Supply models for dynamic transit assignment: a review | 96 |
| 4.2.1. | Network Flow Propagation Model for Dynamic Transit Assignment | 96 |
| 4.2.2. | Arc Performance Functions | 100 |
| 4.3. | Supply model: notation and definitions | 102 |
| 4.3.1. | Network representation | 102 |

| | | |
|--------|---|-----|
| 4.3.2. | Supply models: basic supplementary nomenclature | 105 |
| 4.4. | Supply model for strategy-based dynamic transit assignment | 107 |
| 4.4.1. | NFPM for strategy-based dynamic transit assignment | 107 |
| 4.4.2. | Flow-independent APF | 109 |
| 4.4.3. | Bottleneck Queue Model with variable exit capacity | 109 |
| 4.5. | Demand–supply interaction model: dynamic User Equilibrium..... | 112 |
| 4.5.1. | Formulation of the strategy-based dynamic transit-assignment model as a User Equilibrium | 112 |
| 4.5.2. | Characterisation of the network equilibrium | 115 |
| 4.6. | Discussion | 116 |
| 5. | Model implementation | 118 |
| 5.1. | Solution algorithm for static and uncongested strategy-based transit assignment.. | 118 |
| 5.2. | Solution algorithm for dynamic and congested strategy-based transit assignment | 123 |
| 5.2.1. | Decreasing Order of Time (DOT) method: extension | 124 |
| 5.2.2. | Model graph of the solution algorithm | 125 |
| 5.3. | Algorithm Structure..... | 126 |
| 5.3.1. | Part 1: Demand model (RCM and SM) | 130 |
| 5.3.2. | Part 2: Supply model (NFPM and APF) | 132 |
| 5.3.3. | Part 3: MSA (FPP) | 136 |
| 5.3.4. | Part 4: Convergence check and stop criterion..... | 138 |
| 5.4. | Worked Examples | 140 |
| 5.5. | Software implementation | 152 |

| | | |
|--------|--|-----|
| 5.5.1. | Automatic creation of the model graph | 152 |
| 5.5.2. | Definition of the attractive set..... | 155 |
| 5.6. | Case study | 158 |
| 5.6.1. | Data description | 160 |
| 5.6.2. | Results analysis..... | 165 |
| 5.6.3. | Limitations of the case study and implementation issues..... | 168 |
| 5.7. | Discussion | 170 |
| 6. | Conclusions and recommendations for further research..... | 172 |
| 6.1. | Conclusions..... | 172 |
| 6.2. | Recommendations for further research | 176 |
| 7. | References..... | 189 |
| 8. | Appendix..... | 199 |

TABLE OF FIGURES

| | |
|-----------------|-----|
| Figure 1-1..... | 23 |
| Figure 2-1..... | 43 |
| Figure 3-1..... | 53 |
| Figure 3-2..... | 56 |
| Figure 3-3..... | 57 |
| Figure 3-4..... | 71 |
| Figure 3-5..... | 73 |
| Figure 4-1..... | 96 |
| Figure 4-2..... | 102 |
| Figure 4-3..... | 103 |
| Figure 4-4..... | 113 |
| Figure 5-1..... | 126 |
| Figure 5-2..... | 126 |
| Figure 5-3..... | 142 |
| Figure 5-4..... | 143 |
| Figure 5-5..... | 144 |
| Figure 5-6..... | 145 |
| Figure 5-7..... | 146 |
| Figure 5-8..... | 147 |
| Figure 5-9..... | 147 |

| | |
|------------------|-----|
| Figure 5-10..... | 148 |
| Figure 5-11..... | 148 |
| Figure 5-12..... | 149 |
| Figure 5-13..... | 150 |
| Figure 5-14..... | 150 |
| Figure 5-15..... | 151 |
| Figure 5-16..... | 151 |
| Figure 5-17..... | 162 |
| Figure 5-18..... | 162 |
| Figure 5-19..... | 163 |
| Figure 5-20..... | 164 |
| Figure 5-21..... | 166 |
| Figure 5-22..... | 166 |
| Figure 5-23..... | 167 |
| Figure 5-24..... | 167 |
| Figure 5-25..... | 168 |
| Figure 6-1..... | 180 |
| Figure 6-2..... | 182 |

TABLE OF TABLES

| | |
|-----------------|-----|
| Table 2-1 | 28 |
| Table 2-2 | 43 |
| Table 2-3 | 46 |
| Table 3-1 | 80 |
| Table 3-2 | 80 |
| Table 3-3 | 83 |
| Table 3-4 | 83 |
| Table 3-5 | 87 |
| Table 3-6 | 88 |
| Table 5-1 | 140 |
| Table 5-2 | 141 |
| Table 5-3 | 146 |
| Table 5-4 | 161 |
| Table 6-1 | 181 |

TABLE OF EQUATIONS

| | |
|------------|----|
| 2-1 | 37 |
| 2-2 | 37 |
| 2-3 | 38 |
| 2-4 | 39 |
| 3-1 | 58 |
| 3-2 | 61 |
| 3-3 | 61 |
| 3-4 | 61 |
| 3-5 | 62 |
| 3-6 | 62 |
| 3-7 | 62 |
| 3-8 | 63 |
| 3-9 | 63 |
| 3-10 | 63 |
| 3-11 | 63 |
| 3-12 | 63 |
| 3-13 | 64 |
| 3-14 | 64 |
| 3-15 | 66 |

| | |
|------------|-----|
| 3-16 | 66 |
| 3-17 | 70 |
| 3-18 | 71 |
| 3-19 | 71 |
| 3-20 | 72 |
| 3-21 | 73 |
| 3-22 | 73 |
| 3-23 | 74 |
| 3-24 | 76 |
| 3-25 | 76 |
| 3-26 | 77 |
| 3-27 | 78 |
| 3-28 | 79 |
| 3-29 | 79 |
| 3-30 | 81 |
| 3-31 | 82 |
| 3-32 | 86 |
| 3-33 | 86 |
| 3-34 | 87 |
| 3-35 | 87 |
| 3-36 | 90 |
| 3-37 | 90 |
| 3-38 | 91 |
| 4-1 | 98 |
| 4-2 | 107 |

| | |
|------------|-----|
| 4-3 | 108 |
| 4-4 | 108 |
| 4-5 | 108 |
| 4-6 | 109 |
| 4-7 | 109 |
| 4-8 | 109 |
| 4-9 | 109 |
| 4-10 | 110 |
| 4-11 | 111 |
| 4-12 | 111 |
| 4-13 | 111 |
| 4-14 | 111 |
| 4-15 | 112 |
| 4-16 | 114 |
| 4-17 | 114 |
| 4-18 | 114 |
| 4-19 | 114 |
| 4-20 | 114 |
| 4-21 | 114 |
| 4-22 | 115 |
| 4-23 | 115 |
| 4-24 | 115 |
| 4-25 | 115 |

1. INTRODUCTION

1.1. BACKGROUND AND OBJECTIVES

The challenge of sustainability is encouraging a shift in the demand for mobility from individual to collective means of transport, thus creating a requirement for more attractive public transport systems, above all in urban contexts. On the other hand, recurrent passenger congestion and oversaturation on urban public transport systems, as defined by Nuzzolo et al. (2012), are nowadays a severe problem both in developed and developing countries (HoC Transport Committee, 2003; Pucher et al., 2004; Sohail et al., 2006), and are bound to worsen due to the increasing urbanisation of many regions of the world.

Overcrowding has major negative effects because it compromises the basic safety and comfort of commuters, raises the risk of an accident, makes passengers more vulnerable in emergency situations and can prevent elderly or disabled people from boarding buses or train carriages during peak hours. Moreover, passenger overcrowding may increase vehicle dwelling times as well as waiting times at stops, when passengers fail to board because of insufficient capacity. Finally, as highlighted by the HoC Transport Committee (2003), public transport irregularity and unreliability bring about a loss of productivity because employees who arrive late or fail to arrive at work cause the cancellation or rescheduling of meetings as well as lost business, which solely for the City of London has been 'conservatively estimated to be [worth] about £230 million a year' (Oxford Economic Forecasting, 2003: p. 3).

When political, financial and environmental constraints limit the possibility of designing and building new infrastructure to alleviate congestion and, in general, to increase the quality of service provided by public transport systems, it is fundamental to have reliable technical tools to evaluate and compare possible scenarios. These may include alternative or complementary measures such as the building of new high-speed and high-capacity public transport infrastructure, the modification of existing line routes, frequencies and timetables, or the purchasing of new vehicles to increase line capacities.

The technical tools usually exploited to this aim are *assignment models* that describe and predict the patterns of network usage by travellers for the different scenarios/projects. More specifically, assignment models evaluate flows on the different arcs of a network, which depend on: the *travel demand* between different zones of the area of study (measured in number of trips); the users' route choice behaviour; and the reciprocal interaction between travel demand and the characteristics of the transportation services that make up the *transport supply* (Cascetta, 2001).

Compared to traffic assignment models, which reproduce a continuously available transport system (such as a road network), transit assignment models reproduce a transport system available only at specific times and locations, according to the routes and timetables of its lines. Therefore, not only the in-vehicle, access and egress times, but also the waiting time at stations as well as the transfer time between different services have to be reproduced. The latter two terms may be easily evaluated in scheduled transportation systems with low frequency and high regularity, using the services' timetable (transfer and waiting times) and assuming that passengers try to synchronise their arrival times at stations with the vehicles' departures thus minimising waiting times.

However, it is less intuitive how the (average) waiting and transfer times for services with high frequency and low regularity should be evaluated. First of all, in this case, it is

reasonable to assume that passengers do not explicitly consider the services' timetable when making their travel choices. Thus, they do not time their arrival at a stop with a specific run departure and have to wait, at least, for the first vehicle of the chosen line that leaves the stop. In this case, the waiting time is a stochastic variable that depends in some way on the arrival rate of passengers and transit vehicles at the stop. Secondly, at some stops passengers might have the choice between a local and an express service or between lines with partially overlapping routes that connect to the same destination. Thus, in this case, the waiting time at the stop is a stochastic variable that depends on the arrival rate of passengers and vehicles of *all* the lines of choice.

The problem of correctly representing the phase of waiting/transferring at a stop is, therefore, crucial in transit assignment models because it may yield very different results in terms of 'generalised travel time' estimation for the travel options available and, ultimately, may distort the way passenger decisions are modelled with reference to certain network conditions. Beyond service frequency and regularity, as briefly considered above, waiting/transferring times, and thus route choices, can also be significantly affected by capacity constraints of the public transport network that lead to the formation and dispersion of passenger queues at stops during peak periods. For instance, when several alternatives are available from the same stop, it may happen that faster or direct services are overcrowded while others are not, and thus users prefer to board slower lines rather than keep queuing.

For all these reasons, in order to be sound and reliable, transit assignment models have to consider, and reproduce adequately, demand-side and supply-side phenomena that may affect passengers' behaviour and, thus, can yield different results in terms of flow estimations. More specifically, this thesis is concerned with modelling recurrent overcrowding, which is one of the major problems faced in large-city transport networks.

Although several *static* models are already available in the literature to study the effect of passengers' oversaturation in a steady-state setting, those allow only an average evaluation of network performances (for example, in terms of passenger loads on each line) during the analysis period, which may not be satisfactory if the travel demand has a sharp peak. By contrast, fully *dynamic* models can reproduce the build-up and dissipation of oversaturation in the public transport network, the temporary unavailability of supplied capacity, as well as the effects on passengers' route choices that are produced by a decrease in the supplied *Level of Service* (LoS) during the peak period (for example, longer waiting time at the stop and discomfort on-board). Notwithstanding the higher degree of accuracy, only a very few dynamic transit assignment models have been proposed for public transport systems with low regularity and high frequency, which include many urban public transport systems; and as clarified in Chapter 3, these can only reproduce some of the congestion phenomena that may occur at transit stops.

This thesis fills in the gap still existing in the literature and presents a new mathematical framework that solves the problem of dynamic transit assignment in high-frequency networks subject to demand peaks and temporary overcrowding.

1.2. CONTRIBUTION OF THE THESIS

Available transit assignment models differ greatly on the assumptions made about demand-side and supply-side phenomena and, as clarified in the previous section, are not suitable for the reproduction of passenger flows in all possible contexts of interest.

For example, there is a fairly broad agreement that, in densely connected transit networks, users would not select *the single best* itinerary available but would choose a bundle

of partially overlapping itineraries diverging at stops (formally known as a *travel strategy* or *hyperpath*); and that they would then go on one or another path depending on events occurring as the trips unfolds (such as, for example, which bus happens to arrive first at their stop). Strategy-based assignment models are therefore applied to reproduce travel choices in transit networks where services are so frequent and/or irregular that users do not perceive any utility in timing their arrival at a stop using a timetable of the lines' services (if any is published).

By contrast, notwithstanding the importance of problems triggered by transit congestion, there does not seem to be a broad agreement in the literature on how to deal with this phenomenon and to reproduce the effects it may have on passengers' travel choices, as well as on the LoS provided.

For example, when travel demand exceeds supplied capacity, passengers may be prevented from boarding a vehicle at their stop because of overcrowding. They are therefore forced to keep waiting and a queue arises. The queue of those remaining at the stop may also increase passenger congestion for subsequent vehicle arrivals, thus leading to great LoS variations that cannot be properly captured by static models, even if capacity constraints are considered. On the other hand, the longer waiting time due to overcrowding may induce some users to change their itinerary, mode of travel, departure time or destination, or even induce them to cancel their trip. In such a complex scenario, this thesis is only concerned with changes in the route choice that are produced by congestion on the transit network and devises an innovative mathematical framework to reproduce them in a dynamic setting.

More specifically, it is assumed that if the capacity supplied is insufficient to accommodate the travel demand, the stops' layout is such that passengers are forced to wait in a *First Come, First Served*, or *First In, First Out* (FIFO), queue and respect the priority of those who are at the front. This is usually the case in urban bus and tram networks, whereas

for metro and light railway systems it is acceptable to assume instead that boarding priority is not respected among those who are at the stop, because stations are designed with large platforms that allow passengers to mingle when there is congestion.

While several strategy-based models considering mingling have been proposed, only a few (static) models consider FIFO queuing and their very complex formulation prevents any extension to a dynamic setting, which would be able to reproduce the formation and dispersion of queues over the analysis period. Moreover, these models all imply that, if all lines are congested, passengers would rather walk than remain waiting even if frequencies are high, so that the extra waiting time due to congestion is, anyhow, short.

Consequently, this thesis proposes an innovative mathematical framework for *strategy-based* dynamic assignment to transit networks, where it is assumed that users may be prevented from boarding the first vehicle of their choice because of on-board congestion and, as a result, would be forced to continue waiting at the stop according to a FIFO queuing discipline. More specifically it is assumed that, in the context of commuting trips, when queues arise, transit users have a good estimate of the number of passages of the same line they must let go before being able to board at a certain stop and, consequently, of the total queuing time they will experience. Thus travel choices will be (temporarily) affected and passengers might be willing to board a slower service or to change their origin and/or transferring stop in order to avoid congestion.

Figure 1-1 schematically presents the structure of the dynamic assignment model developed in this work, which extends to the context of interest the traditional structure of Deterministic User Equilibrium (DUE) for Dynamic Traffic Assignment (DTA) models detailed in Cascetta (2009: p. 467). The main inputs are:

- On the demand side, the time-varying origin–destination (*od*) matrix;

- On the supply side:
 - The network topology;
 - The characteristics of the transit lines in terms of: vehicle capacity; and time-dependent service frequency, dwelling time and in-vehicle travel time (for reasons of simplicity, it is assumed here that these are not affected by congestion).

To develop this modelling framework, the following four components are to be specified:

- The *Arc Performance Function* (APF), which yields the exit time at any given entry time for each arc, depending on the transit lines' characteristics and the passenger flows over the network;
- The *Stop Model* (SM), which yields for any given line choice set (formally known as an *attractive set* – Nguyen and Pallottino, 1988) the rate of passengers boarding each line (*diversion probability* – Cantarella, 1997) as well as the expected waiting time, depending on the transit lines' characteristics and passenger congestion;
- The *Route Choice Model* (RCM), which reflects the behaviour of a rational passenger, travelling from an origin to a destination, for given arc performances (i.e. time-varying travel and waiting/queuing times) – the deterministic route choice is modelled through a dynamic shortest-hyperpath search;
- The *Network Flow Propagation Model* (NFPM), which aims at finding time-varying arc flows that are consistent with the arc travel times for given route choices but not consistent with line capacities. (This is the main difference between the NFPM and the Dynamic Network Loading Problem where, instead, mutual consistency of flows and times is sought through the APF for given route choices.)

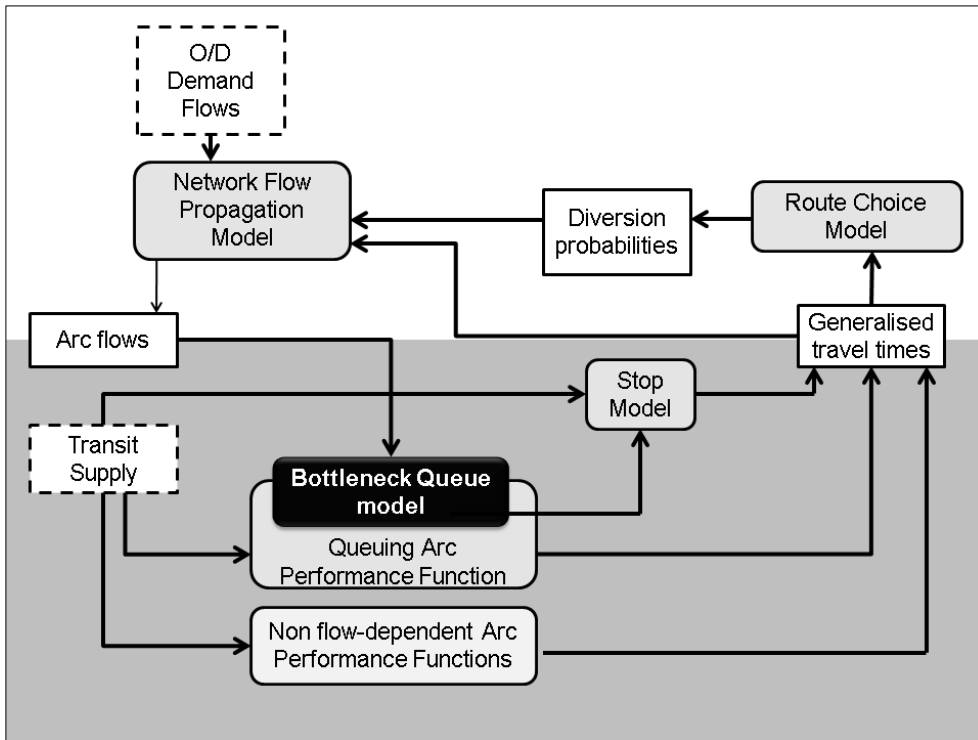


Figure 1-1
Scheme of the fixed-point formulation for the dynamic assignment structure with passenger FIFO queues and without explicit path enumeration.

1.3. THESIS STRUCTURE

The present thesis is organised according to the following structure: Chapter 2 clarifies and details the research background in terms of phenomena that it is necessary to represent in a dynamic transit assignment model. The main methodological innovations required to develop the proposed dynamic User Equilibrium with hyperpaths are described in chapters 3 and 4.

More specifically, Chapter 3 focuses on the *demand model* that associates average values of travel demand to LoS attributes of the transportation system. The two main components of the demand model for dynamic transit assignment are: the *Stop Model* (SM) and the *Route Choice Model* (RCM). The SM is formulated considering the specific

assumption that, in the context of commuting trips, passengers have a good estimate of their lines' average frequencies, travel times upon boarding and congestion levels, expressed as the number of passages of the same line they will miss because of capacity constraints. The RCM, on the other hand, is formulated as a dynamic shortest hyperpath search. This means that, when deciding on their best travel strategy, passengers consider the LoS of the different lines (expressed in terms of frequency, travel time and congestion levels) at the time they expect to board them.

Chapter 4 focuses on the *supply model*, which evaluates network performances (for example, travel times) and flows depending on the travel demand and on the characteristics of the transport systems (for example, the frequencies and capacities of the different lines). The two components of the supply model for dynamic transit assignment are: the *Network Flow Propagation Model* for dynamic assignment (NFPM) and the *Arc Performance Functions* (APF). Beyond the adoption of the *frequency-based* (FB) approach for transit assignment, which implies a line-based supply representation as detailed in Chapter 2, the most relevant characteristics of the supply model are the *continuous-flow* representation and the arc-based *discrete space* representation of the relevant variables. The first assumption means that the flow of passengers is regarded and described as the flow of a fluid, for which the *conservation rule* (Cascetta, 2001: pp. 370–379) holds true; the second assumption implies that variables, such as inflows, outflows, travel times, conditional probabilities and so forth, are defined on a link-basis. Consequently, Chapter 4 extends to a dynamic transit assignment with hyperpaths the *supply model* of Meschini et al. (2007), which also makes use of the *continuous-flow* representation and the arc-based *discrete space* representation. The methodological implications of such an extension are also explained in the same chapter, together with details of the *demand–supply interaction model*, formulated as a Fixed-Point Problem (FPP)

Chapter 5 explains the algorithm implemented to solve the dynamic assignment problem, which extends to the context of interest the *Decreasing Order of Time* (DOT) method originally devised by Chabini (1998) solely for the many-to-one dynamic shortest-path search. Moreover, several worked examples are examined to clarify the effects of the model's assumptions and a case study that uses the tram network of Cracow is presented.

Finally, conclusions are drawn in Chapter 6.

2. RESEARCH BACKGROUND

This chapter reviews the past major methodological achievements in the field of transit assignment, highlighting the assumptions made by different models and, thus, the different contexts of application.

It is organised as follows. First of all, Section 2.1 gives a general classification of existing assignment models and explains the main differences, pros and cons of the two alternative modelling frameworks developed so far for transit assignment: *frequency-based* and *schedule-based* approaches. Then, Section 0 and Section 2.3 detail demand-side and supply-side phenomena, which are mostly relevant to strategy-based transit assignment, such as the one presented in this thesis, and give a brief overview of the methodologies that have been used to represent them. Finally, with respect to the existing literature, Section 2.4 identifies the most important improvements and innovations proposed in this research, which will then be analysed in the following chapters.

2.1. APPROACHES TO TRANSIT ASSIGNMENT

Simulation models create prototypes of complex systems in order to analyse and predict their performance. More specifically, in transport applications, they do so by reproducing and predicting flows (of cars, passengers, pedestrians etc.) in a given network.

This thesis is concerned with *analytical simulation models*, which try to formulate mathematical expressions to reproduce physical and behavioural aspects of the system of interest. More specifically, this work deals with assignment models that, together with trip generation, trip distribution and mode choice, make up the classic four-step structure of analytical simulation models. As the word implies, assignment deals with the problem of assigning objects to predefined categories (Azibi and Vanderpooten, 2002). In the specific case of transit assignment, the objects considered are *passengers* travelling between an origin and a destination (an origin–destination, or *od*, pair), while the categories are *routes* and itineraries connecting the *od* pair in the public transport network.

Research on (traffic and transit) assignment has been carried out for about 60 years and models developed so far may differ remarkably in terms of methodological assumptions and, thus, also in their context of application. In a very recent review (2012), Szeto and Wong point out that, at least for car transport, a general classification may consider criteria such as: the model dynamics; choice dimension modelling; the mathematical formulation approach; and time dimension modelling.

These criteria, summarised in Table 2-1, are quite general and would apply to transit-assignment models as well, as proven by the examples of transit applications that are listed in the same table. With reference to this classification, the model presented in this work is a deterministic within-day dynamic model with continuous time representation and considers the route-choice dimension only with rigid demand. Finally, the dynamic assignment is formulated as an FPP.

A fully detailed discussion of all the criteria that should be considered for a general classification is beyond the scope of this document. On the other hand, because public transport services are ‘discrete both in time and space, as they can only be accessed at certain times and locations’ (Nuzzolo, 2003), beyond the criteria mentioned in Table 2-1, there are

some further distinctions that need to be discussed in terms of modelling frameworks, as is done in the next sub-section.

Table 2-1
Criteria for classifying assignment models

| Criteria | Sub-criteria |
|---|--|
| <p><i>Model dynamics</i> (Szeto and Wong, 2012)</p> | <p>Static assignment</p> <p>Determines the flow in a specific area of the transport system and for a specific period of time for given travel demand and behavioural assumptions (Nguyen and Pallottino, 1988; Spiess and Florian, 1989; De Cea and Fernandez, 1993; Nguyen et al., 1998; Marcotte et al., 2004).</p> |
| | <p>Dynamic assignment</p> <p>Generalises the static assignment problem by considering also the variation in the number of users and service performances over the analysis period, and determines time-varying flows on the network (Poon et al., 2004; Meschini et al., 2007; Sumalee et al., 2009a; Hamdouch and Lawphongpanich, 2008).</p> |
| <p><i>Choice dimension modelling</i> (Szeto and Wong, 2012)</p> | <p>Route and departure time choices</p> <p>Some studies may assume that, for given network condition, users would only select different routes (Nguyen et al., 1998; De Cea and Fernandez, 1993; Marcotte et al., 2004; Meschini et al., 2007), or would only select a different departure time, or would simultaneously select their route and departure time (Nuzzolo et al., 2012; Sumi et al., 1990).</p> |
| | <p>Demand elasticity</p> <p>In models with rigid demand, it is assumed that travellers are bound to travel, whatever the network conditions. On the other hand, in the case of elastic demand, travellers can decide to change mode, destination or cancel their trip (Cantarella, 1997; Huang, 2002).</p> |

| Criteria | Sub-criteria |
|--|---|
| <p data-bbox="204 1003 442 1211">Choice dimension modelling (Szeto and Wong, 2012)</p> | <p data-bbox="464 275 756 304">Travel choice component</p> <p data-bbox="464 338 1396 913">Traditionally, the travel choice is modelled by assuming that passengers are rational decision makers, who have a good knowledge of supply characteristics and try to minimise their total travel time or, equivalently, maximise the <i>travel utility</i> on the basis of this knowledge (Poon et al., 2004). Consequently, in this category of models (<i>deterministic</i> models) all the travel demand between a certain <i>od</i> pair is assigned to the routes with maximum average utility; but if there are two or more alternatives with (equal) maximum utility, there are infinite feasible combinations of demand split among these alternatives and, thus, the demand model does not yield a function that <i>univocally</i> links passenger decisions with network conditions – rather, it yields a <i>one-to-many map</i> (Cascetta, 2009: p. 138).</p> <p data-bbox="464 947 1396 1155">The idealisation that passengers are rational decision makers with a good estimate of supply characteristics (at least, <i>average</i> supply characteristics) does not consider the effect on travel choices determined by personal tastes or distorted perception (Fonzone and Bell, 2010).</p> <p data-bbox="464 1189 1396 1464">Moreover, the uncertainty about network conditions, users’ preferences and other factors that can affect route choices may necessitate the modelling of travel demand by explicitly considering the variance and covariance of the perceived utility/disutility attached to each possible travel alternative (<i>stochastic</i> models) (Yang and Lam, 2006; Sumalee et al., 2009b; Nuzzolo et al., 2012).</p> |

| Criteria | Sub-criteria |
|--|--|
| <p data-bbox="236 954 411 1043"><i>Methodological approaches</i></p> <p data-bbox="217 1077 427 1227">(Boyce et al., 2001; Szeto and Wong, 2012)</p> | <p data-bbox="467 275 724 304">Analytical approaches</p> <p data-bbox="467 349 1396 685">Analytical approaches normally consider the macroscopic travel behaviour of the flow of passengers (which is usually regarded as a fluid) and try to define functions and maps that, at least under some simplifying assumptions, capture the physical and behavioural essence of the system. As pointed out by Szeto and Wong (2012), ‘the main difficulty with the analytical approaches is adding realistic traffic dynamics... to already complicated formulations’.</p> <p data-bbox="467 734 1396 1070">Among analytical approaches, the <i>mathematical programming</i> is mainly used for static assignment. For dynamic assignment to congested networks, the main pitfall of this approach is the inclusion of integrals of time that are path dependent (because the link travel times are non-symmetric functions of the link flows) in its optimisation formulation (Nguyen and Pallottino, 1988; Spiess and Florian, 1989; Lam et al., 1999).</p> <p data-bbox="467 1104 1396 1440">Another possible analytical approach consists in the use of <i>variational inequalities</i>, which can also be seen as a generalisation of the constrained optimisation and fixed-point problem. This is usually the preferred approach for the formulation of dynamic assignment problems because of the relative ease of illustrating mathematical properties, such as the existence and uniqueness of a solution (Hamdouch et al., 2004; Marcotte et al., 2004; Papola et al., 2009).</p> |

| Criteria | Sub-criteria |
|---|---|
| <p data-bbox="204 633 379 723"><i>Methodological approaches</i></p> <p data-bbox="204 757 419 904">(Boyce et al., 2001; Szeto and Wong, 2012)</p> | <p data-bbox="464 275 783 304">Simulation-based approach</p> <p data-bbox="464 338 1394 546">The simulation-based approach emphasises microscopic characteristics of the transport system and tries to simulate the reaction that each single passenger (agent) can have when interacting with the environment (the transport network) as well as other agents.</p> <p data-bbox="464 580 1394 1160">Simulation-based (or, equivalently, agent-based) models are more flexible and provide a more realistic description of the system; however, they also have some major drawbacks. Firstly, they are essentially descriptive and not prescriptive tools because ‘they simulate the probable results of certain... management strategies, but do not prescribe a particular strategy’ (Szeto and Wong, 2012). Secondly, in each computer simulation, agent-based models yield <i>one</i> realisation of route choices out of a large range of possible values and therefore generalisation and transfer of results are not usually possible. Thirdly, agent-based models lack specific and precise properties through which to prove the existence and (possible) uniqueness of the solution or analyse its optimality.</p> <p data-bbox="464 1193 1394 1283">Examples of the simulation-based approach may be found in Rieser et al. (2009) and Cats (2011).</p> |
| <p data-bbox="204 1541 379 1630"><i>Time dimension modelling</i></p> <p data-bbox="204 1664 403 1744">(Szeto and Wong, 2012)</p> | <p data-bbox="464 1395 868 1424">Within-day and day-to-day models</p> <p data-bbox="464 1458 1394 1850">The first group of models assume that travellers make their choices depending on their past experience about the network conditions. There is no learning process involved and the travel choice is considered for a typical day (Spiess and Florian, 1989; De Cea and Fernandez, 1993; Cominetti and Correa, 2001; Kurauchi et al., 2003; Meschini et al., 2007). On the other hand, day-to-day models are concerned with the adjustment of travel decisions (mainly route and departure time) from one day to another (Nuzzolo et al., 2001; Teklu, 2008; Nuzzolo et al., 2012).</p> |

| Criteria | Sub-criteria |
|---|--|
| <i>Time dimension modelling</i> (Szeto and Wong, 2012) | <p>Time representation</p> <p>The analysis period can be represented in a continuous time setting or in discrete time settings. Normally, the continuous time representation (Meschini et al., 2007) is chosen for an accurate mathematical formulation of the problem, while solution methods of assignment models usually require time discretisation. Models formulated with a discrete time representation are also available (Schmöcker et al., 2008).</p> |

2.1.1. Basic modelling frameworks

Two main modelling frameworks are available for transit assignment: *frequency-based* and *schedule-based* assignment. They have different representations of the public transport network and, thus, the choice of framework can have a substantial impact on the route-choice representation. An exhaustive review of FB and SB assignment modelling approaches is provided by Bell and Lam (2003), and only the aspects that are most relevant to this thesis are analysed in the following.

Frequency-based (FB) assignment

FB assignment relies on a line-based supply representation, where each service is considered as a unitary supply facility, with time-dependent performances, such as in-vehicle travel time and service frequency. All the runs of a service are graphically represented together by means of one *line* sub-graph (Nuzzolo, 2003).

From a behavioural perspective, FB models are based on the assumption that passengers perceive a public transport service as a unitary supply facility with a certain expected frequency and in-vehicle travel time. They therefore do not see any advantage in

timing their arrival at a stop or station with the service arrival/departure and, when making their travel decision, would not distinguish between different runs of the same service.

In order to reproduce different levels of service regularity, different assumptions can be made about the Probability Distribution Function (PDF) of headways between two consecutive arrivals of the same line. The vast majority of models assume that the exponential distribution may be used for highly irregular services (Chriqui and Robillard, 1975; Marguier and Ceder, 1984; Spiess and Florian, 1989; De Cea and Fernandez, 1993; Cominetti and Correa, 2001; Cepeda et al., 2006; Kurauchi et al., 2003; Schmöcker et al., 2008; Leurent et al., 2011) and the uniform distribution for regular services (Spiess and Florian, 1989; Billi et al., 2004; Gentile et al., 2005). Some other models (Gendreau, 1984; Bouzaïene-Ayari et al., 2001; Gentile et al., 2005; Noekel and Wekeck, 2008) have also used the Erlang distribution, which has the advantage of major flexibility because, by changing its parameters, it is possible to reproduce different levels of regularity; but this lacks the analytical tractability of the other two PDFs.

From a modelling perspective, FB assignment is advantageous because it reproduces more realistically the choice process of passengers who travel in densely connected transit networks, with high-frequency and/or low-regularity services. Moreover, it requires only a ‘relatively detailed network representation which involves the walking time to a stop, the waiting time for a transit vehicle, the transfers between lines if more than one line is taken and the in-vehicle time’ (Florian, 2003). Finally, such an approach is suitable for strategic and long-term planning of large transit networks, when the detailed schedule of every service is not defined.

For these reasons, FB models are also widely applied in commercial packages, such as EMME/2 and TransCAD.

On the other hand, FB assignment has raised concerns regarding behavioural assumptions and the level of detail of the output produced. Firstly, the assumption that passengers do not know or do not explicitly consider the lines' timetable is unrealistic in networks with low-frequency and/or high-regularity services or in networks where the use of Advanced Traveller Information Systems (ATIS) is so high and reliable that travellers may access, in every point of the network, timely, accurate and exhaustive information about the whole transit system. Secondly, in spite of some attempts (Pyrga et al., 2008), the approach is not suitable for estimating expected transfer times, especially if the interchange occurs between low-frequency services, because the *line-based* network representation does not allow the explicit calculation of run-specific service attributes (such as exact arrival and departure times), but only average values relative to the lines (such as headways between consecutive runs). For the same reason, the FB approach is not capable of: considering scheduled penalties with respect to the desired arrival and/or departure time; analysing service synchronisation; evaluating lines with deviation and limitation of specific runs; or calculating loads and performances of each single run of the service.

The latter analysis can be conducted exclusively through an SB approach and may become critical when a major influx of passengers generates overloading only on certain runs that correspond to arrival irregularities or transfers.

A demand peak due to service irregularity is what usually happens when the *bus-bunching* phenomenon is observed. An initial perturbation can produce an increase of the vehicles' dwelling time at a given stop; the delayed run is, therefore, likely to encounter a higher-than-average demand at the following stop, which implies longer boarding and dwelling times. Consequently, the delayed run tends to be more and more delayed, up to the point at which the headway between this run and the prior one is doubled, while the headway between this run and the following one is null. Obviously, the load on the delayed run is

usually remarkably higher than normal; while the flow on the following run is considerably lower than average.

Similar peaks of in-vehicle loading may be observed when a transfer occur between a high-capacity and low-frequency service (such as a train) to a low-capacity and high-frequency service (such as a bus route).

Schedule-based (SB) assignment

SB assignment relies on a run-based supply representation, where both the spatial and the temporal pattern of each vehicle trip are explicitly represented. From a behavioural perspective, this approach is based on the assumption that, when making their travel decision, passengers would distinguish individual runs of the same service and thus time their arrival at the stop or station with the scheduled departure.

In order to consider individual runs explicitly, the ‘*most natural and well established*’ (Papola et al., 2009) supply model for SB transit assignment seems to be the *diachronic graph* (Nuzzolo, 2003), where each run is modelled through a specific *run* sub-graph whose nodes have space and time coordinates according to the run’s schedule. Therefore, the diachronic graph has the advantage of being inherently dynamic, thus having the additional benefit that the dynamic assignment problem reduces to a static assignment on the time-expanded network. On the other hand, when applied to congested multimodal urban networks, this supply model is not suited to the representation of congestion effects on travel times since the graph structure itself must vary with the flow pattern; additionally, it presents shortcomings on the algorithm side because the complexity of the assignment problem increases more than linearly with transit line frequencies, due to the grow of graph dimension, as pointed out in Meschini et al. (2007).

Because the run-based supply representation implies perfect punctuality of each individual run, service irregularity has to be somehow forced into the model, either implicitly by adding a random term to the perceived utility function (Nielsen, O. A., 2004), or explicitly by simulating vehicle runs and dwelling times as interdependent random variables (Nuzzolo et al., 2001; Huang and Peng, 2002).

Advantages of the FB approach are mirrored by disadvantages of the SB approach and vice versa. In fact, SB models are as widely used as FB models and commercial packages based on this modelling framework include VIPS, OMNITrans and VISUM.

2.2. DEMAND-SIDE PHENOMENA

2.2.1. Application of Random Utility Theory for route-choice modelling

Traditionally, the travel choice is modelled assuming that passengers are rational decision makers who choose an alternative within a discrete choice set, with the scope of maximising their own perceived utility or, equally, of minimising their own perceived disutility.

The perceived utility is typically a function of objective attributes related to the LoS (e.g. travel times, fares and transfers) and the socio-economic characteristics of the individual (e.g. income level, gender and age). On the other hand, the modeller does not know with certainty the *perceived* utility that each traveller associates to each alternative, but is only able to define/observe a *systematic* utility that '*represents the mean (expected value) utility perceived by all decision-makers having the same choice context*' (Cascetta, 2009: p. 91).

If i is the considered decision maker, k is the considered alternative and K^i is his/her choice set, then his/her perceived utility (U_k^i) is modelled as a *random* utility and is usually expressed as the sum of the systematic utility (V_k^i) and an error term (ε_k^i), as shown in equations 2-1. The error term typically represents modelling errors in the estimation and/or in the definition of objective attributes, as well as variations in tastes and preferences among different decision-makers and in each of them over time.

$$U_k^i = V_k^i + \varepsilon_k^i \quad 2-1$$

$$V_k^i = E[U_k^i] = E[V_k^i], \quad Var[V_k^i] = 0 \quad 2-1b$$

$$E[\varepsilon_k^i] = 0, \quad Var[\varepsilon_k^i] = Var[U_k^i] \quad 2-1c$$

In the assumption of rational travellers, the single alternative k will only be chosen if $U_k^i \geq U_r^i$ for all alternatives r that belong to the choice set. On the other hand, as U_k^i and U_r^i are *random* utilities, the modeller can only evaluate the probability that each alternative is chosen as:

$$p_{k|K^i}^i = \Pr[V_k^i - V_r^i > \varepsilon_k^i - \varepsilon_r^i, \forall r \neq k, r \in K^i] \quad 2-2$$

and this probability will depend on the distribution of the error terms.

For example, *Multinomial Logit* (MNL) models assume that the error terms are independent and identically Gumbel-distributed, with null average and scale parameter λ , which is directly related to the variance of the error terms. MNL models are the most used discrete choice models in practice mainly because of the mathematical properties of the Gumbel variables that, under some assumptions about the scale parameter, allow for evaluating the choice probability in a closed form (equation 2-3). Examples of the application of an MNL route choice model for transit assignment are given, for example, by Nguyen et

al. (1998), Lam et al. (1999), Lam et al. (2002), Meschini et al. (2007), Papola et al. (2009) and Nuzzolo et al. (2012)

$$p_{k|K} = \frac{\exp\left(\frac{V_k}{\lambda}\right)}{\sum_{r \in K} \exp\left(\frac{V_r}{\lambda}\right)} \quad 2-3$$

The main drawback of MNL route choice models is the assumption that the error terms are independent and identically distributed. This is especially questionable when different routes are highly overlapping (as in the case of *common lines*) and, thus, some form of correlations between error terms would be expected.

In order to overcome this drawback of the MNL, different extensions of the model have been proposed to explicitly capture correlation between alternative routes, for example:

- The C-Logit (Cascetta et al., 1996), which adds a correction term to the systematic utility (*commonality factor*) that is directly proportional to the degree of overlapping of the considered path with other paths in the choice set;
- The Nested Logit (Williams, 1977), which assumes that routes can be grouped in several nests and the error term of each route is made up by a term common to all alternatives in the same nest and a second, alternative-specific, term;
- The Cross-Nested Logit (Vovsha, 1997), which can be seen as a generalisation of the one-level Nested Logit as it assumes that a choice alternative may belong to several groups with different degrees of membership.

