



Modelling congestion in passenger transit networks

Ektoras Chandakas

► To cite this version:

Ektoras Chandakas. Modelling congestion in passenger transit networks. Architecture, space management. Université Paris-Est, 2014. English. NNT : 2014PEST1011 . tel-01148406

HAL Id: tel-01148406

<https://theses.hal.science/tel-01148406v1>

Submitted on 4 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ECOLE DOCTORALE VILLE, TRANSPORTS ET TERRITOIRES

LABORATOIRE VILLE, MOBILITE, TRANSPORTS

THESE DE DOCTORAT

SPECIALITE TRANSPORT

MODELLING CONGESTION IN PASSENGER TRANSIT NETWORKS

Présentée et soutenue publiquement par

Ektoras CHANDAKAS

Le 01/04/2014

Composition du Jury:

M. Eric KROESS	Professeur associé à l'Université VU, <i>président</i>
M. Achille FONZONE	Chercheur à l'Université Napier d'Edinburgh, <i>rapporteur</i>
M. Guido GENTILE	Professeur à l'Université de Rome, La Sapienza, <i>rapporteur</i>
M. Michael FLORIAN	Professeur émérite à l'Université de Montréal, <i>examineur</i>
M. Daniel GRAHAM	Professeur à l'Imperial College, Londres, <i>examineur</i>
M. Jean-Patrick LEBACQUE	Chercheur à l'IFSTTAR, <i>examineur</i>
M. Fabien LEURENT	Professeur à l'Ecole des Ponts ParisTech, <i>directeur de thèse</i>

Στην προσωπική μας Ιθάκη

Short Abstract

A structural model is provided to capture capacity phenomena in passenger traffic assignment to a transit network. That has been founded on a bi-layer representation of the transit network: on the lower layer the model addresses each network sub-system (line, station and access-egress) separately, on the basis of specific capacity effects; on the upper layer a leg-based representation is used with respect to the sub-systems' costs and operating characteristics to address the trip maker's path choices.

The network model, named CapTA (for Capacitated Transit Assignment), involves a line sub-model which amounts to a sophisticated cost-flow relationship. The capacity constraints pertain to the interaction of passenger and vehicle traffic: vehicle seat capacity drives the in-vehicle comfort; vehicle total capacity determines the in-vehicle comfort and also the platform waiting; passenger flows at vehicle alighting and boarding have an impact on the dwell time, which in turn drives track occupancy and service frequency of any service using the station track infrastructure; the service frequency influences the service capacity and platform waiting. These phenomena are dealt with by line of operations on the basis of a set of local models yielding specific flows and costs.

Equivalently, a station sub-model addresses specific capacity constraints and yields the local walking conditions, sensible to the interaction of the passengers in the interior of a station: the instant bottleneck created at the entry of the circulation elements delays the evacuation of the station platforms; the passenger density and presence of heterogeneous passenger flows slows down the passengers who circulate in the station; and the presence of real-time information influences the decision making process of the transit users exposed to. These effects do not only impact locally the in-station path choice, but most notably they modify the choices of transit routes and itineraries on a network level.

The Paris Metropolitan Region provides an ideal application field of the capacity constrained transit assignment model. It is mainly used as a showcase of the simulation capabilities and of the finesse of the modelling approach. The transit network involves 1 500 bus routes together with 260 trains routes that include 14 metro lines and 4 light rail lines. Traffic assignment at the morning peak hour is characterized by heavy passenger loads along the central parts of the railway lines. Increased train dwelling, due to boarding and alighting flows, and reduction in the service frequency impact the route and the line capacity. The generalized time of a transit trip is impacted mainly though its in-vehicle comfort component. Detailed results have been provided for the RER A, the busiest commuter rail line in the transit network.

Résumé Court

Un modèle structurel est fourni afin d'appréhender les phénomènes de capacité dans un modèle d'affectation de flux de voyageurs sur un réseau de transport collectifs. Cela a été fondé sur une représentation du réseau de transports collectifs en deux couches : sur la couche inférieure, le modèle traite séparément chaque sous système du réseau (ligne, station et rabattement) en fonction des effets de capacité spécifiques ; sur la couche supérieure, le choix d'itinéraire d'un voyageur individuel est adressée par une représentation du réseau en leg (ou segment de ligne) en utilisant le coût et les caractéristiques opérationnelles des sous-systèmes respectifs.

Le modèle de réseau, nommé CapTA (pour Capacitated Transit Assignment – Affectation en Capacité des Transports Collectifs), inclut un sous-modèle de ligne qui revient à une fonction de coût sophistiquée. Les contraintes de capacité correspondent à l'interaction entre voyageurs et trafic de véhicule : la capacité en places assises d'un véhicule influence le confort à bord ; the flux de passagers en montée et descente ont une incidence au temps de stationnement qui de son côté agit sur l'occupation de la voie et la fréquence de service de toutes les missions qui occupes l'infrastructure à quai ; la fréquence de service influence la capacité de services et l'attente en plateforme. Ces phénomènes sont traités par ligne d'exploitation sur la base d'un ensemble des modèles locaux qui rendent de flux et de coût spécifiques.

De manière équivalente, le modèle de station traite de contraintes de capacité spécifiques et évalue les conditions locales de marche, qui est sensible aux interactions des voyageurs à l'intérieur de la station : le goulot instantané à l'entrée d'un élément de circulation retard l'évacuation de la plateforme ; la densité de voyageurs et l'hétérogénéité des leur flux ralentit les voyageurs qui circulent dans une station ; la présence de l'information en temps réel

influence le processus de décision des voyageurs. Ces effets n'ont pas seulement un impact sur le choix d'itinéraire à l'intérieure de la station, mais notamment ils modifient les choix de service sur le niveau du réseau.

La Région Ile-de-France fournit un champ d'application idéal pour un modèle d'affectation de flux de voyageurs en transport collectifs sous contraintes de congestion. Plus précisément, il est utilisé dans le cadre du modèle CapTA pour illustrer les capacités de simulation et la finesse de l'approche de modélisation adoptée. Le réseau de transports collectifs contient 1 500 missions de cars et autocars, tout comme 260 missions ferroviaires et inclut 14 lignes de métro et 4 lignes de tramway. L'affectation de trafic à l'heure de pointe du matin est caractérisée d'une charge importante en voyageurs sur les sections centrales de lignes ferroviaires qui traversent la ville. Un temps de stationnement élevé, en raison de flux de montée et descente, et la réduction de la fréquence de service impactent la capacité des missions et des lignes. Le temps généralisé d'un trajet est impacté notamment de sa composante de confort à bord. Des résultats détaillés sont présentés sur le RER A, la ligne la plus chargée du réseau ferroviaire régional.

Acknowledgements

Proposing a PhD research to a young engineer is a great honour. I would like to express my profound gratitude to Fabien Leurent, who directed this thesis, for the confidence he showed on me, his constant availability, his exemplary guidance and his encouragement throughout the course of this thesis.

I would like to thank Achille Fonzone and Guido Gentile for the attention they brought to my work and for their valuable and constructive comments. My grateful thanks are also extended to Michael Florian, Daniel Graham, Jean-Patrick Lebacque, who kindly accepted to participate to my examination board, and to Eric Kroess who accepted to preside over it.

It is impossible to imagine how I can thank my colleagues at the LVMT without at the same time repeating what so others before me have already said. This particular mixture of geographers, sociologists, and engineers, to name some, had been a constant source of inspiration. I owe a deeper debt to Alexis Poulhès. His determination and efficiency over the software implementation highly accelerated this research and had been a key for its success.

I would like to express my gratitude for the fellow researchers of the COST Action TU1004. Our discussions during our meetings were valuable. They helped me formulate and hone many of my views present in this research.

I owe a more intimate debt to my friends who showed an exquisite combination of patience and encouragement throughout these years. I would like to thank David, Valia, Theo, Dimitris, Nastassia, Angel, Javier and so many others.

Finally, the gratitude for the support, patience and trust I owe to my family during this work, and not only, are beyond my limited capacity of expression. I trust they understand how much I owe them. I would like to tell them: « Σας ευχαριστώ απο καρδιάς ».

Table of Contents

SHORT ABSTRACT	7
RESUME COURT	9
ACKNOWLEDGEMENTS.....	11
TABLE OF CONTENTS.....	13
LIST OF FIGURES	17
LIST OF TABLES	19
NOTATIONS.....	21
 ABSTRACT	 25
 1 GENERAL INTRODUCTION.....	 31
1.1 Context and Planning Issues	31
1.2 Problem Statement and Research Objectives.....	34
1.3 Research Approach	36
1.4 Original Contributions	39
1.5 Outline of the Dissertation	41
 PART I: A LITERATURE REVIEW ON TRANSIT ASSIGNMENT MODELLING.....	 45
2 A REVIEW OF THE PASSENGER FLOW ASSIGNMENT MODELS FOR TRANSIT NETWORKS	47
2.1 Introduction.....	47
2.2 First Generation: the Initial Steps	48
2.3 Second Generation: Setting the Foundations	52
2.4 Third Generation: Integrating Capacity in Macroscopic Static Assignment Models	57
2.5 Fourth Generation: Emergence of Dynamic Models	60
2.6 Fifth Generation: Simulation-Based Models	67
2.7 An Overview of the Transit Assignment Models	71
 3 MODELLING CAPACITY EFFECTS IN TRANSIT ASSIGNMENT: A STATE OF THE ART75	
3.1 Introduction.....	75
3.2 A Typology of Capacity Effects in Public Transportation.....	76
3.3 Passenger Total Capacity	79

3.4 Seat Availability and In-vehicle Comfort	90
3.5 Conclusion	94
PART II: RESTRAINED MODELS	97
4 MODELLING THE LINE SYSTEM ON A STRUCTURAL TRANSIT NETWORK	99
4.1 Introduction.....	99
4.2 The Line System and the Model Framework.....	100
4.3 The In-Vehicle Comfort.....	105
4.4 The Total Passenger Capacity of the Vehicles.....	111
4.5 The Station Platform Occupancy	118
4.6 Combinations of Capacity Effects	123
4.7 Conclusion	124
5 A TRANSIT STATION MODEL ON A STRUCTURAL TRANSIT NETWORK	125
5.1 Introduction.....	125
5.2 Systemic Analysis of a Station.....	127
5.3 Network Topology	142
5.4 The Station Model Formulation.....	144
5.5 The Articulation with the Network Model.....	152
5.6 An Application Instance of the Station Model.....	154
5.7 Conclusion	156
PART III: THE NETWORK MODEL – THE CAPTA SIMULATOR.....	157
6 A SYSTEMIC REPRESENTATION OF THE CAPTA MODEL.....	159
6.1 Introduction.....	159
6.2 Systemic Analysis of the Transit Assignment Models	160
6.3 The Bi-layer Representation of Supply and Demand in CapTA.....	162
6.4 Demand-Side Model Components	164
6.5 Supply-Side Model Components	167
6.6 The Hierarchy of Models in CapTA	169
6.7 The Modelling Differences of the CapTA Model and Some Development Perspectives.....	170
7 MATHEMATIC CHARACTERIZATION OF THE CAPTA MODEL.....	173
7.1 Introduction.....	173
7.2 A Literature Review on the Mathematic Formulation and the Characterization of Traffic Equilibrium	174
7.3 Systemic Characterization of the CapTA Model	189
7.4 Mathematical Analysis of the CapTA Model	197
7.5 Conclusion	215
8 ALGORITHMIC IMPLEMENTATION AND COMPUTATIONAL ASPECTS OF THE CAPTA MODEL	217
8.1 Introduction.....	217
8.2 The Transit Network Model.....	219
8.3 The Line Model Algorithm	223
8.4 The Auxiliary Network State	235
8.5 Software Architecture	239
8.6 Conclusion	242

PART IV: APPLICATION OF THE CAPTA SIMULATOR.....	245
9 A SIMULATION OF THE TRANSIT NETWORK OF THE PARIS METROPOLITAN REGION.....	247
9.1 Introduction.....	247
9.2 The Transport Demand in the Paris Metropolitan Region	248
9.3 The Transit Supply in the Paris Metropolitan Region	250
9.4 Simulation Characteristics and Model Variants.....	253
9.5 Passenger Flows on the Transit Network.....	256
9.6 The Operation of Transit Routes Under Capacity Constraints	259
9.7 Impacts on Users.....	261
9.8 The Impact of In-Vehicle Comfort on the Users	263
9.9 Conclusion	265
10 THE RESULTS OF THE CAPTA MODEL ON A RESTRAINED NETWORK.....	269
10.1 Introduction.....	269
10.2 Transit Network, Trip Demand and Supply of Services	270
10.3 Passenger Flows on the Network	274
10.4 Dwell Times and Service Frequency	276
10.5 Residual Capacity and Boarding Flows on a Vehicle.....	279
10.6 Passenger Stock and Probabilities of Immediate Boarding	280
10.7 Average Generalized Cost	284
10.8 The Impact of a New Rolling Stock.....	286
10.9 Conclusion	287
11 GENERAL CONCLUSION	289
11.1 Introduction.....	289
11.2 Main Results and Conclusions	289
11.3 Discussion and Perspectives	295
REFERENCES.....	299
ANNEXES.....	313
ANNEX A: VEHICLE CAPACITY ON THE GREATER PARIS TRANSIT NETWORK....	315
ANNEX B: ON THE IN-VEHICLE COMFORT MODEL WITH FOLDING SEATS.....	323
ANNEX C: RESEARCH ACTIVITY	331

List of Figures

Figure 1: Five Generations of Transit Assignment Modelling.....	71
Figure 2: A schematic representation of the elements of the transit systems and how they influence the capacity effects (adapted from Leurent, 2011a)	78
Figure 3: The effective frequencies function for various values of beta	81
Figure 4: Overview of the line system (K for capacity)	102
Figure 5: The process of cost evaluation of a line leg (i,s) – TT for Travel Time	104
Figure 6: Density of standing passengers on a vehicle in relation to the volume of on-board passengers and the effect of the folding seats	108
Figure 7: Cumulative distribution of in-vehicle Comfort States according to the vehicle volume	110
Figure 8: The service network of the application instance	113
Figure 9: Average Expected Travel Time (blue) and In-Vehicle Travel Time (red) per (a) stock of passengers waiting – left - and (b) exogenous passenger flow – right.	114
Figure 10: Relation between Passenger Stock waiting to board and arriving passenger flow x_{BD} ...	115
Figure 11: The cumulative passenger flow share of each transit service	116
Figure 12: The marginal generalized time of an additional passenger in the AB pair for the AB (red), AD (blue) pairs and the marginal total generalized time (green)	117
Figure 13: Restrained Frequency at station in relation to the increase on Dwell Time for Metro lines M1, M13, M14 and the RER A	122
Figure 14: Flowchart of a transit station	128
Figure 15: Elevators for primary use in Paris, Auber station (source: author)	132
Figure 16: Schematic representation of a station platform.....	134
Figure 17: An example of real-time information: (a) station screens before the fare control barriers at La Defense station (source: author) – left – and (b) smartphone application (source: SNCF Direct) – right	137
Figure 18: Map of day buses from Trafalgar Square (source: TfL)	139
Figure 19: The bottleneck created when exiting a station platform	146
Figure 20: Average Additional Time of an alighting passenger to exit a station platform	148
Figure 21: Speed - flow diagram for the station's circulation elements	149
Figure 22: Travel time multiplier diagram in function of the flow-to-capacity ratio.....	150
Figure 23: A schematic map of the transit system at Nation station (source: RATP).....	154
Figure 24: Systemic Representation of the Variables (boxes) and Relationships (ellipses) of a Network Traffic Assignment Model.....	160
Figure 25: An origin - destination trip on the upper layer.....	163
Figure 26: Structure of dependencies between the model variables in CapTA.....	195

Figure 27: The sub-models and the logical flow between them.....	196
Figure 28: The structure of the CapTA model	220
Figure 29: The representation of the vehicle links of a service with reference to the station platform i	224
Figure 30: The dependencies between libraries of the software (Poulhes and Chandakas, 2011)	239
Figure 31: An example of the text file with the model parameters	240
Figure 32: The VTOC of the Capta_lib architecture (Poulhes and Chandakas, 2011)	241
Figure 33: Modal share of the motorized trips in Greater Paris by geographic relation (source: OMNIL, 2012).....	248
Figure 34: The map of the demand zones of the Paris Metropolitan Region by geographic sector....	249
Figure 35: The total capacity of the transit lines, by transport mode (central sector)	251
Figure 36: The standing comfort penalty as a function of the density of standing passengers	254
Figure 37: Convergence Indicator of the bounded model variants	255
Figure 38: Passenger flows and the ratio of flow to operative capacity for the bounded model.....	257
Figure 39: Number of Arcs in relation to flow to capacity ratio	258
Figure 40: The initial and restrained frequency at the terminal for some rail lines in the Greater Paris transit network.	259
Figure 41: The relative frequency drop due to the restrained frequency in the central sector of the Paris Metropolitan Region.....	260
Figure 42: Passenger Flows and Average Perceived Waiting Time on selected lines for the unbounded (UC) and bounded (CWCV) model.....	262
Figure 43: Average Sitting and Standing time on selected lines for the unbounded (UC) and the bounded model (CWCV).....	264
Figure 44: The RER A (red) and the rest of the transit network (grey)	271
Figure 45: The frequency at departure on the westbound RER A.....	272
Figure 46: The passenger flow throughout the restrained RER A network	274
Figure 47: Boarding and Alighting Flows on the busiest sections of the westbound service	275
Figure 48: Dwell Time in relation to the exchange volume on the rolling stock used in the RER A .	277
Figure 49: Boarding, alighting passenger flows and dwell time per vehicle of the westbound services at Etoile station	278
Figure 50: Frequency at departure for the westbound transit services at the Central Trunk.....	278
Figure 51: The boarding and alighting process on a vehicle of the TEDY service.....	279
Figure 52: The Passenger Stock at the central section	281
Figure 53: The exogenous flow versus the discharge rate and the exit time interval alighting at La Defense	282
Figure 54: Route Proportions at boarding and Probability of Immediate Boarding at Nation.....	284
Figure 55: The average generalized time of the trips ending at La Defense station	285

List of Tables

Table 1: Basic Notations of Chapter 2	48
Table 2: A theoretical comparison of transit assignment models considering total persons' capacity constraints	90
Table 3: Basic Notations of Chapter 4	100
Table 4: Operational characteristics of the service routes	113
Table 5: Operational characteristics of the M1, M13 and M14 metro lines and the RER A.....	121
Table 6: Characteristics of station platforms and the additional waiting time of the platform exit flow	155
Table 7: Model hierarchy and their characteristics	169
Table 8: Basic Notations of Chapter 8	218
Table 9: The characteristics of the Greater Paris transit system.....	250
Table 10: The main characteristics of the modelled transit network.....	252
Table 11: The graph elements of the service network.....	252
Table 12: Multipliers of physical time	254
Table 13: Average generalized time (in minutes) on the Greater Paris transit network.....	261
Table 14: The detailed stopping policy for the westbound transit services.....	273
Table 15: The characteristics of the rolling stock circulating on the RER A	277
Table 16: Boarding Flows for the unbounded and bounded model variants.....	283
Table 17: The passenger capacity of the metro	316
Table 18: The passenger capacity on the RER and commuter rail.....	317
Table 19: The passenger capacity of the light rail and bus rapid transit vehicles	318
Table 20: The bus fleet composition and its characteristics (source: STIF from Optile and RATP)..	318
Table 21: The coach fleet composition and its characteristics (source: STIF from Optile and RATP)	318
Table 22: The passenger capacity of the buses and coaches	319
Table 23: Exchange capacity on the metro.....	319
Table 24: Exchange capacity on the RER and Commuter Rail.....	320
Table 25: Exchange capacity on the light rail and bus rapid transit.....	321

Notations

UPPER LAYER

N	Set of nodes on the upper layer network $G = (N, A)$
A	Set of arcs of the service network $A \equiv A_L \cup A_S \cup A_R$
A_L	Set of line legs (upper layer)
A_S	Set of station legs (upper layer)
A_R	Set of access-egress legs (upper layer)
$[q_{os}]_{o \in O, s \in S}$	Origin-destination trip matrix
$x_{\ell(i,s)}$	Passenger flow on the (i, s) station couple on line ℓ
$g_{\ell(i,s)}$	Average generalized cost of the (i, s) station couple on line ℓ
$w_{\ell(i,s)}$	Actual waiting time for a trip from i to s along line ℓ
x_i	Outgoing flow on node i
$\hat{\phi}_a$	Fictive frequency on link a
β_a	Discontinuity attenuation factor on link a
α	Waiting weight factor
A_i^+	Set of outgoing arcs from node i
A_i^-	Set of incoming arcs at node i
d	Local strategy on node i , $d \subseteq A_i^+$ and $d \in D(i)$
ρ_a^d	Route share on arc a using strategy d
$x_{i(s)}^d$	Node i exit flow to destination s using strategy d
$\xi_{i(s)}^d$	Flow share of strategy d on node i to destination s
$\tau_{i(s)}^d$	Generalized cost from node i to destination s using strategy d
$\bar{\tau}_{i(s)}$	Minimum generalized cost from node i to destination s

LINE MODEL

ℓ	Transit line, $\ell \in L$
--------	----------------------------

z	Service route $z \in Z_\ell$
$G^{(\ell)}$	Supply graph of line ℓ , $G^{(\ell)} = (N_\ell, A^{(\ell)})$
A_I	Set of interstation track links (lower layer)
A_D	Set of sojourn track links (lower layer)
A_B	Set of boarding links (lower layer)
A_A	Set of alighting links (lower layer)
$A^{(\ell)}$	Set of links of transit line ℓ , $A^{(\ell)} \equiv A_I^{(\ell)} \cup A_D^{(\ell)} \cup A_B^{(\ell)} \cup A_A^{(\ell)}$
N_ℓ	Set of stations on line ℓ
P_z	Set of track links of route z , $P_z \subset A^{(\ell)}$
N_z	Set of stations serviced along the route z , $N_z \subset N_\ell$
y_{za}^{sr}	Passenger load in a vehicle of service z along link a with egress station s and suffering a comfort state r
φ_{za}	Service frequency of service z along link a

IN-VEHICLE COMFORT

\underline{r}	Sitting comfort state
\tilde{r}	“Folding seats” comfort state
\overline{r}	Standing comfort state
k_{za}^r	Capacity of comfort state r for a vehicle of service z along link a
χ_{za}^r	The discomfort factor related to r
p_{za}^{or}	Probability of an on-board passenger at link a to occupy comfort state r
p_{za}^{+r}	Probability of a boarding passenger at link a to occupy comfort state r
d_{za}	Density of standing passengers in a vehicle of service z along link a
$d^{a(d)}$	Density of standing passengers for the activation (deactivation) of the folding seats
g_{za}^r	Average generalized time of comfort state r for service z along link a
$\gamma_{z(i,s)}$	Auxiliary cost, if not seated at station i , to egress station s
$\tilde{g}_{z(i,s)}$	Average generalized in-vehicle time of the service leg $z(i,s)$

TRANSIT BOTTLENECK

H	Simulation reference period
k_{za}^R	Total capacity of a vehicle of service z along link a
σ_{is}	Partial stock of passengers from access station i to egress station s
n_{zi}	Stock of candidate passengers to board a vehicle of service z at platform i
π_{zi}	Probability of immediate boarding of passengers waiting at i for service z

H_{is}	Exit time interval of the passenger flow on the (i, s) station couple
$(\varphi\pi)_{is}$	Available frequency for the (i, s) station couple

RESTRAINED FREQUENCY

HW_{zi}	Average headway of two consecutive vehicle of service z at station i
b_{zi}	Passenger volume boarding a vehicle of service z at station i
e_{zi}	Passenger volume alighting a vehicle of service z at station i
T_{zi}	Dwell time of a vehicle of service z at station i
ω_{zi}	Safe separation time of a vehicle z at station i
H'_i	Temporal occupation of the track infrastructure at station i
η_i	Frequency reduction factor at station i
φ_{zi}^-	Restrained frequency of service z at station i

STATION MODEL

s	Transit station, $s \in S$
$G^{(s)}$	Supply graph of station s , $G^{(s)} = (N_s, A^{(s)})$
A_V	Set of vertical circulation links (lower layer)
A_H	Set of horizontal circulation links (lower layer)
A_P	Set of platform links (lower layer)
$A^{(\ell)}$	Set of links of transit station s , $A^{(s)} \equiv A_V^{(s)} \cup A_H^{(s)} \cup A_P^{(s)}$
N_s	Set of nodes on station s
N_P	Set of platform edges
y_a	Passenger flow at link a
ℓ_i^P	Length of a platform at $i \in N_P$
E_i	Number of exits from platform $i \in N_P$
k_i^e	Nominal capacity of an elementary exit from platform $i \in N_P$
V_c	Average walking speed on the platform
t_a^o	Free flow travel time of link a
k_a	Pedestrian capacity of link a

Abstract

In the transit network of large urban areas, it frequently occurs that the transit system is submitted to heavy congestion at the peak hours on working days, especially so at the morning peak in the central part of the urban area. Under that circumstances, not only may the passengers experience the discomfort of crowding, delay and unreliability, but also the operation of services may be disrupted by increased dwelling times, vehicle bunching and delays, leading to reductions in service frequency. Along with the congestion on the transit services, massive passenger flow influences the in-station walking time for the access-egress trips and the transfers of the passengers. Thus, traffic is inconvenienced and disrupted at both levels of mobile units, passengers and vehicles, with a straightforward effect in the quality of service offered by the transportation system.

We establish a novel framework for modelling capacity effects and develop the CapTA network model (for Capacitated Transit Assignment). It is systemic and modular and addresses in particular the following capacity phenomena,

- The in-vehicle quality of service is linked to the comfort of the passengers on-board. The occupation of heterogeneous comfort states (seats, folding seats and standing at different passenger densities) influences the perceived arduousness of the travel;
- The vehicle capacity at boarding influences the waiting time of the passengers and their distribution to the transit services;
- The track infrastructure capacity relates the dwelling time of the vehicles (and by extent the alighting and boarding flows) with the performance of the transit services and their service frequency;

- The in-station quality of service is linked to the circulation and waiting conditions of the passengers.

The capacity effects listed previously modify the local conditions of a transit trip for each individual passenger. However, these should be addressed within the transit network in order to capture their effect on the network path choices; essentially the economic trade-offs that influence the choice between different network itineraries. Their treatment in a network level assures the coherence of the path choice.

A fundamental characteristic, of the CapTA model is the bi-layer representation of the transit network. In that way, we can distinguish a demand-oriented upper layer of network flowing for the passenger, on the basis of the travel conditions (cost, operational frequency, information), and a supply-oriented lower layer of enforcing the capacity constraints, on the basis of the network flows.

While the lower layer is based on an exhaustive arc-based representation (with boarding, alighting arcs, etc.), a different representation is used on the upper layer. There, the network is described by trip segments, called network legs, based on the network's subsystems. Indeed, a trip will be composed by a number of line, station and access/egress legs. A line leg corresponds to the average conditions of an individual trip-maker for an in-vehicle trip between two stations of a transit line. Respectively, the station leg represents the in-station circulation of an individual and, depending on the phase on his trip, may be a part of the initial access to (or final egress from) the line or of a transfer between two lines. The access/egress leg describes the individual's trip from his origin to the station's entrance (or stop), and on the other end, from the station's exit to his final destination.

This network duality – with a network model for the assignment on the upper layer and the local models for enforcing capacity constraints on the lower layer – requires an intermediate level: the system sub-model. Its main purpose is to facilitate the twofold relationship between the two layers, by network element: top-down, a vector of passenger flows by leg is imputed to each sub-model; bottom-up the sub-model yields the vectors that characterize the conditions on the leg: an average generalized cost and, if applicable, the waiting time and the operating frequency. In other words, the system sub-models amount to a sophisticated cost-

flow relationship in a vector form on the upper layer of the network. Moreover, they manage the modelling process on the lower layer and call the local models.

The CapTA model considers three types of system sub-models, equal to the number of network leg types: line, station and access/egress. It addresses each network element – line, station, access/egress – as a particular subsystem on the basis of these specific models. This particular model architecture implies that the relations among lines, stations or other network elements are handled on the upper layer, since the lower layer sub-networks are not connected with each other.

A line of operations is defined as a sub-graph of the transit network in a single direction of traffic, composed of one or more transit services with different operational characteristics. The line topology of the line is useful to establish the chronological order of the traffic operations. In fact, there are five parallel and related processes of: (i) passenger alighting; (ii) passenger traffic within the vehicle; (iii) passenger waiting on the station platform and boarding a vehicle with available capacity and servicing their egress station; (iv) dwell time and track occupancy that determine the vehicle operations, hence in turn the service operations and their frequency during the period of reference, and (v) interaction with external traffic on the interstation links.

In the level of the line there are two main models: a physical model of flow loading in vehicles and service traffic – the line flow loading model – and an economic model of cost evaluation in the setting of the individual passenger that would use the line on a given leg – the line leg costing model.

- The line flow loading model is used for the estimation of the physical interactions of the passenger flow on the elements of the transit in and the effect of the capacity constraints. The flow loading algorithm proceeds in the direction of traffic along the line, by handling the track links in forward topological order. The treatment of each station involves a number of successive steps, following the chronological order of traffic operation described previously. A local model is developed for each operation that addresses the capacity phenomena.
- The line leg costing model proceeds in a reverse topological order by evaluating at each egress station the trips from the upstream access stations. The formation of the

costs of the in-vehicle trips on each service is done by backward accumulation on the basis of the physical elements provided by the local models.

At the local level, at each station the model captures the capacity phenomena related to the vehicle seat capacity, the total capacity and the interplay of passenger flows at access and egress with the dwell time and service frequency. These phenomena are addressed by line of operations on the basis of a set of local models yielding specific flows or costs. The in-vehicle comfort local model is adapted from Leurent (2012a), where the passenger flows are faced with the available seat capacity at every stage of the route; seated alighting passengers that exit at a given station yield residual capacity, while a two level competition (with a priority for on-board standing passengers) takes place and the passengers from the same competition have equiprobability of occupying a seat. Thereafter, the in-vehicle comfort is dealt at the vehicle level.

The platform waiting and boarding is addressed with the transit bottleneck model described in Leurent and Chandakas (2012b). The modelling approach is based on the explicit description of the passengers waiting to board a vehicle of an attractive service, who constitute a passenger stock by egress station. The passenger stock by service is confronted to the vehicle's available capacity at boarding, which yields the probability of immediate boarding. The average waiting time on a platform for a particular egress station is similar to that of the traffic bottleneck.

The restrained frequency model establishes an interplay between the passenger flows and the line operation at the station. Indeed, the passenger exchange flows influence the dwell time of the vehicles of the transit services at the station. Prolonged platform occupancy of the vehicles of a transit service may lead to the occupation time of the planned service at that section to exceed the reference period. Hence, the service frequency of all the services using the track infrastructure is restrained. This restrained frequency is propagated downstream.

A transit station is a bounded area used to access the vehicles of the transit services at the beginning of a passenger's trip or at any intermediate point, when transferring between services. The station model assigns the in-station passenger flows on the basis of the costs of the paths and employs local models for the evaluation of the physical interaction among the passengers and the station elements. The local models address the bottleneck created during

the nominal evacuation of a platform and the walking delays on the horizontal, vertical circulation elements and the station areas. These are related to the density of passengers and the presence of heterogeneous passenger flows.

While the conceptual model is adequately defined, the research involves a major challenge: develop a software simulator, programmed in C++, efficient enough to deal with large-scale networks. A sequencing of development phases led to the CapTA simulator for passenger flow assignment on large-scale transit networks under capacity constraints. In parallel, a number of networks were adapted or built in order to assess the behaviour of the various models developed: the transit network of the Paris Metropolitan Region; an detailed RER A network extracted from the Greater Paris network and a Nation station network for the line and station models, respectively; and a series of simple test cases for the local models.

The Paris Metropolitan Region provides an ideal application field of the capacity constrained transit assignment model. It is mainly used as a showcase of the simulation capabilities and of the finesse of the modelling approach in CapTA. The transit network involves 1 500 bus routes together with 260 trains routes that include 14 metro lines and 4 light rail lines. Traffic assignment at the morning peak hour is characterized by heavy passenger loads along the central parts of the railway lines. Increased train dwelling, due to boarding and alighting flows, and reduction in the service frequency impact the route and the line capacity. The generalized time of a transit trip is impacted mainly though its in-vehicle comfort component. Detailed results have been provided for the RER A, the busiest commuter rail line in the transit network.

Keywords:

Traffic Equilibrium; Transit Assignment; Bi-layer Assignment; Transit Capacity; Seat Capacity; Vehicle Capacity; Transit Bottleneck; Platform Stock; Station Capacity; Track Occupancy; Line Model; Station Model

Chapter 1:

General Introduction

1.1 Context and Planning Issues

Economic and political developments, accompanied by an increased environmental awareness of the population, lead us towards a call for a sustainable form of development. Assisted by high oil prices, current policies take positive action for a modal shift towards public transportation and they encourage the reduction of oil and energy consumption in the transportation sector and the decrease in urban congestion. The increase of the intervention area of local authorities to urban transportation makes them a key player for the encouragement and development of public transportation, while reorganising the public space in favour of pedestrians. These issues influence and determine the recent success of public transportation. Nonetheless, issues such as congestion appear in the horizon and undermine the success of the transit system.

Transit systems are essential for the efficient operation of large cities and usually are the backbone of their transportation network (Merlin, 1991). However, high population and employment densities in the cities produce massive passenger flows that exert pressure on the local transit systems. Mature cities, with a developed structural transit network such as Paris, London and New York, together with developing cities, where the infrastructure development lags behind economic growth, face increasing transportation demand. In these large agglomerations it becomes the norm for the transit systems to be at saturation, or on the verge of it, for sometimes many hours during the day.

Congestion – occurring when the demand for travel exceeds the transit supply for a certain period – can be observed in the largest transit systems in Europe and North America, as well

as in Asia and Latin America. Transit congestion gives rise to important costs for the society due to the additional time spent travelling in crowded conditions and the stress associated. According to a conservative estimate for the London Corporation (Oxford Economic Forecast, 2003) the economic effects of congestion only on the City of London (where public transport enjoys an 88% modal share) amount to at least 230 million pounds each year. This number alone demonstrates the importance of taking into consideration the effect of congestion in the evaluation of transportation projects.

Transport authorities and public transport operators are increasingly interested in these issues. Along with a remedy to congestion, they are interested in the efficient operation of transit systems to ensure a rational use of public money and in the adequate dimensioning of the transit systems to make them capable of providing an acceptable quality of service under an increasing patronage. The performance and the organization of the transit systems have been widely studied from scientists and engineers (Vuchic, 2005). The travel conditions – grouped under the term of quality of the service – provided on the transit system affect significantly the behaviour of the transit – and non-transit – users and their improvement guarantees the shift towards a sustainable city.

Knowing the demand and understanding the behaviour of the transit users within a city is beneficial. That facilitates the conception of the transportation projects by the transport planners. They interact with the project in two levels. On the elementary level, the transport planners are given a wide range of methodological tools – such as the ones described into the Transit Capacity and Quality of Service Manual (TRB, 2003) – for the efficient design and dimensioning of a given line. Yet, on the network level, a single project or a service change can affect other parts of the transit network, due to the indirect relations with its network elements, the network externalities. At that level, the practitioners do not dispose of widely accepted methodological tools, even though these effects are harder to take into consideration.

The complexity of the transit networks makes the use of transit models necessary. Ortuzar and Willumsen (2001) define a model as “a simplified representation of a part of the real word – the system of interest – which concentrates on certain elements considered important for its analysis from a particular point of view”. In transit assignment models various methods are used for the allocation of a given set of origin-destination trips to the transportation network (Thomas, 1991). The transit assignment models initially used tools and methods developed

for road traffic. However, transit modelling quickly gained its autonomy and became a distinct sector of transportation modelling. The particularities of the public transport systems and their intrinsic complexity, compared to road traffic, are efficiently dealt with breakthrough research, which set a coherent framework for current and future research. Nonetheless, the transit assignment models used only rarely address the effects related to the capacity constraints or consider them ex-post in the planning process. Modelling the capacity effects in transit networks requires the development of new tools and different approaches, adapted to the particularities of public transportation; their temporal and spatial availability.

The temporal availability has to do with the interaction between the passenger arrival and the discrete service availability. Indeed, a passenger who uses a transit line is subject to the line's timetable or its vehicle frequency. Therefore, the waiting time is associated with the transition from pedestrian mode to transit according to the operating characteristics, for a passenger's access and transfer in the network.

The spatial availability is quite straightforward. A transit trip does not include only the in-vehicle travelling. It also includes the journey from the trip origin (such as a residence for a home-based trip) to access the transit network at a stop or station (usually by walking). On the other side of the trip, we include the egress journey to reach the final destination. In addition to that access – egress trips, a trip may also demand a transfer between lines, thus incurring additional walking (for the line change) and waiting. Whereas a road trip can be described simply, a transit trip needs an appropriate description of the access-egress, transfer and waiting elements. The choice of a path from an origin to the destination has to abide by these particularities.

Current research trends include the development of dynamic transit assignment models for capturing the choice of departure time and the temporal variability of the demand and the supply; the modelling of the capacity effects and the quality of service of the lines and stations, such as seat occupation, for static and dynamic transit assignment models; and the emergence of microscopic dynamic models for the simulation of passenger and vehicle movement on the network with a very finite description of the space of the movement and the behaviour of the agents.

The LVMT (Laboratoire Ville, Mobilité, Transports – Laboratory of City, Mobility, Transport) has invested highly in the modelling of the quality of service and the capacity constraints in the transit network with various research, publications (Leurent and Liu 2009; Leurent and Askoura, 2010; Leurent and Benezech, 2011; Leurent et al, 2011, 2012a; Leurent, 2012a; Aguilera et al, 2012; Leurent and Chandakas, 2012; Benezech et al, 2013; Chandakas et al., 2013b) and PhD Thesis (Liu, 2009; Benezech, 2013). It affirms its position with the support of the STIF and the Research and Education Chair on “The socio-economics and modelling of urban public transport”.

1.2 Problem Statement and Research Objectives

A variety of transit assignment models have been proposed throughout the years for modelling capacity constraints – some of which are already implemented in commercial software – and this research sector seems to gain in volume. These researches adopt various approaches: some are similar, while others differ significantly. It is a research sector, which has not yet reached maturity and currently faces a set of challenges, for a number of reasons listed below:

- Research has been restrained so far on two capacity constraints: total passenger capacity and in-vehicle comfort. Nevertheless, a thorough examination of the public transportation system reveals numerous capacity constraints that are essential for accurately depicting the conditions of a transit trip; whether the line and station constraints are concerned, or the interaction between transit supply and passenger demand.
- A multitude of transit assignment models have been proposed by the scientific community and treat some specific capacity constraints: total passenger capacity of a line and in-vehicle comfort. A thorough literature review accompanied by a theoretical and numerical comparison (Leurent and Askoura, 2010) of the main models reveal completely different model behaviours and some lack of coherence of their behaviour faced to the same capacity constraints. Moreover, there is no empirical work

accompanied with the development of these models, since they have not been confronted to observations.

- The research community is currently focused on the development of efficient models for transit assignment under capacity constraints. This work often involves tampering with some fundamental elements like network representation, traffic flowing or route choice. That leads in developing models that are not compatible with each other. Some effort has been done recently for combining different models, especially those concerning total capacity and in-vehicle comfort. Nevertheless, there is a lack of a modelling framework for integrating current models, or models to be developed in the future. That integration should be made in a coherent way.

This research seeks to respond to the issues stated above. More precisely, the CapTA network model focuses on the following capacity phenomena:

- *In-vehicle quality of service:* In public transportation a trip on a vehicle may be done in different comfort states that depend on whether a passenger occupies a seat or a folding seat, or he is standing in crowded conditions or not. These conditions, linked to the time exposed to each one of them, lead to different perception of the trip that influence the passenger's route choice, especially so in a transit network with line redundancies;
- *Vehicle capacity:* When the waiting passengers on a platform are confronted to a vehicle with insufficient capacity at boarding, they cannot board with success the first arriving vehicle. Therefore, these passengers face an increased waiting time, which influences the shares of passengers on each transit service;
- *Track infrastructure capacity:* In a transit line with a segregated right-of-way, the number of vehicles that cross a given line section depends on each vehicle's temporal occupation. Therefore, in the case of the vehicle dwelling at the station, the number of alighting and boarding passengers influences the performance of the transit service;
- *In-station quality of service:* A transit station has given circulation and passenger storage capacity constraints. Thus, an increased passenger demand can degrade the walking and waiting conditions when accessing to or transferring in a station.

The capacity effects listed previously modify the local conditions of a transit trip for each individual passenger. However, these should be addressed within the transit network in order to capture their effect on the network path choices; essentially the economic trade-offs that lead to the choice between different transit lines. Their treatment in a network level assures the coherence of the path choice. The objective of this dissertation is to develop a modular framework for passenger flow assignment modelling under capacity constraints. It is based on a systemic representation of the public transportation system, where each section of a transit trip is managed through a specific submodel. That approach allows building a consistent transit assignment model, while keeping a reasonable level of complexity and computation effort.

1.3 Research Approach

Based on the previous description, this research concludes to a passenger flow assignment model, named CapTA, for Capacitated Transit Assignment. This modelling approach announces the multi-disciplinarity required by this project. A multitude of skills and theoretical tools are needed throughout this research. Hereby, we list the most important ones:

- *Operational Research*. A model, as an optimization problem, needs solid background in this domain. Especially so, when creating new models and providing the mathematical characterization of the traffic equilibrium.
- *Transport Economics*. A passenger is considered an economic agent who faces various choices. Therefore his decisions are most often economic trade-offs inspired from economic theory.
- *Probability and Statistics*. As a part of the process for the formulation of consistent models, wide knowledge in these domains is required.
- *Programming*. The development of an algorithm and the construction of a software for transit assignment modelling in C++, requires sufficient knowledge in that language and especially of structured programming.

A fundamental characteristic of the modelling approach of the CapTA model is that we distinguish the particular subsystems of the transit system. On the one hand, we address each

one of them in a specific way, with respect to their physical characteristics and the capacity constraints detected. On the other hand, we deal with them inseparably in the context of the transit network. The subsystems we refer to in the CapTA model are the following:

- *The Line Subsystem.* It constitutes the basic subsystem of a transit network. The operations of a line are autonomous, especially so for lines with exclusive right-of-way. Each direction of operations can be addressed separately and the vehicle operations of different lines rarely interact. The processes included in a transit line concern the passenger movement within a vehicle, the passenger alighting at the station and the passenger waiting on a platform and boarding a vehicle. In addition, the vehicle operations are limited to the dwelling time, linked to the vehicle traffic and the interaction of the vehicles with the external traffic.
- *The Station Subsystem.* It is an essential component of the transit system and fulfils two primary functionalities. First, it acts as an interface between the city and the transit services, providing access to and egress from the transit network. Second, when two or more lines serve a transit station, a transfer of passengers between them is available, leading to an increase in the accessibility of the transit system. Thus, the circulation and waiting conditions of the passengers in the transit station are specifically addressed.
- *The Access-Egress Subsystem.* This subsystem covers the extremities of a transit trip, namely the part of the passenger trip where he uses a private mode (such as walking, car, or bike). It concerns the access trip from the origin to the first station and the egress trip from the final station to the destination of his trip. The conditions of these trips vary, with respect to the mode used, the quality of the respective network (pedestrian, road and bicycle networks) and the effects linked to the interface of the private modes to the transit station.

The research approach for the elaboration of a traffic network assignment model is based on specific stages. The development of the CapTA network model follows the stages listed hereby:

- *Systemic Analysis of the Transit System.* A transit system is composed of various components which can be grouped in subsystems and interact with each other. A

systemic analysis is therefore necessary in order to define these subsystems and identify the relations between them and the systems' components.

- *Physical and Economic Modelling of the System's Components and Relations.* A transit system is not abstract, but it is rather composed of physical entities, such as passengers, vehicles and platforms. Therefore, the modelling process focuses on the detailed representation of the physical interactions among these entities. Furthermore, these physical interactions yield certain travel conditions which are on the basis of the economic trade-offs modelled in a network level.
- *Algorithmic and Computational Treatment of the Model.* The mathematical formulation of the models is followed by the elaboration of efficient algorithms. These are included in a coherent modelling framework at the network level. At this stage, various mathematical tools are used, in particular probabilities and graph theory. Furthermore, a software simulator is developed, programmed in C++, which is efficient enough to deal with large-scale transit networks.

The modelling process presented previously implies a certain number of choices, related to the elaboration of the CapTA network model. The priority of the modelling approach is given in the exhaustive representation of the public transportation system, without restraining the system representation. Based on operational research methods, we privilege the systemic representation of the transit system, in order to assure the coherence between its components and subsystems, and we address in detail the physical and economic aspects of the component's relations. To these priorities, the mathematic characterization of the traffic equilibrium and the elaboration of a software simulator are added. Therefore, the modelling approach lacks some important steps, namely, the confrontation of the model assumptions with empirical data and the validation of the network model. Although that should not be seen as a weakness of this particular model, compared to others, that step should be considered in the future for validating the particular assumptions adopted.

That approach allows developing a passenger flow assignment model, which is consistent with the phenomena, observed. The principles of the CapTA model are outlined:

- A modular framework, where some parts can be replaced by more appropriate models or new models can be added;

- A bi-layer network representation along with three levels of modelling from network to local scale. Each level handles specific aspects of the transit system;
- A network assignment model, yielding coherence among the particular subsystems and addressing the network externalities in particular. In order to handle large-scale networks, computational efficiency is required;
- Particular local models, based on the physical flow of passengers and vehicles, handle the effects of the capacity constraints on the travel conditions.

The transit assignment model developed in this research is based on a stationary user equilibrium. While a dynamic transit assignment model (DTA) is the appropriate model to apprehend the demand and supply temporal variability within the peak hour, this research does not follow that direction for a number of reasons. First, the current dynamic transit assignment models require an important computational effort and, thus, they are generally limited to small or medium-scale transit networks. That approach does not allow yet simulating large-scale networks, one of the main objectives our work. Second, as numerous innovations are included into the model, it seems fundamental to concentrate our effort on understanding the behaviour of the models on the more comprehensible static context. Thereby, we can apprehend the primary and secondary effects of these capacity constraints and their relations, while acknowledging the limitations of the static assignment. Future research can be focused on adding the temporal variation of the transit supply and demand under such models for accounting for capacity constraints on the transit system.