Alternatively, it is possible to assume that the error terms of the perceived utility are distributed according to a Multi Variate Normal (MVN) distribution, such that their mean is null and their variance and covariance are fully general (equation 2-4).

$$E[\varepsilon_k^i] = 0, \quad Var[\varepsilon_k^i] = (\sigma_k^i)^2, \quad Cov[\varepsilon_k^i, \varepsilon_h^i] = \sigma_{k,h}^i \quad 2-4$$

This assumption results in the *Probit* model (Daganzo and Sheffi, 1977; Sheffi and Powell, 1981), which has the major advantage of overcoming completely the assumption of independent and identically distributed error terms but does not allow the choice probability to be defined in a closed form and thus is solved by numerical approximations. The most commonly used numerical approximation to calculate the choice probability is known as the *Monte Carlo* simulation and it implies the generation of a sequence of pseudo-random numbers that simulates a sample of perceived utilities. The probability $p_{k|K}$ can thus be calculated as the fraction of times that k is the alternative of highest perceived utility in the simulation.

Although several applications also exist for transit assignment (Nielsen, 2000; Sumalee et al., 2009a; Sumalee et al., 2009b), the Probit model has the major flaw of being cumbersome and time consuming because very large sequences of pseudo-random numbers need to be generated in order to obtain stable values of choice probability.

Another important family of demand models is that of the *deterministic* models, which assume that the error terms are null and that the perceived utility is equal to the (deterministic) systematic utility. In such a setting, all the decision makers select the alternatives of highest utility (lowest disutility) and an alternative has full probability (i.e. $p_{k|K} = 1$) of being chosen if and only if it has the maximum utility (minimum disutility).

While in deterministic models the formulation of the choice problem is extremely simplified, as it reduces to a utility maximisation (or disutility minimisation) problem, they do not usually yield a unique result in terms of the best choice to be taken. Indeed, because in these models the utility is a deterministic variable, if two or more alternatives exist with highest utility (lowest disutility) they can all be chosen. When the travel demand is assigned

to the network, this implies that several path combinations may be optimal and, thus, that the number of passengers who choose each ‘best’ alternative is not uniquely defined. As will be discussed in Section 5.3, the Method of Successive Averages (MSA) can be successfully exploited to solve assignment models with deterministic route choice and load the travel demand on the set of maximum utility (minimum disutility) alternatives for each *od* pair.

It should be acknowledged here that the use of a deterministic demand model entails a number of simplifying assumptions, and it has been argued in the literature (see for example Lam et al., 1999; Lam et al., 2002; Sumalee et al., 2009a; Sumalee et al., 2009b) that travel choices may be more realistically represented through a stochastic demand model. Still, deterministic models present several advantages. First, the flexibility and accuracy of stochastic models usually depend on the accurate calibration and validation of a considerable number of behavioural parameters, while no parameter of this sort is included in deterministic models.

Furthermore, deterministic models are easier to understand from a theoretical point of view and, in general, their results are easier to interpret and analyse. Thus, although not extremely refined, deterministic models are very robust and, if used in a sensitivity analysis to compare different project scenarios, they are more reliable. Indeed, in this case the different results are entirely due to the effects that changes in the supplied LoS produce on route choices and are not due to stochastic perceptions and/or user choices.

Finally, it is important to note here that, when the considered network is very congested, deterministic and stochastic models give very similar results (Cascetta, 2009: p. 329) because a configuration of link flows that is very different from the one induced by a deterministic route choice produces very large differences in the disutilities associated with different paths. Thus, it is most likely that those different disutilities are perceived correctly by the users.

Because of their easy mathematical formulation, deterministic route choice models have been widely applied in the realm of transit assignment (Fearnside and Draper, 1971; Last and Leak, 1976), especially when passenger overcrowding (Poon et al., 2004; Hamdouch et al., 2004; Meschini et al., 2007) and/or travel strategies are considered (Spiess and Florian, 1989; De Cea and Fernandez, 1993; Cominetti and Correa, 2001; Kurauchi et al., 2003; Cepeda et al., 2006; Schmöcker et al., 2008).

2.2.2. Travel strategies

The early approaches to transit assignment, such as those of Dial (1967), Fearnside and Draper (1971) and Last and Leak (1976), tried to extend methods developed for traffic assignment to public transport systems. Therefore, in these works, it is assumed that the route choice process resembles that of a car driver, who selects a single path from the set of all the available alternatives connecting origin to destination.

This assumption is perfectly acceptable when passengers have full information about the transit supply, for example because line timetables are published and the itinerary is chosen on this basis. As such, a large number of SB models are founded on this hypothesis (Tong and Richardson, 1984; Wong and Tong, 1999; Nachtigall, 1995; Nuzzolo et al., 2001; Huang and Peng, 2002; Poon et al., 2004; Zografos and Androutsopoulos, 2008; Papola et al., 2009).

By contrast, in FB assignment it is assumed that travel choices are driven by the knowledge of in-vehicle travel time and service frequency, while the timetable is not explicitly considered. However, if passengers do not synchronise their arrival at a stop with the vehicles' arrivals/departures (because services are very frequent and/or irregular), they

may face uncertainty about whether it is best to board the first vehicle arriving at the stop or keep waiting for one on a faster line that connects to the same destination.

This problem mainly arises in densely connected networks with partially overlapping services (*common lines* – Chriqui and Robillard, 1975) and is due to the inherent uncertainty on the supply side. Since the early eighties, the ‘common-lines dilemma’ has been efficiently solved in FB assignment by modelling the (deterministic) route choice as an *optimal travel strategy* (Spiess, 1983; Spiess and Florian, 1989) or, from a graphic-theory point of view, a *shortest hyperpath* (Nguyen and Pallottino, 1988; Nguyen and Pallottino, 1989), namely a set of potentially optimal itineraries that, considered together, allow passengers to arrive at their destination in the shortest possible time. By contrast, only a few instances are available where the route choice is modelled as a *shortest single path* search (Schmöcker et al., 2002; Meschini et al., 2007).

In the traditional formulation (Spiess, 1983; Spiess, 1984; Nguyen and Pallottino, 1988; Spiess and Florian, 1989), it is assumed that the hyperpath is chosen before the trip begins and that, starting from the origin, it involves the iterative sequence of: walking to a public transport stop or to the destination; then selecting the potentially optimal lines to board (*attractive lines* – Nguyen and Pallottino, 1988) and, for each of them, the stop at which to alight. If the only information available to passengers waiting at a stop is which bus arrives first and two or more attractive lines are available, the best option is to board the first approaching (Spiess, 1983; Spiess, 1984). As clarified by Bouzaïene-Ayari et al. (2001), ‘*the outcome of such a choice is a set of simple itineraries that can diverge, only at bus stops, along the attractive lines.*’

The following example will help to clarify the concept of travel strategy and the effects brought about by the consideration of the shortest hyperpath to destination, rather than the shortest single path. Consider the example network depicted in Figure 2-1, the supply

characteristics listed in Table 2-2 and a passenger who wants to travel from Stop 3 to Stop 4. There are two available alternatives: to board Line 3 or to board Line 4.

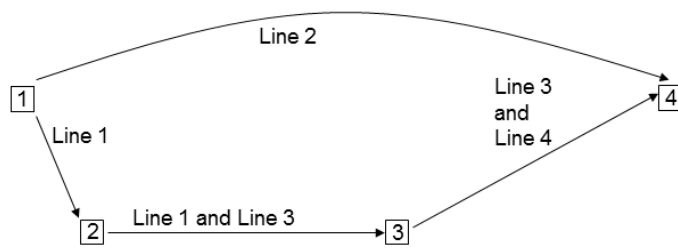


Figure 2-1
Example network

Table 2-2
Example network: frequencies and in-vehicle travel times of Line 3 and Line 4 between Stop 3 and Stop 4

| Line | Arc | Frequency (min ⁻¹) | Travel time (min) |
|------|--------|--------------------------------|-------------------|
| 3 | (3, 4) | 1/15 | 4 |
| 4 | (3, 4) | 1/3 | 10 |

Assuming the services are irregular, with exponentially distributed inter-arrival times, the average waiting time before the first bus of a certain line arrives at the stop is equal to the average headway of the same line. Therefore, considering Line 3 only, the total travel time to destination is $15' + 4' = 19'$. On the other hand, considering Line 4 only, the total time to destination is $3' + 10' = 13'$. The shortest path consists in boarding Line 4 and, on average, the total travel time between stops 3 and 4 accounts for $13'$.

On the other hand, because Line 3, although less frequent than Line 4, is considerably faster, the ideal passenger would be better off boarding the vehicle that arrives first, whether on Line 3 or Line 4, rather than one on Line 4 only. Indeed, if this is done, the expected waiting time at the stop decreases to $2.5'$ and the total expected in-vehicle time to $9'$. Thus the total expected travel time from Stop 3 to Stop 4 is of $11.5'$, a decrease of 23.33% with respect to the value calculated considering the shortest path only.

A detailed explanation of the method applied in order to calculate the total travel time in cases where hyperpaths are considered will be given in Chapter 3.

2.3. SUPPLY-SIDE PHENOMENA

Among all supply-side phenomena that may affect user choice, a leading role is played by: service information and regularity; and passenger congestion and capacity constraints.

2.3.1. Service information and regularity

The application of Intelligent Transport Systems (ITS) has broadened the quantity, quality and frequency of information that passengers can benefit from and, together with service regularity, may have an important impact on the route-choice mechanism of public transport users.

So far, models have mainly concentrated on evaluating the effect on travel strategies of service regularity and information provided at transit stops (*wayside information* – Grotenhuis et al., 2007) in uncongested networks only, where vehicle capacity constraints and queuing are not considered. The main concept is that, if travellers have reliable information on the arrival time of the vehicles – either because there are countdown displays, as in Hickman and Wilson (1995) and Gentile et al. (2005), or the headways between consecutive runs are constant, as in Billi et al. (2004) and Noekel and Wekeck (2008) – they might choose the route ‘intelligently’ and not just select the next arrival from their choice set.

On the other hand, the effect of transport information provided on-board has been less studied (Noekel and Wekeck, 2009).

According to some authors (e.g. Nuzzolo, 2003), as the use of Advanced Traveller Information Systems (ATIS) becomes more widespread, the route-choice mechanism can only be reproduced with SB models. Let us consider a traveller with a handheld navigation device that is capable of showing the scheduled arrival times and in-vehicle travel times of all the available alternatives. If the information is reliable, even if this passenger navigates in a densely connected network, no common-lines dilemma occurs at any transit stop. Indeed, because there is no uncertainty about the supply, the most rational option is not to select a bundle of attractive lines and board one of them depending on in-trip events, but rather to select the shortest single itinerary. However, it is not certain that the majority of passengers would use navigation devices on the transit network; and if they did, it is not certain they would trust the schedule, because transport services such as bus lines are affected by recurrent and non-recurrent road congestion, and thus prone to delays and irregularities that are not captured easily in real time by ATIS.

2.3.2. Passenger congestion and capacity constraints

‘In the context of transit networks, congestion usually refers to the decrease in on-board comfort as the on-board load increases up to a maximum threshold (vehicle capacity), after which users are not allowed to board (oversaturation) and have to wait for the next arriving vehicle’ (Nuzzolo et al., 2012). As such, passenger congestion in transit assignment is not the same as road congestion in traffic assignment since the cost function of public transport does not increase continuously: because transit carriers have a finite capacity, it is a step function.

Additionally, capacity problems are not symmetric, in the sense that they are only experienced by boarders, who may face the formation of queues at stops, where they have to wait for the first run actually available.

Implicit models (Nuzzolo et al., 2012) cannot capture capacity constraints because they simply assume that discomfort is affected by on-board congestion and represent the phenomenon by means of strictly non-decreasing continuous link cost functions with respect to the passenger flow on-board. In this case, all users are affected by congestion in the same way and, thus, capacity constraints are not captured (Spiess, 1983; Wong and Tong, 1999; Nuzzolo et al., 2001; Nuzzolo et al., 2003). On the other hand, *explicit* models (Nuzzolo et al., 2012) differentiate the effect of congestion suffered by those already on-board from that suffered by those waiting to board. The most common approaches used to deal with the problem are summarised in Table 2-3.

Table 2-3
Classification of transit-assignment models with explicit capacity constraints problems

| Approaches | Overview and References |
|----------------------------------|--|
| <i>Effective Frequency</i> | <p>This method is applied only in FB strategy assignment models.</p> <p>The main concept is that the waiting time at stops is a strictly monotone function of passenger flow. The effective frequency, then, is calculated as the inverse of such waiting time.</p> <p>(De Cea and Fernandez, 1993; Cominetti and Correa, 2001; Cepeda et al., 2006; Spiess and Florian, 1989)</p> |
| <i>Fail-to-board Probability</i> | <p>This method is applied to transit networks where the stop layout is such that, if a queue arises, passengers mingle and, therefore, all have the same probability to board or fail-to-board.</p> <p>The main applications of this method are in FB strategies assignment models.</p> <p>(Bell, 2003; Kurauchi et al., 2003; Schmöcker et al., 2008)</p> |

| Approaches | Overview and References |
|-------------------------------|---|
| <i>Ordered Preferences</i> | <p>This method is applied only in SB assignment models.</p> <p>The main idea is that, although passengers know the service timetable, and this is reliable, it is uncertain if they will be able to board and/or sit on the next coming run. This uncertainty leads travellers to build ranked choice sets of alternative runs and take the first which is actually available to them.</p> <p>(Hamdouch et al., 2004; Hamdouch and Lawphongpanich, 2008; Hamdouch et al., 2011)</p> |
| <i>Adaptive Routing</i> | <p>This method is applied both in SB and in FB assignment.</p> <p>The idea is that passengers would choose a specific itinerary or hyperpath. However, once at the stop, if the congestion level on the line(s) of their choice is such that they cannot board the first vehicle, they might re-route and consider also different lines.</p> <p>(Leurent and Benezech, 2011; Nuzzolo et al., 2012)</p> |
| <i>Residual Capacity</i> | <p>This method is applied in FB strategy assignment, where it is assumed that – in cases of overcrowding – FIFO passenger queues would arise. The passengers' split among different attractive lines is assumed to be a function not of the waiting time, as is usually assumed in strategy-based route choice, but of the residual capacity, while the waiting time before boarding is calculated using a bulk queue approximation.</p> <p>(Gendreau, 1984; Bouzaïene-Ayari, 1988)</p> |
| <i>Bottleneck Queue Model</i> | <p>This method is applied both in SB and in FB assignment but, so far, practical formulations have been developed only without considering travel strategies.</p> <p>The main idea is to calculate the time necessary for the last passenger in the queue to reach the front. This time, queuing time, increases the waiting time that would be normally experienced in the absence of passenger congestion.</p> <p>(Poon et al., 2004; Meschini et al., 2007; Papola et al., 2009)</p> |

2.4. DISCUSSION

The brief analysis of the research background given in this chapter highlights the key phenomena to consider when building the mathematical framework of an assignment model for public transport systems.

On the one hand, there is the problem of correctly modelling users' perceptions. To what extent are these perceptions distorted by personal tastes or other sources of errors? Is it possible to assume that passengers are rational decision makers who try to minimise the travel time to destination (or, at least, its expected value)?

Different answers are given by deterministic and stochastic demand models. The first family of models make use of Wardrop's first principle, according to which each traveller knows exactly the travel time he/she will encounter and selects the minimal route. The second family of models imply that errors and uncertainties are attached to the evaluation of the travel time of different alternatives and, thus, only alternatives with minimum *perceived* travel time are actually chosen by the travellers.

Additionally, another important demand-side phenomenon to consider is the correct way of modelling passenger route choices in networks with several (partially) overlapping alternatives. Would users select the single best itinerary to destination? Or would they face the so-called dilemma of 'common lines'? How should the waiting and transfer time be accounted for when common lines are available?

On the other hand, there is also the problem of the best method for reproducing supply-side phenomena, such as the effects of information, service reliability and overcrowding. Consider the case of a completely reliable transit service with full travel information (e.g. in-vehicle travel times and scheduled arrivals and departures). Would

passengers make their choices by considering runs? Or would they have a line-based perception of the supply? Moreover, consider highly congested networks. What is the effect of overcrowding on waiting times at stops? How is it represented? Is it possible to estimate precise vehicle loads for each run, or is it only possible to estimate average values? If wayside information is available – for example, by means of countdown displays – is there a more even spread of passenger flows across the network? If the network performances are subject to some kind of stochasticity (especially waiting times), what is the effect on route choices?

The answers given to these questions are numerous and highly varied, and an exhaustive review and classification of the existing literature is difficult to achieve. However, it is possible to distinguish two different families of models, depending on the modelling approach adopted.

Schedule-based assignment models make use of a run-based supply representation and implicitly assume that passengers would distinguish each run of the same line when making their travel choices. Because SB models enable the precise calculation of waiting and transfer times, they can easily handle the problem of representing the effects of capacity constraints on route choices. On the other hand, because they imply perfect punctuality of each individual run, the effect of service irregularity has to be included in the model, either implicitly by adding a random term to the perceived utility function (Nielsen, O. A., 2004), or explicitly by simulating vehicle runs and dwelling times as interdependent random variables (Nuzzolo et al., 2001; Huang and Peng, 2002).

Frequency-based assignment models make use of a line-based supply representation and assume that, because of high frequency and low regularity of services, passengers would make their travel choices considering *average* characteristics of the service, such as frequency. This family of models has the advantage of more naturally representing the route

choice mechanism in densely connected urban networks. On the other hand, there does not seem to be a broad agreement in the literature on how to deal with congestion and capacity constraints in a fully dynamic setting, especially when these lead to FIFO queues of passengers at the stops.

The present research applies to transit networks with overlapping and highly frequent and/or irregular services, where passengers would not perceive any advantage in considering and respecting the lines' schedule (if any is published). Consequently, an FB assignment with travel strategies is developed, where the effect of supplied uncertainty (especially in terms of waiting times) is dealt with implicitly by assuming, as usual, that passengers choose the alternative with minimum expected travel time. Additionally, the model explicitly considers supply capacity constraints due to overcrowding (in the form of FIFO queues of passengers at stops) and represents the effects of such phenomena on the route choice.

In order to attain this result, an innovative mathematical framework is developed to model travel demand in a dynamic context. Moreover, in order to embed the proposed demand model in a dynamic assignment model for transit networks, the supply model is obtained by extending to strategy-based assignment an existing Network Flow Propagation Model and a Bottleneck Queue Model, originally deployed for dynamic assignment without hyperpaths.

Details of those main methodological contributions are given in the next two chapters.

3. DEMAND MODEL FOR STRATEGY-BASED TRANSIT ASSIGNMENT WITH CAPACITY CONSTRAINTS

3.1. INTRODUCTION

Demand models used in dynamic assignment express the time-dependent relationship between path flows and generalised travel times in terms of route choices (Cascetta, 2001: p. 398). The process of route choice in public transport differs significantly from that in private car travel due to the character of transport *supply*. A parallel can be drawn between the capacity of carriers in public transport and available road capacity in car travel. However, the major difference is that access to that capacity in public transport networks is restricted to specific locations and strictly determined by the schedule and/or frequency of services.

As such, passenger behaviour at stops is the key aspect of modelling demand phenomena in transit assignment especially in FB transit assignment, where it is accepted that passengers might be willing to board more than one line from the same stop (strategy-based assignment). In this case, the study of the Stop Model allows for ‘estimating the passenger distribution among attractive lines and the expected waiting time at bus stops’ (Bouzaïene-Ayari et al., 2001).

Many solutions have been proposed in the literature that either disregard congestion or consider its effects on passenger distribution among attractive lines (*diversion probabilities* – Cantarella, 1997), on expected waiting time or on both of these variables. This

chapter gives an up-to-date critical review of the most representative Stop Models (SM) and Route Choice Models (RCM) developed for hyperpath-based demand modelling in transit assignment, with particular attention given to the results attained in the congested case where it is assumed that passengers may be unable to board the first carrier of their attractive set, due to overcrowding. Moreover, a demand model is presented which includes a completely new SM, as well as an RCM formulated as a dynamic hyperpath search.

The remainder of this chapter is organised as follows. Section 3.2 introduces the original formulations of the SM and RCM for static networks without capacity constraints. Section 3.3 explains the implications of developing the SM and RCM in networks affected by passenger congestion. Finally, Section 3.4 presents the new SM and RCM, which are key elements of the transit assignment proposed in the present work.

Before proceeding to the review of existing models and analysis of the new one, a general network representation, with basic notation, is provided in the following two subsections and will be used to describe and compare the demand models considered in this chapter.

3.1.1. Network representation

$\lambda_e \in \mathfrak{R}_+$: *edge length* (where \mathfrak{R}_+ is the set of non-negative real numbers);

$\rho_e \in \mathfrak{R}_+$: *pedestrian speed* – if $\rho_e = 0$, this means that a connection is unavailable;

$L \subset \mathfrak{N}$: set of lines included in the transit network;

$\ell \in L$: generic line;

$R_\ell \subseteq V$: route of line ℓ ; an ordered sequence of $\sigma_\ell \in \mathbb{N}$ (not repeated) vertices, each of which is denoted as $R_{\ell,i} \in V$ with $i \in [1, \sigma_\ell]$;

$\mu_{\ell,i} \in (0, 1)$: function expressing whether or not a *stop* is made at the i -th vertex along the route of line $\ell \in L$, with $i \in [1, \sigma_\ell]$;

$\mu_i \in (0, 1)$: function expressing if the i -th vertex corresponds to a *stop*;

$\chi_\ell \in \mathbb{R}_+$: the *vehicle capacity* of line ℓ ;

$\varphi_\ell \in \mathbb{R}_+$: *base frequency* – instantaneous flow of departures from the origin terminal $R_{\ell,1}$ at time $\tau \in \mathbb{R}$;

$\theta_{\ell,i}(\tau) \in \mathbb{R}_+$: *line time* – the time when a carrier of line $\ell \in L$, departed from $R_{\ell,1}$ at time $\tau \in \mathbb{R}$, reaches the i -th vertex along its route, with $i \in [1, \sigma_\ell]$.

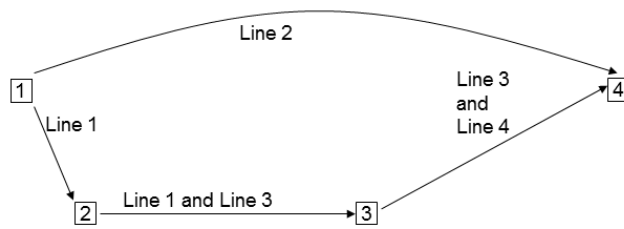


Figure 3-1
Base graph representation of a small network

The topology of the network, including the line routes and the pedestrian network, is described through a directed *base graph* $B = (V, E)$, where $V \subset \mathbb{N}$ is the set of *vertices* (\mathbb{N} is the set of positive integer numbers) and $E \subseteq V \times V$ is the set of *edges* (Figure 3-1).

The generic edge $e \in E$ is univocally identified by its initial vertex $TL_e \in V$, or *tail*, and its final vertex $HD_e \in V$, or *head*; that is: $e = (TL_e, HD_e)$. The generic vertex $v \in V$ is associated with a location in space and is thus characterised by geographic coordinates, while the generic edge $e \in E$ is characterised by λ_e and ρ_e .

The topology of each line $\ell \in L$ is defined by its *route* R_ℓ . The generic section of a route is referred to as $(R_{\ell,i-1}, R_{\ell,i}) \in E$, with $i \in [2, \sigma_\ell]$, and corresponds to an edge of the base graph. For any given vertex $v \in V$ and line $\ell \in L$, the function $s(v, \ell) \in [0, \sigma_\ell]$ yields, if it exists, the index such that $R_{\ell,s(v, \ell)} = v$, and 0 otherwise.

The physical topology of the transit network represented by B is insufficiently detailed for modelling purposes. Indeed, it only allows the representation of movements (on-board a vehicle or on foot) across the network and lacks graphical entities that represent other actions, such as waiting at a stop, boarding, alighting or staying on-board while the vehicle dwells at the stop. As such, a *hypergraph* $H = (N, F)$ is introduced, where N is the set of *nodes* and $F \subseteq N \times N$ is the set of *forward hyperarcs* (Gallo et al., 1993), henceforth simply referred to as *hyperarcs*, included in hypergraph H .

The hypergraph is built from the base graph, which is usually organised in a GIS database, considering the transit-line routes, with their travel times, and the pedestrian speeds. Each node $i \in N$ is indeed the triplet of a vertex $V_i \in V$, a *type* $T_i \in \{P, S, B, A, W\}$ and a line $L_i \in L \cup 0$: $N \subset (V \times \{P, S, B, A, W\} \times L \cup 0)$.

Specifically, the node set and hyperarc sets are constructed as the union of the following subsets:

$$N = N^P \cup N^S \cup N^B \cup N^A \cup N^W;$$

$$F = A^P \cup A^L \cup A^D \cup A^Z \cup A^A \cup A^H \cup A^B;$$

$$N^P: \text{pedestrian nodes, } N^P = \{(v, P, 0) : v \in V\};$$

$$N^C: \text{centroid nodes, } N^C = \{(v, C, 0) : v \in V\} \text{ and it is also assumed } N^C \subseteq N^P;$$

$$N^S: \text{stop nodes, } N^S = \{(R_{\ell,i}, S, 0) : \ell \in L, i \in [1, \sigma_\ell - 1], \mu_{\ell,i} > 0\};$$

$$N^B: \text{boarding nodes, } N^B = \{(R_{\ell,i}, B, \ell) : \ell \in L, i \in [1, \sigma_\ell - 1]\};$$

N^A : *alighting nodes*, $N^A = \{(R_{\ell,i}, A, \ell) : \ell \in L, i \in [2, \sigma_\ell]\}$;

N^W : *waiting nodes*, $N^W = \{(R_{\ell,i}, W, \ell) : \ell \in L, i \in [1, \sigma_\ell - 1], \mu_{\ell,i} > 0\}$;

A^P : *pedestrian arcs*, which represent walking time:

$A^P = \{(i, j) : i \in N^P, j \in N^P, e = (V_i, V_j) \in E, \rho_e > 0\}$;

A^L : *line arcs*, which represent in-vehicle travel time:

$A^L = \{(i, j) : i \in N^B, j \in N^A, V_i \equiv R_{\ell,k}, V_j \equiv R_{\ell,k+1}, \ell \in L, i \in [1, \sigma_\ell - 1]\}$;

A^D : *dwelling arcs*, representing the time spent by a bus at a stop while passengers

alight/board: $A^D = \{(i, j) : i \in N^A, j \in N^B, V_i \equiv R_{\ell,k}, V_j \equiv R_{\ell,k}, \ell \in L, i \in [2, \sigma_\ell - 1]\}$;

A^Z : *dummy arcs*, which connect the line stops and the pedestrian network: $F^Z = \{(i, j) : i \in N^P,$

$j \in N^S, V_i \equiv V_j\}$

(dummy arcs are introduced for algorithmic purposes to identify more easily (hyper)arcs

representing the waiting process);

A^A : *alighting arcs*, which represent the time passengers need to disembark:

$A^A = \{(i, j) : i \in N^A, j \in N^P, V_i \equiv V_j\}$;

A^B : *boarding arcs*, representing the time passengers need to board a vehicle:

$A^B = \{(i, j) : i \in N^W, j \in N^B, V_i \equiv V_j\}$;

A^H : *waiting hyperarcs* (Billi et al., 2004), which represent the total expected waiting time for

a specific set of *attractive lines* serving a stop: $A^H \subseteq \{(i, j) : i \in N^S, J \subseteq N^W, j \in J, V_i \equiv V_j\}$;

FS_i : *forward star* of node $i \in N \setminus \{N^S\}$, i.e. the set of arcs sharing the same tail i : $FS_i = \{(i, j) :$

$TL_a = i\}$;

BS_j : *backward star* of node $j \in N$, i.e. the set of arcs sharing the same head j : $BS_j = \{(i, j) : HD_a$

$= j\}$;

HFS_i : the *hyper-forward star* of node $i \in N^S$, i.e. the set of hyperarcs sharing the same stop tail i : $HFS_i = \{h \in A^H : TL_h = i\}$.

The generic hyperarc $h \in F$ is univocally identified by a single initial node $TL_h \in N$, or *tail*, and a set of final nodes $HD_h \subset N$, or *head*; that is: $h = (TL_h, HD_h)$. The cardinality of the hyperarc (Nielsen, L. R., 2004), i.e. the number of single nodes included in its head, is notated as $|HD_h|$ and it is assumed that $|HD_h| \geq 1$ only for hyperarcs whose tail is a stop node (i.e. waiting hyperarcs), while in all other cases $|HD_h| = 1$. For reasons of clarity and simplicity, all the hyperarcs for which $|HD_h| = 1$ are referred to as *arcs*, while only those for which $|HD_h| \geq 1$ are referred to as *hyperarcs*. Moreover, for the same reasons, a distinction is made between the forward star of a node $i \in N \setminus \{N^S\}$ and the hyper-forward star of a node $i \in N^S$.

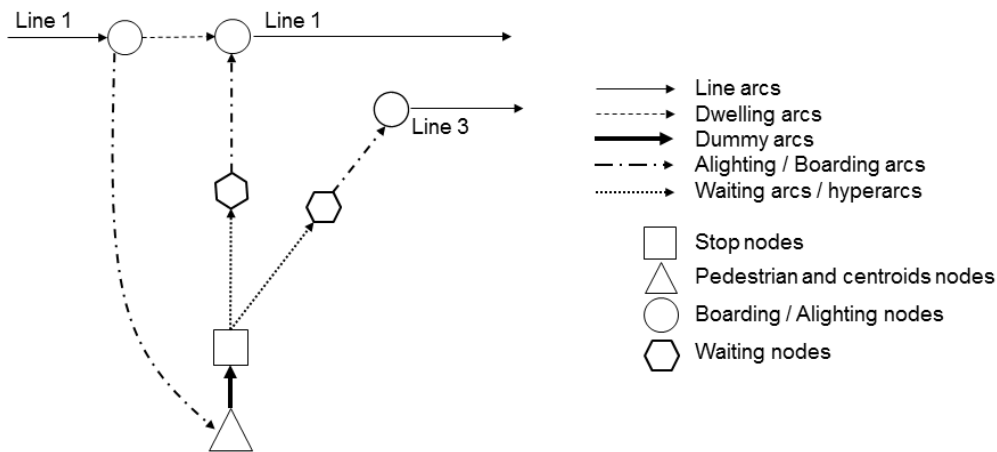


Figure 3-2
Hypergraph representation of a portion of the Stop 2 depicted in Figure 3-1

Because each waiting hyperarc $h \in A^H$ is univocally identified by a singleton *tail* $TL_h \in N^S$ and by a set *head* $HD_h \subseteq N^W$, it can also be indicated as $h = \{(TL_h, j) : j \in HD_h\}$. Therefore, the waiting hyperarc can be seen as a set of ‘branches’, or simple waiting arcs a ,

each of which has the same tail node of h ($TL_a = TL_h$) and a head node belonging to the head set of h ($HD_a \in HD_h$). Moreover, the head node of a branch of a hyperarc h ($a \in h$) is associated with one particular line (L_{HDa}) among those who share the stop represented by $TL_a = TL_h$.

For example, the hyper-forward star of the stop node depicted in Figure 3-2 includes a null hyperarc (hyperarc with no branches) and the three hyperarcs shown in Figure 3-3:

- *hyperarc 1* = $\{a', a''\}$
- *hyperarc 2* = $\{a''\}$
- *hyperarc 3* = $\{a'\}$

where $L_{HDa'} = \text{Line 1}$ and $L_{HDa''} = \text{Line 3}$.

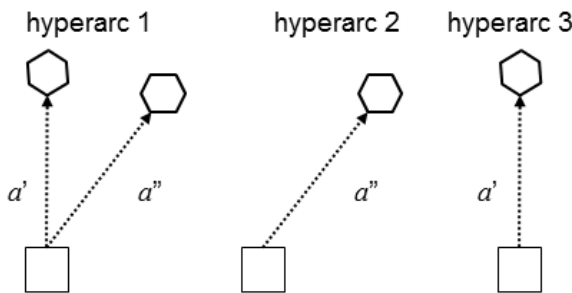


Figure 3-3
Hyperarcs belonging to the hyper-forward star of the stop node depicted in Figure 3-2

3.1.2. Demand models: basic nomenclature

With reference to the generic $a \in h$ and $h \in A^H$, the following variables are defined:

$\varphi_a(\tau)$: *instantaneous frequency* (instantaneous flow of carriers) of the line L_{HDa} evaluated at the vertex of the base graph corresponding to TL_a (V_{TLa}).

The instantaneous frequency can be an external input, or it can be calculated by

propagating in time the base frequency. Since the frequency is regarded here as a continuous flow of carriers with instantaneous capacity, its propagation in time can be derived by applying the FIFO and conservation rules (Cascetta, 2009: p. 437).

More specifically, if $\varphi_\ell(\tau) \cdot \partial\tau$ is the number of carriers that leaves the origin terminal in $\partial\tau$ and $\theta_{\ell,s}(\tau)$ is the arrival time at the s^{th} vertex of route R_ℓ , then under the assumption of stationarity:

$$\varphi_\ell(\tau) \cdot \partial\tau = \varphi_{\ell'}(\theta_{\ell',s}^{-1}(\tau)) \cdot \partial\theta_{\ell',s}^{-1}(\tau) \quad 3-1$$

Or equivalently:

$$\varphi_a(\tau) = \varphi_{\ell'}(\theta_{\ell',s}^{-1}(\tau)) \cdot \left[\frac{\partial\theta_{\ell',s}^{-1}(\tau)}{\partial\tau} \right] \quad 3-1b$$

where $s = s(L_{HDa}, V_{HDa})$;

$\kappa_a(\tau)$: *congestion parameter*, expressed as the total number of vehicle arrivals that passengers, reaching the stop of vertex V_{TLa} at time τ , are unable to board before they board the line L_{HDa} ;

$w_h(\tau)$: *expected waiting time* for passengers reaching the stop of vertex V_{TL_h} at time τ and considering the set of attractive lines represented by $h \in A^H$;

$w_{a|h}(\tau)$: *conditional waiting time* before boarding the line L_{HDa} associated with $a \in h$ for passengers reaching the stop of vertex V_{TLa} at time τ ; its value depends on the set of attractive lines considered, which is represented by $h \in A^H$;

$t_{a|h}(\tau)$: *conditional boarding time* on the line L_{HDa} for passengers reaching the stop of vertex V_{TLa} at time τ – namely $t_{a|h}(\tau) = \tau + w_{a|h}(\tau)$, and its value depends on the set of attractive lines considered, which is represented by $h \in A^H$;

$p_{a|h}(\tau)$: *diversion probability* (Cantarella, 1997): ratio of passengers that board line L_{HDa} to those who reach the stop of vertex V_{TLa} at time τ and whose set of attractive lines is represented by $h \in A^H$.

Moreover, with reference to the generic $a \in F \setminus \{A^H\}$ and $i \in N$, the following variables are also defined:

$c_a(\tau)$: *travel time* of arc a for users entering it at time τ ;

$t_a(\tau)$: *exit time* from arc a for users entering it at time τ – namely, $t_a(\tau) = \tau + c_a(\tau)$;

$t_a^{-1}(\tau)$: *entry time* to the arc a for users exiting it at time τ ;

$g_{i,d}(\tau)$: *total travel time* from node i to destination $d \in N^C$ at time τ ;

$g_{i,d}^*(\tau)$: *minimum total travel time* from node i to destination $d \in N^C$ at time τ .

3.2. DEMAND MODEL FOR STATIC STRATEGY-BASED ASSIGNMENT WITHOUT CONGESTION EFFECTS

3.2.1. Stop Model: original formulation¹

In their seminal work on travel strategies, Nguyen and Pallottino (1988) and Spiess and Florian (1989) prove that in FB networks with *common lines*, i.e. competing lines with partially overlapping itineraries (Chriqui and Robillard, 1975), passengers can minimise their

¹ The notation used with reference to static models disregards time dependency of travel variables.

total travel time to destination by selecting, before the beginning of the trip, an optimal strategy that involves the iterative sequence of: walking to a transit stop or to the destination; then selecting the *attractive set* of lines to board and, for each of them, the stop at which to alight. The optimality of the strategy stems from the choice of the attractive set at each stop, namely the group of lines that, considered together, minimise the total travel time from the current stop to destination.

Billi et al. (2004) and Noekel and Wekeck (2007) summarise the conditions under which such strategic behaviour is considered rational:

1. Passengers arrive at stops randomly, at a constant rate, independently of carriers' arrivals;
2. Carriers' arrivals of different lines are not synchronised and, for each line, follow a Poisson distribution, with parameter the frequency of the line;
3. No information is provided at the stop on actual waiting times and on the available capacities of arriving carriers;
4. Passengers always board the first-arriving carrier of their attractive set;
5. There are no capacity constraints and travellers are always able to board the first attractive line approaching the stop.

In this case the SM is extremely simple and leads to the well-known equations 3-3 and 3-7 (Nguyen and Pallottino, 1988; Nguyen and Pallottino, 1989; Spiess and Florian, 1989), as proved in the following.

Assume that conditions (4) and (5) hold true. Moreover, assume that the attractive set at the considered stop is known and graphically represented by hyperarc h . In this case, the diversion probability on a specific line is equivalent to the probability that this is the first line to appear at the stop, and is expressed as:

$$p_{a|h} = \int_0^{\infty} \left[\text{PDF}_a(w) \cdot \prod_{a' \in h \setminus \{a\}} \overline{\text{CDF}}_{a'}(w) \right] dw \quad 3-2$$

where $\text{PDF}_a(w)$ is the probability distribution function of the waiting time for the first arrival of line L_{HDa} and $\overline{\text{CDF}}_{a'}(w)$ is the complement of the cumulative distribution function (or *survival function*) of the waiting time for line $L_{HDa'}$.

When the aforementioned assumptions (1) to (3) are made, given the properties of Poisson and Uniform PDFs, vehicle inter-arrival times as well as passengers' waiting times have an Exponential distribution with mean equal to $1/\phi_a$. Consequently, the diversion probability is given by equation 3-3:

$$p_{a|h} = \frac{\phi_a}{\sum_{a' \in h} \phi_{a'}} \quad 3-3$$

A well-established result of Statistics is that, for a stochastic variable x , $\text{PDF}(x) = -\overline{\text{CDF}}'(x)$, the expected value ($E[x]$) can always be expressed as $\int_0^{\infty} x \cdot [-\overline{\text{CDF}}'(x)] dx$. Integrating by parts

the latter formula, the following result is obtained:

$$E[x] = -\left[x \cdot \overline{\text{CDF}}(x) \right]_0^{\infty} + \int_0^{\infty} \overline{\text{CDF}}(x) dx = \int_0^{\infty} \overline{\text{CDF}}(x) dx \quad 3-4$$

Therefore, if the considered stochastic variable indicates the waiting time before the first bus of a certain line arrives at the stop and $\text{PDF}(x) = -\overline{\text{CDF}}'(x)$ is its density distribution, then equation 3-4 is the average time to wait before observing the event 'bus arrival'.

Similarly, if passengers consider boarding two or more attractive lines from the same stop, then the same formula can be applied to calculate the total waiting time at the stop,

where $\overline{\text{CDF}}(x)$ is substituted with the joint probability that an attractive line has not shown up until time w $\overline{\text{CDF}}_h(w)$. Because vehicle arrivals of different lines are stochastically independent, this probability can be also expressed as:

$$\overline{\text{CDF}}_h(w) = \prod_{a \in h} \overline{\text{CDF}}_a(w) \quad 3-5$$

Thus, the following equation is obtained:

$$w_h = \int_0^{\infty} \prod_{a \in h} \overline{\text{CDF}}_a(w) dw \quad 3-6$$

In case of exponential PDF, equation 3-6 simply becomes:

$$w_h = \frac{1}{\sum_{a' \in h} \varphi_{a'}} \quad 3-7$$

3.2.2. Route Choice Model: original formulation

The alternatives considered in the RCM are strategies, or hyperpaths, and are defined as follows: a *hyperpath* k connecting origin $o \in N^C$ to destination $d \in N^C$ is a sub-hypergraph $H_{k,o,d} = (N_k, A_k)$ of H , where $N_k \subset N$, $A_k \subset A$, such that:

- $H_{k,o,d}$ is acyclic;
- o has no predecessors and one successor arc;
- d has no successors and at least one predecessor arc;
- For every node $i \in N_k \setminus \{o, d\}$ there is at most one immediate successor arc if $i \notin N^S$, otherwise the successor is a hyperarc with cardinality equal or greater than one;

- For each hyperarc $h \in H_{k,o,d}$ a characteristic vector $\mathbf{p} = (p_{a|h})$ is defined where \mathbf{p} is a real value vector of dimension $(|HD_h| \times 1)$ such that:

$$\begin{aligned} \sum_{a \in h} p_{a|h} &= 1 \\ p_{a|h} &> 0 \end{aligned} \quad 3-8$$

The total travel time of the generic hyperpath $H_{k,o,d}$ can be computed by explicitly taking into account all the elemental paths l forming it (Nguyen and Pallottino 1988; Nguyen and Pallottino 1989). Therefore, if Q_k is the set of such paths, λ_l is the probability of choosing the elemental path l , and n_l is its travel time, then the travel time of hyperpath $H_{k,o,d}$ is:

$$g_k = \sum_{l \in Q_k} \lambda_l \cdot n_l \quad 3-9$$

On the other hand, n_l can be expressed as the sum of travel and waiting times on the path's arcs and nodes:

$$n_l = \sum_{a \in F_k \setminus \{F^H\}} c_a \cdot \delta_{al} + \sum_{\substack{h \in F_k \cap F^H \\ i = TL_h}} w_h \cdot \delta'_{il} \quad 3-10$$

where $\delta_{al} = 1$ if arc a belongs to path l and $\delta_{al} = 0$ otherwise; and $\delta'_{il} = 1$ if path l traverses node i , otherwise $\delta'_{il} = 0$. Thus the following expression of the hyperpath's total travel time can be obtained:

$$g_{k,o,d} = \sum_{l \in Q_k} \lambda_l \cdot \left[\sum_{a \in F_k \setminus \{F^H\}} c_a \cdot \delta_{al} + \sum_{\substack{h \in F_k \cap F^H \\ i = TL_h}} w_h \cdot \delta'_{il} \right] \quad 3-11$$

And the RCM can be formulated as:

$$g_{o,d}^* = \min \{ g_{k,o,d} : k \in H_{o,d} \} \quad 3-12$$

Where $H_{o,d}$ is the sub-hypergraph containing all the hyperpaths connecting the same od pair.

Definition:

A shortest hyperpath $k \in H_{k,o,d}$ is said to satisfy the *concatenation property* if the two sub-hyperpaths of k from o to any intermediate node i and from i to d are themselves shortest hyperpaths.

In the static case, the principle defined above always holds true. Thus, in order to avoid explicit path enumeration, Nguyen et al. (1998) and Nguyen and Pallottino (1989) propose to solve equation 3-12 by applying a local recursive formula (formally known as the *generalised Bellman equation*) that sequentially defines the shortest hyperpath from each intermediate node to destination as well as its travel cost.

$$g_{k,i,d} = \begin{cases} 0, & \text{if } i = d \\ c_a + g_{HD_{a,d}} : a \in FS_i, & \text{if } i \notin N^S \\ w_h + \sum_{a \in h} p_{a|h} \cdot g_{HD_{a,d}} : h \in HFS_i, & \text{if } i \in N^S \end{cases} \quad 3-13$$

Consequently, the optimality of the travel strategy stems from the correct definition of the attractive set, or equivalently the waiting hyperarc that represents it, at each intermediate stop. More specifically, the waiting hyperarc representing the attractive set must satisfy the following condition (Nguyen and Pallottino, 1988; Nguyen and Pallottino, 1989):

$$\begin{aligned} & \exists h \in HFS_i : \\ & w_h + \sum_{a \in h} p_{a|h} \cdot g_{HD_{a,d}} = \min \left\{ w_{h'} + \sum_{a \in h'} p_{a|h'} \cdot g_{HD_{a,d}} : h' \in HFS_i \right\} \end{aligned} \quad 3-14$$

Equation 3-14 represents a combinatorial problem as it requires the computation of $g_{k',i,d}$ for all the possible hyperarcs $h' \in HFS_i$. However, at least for the uncongested static case where

the waiting times are exponentially distributed, it is counter-intuitive to exclude a line from h if it has a shorter remaining travel time upon boarding than any other attractive service. Therefore, it is possible to solve the above combinatorial problem through a greedy approach (Chriqui and Robillard, 1975; Nguyen and Pallottino, 1988; Spiess and Florian, 1989). Namely, the lines are processed in ascending order of their travel time upon boarding and the progressive calculation of the values of $p_{a|h}$, w_h and $g_{k,i,d}$ is stopped as soon as the addition of the next line increases the value of $g_{k,i,d}$. At this point, the cost from the stop node to destination is minimal ($g_{i,d}^*$) and the hyperarc h corresponds to the attractive set.

This formulation of the RCM remains always valid in the static context. However, the SM needs to be expanded in order to consider relevant supply-side phenomena, such as the availability of wayside information and service regularity. Two important SM extensions, for static assignment to uncongested networks, are detailed in the following sub-sections.

3.2.3. Stop Model extension: wayside information

When information about actual waiting times is made available at the stop, the travel behaviour hypothesised in the original formulation of the SM is not rational. Travellers would use the information provided in order to minimise their expected total travel time to destination. Therefore, it is more sensible to assume that, when a carrier approaches the stop, a waiting passenger does not board it simply because it is the first attractive line arriving but would compare its expected travel time to destination upon boarding with the expected total travel time of later arrivals.

Given this assumption, Hickman and Wilson (1995) as well as Gentile et al. (2005) propose that the probability of boarding line L_{HDa} is equal to the probability that it is the line with the best total time (waiting at the stop + travel time upon boarding). Therefore:

$$P_{a|h} = \int_0^{\infty} \left[\text{PDF}_a(w) \cdot \prod_{a' \in h \setminus \{a\}} \text{Prob}(w_{a'} \geq w + g_{HDa,d} - g_{HDa',d}) \right] dw \quad 3-15$$

Or, equivalently:

$$P_{a|h} = \int_0^{\infty} \left[\text{PDF}_a(w) \cdot \prod_{a' \in h \setminus \{a\}} \overline{\text{CDF}}_{a'}(w + g_{HDa,d} - g_{HDa',d}) \right] dw \quad 3-15b$$

While the expected waiting time may be calculated as:

$$w_h = \sum_{a \in h} \left[\int_0^{\infty} w \cdot \text{PDF}_a(w) \cdot \prod_{a' \in h \setminus \{a\}} \overline{\text{CDF}}_{a'}(w + g_{HDa,d} - g_{HDa',d}) dw \right] \quad 3-16$$

3.2.4. Stop Model extension: service regularity

Assumption (2) of the SM is generally supported by empirical evidence for bus services with an average headway equal or inferior to 12 minutes (O'Flaherty and Mangan, 1970; Seddon and Day, 1974). However, less-frequent services and other transport modes, such as light and underground railways, tend to be regular or, at least, more regular.

In highly connected networks with very frequent services, it may always be possible to assume that travellers do not consider timetables explicitly and make their travel choices according to an FB paradigm. On the other hand, service regularity implies that, although the general formulas given for calculating diversion probabilities (3-2) and expected waiting time (3-6) remain valid, waiting times are not exponentially distributed.

Different authors (Gendreau, 1984; Bouzaïene-Ayari et al., 2001; Gentile et al., 2005) recognise that the Erlang distribution is more flexible because, by having a shape parameter that can be changed, it allows the description of inter-arrival times for both completely regular services (i.e. constant headways) and completely irregular services (i.e. headways with exponential distribution), as well as for services with an intermediate level of regularity. On the other hand, it lacks the mathematical properties that allow the easy modelling of congestion effects on waiting time and diversion probabilities if exponential distributions are considered (see Section 3.4).

Finally, it should be noted that if services with constant headways are available, passengers may make use of the knowledge of service regularity and elapsed waiting time to revise their estimate of remaining waiting time, and hence remaining travel time, as they wait (Billi et al., 2004; Noekel and Wekeck, 2007).

Therefore, before starting their trip, users have already defined the choice set and, for each line, the waiting period in which the line is considered attractive. Once they reach the stop, their *dynamic attractive set* varies with the time spent waiting in vain.

The following example helps to clarify this concept. Consider Stop 3 of Figure 3-1 and assume that Line 3 arrives every 10 minutes with constant headways and has a travel time upon boarding of 30 minutes, while Line 4 arrives every 20 minutes with constant headways and has a travel time upon boarding of 15 minutes. As soon as the user reaches the stop, the expected waiting time for Line 3 is 5 minutes and the total expected travel time is 35 minutes, while the expected waiting time for Line 4 is 10 minutes and the total expected travel time is 35 minutes.

However, after 5 minutes of waiting in vain, the expectation of remaining waiting time for Line 3 decreases to $(10 - 5) / 2 = 2.5$ minutes, while for Line 4 it becomes $(20 - 5) / 2$

= 7.5 minutes. At this point, the total travel time for Line 4 accounts for 22.5 minutes, which is less than the travel time upon boarding Line 3. As a result, Line 3 is excluded from the dynamic attractive set.

3.3. DEMAND MODEL FOR STRATEGY-BASED ASSIGNMENT WITH CONGESTION EFFECTS

The SMs and the hyperpath-based RCM reviewed in the previous section disregard congestion effects on passenger distribution among attractive lines as well as on expected waiting time. However, because recurrent passenger congestion is one of the major problems faced by public transport in large cities, in the last three decades several models for FB strategy assignment have been proposed to overcome this flaw (De Cea and Fernandez, 1993; Cominetti and Correa, 2001; Bouzaïene-Ayari et al., 2001; Cepeda et al., 2006; Schmöcker et al., 2008).

In general, when passenger congestion occurs, the queuing mechanism followed by travellers is determined by the stop layout. For example, for stations and stops with long platforms, it is correct to assume that passengers mingle, which implies that no waiting priority is respected. Thus, in cases of oversaturation, a passenger who reaches the stop last may be lucky and board the first approaching vehicle, while those who arrived earlier may be unlucky and continue waiting. In general, a common modelling assumption is that all passengers waiting along the platform have the same probability of boarding the next approaching vehicle provided it is attractive.

On the other hand, it may also happen that FIFO queues arise at the stop. In this case, the calculation of diversion probabilities and waiting time needs to consider the priority of those who are at the front of the queue with respect to those who are at the back.

3.3.1. Models with mingling

The basic assumption of these models is that, should overcrowding occur, the stop layout is such that passengers mingle at the stop without respecting any boarding priority.

The two most relevant methods proposed in this case are formally known as *effective frequency* and *fail-to-board probability*.

Effective frequency

The fundamental idea behind this method is that, with more buses arriving full, the waiting time will increase on average, because it is harder to get onto the vehicle. On the other hand, because passengers mingle, they all have the same likelihood of boarding an approaching bus. Therefore, rather than the nominal frequency (φ_a), it is assumed that passengers will consider an effective frequency (φ'_a) that, in the case of congestion, is lower than the nominal one. The split of passengers among attractive lines and the expected waiting time at the stop is, hence, calculated by applying equations 3-3 and 3-7, where the nominal frequency is substituted with the effective frequency. The route-choice model is defined similarly to the original one.

The first research where the concept of effective frequency was defined and exploited for strategy-based assignment is De Cea and Fernandez (1993). In this work, the effective

frequency is calculated as the inverse of the waiting time for the single line, which not only includes the delay due to discontinuous availability of transit services but also a sort of empirical volume-delay function that estimates the effect of capacity constraints, expressed in terms of ratio between total flow on-board plus wishing to board (*through flow* – Kurauchi et al., 2003) and the supplied capacity.

Equation 3-17 reproduces the formula suggested by De Cea and Fernandez where, with reference to Figure 3-4, DA_a is the dwelling arc corresponding to the line L_{HDa} and χ'_a ² is the *practical capacity* of the same line. The practical capacity is such that a line will never be totally full and, if attractive, the probability of using it will continually decrease as crowding increases, but will never be equal to zero.

Once the expected waiting time for the single line is calculated, the effective frequency (φ'_a) is determined as the inverse of this value (equation 3-17b).

$$w_a = \frac{1}{\varphi_a} + \beta \cdot \left(\frac{q_a + q_{DAa}}{\chi'_a} \right)^n \quad 3-17$$

$$\varphi'_a = \frac{1}{w_a} \quad 3-17b$$

β and n are calibration parameters, while q_{DAa} and q_a are respectively the flow already on-board and the flow of those who want to board.

² In order to improve readability, the notation χ_a is henceforth used in place of $\chi_{L_{HDa}}$.

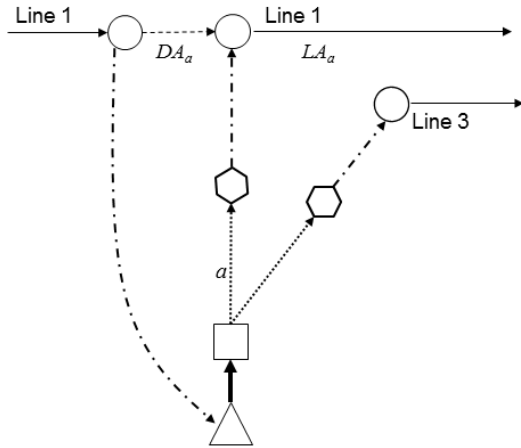


Figure 3-4
Hypergraph representation of Stop 2 where a is a branch of the waiting hyperarc, in this case such that $L_{HDA} = \text{Line 1}$; DA_a is the dwelling arc corresponding to a ; and LA_a is the line arc immediately downstream from a

The research of De Cea and Fernandez may be considered the first computationally tractable model to incorporate capacity constraints; however, it leads to overload of some services because practical capacities are used rather than strict capacities.

In order to overcome this fault, Cominetti and Correa (2001) use an alternative formulation of the effective frequency (equation 3-18) that incorporates congestion functions obtained from queuing models:

$$\varphi'_a = (q_a + q_{DAa}) \cdot \left[\frac{1}{\rho_a (q_a + q_{DAa})} - 1 \right] \quad 3-18$$

where $\rho_a(q_a + q_{DAa})$, $\rho_a \in [0, 1)$, is the unique solution of the following equation:

$$q_a + q_{DAa} = \varphi_a \cdot (\rho_a + \rho_a^2 + \dots + \rho_a^{z_a}) \quad 3-19$$

In this model, because strict capacity constraints are enforced, line loads never exceed capacity and, when the ‘through flow’ (on-board and wishing-to-board flow, taken together) approaches the line capacity, the effective frequency becomes null and the waiting time infinite.

Cepeda et al. (2006) continue the work of Cominetti and Correa (2001) and describe an alternative, more tractable formulation that may also be applied to large networks. The formula of effective frequency adopted in their numerical test is:

$$\varphi'_a = \varphi_a \cdot \left(1 - \left(\frac{q_a}{\chi_a \cdot \varphi_a - q_{L_{Da}} + q_a} \right)^\beta \right) \quad 3-20$$

where β is a calibration parameter and $\chi_a - q_{D_{Da}}$ is the available capacity on line L_{Da} .

Although the introduction of strict capacity overcomes the problem of overloading some services, it may produce problems in finding the equilibrium because the network capacity can be insufficient. Thus, Cepeda et al. (2006) suggest using a dummy network that connects all destinations with walking links.

It is important to notice that the models reviewed so far, which are based on the effective frequency method, all have the disadvantage of being static and unable to describe dynamic phenomena, such as the progressive formation and dispersion of queues over time.

Fail-to-board probability

Schmöcker et al. (2002) were the first to develop the method of fail-to-board probability, which was extended to FB strategy assignment by Kurauchi et al. (2003) and then to dynamic strategy-based assignment by Schmöcker et al. (2008). The following review refers to the latter work only.

The main idea for dealing with capacity constraints is that, in cases of oversaturation, some passengers fail to board a line L_{Da} with a probability (z_a) that depends (equation 3-21) on the capacity available on-board ($\chi_a - q_{D_{Da}}$) and on the flow of passengers who wish to board (q_a) at the time interval when z_a is to be evaluated (ζ^r).

$$z_a(\xi^\tau) = 1 - \max \left(0, \min \left(\frac{\chi_a(\xi^\tau) - q_{DAa}(\xi^\tau)}{q_a(\xi^\tau)}, 1 \right) \right) \quad 3-21$$

In order to represent this event graphically, failure arcs (A^F) are included in the hypergraph, as in Figure 3-5. In the case of overcrowding, the amount of passengers exceeding the available capacity is transferred back to the stop node via the failure arc. They therefore have to wait, again, for the first attractive line approaching the stop.

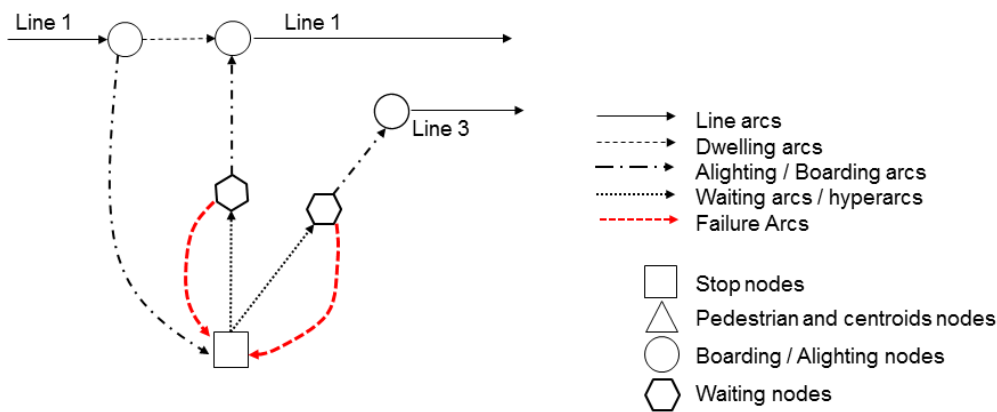


Figure 3-5
Hypergraph representation of Stop 2, depicted in Figure 3-1, in the spirit of Schmöcker et al. (2008)

Because passengers mingle at stop node i , they all have the same fail-to-board probability $z_a(\xi^\tau)$ and suffer the expected delay given by formula 3-22, where τ_{INT} is the duration of each time interval ξ^τ .

$$\tau_{INT} \cdot \left(\frac{z_a(\xi^\tau)}{1 - z_a(\xi^\tau)} \right) \quad 3-22$$

The fail-to-board SM resembles the static and uncongested model, and diversion probabilities and total expected waiting time may be calculated using equations 3-3 and 3-7. On the other hand, the RCM is innovative because it is assumed that the perceived generalised cost would increase due to the fail-to-board probability somewhere along the hyperpath (equation 3-23).

$$g_{k,i,d} = \sum_{l \in Q_k} \lambda_l \cdot \left(\sum_{a \in F_k \setminus \{F^H\}} c_a \cdot \delta_{al}(\xi^\tau) + \sum_{\substack{a \in F_k \cap F^H \\ i = TL_h}} w_h \cdot \delta'_{il}(\xi^\tau) + \beta \cdot \sum_{a \in F_k \cap F^F} \tau_{INT} \cdot \left(\frac{z_a(\xi^\tau)}{1 - z_a(\xi^\tau)} \right) \right) \quad 3-23$$

More specifically, the authors assume that the travel time on $a \in A^P \cup A^B \cup A^A \cup A^D \cup A^L$ stays constant during the analysis period and the same is assumed for the waiting cost on $h \in A^H$ (which depends only on the – constant – frequencies of attractive lines). Dynamic congestion effects are instead considered by means of the sum of expected delays, which is weighted for a calibration parameter (β) introduced to represent the person's value risk-averseness towards delays. Thus, if $\beta = 0$, the passenger is risk prone and would disregard delays due to overcrowding, when making his/her travel choices.

The method of fail-to-board probability for stop and assignment models has the major advantage of describing the progressive formation and dispersion of queues over time due to demand peaks; however, it does not include other dynamic phenomena, such as variation over time of instantaneous frequencies and travel times. Above all, the method suggested by Schmöcker et al. (2008) has the disadvantage of not considering the effect of congestion on diversion probabilities, as those solely depend on the nominal frequencies of the attractive lines.

3.3.2. Models with FIFO queues

In urban surface transport networks, the stop layout is usually such that passengers have to join a FIFO queue and respect the boarding priority of those who arrived before them. Models based on the mingling queuing protocol are clearly not applicable to this scenario.

To the best of the author's knowledge, all models developed so far for FIFO queuing make use of the following *stability condition* (Bouzaïene-Ayari et al., 2001): passengers waiting at a stop node would consider an attractive set that is never completely saturated and therefore each of them would be able to board the first vehicle coming, for at least one of the attractive lines. Two implicit consequences of this assumption are that:

- As congestion increases, more (and hence slower) lines are included in the attractive set;
- If all lines are congested, passengers would rather walk than remain waiting (even if frequencies are high, so that the extra waiting time due to congestion is, anyhow, short).

Gendreau (1984), Bouzaïene-Ayari (1988) and Bouzaïene-Ayari et al. (2001) develop similar SMs and RCMs where the travel strategy is selected before the beginning of the trip taking into account the expected residual capacity on-board.

More specifically, in the first two works, diversion probabilities are calculated by an empirical extension of equation 3-3 (where frequencies are substituted with residual capacities), while the expected waiting time is calculated by means of queuing model approximations derived by Kleinrock (1975) and Powell (1981).

The complexity of such models has prevented any application to real-scale networks. Consequently, in a later study, Bouzaïene-Ayari et al. (2001) try to simplify the model by assuming that headways are Erlang distributed with shape parameter k_a and rate parameter equal to the frequency φ_a . Therefore, the expected value of the waiting time before boarding line L_{HDa} is approximated as:

$$w_a = \frac{1}{2\varphi_a} \cdot \left(1 + \frac{1}{k_a}\right) + \frac{\beta}{2\varphi_a} \cdot \left(\frac{1}{k_a} + \frac{\varphi_a}{\chi_a + \varphi_a}\right) \cdot \frac{q_{LAa} - q_{DAa}}{\chi_a - (q_{LAa} - q_{DAa})} \quad 3-24$$

where β is a calibration parameter and q_{LAa} is the on-board flow on the line arc LAa .

An *attraction factor* is then defined, similarly to the effective frequency, as the inverse of the expected waiting time for the single line. Diversion probabilities and the total expected waiting time are calculated through equation 3-3 and equation 3-7, where frequencies are substituted with the attraction factors.

Although this formulation is meant to be simpler than the one proposed by Gendreau (1984) and Bouzaïene-Ayari (1988), the authors admit that the integration of such an SM with strict capacity constraints into assignment procedures for heavily congested transit networks is complicated. For this reason, a formulation without strict capacity constraints is suggested.

Also, Leurent and Benezech (2011) propose an SM that respects the stability condition. If a user arrives at the stop and the queue is very long, the probability that he may board a fast (or ‘more attractive’) service is low because passengers at the front of the queue have priority. Therefore, he might consider boarding a slower (or ‘less attractive’) service if it arrives first and has residual capacity. The level of congestion from which a line L_{HDa} becomes attractive is defined as the *attractivity threshold* and is indicated as v_a .

Consequently, the passenger distributions (equation 3-25) and the total expected waiting time (equation 3-26) depend on the frequencies of the attractive lines, as well as on the length of the queue (n) and the number of places available on each line ($\chi_a - q_{DAa}$).

$$P_{a|h} = \frac{\varphi_a \cdot \min((\chi_a - q_{DAa}), (n - v_a))}{\sum_{a \in h} \varphi_a \cdot \min((\chi_a - q_{DAa}), (n - v_a))} \quad 3-25$$

$$w_a = \frac{n}{\sum_{a \in h} \varphi_a \cdot \min((\chi_a - q_{DAa}), (n - v_a))} \quad 3-26$$

The main flaw of this model is that it is proposed for an isolated stop only and considers passengers travelling between the same *od* pair, disregarding interactions and/or overtaking between those who have different destinations.

3.4. THE PROPOSED DYNAMIC DEMAND MODEL

The literature reviewed shows that the formulation of SM for FB strategy assignment, where passenger congestion is considered in the form of FIFO queues, has led to intractable or unrealistic formulations. Finally, it should be noted that the proposed models are mainly static and thus unable to capture changes in network conditions over time.

Consequently, a new SM and RCM are proposed for transit networks where overcrowding may lead to the formation and dispersion of FIFO queues at stops. More specifically, although the same stop is shared by several lines, it is assumed that all passengers join a unique, mixed queue, regardless of their choice set. Therefore, if the first passenger in a queue does not board the first bus arriving at the stop because it is not attractive to him/her, the first passenger can be overtaken by the second, third, etc. in the queue, if the service is in their attractive set and there is available capacity on-board.

The new models developed to represent such phenomena are fully dynamic because travel variables are continuous functions of the time of day at which they are evaluated (*temporal profiles* – Bellei et al., 2005). Moreover, the proposed formulation can easily incorporate effects of wayside travel information and highly regular services – which is not

the case for the majority of other models, where the probability distributions and the total expected waiting time are given by various modifications of the standard formulas for the static uncongested case, 3-3 and 3-7.

3.4.1. Stop Model

In the proposed model, the basic hypotheses about carrier and passenger arrivals (Nguyen and Pallottino, 1988; Spiess and Florian, 1989) are not changed but it is assumed that passengers waiting at a stop may be prevented from boarding an approaching carrier by overcrowding. In this case, passengers join a FIFO queue and wait to board the first line of their attractive set that becomes available.

In the context of commuting trips, passengers know by previous experience the number of carrier passages they must let go before being able to board each line from each stop. Therefore, with reference to Figure 3-4, they would know that before boarding Line 1 (L_{HDa}) from Stop 2 (V_{TLa}), they should wait for the $\kappa_a(\tau)$ -th vehicle passage.

If τ is the time when the generic passenger reaches the stop and it is assumed that during the time spent waiting at the stop the frequency of a line is constantly equal to $\varphi_a(\tau)$, then the waiting time before the $\kappa_a(\tau)$ -th carrier arrival of line L_{HDa} occurs is a stochastic variable (w) with Erlang PDF of parameters $\kappa_a(\tau)$ and $\varphi_a(\tau)$ (Larson and Odoni, 1981: p. 54).

$$\text{PDF}_a(w, \tau) = \begin{cases} \frac{\varphi_a(\tau)^{\kappa_a(\tau)} \cdot \exp(-\varphi_a(\tau) \cdot w) \cdot w^{[\kappa_a(\tau)-1]}}{[\kappa_a(\tau)-1]!}, & \text{if } w \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad 3-27$$

As such, the diversion probabilities are expressed by equation 3-2 and the total expected waiting time by equation 3-6, where the PDF and $\overline{\text{CDF}}$ are Erlang.

As will become clear in the following (Section 3.4.3), the correct definition of the hyperpath's travel time (equation 3-37) requires the evaluation of g_{HDa} – the travel time upon boarding L_{HDa} – at the end of the expected waiting time for the considered line. Therefore, in order to calculate $t_{a|h}(\tau)$, an additional variable is defined ($w_{a|h}(\tau)$) that is the *conditional* expected value of the waiting time before boarding L_{HDa} . This expected value is conditional because it is subject to the event that L_{HDa} is the first line with available capacity to show up at the stop among those lines included in the attractive set (or, equivalently, the waiting hyperarc h). Recalling the definition of conditional expected value (Loève, 1978; Melotto, 2004), it is also possible to write:

$$w_{a|h}(\tau) = \frac{1}{p_{a|h}(\tau)} \int_0^{+\infty} w \cdot \text{PDF}_a(w, \tau) \prod_{a' \in h \setminus \{a\}} \overline{\text{CDF}}_{a'}(w, \tau) dw \quad 3-28$$

Moreover, the conditional expected waiting time can also be exploited to evaluate the total expected waiting time, as in equation 3-29:

$$w_h(\tau) = \sum_{a \in h} p_{a|h}(\tau) \cdot w_{a|h}(\tau) \quad 3-29$$

FIFO queues and mingling

Modelling passenger congestion in the form of FIFO queues or mingling may have remarkable impacts on the results obtained by the SM. In order to clarify such effects, consider the connection between Stop 3 and Stop 4 in Figure 3-1 and the four scenarios

summarised in Table 3-1. Services are always assumed to be irregular and inter-arrival times are thus exponentially distributed, with mean equal to the service frequency.

Table 3-1

Average headways, κ values and travel time upon boarding for the two lines in the considered scenarios

| | Line 3 Average frequency [min^{-1}] | Line 3 κ | Line 4 Average frequency [min^{-1}] | Line 4 κ |
|-----|---|--------------------|---|--------------------|
| S 1 | 1/15 | 1 | 1/15 | 1 |
| S 2 | 1/5 | 3 | 1/15 | 1 |
| S 3 | 1/3 | 5 | 1/15 | 1 |
| S 4 | 1 | 15 | 1/15 | 1 |

Table 3-2

Boarding probabilities, conditional expected waiting times and total expected waiting time at the stop for the considered scenarios (for clarity, the dependence of variables from τ is omitted)

| | | Uncongested Model | | Mingling | | | FIFO queues | | |
|-----|--------|-------------------|-------------------|-----------|-----------------------|-------------------|-------------|-----------------------|-------------------|
| | | $p_{a h}$ | $w_h[\text{min}]$ | $p_{a h}$ | $w_{a h}[\text{min}]$ | $w_h[\text{min}]$ | $p_{a h}$ | $w_{a h}[\text{min}]$ | $w_h[\text{min}]$ |
| S 1 | Line 3 | 0.50 | 7.50 | 0.50 | 7.5 | 7.5 | 0.50 | 3.00 | 7.5 |
| | Line 4 | 0.50 | | 0.50 | 7.5 | | 0.50 | 3.00 | |
| S 2 | Line 3 | 0.75 | 3.75 | 0.50 | 7.5 | 7.5 | 0.43 | 4.75 | 8.68 |
| | Line 4 | 0.25 | | 0.50 | 7.5 | | 0.57 | 3.93 | |
| S 3 | Line 3 | 0.73 | 2.18 | 0.50 | 7.5 | 7.5 | 0.40 | 5.00 | 9.00 |
| | Line 4 | 0.27 | | 0.50 | 7.5 | | 0.60 | 4.00 | |
| S 4 | Line 3 | 0.89 | 0.89 | 0.50 | 7.5 | 7.5 | 0.37 | 14.09 | 9.30 |
| | Line 4 | 0.11 | | 0.50 | 7.5 | | 0.63 | 6.29 | |

As shown in Table 3-2, the uncongested model (where $\kappa_a(\tau)$ is always considered equal to 1) disregards capacity constraints and calculates passengers' distributions and waiting times on the grounds of service frequencies *only*. Therefore, as the average headway of Line 3 decreases, its diversion probability increases and the total waiting time decreases. On the other hand, if it is assumed that, when congestion is considered in the form of mingling, passengers perceive an effective frequency $\varphi'_a(\tau) = \varphi_a(\tau)/\kappa_a(\tau)$, then for Line 3 this value is

always equal to $1/15 \text{ min}^{-1}$. Therefore, the diversion probabilities and total waiting time do not change for the four scenarios considered. Finally, in the case of a FIFO queue arising for Line 3, the passenger distribution for this service decreases progressively while the total waiting time increases.

These results may be explained intuitively considering, for example, scenario 2: $\kappa_{line3} = 3$, $\varphi_{line3} = 1/5 \text{ min}^{-1}$, $\kappa_{line4} = 1$ and $\varphi_{line4} = 1/15 \text{ min}^{-1}$. A passenger would board Line 3 only in the case where all the three vehicles of this service pass with a headway shorter than the average value of five minutes. However, this event is less probable than one vehicle of Line 4 arriving before its average inter-arrival time (fifteen minutes) and, consequently, the diversion probability for this line is greater than the diversion probability of Line 3.

The results can also be explained with the properties of the Erlang and Exponential distributions. Indeed, if the mean of the Erlang distribution is constant (in this case, $\kappa/\varphi = 15$) and $\kappa \rightarrow \infty$, then the PDF of the waiting time tends to be highly concentrated around its average value. Thus $w_{a|h}(\tau)$ for Line 3 increases and tends to $\kappa_{line3}/\varphi_{line3}$, while the boarding probability for Line 4 tends to the value expressed by equation 3-30, where, $\kappa_{line3}/\varphi_{line3} = 1/\varphi_{line4}$ is also the expected value of $\text{PDF}_{line4}(w, \tau)$.

$$p_{line4}(\tau) = \int_0^{\frac{\kappa_{line3}}{\varphi_{line3}}} \text{PDF}_{line4}(w, \tau) dw \quad 3-30$$

Because the mean of the Exponential distribution is always greater than its median, the diversion probability of Line 4 increases with κ_{line3} (for constant $\kappa_{line4} = 1$), while p_{line3} progressively decreases.

3.4.2. Extensions of the dynamic Stop Model

If the assumptions about supply-side phenomena are not applicable in the context of study, the proposed SM can be extended to incorporate the cases of interest.

Extension to networks with regular services

When regular services with constant headways are considered, it is not immediately possible to extend the definition of *dynamic attractive set* (Billi et al., 2004) to the scenario of interest because the evaluation of diversion probabilities and expected waiting times requires some integrations over the waiting time w (equations 3-2 and 3-28) during which it is assumed that travel variables – and also the attractive set – remain constant.

However, the model can be extended in the spirit of Gentile et al. (2005): the restructuring of the attractive set with the elapsed time at the stop is disregarded, but the PDF of the waiting time is assumed to be uniform, because headways are fixed, and equal to:

$$\text{PDF}_a(w, \tau) = \begin{cases} \varphi_a(\tau), & \text{if } \frac{[\kappa_a(\tau) - 1]}{\varphi_a(\tau)} \leq w < \frac{\kappa_a(\tau)}{\varphi_a(\tau)} \\ 0, & \text{otherwise} \end{cases} \quad 3-31$$

In order to clarify the combined effects of constant headways and FIFO queues, consider the connection between Stop 3 and Stop 4 in Figure 3-1 and the eight scenarios summarised in Table 3-3.

Table 3-3Average headways, service regularity, κ values and travel time upon boarding for the two lines in the considered scenarios

| | Line 3 Average frequency [min^{-1}] | Line 3 Service regularity | Line 3 κ | Line 4 Average frequency [min^{-1}] | Line 4 Service regularity | Line 4 κ |
|---------|---|---------------------------------|--------------------|---|---------------------------------|--------------------|
| S 5 | 1/15 | Regular | 1 | 1/15 | Regular | 1 |
| S 6 | 1/5 | Regular | 3 | 1/15 | Regular | 1 |
| S 7 | 1/3 | Regular | 5 | 1/15 | Regular | 1 |
| S 8 | 1/1 | Regular | 15 | 1/15 | Regular | 1 |
| S 9 | 1/15 | Regular | 1 | 1/15 | Irregular | 1 |
| S 10 | 1/5 | Regular | 3 | 1/15 | Irregular | 1 |
| S 11 | 1/3 | Regular | 5 | 1/15 | Irregular | 1 |
| S 12 | 1/1 | Regular | 15 | 1/15 | Irregular | 1 |

Table 3-4Boarding probabilities, conditional expected waiting times and total expected waiting time at the stop for the considered scenarios (for clarity, the dependence of variables from τ is omitted)

| | | $p_{a h}$ | $w_{a h}$ [min] | w_h [min] | | $p_{a h}$ | $w_{a h}$ [min] | w_h [min] |
|--------|--------|-----------|--------------------|-------------|---------|-----------|--------------------|-------------|
| Line 3 | S 5 | 0.50 | 5 | 5 | S 9 | 0.63 | 6.27 | 5.52 |
| Line 4 | | 0.50 | 5 | | | 0.37 | 4.22 | |
| Line 3 | S 6 | 0.17 | 11.67 | 7.22 | S 10 | 0.44 | 12.36 | 8.45 |
| Line 4 | | 0.83 | 6.33 | | | 0.56 | 5.42 | |
| Line 3 | S 7 | 0.1 | 13 | 7.4 | S 11 | 0.41 | 13.45 | 8.89 |
| Line 4 | | 0.9 | 6.78 | | | 0.59 | 5.76 | |
| Line 3 | S 8 | 0.03 | 14.33 | 7.49 | S 12 | 0.38 | 14.5 | 9.28 |
| Line 4 | | 0.97 | 7.25 | | | 0.62 | 6.1 | |

The results in Table 3-4 obviously show that, with increasing congestion on Line 3, the total waiting time at the stop (w_h) increases; however – as expected – the service regularity has a positive impact on the LoS in terms of total waiting time. For example, if the case where $\kappa_{line3} = 3$ and $\varphi_{line3} = 1/5 \text{ min}^{-1}$ is considered, the w_h calculated for S2 in cases where FIFO

queues arise (8.68 minutes) is longer than the value calculated for S10 (8.45 minutes) and S7 (7.22 minutes). However, in cases where both services are regular the total waiting time at the stop is constantly well below the values calculated in S1 – S4, by contrast the difference becomes progressively less relevant when comparing S1 – S4 with S9 – S12, because of the properties of the Erlang distribution mentioned in Section 3.4.1.

Similar considerations due to the properties of Erlang and Uniform distributions also apply when considering the effect of congestion on the calculation of diversion probabilities for regular and/or irregular lines. More specifically, if both lines have constant headways and congestion occurs, the passengers' ratio on Line 3 is constantly well below the value obtained in S2 – S4 and S10 – S12, while little difference can be seen among these two sets of scenarios.

Extension to networks with wayside information

The provision of wayside information through countdown displays brings about some important demand-side effects, as discussed here.

Depending on the design of the stop, two important sub-cases of FIFO queues may appear: either the stop is designed to have physically separate queues for each line; or passengers arriving at the stop join a single, mixed queue regardless of their attractive line set.

The first instance is very common in coach terminals. In this case, should congestion occur and no real-time information be available, passengers cannot behave strategically because they must join one specific queue as soon as they reach the stop. It may then be difficult to change queue in order to take advantage of events occurring while they are waiting (e.g. if another line arrives first). Consequently, the stop has to be modelled as a

group of separate stops, each of which is served by one line only. However, if real-time updates on actual arrivals/departures are available and passengers have sufficient experience to predict how many vehicles will pass before being able to board each line, travel behaviour in the case of separate queues can also be modelled as strategic. Indeed, the information ‘anticipates’ the event of a vehicle arrival to the moment when the user reaches the stop; hence, the optimal travel strategy comes true in the moment when the traveller actually chooses which line to board, taking into account the length of the different queues. Thus, if information is provided, this case can be treated as if there were a single mixed queue.

The second type of stop layout (single, ‘mixed’ FIFO queue) is more common in urban public-transport networks and has been the only one considered so far in this chapter. If congestion occurs, users arriving at the stop join the queue and board the first line of their attractive set that becomes available. However, if no real-time information is provided, it is possible that passengers would change their attractive set while they wait, as described by Billi et al. (2004) and Noekel and Wekeck (2007). On the other hand, if information is provided, an attractive-set structuring can more easily be modelled also in the presence of regular services because it can be assumed that passengers know the line they will board as soon as they reach the stop.

Consequently, if countdown displays are available, the inclusion of services with constant headways simply requires one to note that the waiting time is uniformly distributed, as in equation 3-31, and include this PDF and $\overline{\text{CDF}}$ in the time-dependent version of equations 3-15 and 3-16.

Some stops can be shared by regular and irregular services. For example, this can be the case for large bus terminals, where there are some lines whose routes run in segregated lanes (where the absence of interaction with private car traffic and/or road works enhances

the service regularity) and there are also some other lines that are subject to service irregularity because their routes do not run in segregated lanes.

For this reason, the extension of equations 3-15 and 3-16 to the congested and dynamic setting is articulated into two different subcases, depending on whether the line considered for the evaluation of its diversion probability has constant or exponentially distributed headways. As detailed in the following, this has an impact on the evaluation of the survival functions of the waiting times of the ‘competitor’ lines.