1.4 Original Contributions

The fundamental contribution of this study to transportation planning is the development of a coherent modelling framework for passenger flow assignment under capacity constraints. Three subsystems are identified in the transit network and are treated separately at the local level with respect to their intrinsic characteristics. Afterwards, the local conditions are combined in the network layer.

The second contribution involves the development of local models for treating the capacity constraints. Whilst conditioned for the CapTA model, they can easily stand alone on other transit assignment models. Four local models are developed or adapted:

- The restrained frequency model for integrating the line capacity effects on vehicle flowing. Considering the station platform as a scarce resource, the flow dependent dwell time of a vehicle influences the line operation. Thereupon, an interplay between the passenger demand and the supply of services is established;
- A revised in-vehicle comfort model based on vehicle seat capacity (Leurent, 2012a), where – in addition to seat occupation – the in-vehicle travel conditions of the standees depend on their density. An additional algorithm is provided where the use of folding seats is included;
- Local models for treating saturation and capacity constraints on the station level: a systemic approach leads to the development of the platform exit flowing model and the pedestrian circulation model on horizontal and vertical elements;
- The transit bottleneck model for platform waiting and boarding. When the available capacity is not sufficient, faced to the demand, a stock of passengers is formed at the platform. That has an influence on the local route choice and the platform waiting per egress station. The mathematical formulation proposed in Leurent and Chandakas (2012b) is further developed in the CapTA model and added to the simulation software.

A series of important contributions are made on the operational level. Data-wise, this research sought to collecting and enriching data about the transit network of the Paris Metropolitan Region, such as vehicle capacity and characteristics and vehicles' assignment to routes, essential for passing to the capacity-sensible era of transit assignment modelling. In addition, some simple networks, extracted or built, can be the basis for testing alternative models and comparing their behaviour.

This dissertation is part of the research project that concluded in the development of the CapTA simulator. The model development and more precisely the model conception and the software programming were made with the close collaboration of Fabien Leurent and Alexis

Poulhès (Poulhes and Chandakas, 2011). That resulted in a number of scientific papers and presentations, on the basis of this model (Leurent et al 2011, 2012; Leurent and Chandakas, 2012a, 2012b; Chandakas et al, 2013, 2014; Chandakas and Leurent, 2013)

This research provides a state-of-the art simulator for transit assignment under capacity constraints, such as in-vehicle comfort, vehicle's capacity and station track occupancy – and the possibility of activating all of them, or some of them in particular for a simulation. It can contribute to a more comprehensive socio-economic evaluation of investment projects, especially the ones directed to alleviate passenger congestion, such as the replacement of the rolling stock, without planned service changes.

1.5 Outline of the Dissertation

This study consists of 11 chapters grouped into 4 parts, which present respectively, bibliographic reviews, restrained and network models, the CapTA simulator and some applications in the Paris Metropolitan Region.

Part I: Literature Review on Transit Assignment Modelling. This study begins with a wide review of transit assignment modelling from the initial models to the more advanced contemporary models; static and dynamic. A meticulous study of the transit assignment models under capacity constraints is realized.

Chapter 2, *A Review of the Passenger Flow assignment Models for Transit Network*, reviews the five generations of transit assignment models, from their emergence to the advanced static and dynamic models currently under research. The first generation includes the early attempts to model transit assignment and the identification of the particularities of public transportation. These are adequately treated by the second generation, who established an appropriate methodology for the modelling process. The three final generations are developed in parallel and respectively deal with: static transit assignment with capacity constraints; macroscopic dynamic models; and microscopic dynamic models that address the passenger behaviour in limited spaces and the interaction with other individuals as well as simulation-based models for traffic flow assignment.

Chapter 3, *Modelling Capacity Effects in Transit Assignment: a State of the Art*, is aimed at investigating the capacity constraints in public transportation and the research models developed so far. We can identify seven capacity effects, which emerge from the interaction between the four components of the transit system: line, vehicle, passenger and station. Nevertheless, only two of these effects are included in transit assignment models: total vehicle capacity and in-vehicle quality of service. A description and a comparative analysis is provided for the main groups of capacity-constraint assignment models: Effective Frequencies, Fail-to-Board, User Preference Set and Dual Penalty for passenger total capacity; and Seat Capacity, Failure-to-Sit and Seat Availability for in-vehicle quality of service.

Part II: Restrained Models. The second part examines the development and evaluation of the system sub-models, the line and station models.

Chapter 4, *Modelling the Line System on a Structural Transit Network* is aimed at presenting the line system, its modelling framework and the behaviour of the local models. The chapter begins with a description of the line model and the characteristics of the line system. These provide the outlines of the line model, developed on the basis of the topological order of the network and the chronological order of the processes at a station. The second part of the chapter is dedicated to the description of the local models for enforcing the capacity constraints: in-vehicle comfort, transit bottleneck and restrained frequency. Simple test cases are provided for evaluating the behaviour of each local model.

Chapter 5, *A Transit Station Model on a Structural Transit Network* is focused in the development of a system sub-model for the transit station. The chapter begins with a systemic analysis of the transit station, the thorough description of its components, the relations among them and the role of passenger information, as means to discover the appropriate modelling approach. Then, the station model is defined; the lower layer representation and the local models are formulated and explained. The final part involves an application of the station model in a busy station of the Paris Metropolitan Region, the Nation station. The limitations of the model are outlined, along with the effects of congestion on the in-station transit paths.

Part III: The network model – the CapTA simulator. The third part is aimed at presenting the network model for the transit assignment. In addition, the mathematic formulation of the CapTA model is addressed, along with a detailed description of the solution algorithm and some computational aspects of the model.

Chapter 6, A Systemic Representation of the CapTA Model, makes an overview of the network model and its components. It examines each modelling component separately and describes their characteristics and the relationship with each other. Furthermore, key components of the CapTA model are thoroughly explained: the bi-layer representation of the transit network and the structuring of the models in three hierarchical levels.

Chapter 7, The Mathematic Characterization of the CapTA model, is aimed at the mathematic characterization of the network equilibrium. The chapter starts with a through review of the optimization models and the characterization of the traffic equilibrium for various existing transit assignment models. The following section details the dependencies among the variables manipulated in the model and provides a logical structure of these dependencies. A thorough mathematic treatment of the CapTA model is pursued. There, the mathematical properties of the leg cost-flow relationship are demonstrated. The traffic equilibrium is expressed as a twofold vector of arc flows and strategy shares at the network layer, which is characterized as a variational inequality problem.

Chapter 8, Algorithmic Implementation and Computational Aspects of the CapTA Model, presents the algorithms of the CapTA simulator, as programmed in the simulator. The structure and the complexity of these algorithms are discussed. The algorithms concern the models presented in the previous chapters: the network model with the algorithms for the optimal strategy and the flow assignment; the line model with the ZIP and UNZIP algorithms; and the local models, i.e. the in-vehicle comfort, the transit bottleneck and the restrained frequency. The chapter is completed with a description of the computational aspects of the simulator, its structure and a brief description of its components.

Part IV: Application of the CapTA simulator. The final part of this dissertation is aimed at presenting the results of the simulation at various networks. These applications provide the indications that validate its behaviour on such network.

Chapter 9, *A Simulation of the Transit Network of the Paris Metropolitan Region* presents the results of the application of the CapTA model on a large-scale network. First, the transit demand and supply are described, along with the transformed calculation network. Four model variants – in relation with the capacity constraints activated – are tested and their behaviour is compared on the basis of various indicators. The main model variant contains the transit bottleneck, the restrained frequency and the in-vehicle comfort, where the comfort of the standees is linked to their density. This variant is compared to the unbounded model on the basis of some macro-indicators: the operating frequency, the waiting time and the average sitting and standing time for a number of representative transit lines.

Chapter 10, *The Results of the CapTA Model on a Restrained Network* provides an application for the model at the line level, which serves both as a showcase of the capabilities of the model and of the finesse of the modelling approach and as a means to further investigate its behaviour. The RER A – the busiest commuter rail lines in the Paris Metropolitan Area transit network – is chosen as an application instance for the model. Various results are presented in relation to the local models: the dwell time of the vehicles and its relation with the operating frequency; the formation and behaviour of the passenger stock; and the relative effects of the generalized time components. Finally, the effect of the replacement of the one level rolling stock with duplex ones is provided, as an input to a cost-benefit evaluation.

Chapter 11, *General Conclusion* provides the final remarks of the dissertation. First, it summarizes the main results of this research: the development of a modular framework together with a bi-layer representation of the transit network; the flexibility given by this architecture allows addressing the capacity phenomena. Second, it sketches out some future work. Besides issues relevant to the mathematical characterization of the equilibrium, a validation of the modelling assumptions is required. Finally, the upgradability of the CapTA model allows the enrichment of its capabilities and the improvement of the models included.

PART I: A LITERATURE REVIEW ON TRANSIT ASSIGNMENT MODELLING

Chapter 2:

A Review of the Passenger Flow Assignment Models for Transit Networks

2.1 Introduction

The contemporary transit assignment models present in the literature have acquired a high level of complexity and account for a variety of phenomena. In addition to the transit-related phenomena, such as waiting, there are efforts to address phenomena, such as demand and supply uncertainty, temporal variability, route choice behaviour, vehicle and line capacity constraints and crowding. A consistent way to understand current modelling approaches is to follow the evolution of this sector from its initial steps with simple road traffic-based models to the complex and consistent actual models. In 60 years of developing many solutions have been attempted to fulfil the modelling needs.

The objective of the chapter is to make a historical review of the transit assignment models. For that purpose, the models are grouped into five generations with similar characteristics and techniques. The first two generations underline a continuous improvement of the models, while the other three generations branch out and correspond to the three modelling approaches developed in parallel from the 1990's. The chapter is composed of five sections, one for each generation. We conclude with an overview of the evolution of the models and the main points discussed in them.

The first generation of transit models (section 2.2) includes the early attempts to model transit assignment by adapting the existing road traffic assignment models (Beckmann, 1956). The researchers immediately identified the particularities of public transportation: discrete service

availability, transfers and overlapping lines. The second generation of models (section 2.3) offered solutions and some fundamental approaches for dealing with overlapping lines, passengers' behaviour and the network description along with a framework, capable of modelling large-scale networks. Some concepts, such as the optimal strategies and the hyperpaths are still widely used. While each generation marked significant improvements, in relation to the previous ones, the generations 3, 4 and 5 are developed simultaneously and each one deals with a different object of transit modelling. The third generation (section 2.4) of transit models deals with various capacity effects, such as total passenger capacity and seating capacity, for large-scale networks in a macroscopic static state environment. The fourth generation (section 2.5) covers the development of macroscopic dynamic models. The researchers' interest lies in the first place in defining the appropriate representation for demand and transit services and feasible solution algorithms to capture the dynamic effects of transit assignment. Some dynamic models also consider simple capacity constraints. Finally, the fifth generation (section 2.6) assembles micro-dynamic models, simulating the passenger behaviour in limited spaces with the interaction from the other individuals and other sources.

Table 1: Basic Notations of Chapter 2

Variable	Definition
w_i	Waiting Time for line i
φ_i	Frequency at node i
t_o	Time a user has already waited
$RW(t_o)$	Remaining waiting time for a line
$\hat{h}_a(A)$	Route probabilities for a link a on strategy A
x_a	Passenger flow of link a

2.2 First Generation: the Initial Steps

The first generation of transit assignment models were an attempt to adapt the assignment models developed previously for road traffic (Beckmann, 1955) to the context of public transportation. The particularities of public transportation were first identified and then some simple algorithms (Dial, 1967, Le Clercq, 1972, Last and Leak, 1976) were proposed to address them. Its main contribution is the recognition of the complexity of transit assignment

modelling, compared to the route and the formulation of simple models for the modelling of transfers between services and for the assignment of passenger flows into multiple services for the same route sections.

2.2.1 Modelling Overlapping Lines and Transfers

The main difference between public transportation and road transport is the discrete availability of the former. In other words, a transit user is subject to waiting before boarding a vehicle. This probabilistic waiting time should be included to the total travel time. The Transept model (Last and Leak, 1976) introduces the waiting time in the calculation of the generalized cost of a trip. Waiting time for accessing to a line is also introduced in Le Clercq (1972), where the author assumes that the waiting time, w_i , of a service or a group of services equals half the inverse of the frequency, φ_i :

$$w_i = 0,5 / \varphi_i \quad (1)$$

In addition to the waiting time, Last and Leak (1976) replace the actual travel time with the generalized cost of a path. That cost is deduced by assigning different weights to the components of the journey and adding specific penalties, according to the preferences of the transit users.

Attention was also given in addressing the sections where overlapping lines serve the same stops. The authors propose different solutions, mainly based on an aggregation of lines for each section. In Dial (1967) the characteristics (in-vehicle travel time and frequency) of the lines with common sections are expressed by a single section, the trunk-line-link. Le Clercq (1972) used a different method without aggregation. Finally, Last and Leak (1976) use the section concept with an iterative process in order to assign the trips to the common lines, considering capacity and waiting times, through seat competition and a dynamic relationship to obtain a stable solution.

One of the most thoroughly discussed subjects in these papers is the effect of the line transfers on the path choice and flow assignment. To capture the effect, a widely used approach is to introduce a transfer penalty, equal to the waiting time to which passengers are exposed beforehand. This is the approach adopted by Dial (1967) in the Transit Pathfinder Algorithm. Le Clercq (1972) suggests another approach, where transfers are modelled as left turn

penalties; a concept imported from road traffic assignment models. Last and Leak (1976) opt for another methodology, clearly privileging direct routes over transfers. By the generalized cost function, they include a constant penalty for every transfer made. Furthermore, they apply a simple rule according to which the transfer would take place only in the end of the section with the common lines, for facilitating the flow assignment.

Contrary to the flow assignment in road traffic, transfers increase the complexity of the shortest path algorithms. Indeed, in order to reach a node upstream, it is essential to know the history of the path. Dial (1967) use an adapted Moore algorithm to find the shortest path. He introduces a backward loop, which accounts correctly for the effect of the transfer, but it is expensive in computation power. Le Clercq (1972) differs in two ways: first the algorithm minimizes the travel time and second it uses a 'once-through' forward algorithm, which calculates the cost of the next points to be reached at each advancement step. These algorithms have essential differences: In Dial (1967) a label correcting step is included when a shortest path is added, while in Le Clercq (1972) a shortest path tree is added at each node.

Accordingly, after the route choice, the flow is assigned to the selected path. The solution algorithms proposed by Dial (1967) and Le Clercq (1972) minimize the actual travel time for each origin - destination. On the other hand, Last and Leak (1976) use a trip loading process, by performing a link-by-link loading for each route in a topological order which is appropriate only for radial service. It should be noted that all these models provided a heuristic solution for the path assignment, without an optimization process.

2.2.2 The Transept Model

We hereby present the main lines of the Transept model, which is the first complete transit assignment model. Transept, developed in Last and Leak (1976), is a bus model capable of evaluating and testing different bus networks, and serving both the transport planner and the bus operator. We identify four main innovations of the model. First, it takes into account the preferences of the passengers by introducing a generalized cost function and a leg-based network description. The generalized cost becomes an essential criterion for the route choice. Secondly, in addition to the shortest travel paths, it includes a modal split module, which is based on the costs of the alternative modes, such as walking and car, in order to adjust the demand for bus services in relation to the alternatives. The authors also develop a method to

assign users of each transit path (a sequence of links from origin to destination with one or multiple routes) into individual routes. Finally, the model offers an evaluation of the costs and benefits for each route and a module to optimize bus frequencies.

Let us now review the main components of Transept: the network representation, the route choice, the flow assignment and the post-modelling reporting. The network is composed of three hierarchical levels. First, the bus stops: the points on the space for accessing to the public transportation network. Second, the transit links: sections of a path, containing a sequence of stops served by one or more transit routes. Third, the transit routes: they serve a sequence of stops and their characteristics depend on the bus type, the stopover time and the number of vehicles allocated (for the estimation of the headways).

The route choice relies in selecting the path – composed by a series of transit links – of minimum cost for a given origin – destination pair, while privileging the paths including direct routes over transfers. That is done by a modified version of the Floyd's algorithm. Alternative paths, requiring at least one transfer, are also considered, for comparison with direct paths. Mode choice is done by a multinomial logit model, which also considers walking, alternative bus paths and car mode as the alternative modes. The mode choice is sensible to car ownership and produces separate origin – destination matrices by mode: car, walk and bus.

In Transept, a passenger is assigned to specific routes from the various paths, on the basis of their physical and economic behaviour. Designed for modelling radial networks, the passenger flows are assigned to each link, while keeping the upstream priority of passengers concerning vehicle and seat capacity. Once the passengers are assigned to transit legs, they are dispatched into the transit routes by a two-fold competition: between routes for passengers on the one hand and among passengers for seats and capacity on the other. The competition for seats and capacity is handled by an iterative loop for every link, involving a probability of overloading. The waiting time is recalculated on the basis of the updated conditions.

The structure and implementation of Transept makes it useful for both transport planners and bus operators. For the former, the model assesses the total user benefit in generalized cost and its distribution within the study area. By using the modal split model, Transept can evaluate the economic benefits from modal shift, through the estimation of the consumers' surplus. For

the latter, the model provides detailed predictions of loading for each route to the operator, for adapting accordingly the capacity, and a financial analysis for each route, by comparing fares and revenues with the route costs. It can also provide an optimization of the level of service by adapting the frequency of the bus routes to the passenger demand, subject to a fixed number of buses.

With Transept, Last and Leak (1976) mark a major breakthrough. In terms of transit engineering, a complete transit assignment model can be used in operational level, adapted to the computational power available. Its main advantages remain in the modelling of the economic behaviour of an agent and in its structure, which provides useful output for the bus operator and the transport planner of a given public transport network.

2.3 Second Generation: Setting the Foundations

The second generation of transit assignment models makes an additional step towards a more realistic representation of public transportation. This is achieved by representing waiting time as a stochastic phenomenon related to the passenger and vehicle arrival at the station. In addition, these models laid the necessary framework, on which more complex contemporary models have been developed.

2.3.1 *The Local Route Choice Problem*

The local route choice problem – approached as the combined problem of selecting the attractive lines between two stops nodes and assigning the section's transit flow – is the focus of an extensive literature. Andreasson (1976) and Hasselström (1981) proposed a heuristic procedure to distribute the passengers to overlapping lines. Chriqui and Robillard (1975) are the first who clearly defined explicitly the common lines problem. The objective of their model is to define the attractive set of routes between two consecutive nodes. The attractive set minimizes the user's expected travel time – the sum of expected waiting time and in-vehicle travel time.

Chriqui and Robillard (1975) calculate the expected waiting time for each line on the basis of the distributions of the passenger and vehicle arrivals at the station. They assume that a transit user takes the first approaching vehicle from a subset of attractive lines. They also suppose

that the arrivals of the vehicles of the different transit lines are independent. Based on these assumptions, they formulate a generic mathematical function for the expected total travel time. While they propose a Boolean hyperbolic program (Robillard, 1971) that gives the optimal solution, they design a heuristic solution algorithm, applied for the uniform and exponential distribution of vehicle arrivals.

Marguier and Ceder (1984) add a sophistication level by taking into account the elapsed waiting time on the local route choice. Indeed, in general, the expected waiting time for a vehicle of a given line depends on the time during which the passenger has already waited. Another virtue of Marguier and Ceder (1984) is to consider explicitly the relations between three types of distributions; the bus regularity, through the headway distribution, the bus arrival variability between days and the passengers' arrival distribution.

Marguier and Ceder (1984) derive analytical function for the interdependence of the arrivals of the transit services and the passengers. The analytical solution developed concerns the simple case of two alternative lines, with either deterministic or exponential interarrival distribution of the services. Nevertheless, the mathematical form of the local route problem is complex and an analytical solution is practically impossible for three overlapping routes or more. Therefore they are unable to propose a solution algorithm for a general case.

2.3.2 Modelling at the Network Scale

Once the local assignment problem is treated, it is essential to place it in the frame of an entire network. On that issue, the first models were based on an all-or-nothing assignment procedure, directly inspired from road traffic assignment, with some adaptations to include the temporal availability of transit services. At that phase, the scientific community was focused on the mathematical expression of the preferences and decisions of travellers regarding entire trips (i.e. possibly involving multiple transfers). Various approaches emerged. The most popular is the optimal strategy concept, introduced in Spiess (1984) and formally presented by Spiess and Florian (1989). That concept is also adopted by De Cea and Fernandez (1989), with a different network description.

Spiess and Florian (1989) first observed that the shortest path algorithms are not optimal in regards to the minimization of the expected travel time. Based on this assumption, and the random nature of the waiting phenomenon, they deduce an alternative definition of local route

choice strategies. The authors define a strategy as a set of rules that, when applied, allow the transit users to reach their destination. The type of strategy that travellers can use depends on the information that is available to them during the trip. Among the feasible strategies exists an optimal strategy, which minimizes the expected total travel time.

The transit network is modelled as a graph with arcs and nodes. The transit routes are represented as a set of arcs between the stops. On that basis, a strategy to reach destination s is associated with a partial network $G_s = (N, \bar{A})$, where $\bar{A} \subseteq A$ is the subset of links that will be used.

Spiess and Florian (1989) formulate a wardropian transit equilibrium as an extremal optimization problem with a nonlinear objective function. They demonstrate that it can be reduced to an equivalent linear programming problem and they prove that there is an optimal solution (see chapter 7). The mathematical treatment allows defining at each stop the probability that a link a is served first among all outgoing links of a node, \bar{A}_i^+ (the forward star of node i on strategy \bar{A}). At the local level, the route proportions designate that probability or, alternatively, the share of the passenger flow for each outgoing line. At each stop, the formulations of the expected combined waiting time, $w_i(A)$, and the link probabilities, $\hat{h}_a(A)$, are:

$$w_i(\bar{A}_i^+) = \alpha / \sum_{a \in \bar{A}_i^+} \varphi_a, \alpha > 0 \quad (2)$$

$$\hat{h}_a(\bar{A}_i^+) = \varphi_a / \sum_{a \in \bar{A}_i^+} \varphi_a, a \in \bar{A}_i^+ \quad (3)$$

Where, α is a positive constant related to the arrival distribution of the vehicles and the passengers and the perception of the waiting time as opposed to the in-vehicle travel time (reference time).

In addition, the authors develop a solution algorithm with a sequential treatment per destination, whose objective is to reduce memory needs and increase calculation speed. The optimal strategy algorithm starts from the destination and examines all the arcs upwards, by order of increasing cost to destination order. The attractive set consists then of an ordered list of the arcs by ascending order to destination which yields the minimum cost from each node to the destination. The flow assignment algorithm uses that ordered list to assign the

passenger flow according to route frequencies. Spiess and Florian (1989) prove that the solution given by the algorithm is the optimum of the dual linear programming problem and they evaluate the complexity of the algorithm. If $|A|$ is the cardinal of the set A of the links of the strategy, the complexity of the algorithm for $|S|$ destinations would be $O(|S||A|\log|A|)$.

Spiess and Florian (1989) lay the foundations of modern transit assignment. They adopt an exhaustive mathematical treatment of the optimizing problem, making the connection between qualitative description of an economic behaviour and its quantitative formalization. In addition they prove the optimality of the solution algorithm and calculated the complexity of the entire algorithm.

Nguyen and Pallottino (1988) propose an alternative formulation of the optimal strategies, based on a graph theoretical approach. The hyperpath is formally defined as an acyclic graph, which contains a set of nodes and directed arcs related with a strategy. It is composed by elementary paths, r , each of them characterised by a path probability. They assume that a traveller chooses the attractive set of lines at each station before his departure.

Nguyen and Pallottino (1988) assume a standard additive cost structure, where the cost of a path is simply the sum of the costs of its arcs and nodes:

$$g_r = \sum_{a \in r} g_a + \sum_{i \in r} w_i^r, \quad \forall r \in R \quad (4)$$

Considering the hyperpath h as a set of paths R where each path has \hat{h}_r routing proportions, its cost is then:

$$g_R = \sum_{r \in R} \hat{h}_r g_r \quad (5)$$

The cost structure adopted above by Nguyen and Pallottino (1988) along with the hyperpath property, as a directed acyclic graph, allows the cost of a hyperpath to be computed by a recursive formula from the destination in a reverse topological order. The shortest hyperpath is an application of the extended Bellman's equations. From a given origin node i , to a given destination s , the length τ_i of the shortest hyperpath satisfies:

$$\tau_i = \begin{cases} 0 & , \text{ if } i = s \\ \min_{j \in A_i^+} \{ \tau_j + c_{ij} \} & , \text{ if } i \notin R \\ \min_{\bar{A}_i^+ \subseteq A_i^+} \left\{ \sum_{j \in \bar{A}_i^+} \varphi_{ij} \tau_j + 1 \right\} / \sum_{j \in \bar{A}_i^+} \varphi_{ij} & , \text{ if } i \in R \end{cases} \quad (6)$$

If the combinational problem of determining the optimal subset of boarding arcs at stop node i can be solved efficiently, the algorithm associated with the generalized Bellman's equation produces a shortest hyperpath tree (Gallo et al, 1993).

Nguyen and Pallottino (1988) propose an equilibrium assignment model for the hyperpath framework without congestion. The equilibrium model relies on the assumption that all transit users travel on shortest hyperpaths. Thus, by applying the conservation of flow equations and the wardropian condition for equilibrium (Wardrop, 1952), the optimal flow can be estimated as a solution of a variational inequality problem, in which the cost of a hyperpath and its flow are associated. If the cost is a positive continuous strictly monotone function, then the existence of an equilibrium hyperpath flow is ensured, though it is not necessarily unique.

The hyperpath framework seems appropriate for the simulation of the particularities of public transport such as waiting times and transfers. The algorithm of Nguyen and Pallottino (1988) is flexible and it allows for a diversity of approaches regarding cost formation and equilibrium, provided that the hyperpath is characterized by an additive cost structure; some occasional limitations of the framework exist.

2.3.3 Capacity Constraints and Congestion

The models reviewed until now are focused on waiting times, transfers and passenger strategies. They let aside the capacity constraints of public transportation networks and their consequence: congestion. This topic is not entirely ignored. Early papers addressed it, such as Last and Leak (1976) who implements a local equilibrium for each route section. But their solution includes various restrictions and do not apply to non-radial networks and other types of transport modes.

Spiess and Florian (1989) propose a non-linear version of the optimal strategy model in order to model congestion. They make a parallel with road congestion assuming that the in-vehicle

travel time (or generalized cost) increases with the traffic flow of the link. This model is more appropriate regarding in-vehicle discomfort, since it does not include any increase in waiting times. Previously, Gendreau (1984) took another path by addressing congestion effects on both the in-vehicle travel time and waiting time. With that objective, Gendreau (1984) develops a theoretical model based on queuing theory where both waiting times and the distribution probabilities of passengers among attractive lines depend on the traffic flow.

London Underground (LUL) used a transit assignment model with capacity constraints. According to Harris (1989) the model includes both the effect of congestion in the in-vehicle comfort and in waiting time. Namely, it applies a penalty factor in the generalized cost calculation, dependent on the traffic flow. A level of service calculation can be also applied, via the same mechanism.

Finally, De Cea and Fernandez (1993) present the effective frequencies model to deal with congestion in public transportation vehicles. They use a route choice mechanism derived from the optimal strategies of Spiess and Florian (1989). They proposed that waiting time is specified by a BPR function of the boarding and in-vehicle flows. The effective frequencies at a given station express the local frequency (and the associated waiting time) perceived by the boarding passengers. The solution of that asymmetrical problem is obtained by a diagonalization method, where an equivalent convex cost optimization problem is solved through a Frank-Wolfe algorithm.

One of the objectives of the authors was to be able to address the common lines problem and to obtain the attractive line set. The congestion can be included in the generalized cost calculation, with an additional variable linked to the effective frequencies. Nevertheless, in the model of De Cea and Fernandez (1993) the boarding flow is not restricted and it is possible to locally exceed the service capacity.

2.4 Third Generation: Integrating Capacity in Macroscopic Static Assignment Models

The third generation includes transit assignment models, which deal explicitly with capacity effects and congestion in a macroscopic static regime. The impact of capacity constraints on

the performance of the public transportation networks is complex. Leurent (2011) identifies seven types of capacity constraints in public transport, involving the line infrastructure, the stations, the vehicles, the passengers and the interface among them. Two of these effects are addressed in great extent by current transit assignment models: vehicle total capacity and seat capacity.

2.4.1 Vehicle Passenger Capacity

The main path for modelling the impact of total vehicle capacity involved the effective frequencies concept introduced in De Cea and Fernandez (1993). Various approaches followed, such as Wu et al (1994), for relating the route frequency (and therefore the waiting time and routing proportions) with the passenger flow vector. Cominetti and Correa (2001) treat the local route choice under congestion, and offer a formalization of the model and a proof of existence of the network equilibrium. In addition, through the fixed point problem approach in the space of arcs, they provide an extension of the model on large scale networks. The model can easily incorporate the effect of congestion, since it keeps the standard assignment framework. The work of Cepeda et al (2006) takes on the same concept and provides a feasible solution algorithm and a proof of convergence. However, it is based on a BPR type function with no apparent justification.

Lam et al (1999) represents vehicle congestion by applying a penalty to route sections between two stations, where the volume exceeds the section's capacity. A dual variable is linked to each constraint: zero if the constraint is not binding; positive otherwise. In the latter case, it represents an additional time per passenger. This model allows the explicit representation of the congestion costs and it incorporates them in the generalized cost function. Nevertheless, by applying that to all passengers without considering priority rules, it is inconsistent with the way public transportation really functions: when a vehicle is saturated, the waiting passengers at boarding are those who perceive a penalty.

Kurauchi et al (2003) considers strict capacity constraints at each station: the flow waiting to board is confronted to the available capacity. That ratio refers to the probability that a passenger will be able to board in the next available vehicle. The complementary of the probability, p_F , is the failure to board probability. If there is no waiting discipline, such as FIFO, all passengers have an equal probability to board in a given vehicle. That regime is

referred to as the “mingled passenger” hereafter. In that case they have equal probability to board, $1 - p_F$. Passengers who cannot board are diverted towards escape arcs, directly to the destination. This has as consequence to reduce artificially the transit volume on the network and to neglect the effect of congestion in the generalized cost function. Shimamoto et al (2005) adapted this model to include that congestion effect: The waiting time became function of the failure-to-board probability. Nonetheless, Shimamoto et al (2005) disregard the interaction among the common lines, and its influence on the waiting times; particularly, through the changes of the sets of the attractive lines.

Hamdouch, Marcotte and Nguyen (2004) follow a different direction to deal with capacity constraints. They propose a model based on an alternative definition of the strategy of a passenger. According to their definition, on the basis of an ordered set of nodes, the passenger flow is assigned: in priority to the attractive lines and then, when saturation occurs, to the following lines on the list. The expected costs of the strategy are then derived from the non-congested costs of the paths, weighted with their probabilities. The main shortcoming of their approach is the absence of route cost variability related to the passenger flows. However, a solution to that problem is brought by Hamdouch and Lawphongpanich (2008) with the dynamic version of the model.

More recently, Teklu (2007) presented an alternative approach to model the passenger’s total capacity of a vehicle. The main feature of the model is the Monte Carlo simulation for the assignment of the passengers to transit routes, along with the probit model for local route choice. The network representation is based on route sections, adapted to specific particularities of the simulation. Significant differences from the previous models can be observed. The most important one is the extraction of a “democratic cost” as the one that is measured by ghost passengers: they circulate freely, do not take any space and do not contribute to congestion. Furthermore, the variance in passengers’ perceived cost is traced by an error term for each route section. A learning process updates the costs on the basis of a convex combination of the previous and the auxiliary state. Finally, strict capacity constraints are enforced. However they are adapted to the stochastic assignment of the vehicle loads. The interaction between the stochastic demand and supply is modelled by a Monte Carlo simulation.

2.4.2 Seating Capacity

In addition to the models treating total capacity, different approaches were developed to tackle the seating capacity of public transport, a topic with an increasing interest, since it is a main component of the quality of service (LOS) of a route and is proven to affect the passenger behaviour.

The first approach to model explicitly the seat allocation is described in Leurent (2006) and Leurent (2012a). The author introduces a model with seating capacity and the distinction between standing and sitting comfort state of the passenger trips. The model features priority rules for the allocation of passengers to seats. Passengers aboard have the priority over the boarding passengers. Moreover, passengers with the same level of priority have the same probability to find a seat, if there are not enough seats to accommodate them at all. Obviously, the model assumes that a passenger prefers to seat rather than to stand. The segment cost of a transit leg – similar to the route section in De Cea and Fernandez (1989) – from boarding to alighting is a random variable with structural dependency on the passenger flows. Initially, the model concerned the seat allocation mechanism of a single route, but it has been extended to a general network. In the latter case, the assignment involves an equilibrium.

Another approach is investigated by Schmocker et al (2011) who adapted Kurauchi et al (2003) failure-to-board model by adding a failure-to-sit probability for each transition. The constraints only concern the seating capacity and total vehicle capacity is not addressed. They use a markovian loading process, and calculate an expected generalized cost for all the arcs, based on the standing and sitting probabilities; all these in a hyperpath framework. The underlying passenger behaviour rules are the same as in Leurent (2010) (see Chapter 3 for further discussion).

2.5 Fourth Generation: Emergence of Dynamic Models

The static assignment models presented in the previous section were developed to simulate passenger flows on macroscopic context under capacity constraints. They are used for calculating average loads during a predefined time period, assuming a uniform arrival distribution of passengers at stops. Nevertheless, they generally do not consider a variation of

the temporal profile of the demand (constant arrival rates) and the supply during the simulation period. The presence of passenger peaks during the period, together with dynamic phenomena related to transit operation, are some of the effects that cannot be adequately modelled by static transit assignment models.

The dynamic transit assignment models require an adapted demand and supply representation. The demand side translates the temporal variation of passenger arrivals. On the supply side, a transit route is described by a set of identical runs. A run is identified as a single trip of a transit vehicle from the origin to the destination, through a predefined ordered set of stops. Furthermore, the supply representation can use timetables that are closer to the observed one during regular operation. Therefore, these models, by definition, are adapted to a temporal variation of the headway, travel time and quality of service of the transit routes. Although, the dynamic assignment models demand great computational power, recent development in computers and the operational research make a step closer to modelling larger networks and high frequency lines (Tong et al, 2001).

The section's outline consists of four parts. First, various approaches for the demand and supply representation used in research models are described. Then, the route choice models are defined. Finally, some approaches for modelling congestion in dynamic assignment models are introduced.

2.5.1 Demand Representation

In the context of dynamic travel demand modelling, a dynamic representation of the demand is required. Although the spatial distribution follows the assumptions of a static model, the temporal distribution can be either exogenous or endogenous. If it is exogenous, the departure times are fixed and independent of a day-to-day evolution (Nuzzolo et al, 2001). Such time dependent matrix could be calculated with various methods, such as the one described in Wong and Tong (2001) who used data collected from ticketing systems.

In the endogenous case, only the desired arrival time of passengers is exogenous. The departure time would be calculated endogenously with respect to the passenger preferences regarding travel time, discomfort, early or late arrival etc. In that case an additional departure model must be included for each transit user (Sumi et al, 1990). It increases vastly the complexity of the model and the data requirements. However, it makes it more responsive and

behaviourally more realistic. A typical departure choice model is defined in Nguyen et al (2001). For a given class of passengers who travel from an origin to a destination, the latest departure time from the origin is calculated. It corresponds to the latest departure time, if a free-flow shortest path would be followed. That stands for reference, in order to calculate the penalties for earlier than the latest non-congested departure and for late arrival. These costs are considered jointly with other trip attributes in the utility function.

2.5.2 Service Formation

The representation of transit services is crucial for a transit assignment model. In contrast to a static transit assignment – where it is usually assumed that a frequency-based transit service is associated with waiting and where the service headways are random variables – the dynamic transit assignment may be associated to both schedule-based and frequency based transit assignment.

The schedule-based transit assignment approach relies on a disaggregated service description, on the basis of individual vehicle arrival and departure times. The waiting time corresponds to the relation between passenger and vehicle arrival time at the stop, while considering queuing due to congestion. Although, it is conceptually simple, designing an efficient representation of public transportation services consistent with the principles is a challenge. Four alternative methods have been explored for representing transit services.

First, Nuzzolo et al (2001) adopt a diachronic graph Ω – also used by Sumalee et al (2009) and Hamdouch et al (2011) –, composed of three different sub-graphs in which each node has an explicit time and space coordinate. These sub-graphs are:

- a service subgraph Ω_s , in which each run of each line is defined both in space, through its stops, and in time, through its arrival and departure time;
- a temporal centroids subgraph Ω_d , in which each node represents both temporal centroids, in order to simulate space-time characteristics of trips, and passenger arrival and departure times;
- an access-egress subgraph Ω_{ae} , which allows the connection between centroids and stops, and stops between them.

Secondly, Moller and Pedersen (1999) developed a dynamic model using a dual graph representation of a transit system, such as the one introduced in Añez et al (1996). By transforming the original links into nodes, the service network is acquired. It is composed of:

- dual nodes which represent links and retain all the characteristics of the original links;
- dual graph links which represent transit routes or transfers (including a transfer cost);
- centroids which are nodes in the original network and remain unchanged in the dual one.

Third, Nguyen et al (2001) uses a discrete space-time graph, $G = (O \cup D \cup N, A)$: O and D are the set of origins and destinations without any temporal attributes; N is a set of space-time nodes; and A is a set of space-time arcs. A consists of two groups: the first one represents the evolution in space and time; and the second group represents a temporal movement, where the objects stays in the same location. A transit station explodes into a bipartite subgraph of the space-time graph G , which designates the movements allowed in the node.

A similar representation is given by Hamdouch and Lawphongpanich (2008) who adopt a time-expanded (TE) network. It resembles the static network supply representation, but additional dimensions are added for representing time steps. Therefore, a schedule can be associated to each line, indicating the exact times of arrival and departure of each run at each station.

In the case of high frequency transit networks, the explicit description of each run of the line may be both limited by computational requirements and memory. This is one of the reasons, for which frequency-based dynamic transit assignment models have their own relevance. In Schmöcker et al (2008), the reference period is divided into intervals of intermediate size (e.g. 15 minutes), sufficiently large to justify a frequency-based approach, but still short enough for assuming that both supply and demand are static within each interval. The passengers may fail to board during one interval. In that case they are transferred to the subsequent time period, where they attempt to board again. This approach seems to be a good trade-off between the simplicity of static models and the realism of dynamic ones; especially for large-scale networks. In a similar approach Meschini et al (2007) propose a multimodal network, where

the waiting times are replaced by short-term average waiting time. That time can vary during the day, according to a temporal profile. It corresponds to the expected waiting time at stop in a frequency-based transit service. A transit frequency propagation model is also included, considering the effect of route congestion in actual transit frequencies.

In the dynamic setting a path from an origin and destination generally includes both temporal and spatial aspects. In most of the models, the path is accurately described by a space-time sequence; an origin with a departure time, access link and the vehicle runs with their corresponding time and a sequence of nodes or arcs with space and time characteristics.

In dynamic transit assignment models, the cost of a path is not stable in time and depends on the temporal and physical availability of the services and the congestion effects. This is not the only source of variability regarding the quality of service in public transportation. Indeed, this variability can also stem from the variation in the preferences of passengers. To account for such variation, a possibility is to introduce random sensitivity coefficients in the generalized cost function. The values are not constant, but they are distributed along given density functions (Tong and Wong, 1999). The generalized cost may take the form of a time-dependent generalised cost function (Schmöcker et al, 2008) or simply that of a sum of the costs of the relevant arcs.

2.5.3 Path Choice

The main particularity of dynamic assignment is that the total travel time from an origin to a destination is time dependent. It is therefore impossible to consider both arrival and departure time exogenous. For example Nguyen et al (2001) fix the arrival times and deduce from it both path choice and departure times. With the departure time model (described in section 5.1) it is possible to evaluate the influence of congestion on the path choice. An important advantage of dynamic transit assignment models is its conceptual simplicity in enforcing strict capacity constraints, abiding by the flow conservation principle.

Since path costs are time dependent, the path choice is connected to the departure time choice. Nuzzolo et al (2001) propose a dynamic path choice model based on random utility theory and a mixed choice behaviour. In other words, before leaving the origin, the passenger decides his departure time and the path choice according to the available information. En-route, at the stops the user acts in an intelligent adaptive way, choosing if to board at a run

with residual capacity when that vehicle serves the station. The pre-trip choices are influenced by past experiences through a learning mechanism, which takes into account the attributes of the perceived utilities functions, while the en-route choices depend on the information available at stops.

Nguyen et al (2001) develops a *latest departure time without congestion* algorithm in order to calculate the boarding penalty cost for each passenger class, who desire to arrive at destination at time t^* . The algorithm is based on a backward sweep of every origin from each destination and time period. Thus, an *earlier than the latest departure time* penalty cost is assumed as well as a penalty cost for arriving later or earlier at the destination than desired. In addition, vehicle capacity constraints are considered through a penalty cost, which is associated with every boarding or transfer at arc a . Each node $i(m)$ makes reference to the time segment, $m = 1, 2, \dots, n$ of the space-time graph. The boarding penalty, for every arc $a = (i(m), j) \in A_j^-$ of the backward star A_j^- of node j , added to the path cost is estimated by a standard function:

$$BP_a(\mathbf{x}) = \beta_a \left(\left[\sum_{k=0}^m x_{i(k)j} - \rho_a \kappa_a \right]^+ \right)^\theta \quad (7)$$

where x_a and κ_a are respectively the flow and capacity on arc a and β_a, ρ_a, θ are positive parameters.

Tong and Wong (1999) developed a time-dependent optimal path algorithm, introduced in Tong and Richardson (1984), which consists of three stages: a forward pass for the shortest path followed by a backward pass and a branch and bound method. The first two passes follow the Dijkstra shortest path algorithm and they determine the time windows and sub-networks used later in the branch and bound stage. In that stage the algorithm starts from the origin node and the paths branch outward from node i within the predefined sub-network. A path is complete whenever the destination is reached or the weighted arrival at a node is greater than the bound (as defined in the backward pass).

Trozzi et al (2013) propose a frequency based assignment model in a dynamic setting, which extends the concept of hyperpaths to transit networks with passenger congestion. A route

choice algorithm is developed by adapting the Decreased Order of Time (DOT) method presented in Chabini (1998) to the setting of a strategy-based transit assignment.

2.5.4 Capacity Constraints and Congestion

By construction, dynamic transit assignment models are particularly relevant when severe congestion is involved. Since the transit supply is described in disaggregated level of individual vehicles, strict capacity constraints should be easy to be implemented. However, that is not the case. This section is focused on how exactly capacity constraints and congestion can be represented in dynamic transit assignment.

An implicit approach used widely is to associate the level of service with a strictly non-decreasing arc performance function. Although, many similarities with the static context can be found, there is no straightforward link to a more complex path choice behaviour of the passengers.

Various approaches have been developed to deal with the impact of strict vehicle capacity constraints on boarding. Hamdouch and Lawphongpanich (2008) consider the formation of a passenger queue at the station for each service, with a direct impact on waiting times and user decisions – through strategy definition. Note that the model allows for multiple classes; the waiting times and decisions of each class is taken into account. Nuzzolo et al (2012) propose a modelling framework, which explicitly considers vehicles' capacity using a diachronic graph. Additional waiting time due to insufficient vehicle capacity interacts with the departure time model.

Poon et al (2004) proposed a model of route choice with congestion in a schedule-based transit network. The authors make use of an alternative minimum path algorithm already suggested in Tong and Richardson (1984) and also used in Tong and Wong (1999) in a dynamic model without capacity effects. By introducing a bottleneck model at each boarding link, a queuing time is calculated for a passenger at each time increment, through the difference between cumulative arrival and departure curves in a platform node.

In the quasi-dynamic frequency-based assignment model of Schmocker et al (2008), a failure node is added to ensure priority for on-board passengers. In addition, fail-to-board arcs transfer the excess demand back to the stop, ensuring a strict capacity constraint. These

failing-to-board passengers are transferred to the subsequent time interval, suffering a delay, the overcrowding penalty.

Building on the frequency-based multimodal assignment model of Meschini et al (2007), Trozzi et al. (2010) extend the Bottleneck Queue Model with time varying exit capacity and apply it to separate queues. A dummy arc is included for representing queuing when the services are saturated. Therefore, the waiting and queuing time is calculated as the waiting time before a k^{th} vehicle arrival. The stop model is included to the dynamic user equilibrium model presented in Trozzi et al (2013), which extends the shortest hyperpath to the dynamic setting.

Beyond the models addressing service total capacity, other researches focus on in-vehicle comfort. For example, Nuzzolo et al (2001) incorporates a discomfort cost attributed to all the passengers on-board and related to the vehicle load, without dealing with the strict capacity of the vehicle. Although that cost is included in the generalised cost, it does not affect the actual travel time. Tian et al (2007) proposed an alternative model with multiple access nodes and a single egress node. They focused on the trade-off between on-board standing congestion and seating capacity, which has an impact in departure time choice.

Finally, Sumalee et al (2009) develop a model based on a diachronic graph that seeks to differentiate the discomfort level perceived by sitting and standing passengers. Therefore, the seating probability depends on the number of available seats, on the priority rules – on-board passengers over boarding passengers – and on the passenger's willingness to get a seat. The last one depends on the time a passenger has been standing and the remaining journey time. To apprehend the stochasticity of seat availability, a Gaussian random error term is introduced in the disutility for each path. Hamdouch et al (2011) extend this model to include both vehicle capacity and seat availability.

2.6 Fifth Generation: Simulation-Based Models

The fifth generation is characterised by dynamic microscopic modelling. This field of research, corresponding to the simulation-based approach, is focused in dynamically modelling the movement of both vehicles and passengers on a network, with a very detailed

description of the space. It further extends in the modelling of the behaviour of the agents and the randomness of certain aspects. Significant effort is made to develop simulation-based transit models, such as the Mezzo model described in Toledo et al (2010). However, the literature on passenger traffic dynamic modelling mainly concerns the circulation of passengers in confined spaces – such as transit stations (Hoogendoorn et al., 2004) – and the interaction between pedestrians and various objects – such as vehicles, especially when passengers are boarding and alighting (Fiegel et al., 2009). Many approaches have been tested to model the space and the agents. These models are generally based on a detailed representation of space, the abilities and the general attributes of the pedestrians (Kluepfel et al., 2003).