For example, if L_{HD_a} is a service with constant headways, $PDF_a(w, \tau)$ is expressed by means of equation 3-31. Moreover, if:

$$\beta_{a'} = w + \frac{\kappa_a(\tau) - 1}{\varphi_a(\tau)} + g_{HD_{a,d}} - g_{HD_{a',d}} \quad 3-32$$

then $\overline{CDF}_{a'}(w + g_{HD_{a,d}} - g_{HD_{a',d}}, \tau)$ is expressed as in equation 3-33 if $L_{HD_{a'}}$ is a service with exponentially distributed headways; while if $L_{HD_{a'}}$ is a service with constant headways, $\overline{CDF}_{a'}(w + g_{HD_{a,d}} - g_{HD_{a',d}}, \tau)$ is expressed as in equation 3-34.

$$\overline{CDF}_{a'}(w + g_{HD_{a,d}} - g_{HD_{a',d}}, \tau) = \sum_{j=0}^{\kappa_{a'}(\tau)} \frac{\varphi_{a'}(\tau)^{\kappa_{a'}(\tau)-j} \cdot e^{-\varphi_{a'} \cdot \beta_{a'}} \cdot \beta_{a'}^{[\kappa_{a'}(\tau)-j]}}{(\kappa_{a'}(\tau) - j)!} \quad 3-33$$

$$\overline{\text{CDF}}_{a'}(w + g_{HD_{a,d}} - g_{HD_{a',d}}, \tau) = \begin{cases} 1, & \beta_{a'} < \frac{\kappa_{a'}(\tau) - 1}{\varphi_{a'}(\tau)} \\ \int_{\beta_{a'}}^{\frac{\kappa_{a'}(\tau)}{\varphi_{a'}(\tau)}} \varphi_{a'}(\tau), & \frac{\kappa_{a'}(\tau) - 1}{\varphi_{a'}(\tau)} < \beta_{a'} < \frac{\kappa_{a'}(\tau)}{\varphi_{a'}(\tau)} \\ 0, & \beta_{a'} > \frac{\kappa_{a'}(\tau)}{\varphi_{a'}(\tau)} \end{cases} \quad 3-34$$

On the other hand, in the case where L_{HD_a} is a service with exponentially distributed headways, then $\text{PDF}_a(w, \tau)$ is expressed by means of equation 3-27, while $\overline{\text{CDF}}_{a'}(w + g_{HD_{a,d}} - g_{HD_{a',d}}, \tau)$ is expressed by equations 3-33 and 3-34 for irregular and regular services respectively, where $\beta_{a'}$ is defined as:

$$\beta_{a'} = w + g_{HD_{a,d}} - g_{HD_{a',d}} \quad 3-35$$

Table 3-5

Average headways, service regularity, κ values and travel time upon boarding for the two lines in the considered scenarios

| | Line 3 Average frequency [min ⁻¹] | Line 3 Service regularity | Line 3 κ | Line 3 Travel time upon boarding [min] | Line 4 Average frequency [min ⁻¹] | Line 4 Service regularity | Line 4 κ | Line 4 Travel time upon boarding [min] |
|---------|--|---------------------------------|--------------------|---|--|---------------------------------|--------------------|---|
| S 13 | 1/15 | Irregular | 1 | 5 | 1/15 | Irregular | 1 | 10 |
| S 14 | 1/5 | Irregular | 3 | 5 | 1/15 | Irregular | 1 | 10 |
| S 15 | 1/3 | Irregular | 5 | 5 | 1/15 | Irregular | 1 | 10 |
| S 16 | 1/1 | Irregular | 15 | 5 | 1/15 | Irregular | 1 | 10 |
| S 17 | 1/15 | Regular | 1 | 5 | 1/15 | Regular | 1 | 10 |
| S 18 | 1/5 | Regular | 3 | 5 | 1/15 | Regular | 1 | 10 |
| S 19 | 1/3 | Regular | 5 | 5 | 1/15 | Regular | 1 | 10 |

| | Line 3 Average frequency [min ⁻¹] | Line 3 Service regularity | Line 3 κ | Line 3 Travel time upon boarding [min] | Line 4 Average frequency [min ⁻¹] | Line 4 Service regularity | Line 4 κ | Line 4 Travel time upon boarding [min] |
|---------|--|---------------------------------|--------------------|---|--|---------------------------------|--------------------|---|
| S 20 | 1/1 | Regular | 15 | 5 | 1/15 | Regular | 1 | 10 |
| S 21 | 1/15 | Regular | 1 | 5 | 1/15 | Irregular | 1 | 10 |
| S 22 | 1/5 | Regular | 3 | 5 | 1/15 | Irregular | 1 | 10 |
| S 23 | 1/3 | Regular | 5 | 5 | 1/15 | Irregular | 1 | 10 |
| S 24 | 1/1 | Regular | 15 | 5 | 1/15 | Irregular | 1 | 10 |

In order to clarify the combined effects of countdown displays, service regularity and FIFO queues, consider the connection between Stop 3 and Stop 4 in Figure 3-1 and the twelve scenarios summarised in Table 3-5. Notably, in this case, the solution of the SM requires not only that frequencies and congestion levels are known, but also that travel times to destination are known. It is therefore assumed that, while the travel time upon boarding Line 3 is of 5 minutes, for Line 4 it is of 10 minutes. The results of this SM are displayed in Table 3-6.

Table 3-6

Boarding probabilities, conditional expected waiting times and total expected waiting time at the stop for the considered scenarios (for clarity, the dependence of variables from τ is omitted)

| | | $P_{a h}$ | $W_{a h}$ [min] | W_h [min] | | $P_{a h}$ | $W_{a h}$ [min] | W_h [min] | | $P_{a h}$ | $W_{a h}$ [min] | W_h [min] |
|--------|---------|-----------|--------------------|----------------|---------|-----------|--------------------|----------------|---------|-----------|--------------------|----------------|
| Line 3 | S 13 | 0.64 | 8.02 | 7.83 | S 17 | 0.78 | 6.31 | 5.65 | S 21 | 0.81 | 6.62 | 5.96 |
| Line 4 | | 0.36 | 7.5 | | | 0.22 | 3.33 | | | 0.18 | 2.98 | |
| Line 3 | S 14 | 0.59 | 11.19 | 9.08 | S 18 | 0.5 | 12.22 | 8.05 | S 22 | 0.61 | 12.36 | 8.90 |
| Line 4 | | 0.41 | 5.04 | | | 0.5 | 3.89 | | | 0.39 | 3.52 | |
| Line 3 | S 15 | 0.56 | 12.53 | 9.38 | S 19 | 0.43 | 13.38 | 8.23 | S 23 | 0.57 | 13.45 | 9.32 |
| Line 4 | | 0.44 | 5.39 | | | 0.57 | 4.29 | | | 0.43 | 3.87 | |
| Line 3 | S 16 | 0.51 | 14.62 | 9.72 | S 20 | 0.37 | 14.48 | 8.32 | S 24 | 0.53 | 14.5 | 9.70 |
| Line 4 | | 0.49 | 4.58 | | | 0.63 | 4.75 | | | 0.47 | 4.24 | |

When information is provided by means of countdown displays, similar considerations apply on the effect of FIFO queues on values of waiting times and diversion probabilities. However, because passengers prefer to board the fastest service (rather the first that becomes available), they accept having to wait longer at the stop and consequently the value of w_h is constantly higher than the one calculated in the corresponding scenarios where no information is assumed. For the same reason, the diversion probability on Line 3 is constantly higher than the value calculated in the corresponding scenarios without countdown displays and, with increasing congestion, the effect is predominantly relevant when both services have constant headways.

3.4.3. Route Choice Model: dynamic hyperpath search

A sub-hypergraph $H_{k,o,d} = (N_k, A_k)$ of H , where $N_k \subset N$, $A_k \subset A$, is a dynamic hyperpath if:

- $H_{k,o,d}$ is acyclic;
- o has no predecessors and one successor arc;
- d has no successors and at least one predecessor arc;
- For every node $i \in N_k \setminus \{o, d\}$ there is at most one immediate successor arc if $i \notin N^S$, otherwise the successor is a hyperarc with cardinality equal or greater than one;
- For each hyperarc $h \in H_{k,o,d}$ a characteristic vector $\mathbf{p}(\tau)$ is defined where $\mathbf{p}(\tau) = (p_{a|h}(\tau))$ is a real-value vector of dimension $(|HD_h| \times 1)$ such that:

$$\begin{aligned} \sum_{a \in h} p_{a|h}(\tau) &= 1 \\ p_{a|h} &> 0 \end{aligned} \quad 3-36$$

It has been shown that, in a static context, the total travel time of the generic hyperpath $H_{k,o,d}$ can be computed by explicitly taking into account all the elemental paths l forming it (Nguyen and Pallottino, 1988; Nguyen and Pallottino, 1989), as in equation 3-11. In a dynamic setting, such as that considered here, travel times depend on the time the arc is entered. Consequently, it can happen that the same node is traversed by different paths at different times and the travel time associated with it has different values. Hence, the definition of the dynamic hyperpath's total travel time is given only implicitly, by extending the *generalised Bellman equation*:

$$g_{i,d}(\tau) = \begin{cases} 0, & \text{if } i = d \\ \min(c_a(\tau) + g_{HD_o,d}(t_a(\tau)) : a \in FS_i), & \text{if } i \notin N^S \\ \min\left(w_h(\tau) + \sum_{a \in h} p_{a|h}(\tau) \cdot [g_{HD_o,d}(t_{a|h}(\tau))]\right) : h \in HFS_i, & \text{if } i \in N^S \end{cases} \quad 3-37$$

This implicit formulation of the RCM can always be applied in a static scenario because the *concatenation property* always holds true. On the other hand, it should be noted here that this property applies to a dynamic problem only if the FIFO rule is respected (Ziliaskopoulos, 1994), as is the case here for passengers having the same attractive set.

Attractive set definition

If waiting times are not exponentially distributed, the combinatorial problem (equation 3-38) of selecting the attractive set h cannot be solved through a greedy approach.

$\exists h \in HFS_i :$

$$w_h + \sum_{a \in h} p_{a|h}(\tau) \cdot [g_{HD_{a,d}}(t_{a|h}(\tau))] = \min \left\{ w_{h'} + \sum_{a \in h'} p_{a|h'}(\tau) \cdot [g_{HD_{a,d}}(t_{a|h'}(\tau)) : h' \in HFS_i \right\} \quad 3-38$$

Moreover, as the exit time $t_{a|h'}(\tau)$ changes with the h' considered, also $g_{HD_{a,d}}(t_{a|h'}(\tau))$ may change. Hence, it is in general necessary to compute $g_{i,d}(\tau)$ for all the possible $h' \in HFS_i$ and set h equal to the hyperarc that yields the minimum travel time $g_{i,d}^*(\tau)$.

Obviously, applications to large-scale networks require a simplification of this combinatorial problem, as will be discussed in Chapter 5.

3.5. DISCUSSION

This chapter proposes an innovative demand model for dynamic transit assignment, which allows for considering overcrowding at transit stops.

The fundamental hypotheses on demand-side phenomena are that:

1. In the context of commuting trips, passengers have a good knowledge of transit supply, both in terms of line frequencies and average travel time upon boarding, and in terms of the number of vehicles of the same line they will fail to board because of overcrowding.
2. Passengers do not time their arrival at the stop with the service timetable.
3. If the network is densely connected and services are very frequent, passengers select a travel strategy rather than a single path. In the case of congestion, the dynamic strategy is chosen depending on: the expected travel time upon boarding each

attractive line at the time of boarding ($g_{HD_{a,d}}(t_{ah}(\tau))$); the instantaneous frequency of the attractive lines at the time when the stop is reached ($\varphi_a(\tau)$); and the congestion parameter of the attractive lines at the same time ($\kappa_a(\tau)$).

While the hypotheses on supply-side phenomena are that:

- a. In cases of overcrowding, passengers respect a single-file FIFO queuing protocol; therefore, even if a stop is shared by several lines, passengers join a unique queue and, while they respect the priority of those in the queue with the same attractive set, overtaking may be possible among passengers with different choice sets;
- b. No real-time updates are provided on actual vehicle arrivals/departures;
- c. Vehicle arrivals follow a Poisson distribution with rate equal to the instantaneous frequency $\varphi_a(\tau)$.

The core of the demand model for dynamic assignment is the dynamic SM presented in Section 3.4.1. In its principal formulation (equation 3-27), the model considers the usual assumptions on supply-side phenomena (b, c) that are accepted by the majority of frequency-based models for transit assignment (Nguyen and Pallottino, 1988; Spiess and Florian, 1989; De Cea and Fernandez, 1993; Cominetti and Correa, 2001; Cepeda et al., 2006; Schmöcker et al., 2008). However, the model can be easily extended to consider other scenarios, whose effects on waiting time and passenger distribution are discussed in Section 3.4.2.

Unlike the Exponential and Uniform distributions, the Erlang distribution (which may be used to describe waiting times before the arrival of services with medium regularity) cannot be easily convoluted and, thus, the proposed SM cannot incorporate services with Erlang-distributed headways.

The demand model is completed by incorporating the proposed SM into a dynamic RCM with hyperpaths, whose inputs are: time-dependent frequencies; in-vehicle travel times; dwelling times; passengers' boarding and alighting times; and congestion factors ($\kappa_a(\tau)$).

In order to estimate $\kappa_a(\tau)$ accurately for each combination of transit line/stop, the proposed SM and deterministic RCM need to be embedded in a complete dynamic transit assignment procedure, in the form of a dynamic Deterministic User Equilibrium with hyperpaths (Figure 1–1).

As such, beyond the stop and route choice models (which are the fundamental pillars of the demand model), the supply model also has to be specified through its two components:

- The *Arc Performance Function* (APF), which yields the exit time at any given entry time for each arc, depending on the transit lines' characteristics and the passenger flows over the network;
- The *Network Flow Propagation Model* (NFPM), which aims at finding time-varying arc flows that are consistent with the arc travel times for given route choices, but not consistent with line capacities. (This is the main difference between the NFPM and the Dynamic Network Loading Problem, where instead mutual consistence of flows and times is sought through the APF for given route choices.)

The research background and methodological innovation of the proposed supply model and demand–supply interaction model will be detailed in the next chapter.

4. SUPPLY AND DEMAND– SUPPLY INTERACTION MODELS FOR STRATEGY- BASED DYNAMIC TRANSIT ASSIGNMENT

4.1. INTRODUCTION

Recurrent congestion has developed into a major problem affecting the high-frequency transit systems of large cities. For example, during peak hours, passengers often experience what is known as an ‘*oversaturation queuing time*’ at stops (Meschini et al., 2007) because they are unable to board the first vehicle of their choice that arrives. The queue of those who remain at the stop may also increase passenger congestion for subsequent vehicle arrivals, thus leading to high Level of Service (LoS) variations in a short time.

Consequently, the classical framework of static assignment may be an improper analysis tool as it is unable to capture the excess of travel demand with respect to supplied capacity as well as changes in transit supply during peak periods (static models without capacity constraints), or the LoS variations that may follow demand peaks (static models with capacity constraints).

From the passenger perspective, the main factors influencing travel choices are services’ performances, such as: in-vehicle travel time; frequency; regularity; and overcrowding. Thus, in the context of congested networks, Chapter 3 has presented several static and dynamic formulations already available in the literature, as well as a new dynamic

model, for reproducing the effects of overcrowding on passengers' travel choices and on the waiting process at transit stops (demand model).

By contrast, when considering the 'dual problem' of reproducing the effects of travel choices on the network, it is also crucial to capture LoS variations over time. Thus, a *supply model for dynamic assignment* (Cascetta, 2009: p. 425) is introduced in this chapter through its two main components:

- The *Network Flow Propagation Model* (NFPM), which aims at finding time-varying arc flows that are consistent with arc travel times for given route choices;
- The *Arc Performance Function* (APF), which yields the exit and entry times for each arc, depending on the transit lines' characteristics and the passenger flows over the network.

After reviewing existing dynamic supply models (Section 4.2), some additional notation and definitions are introduced in Section 4.3; and Section 0 details the new NFPM and APF proposed.

Finally, Section 4.5 of this chapter describes the assignment model developed to simulate the dynamic demand–supply interaction in the form of a *User Equilibrium* configuration, with some remarks on its characterisation in terms of the existence and uniqueness of the equilibrium and the existence of multiple hyperpaths to destination at the equilibrium.

4.2. SUPPLY MODELS FOR DYNAMIC TRANSIT ASSIGNMENT: A REVIEW

4.2.1. Network Flow Propagation Model for Dynamic Transit Assignment

The NFPM aims at spreading passenger flows across the network, consistently with travel demand, route choices and network performances. If the transit assignment aims at studying dynamic congestion phenomena (such as the formation and dispersion of passenger queues over time), then the flow propagation also depends on the specific time of the day considered, as the following example helps to make clear.

Consider the small network in Figure 4-1 and its performances for the analysis period 07:30–08:30, outside which it is assumed that the travel demand is null.

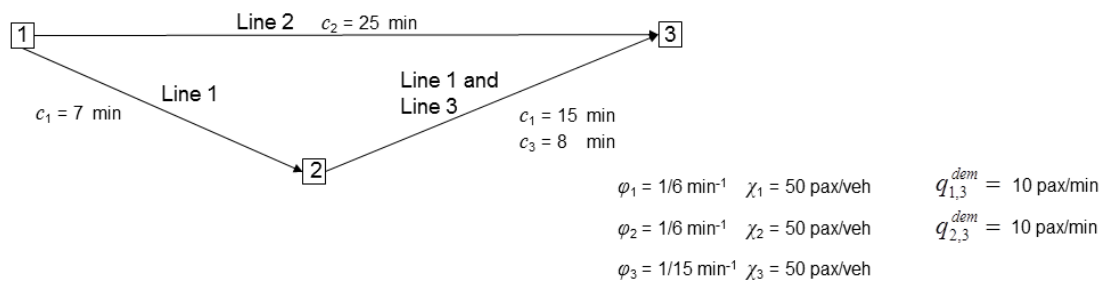


Figure 4-1

Example network and travel variables: c_i indicates the in-vehicle travel time for the generic line i on the corresponding edge; χ_i indicates the capacity of each vehicle of line i ; φ_i indicates the (constant) frequency of line i ; and $q_{j,4}^{dem}$ is the instantaneous travel demand (assumed here to be constant) from the generic node j to node 4 (the only destination considered)

At the beginning, all passengers waiting at Stop 2 can board Line 1 or Line 3, which are both attractive and have full available capacity. Thus passenger distribution on the lines solely depends on the frequency of the services.

On the other hand, as soon as the flow of those who boarded Line 1 at Stop 1 reaches Stop 2, the available capacity on this line decreases and, consequently, the flow that boards Line 1 at Stop 2 decreases to meet the capacity constraint. In other words, the flow of

passengers boarding a specific line at a specific stop is a function of the time of the day that must be consistent with available capacities, are generalised travel times and route choices evaluated at the same time of the day. Those changes are certainly not captured by static uncongested models or even by static congested models.

As highlighted in Chapter 2, this problem is efficiently dealt with in SB transit-assignment models that make use of the *diachronic graph* because the time dimension of the problem is explicitly reproduced by the graph topology.

More specifically, in models with continuous flow where the route choice is dynamically adapted according to congestion levels, so as to minimise the waiting time at the stop, the equilibrium simply reduces to a static assignment on the time-dependent network, as in Hamdouch and Lawphongpanich (2008) and Nuzzolo et al. (2012).

On the other hand, there are examples of discrete-flow supply models (Poon et al., 2004) where such adaptation of the route choice is not considered because passengers are forced to join a FIFO queue and wait until a vehicle of the chosen line comes with capacity available on-board.

In order to reproduce this phenomenon during the network loading, individual *packets of passengers* directed towards the same destination and making use of the same path are moved forward across the network in topological and chronological order until they reach a stop. At this point, the loading of the packets pauses because it is not known whether these passengers can board the transit vehicle at the time of their arrival or not; and it resumes on the *residual path* to destination when the movements of all other passengers have been simulated at least up to this moment in time (i.e. it resumes when it is known whether or not capacity is available on-board). Instead of sequentially solving the NFPM and APF, the procedure of Poon et al. (2004) seeks mutual consistency of flows and generalised travel

times – for given route choices – and can be thought of as a Dynamic Network Loading Problem (DNLP – Wu et al., 1998; Cascetta, 2009) for SB transit assignment.

By contrast, in the *frequency-based* realm, the definition of a supply model for dynamic assignment is not equally simple because different runs of the same service are not distinguished and thus it is not immediately possible to evaluate the capacity available on a certain line/stop at a certain time of the analysis period. Indeed, the majority of available models, tough with capacity constraints, are developed in a static setting only.

One of the few existing dynamic examples is provided by Schmöcker et al. (2008), who propose an NFPM for FB transit assignment with hyperpaths that makes use of a continuous-flow supply representation and, on the assumption that passengers mingle on the platform, is formulated as a Markovian loading process.

The analysis period is divided into time intervals (ξ^τ) and, proceeding in chronological order, the loading process takes the following steps:

1. For a passenger directed towards d , calculate the *transition probability matrix* ($\mathbf{\Pi}^\tau_d$) that he/she may move from node i to j at time interval ξ^τ ;
2. Calculate for each destination d the vector $\mathbf{q}_d^{dem,\tau}$, which includes the demand flow at each intermediate node i (should it be an origin node) and time interval ξ^τ , and also the flow of those who failed to board at previous time intervals (should it be a stop node);
3. Evaluate the vector of flows traversing each intermediate node i and directed towards destination d at time interval ξ^τ as:

$$\mathbf{q}_d^\tau = \left(\left[\mathbf{I} - \mathbf{\Pi}_d^\tau \right]^{-1} \right)' \Delta^\tau \mathbf{q}_d^{dem,\tau} \quad 4-1$$

It should be noted here that Δ^r is a matrix whose elements (δ_{ij}^r) are equal to one if the travel time between i and j is shorter than the length of one time interval and equal to zero otherwise. Obviously, Δ^r depends on the length of time intervals and, as for all models exploiting a discrete time representation, a compromise between result accuracy (short time intervals) and algorithm performances (long time intervals) must be attained. Finally, when equilibrium flows are calculated, fail-to-board probabilities are adjusted, in order to ensure capacity constraints, and failed-to-board trips are re-assigned to the following time interval.

The main drawback of the Markovian loading proposed by Schmöcker et al. (2008) is its inapplicability to the case where passenger congestion results in FIFO queues. Indeed, in this case, past states determine the current position of the passenger within the queue and thus affect their probability of boarding a line. As such, the Markovian loading process cannot be applied.

By contrast, this form of congestion is considered by Meschini et al. (2007) and Papola et al. (2009), where the same NFPM, with continuous flow representation, is proposed for FB and SB dynamic transit assignment, respectively. More specifically, the authors extend an existing approach for Dynamic Traffic Assignment (Bellei et al., 2005) where flows are macroscopic time-continuous functions (*temporal profiles*) and conceive transit services as a continuous flow of supply with ‘instantaneous capacity’ (e.g. 1,000 passengers per hour and not 100 passengers per vehicle), which allows representing the average effect of time-discrete services on the temporal profile of generalised travel times. It should be noticed that this almost ‘continuous availability of the vehicles’, though questionable from a phenomenal point of view, is consistent with the basic assumption of the *frequency-based* modelling framework that passengers conceive all the runs of the same line as a unitary supply facility.

In such a setting, the proposed NFPM simulates how temporal profiles of path flows propagate through the network in topological and chronological order and induce temporal profiles of arc inflows and outflows that are consistent with the arc travel times, for given route choices.

The limitation of these studies, though, is that passenger strategies are not considered in the RCM; moreover, it is not clear how the additional waiting time due to overcrowding should be accounted for when passengers in a queue are willing to board a set of lines and passengers in the same queue have different attractive sets (e.g. because they are bound for different destinations). Therefore, Section 0 discusses the methodological innovations needed to extend this supply model to dynamic assignment with travel strategies.

4.2.2. Arc Performance Functions

Link travel time functions or *arc performance functions* (APF) are a fundamental component of dynamic supply models for congested networks because they express the generalised travel time (or performance) on an arc at a certain time of the day ‘*as a function of link flows on the network*’ (Cascetta, 2001: p. 379) at the same time of the day.

In transit-assignment models, the in-vehicle travel time is usually given as an exogenous input that only depends on the time of the day at which it is evaluated. However, in multimodal assignment models (Meschini et al., 2007), it is a function of the time-dependent car-equivalent flow on the road arc considered.

Also, in the majority of transit-assignment models, the dwelling time spent by a vehicle at the stop as well as passengers’ boarding and alighting times are considered constant or time dependent, but flow independent.

By contrast, when congestion occurs, the extra waiting or oversaturation queuing time spent by passengers at the stop is a function of the flow and cannot be considered as exogenous information.

In Schmöcker et al. (2008), for example, the oversaturation queuing time is a function of the on-board flow and the number of passengers who wish to board through the fail-to-board probability. This formulation takes advantage of the Markovian properties of the mingled queues considered but cannot be immediately extended to the case where passengers queue according to a FIFO protocol.

In this case (Poon et al., 2004; Meschini et al., 2007; Papola et al., 2009), it is possible to exploit a *Bottleneck Queue Model* that explicitly simulates the formation and dispersion of FIFO queues and, thus, determines the oversaturation queuing time. As explained in Cascetta (2009: p. 425), the mathematical formulation of a bottleneck usually considers the cumulative number of passengers arriving at the stop and the cumulative number of passengers leaving the stop, which in turn depends on the capacity available on-board. If the available on-board capacity does not suffice to accommodate passengers arriving at the stop, a queue builds up which will dissipate only if/when the available on-board capacity is greater than the inflow of arriving passengers.

4.3. SUPPLY MODEL: NOTATION AND DEFINITIONS

4.3.1. Network representation

As discussed in Chapter 3, the *Stop Model* and *Route Choice Model* refer to hypergraph H , whose *waiting hyperarcs* graphically represent the process of waiting for the first service with available capacity among a set of lines. When making their travel choices, passengers do not distinguish between the under-saturation delay, due to the inherent transit service discontinuity, and oversaturation queuing time. Thus the total waiting cost w_h for a waiting hyperarc h includes both attributes and depends on which lines are included in the attractive set.

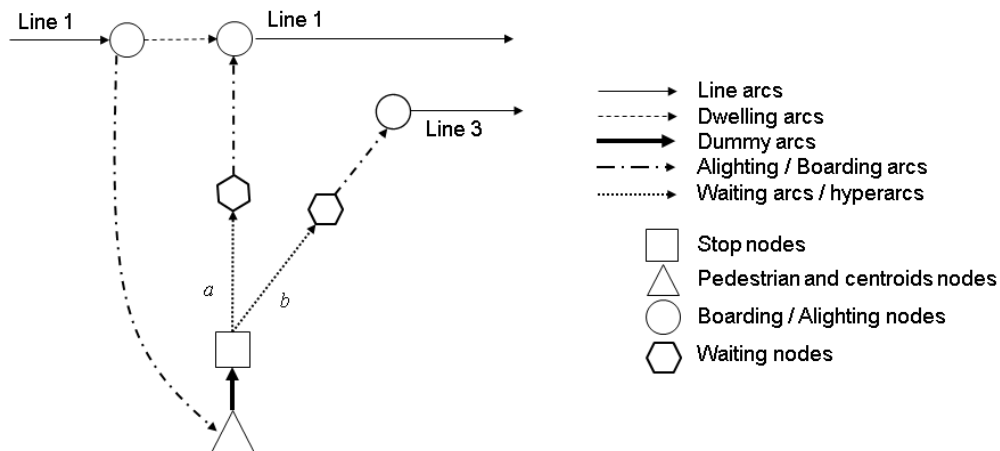


Figure 4-2
Hypergraph representation of Stop 2 depicted in Figure 3-1, according to the hypergraph description given in Chapter 3

On the other hand, in order to ensure that capacity constraints are respected, in the NFPM and APF it is necessary to ensure that inflows/outflows are consistent with arcs' exit times and therefore the two waiting phases – under-saturation delay and oversaturation queuing time – have to be distinguished, as the following example helps to make clear.

Consider the hypergraph stop representation depicted in Figure 4-2 and the hyperarcs $h' = \{a\}$ and $h'' = \{a, b\}$. Clearly, the conditional exit time $t_{a|h}(\tau)$ calculated by means of the *Stop Model* specified in Chapter 3 is different if the hyperarc considered is h' or h'' .

If transit lines are conceived as a continuous flow of carriers, with instantaneous available capacity $e_a(\tau)$, then the capacity available on Line 1 at $t_{a|h}(\tau)$ is generally different from the capacity available on the same service at $t_{a|h''}(\tau)$. As a result, in the case of sudden LoS variations, it could be that $e_a(t_{a|h'}(\tau)) \ll e_a(t_{a|h''}(\tau))$, and therefore those who are at the front of the queue could be loaded, or not loaded, onto the next vehicle arriving on the basis of their choice set.

However, when loading the network, the FIFO rule requires that the event of boarding or not boarding an attractive line depends solely on on-board capacity constraints and on the position in the queue occupied by the passenger, while it does not depend on which other lines are attractive for the passenger.

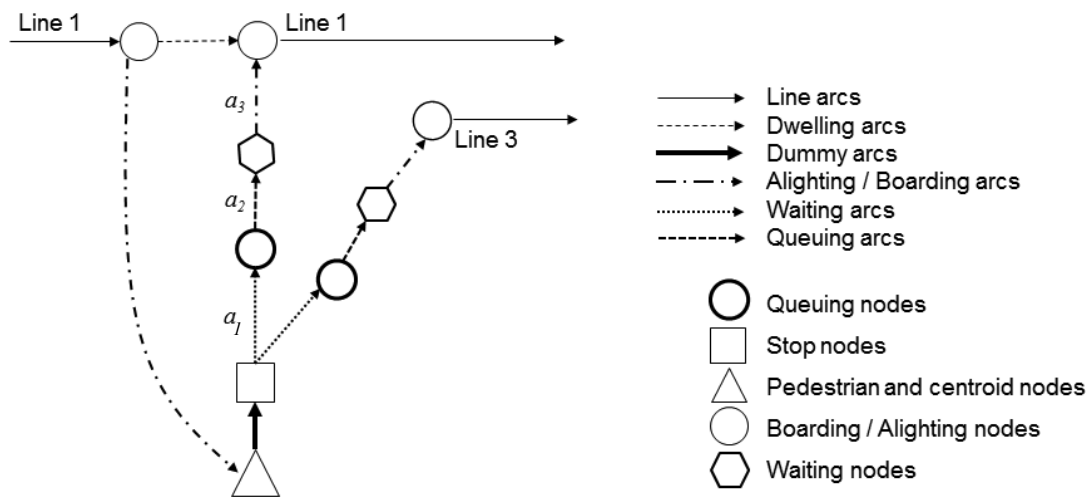


Figure 4-3
Model graph representation of the same stop depicted in Figure 4-2

Thus, in order to ensure that inflows/outflows are consistent with arcs' exit times, the NFPM and APF are referred to a *model graph* $G(A, N)$ such that different graphic structures are used

to model the two waiting phases described above (under-saturation delay and oversaturation queuing time). More specifically, the node set N and arc set A of the model graph are built as the union of the following subsets (Figure 4-3):

$$N = N^P \cup N^S \cup N^A \cup N^W \cup N^Q \cup N^B;$$

$$A = A^P \cup A^L \cup A^D \cup A^Z \cup A^A \cup A^W \cup A^Q \cup A^B.$$

The definitions of the node subset defined in Chapter 3 also apply here and the definitions of A^P , A^L , A^D , A^Z and A^A correspond to those given in Chapter 3 for the equivalent hyperarc subsets. On the other hand, some new subsets of nodes and arcs are introduced for the model graph:

$$N^Q: \text{ queuing nodes, } N^Q = \{(R_{\ell,i}, Q, \ell) : \ell \in L, i \in [1, \sigma_{\ell}-1], \mu_{\ell,i} > 0\};$$

A^W : *waiting arcs*, which represent only the *under-saturation* waiting time, i.e. the average delay due to the fact that the transit service is not continuously available over time: $A^W = \{(i, j) : i \in N^S, j \in N^W, V_i \equiv V_j\}$;

A^Q : *queuing arcs*, which represent only the *oversaturation* queuing time, i.e. the ‘time spent by users queuing at the stop and waiting [until] the next service becomes actually available to them’ (Meschini et al., 2007): $A^Q = \{(i, j) : i \in N^W, j \in N^Q, V_i \equiv V_j\}$;

A^B : *boarding arcs*, which represent the time passengers need to embark on a vehicle:

$$A^B = \{(i, j) : i \in N^Q, j \in N^B, V_i \equiv V_j\}.$$

Finally, the directed hypergraph $H = (N, F)$, to which the *Stop Model* and *Route Choice Model* are referred, can always be built on graph G . In this case, it is assumed that: $F = A \setminus \{A^W\} \cup A^H$ and, as waiting hyperarcs A^H represent the total waiting time for the considered attractive set, the travel cost of queuing arcs $A^Q \subset H$ is always assumed to be null.

This consistency is crucial for algorithmic purposes as only one graphic structure (model graph) has to be built and stored for the network. On the other hand, in the supply model, the assumption of first representing the under-saturation delay through the waiting arc and then the oversaturation delay through the queuing arc, as in Figure 4-3, is questionable from a phenomenal point of view as the exact opposite occurs in reality. However, this is a valid choice from a modelling point of view for three reasons:

- A strategy-based model with separable queues can be developed in this way, while the overtaking among passengers with different attractive sets of lines would violate the FIFO discipline of queues;
- Transit lines are conceived as a continuous flow of carriers and, as such, the representation of the delay due to the inherent service discontinuity is to be anyhow forced into the model; under this consideration, the under-saturation delay can be added wherever it is more convenient from a modelling point of view, in this case before the queuing process;
- In the *Route Choice Model*, the impedance of waiting is considered through a unique process represented by hyperarcs.

4.3.2. Supply models: basic supplementary nomenclature

$r_{a,d}(\tau)$: *conditional probability* of using arc $a \in A$ for passengers entering it at time τ and directed to destination $d \in N^C$, among the arcs of its tail's forward star;

$r_{h,d}(\tau)$: *conditional probability* of using hyperarc $a \in A^H$ for passengers entering it at time τ and directed to destination $d \in N^C$, among the hyperarcs of its tail's hyper-forward star;

$q_{i,d}^{dem}(\tau)$: *instantaneous demand flow* from node i to $d \in N^C$ at time τ ; it is >0 only if $i \in N^C$;

$q_{a,d}^{in}(\tau)$: *instantaneous inflow* of passengers entering arc $a \in A$ at time τ and directed to destination $d \in N^C$;

$q_a^{in}(\tau)$: *instantaneous inflow* of passengers entering arc $a \in A$ at time τ ;

$q_{a,d}^{out}(\tau)$: *instantaneous outflow* of passengers leaving arc $a \in A$ at time τ and directed to destination $d \in N^C$;

$q_a^{out}(\tau)$: *instantaneous outflow* of passengers leaving arc $a \in A$ at time τ ;

$q_a^{IN}(\tau)$: *cumulative inflow* of arc $a \in A$ at time τ , resulting from the network loading;

$q_a^{OUT}(\tau)$: *cumulative outflow* of arc $a \in A$ at time τ , consistent with its time-varying exit capacity;

$e_a(\tau)$: *instantaneous exit capacity* of arc $a \in A$ at time τ ;

$e_a^{CUM}(\tau)$: *cumulative exit capacity* of arc $a \in A$ at time τ ;

\mathbf{q}^{dem} : *vector of demand flows*;

\mathbf{q} : *vector of instantaneous arc (in/out) flows*;

\mathbf{g} : *vector of travel cost to destinations*;

\mathbf{r} : *vector of conditional probabilities*.

4.4. SUPPLY MODEL FOR STRATEGY-BASED DYNAMIC TRANSIT ASSIGNMENT

As anticipated in the previous section, the NFPM and the APF proposed in Meschini et al. (2007) and Papola et al. (2009) are extended here to strategy-based dynamic assignment.

4.4.1. NFPM for strategy-based dynamic transit assignment

The dynamic assignment and supply models are efficiently formulated using an implicit arc-based setting, rather than one based on hyperpaths. To this aim, the deterministic RCM presented in Chapter 3 can be re-formulated in the spirit of the *User Equilibrium* principle of Wardrop (1952), according to which the travel times of all used paths/hyperpaths between the same *od* pair are equal and minimal.

The complementary problem, formally known as *Wardrop inequalities*, used to express route choice is referred to the decision of perfectly rational passengers leaving node $i \in N$ at time τ and directed towards destination $d \in N^C$:

$$\begin{aligned} r_{a,d}(\tau) \cdot \left((t_a(\tau) - \tau) + g_{HD_a,d}(t_a(\tau)) - g_{TL_a,d}(\tau) \right) &= 0, \forall a \in FS_i \\ r_{a,d}(\tau) &\geq 0, \forall a \in FS_i \\ \sum_{a \in FS_i} r_{a,d}(\tau) &= 1 \end{aligned} \quad \begin{array}{l} \text{4-2} \\ \text{if } i \in N^S \end{array}$$

$$\begin{aligned} r_{h,d}(\tau) \cdot \left(w_h(\tau) + \sum_{a \in h} \left(p_{ah}(\tau) \cdot g_{HD_a,d}(t_{a|h}(\tau)) - g_{TL_a,d}(\tau) \right) \right) &= 0, \forall h \in HFS_i \\ r_{h,d}(\tau) &\geq 0, \forall h \in HFS_i \\ \sum_{h \in HFS_i} r_{h,d}(\tau) &= 1 \end{aligned} \quad \begin{array}{l} \text{4-2b} \\ \text{if } i \in N^S \end{array}$$

$$r_{a,d}(\tau) = \sum_{\substack{i=TL_a \\ h \in HFS_i}} (r_{h,d}(\tau) \cdot p_{ah}(\tau)), \forall a \in FS_i \quad \begin{array}{l} 4-2c \\ \text{if } i \in N^S \end{array}$$

For the reasons explained in 4.3.1, the NFPM must refer to the model graph G . Hence, equation 4-2c transforms *hyperarc* conditional probabilities into *waiting arc* conditional probabilities.

At this point, the flow can be propagated forward on the model graph, starting from the origin node(s). Once the intermediate node i is reached, the flow moves along its forward star, according to equation 4-3.

$$q_{a,d}^{in}(\tau) = r_{a,d}(\tau) \cdot \left(q_{i,d}^{dem}(\tau) + \sum_{b \in BS_i} q_{b,d}^{out}(\tau) \right), \quad a \in FS_i \quad 4-3$$

The inflow $q_{a,d}^{in}(\tau)$ entering arc $a \in A$ at time τ and directed to destination $d \in N^C$ is given by the arc *conditional probability* $r_{a,d}(\tau)$ multiplied by the flow on node i (i.e. TL_a) at time τ . The latter is given, in turn, by the sum of: the flows that leave each arc $b \in BS_i$ at time τ ; and the demand flow at the same time, $q_{i,d}^{dem}(\tau)$. Additionally, $q_{b,d}^{out}(\tau)$ is calculated as in equation 4-4 by applying the FIFO and flow conservation rules (Cascetta, 2009: p. 437).

$$q_b^{out}(\tau) = q_b^{in}(t_b^{-1}(\tau)) \cdot \frac{\partial t_b^{-1}(\tau)}{\partial \tau} \quad 4-4$$

Then, the total flow entering or leaving a at time τ is evaluated as in equation 4-5:

$$q_a^{in}(\tau) = \sum_{d \in N^C} q_{a,d}^{in}(\tau), \quad q_a^{out}(\tau) = \sum_{d \in N^C} q_{a,d}^{out}(\tau) \quad 4-5$$

4.4.2. Flow-independent APF

The APF of each $a \in A$ determines the temporal profile of the generalised travel time and, thus, the temporal profile of the exit time for any arc a and entry time τ . The APF depends on the flow of the considered arc and of its adjacent arcs at previous instants resulting from the NFPM. Thus, in general, the travel costs are not separable either in time or in space.

In this particular formulation, except for the queuing arcs, all other arcs have a flow-independent exit time provided by the following equations:

$$t_a(\tau) = \frac{\lambda_{ij}}{\rho_{ij}} + \tau \quad \begin{array}{l} 4-6 \\ a \in A^P \end{array}$$

$$t_a(\tau) = \theta_\ell \left(\theta_\ell^{-1}(\tau) \right) \quad \begin{array}{l} 4-7 \\ a \in A^L \end{array}$$

$$t_a(\tau) = \frac{1}{\varphi_a(\tau)} + \tau \quad \begin{array}{l} 4-8 \\ a \in A^W \end{array}$$

$$t_a(\tau) = \delta_a + \tau \quad \begin{array}{l} 4-9 \\ a \in A^D \cup A^A \cup A^Z \cup A^B \end{array}$$

where: $ij = (V_{TL_a}, V_{HD_a})$; $\ell = L_{HD_a} = L_{TL_a}$ is the line corresponding to HD_a and TL_a ; $s = s(L_{HD_a}, V_{HD_a})$; $s-1 = s(L_{TL_a}, V_{TL_a})$; δ_a is a constant representing alighting, boarding, dwelling time and, for algorithmic purposes, also the travel time on dummy arcs.

4.4.3. Bottleneck Queue Model with variable exit capacity

Note that equation 4-8 only considers the travel cost due to the inherent discontinuity of transit services, while the contribution due to overcrowding is not represented. In fact, the

queuing time is considered in the APF of queuing arcs, which in turn depends on the current length of the queue at the stop by means of a *Bottleneck Queue Model* with time-varying exit capacity. The mathematical framework of the model is significantly different from the physics of the phenomenon, which can be thought of as a ‘gate system’. As soon as passengers reach the stop, they join a queue that can be thought of as being behind a gate: whoever is at the front of the queue passes through the gate (so the queuing time due to congestion is over) and starts waiting for the next arrival (this is the under-saturation delay due to the discontinuity of the service). By contrast, the phenomenon is represented in the model in the inverse order (Figure 4-3):

- At first, passengers experience the under-saturation delay, which corresponds to the waiting time before the first carrier of any attractive line arrives;
- Then, in cases of overcrowding, they suffer a queuing time, which is graphically represented by a queuing arc.

The queuing time is calculated by means of a Bottleneck Queue Model with time-varying exit capacity (equations 4-10 to 4-14).

Although these equations may appear somewhat complicated, their algorithmic implementation is fairly straightforward (Section 5.3). Their conceptual explanation is provided here.