In these models, the reference space is described as a combination of accessible and inaccessible zones. The points of attraction, the obstacles and the events that generate flow have to be also defined (Banos and Charpentier, 2010). In some models, the physical space accessible to passengers is usually represented as a grid of square cells of various surfaces (Blue and Adler, 2001; Banos and Charpentier, 2010; Still, 2000). Other models represent physical space as a two-dimensional space within which the pedestrian is moving with complete freedom of movement (Teknomo, 2006). This space is endowed with specific landmarks and directions, which can influence the path of the pedestrians. The grid representation of the space is usually associated with cellular automata models, while the two-dimensional space is used in the multi-agent systems, which are described right next.

Microsimulation models differ with respect to the way the pedestrian behaviour and route choice are modelled. Two general categories can be distinguished:

- Cellular automata (CA) and
- Multi-agent systems (MAS)

The cellular automata approach models a person by an entity (automatum) with limited intelligence, which is able to fulfil at most basic safety manoeuvres (Still, 2000). It is basically a probabilistic model. The automata can execute three main movements: side stepping, forward movement and conflict mitigation (Blue and Adler, 2001). Its velocity depends on the position of the entities in the immediate environment: the presence of other

entities may result in the adoption of a different velocity from the desired one, in order to avoid collision or passing through other automata.

The multi-agent systems consider a different representation of a person, by an agent. The agents navigate in a dynamic environment composed of a set of constraints and opportunities. It is based on the Social Forces Model (Helbing and Molnár, 1995), according to which the interaction with the environment is the result of social forces: repulsions and attractions. The repulsive force corresponds to the willingness of an agent to create his own private sphere and keep distance himself from other pedestrians or area borders and obstacles. The attractive force is the result of the attraction by other people or objects; though it is decreasing with exposure. The path of a pedestrian is therefore subject to the vector of attractive and repulsive forces. Additional randomness may be considered by adding a fluctuation term that produces oscillations in individual trajectories (Helbing and Molnár, 1995).

In the approach of Hoogendoorn and Bovy (2004) three levels of pedestrian behaviour are distinguished: strategic, tactical and operational level. They particularly describe the pedestrian behaviour at the tactical level: specifically how a pedestrian chooses his activity schedule and area and the route to reach it. The authors consider the pedestrian as an economic agent who seeks to minimize his expected disutility and they propose an optimization program based on solving a dynamic programming equation for decision-making in continuous time and space under uncertainty. This optimization program can be solved by approximate solutions in a discrete space and discrete time framework and it is compatible with the architecture of the pedestrian flow simulation model NOMAD (Hoogendoorn, 2001).

A validation method for the stochastic behavioural models, that model explicitly the interactions of the pedestrians with their environment, is that these models are able reproduce the macroscopic flow behaviour of the pedestrians. Validation is required in all possible environments. In particular in high pedestrian densities, the presence of various phenomena such as pedestrian auto-organization into lanes or cluster formations is expected, as described at a macroscopic level in Hoogendoorn and Daamen (2005). The microscopic models aim to reproduce pedestrian trajectories within a certain area, under certain conditions and capacity constraints. Even though some are used for modelling large transit centres, they have fixed origin – destination matrices and they cannot consider the influence of station circulation on transfer quality and particular transit route choices.

Simulation-based transit assignment models are developed mainly to offer a dynamic aspect on transit operations, especially in interaction with other components of the transit system, such as passengers. In order to curb the extensive computational effort required on microscopic simulations, Toledo et al (2010) develop a multi-agent system, the Mezzo model, which can be qualified as a mesoscopic model, since it uses a simplified supply network.

The bus vehicle is the main modelling entity of the model, characterized by a number of attributes: the bus type with specific characteristics, the bus line for which it is generated and a bus route (sequence of stops) to be executed. The bus vehicles circulate on sections between stops, expressed as network links. These are made up of two parts: a running part where the vehicles do not suffer any delay and a queuing part, which corresponds to the additional time needed for exiting the link, calculated with respect to the link's exit capacity and vehicle density. Since Mezzo models cars as an additional entity, the network links can be used to model segregated or mixed traffic conditions.

Although transit operations are described with a high level of detail, passenger demand is sufficiently represented. Passengers board individual vehicles to arrive to their destination and choose their path on the basis of the network attributes: transit routes, access and egress links, estimated travel times and schedules or frequencies. The generation of the path choice set is computationally demanding. However specific algorithms have been developed for constraining the choice set and facilitating the passengers' path choice.

Simulation-based transit models are similar to the microscopic dynamic models for pedestrians in confined spaces, in the sense that they are generally agent-based simulations containing specific entities (bus vehicles, cars and passengers in Mezzo), each one with particular characteristics. Their main differences lay in the representation of space (or supply) and the level of complexity of passenger's choice in the respective networks. Furthermore, we remark that in the case of simulation-based transit models, the case studies presented are of small size. Large-scale applications are not sufficiently addressed.

2.7 An Overview of the Transit Assignment Models

This chapter describes the historical development and the subsequent branching of the transit assignment models. Initially, these models (1st generation) were based on pre-existing road traffic models, adapted for considering transit specific phenomena, such as waiting. However, with the advent of more robust models (2nd generation) focusing on the individual route choice behaviour, the research community focused on more complex problems, such as enforcing capacity constraints and considering the temporal variation of supply and demand. Different modelling motivations and objectives led to branching into three main generations: static transit assignment (3rd generation), largely dominated by frequency-based modelling that considerer various capacity constraints; dynamic transit assignment (4th generation) where models are more often schedule-based; and an alternative simulation-based approach. The latter includes microscopic models developed specifically to capture the movement of pedestrians in closed areas. Figure 1 illustrates that historical evolution.

The first generation of transit assignment models is characterised by a clearly visible inheritance from road traffic assignment models: at this stage transit assignment was still in the process of acquiring the autonomy as a research field. The particularities of public transport are progressively taken into account, by more or less relevant means. Among them, particular interest was given in the influence of transfers in route choice; or the allocation of passengers among overlapping lines. These models are characterised by an engineering approach, avoiding complex mathematical formalisation, and trying to make a trade-off between the computational constraints of the time and the complexity of the phenomena in public transportation.

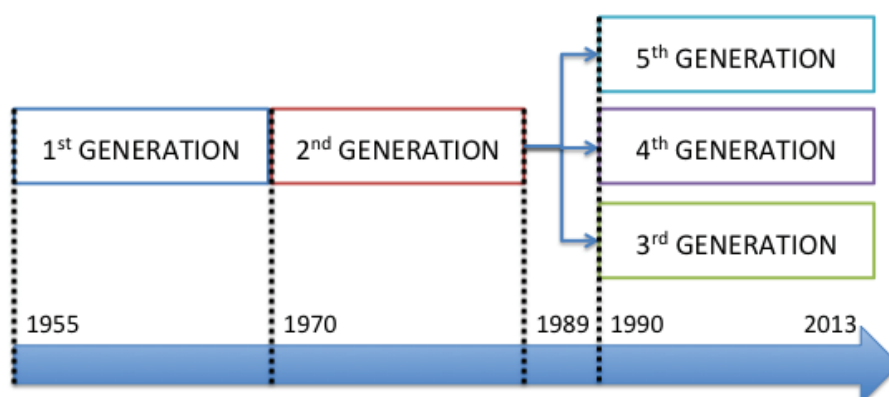


Figure 1: Five Generations of Transit Assignment Modelling

The second generation of transit assignment models brings answers to some of the questions previously raised by the public transportation particularities. These were associated with the robust mathematic formalization of the models developed. It is a fundamental generation, since it overcomes critical technical locks: overlapping lines (Chriqui and Robillard, 1975), passenger's route choice behaviour (Spiess and Florian, 1989) and appropriate supply and demand representation (Nguyen and Pallottino, 1988). These principles are the basis of many modern transit assignment models. Alongside, particular attention was given to the appropriate mathematic formalization of the conceptual models and the development of efficient algorithms.

Current generation of static transit assignment models is mainly based on adjustments of the framework developed priory for passenger flow assignment, except for one major point: the influence of capacity constraints on quality of service and route choice. Various ways to model congestion are considered. Generally, the contributions develop a conceptual model and then build a relevant optimization problem. They define the appropriate demand and supply representation and propose an efficient algorithm. The models developed do not include all the capacity effects identified in the literature. Their current capabilities include the passenger capacity of vehicles and lines and the seating capacity. These have an impact on waiting time and quality of service and therefore on route choice.

At the same time, new methodological approaches emerged and progressively took importance. A dynamic aspect is added, as they include a temporal variation of both demand and supply. They mainly concern schedule-based models that assume a stochastic behaviour of the demand but a deterministic supply of services. That is the reason that some frequency-based dynamic models have been developed, due to the efficiency in capturing the stochastic behaviour of transit supply. Recent dynamic models deal more thoroughly with capacity constraints. Their main advantages laid in their ability to capture dynamic phenomena, such as queuing and departure time choice. Despite their recent advances, they are still computationally demanding, and they seem more appropriate for medium-sized networks.

Specific simulation-based models also emerged for solving particular problems. These models are applied in two particular circumstances: simulating the passenger and vehicle flows in

transit networks and the movement of pedestrians in confined spaces, such as stations. The latter models focus on the pedestrian's behaviour at an individual level. While they present many limitations, they can have a wide variety of applications in transit modelling: from modelling the circulating conditions within a large transit station – in conjunction with the path assigned from a transit assignment model – to modelling the variability of the dwelling time and the boarding and alighting procedure.

We observe that once the fundamental framework, dealing with the basic phenomena in public transportation was developed, a variety of models have been proposed. That led to a multitude of models that can be grouped into the three recent generations. The generations continue to evolve independently in order to address generation-specific problems. They also evolve in conjunction with each other in order to deal with new modelling issues, such as the behavioural changes from the development of intelligent transit systems.

Chapter 3:

Modelling Capacity Effects in Transit

Assignment: a State of the Art

3.1 Introduction

In densely populated areas massive passenger flows over a public transportation network may degrade the quality of service of the transit system and, thus, reduce its overall attractiveness. Indeed, the perception of the quality of service is the most important characteristic of the modal choice and route choice of the users. For example, Leurent and Liu (2009) apply a seat availability model of in-vehicle comfort and estimate that the link flows are increased up to 15% at certain metro sections.

The previous chapter demonstrated the important effort to improve the transit assignment models by addressing the impact of the capacity constraints on passengers' decisions. Researchers were primarily interested in making explicit the effect of total passenger capacity of a vehicle or a service on the passengers' route choice. Four families of models emerged; distinguished by the way they consider the total persons' capacity effects: the Effective Frequencies model; the Failure-to-Board model; the User Preference Set model; and the Dual Penalty model.

Furthermore, a parallel category of models focused on the in-vehicle comfort, with regard to the seat capacity of the vehicles. Three main models, or group of models, can be distinguished: the Seated Capacity model; the Failure-to-Sit model; and the Seat Availability model.

This chapter is structured in five parts. First, a panorama of the capacity effects on public transportation is provided. Afterwards, the four main types of models representing total capacity are described theoretically, analysed and then compared. The description and comparison of the three assignment models for seat availability follows. The conclusion focuses on some appreciation of the transit assignment models and current modelling needs.

3.2 A Typology of Capacity Effects in Public Transportation

The first step for addressing the passenger flow assignment under capacity constraints is to provide a consistent view of the capacity effects, which affect public transport networks. A thorough description of the technical characteristics of a transit system is available in the 2nd Edition of the Transit Capacity and Quality of Service Manual (TRB, 2003). The document synthesizes the capacity characteristics and associates them with congestion for all public transportation modes.

A more systematic analysis of these capacity effects was provided in Leurent (2011a). The author proceeds to a systemic analysis of the public transportation system, in two main stages. First, the major subsystems of the public transport system are identified and described. Then, their relations, and particularly those related to capacity effects, are analysed. Five main subsystems, or component categories, are isolated: the line, the station, the vehicle, the passengers and the global management. On the basis of this segmentation, seven capacity effects are identified, as listed below:

- (I) **The vehicle capacity of an infrastructure:** It is a straightforward capacity constraint, since an infrastructure cannot accommodate more than a certain number of vehicles for a fixed period. It is an essential aspect for assuring the efficiency of the services, but in general it is absent from transit assignment models. A variety of mode-specific vehicle traffic models exist, but passengers are usually absent, or of limited importance. A standard capacity method assessment is provided by UIC Leaflet 406, for rail transport; other researches, such as Lai et al (2011), and Abril et al (2008) focus on the effect of train signalling, service mix on a line, etc.

- (II) **Operating capacity of a route:** While vehicle fleet size can be considered variable in long term, it is safe to consider it fix from the perspective of short term planning. Then, the vehicle usage demand could affect the planned transit schedule. Indeed, as modelled by Lam et al (2002), the increase of a journey cycle time due, for example, to a deterioration of commercial speed, will reduce the frequency of the service, for a fixed fleet size. The same consequence occurs if other operational characteristics, such as dwell time, impact the movement of vehicles.
- (III) **The passenger capacity of a vehicle:** At the scale of the vehicle, two types of capacity can be distinguished: the seat capacity and the total person capacity of a vehicle. It is usually expressed as a strict capacity, measured as the maximum volume of passengers that can be accommodated. A third capacity, of a different nature, can be considered: the passenger exchange capacity, i.e. the number of passengers boarding and alighting on a given period, potentially depending on the platform and in-vehicle conditions.
- (IV) **The passenger capacity of a route:** As suggested above the seat capacity and the total person capacity can be considered at a single vehicle. They can also be considered at the level of a service route, which combines the capacity of all relevant vehicles. It also involves the service frequency and represents the maximum flow of passengers that can be transported on a given period.
- (V) **The passenger capacity of a station:** At the station level, the pedestrian traffic and the passenger storage capacity can be determined per type of station element for the different areas of a station. A station can be a complex system – even more if it is a large intermodal facility – and many microscopic station models were developed (Hoogendoorn et al, 2004, Hoogendoorn, 2005). However by construction, they are limited in the route choice of passengers within the station and not to the repercussion on the transit network.
- (VI) **The vehicle storage and movement capacity of a station:** This item regroups capacity effects relevant to the operation of transit services. These are related to the movement or presence of vehicles in stations.

(VII) **The capacity of a station as an interface with personal transport modes:** A great interest has emerged recently, regarding the development of intermodal facilities, especially in relation with individual modes (motorized or not). A given station, with a given layout, has a certain capacity to allow for intermodal trips combining transit and car or bicycle, both in terms of circulation and storage. More precisely, this capacity is related to facilities, such as parking, park and ride equipment and bike shelters.

These capacity effects interact with one another, in a complex but organized way illustrated by Figure 2. They are indicated by arrows that link the subsystems and the components of these subsystems together. An arrow between two items indicates that the first limits the second in terms of capacity. For example, fleet size and service time both limit route frequency; dwell time limits service time etc. The Latin number makes reference to the previous list of capacity constraints.

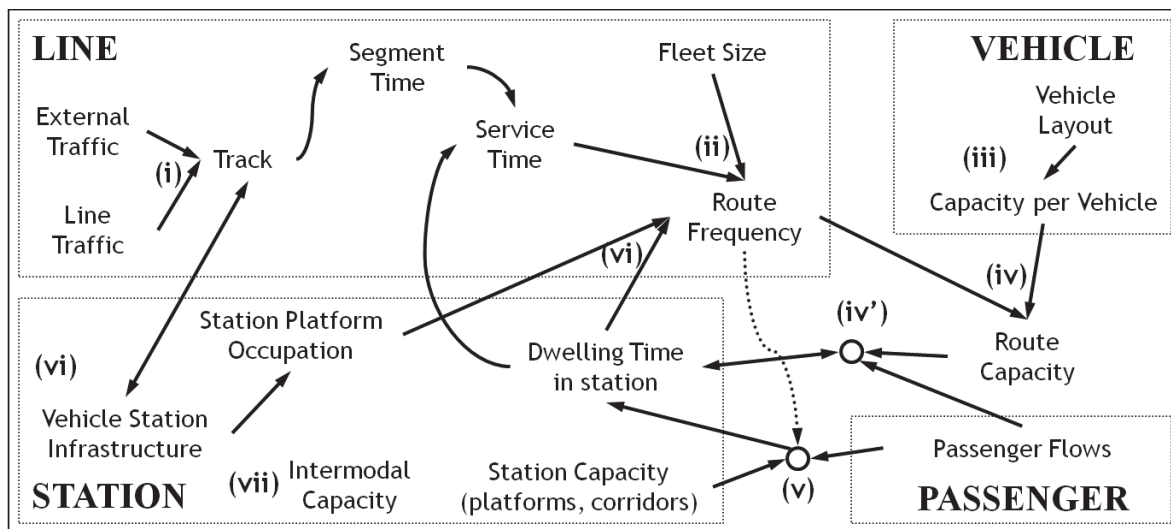


Figure 2: A schematic representation of the elements of the transit systems and how they influence the capacity effects (adapted from Leurent, 2011a)

The capacity effects have given rise to a number of scientific studies and models (see I-VII). Nonetheless, this chapter is focused on the capacity constraints more present to the transit assignment models: passenger total capacity and seat capacity.

3.3 Passenger Total Capacity

Throughout the development of transit assignment models, modelling the effect of passenger total capacity of a route or a vehicle progressively became a major topic of research. Indeed, a large number of static assignment models, and some dynamic or quasi-dynamic ones have been proposed to deal with large-scale transit networks. These models are based on both physical and economic approaches: some of them assume strict capacity constraints; other not.

This section describes the main models, which address the passenger total capacity effect. First, the “Effective Frequencies” model provides an economic approach of passenger congestion, where the increased waiting time (i.e. the time interval between the passenger arrival at the station and the moment they are able to board in a vehicle) is modelled as function of a perceived local frequency that decreases with the residual capacity of vehicles (Cepeda et al., 2006). Second, the “Fail-to-Board” model initially presented in Kurauchi et al (2003) addresses the common lines problem by adapting results from queuing theory. Third, the “User Preference Set” introduces an alternative notion of a strategy. Finally, the “Dual Penalty” model suggests a penalty introduced on congested segments, while imposing a strict capacity constraint.

3.3.1 The Effective Frequencies Approach

3.3.1.1 Model Description

The effective frequencies model is initially introduced by De Cea and Fernandez (1993). It is currently the most widely accepted of the static assignment models treating congestion. The vehicle capacity is tackled by a simple mechanism: if a passenger wants to board a heavily loaded vehicle, he may have to wait for several vehicles of the service to stop before succeeding to board. That is modelled as a local reduction of the service frequency. It has also a direct impact on the waiting time and route choice of the passengers. The detail of the specification is discussed below. Similarly, Wu et al (1994) suggest a model, which modifies only the waiting time needed until boarding a vehicle. Other approaches have been proposed; for example Gendreau (1984) addresses congestion using results from queuing theory. That implies an unbounded increasing delay function of passenger flows. Bouzaiene-Ayari et al (2001) replaces the frequencies with an attraction factor, a strictly decreasing function of

passenger flows. That attraction factor is used for the calculation of the local route choice probabilities and the waiting time.

An additional aspect of the effective frequencies approach is that only alters the value of the service frequency. Therefore, a frequency-based model jointly affects the waiting time and the route choice. In the same time it adheres by the architecture of frequency-based transit assignment modes, which makes this model particularly attractive.

In Cepeda et al (2006), the authors propose a general modelling framework assuming a flow dependent frequency function. They use an extremal formulation of the local route choice to derive a duality gap function for the network assignment. That guarantees that the equilibrium is reached when the duality gap is null. We consider k as the vehicle capacity, x_a^o and x_a^+ as the total flow on-board after the stop and the flow boarding respectively, φ_a^0 and $\hat{\varphi}_a$ the nominal service frequency and the effective frequency. The available capacity after the stop is given by $\varphi_a k - x_a^o$. Cepeda et al (2006) specify the effective frequencies function on the basis of a BPR function with a parameter β , as follows:

$$\varphi_a'(\mathbf{x}) = \begin{cases} \varphi_a [1 - (\frac{x_a^+}{\varphi_a k - x_a^o + x_a^+})^\beta] & \text{if } x_a^o < \varphi_a k \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

3.3.1.2 Implementation

Although various algorithms have been defined in some of the articles cited previously, it is in Cepeda et al (2006) that a general solution algorithm with a convergence proof is proposed. A heuristic minimization method is proposed for solving the equilibrium problem on the basis of the method of successive averages (MSA). At each iteration the frequency at boarding can be updated without modifying the general architecture of the model. Other algorithms exist, such as in Cominetti and Correa (2001).

The initialization step of the MSA algorithm needs a feasible initial flow. It can be calculated by a free flow assignment on the shortest hyperpaths, using nominal frequencies, as a free-flow assignment. In order to ensure that the flows are feasible, two solutions are suggested. The first one is to provide a subgraph with infinite capacity, such as a pedestrian network, to

absorb any excess flow. The second one is to make sure that the effective frequencies never go below a certain, given minimum. In other words, $\varphi_a^\varepsilon(x_a) = \max\{\varphi'_a(x_a); \varepsilon\}$, where ε is a small positive parameter.

3.3.1.3 Discussion

It was demonstrated in Cepeda et al (2006) that whenever there is sufficient overall capacity (or the previous uniqueness curbing techniques are applied) the model reaches a steady state. However, there is no proof that flow does not exceed capacity at convergence.

An additional issue arises with the type of BPR function like Equation (8). In fact, with the parameters proposed in Cepeda et al (2006), a small boarding flow in relation with the available capacity results in a significant reduction in the effective frequency – and a respective increase of the waiting time. Figure 3 illustrates the effective frequency – as a portion of the nominal frequency – in relation to the boarding flow – as a portion of the available capacity, for various values of β .

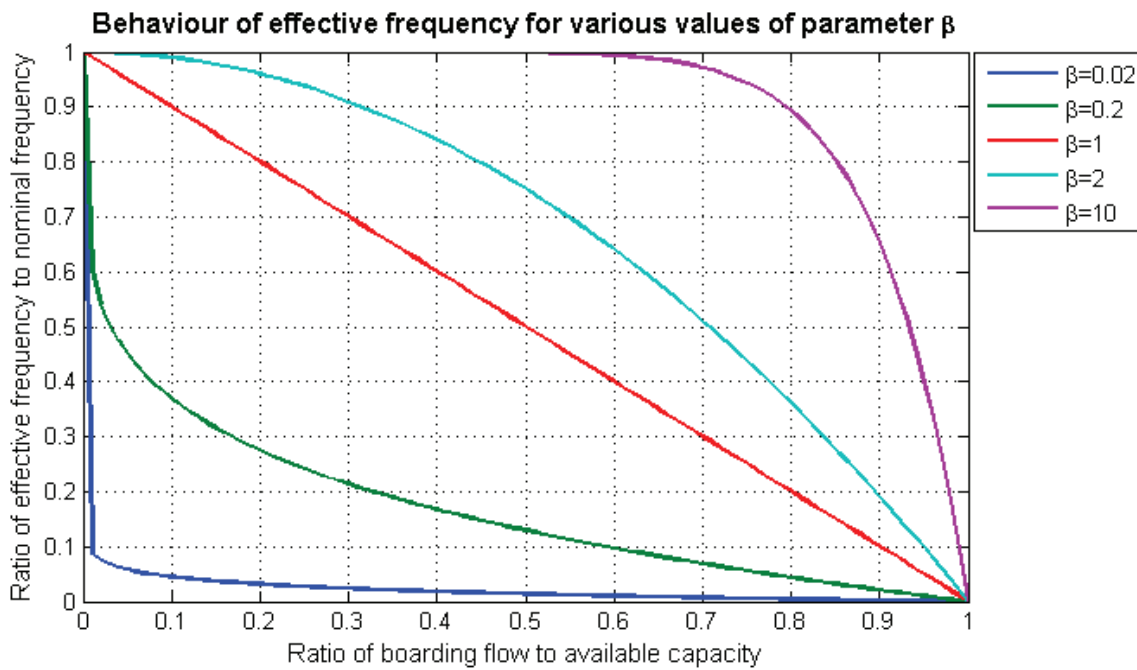


Figure 3: The effective frequencies function for various values of beta

The selection of a small β (for example $\beta = 0,2$ in the numerical example in Cepeda et al., 2006) reduces very significantly the effective frequency and is not realistic. However it results

in a quick convergence to the equilibrium. A bigger value of β would be more acceptable, despite its large impact on convergence speed.