With reference to Figure 4-3, consider the boarding arc a_3 . Its exit capacity at τ coincides with the instantaneous capacity available on-board at the same time and may be calculated as in equation 4-10, where DAa is the dwelling arc that enters the same boarding node of a_3 , and the outflow is defined as in equation 4-4.

$$e_{a_3}(\tau) = \chi_a \cdot \varphi_a(\tau) - q_{DAa}^{out}(\tau) \quad 4-10$$

$$e_{a_2}(t_{a_3}^{-1}(\tau)) = \frac{e_{a_3}(\tau)}{\frac{\partial t_{a_3}^{-1}(\tau)}{\partial \tau}} \quad 4-11$$

For those who leave $a_2 \in A^Q$ at $t_{a_3}^{-1}(\tau)$ and leave $a_3 \in A^B$ at τ , the exit capacity $e_{a_2}(t_{a_3}^{-1}(\tau))$ does not coincide with the instantaneous capacity available on-board at the same time $e_{a_3}(t_{a_3}^{-1}(\tau))$; instead, the exit capacity $e_{a_3}(\tau)$ needs to be propagated backwards in time (4-11). The temporal profiles of the exit capacity and inflows of the queuing arc $a_2 \in A^Q$ are then used to obtain the cumulative values of exit capacity and inflows (4-12).

$$q_{a_2}^{IN}(\tau) = \int_0^{\tau} q_{a_2}^{in}(\mathcal{G}) d\mathcal{G}, \quad e_{a_2}^{CUM}(\tau) = \int_0^{\tau} e_{a_2}(\mathcal{G}) d\mathcal{G} \quad 4-12$$

$$q_{a_2}^{OUT}(\tau) = \min\{q_{a_2}^{IN}(\mathcal{G}) + e_{a_2}^{CUM}(\tau) - e_{a_2}^{CUM}(\mathcal{G}) : \mathcal{G} \leq \tau\} \quad 4-13$$

$$q_{a_2}^{IN}(\tau) - q_{a_2}^{OUT}(\tau) = \int_{t_{a_2}^{-1}(\tau)}^{\tau} e_{a_2}(\mathcal{G}) d\mathcal{G} \quad 4-14$$

At this point, the cumulative flow leaving arc $a_2 \in A^Q$ is calculated in the spirit of the Bottleneck Queue Model (4-13). If it is assumed that the queue at time τ began at a previous instant $\mathcal{G} \leq \tau$, then $q_{a_2}^{OUT}(\mathcal{G}) = q_{a_2}^{IN}(\mathcal{G})$, and from \mathcal{G} to τ the cumulative flow of passengers that leave the arc $a_2 \in A^Q$ is $e_{a_2}^{CUM}(\tau) - e_{a_2}^{CUM}(\mathcal{G})$, then equation 4-13 yields the cumulative number of passengers that have left the queue at time τ as the minimum among each cumulative outflow that would occur if the queue began at a previous instant $\mathcal{G} \leq \tau$.

On the other hand, the number of passengers queuing on arc $a_2 \in A^Q$ at time τ , which is $q_{a_2}^{IN}(\tau) - q_{a_2}^{OUT}(\tau)$, can be expressed (using equation 4-14) also as the integral of the exit

capacity e_{a_2} from $t_{a_2}^{-1}(\tau)$ to the exit time τ . The queuing time $t_{a_2}^{-1}(\tau) - \tau$ is consistent with the temporal profile of e_{a_2} and is the output of the above bottleneck model.

During this period of time, some vehicles of the line associated with $a_2 \in A^Q$ approach the stop, but queuing passengers cannot board them because of capacity constraints. Therefore, the number of vehicle passages that the passengers must let go before boarding is given by equation 4-15, where $\lfloor x \rfloor$ indicates the floor function of x .

$$\kappa_{a_2}(\tau) = 1 + \left\lfloor \int_{\tau}^{t_{a_2}^{-1}(\tau)} \varphi_{a_2}(\vartheta) d\vartheta \right\rfloor \quad 4-15$$

As seen in Chapter 3, this result is used as an input of the SM. Consider the arc $a \in h$ ($h \in A^H$) in the hypergraph of Figure 4-2 and consider the equivalent $a_1 \in A^W$ in the model graph of Figure 4-3. Then $\kappa_a(\tau) = \kappa_{a_2}(\tau)$.

4.5. DEMAND-SUPPLY INTERACTION MODEL: DYNAMIC USER EQUILIBRIUM

4.5.1. Formulation of the strategy-based dynamic transit-assignment model as a User Equilibrium

The extension of the first principle of Wardrop to a dynamic scenario allows for the formulation of the strategy-based dynamic transit-assignment model as a *User Equilibrium* (UE) that represents configurations in which no user can improve his/her travel cost at the time he/she is travelling by unilaterally changing hyperpath. The dynamic UE can be

specified as a Fixed-Point Problem (FPP) by combining the supply and demand models or, equivalently, the *Uncongested Network Assignment Map* (which combines the results of the RCM and the NFPM) and APF (Cascetta, 2009: pp. 305, 464–467), as done here.

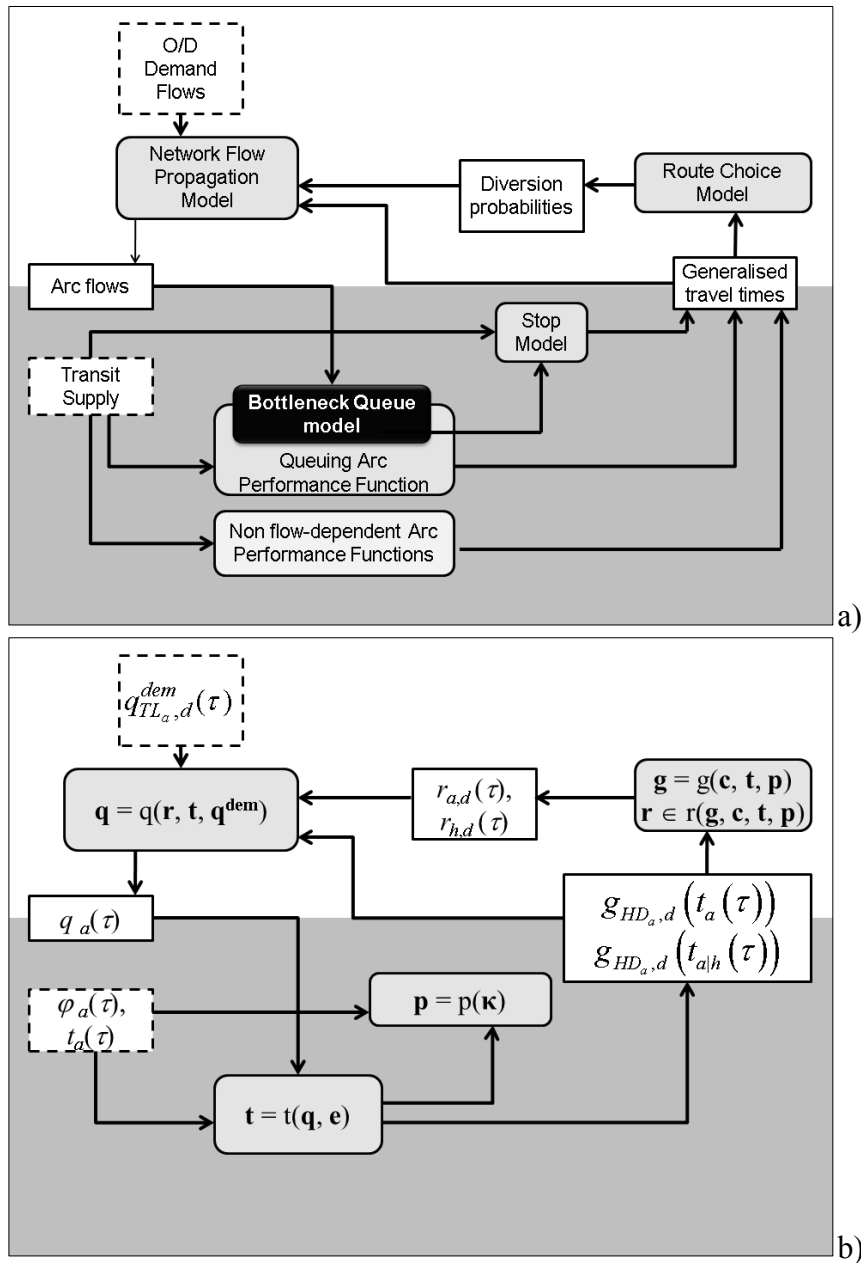


Figure 4-4
a) Scheme of the fixed-point formulation for the strategy-based dynamic transit-assignment model
b) Variables and models of the fixed-point formulation for the strategy-based dynamic transit-assignment model

Figure 4-4 shows the scheme, variables and models of the fixed-point formulation for the strategy-based dynamic transit assignment.

First of all, in order to formulate the dynamic UE with implicit path enumeration, the RCM (3-11) and the NFPM (4-3) are expressed in compact form respectively by equations 4-16, 4-17 (RCM) and 4-18, 4-19 (NFPM).

$$\mathbf{g} = \mathbf{g}(\mathbf{t}, \mathbf{p}) \quad 4-16$$

$$\mathbf{r} \in \mathbf{r}(\mathbf{g}, \mathbf{t}, \mathbf{p}) \quad 4-17$$

$$\mathbf{q} = \mathbf{q}(\mathbf{r}, \mathbf{t}, \mathbf{q}^{\text{dem}}) \quad 4-18$$

$$\mathbf{e} = \mathbf{e}(\mathbf{q}, \mathbf{t}) \quad 4-19$$

It should be noticed here that, because the RCM under consideration is deterministic, when more than one arc in the forward star of a node minimises the total travel time from that node to the destination, then the set of arc conditional probabilities solving equations 4-2 is not unique. Consequently, in equation 4-17, the symbol ‘=’ is substituted by ‘ \in ’. Also, the compact formulation of the RCM clearly shows that the vector of conditional probabilities (\mathbf{r}) also depends on the vector of diversion probabilities (\mathbf{p}).

The combination of the RCM and NFPM yields the Uncongested Network Assignment Map (Cascetta, 2009: p. 279), as in equation 4-20:

$$[\mathbf{q}, \mathbf{e}] \in \xi(\mathbf{t}, \mathbf{p}, \mathbf{q}^{\text{dem}}) \quad 4-20$$

On the other hand, the APFs also imply equations 4-21 and 4-22, while the diversion probabilities calculated by the SM may be expressed in compact form, as in 4-23. Thus, exit times and diversion probabilities, which represent the supply, are expressed as in equation 4-24.

$$\mathbf{t} = \mathbf{t}(\mathbf{q}, \mathbf{e}) \quad 4-21$$

$$\boldsymbol{\kappa} = \boldsymbol{\kappa}(\mathbf{t}) \quad 4-22$$

$$\mathbf{p} = \mathbf{p}(\boldsymbol{\kappa}) \quad 4-23$$

$$[\mathbf{t}, \mathbf{p}] = \mathbf{v}(\mathbf{q}, \mathbf{e}) \quad 4-24$$

Finally, the UE is obtained as a Fixed-Point Problem combining equation 4-20 and equation 4-24:

$$\mathbf{q} = \boldsymbol{\zeta}(\mathbf{q}) \quad 4-25$$

4.5.2. Characterisation of the network equilibrium

The existence of the equilibrium may be proved following Cascetta (2009: p. 378), because all the maps and functions, defined over the non-empty, compact and convex set of arc flows, are upper semi-continuous. Conversely, it is not possible to prove mathematically the uniqueness of the equilibrium because the problem does not have separable APFs (as in Cantarella et al. (2010)); but the generalised travel cost for queuing arcs depends on the link flow on the queuing arc considered, as well as on adjacent dwelling arcs (equations 4-10 to 4-14).

Finally, it is important to note that the formulation of a DUE, as in this case, implies the assumption that users have a full and correct perception of generalised travel times and choose travel alternatives with minimum cost. Thus, at the equilibrium, the same *od* pair may be connected by several minimal hyperpaths or the total generalised travel cost may be minimised through a split of the demand flows among different strategies, as shown in Cominetti and Correa (2001) and Schmöcker (2006).

4.6. DISCUSSION

This chapter proposes a supply model for strategy-based dynamic transit assignment and a demand–supply interaction model, for the same problem, in the form of a DUE.

As discussed in Section 4.2, the vast majority of models for FB assignment with capacity constraints do not consider within-day dynamics and thus are only able to produce *average* results during the analysis period that are not able to capture the build-up and dissipation of passengers' queues, and whose distortion becomes progressively more relevant as the demand is more peaked.

A relevant example of a supply model for dynamic transit assignment with hyperpaths, which overcomes the afore-mentioned flaw, is given by Schmöcker et al. (2008) (though it adopts a *discrete* time representation, for which a compromise between result accuracy and algorithm performances must be attained). The model relies on a Markovian process to represent the network loading and thus can be applied only if queuing passengers mingle at the stop/station, while it cannot reproduce the priority of those who are at the front of a FIFO queue with respect to those who are at the back.

Therefore, in this thesis the supply model is developed by extending to strategy-based transit assignment the supply model originally proposed by Meschini et al. (2007) for FB assignment *without* hyperpaths.

The main assumption of the supply model is that flows are macroscopic time-continuous functions and transit services are conceived as a continuous flow of supply with 'instantaneous capacity'. Therefore, the supplied capacity is not accounted for in terms of 'passengers per vehicle' (e.g. 120 passengers per vehicle), but in terms of passengers per time interval (e.g. 2 passengers per minute), and this hypothesis, though questionable from a

phenomenal point of view, is consistent with the basic assumption of the *frequency-based* models that passengers conceive all the runs of the same line as a unitary supply facility.

The extension of the supply model to the context of interest (strategy-based dynamic assignment) requires the following main methodological innovations:

- In order to ensure that inflows/outflows are consistent with arcs' exit times, the supply model is referred to a different graph structure (model graph G) than the demand model; indeed, in G , the waiting arcs $a \in A^W$ only represent the under-saturation delay, while the queuing time, as it results from the *Bottleneck Queue Model*, is represented by queuing arcs $a \in A^Q$;
- The queuing time calculated in the APF is exploited to calculate congestion parameters $\kappa_a(\tau)$, which are used, in turn, in the SM to calculate *diversion probabilities* and *total expected waiting time*;
- In order to guarantee a smooth transition from RCM to NFPM, the queuing phenomenon is represented in a reverse order (the queuing time is after the under-saturation delay); in this way, a model with separate queues can be developed and overtaking among passengers directed towards different destinations may be disregarded.

The dynamic assignment is regarded as a UE and formalised as a system between UNAM and APF, where the reciprocal consistency between flows and travel times is attained jointly at the equilibrium.

5. MODEL IMPLEMENTATION

In order to implement the strategy-based dynamic assignment procedure detailed in chapters 3 and 4, a solution algorithm is devised here, which extends to the dynamic setting the original formulation given by Nguyen and Pallottino (1988) and Spiess and Florian (1989) in their seminal works on static strategy-based transit assignment (Section 0).

The solution algorithm is detailed in sections 5.2 and 5.3. Then, in Section 5.4, it is applied to develop some worked examples to highlight the dynamic effects of passenger congestion on route choices and compare the different flow patterns on the network when different assumptions about the queuing mechanism and information provision are considered. Finally, the model implementation is complemented with an application to a real-scale network, which proves that the complexity of the mathematical framework devised in this thesis is compatible with the analysis of real scenarios.

5.1. SOLUTION ALGORITHM FOR STATIC AND UNCONGESTED STRATEGY-BASED TRANSIT ASSIGNMENT

The solution algorithm proposed for the strategy-based dynamic transit assignment inherits the general structure suggested by Spiess and Florian (1989) for the static and uncongested case. The original algorithm includes two parts:

- Part 1: *Hyperpath search*

For every possible destination, shortest all-to-one hyperpaths (or hypertrees) are found by scanning the network in reverse topological order, starting from the destination;

- Part 2: *Assign demand according to shortest hyperpaths*

For every possible origin, the travel demand is loaded by scanning the network in topological order, proceeding from each origin to the destination.

The variable list of the algorithm includes:

- d : destination node;
- o : origin node;
- i : generic node;
- N^O, N^D : set of origin and destination nodes; $N^O, N^D \subseteq N^C$;
- $H_{k,i,d}$: sub-hypergraph representing the hyperpath k connecting i to d (i may also be an origin node);
- FS_i : set of arcs belonging to the forward star of node i ;
- HFS_i : set of hyperarcs belonging to the hyper-forward star of node i , $i \in N^S$;
- a : generic arc / branch of hyperarc $a \in h$;
- b : generic arc belonging to the backward star of a node;
- h : generic hyperarc;
- $suc(i)$: successor arc/hyperarc of the generic node i ;
- φ_a : average service frequency corresponding to waiting arc a ;
- φ_i : cumulative frequency at node i , $i \in N^S$;
- c_a : travel time on arc a ;
- $p_{a|h}$: diversion probability;
- $g_{i,d}$: current travel cost from generic node i to destination d ;

- $g^*_{i,d}$: minimum travel cost from generic node i to destination d ;
- $q_{a,d}$: flow on arc a directed to destination d ;
- $q_{i,d}$: flow traversing node i and directed to destination d .

The solution algorithm for the static strategy-based transit assignment is detailed here.

- 1 Part 1 (Hypertree search)
- 2 $\forall d \in N^D$
- 3 Step 1.0 (Initialisation):
- 4 Set $g^*_{d,d} = 0, suc(d) = \emptyset$
- 5 $\forall i \in N \setminus \{d\}$
- 6 Set $g^*_{i,d} = \infty$
- 7 Step 1.1 (RCM):
- 8 $\forall i \in N \setminus \{d\}$ in reverse topological order
- 9 If $i \notin N^S$ AndAlso $g^*_{i,d} > g^*_{j,d} + c_a, a = (i, j) \in FS_i$, Then
- 10 $g_{i,d} = g^*_{j,d} + c_a$
- 11 If $g^*_{i,d} > g_{i,d}$ Then
- 12 $g^*_{i,d} = g_{i,d}$ And $suc(i) = a$
- 13 If $i \in N^S$ Then
- 14 Step 1.1.1 (determining the attractive set and waiting hyperarc h)

15 Sort waiting arcs $a \in FS_i$ in increasing order of $g_{HDa,d}$:

$$16 \quad g_{HDa_1,d} \leq g_{HDa_2,d} \leq g_{HDa_3,d} \dots \leq g_{HDa_n,d}, \quad n = |FS_i|$$

17 Set: $h = a_1$; $g^*_{i,d} = g_{j,d} + 1/\varphi_{a_1}$; $\varphi_i = \varphi_{a_1}$; $k = 2$

18 While ($k \leq n$ And $g_{HDa_k,d} < g^*_{i,d}$) Do

$$19 \quad \varphi_i = \varphi_i + \varphi_{bk}$$

$$20 \quad g^*_{i,d} = \frac{1}{\varphi_i} + p_{a_k|h} \cdot g^*_{HDa_k,d}$$

$$21 \quad h = h \cup \{a_k\}$$

$$22 \quad k = k + 1$$

$$23 \quad g_{i,d} = g^*_{i,d} \text{ And } suc(i) = h$$

24 Repeat Step 1.1 until no label can be further decreased

25 Part 2 (Assign demand according minimal hyperpath)

$$26 \quad \forall d \in N^D$$

27 Step 2.0 (Initialisation):

$$28 \quad \forall o \in N^O$$

$$29 \quad q_{o,d} = q_{o,d}^{dem}$$

$$30 \quad \forall a \in A: a = succ(o)$$

$$31 \quad q_{a,d} = q_{o,d}$$

32 Step 2.1 (Loading):

$$33 \quad \forall H_{k,o,d}$$

34 $\forall i \in N \cap H_{k,o,d}$ in topological order

35 If $i \notin N^S$ Then

36 $\forall b \in BS_i$

37 $q_{i,d} = q_{i,d} + q_{b,d}$

38 $a = succ(i)$

39 $q_{a,d} = q_{i,d}$

40 Else

41 $\forall b \in BS_i$

42 $q_{i,d} = q_{i,d} + q_{b,d}$

43 $h = succ(i)$

44 $\forall a \in h$

45 $q_{a,d} = q_{i,d} \cdot p_{a|h}$

46 $q_a = q_a + q_{a,d}$

5.2. SOLUTION ALGORITHM FOR DYNAMIC AND CONGESTED STRATEGY-BASED TRANSIT ASSIGNMENT

In the problem of interest, the presence of congestion requires the structure of the solution algorithm to be changed and the introduction of an additional feedback loop, to express the dependency of generalised travel times on flows. This loop, formalised with the fixed-point problem illustrated in Chapter 4, can be solved, as usual, by means of the Method of Successive Averages (MSA). Thus, the algorithm includes three parts:

- Part 0: *Initialisation of equilibrium arc flows to zero*
- Part 1: *Hypertree search (RCM and SM)*

For every possible destination, shortest all-to-one hyperpaths (or hypertrees) are found scanning the network in reverse topological order;

- Part 2: *Assign demand according to shortest hyperpaths (NFPM)*

For every possible origin, the travel demand is loaded by scanning the network in topological order; loading flows are obtained;

- Part 3: *MSA (FPP)*

Equilibrium flows are updated by means of an MSA; when equilibrium flows are known, travel times along queuing arcs and congestion parameters are updated;

- Part 4: *Convergence checks and stop criterion*

If the gap between the *total equilibrium travel time* (namely the sum of all travel times experienced by passengers moving across the network) and *loading travel time* (namely the sum of travel times calculated assuming that all passengers travel along optimal strategies only) is lower than a fixed quantity, then exit the loop; otherwise, repeat Part 1, Part 2 and Part 3 considering the updated values of equilibrium flows and congestion parameters.

Mutual consistency between link flows and arc performances is attained only at the equilibrium.

Beyond the feedback loop, which is needed to consider the effects of passenger congestion, the proposed algorithm differs from the original one because the considered problem is time-dependent. In order to consider explicitly the time dimension of supply and demand variables, which may change with the time of day, the *Decreasing Order of Time* (DOT) method, developed by Chabini (1998) solely for the Dynamic Shortest Path Problem (DSPP), is opportunely extended to solve the dynamic Deterministic User Equilibrium proposed for transit networks.

5.2.1. Decreasing Order of Time (DOT) method: extension

In assignment applications, the shortest (hyper)path problem needs to be solved for every possible *od* pair and arrival time. Thus, Chabini's DOT method, which has been analytically proved to be the most efficient solution method for the all-to-one search for every possible arrival time, is extended to the time-dependent shortest hyperpath problem.

Although the proposed model has a continuous time representation, a discrete time representation is adopted for its numerical solution. The main idea is to divide the analysis period $AP = [0, \Theta]$ into T time intervals, such that $AP = \{\xi^0, \xi^1, \dots, \xi^\tau, \dots, \xi^{T-1}\}$, with $\xi^0 = 0$ and $\xi^{T-1} = \Theta$, and to replicate the network along the time dimension, forming a time-expanded hypergraph H_T containing vertexes in the form (i, ξ) and edges in the form $((i, \xi), (j, t_{ij}(\xi)))$.

If time intervals are short enough to ensure that the exit time of a generic edge $t_{ij}(\xi^\tau)$ is not earlier than the next interval $\xi^{\tau+1}$, for $\tau \leq T-2$, it is ensured that the network is cycle-free

and the vertex chronological ordering is equivalent to the topological one. Thus, H_T is scanned starting from the last temporal layer to the value assumed for $\xi = \xi^0$ and, within the generic layer, no topological order needs to be respected. It is important to note here that, by processing the analysis period ‘layer by layer’ in reverse chronological order, the DOT method ensures that H_T does not need to be explicitly constructed and stored. Finally, because short time intervals are chosen, time-dependent variables are set to be constant over the same interval.

When a generic vertex (i, ξ^r) is visited, its (hyper-)forward star is scanned in order to set the minimal travel cost to destination and the successive (hyper-)edge by means of the generalised Bellman equation (equation 3-37).

Like the RCM, the NFPM is also solved efficiently by taking advantage of the absence of cycles in the time-expanded network. Therefore, when the demand flows are loaded, the network is scanned in chronological order, while no topological order needs to be respected within the same time interval (or layer).

5.2.2. Model graph of the solution algorithm

As specified in Chapter 4, although the demand and supply models are referred to two different graphs, they are conceived in such a way that only one graphic structure has to be built and stored for algorithmic purposes. It is displayed in Figure 5-2.

The main idea is that one type of arc (the queuing arc) is exploited in the supply model to represent the queuing time (an activity that in our model *follows* the under-saturation waiting time) and boarding time, while in the demand model the set of queuing

arcs represents the boarding time only (indeed, in the demand model, the queuing time is considered within the *total* waiting time at the stop w_{alh}).

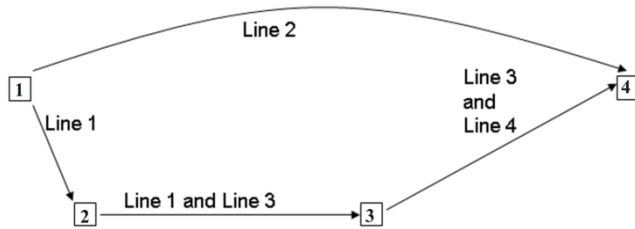


Figure 5-1
Base-graph representation of the small example network

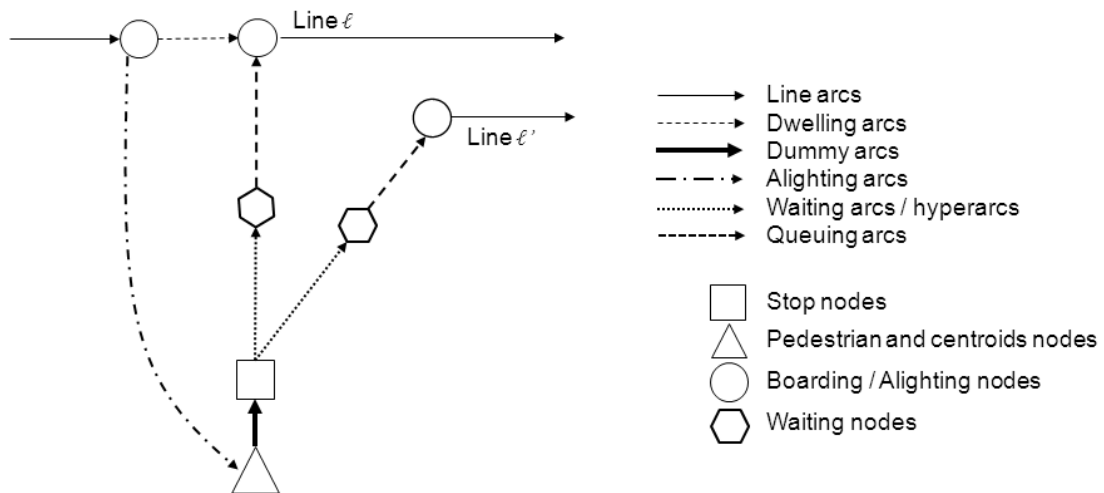


Figure 5-2
Graph representation of a stop shared by two lines
(While this is an intermediate stop for Line l , notice that there is no dwelling arc for Line l , as this is its departure terminal stop. This stop node might be equivalent to Stop 2 in Figure 5-1.)

5.3. ALGORITHM STRUCTURE

The following is a list of the algorithm's variables:

- τ : time interval index;

- τ_{INT} : time interval length;
- \mathcal{G} : index of the time interval when a queuing arc, entered at time interval τ , is left;
- d : destination node;
- i : generic node;
- N^D : set of destination nodes, $N^D \subset N^C$;
- N^O : set of origin nodes, $N^O \subset N^C$;
- FS_i : set of arcs belonging to the forward star of node i ;
- HFS_i : set of hyperarcs belonging to the hyper-forward star of node i ;
- $a = (i, j)$: generic arc and/or branch of hyperarc $a \in h$;
- b : generic arc belonging to the backward star of a node;
- χ_a : vehicle capacity of the line associated with arc $a \in A$;
- $DA_{a \in A^D}$: if $a \in A^O$, DA_a is the dwelling arc corresponding to it; if the considered queuing arc refers to the first stop of a line, then it has no corresponding dwelling arc;
- $WA_{a \in A^W}$: if $a \in A^O$, WA_a is the waiting arc corresponding to it (namely, the head node of WA_a corresponds to the tail node of a);
- h : generic hyperarc;
- $suc(i, \xi^\tau)$: successor arc and/or hyperarc of the generic node i at time interval ξ^τ ;
- $c_a(\xi^\tau)$: generalised travel time on arc a at time interval ξ^τ ;
 $\forall a \in A^D \cup A^Z \cup A^A \cup A^B \cup \delta_a = \tau_{\text{INT}}$
- $\varphi_a(\xi^\tau)$: instantaneous frequency corresponding to the line associated with arc a at time interval ξ^τ ;
- $z_a(\xi^\tau)$: queuing time on arc a at time interval ξ^τ ;
- $t_a(\xi^\tau)$: exit time from arc a for users entering it at time interval ξ^τ ;

- $t_a^{-1}(\tau)$: entry time to the arc a for users exiting it at time τ ;
- $\kappa_a(\xi^\tau)$: congestion parameter at time interval ξ^τ for the line L_{HDa} associated with the arc $a \in A^Q$;
- $p_{a|h}(\xi^\tau)$: diversion probability at time interval ξ^τ ;
- $w_{a|h}(\xi^\tau)$: conditional expected waiting time at time interval ξ^τ ;
- $w_h(\xi^\tau)$: waiting time at node $i = TL_h$ at time interval ξ^τ ;
- $g_{i,d}(\xi^\tau)$: current travel cost from generic node i to destination d at time interval ξ^τ ;
- $g_{i,d}^*(\xi^\tau)$: minimum travel cost from generic node i to destination d at time interval ξ^τ ;
- $g_{i,d}^{* \text{ stat}}$: minimum travel cost from generic node i to destination d at time interval $\xi^\tau \geq \xi^{T-1}$ (the value is calculated following Step 1.1 of the static and uncongested strategy-based transit assignment, as detailed in 0);
- $r_{a,d}(\xi^\tau)$: conditional probability of using arc $a \in A$ for passengers entering it at time interval ξ^τ and directed to destination $d \in N^C$, among the arcs of its tail's forward star;
- $r_{h,d}(\xi^\tau)$: conditional probability of using hyperarc $a \in F$ for passengers entering it at time interval ξ^τ and directed to destination $d \in N^C$, among the hyperarcs of its tail's forward star;
- $q_{i,d}^{\text{dem}}(\xi^\tau)$: instantaneous demand flow from $i \in N$ to $d \in N^C$ at time interval ξ^τ ; this can be greater than 0 only if i is an origin;
- $q_{i,d}(\xi^\tau)$: instantaneous loading flow traversing node i and directed to $d \in N^C$ at time interval ξ^τ ;
- $q_i(\xi^\tau)$: total instantaneous loading flow traversing node i at time interval ξ^τ ;
- $q_{a,d}^{\text{in}}(\xi^\tau)$: instantaneous loading inflow of passengers entering arc $a \in A$ at time interval ξ^τ and directed to destination $d \in N^C$;

- $q_{a,d}^{out}(\xi^\tau)$: instantaneous loading outflow of passengers leaving arc $a \in A$ at time interval ξ^τ and directed to destination $d \in N^C$;
- $q_a^{in}(\xi^\tau)$: instantaneous loading inflow of passengers entering arc $a \in A$ at time interval ξ^τ ;
- $q_a^{out}(\xi^\tau)$: instantaneous loading outflow of passengers leaving arc $a \in A$ at time interval ξ^τ ;
- $\hat{q}_a^{in}(\xi^\tau, iter)$: instantaneous equilibrium inflow of passengers entering arc $a \in A$ at time interval ξ^τ and iteration $iter$;
- $\hat{q}_a^{out}(\xi^\tau, iter)$: instantaneous equilibrium outflow of passengers leaving arc $a \in A$ at time interval ξ^τ and iteration $iter$;
- $q_a^{IN}(\xi^\tau, iter)$: cumulative equilibrium inflow of passengers entering arc $a \in A$ at time interval ξ^τ and iteration $iter$;
- $q_a^{OUT}(\xi^\tau, iter)$: cumulative equilibrium outflow of passengers leaving arc $a \in A$ at time interval ξ^τ and iteration $iter$;
- $e_a(\xi^\tau)$: instantaneous exit capacity for arc $a \in A$ at time interval ξ^τ ;
- $e_a^{CUM}(\xi^\tau)$: cumulative exit capacity for arc $a \in A$ at time interval ξ^τ ;
- $n_a(\xi^\tau)$: number of passengers queuing on arc $a \in A^Q$ at the end of time interval ξ^τ ;
- \mathbf{q}^{in} : vector of instantaneous loading inflow;
- $\hat{\mathbf{q}}^{in}(iter)$: vector of instantaneous equilibrium inflow;
- \mathbf{q}^{out} : vector of instantaneous loading outflow;
- $\hat{\mathbf{q}}^{out}(iter)$: vector of instantaneous equilibrium outflow;
- \mathbf{c} : vector of generalised travel times.

5.3.1. Part 1: Demand model (RCM and SM)

The solution algorithm for the time-dependent all-to-one shortest hyperpath problem for every possible arrival time is detailed here.

Considering the supply input in terms of line frequencies and in-vehicle travel times, as well as congestion parameters, allows the calculation of diversion probabilities and total waiting times as in equations 3-2 and 3-29; while the generalised travel time from an intermediate node to the destination is determined by means of equation 3-37 and the attractive set is determined by means of equation 3-38.

1 Step 1.0 (Static pre-processing – Initialisation):

2 $\forall i \in N \setminus \{d\}$

3 Calculate $g^*_{i,d}(\xi^{T-1}) = g^*_{i,d}{}^{\text{stat}}$

4 $\forall \tau \in [0, T-2]$

5 Set $g^*_{d,d}(\xi^\tau) = 0, \text{suc}(d, \xi^\tau) = \emptyset$

6 $\forall i \in N \setminus \{d\}$

7 Set $g^*_{i,d}(\xi^\tau) = \infty$

8 Step 1.1 (Hyperarcs' dynamic attributes – SM):

9 $\forall \tau \in [0, T-2]$

10 $\forall i \in N^S$

11 $\forall h \in HFS_i$

12 $\forall a \in h$

13 Calculate $p_{a|h}(\xi^r)$ using equation 3-2

14 Calculate $w_{a|h}(\xi^r)$ using equation 3-28

15
$$w_h(\xi^r) = w_h(\xi^r) + w_{a|h}(\xi^r) \cdot p_{a|h}(\xi^r)$$

16 Step 1.2 (Select the hypertree with minimal generalised travel time – RCM):

17 $\forall \tau \in [T-2, 0]$ Step -1

18 $\forall i \in N \setminus \{d\}$

19 If $i \in N^S, \forall h \in HFS_i$

20 $\forall a \in h$

21
$$t_{a|h}(\xi^r) = \lceil [w_{a|h}(\xi^r) / \tau_{INT}] \rceil + \xi^r$$

22 If $t_{a|h}(\xi^r) \leq \xi^T$ and $g_{HD_{a,d}}(t_{a|h}(\xi^r)) < \infty$ Then

23
$$g_{i,d}(\xi^r) = w_h(\xi^r) + p_{a|h}(\xi^r) \cdot g_{HD_{a,d}}(t_{a|h}(\xi^r))$$

24 Else

25
$$g_{i,d}(\xi^r) = w_h(\xi^r) + p_{a|h}(\xi^r) \cdot g_{HD_{a,d}}^{stat}$$

26 If $g^*_{i,d}(\xi^r) > g_{i,d}(\xi^r)$ Then

27
$$g^*_{i,d}(\xi^r) = g_{i,d}(\xi^r) \text{ And } suc(i, \xi^r) = h$$

28 ElseIf $i \notin N^S, \forall a \in FS_i$ Then

29
$$t_a(\xi^r) = \lceil [c_a(\xi^r) / \tau_{INT}] \rceil + \xi^r$$

30
$$g_{i,d}(\xi^r) = c_a(\xi^r) + g_{j,d}(t_a(\xi^r))$$

31 If $g^*_{i,d}(\xi^r) > g_{i,d}(\xi^r)$ Then

$$g^*_{i,d}(\xi^r) = g_{i,d}(\xi^r) \text{ And } \text{suc}(i, \xi^r) = a$$

5.3.2. Part 2: Supply model (NFPM and APF)

Part 1 of the solution algorithm finds the all-to-one hyperpaths for every possible destination; then Part 2 loads the *od* travel demand on the minimal (single) hyperpath connecting the considered pair of nodes. Although the deterministic RCM implies that *all* the optimal hyperpaths are used at the equilibrium, iteration by iteration only *one* minimal hyperpath is considered in the solution algorithm and loaded accordingly (this is also why mutual consistency between link flows and arc performances is attained only at the equilibrium).

In order to obtain this result, the time-expanded network is scanned in chronological order starting from the first time interval. When a generic vertex (corresponding to the generic node i at time τ) is reached, all the flow directed towards destination d is loaded on its successor arc (or hyperarc, in cases where the vertex corresponds to a stop).

Finally, if the considered vertex corresponds to a queuing node of the model graph, the Bottleneck Queue Model is applied to evaluate the queuing time and, thus, the time at which the flow entering the successor arc at τ will leave the arc.

The pseudo-code of the supply model is detailed here:

- 1 Step 2.1 – Initialisation and Demand Loading:
- 2 $\forall \tau \in [0, T-1]$
- 3 $\forall a \in A$

4 $q_a^{in}(\xi^\tau) = q_a^{out}(\xi^\tau) = 0$

5 $\forall d \in N^D$

6 $\forall \tau \in [0, T-1]$

7 $\forall i \in N$

8 $q_{i,d}(\xi^\tau) = 0$

9 $\forall a \in A$

10 $q_{a,d}^{in}(\xi^\tau) = q_{a,d}^{out}(\xi^\tau) = 0$

11 $\forall o \in N^O$

12 $q_{o,d}(\xi^\tau) = q_{o,d}(\xi^\tau) + q_{o,d}^{dem}(\xi^\tau)$

13 Step 2.2 – Assign demand according to shortest hyperpaths to obtain loading flows:

14 $\forall \tau \in [0, T-1]$

15 $\forall i \in N$

16 If $i \in N^S$ Then

17 $\forall d \in N^D$

18 $\forall b \in BS_i$

19 $q_{i,d}(\xi^\tau) = q_{i,d}(\xi^\tau) + q_{b,d}^{out}(\xi^\tau)$

20 $h = succ(i)$

21 $\forall a \in h$

22 $q_{a,d}^{in}(\xi^\tau) = q_{i,d}(\xi^\tau) \cdot p_{a|h}(\xi^\tau)$

$$23 \quad q_{a,d}^{out} \left(\xi^\tau + \left\| \frac{\varphi_a(\xi^\tau)}{\tau_{INT}} \right\| \right) = q_{a,d}^{in}(\xi^\tau), \text{ where } \|\cdot\| \text{ indicates the 'whole part' of a number}$$

$$24 \quad q_a^{in}(\xi^\tau) = q_a^{in}(\xi^\tau) + q_{a,d}^{in}(\xi^\tau)$$

$$25 \quad q_a^{out} \left(\xi^\tau + \left\| \frac{\varphi_a(\xi^\tau)}{\tau_{INT}} \right\| \right) = q_a^{out} \left(\xi^\tau + \left\| \frac{\varphi_a(\xi^\tau)}{\tau_{INT}} \right\| \right) + q_{a,d}^{out} \left(\xi^\tau + \left\| \frac{\varphi_a(\xi^\tau)}{\tau_{INT}} \right\| \right)$$

26 ElseIf $i \in N^Q$ Then

$$27 \quad \forall d \in N^D$$

$$28 \quad \forall b \in BS_i$$

$$29 \quad q_{i,d}(\xi^\tau) = q_{i,d}(\xi^\tau) + q_{b,d}^{out}(\xi^\tau)$$

$$30 \quad a = succ(i)$$

$$31 \quad q_{a,d}^{in}(\xi^\tau) = q_{i,d}(\xi^\tau)$$

$$32 \quad q_a^{in}(\xi^\tau) = q_a^{in}(\xi^\tau) + q_{a,d}^{in}(\xi^\tau)$$

33 Calculate $q_a^{out}(\xi^{\tau+1})$ by means of the Bottleneck Queue Model

$$34 \quad \forall d \in N^D$$

$$35 \quad q_{a,d}^{out}(\xi^{\tau+1}) = q_a^{out}(\xi^{\tau+1}) \cdot \frac{q_{i,d}(\xi^\tau)}{q_i(\xi^\tau)}$$

36 Else

$$37 \quad \forall d \in N^D$$

$$38 \quad \forall b \in BS_i$$

$$39 \quad q_{i,d}(\xi^\tau) = q_{i,d}(\xi^\tau) + q_{b,d}^{out}(\xi^\tau)$$

40 $a = succ(i)$

41 $q_{a,d}^{in}(\xi^\tau) = q_{i,d}(\xi^\tau)$

42 $q_{a,d}^{out}(t_a(\xi^\tau)) = q_{a,d}^{in}(\xi^\tau)$

43 $q_a^{in}(\xi^\tau) = q_a^{in}(\xi^\tau) + q_{a,d}^{in}(\xi^\tau)$

44 $q_a^{out}(t_a(\xi^\tau)) = q_a^{out}(t_a(\xi^\tau)) + q_{a,d}^{out}(t_a(\xi^\tau))$

Bottleneck Queue Model on queuing arc $a \in A^Q$ at time interval ξ^τ .