The unrealistic reduction of the service frequency which results in large increase in the waiting time implies that at convergence the passenger flows correspond to the optimal solution, but the cost is far from realistic. Therefore, the waiting times cannot be used for the calculation of a congestion cost. Nevertheless, the great advantage of the Effective Frequencies model is its easy implementation, since it is merely an extension of the basic frequency based static assignment model.

3.3.2 The Failure-to-board Model

3.3.2.1 Model Description

The failure-to-board modelling approach was initially introduced in Kurauchi et al (2003). It is an adaptation of the queuing theory to the problem of common lines. Indeed, making explicit waiting times is not easy when multiple services are attractive. In fact the attractiveness of a line depends also on the expected waiting time until boarding, and that in turn depends on the number of passengers finding this line attractive – this volume is related to its available capacity. As a matter of fact, queuing theory is relevant to describe the saturation phenomena of a demand sensitive to waiting and in-vehicle travel times. Markov chains are used for the loading process, using the route shares (or in that case transition probabilities) obtained by the physical flow conditions and the shortest hyperpaths. The states of this Markov process correspond to the origin, intermediate and destination nodes. Multiplying the transition probability matrix by destination, yields the probabilities that passenger traffic traverses a given node. That, in turn is used for calculating the passenger traffic by destination.

With regard to the boarding process, the model considers the line's available capacity after passengers have alighted at a station (supply) and confronts it to the volume wanting to board (demand). As a result, passengers have a given probability to be able to board the first vehicle to arrive at the stop. In the contrary case, they fail, with a complementary probability, the failure-to-board probability, or p_F , of which the model takes its name. The probability of a passenger to board the first vehicle is then $1 - p_F$.

In Kurauchi et al (2003), which will be referred to as the “static” model, the excess flow of passengers who fail to board are directed towards an escape arc, leading them directly to their destinations, and facing a significant cost. Thus, the capacity constraints are clearly satisfied, at the expense of the conservation of flow. To overcome this major limitation, Schmocker et al (2008) proposed a quasi-dynamic version of the model, where a waiting passenger who fails to board the line at that time interval is transferred to the following time interval, and is submitted to the corresponding waiting time.

A penalty is associated with the event of failing to board and thus it is integrated in the generalised cost. That penalty can be interpreted as the cost of passengers to be exposed to risk (implicitly assuming that they are risk averse). Whatsoever, Kurauchi et al (2003) does not explain clearly how this penalty should depend on the failure-to-board probability. In the quasi-dynamic version, Schmocker et al (2008) transform that cost into an additional waiting time related to the possibility of failing to board. If T is the duration of a time interval, a passenger faces an additional waiting time before boarding d_B given by the function:

$$d_B = \frac{p_F}{1 - p_F} T$$

The model adds that penalty to the generalised cost of the passenger; however, the model does not take into account the possibility for the passenger to modify his route choice at the stop and reduce his travel time.

3.3.2.2 Implementation

A classic hyperpath framework, as described in Nguyen and Pallotino (1988), is adopted in order to assign the demand in a transit network by minimising the expected travel time for each user. Once the attractive set of lines is defined, the passenger flow is assigned with respect to the line frequencies. Due to the additive cost structures of the arcs, the Bellman equation can be used for the construction of the minimum cost hyperpath.

An extension of the model is described in Shimamoto et al (2005) for the optimization of transit fares. The failure-to-board assignment approach is used within a multi-objective optimization problem. The lower lever corresponds to the equilibrium assignment of passenger flows with the failure-to-board model. The upper level refers to the total cost and

the connectivity reliability for each origin-destination pair – based on the fail-to-board probabilities. A non-dominated sorting genetic algorithm (NSGA-II) is used to resolve the multi-objective optimisation problem. The minimization problem associated with the transit assignment is a result of the combination of the classic method of successive averages (MSA) and the absorbing Markov chains.

3.3.2.3 Discussion

As mentioned above, the failure-to-board model uses a queuing model approach in order to represent the passengers' behaviour when dealing with a lack of vehicle capacity, while keeping the classic hyperpath framework. Although the model avoids a detailed physical description of the waiting process, a failure penalty is calculated. That is considered jointly with a risk averseness parameter, θ . If that parameter is zero, the passengers choose the hyperpath with the shortest travel time. When θ increases, so does the effort to avoid the risk of failing to board a vehicle.

Although, the quasi-dynamic model transfers the fail-to board volumes to the following time intervals, the input demand needs to be inferior to the overall capacity in order to be entirely assigned to the transit routes. That implies some special border conditions, with a cooling-off period at the end of the period. In addition, the quasi-dynamic model does not include a departure time choice model. Therefore, with a fixed temporal profile for the demand, the capacity constraints will only result in a delay at the destination. In that sense the trade-off between departure time and travel conditions is not captured, as it is in other dynamic models.

3.3.3 The User Preference Set Model

3.3.3.1 Model Description

Hamdouch et al (2004) argues that when transposing Wardropian equilibrium in transit networks with rigid capacities or fixed schedules, the strategy defined in Spiess and Florian is not sufficient and a different definition of strategy is needed. Thus, the authors adopt the concept of Marcotte and Nguyen (1998), which specifies a set of attractive lines and further orders them by attractiveness (in comparison to the classic concept where the common lines at a station are simply divided to attractive and not attractive ones). Applied to a congested transit network, this results in creating for each node and each destination an ordered list of

successor nodes. Therefore, a user defines a preference set of lines at each node. The model follows a classic assignment procedure by evaluating the arc costs, constructing their optimal strategies and then assigning flows on the arcs.

This model, named User Preference Set (UPS) is based on an access probability, π_a , which is the probability of a passenger to access each successor node. If we consider κ'_a the available capacity of the arc a and x_a the flow, then the access probability, π_a , is:

$$\pi_a = \min\{1; \frac{\kappa'_a}{x_a}\}$$

The calculation of the access probabilities for each outgoing arc leads to the calculation of the cost in the construction of the optimal strategy.

The authors' definition of strategy assumes that the flow is assigned to the attractive arcs as long as there is available capacity, according to the access probabilities calculated as in the optimal strategy. Once, their capacity is reached, the excess flow is assigned to the next line in the preference set and so on. In the dynamic version of the model, Hamdouch and Lawphongpanich (2008) adapt this ordered preference list of the successor nodes and make it sensitive to the temporal availability of the supply.

In addition, the dynamic model considers the priority of on-board passengers in a vehicle over the boarding passengers, by distinguishing two classes of users. The assignment mechanism deals firstly with the priority passengers, who have an access probability $\pi_a^1 = 1$ if they don't violate the capacity constraints. Then, the access probabilities are calculated for the non-priority users. The passengers who fail to board a given vehicle are assigned to the following vehicle arrival, where the boarding mechanism is repeated. The cost of each strategy is derived from the path costs and the access probabilities for each class of users, while no penalties are included.

3.3.3.2 Implementation

As mentioned beforehand, the assignment procedure of the model seems classic. First, the access probabilities are calculated for each boarding arc, along with the arc cost. Then, the optimal strategy is constructed and the flow is loaded. The loading mechanism is based on the

ordered preferences set for each node. After the first iteration, where the strategy is reduced to the shortest path, the saturated arcs are removed from the preference set, and the residual capacity is updated. Then, the access probabilities for the arcs in the preference set are being recalculated and the excess flow is assigned.

The authors use a recursive process for the definition of the optimal strategy. Indeed, the optimal preference order for each node i is calculated by sorting the expected travel costs from i to the destination. The construction process goes through the nodes in a reverse topological order, in accordance to Nguyen and Pallottino (1989). After the local assignment, a dynamic programming equation is used to compute the costs from the node to the destination for each user class.

In order to reach the strategic equilibrium two approaches are considered, whilst the convergence is measured through the relative gap function; a non-negative function whose minimum is achieved for the equilibrium flow. The first approach is to reach the equilibrium flows by a Frank and Wolfe (1956) algorithm of linear programming. The flow vector of each iteration is computed through a convex combination where a step-size, specific to each OD pair, is calculated with respect to the distance of the current to the auxiliary flow vector and the number of iterations already executed. The second approach consists in a projection algorithm using the Euclidean projection of the flow vector on the convex set of feasible flows.

3.3.3.3 Discussion

The User Preference Set model adapts the strategy concept to a congested transit network. It is based on an explicit in-vehicle capacity and provides an according definition of the strategy. Nevertheless, contrary to what was proposed in Spiess and Florian (1989), the path cost does not include any penalties due to the flow saturation or the failing of the candidate passengers to board a given vehicle. That aspect is modified in the dynamic model in Hamdouch and Lawphongpanich (2008).

The model addresses efficiently the priority of the passengers who are on board, over those waiting to board. However, at the level of strategies, the model fails to deal with the combined frequency problem of common routes and the effect of congestion on their availabilities. Without saturation, the model is reduced to an optimal path assignment using average waiting

times, rather than an optimal strategy of Spiess and Florian (1989). In addition, the authors seem not to deal with the combined treatment of different destinations. Doing so is indeed a problem, since the ordered preference list depends on the destination and on the intermediate nodes, thus causing ambiguities in the definition of the probability of boarding. Finally, the case of a cyclical network is also disregarded, reducing the model's domain of application.

3.3.4 The Dual Penalty Model

3.3.4.1 Model Description

The last of the modelling approaches presented in this section is the work of Lam et al (1999b). The authors address the capacity constraints on a transit network by imposing a penalty on the congested segments. Indeed, they assume that when the demand for a service exceeds the available capacity, only a part of the passengers will be able to board in the first vehicle at a station, or will opt for an alternative route. Whatsoever, these boarding passengers will suffer a passenger overload delay due to the increase of their waiting time.

The delay is determined endogenously at the equilibrium on the basis of the characteristics of the congested transit network. The authors do not propose a direct formulation of the congestion penalty, but rather that is implicitly calculated in conjunction with alternative route choices. In fact, contrary to the previous models, where a delay is calculated in a straightforward way from the relation of the volume of candidate passengers and the available capacity offered, in Lam et al (1999b) the delay can be expressed as the penalty for which passengers are dispatched to alternative paths.

It is further assumed that the passengers waiting at a station create distinct queues for every line, and that they also form groups according to their egress stations. The service rate for each queue corresponds to the capacity available on the line. The model suggests an explicit representation of the in-vehicle capacity constraints. Even though some arcs may be locally at capacity, a sufficient network capacity for every OD path flow is necessary for the convergence of such a network with passenger queues (Bell, 1995).

3.3.4.2 Implementation

In Lam et al (1999b) a multipath assignment procedure is used. The procedure is adapted to account for the variability in the perception of time by passengers, and also their limited knowledge of the actual vehicle timetables. It is also adapted to account for congestion, on the basis of the dual penalty approach of which the principle is described above. Let us now look at it in detail.

Let us define the passenger overload delay d_a at an in-vehicle link a connecting a station i . When the flow is lower than the available capacity, $x_a \leq \kappa'_a$, the passengers do not suffer any additional cost. On the other hand, when $x_a > \kappa'_a$, insufficient capacity prevents some passengers to board the first vehicle. The congestion penalty is integrated in the arc cost. With t_a the in-vehicle travel time and w_a the waiting time of the attractive set, the cost function is:

$$g_a = t_a + w_a + d_a$$

Formally, the passenger overload delay of a route is calculated through the Lagrange multiplier of the Lagrangian of the objective function. It is therefore the value for which, the passenger flow is inferior to the capacity constraints.

The waiting time is calculated as a combination of the frequencies of the attractive lines, as in Chriqui and Robillard (1975). Nevertheless, when addressing the congested arcs, the authors implement a logit model for traffic assignment on alternative arcs. Therefore, the ratio among alternative routes stems from the difference of their cost. The logit scale, parameter θ is associated, according to the authors, to the degree of passengers' knowledge of the services' actual timetables. Increasing the value of the parameter implies possessing better knowledge, up to considering perfect knowledge ($\theta \rightarrow \infty$), which reduces the Stochastic User Equilibrium (SUE) to a Deterministic User Equilibrium (DUE). The equilibrium is achieved through an iterative process, using a method of successive averages (MSA).

3.3.4.3 Discussion

By developing a stochastic approach for the equilibrium assignment problem and an explicit in-vehicle capacity constraint, the dual penalty model expands the conventional transit assignment framework. It provides an explicit calculation of the cost of the capacity effect, in

competition with other alternative lines. Furthermore, this cost is determined using a stochastic user equilibrium and it is incorporated in the path cost of a passenger.

Nevertheless, even though the authors establish that cost as a penalty for passengers wanting to board, it is assigned to the in-vehicle arcs of the transit network. Thus, it is applied to all the passengers on-board the vehicle and although it can be interpreted as a discomfort penalty, due to the overcrowding of the vehicle, it is incorrect to consider it as a waiting time before boarding.

3.3.5 A Theoretical Comparison of the Total Person Capacity Models

All the approaches discussed above share the common objective to account for the passenger capacity of vehicles and transit services. These approaches share some similarities: queuing theory is often involved to analyse the relationship between passenger stocks (queues) and flows; classical transit assignment models are also often the ground on which these original approaches are based on.

However, they differ a lot with respect to passenger behaviour. In order to compare them in a systematic way, the criteria set of Leurent and Askoura (2010) is used below. Table 2 regroups the characteristics of the capacity models, based on the representation of time and the following other criteria: first, the representation of the capacity constraint, in other words whether it is modelled explicitly or implicitly; second, the presence of a priority rule for the passengers on-board over the boarding; third, the method the waiting time at station is calculated; fourth, the distribution of the boarding passengers among the attractive routes for a given destination.

Table 2: A theoretical comparison of transit assignment models considering total persons' capacity constraints

Capacity Model	Effective Frequencies (FF)	Dual Penalty (DP)	Failure-to-Board (FtB)	User Preference Strategy (UPS)
Representation of time	Static	Static	Static and Pseudo-dynamic	Static and Dynamic
Representation of the capacity constraint	Implicit through reduced frequency at boarding	Explicit, IVTT penalty when reaching capacity	Explicit. A Failure arc to carry excess flow	Explicit. Filling arcs one by one
Priority of passengers on-board over boarding passengers	Explicit. Capacity constraints only at boarding arc	No. Penalty at interstation arc	Explicit	Explicit Flow charging upstream
Waiting Time for a route for a boarding passenger	Waiting time increase through reduced frequency	Penalty due to IVTT increase	Additional waiting time added to the journey time	No penalties due to excess volume
Distribution of boarding passenger volumes in a station between the attractive routes for a given destination	Uncontrolled effects due to frequency and capacity reduction at all load rates (BPR function)	Nested model: - Strategy of assignment to attractive routes - MNL between strategies	No interaction with other attractive lines	Not dealing with discrete availability of routes

3.4 Seat Availability and In-vehicle Comfort

This section focuses on the second main capacity constraint represented in transit assignment models: the seat capacity. Although, the first in-vehicle comfort models date back from Last and Leak (1976), in-vehicle comfort modelling had not acquired enough attention until recently. Gendreau (1984) suggested accounting for in-vehicle comfort by applying a discomfort penalty to all on-board passengers, without yet distinguishing seated and standing ones. In the 2000's the scientific community led a significant research effort in order to include the effect of seat availability on quality of service and passenger route choice. Transit assignment models dealing with seat capacity are mainly static models adapted to large-scale transit network modelling. They are heterogeneous, regarding network representations and modelling approaches. They are all based on a formal description of the seat allocation

mechanism, notably assuming that on-board passengers have a seating priority over boarding passengers at a stop. They distinguish the two classes of on-board passengers (seating and standing). Three representative seat availability models are discussed below: the static models of Seated Capacity (Leurent, 2012a), the Failure-to-Sit (Schmocker et al., 2011) and the dynamic Seat Availability model (Sumalee et al., 2009) extended in Hamdouch et al (2011).

3.4.1 The Seated Capacity Model

Interested in the transit characteristics that constitute the quality of service in public transportation Leurent (2012a) and Leurent and Liu (2009) propose a transit assignment model which could capture the fact that upstream passengers have the priority over those boarding to obtain a seat, all this in a framework that can efficiently deal with a large scale network. This priority rule is consistent with empiric observations, which suggest that passengers consider standing to be more penalizing than seating (Kroess et al, 2007), and therefore compete to occupy a seat whenever one becomes available.

The authors model the assignment of passengers to seats as a two-stage mechanism. At the arrival of a vehicle in a station, the sitting passengers who are alighting free their seats, and the on-board standing passengers compete for them. Then, the passengers boarding compete for the remaining ones, if there are any. Formally, at each stop this mechanism yields sitting probabilities for the on-board standing and then for the boarding passengers.

They combine the optimal strategies framework in Spiess and Florian (1989) and the network representation in De Cea and Fernandez (1993) to adopt a network description based on route sections, or service legs; one service leg is created per transit service and boarding-alighting station couple. Since a seated passenger will only stand for alighting, we can define a service mode for each boarding-alighting trip with respect to the station the passenger sat. At any point of the journey there is a given probability to sit. Hence the cost of a boarding-alighting leg is a random variable linked to the probability of having a certain service mode. However, a mean cost of a given service leg can be calculated.

Despite the modifications in the network description and a leg-costing algorithm, Leurent (2012a), the route choice model follows standard transit assignment model architecture. The route section approach leads to an increase in calculation time. However, the development of a cost-flow relationship at the route-level allows dealing with each line separately and, thus,

reduces the total number of legs needed. The network assignment problem relies in finding the transit network traffic equilibrium on the basis of the calculation of the leg costs. That approach can handle with great efficiency a large-scale transit network, such as the transit network of the Paris Metropolitan Region, on which the model was applied in Leurent and Liu (2009).

3.4.2 Failure-to-Sit Model

In Schmocker et al (2011) the authors describe the failure-to-sit model, developed for frequency-based assignment, where on-board congestion is related to seat capacity. Through the fail-to-sit probabilities at each station, the authors wish to explicit the effect of seat availability on route choice, with a method inspired by the Failure-to-board model of Kurauchi et al (2003).

In addition to the fail-to-sit probabilities, the authors propose an alternative network description, where a line is modelled with two parallel arcs; one for standing passengers and one for sitting passengers. Stations consist of 5 nodes and 8 arcs for each line. At the station, simple priority rules are applied to all passengers, where the on-board passengers would have priority over boarding passengers to fill the available seats. The fail-to-sit probabilities are updated on the basis of these flows.

The assignment procedure is based on the hyperpath assignment framework with many adaptations. Once the fail-to-sit probabilities are calculated, the optimal hyperpath is found from each destination to the origins. Simultaneously, the path costs are updated, from downstream to upstream, according to the fail-to-sit probabilities. After constructing the optimal hyperpath, an algorithm determines the route shares at boarding on the basis of the combined frequency of the attractive lines and the fail-to-sit probabilities. The assignment algorithm is based on absorbing Markov chains.

3.4.3 Seat Availability Model

In Sumalee et al (2009), the authors present a dynamic transit assignment model with two levels of discomfort perceived by the passengers based on whether they are sitting or standing. In addition, the passengers consider the stochastic nature of the quality of service, due to the probability of getting a seat. Departure time choice and route choice are based on

the expected travel disutility or generalized cost, and the authors focus specifically on the effect of seat allocation on the temporal distribution of travel flows. Aside from these elements, the generalized cost also depends on the perceived travel time, the waiting and transfer time, the walking access and egress time, the late or early arrival and the total monetary travel cost.

The model is based on the sitting probability for on-board standing and boarding passengers. That depends on the number of available seats, the priority of on-board and boarding passengers and a passenger motivation to get a seat, related to the time already standing and the remaining journey time. The sitting allocation mechanism at a station consists of two stages. The first stage concerns the standing passengers on board a vehicle, after the alighting passengers left the vehicle. The algorithm assigns randomly a passenger to an empty seat based on his willingness to get a seat. This procedure is repeated until all seats are filled. The second stage is addressed to passengers boarding according to their arrival time, following the bulk-queue model. The passengers of a group will be randomly allocated a seat until all seats will be filled.

The network is represented by a diachronic graph, which contains a service, a demand and an access/egress subgraph. According to the authors, the route choice procedure can be assumed to be predefined. Alternatively an MSA algorithm can be applied to solve the fixed-point problem of the probit stochastic user equilibrium defined in Sumalee et al (2009). Once the passenger flows have been loaded on the diachronic graph, an algorithm simulates the run of each vehicle from its origin to its destination in order to compute the passenger flows on board and execute the two stage seat allocation procedure, as described previously. The convergence to equilibrium is monitored through the maximum gap of an auxiliary path flow to its updated path flow.

The seat availability approach is innovative in the sense that it includes a motivation of a standing passenger to get a seat - sensible to the time travelled standing and his egress station – and all than with a standard seat allocation mechanism in a dynamic context. However, the numerical examples are based on simple lines and it seems that the complexity of the approach does not allow its implementation on a large-scale network. Recently, Hamdouch et al (2011) integrated the seat availability approach into a dynamic assignment model, which addresses total and seat vehicle capacity.

3.5 Conclusion

Transit assignment is a recent sector of research, compared to its road counterpart. The particularities (spatial and temporal availability) and the complexity (interactions among physical line, vehicles, passengers and stations) of public transportation led to a combination of unique modelling concepts and generalized modelling techniques. The scientific community is focused on three major sectors of research in transit modelling: modelling capacity effects on static macroscopic and dynamic assignment models, developing more efficient dynamic assignment models for implementation on large-scale networks and focusing on the dynamic simulation of vehicle and passenger movements on the network.

Modelling the capacity effects is of great concern for the realism of transit assignment models, as well as for the socio-economic assessment of transit infrastructure projects; and a plethora of transit assignment models has been proposed, stretching from small to large-scale network implementation and from static to dynamic models. At this stage, only two main capacity effects have been widely addressed: the total person's capacity, affecting route choice and waiting time for a given service or group of services, and seat availability, a determining factor of the quality of service of public transportation networks.

Regarding the models addressing the total person capacity, various approaches were applied, usually inspired from queuing theory. These models generally adopt the classic framework of optimal strategies along with their graph representation as hyperpaths and consider the priority of the passengers, on-board over boarding. The dynamic models of this category generally seem to lack the capability of modelling a large-scale network. Even though there is a great variety of models, they do not seem efficient to realistically reproduce both the route choices and the congestion cost on the network.

Recently, more attention has been given to transit assignment models that deal with seat availability in both static large-scale network and dynamic. We observe that the models representing the phenomenon share a similar loading mechanism, consisting of two stages at boarding, and guaranteeing the priorities of on-board passengers over boarding ones. It should be noted that the model in Sumalee (2009) is unique in its inclusion of a motivation of passengers to get a seat, which depends on the time standing and the egress station.

Three main trends of research are being drawn for the future. First, it is essential to overcome the great complexity of the transit network and develop efficient algorithms that capture the various phenomena, especially for developing dynamic models to simulate large-scale networks. Second, the interaction among the transit system's elements ought to be thoroughly examined to grasp the capacity effects in an entire network. The third main axis for future research lies in including the effect of an intelligent transportation system to the transit assignment models and more precisely of the effect of real-time information constantly available.

PART II: RESTRAINED MODELS

Chapter 4:

Modelling the Line System on a Structural Transit Network

4.1 Introduction

The transit network of a large metropolitan area is frequently subject to heavy congestion during the day. That congestion, not only affects the quality of service experienced by the transit users (for instance in-vehicle crowding), but also it can have an effect in the operation of the transit services. That double effect is captured by the line model, introduced in Leurent et al (2011). The described framework offers an approach to make a connection between passenger traffic assignment models for transit networks and transit operation models.

The line model is established with two main objectives. On the one side, a modular framework is developed for treating the capacity constraints within a transit line, based on a two-layer representation of the transit network. These upper and lower layers correspond to the demand and supply side with a two-way communication. On the upper layer, the demand is assigned to paths based on the local costs calculated by the line model on the lower layer. Respectively, on the lower layer, each line is treated separately and the cost of each trip segment is calculated, on the basis of the upper layer passenger flows and of the local constraints at a stop or on a vehicle.

On the other side, the line model captures certain vehicle and station capacity effects that influence the passengers' route choice, such as passengers' seat availability, total vehicle capacity and station platform occupancy. Each effect is addressed at a local level with a specific model. While the seat capacity and total vehicle capacity, modelled by the seat

availability (Leurent, 2012a) and transit bottleneck (Leurent, 2011b) models respectively, are focused exclusively on the passenger flow, the station platform occupancy – addressed in the restrained frequency model – makes a connection between the passenger flow at the station affecting the dwell time of a vehicle and the operation of the transit services.

This chapter is composed of four sections and a conclusion. Section 2 introduces the line system and sketches the general line model framework. Sections 3 and 4 deal with the in-vehicle comfort and transit bottleneck model. Finally, section 5 describes the restrained frequency model.

Table 3: Basic Notations of Chapter 4

Variable	Definition
N	Set of nodes of the service network, $G = (N, A)$
A	Set of the line legs of service network, $A \equiv A_P \cup A_L$
ℓ	Transit line, $\ell \in L$ the set of lines of the transit network
z	Service route $z \in Z_\ell$
φ_{z0}	Nominal frequency of route z
k_z^R	Total capacity of a vehicle of the route z
$x_{\ell(i,j)}$	Passenger flow from the station couple $i \rightarrow j$ on line ℓ
H	Reference assignment period $H = 1$
η_i	Service frequency reduction factor at station platform i
n_{zi}	Stock of passengers willing to board route z at station i
π_{zi}	Probability of immediate boarding of stock n_{zi} on route z at station i
$g_{\ell(i,j)}$	Generalized cost of the trip $i \rightarrow j$ on line ℓ

4.2 The Line System and the Model Framework

The increase in the realism of the transit assignment models and the integration of various capacity effects and their interactions achieved in the line model results in these models gaining in detail but also in complexity. Our modelling approach is based on a systemic analysis of the transit system. In Leurent (2011) the author identifies the components of a transit system and along with their interactions.

A transit line constitutes the elementary subsystem of a transit network. The operations of the line are autonomous, especially so for lines with exclusive right-of-way. Each direction can be addressed separately and the vehicle operations of different lines rarely interact. Furthermore, the boundaries between transit lines include weak interactions between passengers accessing or transferring to the line. The capacity effects internal to the transit line and the cost of the individual trips of an origin – destination station couple within the line are addressed within this line model.

The concept of the transit line is reduced to a subset of services that cannot overtake one another, especially because they share the infrastructure (station platforms, section tracks). One transit line is defined per direction of the service.

A line of operations is formally defined as a connected, arborescent, acyclic network in a single direction of traffic. The link set includes track links either of interstation run or station sojourn, together with pedestrian links for egress and access at stations. Let us denote by $\ell \in L$ a line and $A^{(\ell)}$ its set of links, which is composed of: $A_I^{(\ell)}$ the subset of interstation track links, $A_S^{(\ell)}$ that of station track links – they both represent the vehicle traffic – and $A_B^{(\ell)}$ the subset of boarding links and $A_A^{(\ell)}$ that of alighting – these sets represent the passenger access and egress. The line is operated by one or several transit services, denoted $z \in Z_\ell$: each service has a given track route i.e. an acyclic path denoted $P_z \subset A^{(\ell)}$ and a node set $N_z \subset N_\ell$ of stations serviced along the route.

The line and service topology of links and nodes is useful to model not only the topology of service legs and line legs, but also the chronological order of traffic operations. Figure 4 depicts the process of operations. In fact there are five serial and related processes hereby listed:

- i. passenger alighting;
- ii. passenger movement within a vehicle;
- iii. passenger waiting on station platform and boarding vehicles with available capacity and servicing their egress station;

- iv. dwell time and track occupancy that determine vehicle operations hence in turn service operations and their frequency during the period of reference and
- v. Interaction with external traffic on interstation links.

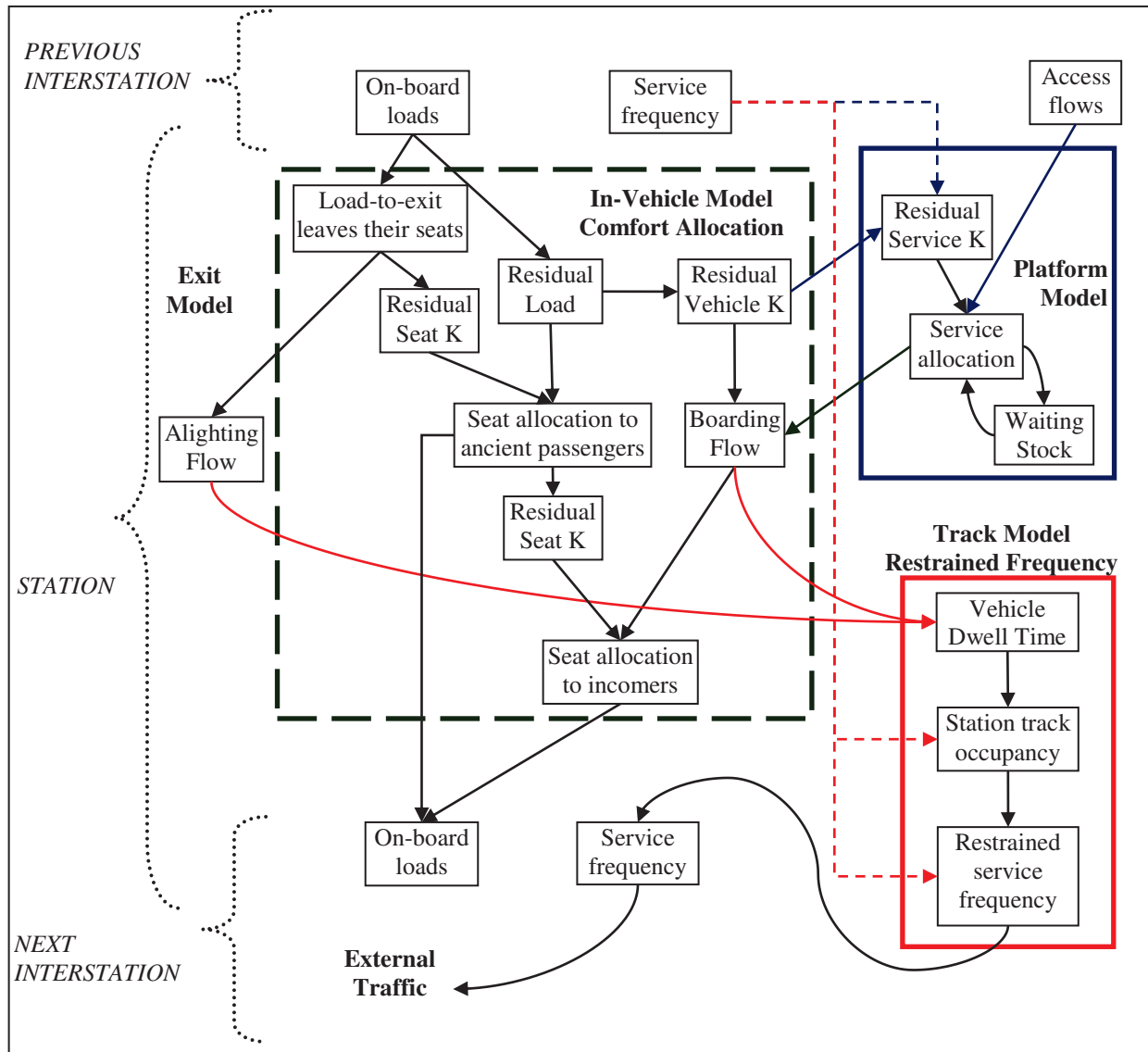


Figure 4: Overview of the line system (K for capacity)

Each process is treated with a specific local model, where the output vector of one model feeds the successor model. Their coordination is achieved by the line model, which calls two main models at that level (see chapter 8 for detailed algorithms):

- a physical model of flow loading in vehicles and of service traffic treating the line in the topological order of the stations, and

- an economic model of cost evaluation in the setting of the individual passenger that would use the line on a given leg.

The line model manipulates mainly two variables. First, the passenger load y_{za}^{sr} in a vehicle of service z along link a with egress station s and suffering a comfort state r . Second, the frequency ϕ_{za} of service z on link $a \in P_z$. The latter identifies the number of vehicles that operate on service during the assignment period.

It would be possible to model one sub-network link by passenger state hence by link a , service z , comfort state r and egress station s . However, it is sufficient to model the service topology in relation to the line topology on the basis of the sets P_z and N_z , and to identify the relevant flow state variables y_{za}^{sr} and ϕ_{za} .

The following sections focus on some particular local models, applied at each station: the in-vehicle comfort for the passenger traffic within the vehicle; the transit bottleneck for the passenger waiting on station platform and boarding; and the restrained frequency for track temporal occupancy and demand-service interplay.

The line model manipulates an additional set of local variables. In order to model seat capacity (see section 4.3), two comfort states of sitting versus standing are identified by index r : incoming passengers in service z at station link a have a probability p_{zas}^{+r} of getting comfort state r with respect to their destination station s . On the previous track link, on-board standees that do not exit at the station have a probability $p_{zas}^{o(\rho)r}$ of getting state r from previous state ρ . The bottleneck created from the passengers at boarding is expressed by the probability of immediate boarding (PIB), π_{zis} , for a vehicle of service z arriving at station i for passengers destined to station s (see section 4.4). The three probability vectors are determined within the line flow model together with the flows y_{za}^{sr} and the frequencies ϕ_{za} .

Considering the equiprobability to get a seat among boarding passengers, whatever their egress station, the previous probability vectors can be expressed with respect to the boarding station, regardless of their egress station. All these variables are taken as exogenous in the line cost model. The vectors of the probabilities of getting a comfort state $\mathbf{p}_z^{+r} = [p_{za}^{+r} : a \in P_z]$ and

$\mathbf{p}_z^{o(\rho)r} = [p_{za}^{o(\rho)r} : a \in P_z]$ determine the in-vehicle travel conditions for a given service leg $z(i, s)$. The vector of probabilities of immediate boarding $\boldsymbol{\pi}_z = [\pi_{zi} : i \in N_z]$ in relation with the vector of service frequencies $[\varphi_{za}]_{a \in P_z}$ determine the boarding share of each service $z : i \in N_z$.

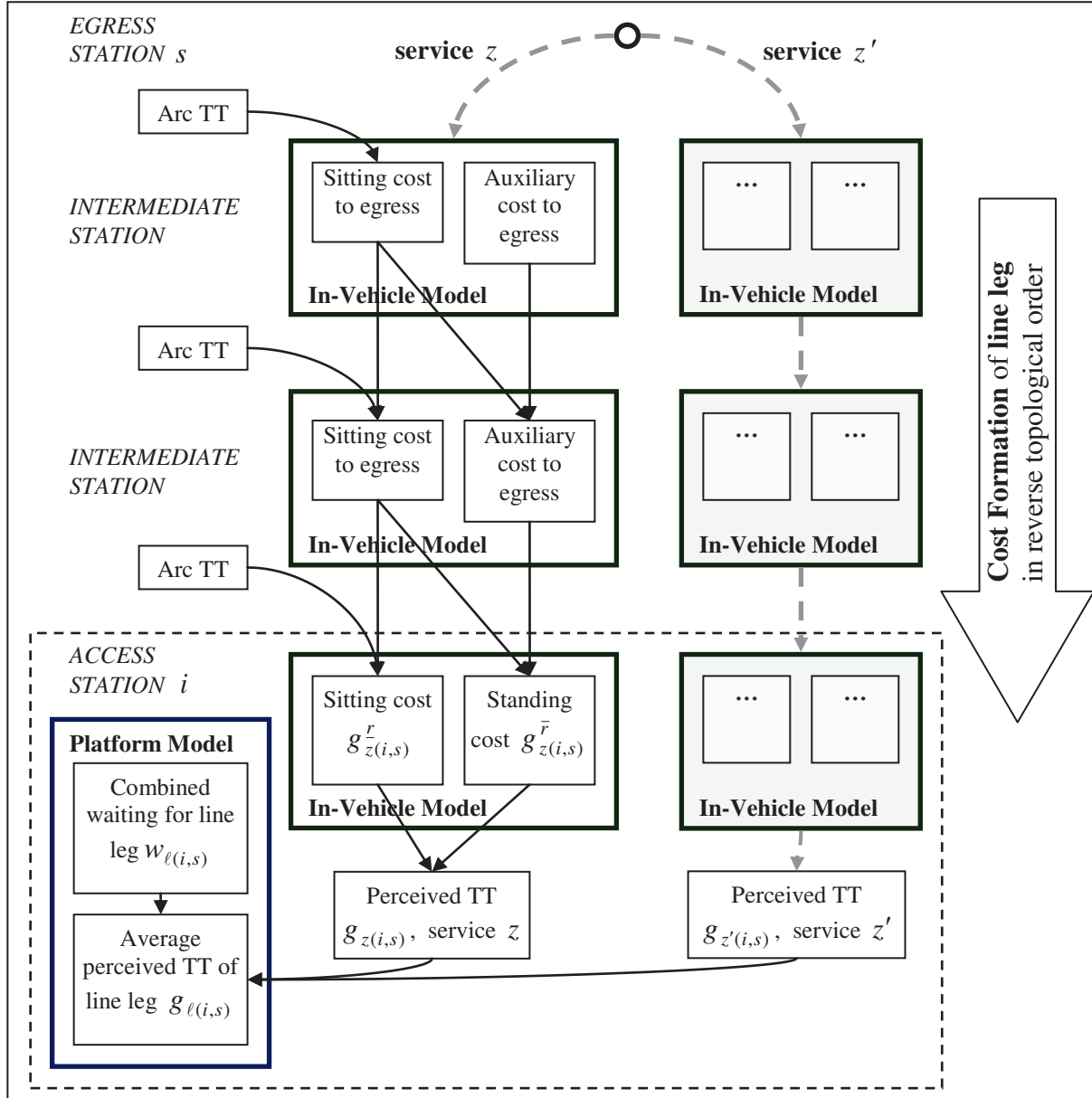


Figure 5: The process of cost evaluation of a line leg (i, s) – TT for Travel Time

Figure 5 illustrates the process of cost formation of a trip on access-egress station couple (i, s) by two alternative services, z and z' , that makes reference to the line leg $\ell(i, s)$. The

arrows denote the formation of cost the line leg $\ell(i, s)$. That is done in a reverse topological order, from egress station s to access station i by means of three successive processes.

From the egress station, a recursive formula for the formation of the service leg cost $g_{z(i,s)}$ is used (see sections 4.3 and 8.3.3.2), based on the sitting cost $[t_{za}]_{a \in P_z}$, the standing cost $[g_{za}^r]_{a \in P_z}$ and the calculated auxiliary cost $[\hat{\gamma}_{z(j,s)}]_{j,s \in N_z, j < s}$ – subject to the probabilities of improving the comfort state downstream. The average in-vehicle travel time $\tilde{t}_{\ell(i,s)}$ and average generalized in-vehicle travel time $\tilde{g}_{\ell(i,s)}$ are calculated as a weighted average of the alternative services – given the probabilities of immediate boarding $[\pi_{zi}]_{i \in N_z}$ and the service frequencies $[\varphi_{za}]_{a \in P_z}$ of the services at i . The previous state variables determine the combined waiting time, $w_{\ell(i,s)}$, and the average generalized travel time, $g_{\ell(i,s)}$, for a trip on the line leg $\ell(i, s)$.

4.3 The In-Vehicle Comfort

The in-vehicle travel time is a major component of the journey time and its quality is greatly affected by the passenger volumes. The line model uses a local comfort model to assess the in-vehicle comfort of an interstation at the station after boarding. The in-vehicle comfort of a passenger depends on whether he occupies a seat or not and on the density of the standing passengers if he is standing. The in-vehicle comfort model stems from the model for seat capacity introduced in Leurent (2006) and developed in Leurent (2012a). Some modifications are made to adapt it to the CapTA model and to make the standing comfort state sensible to the density of standing passengers. These are detailed in the first section. The second section discusses the impact of the folding seats on the in-vehicle comfort, such as discussed in Chandakas (2009).

4.3.1 The Model for Seat Capacity

The Seated Capacity model, developed in Leurent (2012a), treats the in-vehicle comfort by addressing the seat availability during a trip. Two comfort states are defined, sitting and standing, where the latter is considered more onerous than the former. That agrees with the

founding of stated preference surveys, such as in Kroess (2007) for the Paris Metropolitan Region. A rational economic agent seeks to minimize his perceived journey time and therefore prefers sitting rather than standing. He will quit the sitting comfort state at the alighting station.

When a vehicle stops at a station i , the alighting passengers free a number of seats, which then become available to the standing passengers. Two priority rules are established; first, on-board passengers have a physical priority over boarding passengers, second, the passengers of the same class have equal probability to succeed in sitting, p_{za}^{or} and p_{za}^{+r} for on-board and boarding passengers respectively.

At each level of the competition, let k_{za}^{r} denote the available seat capacity and y_{za}^{sr} the vehicle flow of incoming passengers at link a , of service z , of each competition stage with destination s and comfort state r . The candidate passengers with comfort state r are $y_{za}^r = \sum_{s>i} y_{za}^{sr}$. The symbols o , $+$ state the competition stages, for on-board and boarding passengers respectively. The probability to occupy a seat is:

$$p_{za}^{(o/+)r} = \min\left\{1; \frac{k_{za}^{r}}{y_{za}^r}\right\}, \text{ or } p_{za}^{(o/+)r} = 1 \text{ when } y_{za}^r = 0 \quad (9)$$

The seat competition is a random procedure. A seated passenger will stand only to alight. We can therefore define a service mode: a given trip is identified by the boarding and alighting stations and the station of comfort state change to sitting. For a passenger boarding at a station i , the average comfort state corresponds to the weighted average of all the service modes applicable: its cost will be the weighted average of the cost of the service modes.

4.3.2 Interaction Between the Seat Capacity and Line Models

The solution algorithm of the Seat Capacity Model described in Leurent (2012a) holds many similarities with the architecture of the line model and its integration is straightforward. The two stage treatment of the seat capacity model corresponds to the two main models at the level of the line. The passenger flowing is integrated in the line flowing model of the line model. Indeed, by considering the chronological order of the operations for every service at a stop, the alighting and the updated in-vehicle conditions are followed by the calculation of the

on-board sitting probability, while the boarding sitting probability is calculated after the platform waiting and boarding local model. Respectively, the service leg cost evaluation is performed at each station and for every stopping station throughout the line leg cost evaluation model. However, the line model focuses on the individual travel conditions and the passenger flows correspond to the vehicle flow vector. The integration of the in-vehicle comfort model to CapTA is further described in the algorithmic section of Chapter 8.

Furthermore, an extension of the model for in-vehicle comfort is the addition of the average density of standing passengers. In that way, a linear cost may be imputed to the standing passengers, according to the standing conditions. The line model acts as a central module on the simulation procedure, which coordinates the input and output of the various local models. In that case, the Seat Capacity Model affects the perceived in-vehicle travel time of the services. In addition, the local models at the station interact via the line model and produce various secondary effects, which are discussed in section 4.6.

4.3.3 The Effect of Folding Seats on the In-Vehicle Comfort

While a seat capacity model describes the occupation process of the available seats, for simplicity it omits the use of folding seats, due to their complex operating characteristics. Indeed, these seats are used at medium vehicle loads – when the load just exceeds the nominal seat capacity. At unfolded position, they occupy a part of the vehicle surface used for standing. They are still used at low standing densities and they cease to be used when the standing densities increase over a critical value. The folding seats and their temporary usage are discussed in Chandakas (2009) as a means to accurately provide a measure of the variations of the in-vehicle comfort in the commuter rail of the Greater Paris transit network.

Their modelling is essential for the estimation of the quality of service, since they add roughly 80-120% and 0-35% of seat capacity to the metro and suburban rail vehicles respectively. Furthermore, the use of folding seats at the sides of the door may influence the passenger exchange rate. If they are used, the available door width and the number of flow streams may be reduced at certain vehicle seat arrangements. It has therefore an impact on the nominal exchange capacity of a door and leads to an increased dwell time. In the following section we make a description of their operation and the modifications which concern the in-vehicle comfort.

4.3.3.1 The Attributes of the Folding Seats

The folding seats are seats usually positioned near the doors of the vehicles or other surfaces that are assigned as potential surface for passengers to stand at high congestion levels. When a folding seat is used, it covers a given vehicle surface, which is inaccessible for standing passengers, therefore reducing the total surface available for standing passengers. When folding seats are occupied, a smaller surface is available for the standing passengers, who stand denser. The standing density d_i is defined as the ratio of the number of standing passengers to the surface available for standing at station i .

Moreover, whether the folding seats are occupied or not is related to the passenger load of the vehicle. It is assumed that by civil obligation the sitting passengers liberate the folding seats when the vehicle is congested. Moreover, the penalty for occupying a folding seat is greater than that of a normal seat, due to the risk of standing later at the trip. The state transition of the vehicle between active or inactive folding seats takes place at a certain standing passenger density, defined as a critical load. Assuming a different critical density for occupying or liberating a folding seat; a higher density is needed for liberating a seat rather than occupying it – if unused.