1 If $DA_a \neq \emptyset$ Then

2 $e_a(\xi^{\tau+1}) = \chi_a \cdot \varphi_a(\xi^{\tau+1}) - q_{DA_a}^{in}(t_{DA_a}^{-1}(\xi^{\tau+1}))$

3 Else

4 $e_a(\xi^{\tau+1}) = \chi_a \cdot \varphi_a(\xi^{\tau+1})$

5 If $n_a(\xi^{\tau-1}) = 0$ Then

6 If $q_a^{in}(\xi^\tau) \leq e_a(\xi^{\tau+1})$ Then

7 $q_a^{out}(\xi^{\tau+1}) = q_a^{in}(\xi^\tau)$

8 Else

9 $q_a^{out}(\xi^{\tau+1}) = e_a(\xi^{\tau+1})$

10 $n_a(\xi^\tau) = \max \left\{ 0, \left(q_a^{in}(\xi^\tau) - q_a^{out}(\xi^{\tau+1}) \right) \cdot \tau_{INT} \right\}$

11 Else

$$12 \quad q_a^{out}(\xi^{\tau+1}) = \min \left\{ e_a(\xi^{\tau+1}), \left(\frac{n_a(\xi^{\tau-1})}{\tau_{INT}} + q_a^{in}(\xi^\tau) \right) \right\}$$

$$13 \quad n_a(\xi^\tau) = \max \left\{ 0, \left(n_a(\xi^{\tau-1}) - q_a^{out}(\xi^{\tau+1}) \cdot \tau_{INT} \right) \right\}$$

5.3.3. Part 3: MSA (FPP)

The Method of Successive Averages has been extensively exploited to solve assignment problems formulated as fixed-point problems (FPP), and available examples include Cepeda et al. (2006), Meschini et al. (2007) and Schmöcker et al. (2008). The MSA is therefore also used in this case to solve the FPP that combines the Uncongested Network Assignment Map (UNAM) and the Arc Performance Functions (APF) (Cascetta, 2009: pp. 305, 464–467).

At each iteration, total flows across the network (equilibrium flows) are updated by averaging flows loaded in the current iteration on shortest hyperpaths only (loading flows) with the results obtained over all the past iterations (Step 3.1). Then, equilibrium flows are used to evaluate congestion parameters and travel times on queuing arcs (Step 3.2).

The pseudo-code of the MSA and the update of network performances ($\kappa_a(\tau)$) are detailed here:

1 Step 3.1 (MSA):

2 $\forall a \in A$

3 $\forall \tau \in [0, T-1]$

4 $\hat{q}_a^{in}(\xi^\tau, iter) = \hat{q}_a^{in}(\xi^\tau, iter-1) + (1/iter) \cdot (q_a^{in}(\xi^\tau) - \hat{q}_a^{in}(\xi^\tau, iter-1))$

$$5 \quad \hat{q}_a^{out}(\xi^\tau, iter) = \hat{q}_a^{out}(\xi^\tau, iter-1) + (1/iter) \cdot (q_a^{out}(\xi^\tau) - \hat{q}_a^{out}(\xi^\tau, iter-1))$$

6 Step 3.2 – Calculation of congestion parameters and travel time on queuing arcs:

$$7 \quad \forall a \in A$$

$$8 \quad \hat{q}_a^{IN}(\xi^0, iter) = \hat{q}_a^{in}(\xi^0, iter)$$

$$9 \quad \hat{q}_a^{OUT}(\xi^0, iter) = \hat{q}_a^{out}(\xi^0, iter)$$

$$10 \quad z_a(\xi^0) = 0$$

$$11 \quad \forall \tau \in [1, T-1]$$

$$12 \quad \hat{q}_a^{IN}(\xi^\tau, iter) = \hat{q}_a^{IN}(\xi^{\tau-1}, iter) + \hat{q}_a^{in}(\xi^\tau, iter)$$

$$13 \quad \hat{q}_a^{OUT}(\xi^\tau, iter) = \hat{q}_a^{OUT}(\xi^{\tau-1}, iter) + \hat{q}_a^{out}(\xi^\tau, iter)$$

14 If $\tau < T-1$ Then

$$15 \quad \mathcal{G} = \tau + 1$$

$$16 \quad \hat{q}_a^{OUT}(\xi^\mathcal{G}, iter) = \hat{q}_a^{OUT}(\xi^\tau, iter) + \hat{q}_a^{out}(\xi^\mathcal{G}, iter)$$

17 Do Until $\hat{q}_a^{OUT}(\xi^\mathcal{G}, iter) \geq \hat{q}_a^{IN}(\xi^\tau, iter)$ or $\tau \geq T-1$

18 If $\tau < T-1$ Then

$$19 \quad \mathcal{G} = \mathcal{G} + 1$$

$$20 \quad \hat{q}_a^{OUT}(\xi^\mathcal{G}, iter) = \hat{q}_a^{OUT}(\xi^{\mathcal{G}-1}, iter) + \hat{q}_a^{out}(\xi^\mathcal{G}, iter)$$

21 If $\mathcal{G} < T-1$ Then

$$22 \quad z_a(\xi^\tau) = (\xi^\mathcal{G} - \xi^\tau) \cdot \tau_{INT}$$

$$23 \quad \kappa_a(\xi^\tau) = 1 + \lfloor z_a(\xi^\tau) \cdot \varphi_a(\xi^\tau) \rfloor$$

24 Else

$$25 \quad n_a(\xi^{T-1}) = \max \{0, \hat{q}_a^{IN}(\xi^\tau, iter) - \hat{q}_a^{OUT}(\xi^{T-1}, iter)\}$$

$$26 \quad e_a(\xi^T) = \chi_a \cdot \varphi_a(\xi^T)$$

$$27 \quad z_a(\xi^\tau) = (\xi^{T-1} - \xi^\tau) \cdot \tau_{INT} + \left\| \frac{n_a(\xi^{T-1})}{ne_a(\xi^T)} \right\|$$

$$28 \quad \kappa_a(\xi^\tau) = 1 + \lfloor z_a(\xi^\tau) \cdot \varphi_a(\xi^\tau) \rfloor$$

29 Else

$$30 \quad n_a(\xi^{T-1}) = \hat{q}_a^{IN}(\xi^{T-1}, iter) - \hat{q}_a^{OUT}(\xi^{T-1}, iter)$$

$$31 \quad e_a(\xi^T) = \chi_a \cdot \varphi_a(\xi^T)$$

$$32 \quad z_a(\xi^\tau) = (\xi^T - \xi^\tau) \cdot \tau_{INT} + \left\| \frac{n_a(\xi^{T-1})}{ne_a(\xi^T)} \right\|$$

$$33 \quad \kappa_a(\xi^\tau) = 1 + \lfloor z_a(\xi^\tau) \cdot \varphi_a(\xi^\tau) \rfloor$$

5.3.4. Part 4: Convergence check and stop criterion

The total generalised travel time on shortest hyperpaths is calculated by summing the generalised travel time of all arcs included in the shortest hyperpaths, weighted by the loading flows – as resulted from the last iteration – on the same arcs. The total generalised travel time on the network is calculated by summing the generalised travel time of all arcs of the network, weighted by the equilibrium flows on the same arcs.

If the total generalised travel time on shortest hyperpaths is equal or close to the total generalised travel time on the network, a solution to the UE assignment, formulated as an FPP, has been found. Indeed, the algorithm has reached a point where the network performances, as expressed by frequencies, travel times and congestion levels, produce mutually consistent route choices and, thus, arc flows. On the other hand, if this is not the case, the new values of network performances (more specifically, queuing times and congestion parameters) are exploited in the following iteration as an input of the RCM.

Thus, if the network is not congested, the algorithm will terminate immediately because the generalised travel times calculated with the empty network have the same value as those calculated when the network is loaded. However, if the demand exceeds the supply, the MSA may take many iterations before attaining convergence and, for this reason, for practical applications, ε is usually set equal to a value between 0.001 (Cepeda et al., 2006; Meschini et al., 2007) and 0.1 (Schmöcker et al., 2008).

The pseudo-code of the convergence check and stop criterion is detailed here:

```

1   If  $|\hat{\mathbf{q}}^{in}(iter) \cdot \mathbf{c}(iter) - \mathbf{q}^{in} \cdot \mathbf{c}(iter)| \leq \varepsilon$  Then
2       STOP
3   Else
4        $\forall a \in A^Q$ 
5            $\forall \tau \in [1, T-1]$ 
6                $\kappa_{WAa}(\xi^\tau) = \kappa_a(\xi^\tau)$ 
7       Repeat Part 1, Part 2 and Part 3

```

5.4. WORKED EXAMPLES

The solution algorithm detailed in the previous section has been applied to solve strategy-based dynamic assignment problems for the example network depicted (base graph) in Figure 5-1.

In order to highlight the different hyperpath selection when passenger queues arise, travel times and frequencies are assumed to stay constant during the analysis period [07:30 – 09:00] and are displayed in Table 5-1. The length of the time intervals ξ^r is set to be one minute and it is assumed that time-dependent variables stay constant over each time interval. Thus, in order to ensure algorithm precision, dummy arcs are also supposed to have a travel time of one minute.

Table 5-1, together with the vehicle capacity. Moreover, it is assumed that the alighting, boarding and dwelling time is one minute.

The length of the time intervals ξ^r is set to be one minute and it is assumed that time-dependent variables stay constant over each time interval. Thus, in order to ensure algorithm precision, dummy arcs are also supposed to have a travel time of one minute.

Table 5-1
Frequencies, in-vehicle travel times and vehicle capacity of the lines in the small example network of Figure 5-1

| Line | Connection | Frequency (vehicles/min) | In-vehicle travel time (min) | Vehicle capacity (places) |
|------|-----------------|--------------------------|------------------------------|---------------------------|
| 2 | Stop 1 – Stop 4 | 1/6 | 25 | 50 |
| 1 | Stop 1 – Stop 2 | 1/6 | 7 | 50 |
| 1 | Stop 2 – Stop 3 | 1/6 | 6 | 50 |
| 3 | Stop 2 – Stop 3 | 1/15 | 4 | 50 |
| 3 | Stop 3 – Stop 4 | 1/15 | 4 | 50 |
| 4 | Stop 3 – Stop 4 | 1/3 | 10 | 25 |

Table 5-2

Time-dependent *od* matrix1 during the analysis period (the travel demand is expressed in passenger/min)

| Origin Centroid | Destination Centroid | Travel Demand [07:30 – 09:00] |
|-----------------|----------------------|-------------------------------|
| 17 | 16 | 5 |
| 18 | 16 | 7 |
| 19 | 16 | 7 |
| 20 | 16 | 0 |

In the first instance studied, the only destination considered is node 16 (see Figure 5-3) and it is assumed that the *od* matrix is in the form given by

Table 5-2.

In this setting, at the beginning of the analysis period [07:30 – 07:55], no congestion phenomenon occurs in the network and the model yields the same results that could be obtained by applying static models as in Nguyen and Pallottino (1988) and Spiess and Florian (1989).

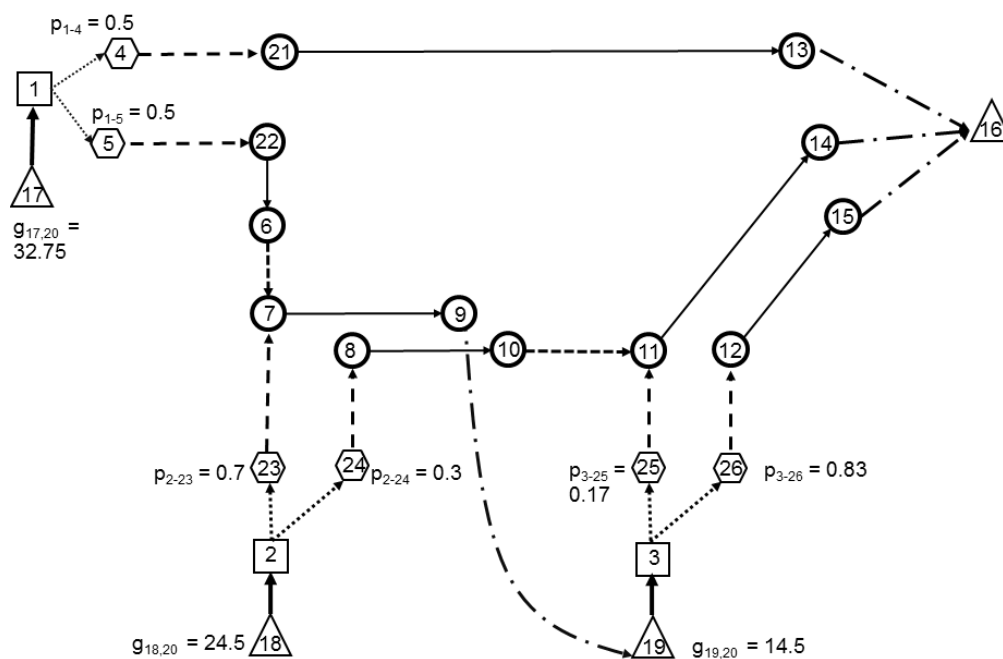


Figure 5-3

Shortest hypertree, diversion probabilities and travel times to destination (16) when the network is not congested [07:30 – 07:55] (For clarity reasons, the diversion probability is here indicated as $p_{i,j}$, where i is the stop node at which passengers are waiting and j is the waiting node corresponding to the attractive line considered.)

Figure 5-3 shows the hypergraph representation of the all-to-one shortest hyperpaths in uncongested conditions. Nodes $1, 2, 3 \in N^S$ and represent, respectively, Stop 1, Stop 2 and Stop 3 of Figure 5-1 while nodes $17, 18, 19, 16 \in N^C$ and represent centroids connected, respectively, to stops 1, 2, 3 and 4 of Figure 5-1. Also, the two different route sections connecting Stop 2 and Stop 3 are represented by distinct line arcs (7, 9) and (8, 10), and similarly the two route sections connecting Stop 3 and Stop 4 are represented by distinct line arcs (11, 14) and (12, 15).

When congestion occurs, un-congested and static models would not be able to reproduce the dynamic phenomenon of formation and dispersion of FIFO queues, nor its effect on route choice, as detailed in the following.

The passengers at origin node 17 at 07:30 who board Line 1 will reach node 7 at 07:46, where they will be joined by those passengers who, from stop node 2, board the same line. These travellers have to disembark through arc (9, 19) and reach Stop 3 at 07:54. Therefore, from this moment onwards, the total flow from Stop 3 to destination exceeds the total available capacity and, when the successive vehicles of Line 3 and Line 4 arrive, these two services become heavily congested (Figure 5-4a and Figure 5-4c).

The decreased available capacity of Line 3, combined with a lower frequency of the service, determines a fall of its diversion probability, while the diversion probability of Line 4 increases (Figure 5-4b).

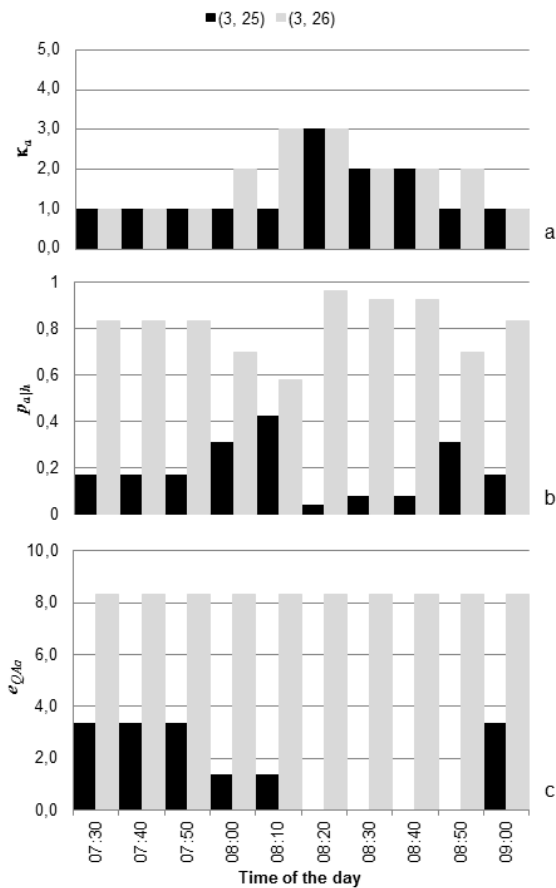


Figure 5-4 Variation of congestion factor κ_a (a), diversion probability $p_{a/h}$ (b) and instantaneous exit capacity $e_{Q_{Aa}}$ (c) at Stop 3; a represents the waiting arcs for Line 3 and Line 4 and Q_{Aa} the corresponding queuing arcs for the same lines at Stop 3

It is important to notice here that the value of the diversion probability solely depends on the frequency of the line and on its congestion level at the considered stop. On the other hand, the inclusion of a line in the attractive set depends on its total travel time upon boarding.

The analysis of congestion patterns at Stop 3 suggests that the model is able to simulate ‘forward effects’ – namely effects produced by what happened upstream in the network at an earlier time of the day (passengers boarding Line 1 at 07:30) – on what happens downstream at a later time (the queue of passengers, wishing to board Line 3, that occurs at Stop 3 at 07:55).

Additionally, the model also simulates ‘backward effects’ – namely effects produced by what is expected to happen downstream in the network at a later time – on what happens upstream at an earlier time. The analysis of Line 1 helps to clarify this concept.

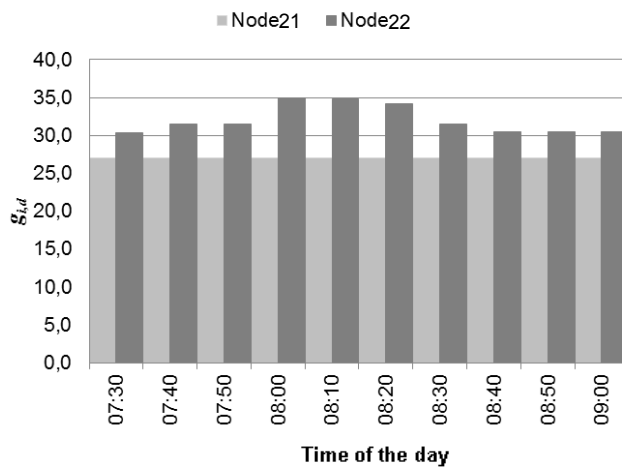


Figure 5-5
Travel time to destination ($g_{i,d}$) upon boarding Line 1 (boarding node 21) or Line 2 (boarding node 21) from Stop 1 during the analysis period [07:30 – 09:00]

Line 1 never becomes congested at stops 1 or 2 (Figure 5-6a and Figure 5-7a). However, because a long queue for Line 3 arises at Stop 3 at 08:12, then, after 07:53, the travel time upon boarding Line 1 from Stop 1 increases to 35 minutes (Figure 5-5) and this line is thus excluded from the attractive set of Stop 1 (Figure 5-6b). Line 1 is included again from 08:25 onwards, namely when the travel time upon boarding decreases again, because by the time Stop 3 is reached (08:44), congestion on Line 3 will have dissipated.

Similarly, at Stop 2 (Figure 5-7b) starting from 08:00 passengers would board only Line 3. Should they board Line 1, they would reach Stop 3 at 08:09, when a queue for boarding Line 4 arises and, consequently, the travel time upon boarding Line 1 increases to 23.4 minutes. Afterwards (at 08:33, Figure 5-7a), because Line 3 becomes congested at Stop 2, Line 1 is reintroduced into the attractive set of Stop 2 (Figure 5-7b).

Figure 5-6c and Figure 5-7c complete the example and respectively depict the available capacity of Line 1 and Line 2 at Stop 1 and Line 1 and Line 3 at Stop 2.

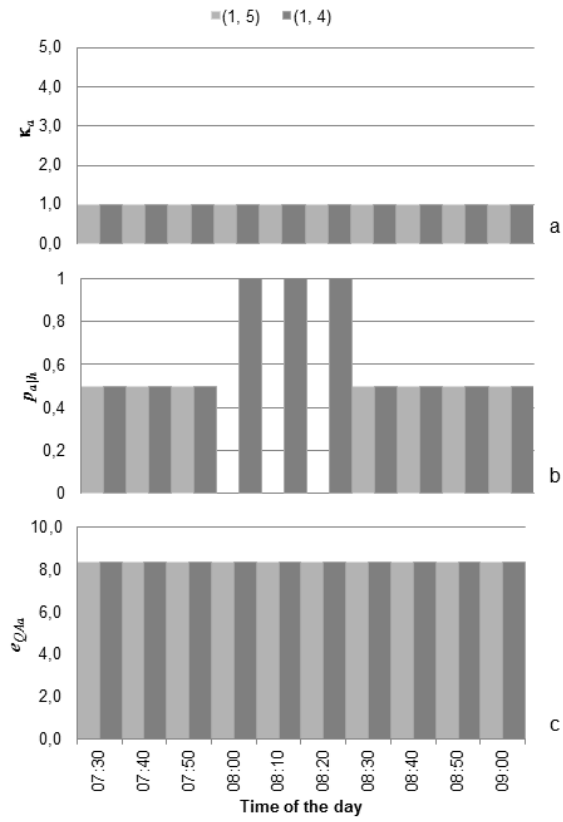


Figure 5-6

Variation of congestion factor κ_a (a), diversion probability $p_{a|h}$ (b) and instantaneous exit capacity e_{QAa} (c) at Stop 1; a represents the waiting arcs for Line 1 and Line 2 and QA_a the queuing arcs for the same lines at Stop 1

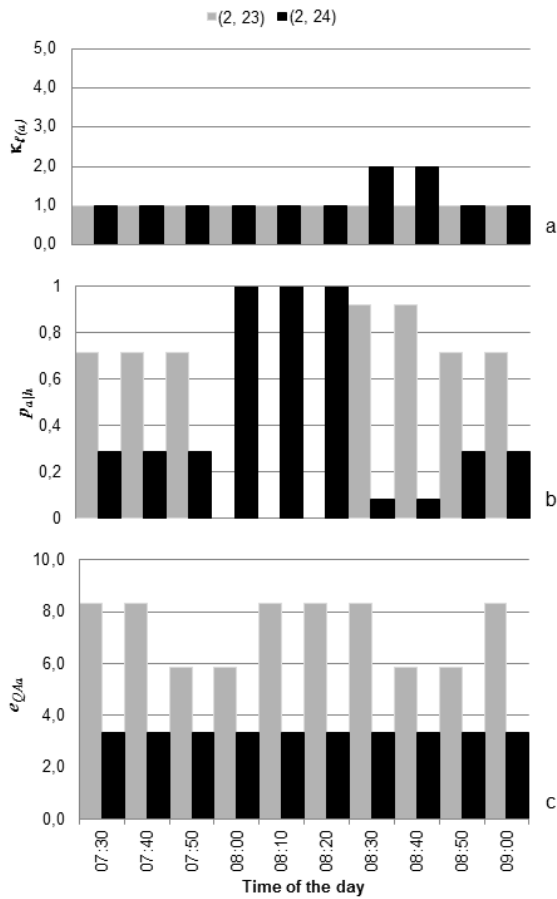


Figure 5-7 Variation of congestion factor κ_a (a), diversion probability p_{alh} (b) and instantaneous exit capacity $e_{Q_{Aa}}$ (c) at Stop 2; a represents the waiting arcs for Line 1 and Line 3 and Q_{Aa} the queuing arcs for the same lines at Stop 2

A second instance of the problem is also studied, where the destinations considered are nodes 16, 18 and 19 and it is assumed that the od matrix is in the form given by Table 5-3. In such a setting, the solution algorithm (Figure 5-8) converges to $\varepsilon = 0,001$ in 30 iterations and to $\varepsilon = 0,0001$ in 67 iterations.

Table 5-3 Time-dependent od matrix2 during the analysis period; the travel demand is expressed in passenger/min

| Origin Centroid | Destination Centroid | | |
|-----------------|----------------------|----|----|
| | 16 | 18 | 19 |
| 17 | 5 | 5 | 1 |
| 18 | 4 | 0 | 3 |
| 19 | 6 | 0 | 0 |

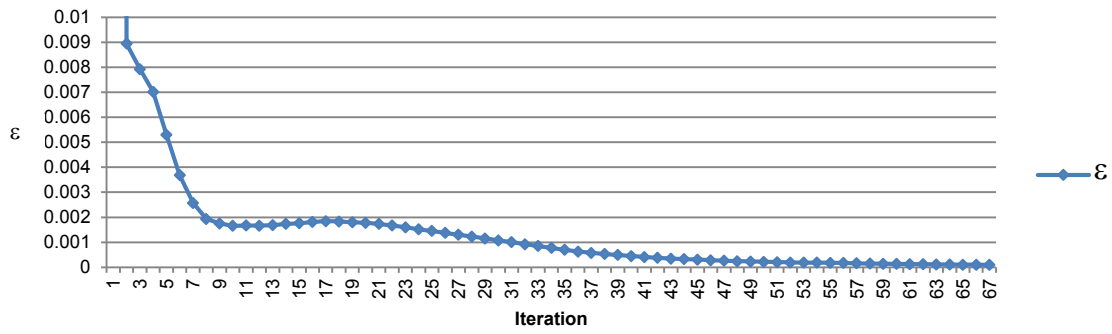


Figure 5-8
Algorithm convergence

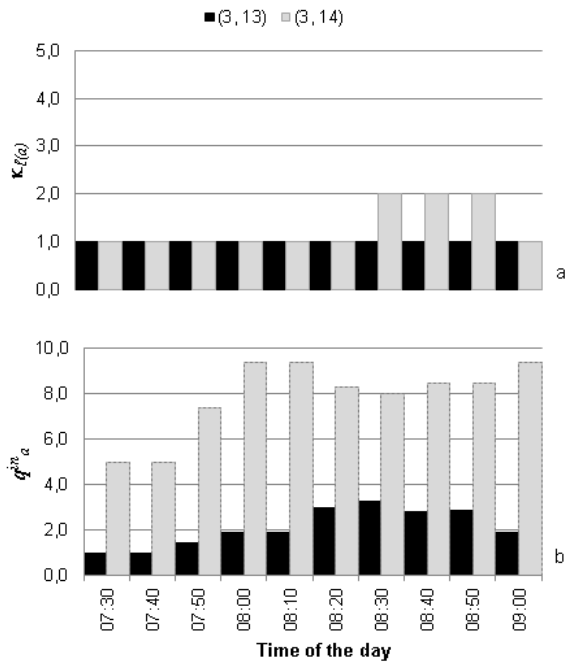


Figure 5-9
Variation of congestion parameter $\kappa_{l(a)}$ (a) and inflow q_a^{in} (b) at Stop 3; a represents waiting arcs corresponding to Line 3 and Line 4

The results show that the only queue in the network occurs at Stop 3 where, between 08:25 and 08:55, passengers have to wait for the second passage of Line 4 if they want to board this service. Therefore, the diversion probabilities and, thus, the inflow on waiting arcs at Stop 3 are greatly affected by congestion, as depicted in Figure 5-9; and the inflow on arc (3, 13) increases [08:20 – 08:50] when passengers know that, by the time the next carrier of Line 4 arrives, it will be full and no place will be available on-board.

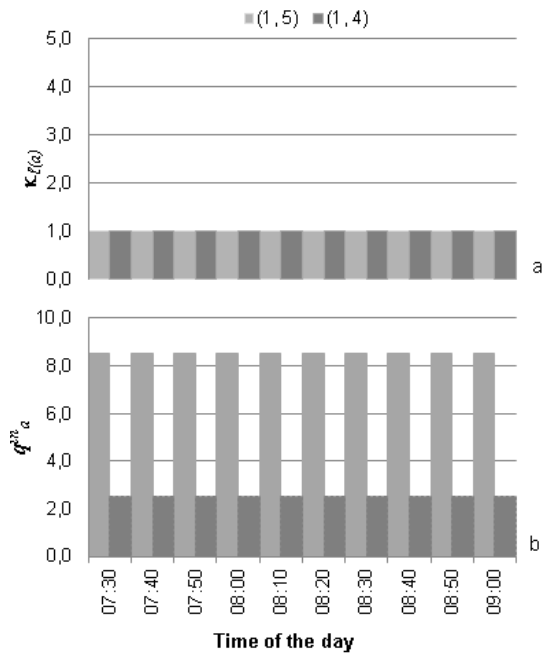


Figure 5-10
Variation of congestion parameter $\kappa_{l(a)}$ (a) and inflow q_a^{in} (b) at Stop 1; a represents waiting corresponding to Line 1 and Line 2

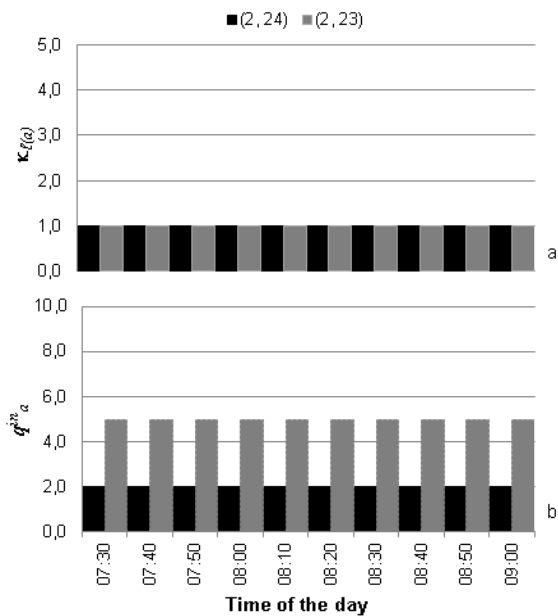


Figure 5-11
Variation of congestion parameter $\kappa_{l(a)}$ (a) and inflow q_a^{in} (b) at Stop 2; a represents waiting arcs corresponding to Line 1 and Line 3

Notwithstanding the queue at Stop 3, the increase in total travel time from node 22 and node 7 to node 16 is not remarkable and, as opposed to the first example, Line 1 is always kept in

the attractive sets of Stop 1 and Stop 2 whichever is the final destination node considered. As a consequence (Figure 5-10 and Figure 5-11), the inflow on waiting arcs (1, 4), (1, 5), (2, 23) and (2, 24) stays constant throughout the analysis period.

It is also important to notice here that, at Stop 1, only half of the flow directed towards node 16, namely 2.5 passengers per minute, is propagated on arc (1, 4) because, in the spirit of ‘hyperpaths’, diversion probabilities are computed solely on the grounds of waiting times at the current stop (Figure 5-12). As a consequence, not all the capacity available on Line 2 is used but the inflow of 8.5 passengers per minute that boards Line 1 contributes to produce congestion further down the network.

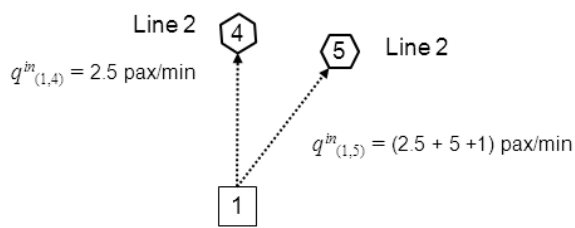


Figure 5-12
Instantaneous inflow on Line 1 and Line 2 at Stop 1

However, if it is assumed that real-time information is provided at stops by countdown displays the RCM changes in such a way that a more even spread of flows across the network is attained and, thus, the usage of the supplied capacity is optimised. Indeed, in this scenario, travel times upon boarding do not only affect the inclusion/exclusion of a line from the attractive set but also the evaluation of diversion probabilities.

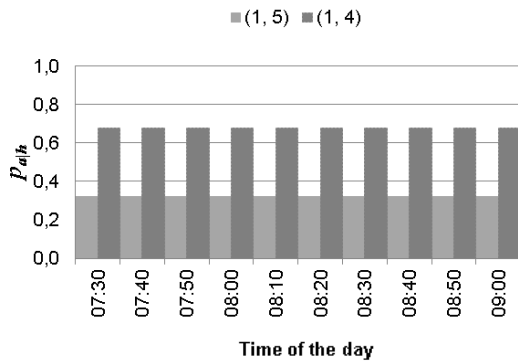


Figure 5-13
 Diversion probabilities at Stop 1, for passengers directed to node 16

Consequently, when the shortest hypertree is calculated for destination node 16, the diversion probability of arc (1, 4) increases to 0.67 because the travel time upon boarding Line 2 (27 minutes) is inferior to the total travel time upon boarding Line 1 (29.62 minutes) (Figure 5-13). As a result, the inflow and congestion parameters for the three stops are those depicted in Figure 5-14, Figure 5-15 and Figure 5-16. As the flow spread does not produce congestion phenomena, the algorithm converges in one iteration only.

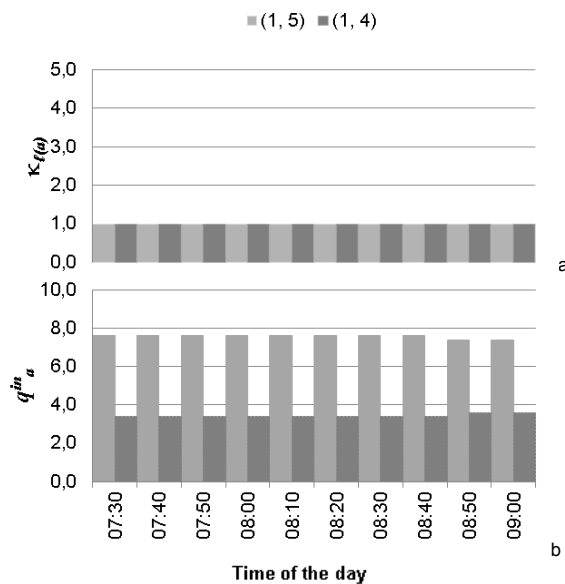


Figure 5-14
 Variation of congestion parameter $\kappa_{l(a)}$ (a) and inflow q_a^{in} (b) at Stop 1 when countdown displays are considered

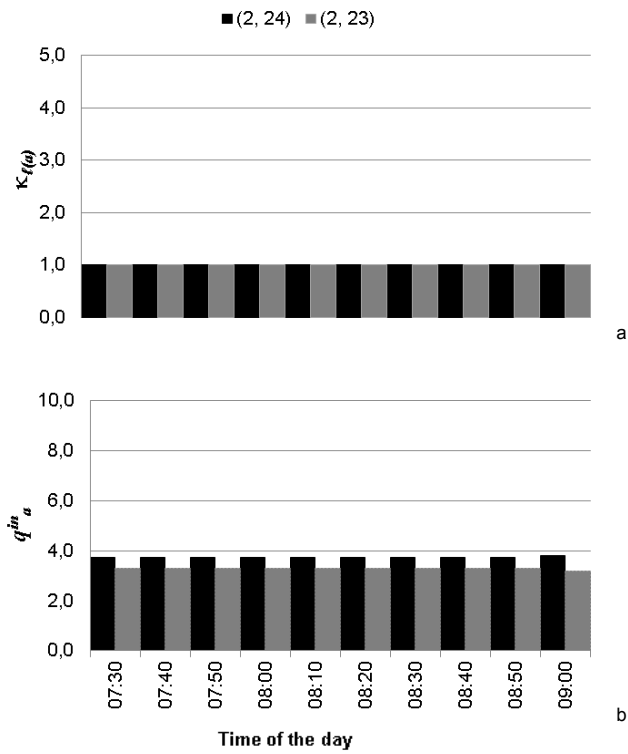


Figure 5-15
Variation of congestion parameter $\kappa_{l(a)}$ (a) and inflow q_a^{in} (b) at Stop 2 when countdown displays are considered

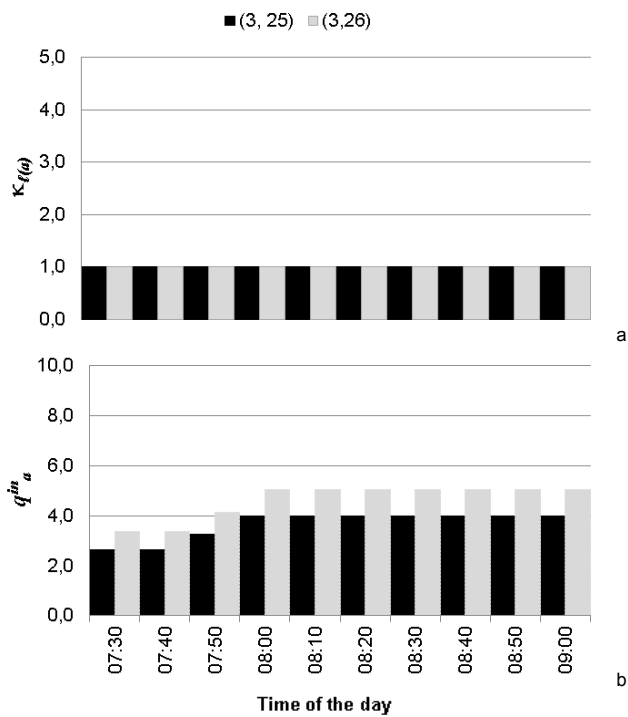


Figure 5-16
Variation of congestion parameter $\kappa_{l(a)}$ (a) and inflow q_a^{in} (b) at Stop 3 when countdown displays are considered

5.5. SOFTWARE IMPLEMENTATION

In order to apply the proposed methodology to a real-scale network, the solution algorithm detailed in Section 5.3 has been implemented using the programming language *Microsoft Visual Basic 2010*, which allows the use of OPTIMA (©SISTeMA – Soluzioni per l'Ingegneria dei Sistemi di Trasporto e l'infoMobilità s.r.l.) to manage the import of the base graph as well as the export of assignment results from/to various database formats, including the VISUM format.

The model graph described in Section 5.2 requires a multiplicity of nodes and arcs which make it impossible to create such a network manually from the base graph. Therefore, a procedure is implemented in the software to perform the task automatically, as detailed in 5.5.1.

Moreover, when real-scale networks are considered, heuristics are needed to select the set of attractive lines considered by users at each stop. The solution proposed is detailed in 5.5.2.

5.5.1. Automatic creation of the model graph

The network data, available in VISUM format, are imported by OPTIMA into *Microsoft Visual Basic 2010* and used to build the base graph of the network, which includes all the basic information listed in 3.1.1:

- Edge length: λ_e ;
- Pedestrian speed: ρ_e ;

- Set of lines: L ;
- Route of the generic line ℓ , defined as an ordered sequence on not repeated σ_ℓ vertices:

$$R_\ell = \{ R_{\ell,i} \}, i \in [1, \sigma_\ell]$$

(The generic section of a route is referred to as $(R_{\ell,i-1}, R_{\ell,i}) \in E$, with $i \in [2, \sigma_\ell]$, and corresponds to an edge of the base graph; for any given vertex $v \in V$ and line $\ell \in L$, the function $s(v, \ell) \in [1, \sigma_\ell]$ yields, if it exists, the index such that $R_{\ell,s(v, \ell)} = v$, and 0 otherwise);

- Function expressing if a stop is made or not at the v -th vertex along the route of line ℓ :

$$\mu_{\ell,v};$$

- Function expressing if the v -th vertex corresponds to a stop: μ_v ;
- Vehicle capacity and base frequency of line ℓ : χ_ℓ, φ_ℓ ;
- Line time of line ℓ : $\theta_{\ell,i}(\tau), i \in [1, \sigma_\ell]$.

The model graph is therefore automatically built on the basis of the base graph, as detailed in the following:

Automatic creation of the nodes and arcs of G

1 $\forall v \in V$

2 If $\mu_v > 0$ Then

3 $N = N \cup \{i\}, N^P = N^P \cup \{i\}, i = (v, P, 0)$

4 $N = N \cup \{i\}, N^S = N^S \cup \{i\}, i = (v, S, 0)$

5 $A = A \cup \{a\}, A^Z = A^Z \cup \{a\}, a = (i, j): i \in N^P, j \in N^S, V_i \equiv V_j = v$

6 $\forall \ell \in L$

7 If $\mu_{v,\ell} > 0$ Then

8 $N = N \cup \{i\}, N^W = N^W \cup \{i\}, i = (v, W, \ell)$

9 $N = N \cup \{i\}, N^Q = N^Q \cup \{i\}, i = (v, Q, \ell)$

10 $N = N \cup \{i\}, N^B = N^B \cup \{i\}, i = (v, B, \ell)$

11 $A = A \cup \{a\}, A^W = A^W \cup \{a\}, a = (i, j): i \in N^S, j \in N^W, V_i \equiv V_j = v, L_{HD_a} = \ell$

12 $A = A \cup \{a\}, A^Q = A^Q \cup \{a\}, a = (i, j): i \in N^W, j \in N^Q, V_i \equiv V_j = v, L_{HD_a} = \ell$

13 $A = A \cup \{a\}, A^B = A^B \cup \{a\}, a = (i, j): i \in N^Q, j \in N^B, V_i \equiv V_j = v, L_{HD_a} = \ell$

14 For $s = 2$ to $\sigma_\ell - 1$

15 $N = N \cup \{i\}, N^A = N^A \cup \{i\}, i = (v, A, \ell)$

16 $A = A \cup \{a\}, A^A = A^A \cup \{a\}, a = (i, j): i \in N^A, j \in N^P, V_i \equiv V_j = v, L_{TL_a} = \ell$

17 $A = A \cup \{a\}, A^D = A^D \cup \{a\}, a = (i, j): i \in N^A, j \in N^B, V_i \equiv V_j = v, L_{HD_a} = \ell$

18 $s = \sigma_\ell$

19 If $1(v, \ell) = \sigma_\ell(v, \ell)$ Then (the line is circular)

20 $N = N \cup \{i\}, N^A = N^A \cup \{i\}, i = (v, A, \ell)$

21 $A = A \cup \{a\}, A^A = A^A \cup \{a\}, a = (i, j): i \in N^A, j \in N^P, V_i \equiv V_j = v, L_{TL_a} = \ell$

22 $A = A \cup \{a\}, A^D = A^D \cup \{a\}, a = (i, j): i \in N^A, j \in N^B, V_i \equiv V_j = v, L_{HD_a} = \ell$

23 Else

24 $N = N \cup \{i\}, N^A = N^A \cup \{i\}, i = (v, A, \ell)$

25 $A = A \cup \{a\}, A^A = A^A \cup \{a\}, a = (i, j): i \in N^A, j \in N^P, V_i \equiv V_j = v, L_{TL_a} = \ell$

26 Else

$$27 \quad N = N \cup \{i\}, N^P = N^P \cup \{i\}, i = (v, P, 0)$$

28 $\forall e \in E$

$$29 \quad A = A \cup \{a\}, A^P = A^P \cup \{a\}, a = (i, j): i \in N^P, j \in N^P, e = (V_i, V_j) \in E, \rho_e > 0$$

30 $\forall a \in A^Q$

$$31 \quad BA_a \in A^B : TL_{BA_a} = HD_a$$

$$32 \quad DA_a \in A^D : HD_{BA_a} = HD_{DA_a}$$

$$33 \quad WA_a \in A^W : TL_a = HD_{WA_a}$$

5.5.2. Definition of the attractive set

As mentioned in Section 3.4.3, the application of the RCM proposed in this thesis to real-scale networks requires that heuristics are devised for the definition of the attractive set at each stop, because the exact solution implies a combinatorial problem with factorial complexity.

On the other hand, results of a stated preference survey conducted by Fonzone et al. (2010, 2012) seem to suggest that, even when several competing alternatives exist, passengers tend to simplify the portfolio of available options. More specifically, the authors ask different passenger groups to describe their actual travel patterns as well as to choose their strategy in hypothetical bus networks. In both cases, only some passengers choose the hyperpaths predicted by the Spiess and Florian model, while a significant percentage seem to prefer simpler choice sets.

Therefore, in order to define the attractive set in real-scale applications, when a stop is reached only the three ‘best lines’ (namely lines with shortest travel times upon boarding) are selected and equation 3-38 is applied to this subset. Consequently, Part 1 of the algorithm (SM and RCM) is modified as follows:

1 Step 1.0 (Static pre-processing – Initialisation):

2 $\forall i \in N \setminus \{d\}$

3 Calculate $g^*_{i,d}(\xi^{T-1}) = g^*_{i,d}{}^{\text{stat}}$

4 $\forall \tau \in [0, T-2]$

5 Set $g^*_{d,d}(\xi^\tau) = 0, \text{ suc}(d, \xi^\tau) = \emptyset$

6 $\forall i \in N \setminus \{d\}$

7 Set $g^*_{i,d}(\xi^\tau) = \infty$

8 $w_h(\xi^\tau) = w_h(\xi^\tau) + w_{a|h}(\xi^\tau) \cdot p_{a|h}(\xi^\tau)$

9 Step 1.1 (Select the hypertree with minimal generalised travel time – RCM):

10 $\forall \tau \in [T-2, 0]$ Step -1

11 $\forall i \in N \setminus \{d\}$

12 If $i \in N^S$, Then perform Step 1.2

13 $\forall a \in h$

14 $t_{a|h}(\xi^\tau) = \lceil w_{a|h}(\xi^\tau) / \tau_{\text{INT}} \rceil + \xi^\tau$

15 If $t_{a|h}(\xi^\tau) \leq \xi^T$ and $g_{HD_{a,d}}(t_{a|h}(\xi^\tau)) < \infty$ Then

$$16 \quad g_{i,d}(\xi^r) = w_h(\xi^r) + p_{ah}(\xi^r) \cdot g_{HDa,d}(t_{ah}(\xi^r))$$

17 Else

$$18 \quad g_{i,d}(\xi^r) = w_h(\xi^r) + p_{ah}(\xi^r) \cdot g_{HDa,d}^{\text{stat}}$$

19 If $g_{i,d}^*(\xi^r) > g_{i,d}(\xi^r)$ Then

$$20 \quad g_{i,d}^*(\xi^r) = g_{i,d}(\xi^r) \text{ And } \text{suc}(i, \xi^r) = h$$

21 ElseIf $i \notin N^S, \forall a \in FS_i$ Then

$$22 \quad t_a(\xi^r) = \lceil c_a(\xi^r) / \tau_{\text{INT}} \rceil + \xi^r$$

$$23 \quad g_{i,d}(\xi^r) = c_a(\xi^r) + g_{j,d}(t_a(\xi^r))$$

24 If $g_{i,d}^*(\xi^r) > g_{i,d}(\xi^r)$ Then

$$25 \quad g_{i,d}^*(\xi^r) = g_{i,d}(\xi^r) \text{ And } \text{suc}(i, \xi^r) = a$$

26 Step 1.2 Hyperarc definition and SM:

$$27 \quad \forall a \in FS_i$$

$$28 \quad t_a(\xi^r) = \kappa_a(\xi^r) / \varphi_a(\xi^r)$$

29 Sort $a \in FS_i$ in increasing order of $g_{j,d}(t_a(\xi^r))$:

$$30 \quad g_{j_1,d}(t_{a_1}(\xi^r)) \leq g_{j_2,d}(t_{a_2}(\xi^r)) \leq \dots \leq g_{j_n,d}(t_{a_n}(\xi^r)), n = |FS_i|$$

$$31 \quad FS'_i = FS_i \cap \{a_1, a_2, a_3\}$$

$$32 \quad \forall h \in HFS'_i$$

$$33 \quad \forall a \in h$$

34 Calculate $p_{ah}(\xi^r)$ using equation 3-2

5.6. CASE STUDY

Section 5.4 has proven the methodological validity of the proposed approach by showing that it properly reproduces dynamic effects of congestion at stops. In this section, a larger case study is presented which confirms the scalability of the model and its applicability also to realistic networks.

In the presence of time-dependent travel demand, the model should capture congestion phenomena during the peak periods of travel demand. Therefore, the temporal profile of the number of queuing passengers on arc $a \in A^Q$ ($n_a(\tau)$) should increase when the supplied capacity no longer meets the travel demand and decrease down to zero when the latter decreases and becomes lower than the supplied capacity (off-peak periods). Similarly, if WA_a is a branch of hyperarc ($WA_a \in h$) corresponding to $a \in A^Q$ (see for example the branch a depicted in Figure 4-2 and $a_2 \in A^Q$ depicted in Figure 4-3), then it is expected that the temporal profile of the waiting time before boarding the line associated with this branch of hyperarc ($w_{WA_a|h}(\tau)$) should somehow follow the curve of $n_a(\tau)$ for every possible hyperarc h that is considered (although its magnitude will depend also on the congestion levels of the *other* lines included in the choice set represented by h).

Furthermore, when the demand peak produces queuing at some stops and, thus, the total travel time increases on some routes, travel choices are affected and a modification in the flow pattern is expected to be seen. In other words, it is possible that express lines, which

potentially offer a fast connection but force passengers to queue at the origin/transferring stop and wait for the following run(s), become less attractive while passenger flows on slower but uncongested lines increase.

Finally, it is expected that this phenomenon will be more evident in a densely connected network. This is because passengers have several alternative lines connecting to the same destination (directly or indirectly) that depart from their stop and, thus, can re-route very easily to board a less-congested line.

Consequently, in order to observe all the phenomena described above, the ‘ideal’ case study should consider a public transport network with the following characteristics:

1. Transit lines with high frequency and/or low reliability, so that it is fair to assume that passenger would not explicitly consider the lines’ timetable when making their travel choices (i.e. they do not time their arrival at the stop with the timetable), but would only take into account: the average frequency, expected travel time to destination upon boarding and congestion levels (namely, the number of runs they expect to miss because of overcrowding);
2. Network densely connected with partially overlapping lines, so that passengers have several available alternatives that connect the same *od* pair and, thus, can choose travel strategies rather than simple itineraries;
3. Time-dependent travel demand, whose peaks temporarily exceed the supplied capacity and produce severe overcrowding with long passenger queues at several stops.

Unfortunately, for the purpose of the case study presented here, it was not possible to consider a public transport network with all these features. More specifically, as will be detailed in sub-section 5.6.1, the requirement on the travel demand is not met, hence the

results analysed in 5.6.2 have some limitations, which are discussed in sub-section 5.6.3.

5.6.1. Data description

This case study examines the results obtained from applying the solution algorithm to the tram network of Cracow, whose base graph has been kindly provided in VISUM format by Cracow University (Rudnicki et al., 2011) and includes:

- 23 lines and 157 stops (Figure 5-17);
- 136 traffic zones, whose centroids are displayed in Figure 5-18;
- 826 line segments ($R_{\ell,i-1}, R_{\ell,i}$).

Instead of the exact timetable, the frequency of each line has been considered, as displayed in Table 5-4, together with the carrier capacity.

Table 5-4Frequencies [min^{-1}] and capacities [number of places] of the 23 lines considered in the case-study network

| Line | Frequency | Carrier Capacity |
|------|-----------|------------------|
| 1 | 1/5 | 160 |
| 2 | 1/10 | 160 |
| 3 | 1/5 | 160 |
| 3 | 1/5 | 160 |
| 4 | 1/5 | 200 |
| 5 | 1/10 | 160 |
| 6 | 1/2 | 160 |
| 7 | 1/10 | 160 |
| 8 | 1/5 | 160 |
| 9 | 1/5 | 160 |
| 10 | 1/5 | 160 |
| 11 | 1/10 | 90 |
| 13 | 1/5 | 200 |
| 14 | 1/5 | 160 |
| 15 | 1/10 | 160 |
| 16 | 1/5 | 160 |
| 17 | 1/15 | 160 |
| 19 | 1/5 | 160 |
| 20 | 1/10 | 90 |
| 21 | 1/10 | 160 |
| 22 | 1/5 | 200 |
| 50 | 1/3 | 160 |
| 51 | 1/2 | 160 |

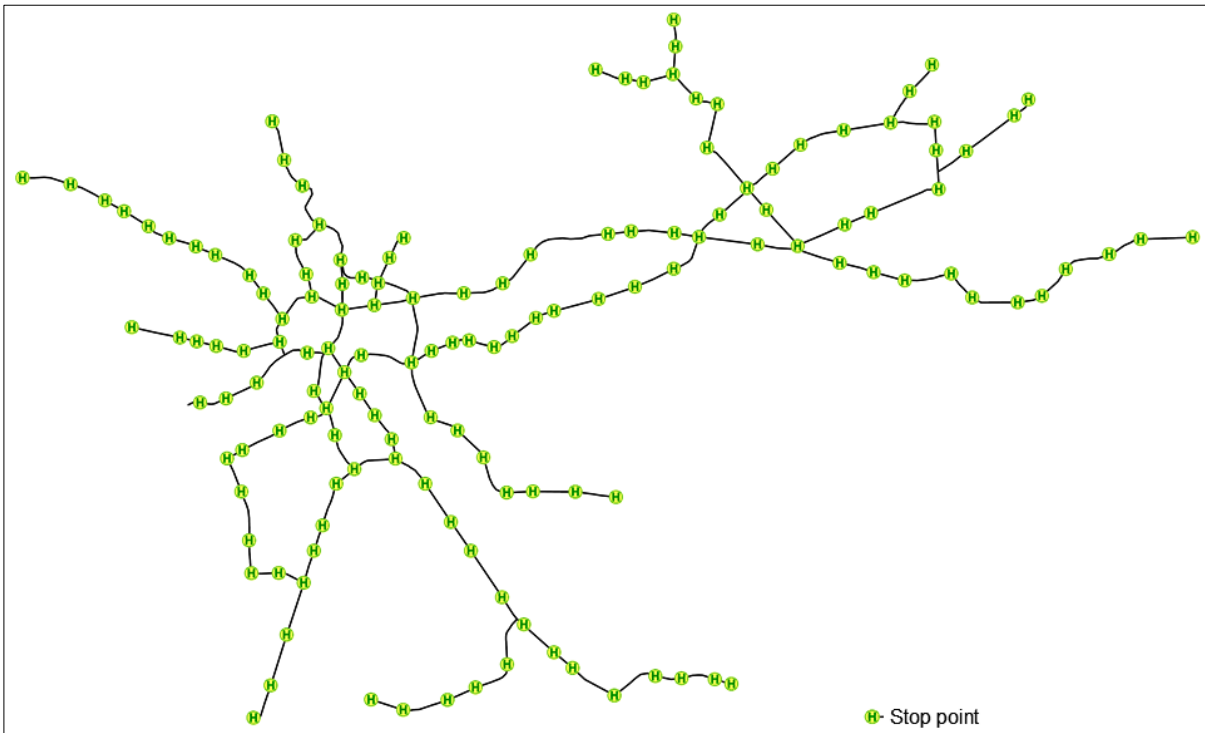


Figure 5-17
Base-graph representation of the case-study network

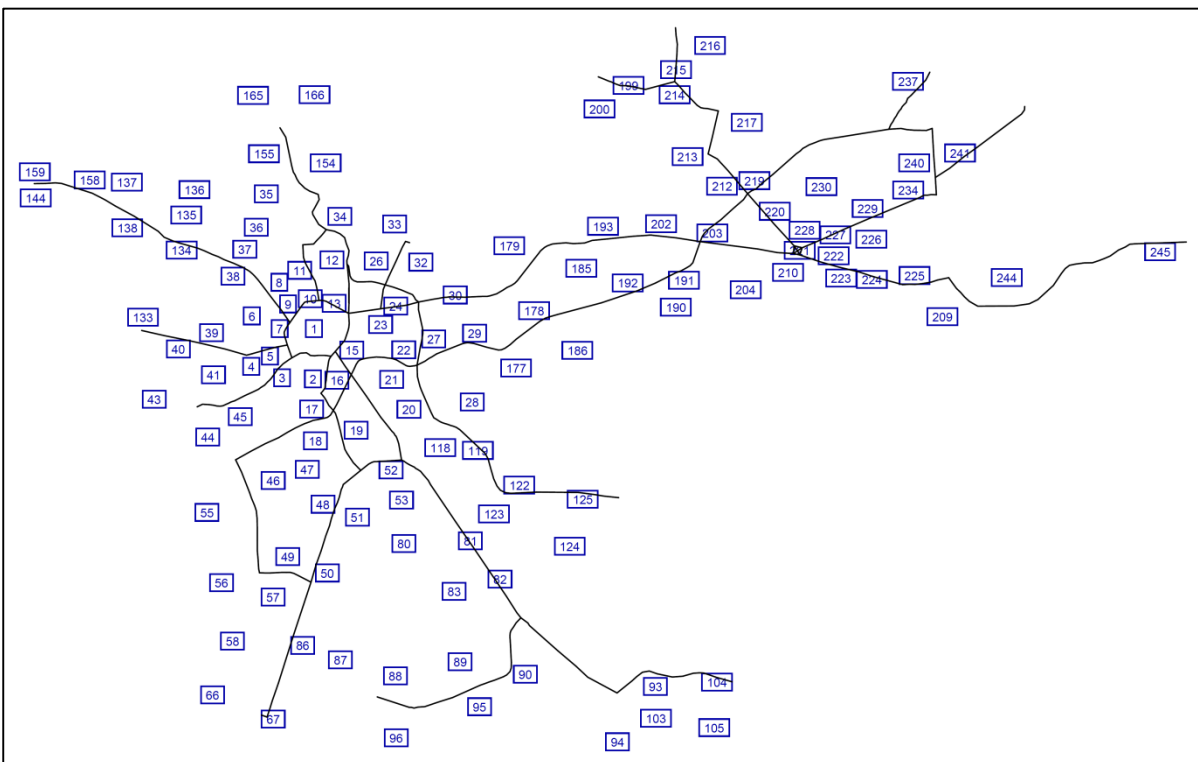


Figure 5-18
Centroids of the 136 traffic zones in the case-study network

While the *od* matrix under consideration is detailed in the Appendix, a summary of it is given in Figure 5-19a and b.

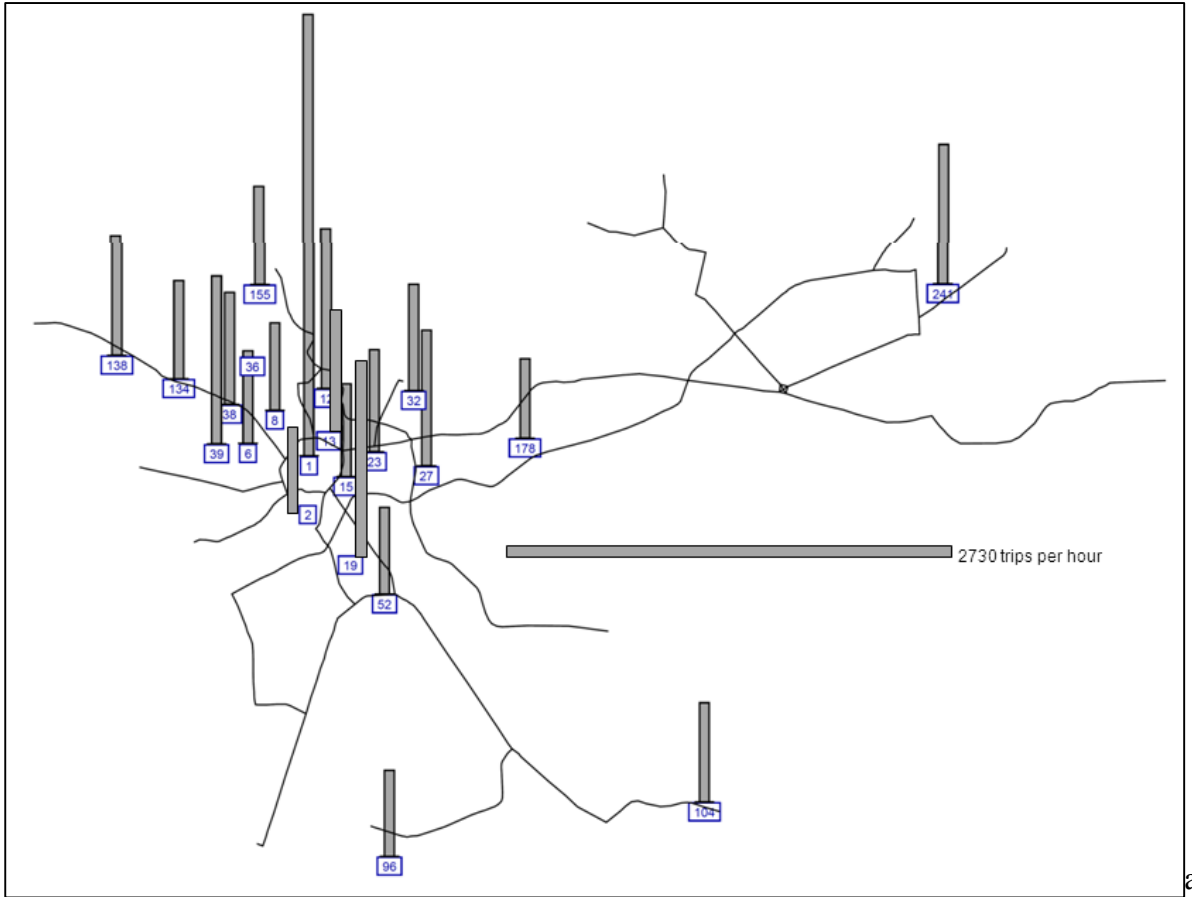


Figure 5-19
a: Zones of the network with highest demand attraction: the chart highlights the zones that attract more than 450 trips during the analysis period

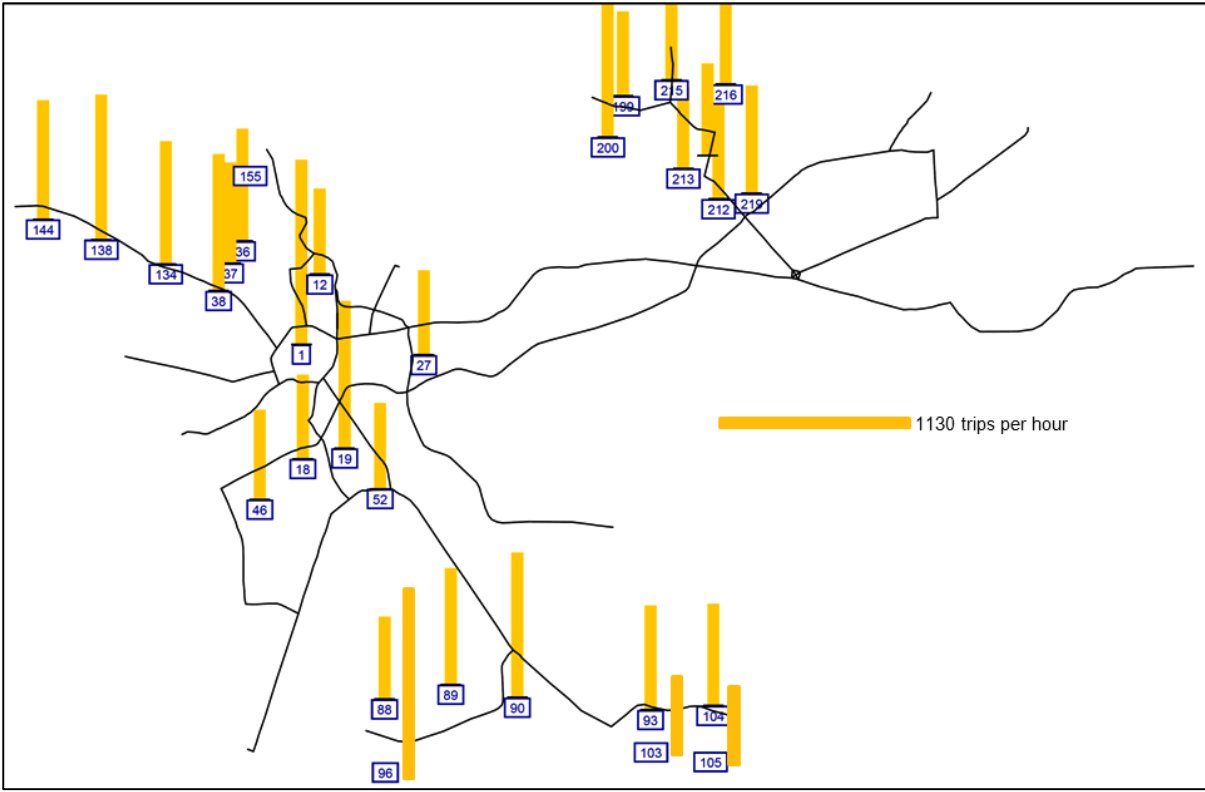


Figure 5-19
 b: Zones of the network with highest demand generation: the chart highlights the zones that generate more than 450 trips during the analysis period

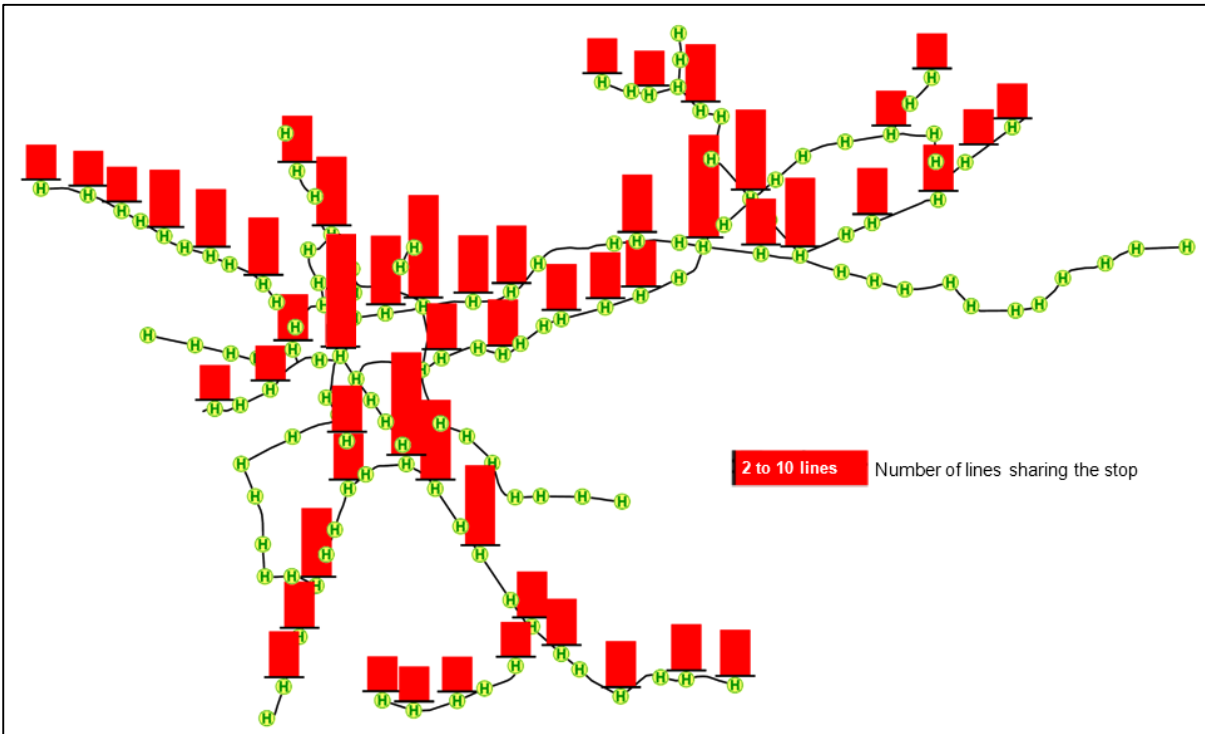


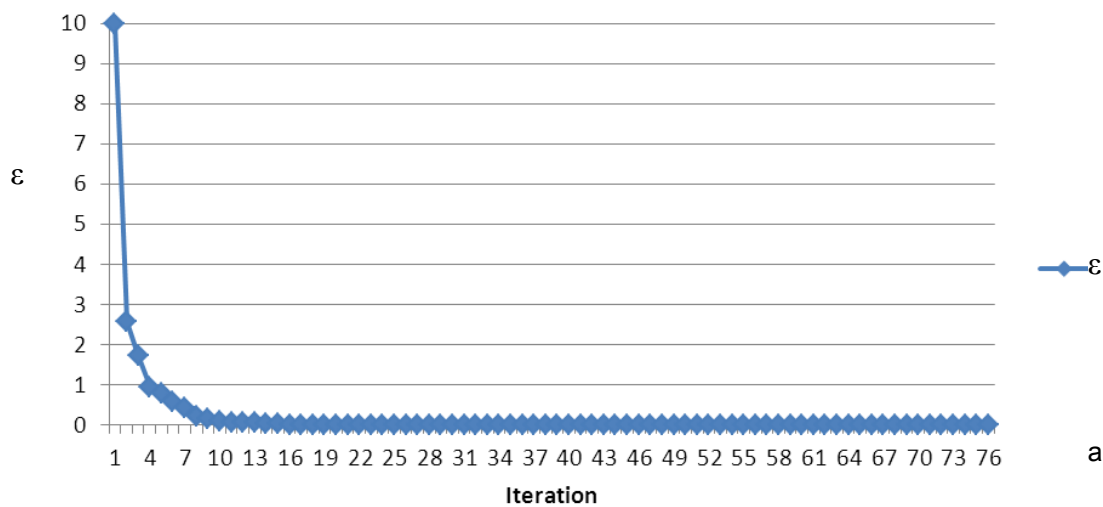
Figure 5-20
 Number of lines sharing each stop

Because 95.65% of the lines have an average frequency of six or more vehicles per hour, the assumption that passengers would not explicitly consider the services' timetable seems to be rational. Similarly, with most of the stops shared by two or more lines (Figure 5-20), it is reasonable to assume that users would take advantage of partially overlapping line routes by choosing an optimal strategy to destination, rather than a simple path.

5.6.2. Results analysis

The analysis period of one hour has been divided into 60 time intervals of one minute each. Alighting, boarding and dwelling time are each assumed to be of one minute and, for computational reasons, the travel time on dummy arcs is also set to be of one minute. Moreover, the temporal profile of the total travel demand is assumed to be constant during the analysis period and the same applies to line frequencies.

The maximum value of ϵ accepted for the stop criterion is 0.001 and the algorithm converges in 76 iterations (Figure 5-21).



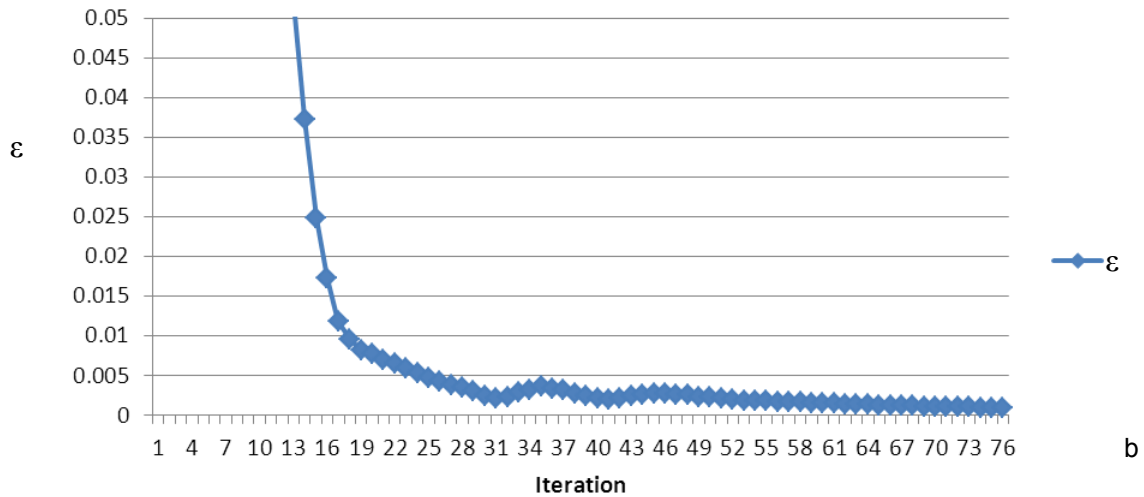


Figure 5-21
a: Algorithm convergence
b: algorithm convergence – detail

Results of the assignment procedure are summarised in Figure 5-22 to Figure 5-25 that clearly show the variation in the flow pattern during the analysis period, with a progressive increase of total flows in areas of the network with highest demand attraction.



Figure 5-22
Flow pattern across the network during the 15th time interval

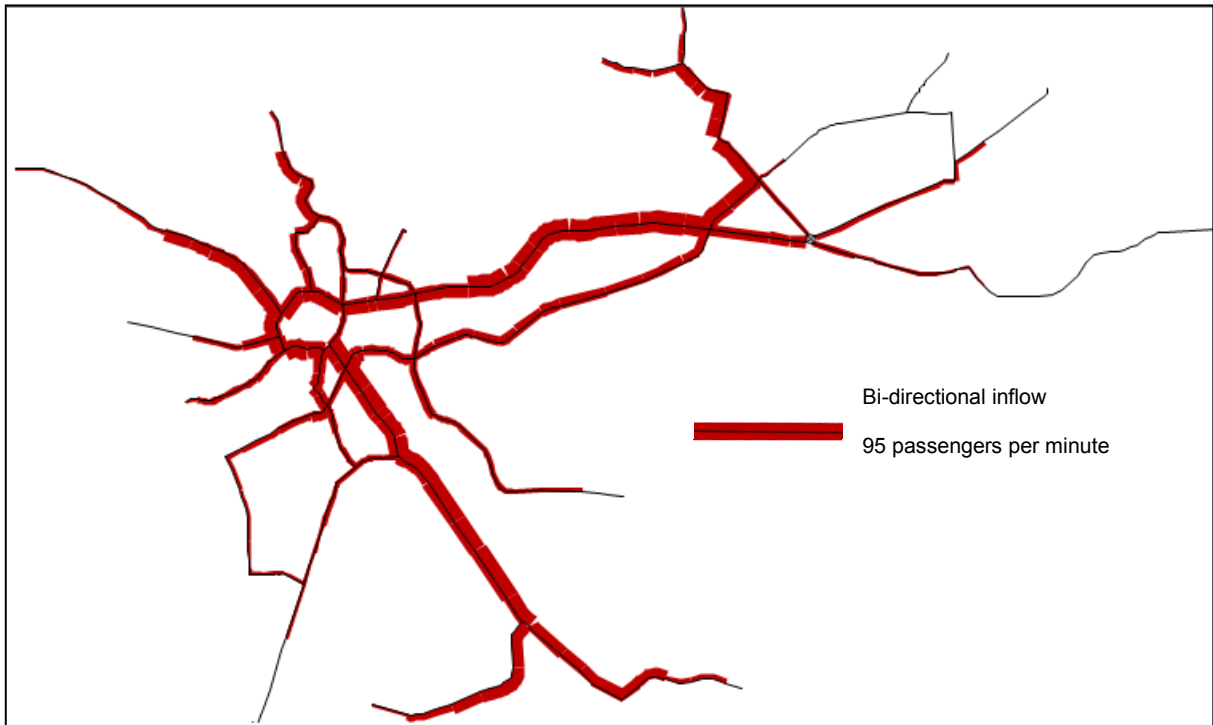


Figure 5-23
Flow pattern across the network during the 30th time interval

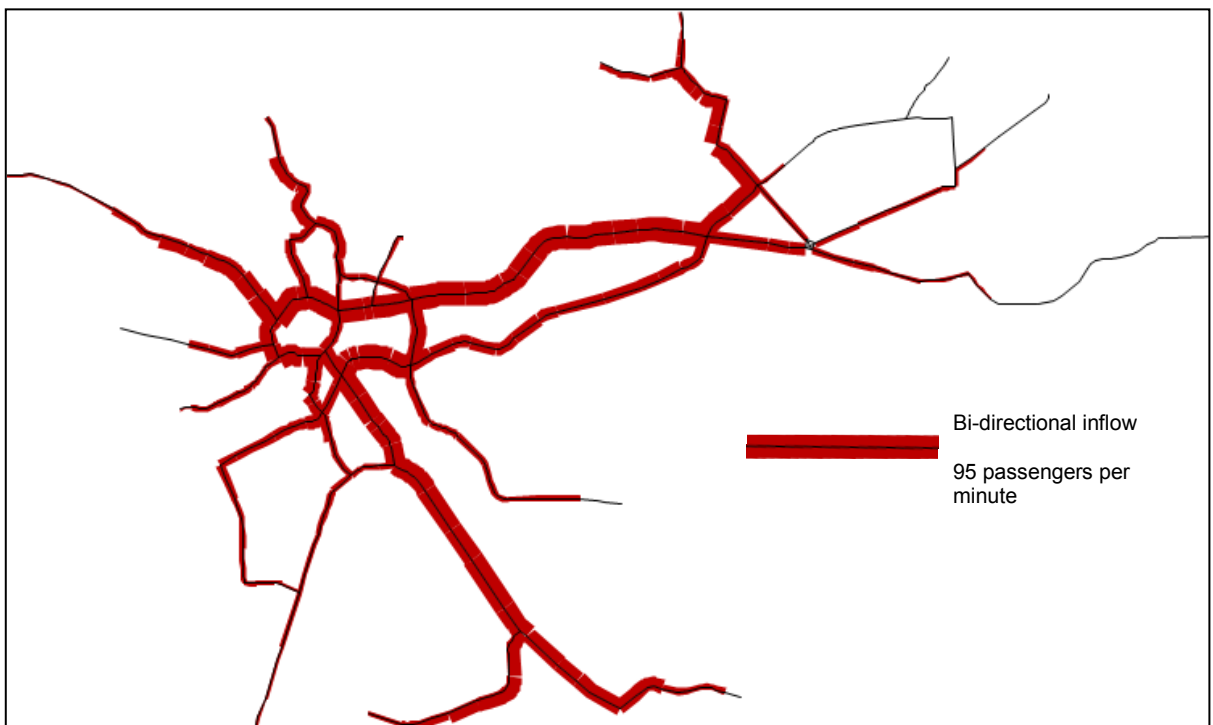


Figure 5-24
Flow pattern across the network during the 45th time interval



Figure 5-25
Flow pattern across the network during the 60th time interval

5.6.3. Limitations of the case study and implementation issues

The results of the case study clearly show that the network is dynamically loaded (Figure 5-22 – Figure 5-25) and that passenger flows move progressively from their origins towards their destination. On the other hand, the absence of a temporal profile for the travel demand (in fact, only the *total* demand during the analysis period is known) does not allow for a real evaluation of congestion effects on the algorithm efficiency as well as on the users' choices.

More specifically, the peak-less travel demand considered in this case study does not trigger congestion phenomena at stops and, hence, it is not possible to analyse the variation of running time due to the longer computation needed to solve the Bottleneck Queue Model at stops where the number of queuing passengers exceeds the number of places available on-

board the arriving lines. Furthermore, as queuing times do not increase, it cannot be clearly observed that there is any change in the chosen routes due to network congestion; and it also cannot be analysed whether or not (or by how much) having the possibility of choosing travel strategies rather than simple itineraries reduces the overall congestion and leads to an optimal usage of the supplied capacity.

The availability of reliable and detailed inputs is one of the main issues to be considered if such a dynamic model is to be applied to a vast network. More specifically, although the frequency-based framework does not require very detailed input on the supply-side, the *instantaneous* travel demand has to be known in order to analyse the effects of its peaks on network conditions. On the other hand, the application of ITS – such as the use of smart cards that record the origin and destination of all journeys – is now remarkably simplifying the task of collecting detailed and reliable information on travel demand; and it is thus thought that in the near future the provision of time-dependent demand data will be a common practice.

A second issue that should be further analysed before any practical implementation of this model is its sensitivity to the heuristic adopted for the calculation of the attractive set. In fact, the proposed method implies that passengers tend to simplify the routing problem and thus, in their normal day-to-day life, would only consider a sub-set of all the available alternatives from the same stop. The assumption would seem generally realistic; however, there are certain instances, especially when the origin and destination points are quite close to each other, where *all* the common lines share the same route and thus have almost the same travel time. In this case, it is unlikely that passengers would simplify their choice set and consider only the three best options, as done here, but would rather include them all in the attractive set and consider their total combined frequency.

5.7. DISCUSSION

This chapter presents the implementation of the proposed assignment model and gives details of the solution algorithm, which is developed by extending to the scenario of interest the DOT method, originally devised by Chabini (1998) to solve all-to-one searches for every possible arrival time in a dynamic setting (Section 5.2.1). Although the demand and supply models are referred to two different graphic structures, those are conceived in such a way that only one model graph needs to be built and stored in order to implement the solution algorithm, as explained in sections 5.2.2 and 5.5.1.

The solution algorithm was firstly applied to a small example network (Section 5.4) to show clearly the effects of the build-up and dissipation of passengers' queues on route choices, as reproduced by the demand model, as well as the 'feedback' effect of route choices on the build-up and dissipation of queues at stops, as reproduced by the supply model. Although congestion phenomena further down in the network may decrease the 'attractiveness' of one line and determine its exclusion from the choice set, if a line is attractive its diversion probability solely depends on its frequency and congestion at the current stop. On the other hand, if it is possible to assume that diversion probabilities are also affected by the travel time upon boarding, as hypothesised in the dynamic versions of equation 3-15 for the case when countdown displays are available, a more even spread of passenger flows is seen across the network (Figure 5-14, Figure 5-15 and Figure 5-16). Unfortunately, from a computing point of view, the latter result is attained with a remarkable increase of memory consumption because each value of the diversion probability ($p_{a|h}$) not only depends on the particular choice set considered at the transit stop and on the parameters of the attractive line (κ_a, φ_a) but also on the particular *destination* considered.

The solution algorithm is also applied to a larger case study to demonstrate the scalability of the proposed model to real-world networks. Although the implementation of the solution algorithm has not been optimised, the MSA that solves the FPP formulation of the dynamic DUE converges to 0.001 in 25 iterations, with each iteration taking 30 minutes to complete on Processor Intel® Core™ i7-3939K CPU@ 3.20GHz.

The case study clearly shows the change in the flow pattern across the network with the elapsed time. However, as a peak-less demand matrix was considered, severe congestion phenomena were not observed. As a consequence, the case study does not allow for a real evaluation of congestion effects on the algorithm efficiency or on the users' choices.

6. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

6.1. CONCLUSIONS

This thesis has presented an analytical approach to transit assignment that explicitly considers the dynamic interaction between travel demand and public transport network supply, which determines – in turn – the performance of the transport system with regards to congestion.

The main task of the assignment model is to distribute passenger flows over transit routes of the public transport network in accordance with the *od* travel demand matrix. Therefore, the core of this mathematical framework is the demand model, made up of the Stop Model (SM) and the Route Choice Model (RCM), which links passenger decisions with network conditions at the time of the day when the trip is made.

The demand model is assumed to be *deterministic*, which means that passengers are perfectly rational decision makers and fully informed about average line frequencies and travel times upon boarding. They therefore all choose to travel along the *best* option(s) available, i.e. the option(s) with the shortest expected travel time.

It should be acknowledged here that the use of a deterministic demand model entails a number of simplifying assumptions, and it has been argued in the literature (see for example Lam et al., 1999; Lam et al., 2002; Sumalee et al., 2009a; Sumalee et al., 2009b) that travel choices may be more realistically represented through a stochastic demand model. Still, deterministic models present several advantages. First, the flexibility and accuracy of

stochastic models usually depend on the accurate calibration and validation of a considerable number of behavioural parameters, while no parameter of this sort is included in deterministic models.

Furthermore, deterministic models are easier to understand from a theoretical point of view and, in general, their results are easier to interpret and analyse. Thus, although not extremely refined, deterministic models are very robust and, if used in a sensitivity analysis to compare different project scenarios, they are more reliable. Indeed, in this case the different results are entirely due to the effects that changes in the supplied LoS produce on route choices and are not due to stochastic perceptions and/or user choices.

Lastly, as recognised by Cascetta (2009: p. 329), ‘deterministic and stochastic models give similar results in the case of very congested networks’ because a distribution of flows that is very different from the one found in DUE would yield such evident differences between the generalised travel time on the different (hyper)paths that ‘these differences are likely to be correctly perceived by almost all the users’.

In the assumption that public transport services are frequent and/or irregular enough for passengers not to consider explicitly the lines’ timetable when making their travel choice – this is the major assumption of the *frequency-based* (FB) modelling approach – the deterministic RCM is formulated as a shortest hyperpath search. Unlike other examples available in the literature of strategy-based transit assignment, this model is *fully* dynamic, which implies that all the variables considered (frequencies, travel times upon boarding, diversion probabilities, etc.) are *temporal profiles*, i.e. continuous functions of the time of day at which they are evaluated. Moreover, the inclusion/exclusion of a line from the attractive set, as well as the computation of diversion probabilities and waiting times at the stop, depend respectively on the travel time upon boarding, and on frequency and congestion levels at the time of day at which the passenger actually travels.

If the stop layout is such that passengers respect a FIFO queuing mechanism (as is the case for bus stops, for example), congestion levels are estimated by commuting users as the number of vehicle passages of the same line that they must let go before actually being able to board. In the devised assignment model, this parameter ($\kappa_a(\tau)$) is evaluated by means of the *Bottleneck Queue Model* with time-varying exit capacity and is one of the outputs of the supply model for dynamic assignment.

Whereas, in *schedule-based* (SB) assignment models, the graph representation of the supply model (*diachronic graph*) allows for implicit consideration of the time dimension of the problem, the vast majority of FB models are developed in a static setting, even when capacity constraints are considered. Therefore, because transit lines are conceived as a unitary supply facility, with no explicit difference among runs of the same service, only an approximated evaluation of service loads is attained, with average values calculated over the analysis period.

Obviously, the result distortion produced by such approximation increases if the demand profile during the analysis period is very peaked. To overcome this flaw, Schmöcker et al. (2008) develop a quasi-dynamic strategy-based assignment model, where capacity constraints are explicitly considered and passengers who fail to board because of on-board congestion are forced to remain at the stop and queue in a random fashion (mingling). The assumption of mingling allows the formulation of the Network Flow Propagation Model (NFPM) as a Markovian loading process; however, for the same reason, the model cannot be applied to the case where passengers respect a FIFO queuing protocol. Moreover, the proposed model is quasi-dynamic in the sense that it is conceived as a series of steady-state assignments over short intervals of time (15 minutes).

In order to develop a fully dynamic model that can reproduce the build-up and dissipation of FIFO queues of passengers, this thesis has extended to the *strategy-based*

transit assignment the supply model proposed by Meschini et al. (2007). Flows are macroscopic temporal profiles and transit services are represented as a continuous flow of supply with ‘instantaneous capacity’. The latter representation is consistent with the basic assumption of *frequency-based* models that passengers conceive all the runs of a particular line as a unitary supply facility.

Such extension requires the development of a model graph in which the oversaturation queuing time is represented *after* the ‘under-saturation’ waiting time. Although this may be questionable from a phenomenal point of view, it is a valid modelling choice because:

- A strategy-based model with separable queues can be developed in this way, while the overtaking among passengers with different attractive sets of lines would violate the FIFO discipline of queues;
- Transit lines are conceived as a continuous flow of carriers and, as such, the representation of the delay due to the inherent service discontinuity is to be anyhow forced into the model; under this consideration, the under-saturation delay can be added wherever it is more convenient from a modelling point of view, in this case before the queuing process;
- In the *Route Choice Model*, the impedance of waiting is considered through a unique process represented by hyperarcs.

Although the model graph to which the supply model is referred is different from the hypergraph used to model demand phenomena, they are conceived in such a way that only one graphic structure needs to be built and stored for algorithmic purposes.

The proposed assignment model fills in the gap existing in the literature of strategy-based dynamic assignment, clearly reproducing the dynamic changes in route choices produced by variations in network performance due to temporary oversaturation.

Although the consideration of congestion effects in the form of FIFO queues significantly complicates the demand model, because the exponential distribution can no longer be exploited to describe inter-arrival times for irregular services, the new mathematical formulation can still be solved analytically and the existence of at least one equilibrium configuration can be proved. Moreover, an application to the Cracow tram network demonstrated the scalability of the proposed model to real-world networks.

Finally, although the complete implementation of the assignment model is devised only for the case of irregular services with exponentially distributed headways and no real-time information on actual waiting times, the demand model may also easily incorporate other cases of interest (the presence of constant headways and/or countdown displays) and demonstrate the effects on flow distribution brought about by different LoS supplied in the transport system.

6.2. RECOMMENDATIONS FOR FURTHER RESEARCH

Supply-model refinement

The core of the proposed assignment model is the innovative dynamic demand model, which is able to reproduce the effects of formation and dispersion of queues on passengers' route choice.

The model is developed according to an FB approach, because this seems to be the most suitable paradigm to model demand phenomena in densely connected transit networks with highly frequent and/or irregular services, where passengers perceive runs of the same transit line as a unitary supply facility and do not consider timetables, even if available, when making their travel choices.

On the other hand, the FB approach lacks the precision of SB models and is not capable of reproducing phenomena that are crucial both at planning and operational level: service synchronisation; deviation and limitation of specific runs; evaluation of loads and performances of specific runs (especially if bus bunching is observed or if there are transfers from a low-frequency and high-capacity mode to a high-frequency and low-capacity mode).

Therefore, future research should concentrate on overcoming the traditional dichotomy between FB and SB models in order to develop a unified modelling framework for dynamic assignment to transit networks.

To this aim, the supply model needs to be completely detached from the demand model and referred to a different graphic model of the network, thus renouncing the consistency between hypergraph and model graph, which allows the building and storing of only *one* graphic structure for algorithmic purposes. By doing so, it will become possible to develop a very detailed supply model with highly refined time discretisation that is able to reproduce passenger loads on specific runs as well as supply-side dynamic phenomena such as the increase of dwelling, boarding and alighting times due to congestion on-board and at the stop; the formation and dispersion of passenger queues at stops; and the backward propagation of passenger queues from platforms to other parts of the station. The latter phenomenon can be observed during high-peak periods in very congested urban railway and underground stations and is similar to the *spillback congestion*, as defined by Cascetta (2009: p. 468), that can be observed in road networks. In this case, the Bottleneck Queue Model

cannot capture the backward propagation of congestion and more accurate modelling techniques are needed, so far having been implemented for Dynamic Traffic Assignment (DTA) only. For instance, spillback congestion can be reproduced by means of the *General Link Transmission Model* (Gentile, 2008), which solves the Dynamic Network Loading Problem without requiring the very refined spatial discretisation required by the more traditional *Cell Transmission Model* proposed by Daganzo (1994, 1995).

On the other hand, the representation of the demand model in densely connected transit networks does not need to consider individual runs and, while it is advisable to include in the demand model a sound representation of congestion effects on the route choice, as done here, it seems that a less-refined time discretisation could be used for this purpose.

Demand-model refinement

A second major issue for future research to address concerns the refinement of the demand model, with consideration of: the different values of time that passenger groups attach to different phases of the trip; the unreliability of in-vehicle travel times; the effect of real-time travel information provided by handheld devices.

The values of time that different passenger groups perceive while they are waiting, on-board (sitting or standing), transferring, or walking to their origin stop or final destination may result in different (hyper)path sets connecting the same *od* pair. For example, as pointed out in Schmöcker et al. (2013), values of time are generally perceived as higher while waiting at stops than while travelling on-board a vehicle. Moreover, it is reasonable to assume that the disutility of waiting increases if congestion occurs and passengers, unable to board a vehicle, are forced to keep waiting for the next vehicle arrival. Similarly, the value of in-

vehicle travel time might be perceived as higher by standing passengers with respect to sitting passengers, especially when vehicles are very crowded.

Thus, the new formulation of the RCM should also consider these different values of the generalised travel time, which need to be calibrated and validated against data on travel choices actually made by public transport users. Such data can be collected either via e-ticketing systems, as done by Schmöcker et al. (2013), or by conducting stated-preference surveys, as done by Fonzone et al. (2012).

An additional important factor to notice here is that, while the hyperpath paradigm allows for consideration of the effects on route choices generated by *waiting time* uncertainties, some public transport modes (e.g. buses and trams that do not run in segregated lanes) are also subject to in-vehicle travel time uncertainties due to road congestion, bad weather conditions, road works and so forth. Such variations may cause significant differences in the total travel time to destination that are not considered at all in the traditional formulation of strategy-based RCM and assignment, where only the *average* in-vehicle travel time is considered. By contrast, although in-vehicle travel time is an important factor affecting the traveller's route choice, its time variability can be even more important.

Consequently, future research efforts will include in the proposed RCM the consideration of arcs' *reliability*, defined either as the probability that an arc's traversal time exceeds some pre-set threshold that defines the 'normal conditions' (e.g. Bell and Iida, 1997), or as the amount of delay to be expected with a certain level of confidence (e.g. Kaparias et al., 2008). Following the *Link Penalty Method*, formulated by Chen et al. (2006) and further developed and implemented by Kaparias et al. (2007) and Kaparias and Bell (2009), this can be done by defining weights that are used to penalise arcs that are most prone to in-vehicle travel time variability and, ultimately, to exclude travel options whose travel time variability does not satisfy the constraints imposed.

Finally, future research will concentrate on incorporating into the RCM the effects of travel behaviour brought about by ATIS in the form of real-time travel information provided by handheld devices such as smartphones (henceforth, this will also be referred to as *ubiquitous information*).

Certainly, if the information provided is reliable and complete, passengers would have no uncertainties about the transport supply and an SB approach would better approximate their choices (Nuzzolo, 2003). On the other hand, notwithstanding technological improvements, ATIS are still subject to some degree of errors and delays, and it might be plausible to assume that public transport users would rely on the information provided only if it relates to events close in time and in space, as the following example helps to clarify.

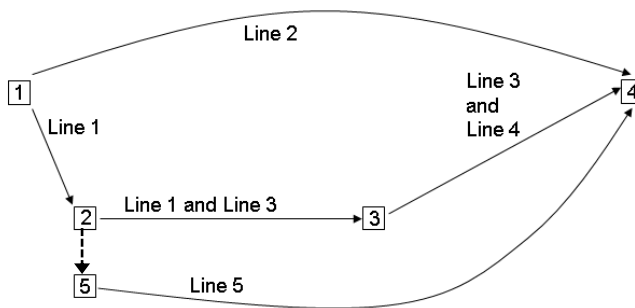


Figure 6-1
Example network: full arrows indicate route sections, while the dashed arrow indicates a pedestrian connection

Let us consider the example network depicted in Figure 6-1 and the supply characteristics listed in Table 2-2.

Table 6-1

Example network: frequencies and in-vehicle travel times; 'Line 0' indicates the pedestrian connection; because pedestrian connections are continuously accessible, the frequency is set equal to infinity

| Line | Section | Frequency (min ⁻¹) | Travel time (min) |
|------|---------|--------------------------------|-------------------|
| 1 | (1, 2) | 1/6 | 25 |
| 1 | (2, 3) | 1/6 | 7 |
| 2 | (1, 4) | 1/6 | 6 |
| 3 | (2, 3) | 1/15 | 4 |
| 3 | (3, 4) | 1/15 | 4 |
| 4 | (3, 4) | 1/3 | 10 |
| 5 | (5, 4) | 1/10 | 14 |
| 0 | (4, 5) | ∞ | 2 |

In an FB assignment, it would be assumed that passengers navigate in the network following their optimal strategy, which means diversions between simple itineraries can occur only at bus stops, as detailed in Chapter 3. For those who board Line 1 at Stop 1, the transfer point selected would be Stop 3, where Line 3 and Line 4 would both be attractive. Line 5 would *not* be included in the strategy.

However, if passengers can access travel information not only when way-side but also whilst on-board, and if such information is reliable when it refers to the neighbourhood of the current position, the optimal strategy expands and includes other decision points, such as where to transfer (Noekel and Webeck, 2009) and possibly also where to begin the trip (i.e. which stop to choose as origin). Therefore, in this 'hyper-strategy', simple itineraries are not restricted to diverging only at bus stops, along attractive lines: rather, each node of the network becomes a diversion node.

Let us follow a hypothetical passenger who travels between nodes 1 and 4 and boards Line 1. The user has installed on his smartphone an application that can provide live bus departures from all the bus stops of the network (Figure 6-2).

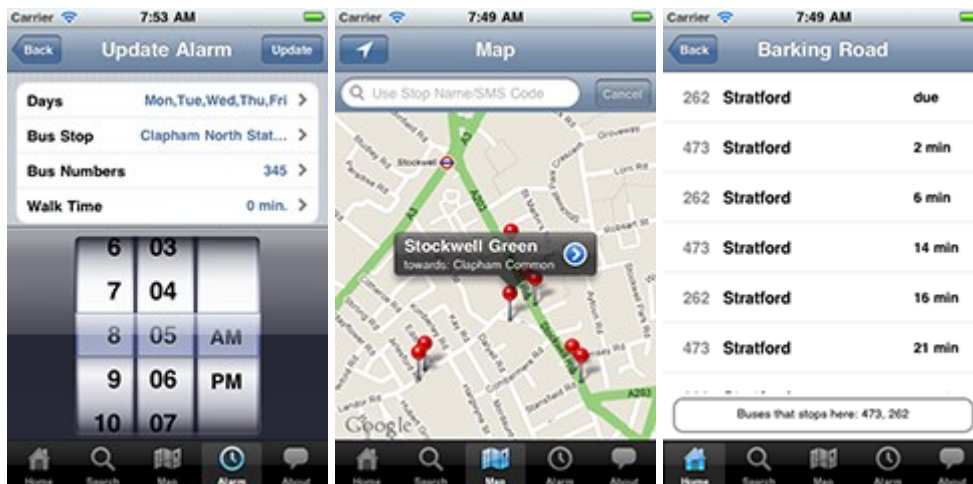


Figure 6-2
Graphic User Interface of the London Bus Stop Live application for iPhone

When approaching Stop 2, instead of passively staying on-board, the passenger would: make a query on his/her smartphone application; see if there is a vehicle of Line 3 approaching Stop 2 or of Line 5 approaching Stop 5; compare these travel options with the one of staying on-board; then decide what to do on the basis of the real-time information acquired. If he/she stays on-board, the expected total travel time to destination is of 17.5 minutes; therefore, he/she might consider alighting if Line 3 is coming in nine minutes or less or if Line 5 is coming in two to three minutes. In other words, the hypothetical passenger would transfer to Line 3 only if the total travel time to destination is of 17 minutes or less and only if, by the time he/she alights and reaches Stop 5, a bus of Line 5 is arriving or will arrive within a minute. Clearly, if no real-time information is provided on-board, the total travel time from Stop 5 to destination – assuming exponentially distributed headways – would be of 24 minutes and this alternative would never be considered attractive.

The assumption underlying the example above is that, notwithstanding the higher degree of information provided, there is some lack of information and uncertainty about the transport supply, which is mainly due to the unreliability of travel predictions. In other words, it is assumed that travellers would trust and exploit to their advantage real-time travel information when they refer to a *close* space–time horizon (i.e. they are approaching Stop 2 on Line 1 and want to know if it is more convenient to stay on-board or transfer). As long as the horizon increases, the prediction becomes less reliable (because unpredicted events, such as vehicle breakdowns or non-recurrent road congestion, may occur) and, thus, passengers would not use ubiquitous information to improve their travel choices.

Consequently, the proposed RCM will be extended in the spirit of Hickman and Wilson (1995) and Gentile et al. (2005) to incorporate this kind of information. In addition, a sensitivity analysis would also be needed to define correctly the boundaries of the space–time horizon within which it is reasonable to assume that ubiquitous travel information affects route choice.

Model validation

Finally, before any practical implementation of the presented model it should be noticed that, in general, its ability to replicate real traffic conditions on the public transport network depends on the adherence of its assumptions to reality, which should be accurately validated.

For example, the significance of the supply assumptions regarding the LoS (namely, the hypothesis of either perfectly irregular services with exponentially distributed headways or perfectly regular services with constant headways) as well as the absence of any synchronisation between the lines' schedule and passengers' arrival at the stop should be re-

validated because the latter studies on the topic date back to the seventies (O'Flaherty and Mangan, 1970; Seddon and Day, 1974).

Also, the assumption underlying the functional form assumed by the demand model (deterministic) should be tested against a sample of real travel choices made by transit users. If the assumption can be accepted, a systematic calibration and validation of the demand-model parameters would not be needed because all the variables considered are 'physical' (for example: the expected number of missed runs of the same line, the average frequency and travel time upon boarding, and the vehicle capacity).

On the other hand, if this is not the case or if, for example, the demand model needs to be refined by considering the different values of time that passenger groups attach to different phases of the trip, then the model would include 'non-physical' parameters that require calibration and validation against two different sets of real data (one for calibration and the other for validation purposes) that directly or indirectly reproduce passengers' choices in the public transport network.

For this purpose, one could, for example, compare the *outputted* temporal profile of the vehicle loads with vehicle loads actually *measured* across the network (for example, through the installation of scales on the vehicles), or the *outputted* temporal profile of the length of the passenger queue at the stops with the length of the queue actually *measured* at stops where CCTV cameras are installed. In any case it is important to notice here that the selection of the real data to consider for calibration and validation purposes should be done on a case-by case basis considering the obvious complications due to data accuracy (for example, correlating the weight of the vehicle with the exact number of people on-board is not a straightforward process) and also a number of different practical factors such as that, because of privacy issues, CCTV footage is usually retained only for a very short period of

time (notably, 14 days at London Underground stations (Transport for London, 2013)), and, thus, an analysis of the data recorded is not always possible.

LIST OF RELEVANT PUBLICATIONS

Parts of this thesis have been previously published in the form of journal publications and scientific conference papers.

Trozzi, V., Bell, M. G. H., Gentile, G. & Haji Hosseinloo, S. (2010a) Dynamic hyperpaths in multimodal transit networks: the stop model with online information. *5th IMA Conference on Mathematics in Transport*, London.

Trozzi, V., Haji Hosseinloo, S., Gentile, G. & Bell, M. G. H. (2010b) Dynamic hyperpaths: the stop model. In: Fusco, G. (ed.), *International Conference on Models and Technology for Intelligent Transportation Systems*, Rome, Italy: Aracne.

Trozzi, V., Gentile, G., Kaparias, I. & Bell, M. G. H. (2012a) Effects of Countdown Displays in Public transport Route Choice under severe overcrowding. *Networks and Spatial Economics* (accepted).

Trozzi, V., Gentile, G., Bell, M. G. H. & Kaparias, I. (2012b) Dynamic User Equilibrium Hyperpaths in Bus Networks with Passengers Queues. *hEART2012 – 1st Symposium of the European Association for Research in Transportation*, 4–7 September 2012, Lausanne, Switzerland.

Trozzi, V., Kaparias, I., Bell, M. G. H. & Gentile, G. (2013a) A dynamic route choice model for public transport networks with boarding queues. *Transportation Planning and Technology*, 36, 44–61.

Trozzi, V., Gentile, G., Bell, M. G. H. & Kaparias, I. (2013b) Dynamic User Equilibrium in public transport Networks with Passenger Congestion and Hyperpaths. *Transportation Research Series* (in press).

Fonzone, A., Schmöcker, J.-D. & Trozzi, V. (2013) Stretching hyperpaths: How ubiquitous real time information can affect transit system use. *hEART2013 – 2nd Symposium of the European Association for Research in Transportation* (submitted).

Gentile, G., Tididi, D., Kucharski, R. & Trozzi, V. (2013) Combining frequency and schedule based approaches in a dynamic assignment model for highly congested transit networks. *hEART2013 – 2nd Symposium of the European Association for Research in Transportation* (submitted).

More specifically, Trozzi et al. (2010a) introduces the new stop model for congested public transport networks and discusses the effect on waiting time and passenger distribution brought about by different layouts of the stop, whereas in Trozzi et al. (2010b) the effect of information and regularity are considered for the isolated stop.

The complete demand model for dynamic transit assignment is formulated in Trozzi et al. (2013a), where considerations about the properties of Erlang and Exponential distributions (as in Section 3.4.1) are also presented. On the other hand, in Trozzi et al. (2012a) the impact of wayside information on the flow distribution is analysed.

Finally, the supply model for dynamic assignment and demand–supply interaction are presented in Trozzi et al. (2012b) and Trozzi et al. (2013b), where details are given of the extension of the supply model, originally presented by Meschini et al. (2007) for dynamic transit assignment without hyperpaths.

Some of the future research streams are also already included in the most recent works listed here. More specifically, Trozzi et al. (2013b) already includes the idea of considering travel costs, as opposed to travel times, weighted by specific parameters which depend on the specific part of the journey under consideration (walking, waiting, queuing etc.).

Likewise, the idea of detaching the graphic models, which represent demand and supply phenomena, in order to allow for greater modelling flexibility is already included in the same paper (Trozzi et al., 2013b) and is further extended in Gentile et al. (2013), where it

is proposed to extend the GLTM to model dynamic supply-side phenomena in the public transport network.

Finally, research on extending the concept of hyperpaths to the case where ubiquitous information is provided by means of handheld devices is presented in Fonzone et al. (2013).

7. REFERENCES

- Azibi, R. & Vanderpooten, D. (2002) Construction of rule-based assignment models. *European Journal of Operational Research*, 138, 274–293.
- Bell, M. G. H. (2003) Capacity constrained transit assignment models and reliability. In: Bell, M. G. H. & Lam, W. H. K. (eds.), *Advanced Modelling for Transit Operations and Service Planning*, Oxford: Pergamon.
- Bell, M. G. H. & Iida, Y. (1997) Network reliability. In: Bell, M. G. H. & Iida, Y. (eds.), *Transportation network analysis*, Chichester: Wiley and Sons.
- Bell, M. G. H. & Lam, W. H. K. (2003) *Advanced Modelling for Transit Operations and Planning*, Oxford: Pergamon.
- Bellei, G., Papola, N. & Gentile, G. (2005) A within-day dynamic traffic assignment model for urban road networks. *Transportation Research Part B: Methodological*, 39, 1–29.
- Billi, C., Gentile, G., Nguyen, S., Pallottino, S. & Barrett, K. (2004) Rethinking the wait model at transit stops. Proceedings of TRISTAN V – Triennial Symposium on Transportation Analysis, 2004.
- Bouzaïene-Ayari, B. (1988) *Modélisation des arrêts multiples d'autobus pour les réseaux de transport en commun*, École Polytechnique de Montréal.
- Bouzaïene-Ayari, B., Gendreau, M. & Nguyen, S. (2001) Modeling Bus Stops in Transit Networks: A Survey and New Formulations. *Transportation Science*, 35, 304–321.
- Boyce, D., Lee, D.-H. & Ran, B. (2001) Analytical Models of the Dynamic Traffic Assignment Problem. *Networks and Spatial Economics*, 1, 377–390.
- Cantarella, G. E. (1997) A general fixed-point approach to multimode multi-user equilibrium assignment with elastic demand. *Transportation Science*, 31, 107.
- Cantarella, G. E., Gentile, G. & Veloná, P. (2010) Uniqueness of Stochastic User Equilibrium. *5th IMA Conference on Mathematics in Transportation*, April 2010 London,

UK.

Cascetta, E. (2001) *Transportation systems engineering: theory and methods*, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Cascetta, E. (2009) *Transportation Systems Analysis – Models and Applications*, New York: Springer US.

Cascetta, E., Nuzzolo, A., Russo, F. & Vivetta, A. (1996) A Modified Logit Route Choice Model Overcoming Path Overlapping Problems: Specification and Some Calibration Results for Interurban Networks. In J. B. Lesort (ed.), *13th International Symposium on Transportation and Traffic Theory*, Pergamon.

Cats, O. (2011) *Dynamic Modelling of Transit Operations and Passenger decisions*, PhD thesis, KTH – Royal Institute of Technology.

Cepeda, M., Cominetti, R. & Florian, M. (2006) A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. *Transportation research Part B: Methodological*, 40, 437–459.

Chabini, I. (1998) Discrete dynamic shortest path problems in transportation applications: complexity and algorithms with optimal running time. *Transportation Research Record: Journal of the Transportation Research Board*, 1645, 170–175.

Chen, Y., Bell, M. G. H., Wang, D. & Bogenberger, K. (2006) Risk-averse time-dependent route guidance by constrained dynamic A* search in decentralized system architecture. *Transportation Research Record: Journal of the Transportation Research Board*, 1944, 51–57.

Chriqui, C. & Robillard, P. (1975) Common bus lines. *Transportation Science*, 9, 115–121.

Cominetti, R. & Correa, J. (2001) Common-Lines and Passenger Assignment in Congested Transit Networks. *Transportation Science*, 35, 250–267.

Daganzo, C. F. (1994) The cell transmission model: a dynamic representation of highway traffic consistent with hydrodynamic theory. *Transportation Research, B* 28, 269–287.

Daganzo, C. F. (1995) The cell transmission model, part II: network traffic. *Transportation Research B*, 29, 79–93.

- Daganzo, C. F. & Sheffi (1977) On Stochastic Models of Traffic Assignment. *Transportation Science*, 11, 253–274.
- De Cea, J. & Fernandez, E. (1993) Transit Assignment for Congested public transport Systems: An Equilibrium Model. *Transportation Science*, 27, 133–147.
- Dial, R. B. (1967) Transit pathfinder algorithm. *Highway Research Record*, 205, 67–85.
- Fearnside, K. & Draper, D. P. (1971) Public transport assignment – a new approach. *Traffic Engineering and Control*, 12, 298–299.
- Florian, M. (2003) Frequency based transit route choice models. In: Bell, M. G. H. & Lam, W. H. K. (eds.), *Advanced modelling for transit operations and service planning*, Oxford: Pergamon.
- Fonzone, A. & Bell, M. G. H. (2010) Bounded rationality in hyperpath assignment: the locally rational traveller model. *Transportation Research Board (TRB) 2010 – 89th Annual Meeting*, Washington, D.C.
- Fonzone, A., Schmöcker, J.-D., Bell, M. G. H., Gentile, G., Kurauchi, F., Nokel, K. & Wilson, N. H. M. (2010) Do ‘hyper-travellers’ exist? – Initial results of an international survey on public transport user behaviour. *12th WCTR*, Lisbon, Portugal.
- Fonzone, A., Schmöcker, J.-D., Kurauchi, F. & Hemdam, S. M. H. (2012) Determinants of Hyperpath Choice in Transit Networks. *12th Conference on Advanced Systems for Public Transport (CASPT)*, Santiago de Chile, Chile.
- Gallo, G., Longo, G., Pallottino, S. & Nguyen, S. (1993) Directed Hypergraphs and Applications. *Discrete Applied Mathematics*, 42, 177–201.
- Gendreau, M. (1984) *Étude approfondie d’un modèle d’équilibre pour l’affectation des passagers dans les réseaux de transport en commun*, PhD, Université de Montréal.
- Gentile, G. (2008) The General Link Transmission Model for Dynamic Network Loading and a comparison with the DUE algorithm. *2nd International Symposium on Dynamic Traffic Assignment – DTA 2008*, Leuven, Belgium.
- Gentile, G., Nguyen, S. & Pallottino, S. (2005) Route Choice on Transit Networks with Online Information at Stops. *Transportation Science*, 39, 289–297.

Grotenhuis, J. W., Wiegmansa, B. W. & Rietveld, P. (2007) The desired quality of integrated multimodal travel information in public transport: Customer needs for time and effort savings. *Transport Policy*, 14, 27–38.

Hamdouch, Y., Ho, H. W., Sumalee, A. & Wang, G. (2011) Schedule-based transit assignment model with vehicle capacity and seat availability. *Transportation Research Part B: Methodological*, 45, 1805–1830.

Hamdouch, Y. & Lawphongpanich, S. (2008) Schedule-based transit assignment model with travel strategies and capacity constraints. *Transportation Research Part B: Methodological*, 42, 663–684.

Hamdouch, Y., Marcotte, P. & Nguyen, S. (2004) Capacitated transit assignment with loading priorities. *Mathematical Programming B*, 101, 205–230.

Hickman, M. D. & Wilson, N. H. M. (1995) Passenger travel time and path choice implications of real-time transit information. *Transportation Research Part C: Emerging Technologies*, 3, 211–226.

House of Commons (HoC) Transport Committee (2003) Overcrowding on Public Transport – Seventh Report of Session 2002–03.

Huang, H.-J. (2002) Pricing and logit-based mode choice models of a transit and highway system with elastic demand. *European Journal of Operational Research*, 140, 562–570.

Huang, R. & Peng, Z. (2002) Schedule-based path finding algorithms for transit trip-planning systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1783, 142–148.

Kaparias, I. & Bell, M. G. H. (2009) Testing a reliable in-vehicle navigation algorithm in the field. *IET Intelligent Transport Systems*, 3(3), 314–324.

Kaparias, I., Bell, M. G. H. & Belzner, H. (2008) A new measure of travel time reliability for in-vehicle navigation systems. *Journal of Intelligent Transportation Systems*, 12, 202–211.

Kaparias, I., Bell, M. G. H., Bogenberger, K. & Chen, Y. (2007) An approach to time-dependence and reliability in dynamic route guidance. *Transportation Research Board (TRB) 2007 – 86th Annual Meeting*, Washington, D.C.

- Kleinrock, L. (1975) *Queuing Systems*, New York: John Wiley.
- Kurauchi, F., Bell, M. G. H. & Schmöcker, J. D. (2003) Capacity Constrained Transit Assignment with Common Lines. *Journal of Mathematical Modelling and Algorithms*, 2, 309–327.
- Lam, W. H. K., Gao, Z. Y., Chan, K. S. & Yang, H. (1999) A stochastic user equilibrium assignment model for congested transit networks. *Transportation Research Part B: Methodological*, 33, 351–368.
- Lam, W. H. K., Zhou, J. & Shen, Z.-H. (2002) A capacity restraint transit assignment with elastic line frequency. *Transportation Research Part B: Methodological*, 36, 919–938.
- Larson, R. C. & Odoni, A. R. (1981) *Urban Transport Operation research*, Prentice-Hall.
- Last, A. & Leak, S. E. (1976) A bus model. *Traffic Engineering and Control*, 17, 14–17.
- Leurent, F. & Benezech, V. (2011) The Passenger Stock and Attractivity Threshold model for traffic assignment on a transit network with capacity constraint. *Transportation Research Board (TRB) 2011 – 90th Annual Meeting*, Washington, D.C.
- Leurent, F., Chandakas, E. & Poulhès, A. (2011) User and service equilibrium in a structural model of traffic assignment to a transit network. *Procedia Social and Behavioral Sciences*, 20, 495–505.
- Loève, M. (1978) Chapter 27: Concept of Conditioning. In: Loève, M., *Probability Theory Vol. II* (4th ed.), New York: Springer, 3–12.
- Marcotte, P., Nguyen, S. & Schoeb, A. (2004) A Strategic Flow Model of Traffic Assignment in Static Capacitated Networks. *Operations Research*, 52, 191–212.
- Marguier, P. H. J. & Ceder, A. (1984) Passenger Waiting Strategies for Overlapping Bus Routes. *Transportation Science*, 18, 207–230.
- Melotto, D. (2004) *Calore atteso condizionato e sue applicazioni*, MSc thesis, Università degli studi di Padova.
- Meschini, L., Gentile, G. & Papola, N. (2007) A frequency based transit model for dynamic traffic assignment to multimodal networks. In: Allsop, R., Bell, M. G. H. & Heydecker, B. G. (eds.), *17th International Symposium on Transportation and Traffic Theory*, London:

Elsevier.

Nachtigall, K. (1995) Time depending shortest path problems with applications to railway networks. *European Journal of Operational Research*, 83, 154–166.

Nguyen, S. & Pallottino, S. (1988) Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research*, 37, 176–186.

Nguyen, S. & Pallottino, S. (1989) Hyperpaths and shortest hyperpaths. *Combinatorial Optimization*, Berlin and Heidelberg: Springer.

Nguyen, S., Pallottino, S. & Gendreau, M. (1998) Implicit enumeration of hyperpaths in a logit model for transit networks. *Transportation Science*, 32, 54–64.

Nielsen, L. R. (2004) *Route Choice in Stochastic Time-Dependent Networks*, University of Aarhus.

Nielsen, O. A. (2000) A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B: Methodological*, 34, 377–402.

Nielsen, O. A. (2004) A large-scale stochastic multi-class schedule-based transit model with random coefficients. In: Wilson, N. H. M. & Nuzzolo, A. (eds.), *Schedule-based Dynamic Transit Modelling: Theory and Applications*, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Noekel, K. & Wekeck, S. (2007) Choice models in frequency-based transit assignment. *European Transport Conference 2007*.

Noekel, K. & Wekeck, S. (2008) Choice set structuring in frequency-based transit assignment. *European Transport Conference 2008*.

Noekel, K. & Wekeck, S. (2009) Boarding and Alighting in Frequency-based Transit Assignment. *Transportation Research Record: Journal of the Transportation Research Board*, 2111, 60–67.

Nuzzolo, A. (2003) Transit Path Choice and Assignment Model Approaches. In: Bell, M. G. H. & Lam, W. H. K. (eds.), *Advanced Modelling for transit Operations and Service Planning*, Oxford: Pergamon.

Nuzzolo, A., Crisalli, U. & Rosati, L. (2012) A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transportation Research Part C: Emerging Technologies*, 20, 16–33.

Nuzzolo, A., Russo, F. & Crisalli, U. (2001) A Doubly Dynamic Schedule-based Assignment Model for Transit Networks. *Transportation Science*, 35, 268–285.

Nuzzolo, A., Russo, F. & Crisalli, U. (2003) *Transit Network Modelling: the schedule-based dynamic approach*, Franco Angeli – Collana Trasporti.

O’Flaherty, C. A. & Mangan, D. O. (1970) Bus passenger Waiting Times in Central Areas. *Traffic Engineering and Control*, 11, 419–421.

Oxford Economic Forecasting (2003) *The Economic Effects of Transport Delays on the City of London*. London: Corporation of London.

Papola, N., Filippi, F., Gentile, G. & Meschini, L. (2009) Schedule-based transit assignment: a new dynamic equilibrium model with vehicle capacity constraints. In: Wilson, N. H. M. & Nuzzolo, A. (eds.), *Schedule-Based Modeling of Transportation Networks*, New York: Springer US.

Poon, M. H., Tong, C. O. & Wong, S. C. (2004) A dynamic schedule-based model for congested transit networks. *Transportation Research Part B: Methodological*, 38, 343–368.

Powell, W. B. (1981) *Stochastic delays in transportation terminals: new results in the theory and application of bulk queues*, PhD, M.I.T.

Pucher, J., Korattyswaroopam, N. & Ittyerah, N. (2004) The crisis of public transport in India: overwhelming needs but limited resources. *Journal of public transportation*, 7, 95–113.

Pyrga, E., Schulz, F., Wagner, D. & Zaroliagis, C. (2008) Efficient Models for timetable information in public transportation Systems. *Journal of Experimental Algorithmics*, 12.

Rieser, M., Grether, D. & Nagel, K. (2009) Adding mode choice to multi-agent transport simulation. *Transportation Research Record: Journal of the Transportation Research Board*, 2132, 50–58.

Rudnicki, A., Bauer, M., Kucharski, R. & Szarata, A. (2011) Transportation model for Krakow agglomeration, Cracow University of Technology.

Sheffi, Y. & Powell, W. B. (1981) A Comparison of Stochastic and Deterministic Traffic Assignment over Congested networks. *Transportation Research Part B: Methodological*, 15, 191–207.

Schmöcker, J.-D. (2006) *Dynamic Capacity Constrained Transit Assignment*, PhD, Imperial College London.

Schmöcker, J.-D., Bell, M. G. H. & Kurauchi, F. (2008) A quasi-dynamic capacity constrained frequency-based transit assignment model. *Transportation Research Part B: Methodological*, 42, 925–945.

Schmöcker, J.-D., Bell, M. G. H. & Lee, C. (2002) An application of congested transit network loading with the Markov chain approach. *9th Meeting of the EURO Working Group on Transportation 'Intermodality, Sustainability and Intelligent Transport Systems'*, Bari Italy.

Schmöcker, J.-D., Shimamoto, H. & Kurauchi, F. (2013) Generation and Calibration of Transit Hyperpaths. *Procedia – Social and Behavioral Sciences* (accepted).

Seddon, P. A. & Day, M. P. (1974) Bus passenger waiting times in Greater Manchester. *Traffic Engineering and Control*, 15, 442–445.

Sohail, M., Maunder, D. & Cavill, S. (2006) Effective regulation for sustainable public transport in developing countries. *Transport Policy*, 13, 177–190.

Spiess, H. (1983) *On optimal route choice strategies in transit networks*, Montréal: Université de Montréal, Centre de recherche sur les transports.

Spiess, H. (1984) *Contributions à la théorie et aux outils de planification des réseaux de transport urbain*, PhD thesis, Université de Montréal.

Spiess, H. & Florian, M. (1989) Optimal strategies: A new assignment model for transit networks. *Transportation Research Part B: Methodological*, 23, 83–102.

Sumalee, A., Tan, Z. & Lam, W. H. K. (2009a) Dynamic stochastic transit assignment with explicit seat allocation model. *Transportation Research Part B: Methodological*, 43, 895–912.

Sumalee, A., Uchida, K. & Lam, W. H. K. (2009b) Stochastic Multi-modal Transport network under demand uncertainties and Adverse Weather condition. In: Lam, W. H. K. & Lo, H. K. (eds.), *International Symposium on Transportation and Traffic Theory*, 16–18 July 2009, Hong Kong: Springer Science.

Sumi, T., Matsumoto, Y. & Miyaki, Y. (1990) Departure time and route choice of commuters on mass transit systems. *Transportation Research Part B: Methodological*, 24, 247–262.

Szeto, W. & Wong, S. (2012) Dynamic traffic assignment: model classifications and recent advances in travel choice principles. *Central European Journal of Engineering*, 2, 1–18.

Teklu, F. (2008) A Stochastic Process Approach for Frequency-based Transit Assignment with Strict Capacity Constraints. *Networks and Spatial Economics*, 8, 225–240.

Tong, C. O. & Richardson, A. J. (1984) A Computer Model for Finding the Time-Dependent Minimum Path in a Transit System with Fixed Schedules. *Journal of Advanced Transportation*, 18, 145–161.

Transport for London (2013) How long does TfL keep CCTV or ANPR footage? Available at <http://www.tfl.gov.uk/termsandconditions/22246.aspx>, accessed 10/10/2013.

Vovsha, P. (1997) The cross-nested logit model: application to mode choice in the Tel-Aviv metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board*, 1607, 6–15.

Wardrop, J. (1952) Some theoretical aspects of road traffic research. In: *ICE Proceedings: Part II, Engineering Divisions*, 1, 325–362.

Williams, H. C. W. L. (1977) On the formulation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9, 285–344.

Wong, S. C. & Tong, C. O. (1999) A stochastic transit assignment model using a dynamic schedule-based network. *Transportation Research Part B: Methodological*, 33, 107–121.

Wu, J. H., Chen, Y. & Florian, M. (1998) The continuous dynamic network loading problem: a mathematical formulation and solution method. *Transportation Research Part B: Methodological*, 32, 173–187.

Yang, L. & Lam, W. H. K. (2006) Probit-Type Reliability-Based Transit Network Assignment. *Transportation Research Record: Journal of the Transportation Research Board*, 1977, 154–163.

Ziliaskopoulos, A. (1994) *Optimal path algorithms on multidimensional networks: analysis, design, implementation and computational experience*, PhD, University of Texas.

Zografos, K. G. & Androutsopoulos, K. N. (2008) Algorithms for Itinerary Planning in Multimodal Transportation Networks. *IEEE Transactions on Intelligent Transportation Systems*, 9, 175–184.

8. APPENDIX

The detailed *od* matrix is given in the following. Each entry of the matrix refers to the entire analysis period.