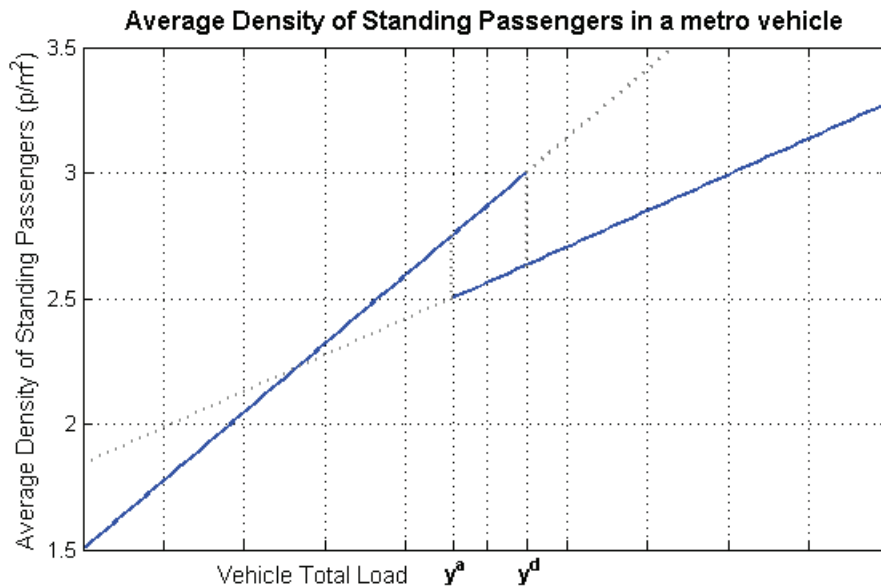


Figure 6: Density of standing passengers on a vehicle in relation to the volume of on-board passengers and the effect of the folding seats

Figure 6 illustrates the density of the standing passengers and the vehicle load for a standard vehicle of the Parisian metro. We observe that standing density increases after the occupation of all normal or folding seats. Furthermore, when the load increases, there is a standing density, d^d , for which the passengers seated in the folding seats liberate them and join the standing passengers. The available space for standing is increased. Therefore, a discontinuity in standing density is provoked, associated with a critical load for deactivation, y^d .

On the opposite direction, when the vehicle load decreases, there is a density for which the standing passengers decide to activate the folding seats and occupy them. We assume a density for activation $d^a \leq d^d$ and a critical load for activation, y^a . At activation, the remaining standing passengers will suffer from a reduced available surface, whereas those sitting at the folding seats will perceive inferior cost than standing.

4.3.3.2 In-Vehicle Comfort States

Previously, the physical characteristics of the folding seats and their activation – deactivation mechanism are described according to the density of the standing passengers. A passenger who is seated on a folding seat is not fully isolated like the other sitting passengers – suffering the passenger congestion – and may face a temporary occupation subject to the density of the standing passengers. He should perceive, then, an increased journey discomfort compared to one occupying an isolated seat. that makes sitting on a folding seat a distinct comfort state which is positioned in terms of travel penalty between the sitting and the standing states.

To summarize, by including the folding seat as a distinct comfort state, the in-vehicle comfort is characterized by three states, with their respective section costs and comfort state transitions as follows:

- **Sitting in normal seat** with a discomfort cost g_a^r for a passenger sitting on route section a . The passenger gives up his state only at alighting;
- **Sitting in folding seat** yields a discomfort cost $g_a^{\tilde{r}}$ for a passenger. A sitting passenger competes equally with the standing passengers for the transition to the sitting comfort state. High vehicle loads may neutralize folding seats;

- **Standing** with a discomfort cost $g_a^{\bar{r}}$. That cost may depend on the density of standing passengers $g_a^{\bar{r}} = G_a(d_a)$. A standing passenger tries a transition to sitting comfort state and if he fails, to sitting in a folding seat comfort state.

It stands for the comfort state costs: $g_a^{\bar{r}} > g_a^{\tilde{r}} > g_a^r$

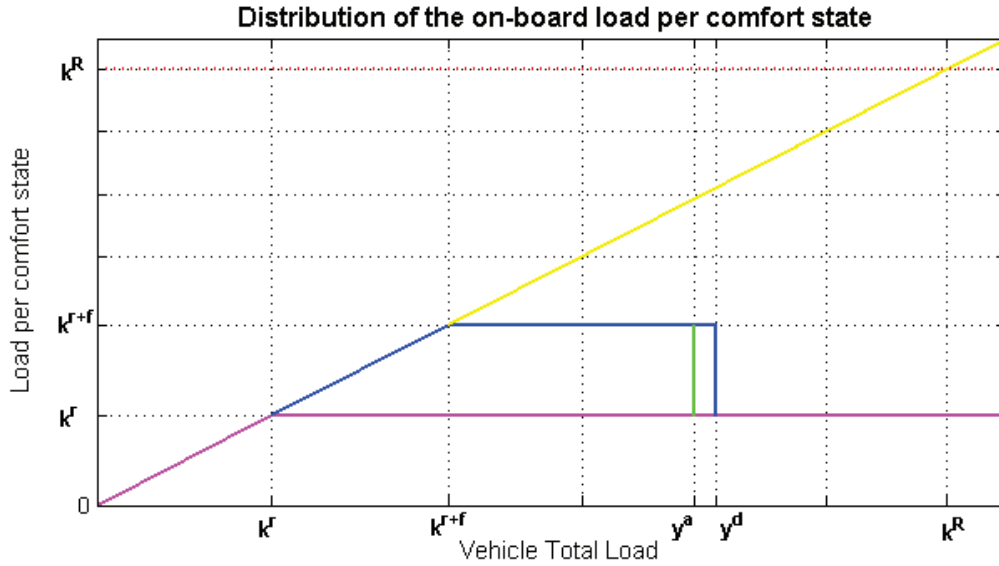


Figure 7: Cumulative distribution of in-vehicle Comfort States according to the vehicle volume

The composition of the in-vehicle passenger volume by comfort state is illustrated by the cumulative passenger flows of a typical vehicle of the Paris metro in Figure 7. The magenta defines the number of passengers sitting on a route section and the yellow, those standing. Therefore, from volume 0 to k^r (horizontal axis) any passenger boarding will get a seat. We observe that for passenger volumes from k^r to y^d (blue line on one direction) and y^a (green line on the opposite direction) there exists a folding seat comfort state with a smaller discomfort than the standing passengers. The red line corresponds to the nominal crush capacity of the vehicle, k^R . For example, a vehicle containing k^{r+f} passengers (horizontal axis) will be composed of k^r passengers with comfort state \underline{r} and $k^{r+f} - k^r$ passengers with comfort state \tilde{r} .

4.4 The Total Passenger Capacity of the Vehicles

When a passenger demand faces insufficient vehicle total capacity, the failure to treat the demand leads to an increased waiting time for the passengers on the platform. Using queuing theory approach, a local routing model with capacity constraint is introduced in Leurent (2011b). The transit bottleneck model deals with the vehicle's passenger capacity by establishing a relation between the exogenous passenger flow, the waiting time and the local route choice. That is realized via the endogenous variables of passenger stock per destination at the platform and the vehicle's residual capacity after alighting.

The transit bottleneck is thoroughly described in Leurent and Chandakas (2012). The brief theoretical description is followed by an application instance, in order to demonstrate the behaviour of the model. A three line and three station simplified network is used to show the various traffic states that emerge, the destination competition when capacity is insufficient and the congestion externalities.

4.4.1 The Transit Bottleneck Model

The transit bottleneck model seeks to capture the effect on waiting time and route choice of the strict passenger capacity constraint. We assume that when the available capacity of a vehicle at a station is sufficient for handling the passenger demand, the model behaves identically to the non-saturated case and produces the same costs. Nonetheless, when the demand exceeds the available capacity of a vehicle, a stock of passengers is formed at the platform. That stock cannot be dissipated over the reference period H .

The objective of the transit bottleneck model is to calculate the actual waiting time when saturation occurs along with the boarding shares for each transit service. Therefore, as input the exogenous flow $x_{\ell(i,s)}$ for each in-vehicle trip boarding at i and alighting at s , the actual frequency of the transit service φ_{zi} and the available capacity of a vehicle, after passengers have alighted at the station, k_{zi}^R are used. The boarding characteristics are estimated through the calculation of the passenger stock waiting at a station platform for a vehicle of the attractive set of routes and its probability to board that vehicle.

A partial stock of passengers, σ_{is} , is identified for the passengers at station i who are destined to s . The stock of passengers who wish to board a vehicle of the service $z \in Z_i$ will be the sum of the partial stocks of the destinations served by z , $n_{zi} = \sum_{s \in N_z} \sigma_{is}$. If the available capacity of the vehicle is not sufficient to accommodate the stock of passengers waiting, then only a fraction of the stock will achieve to board the vehicle. The probability of the stock of passengers waiting for a service z , to board a vehicle of that service will be:

$$\pi_{zi} \equiv \min\left\{1; \frac{k_{zi}^R}{n_{zi}}\right\} \text{ or } \pi_{zi} = 1 \text{ if } n_{zi} = 0 \quad (10)$$

Leurent and Chandakas (2012b) demonstrate that the partial stock of passengers, queuing in a mingled discipline, for each destination $\sigma = [\sigma_{is} : s \in N^{(\ell)}]$ satisfies a Fixed Point Problem (FPP). That partial stock must satisfy the following equation.

$$\frac{2 \cdot \sigma_{is}}{x_{\ell(i,s)} \cdot H^2} = \frac{x_{\ell(i,s)}}{\sigma_{is}} - (\varphi\pi)_{is} \quad (11)$$

Where $(\varphi\pi)_{is} = \sum_{z: s \in N_z} \varphi_{zi} \pi_{zi}$

The partial stock should then be the solution of the equation:

$$\tilde{F}(\sigma_{is}) = \frac{2 \cdot \sigma_{is}}{x_{\ell(i,s)} \cdot H^2} - \frac{x_{\ell(i,s)}}{\sigma_{is}} + (\varphi\pi)_{is} = 0 \quad (12)$$

When multiple destinations are served from a station i the problem cannot be solved analytically. A Newton - Raphson Algorithm is used for solving the problem by coordinating all the destinations from that station. The Fixed Point Problem is shown to have a solution, which is unique.

Let q_{is}^- be the discharge flow. When saturation occurs, the exogenous flow entering the queue exceeds the discharge flow, $x_{\ell(i,s)} > q_{is}^- = \sum_{z: s \in N_z} \varphi_{zi} \pi_{zi} \sigma_{is}$, since not all the passengers can board immediately a vehicle of the transit service. They have to experience a prolonged waiting in a stock. A longer than the reference period is needed for the evacuation of that flow, noted H_{is} . Since the flow is conserved, it holds $x_{\ell(i,s)} \cdot H = q_{is}^- H_{is}$ and we can define the exit time interval H_{is} :

$$H_{is} = \frac{x_{\ell(i,s)} \cdot H}{q_{is}^-} \quad (13)$$

Therefore the actual waiting time becomes:

$$w_{\ell(i,s)} = \frac{\sigma_{is} \cdot H_{is}}{x_{\ell(i,s)} \cdot H} = \frac{\sigma_{is}}{q_{is}^-} = \frac{1}{(\varphi\pi)_{is}} \quad (14)$$

Since $0 < \pi_{zi} \leq 1$, it stands for the perceived frequency at boarding that $(\varphi\pi)_{is} \leq \sum_{z:i,j \in N_z} \varphi_z$.

4.4.2 An Application Instance and the Transit Bottleneck Model Behaviour

A classroom instance is used to demonstrate the behaviour of the transit bottleneck model. The example is similar to that developed in Leurent and Askoura (2010) for the comparison of various assignment models under capacity constraints. Three parallel lines connect an origin A with a destination D. The two (ML2, ML3) also serve node B, which can serve both for an origin and a destination, as shown in Figure 8. The following table resumes their operational characteristics.

Table 4: Operational characteristics of the service routes

Route	Frequency (veh/h)	Vehicle Capacity (pass/veh)	Service Capacity (pass/h)
ML1	6	50	300
ML2	12	100	1200
ML3	6	150	900

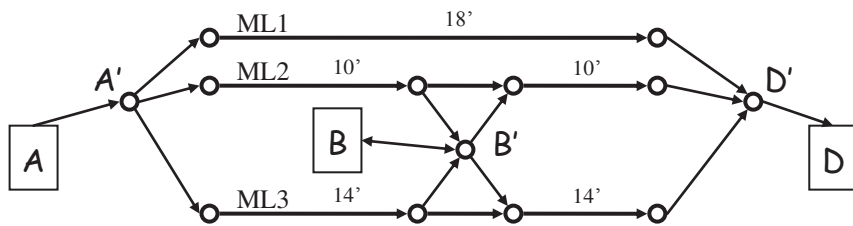


Figure 8: The service network of the application instance

4.4.2.1 Passenger Traffic Flow States

The model presented calculates the expected travel time of an access – egress station couple in the basis of the passenger stock (for the waiting time) and the weighted average of the flow

assigned to the transit services under strict capacity constraints. Therefore, when looking at the combined effect on a line with at least two parallel routes, we can distinguish three flow states, depending on the arriving passenger flow. To demonstrate these states, we focus on the simpler BD OD pair. While, we maintain the flow $x_{AD} = 0$, we increase the flow x_{BD} , in order to saturate first the route ML2 and then the route ML3. Figure 9 (a) and (b) illustrate these three traffic flows of the pair BD, relating the in-vehicle time (t_{BD}), in a dashed line and the journey time ($tt_{BD} = t_{BD} + w_{BD}$), in continuous line, with the passenger stock waiting at station B', σ_{BD} , and the arriving passenger flow x_{BD} .

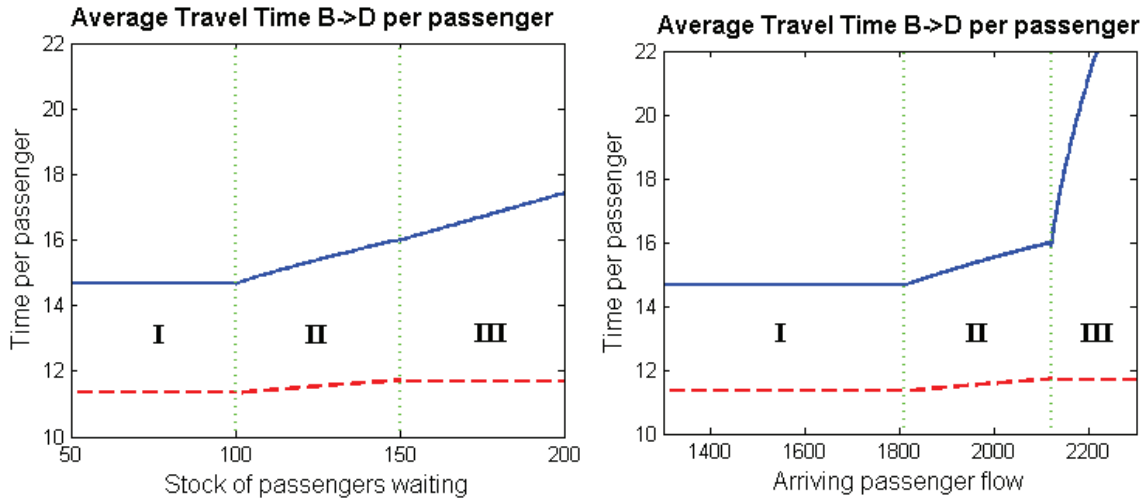


Figure 9: Average Expected Travel Time (blue) and In-Vehicle Travel Time (red) per (a) stock of passengers waiting – left - and (b) exogenous passenger flow – right.

The three flow states can be distinguished on the line, as indicated in Figure 9 and Figure 10:

- (I) The uncongested state, where a weak passenger flow does not face any capacity constraints;
- (II) The semi-congested state, when at least one of the services is saturated, while others have some available capacity to accommodate the additional flow. Even though, the access – egress flow $x_{\ell(i,s)}$ can be transferred within the reference period H , there is an impact in both the expected waiting time and the in-vehicle travel time;

(III) The congested state, (all the services are saturated) where the access – egress flow $x_{\ell(i,s)}$ cannot be served within the reference period, H . The insufficient capacity results in a significant increase of the expected waiting time for the access – egress couple.

Figure 10 illustrates the relation between the arriving passenger flow and the passenger stock. We observe that the congested state begins in a slightly bigger traffic flow than the combined residual capacity of the routes ($\sum_z \kappa_z = 2100 \text{ pass/h}$). That is attributed to the objective function developed due to some mathematical approximation of the queue. Nonetheless, the difference observed is insignificant for the calculation of the expected waiting time and the flow share of the transit services.

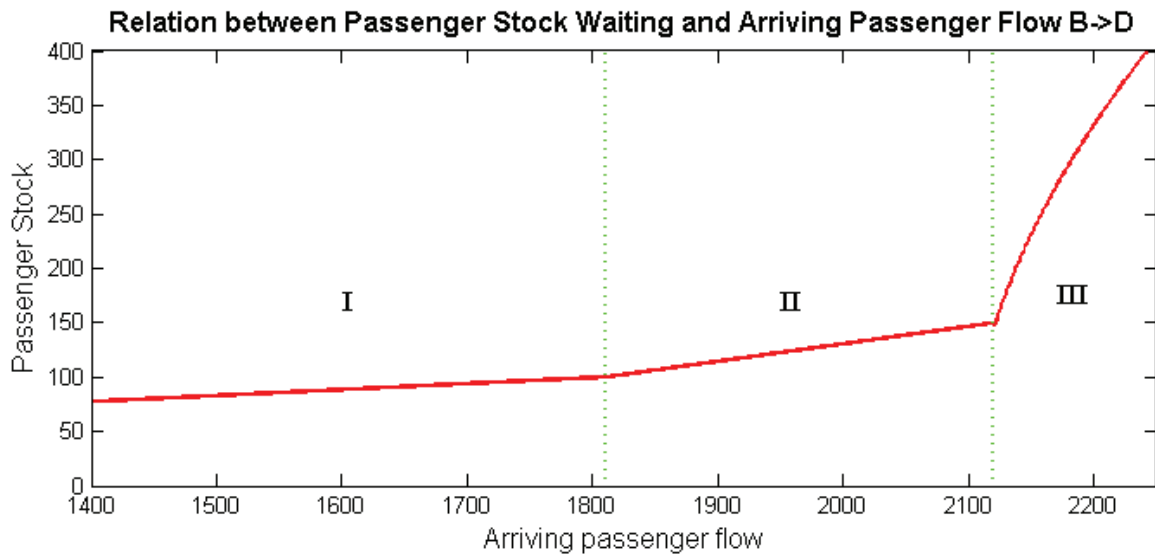


Figure 10: Relation between Passenger Stock waiting to board and arriving passenger flow x_{BD}

4.4.2.2 Destination Competition

When multiple access-egress flows share the same transit service, a competition occurs among their respective partial passenger stocks at boarding. At unsaturated conditions, the route shares correspond to the relative proportions of the service frequencies. When at least one of the routes is saturated, the route shares of the passenger flows are modified, due to competition of the partial passenger stocks per egress station. The behaviour of the transit bottleneck model at congested state is demonstrated using the three-route network described previously. In this example, we keep a constant exogenous passenger flow of the A->D access

– egress station couple at a low passenger flow ($x_{AD} = 800$) and increase the A->B passenger flow, x_{AB} , until all routes are saturated.

Figure 11 illustrates the share of each transit route for the A->D flow. At unsaturated conditions the routes ML1, ML2, ML3 are assigned the 25%, 50%, 25% of the A->D flow, x_{AD} , respectively, according to the route frequencies. Nevertheless, once a transit route is at capacity, an increase in the waiting time and a change in the route shares are observed. For the example developed, ML2 is the first to be saturated for $x_{AB} = 1209$ passengers to destination B. Beyond that flow, the route share of the x_{AD} flow is reduced on ML2 ($\rho_{AD}^{ML2} = 0,40$ for $x_{AB} = 1558$). The saturation of ML3 creates a competition between the B and D partial stocks at A for the transit routes ML2 and ML3. Therefore, ML1 receives the flow unable to board to the other routes and its flow proportion increases from $\rho_{AD}^{ML1} = 0,30$ (for $x_{AB} = 1558$) to $\rho_{AD}^{ML1} = 0,377$ (for $x_{AB} = 1638$). When all services are at capacity, the boarding competition reduces the probability of the stock to board ML2 and ML3 and therefore the proportion of x_{AD} passing from ML1 increases to $\rho_{AD}^{ML1} = 0,441$ for $x_{AB} = 2500$.

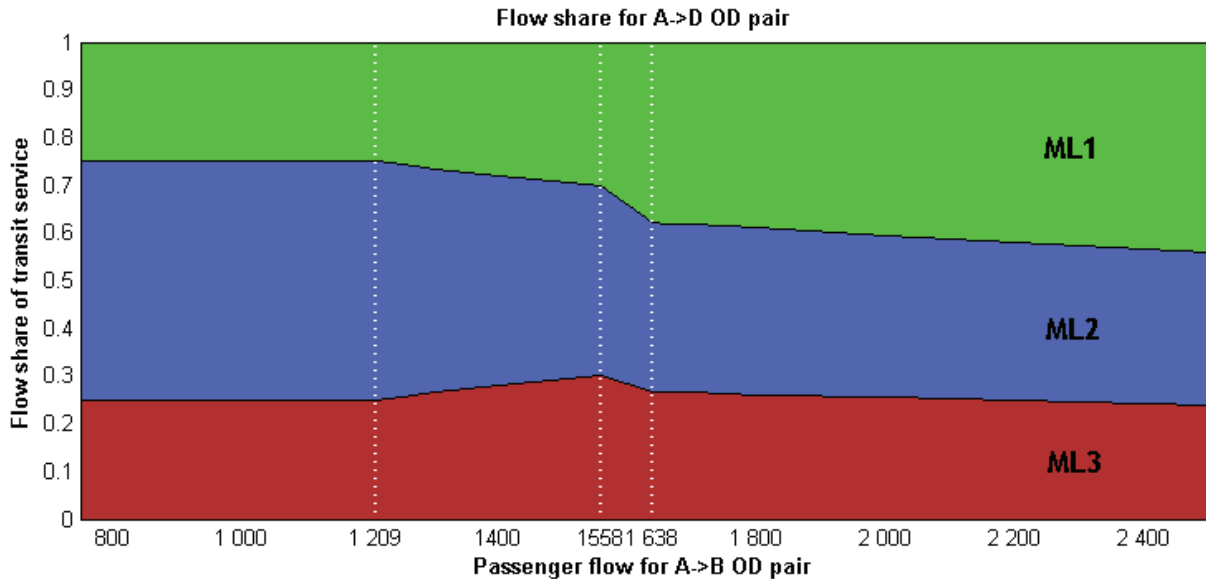


Figure 11: The cumulative passenger flow share of each transit service

4.4.2.3 Congestion Externalities

A passenger using a transit line will induce an extra travel cost to the line and the other passengers, because of the change in the conditions of their trip. The additional cost of a single arriving passenger will then depend on the level of congestion. We note $G = G_{AB} + G_{AD}$ the total cost on the line, where $G_{AB} = g_{AB} \cdot x_{AB}$ and $G_{AD} = g_{AD} \cdot x_{AD}$ are the total generalized costs for each OD pair and g_{AB} , g_{AD} are the average generalized times of each access – egress station couple. The effect of an additional passenger on the pair AB on the line's cost will be given by:

$$\frac{\partial G}{\partial x_{AB}} = \frac{\partial G_{AB}}{\partial x_{AB}} + \frac{\partial G_{AD}}{\partial x_{AB}}$$

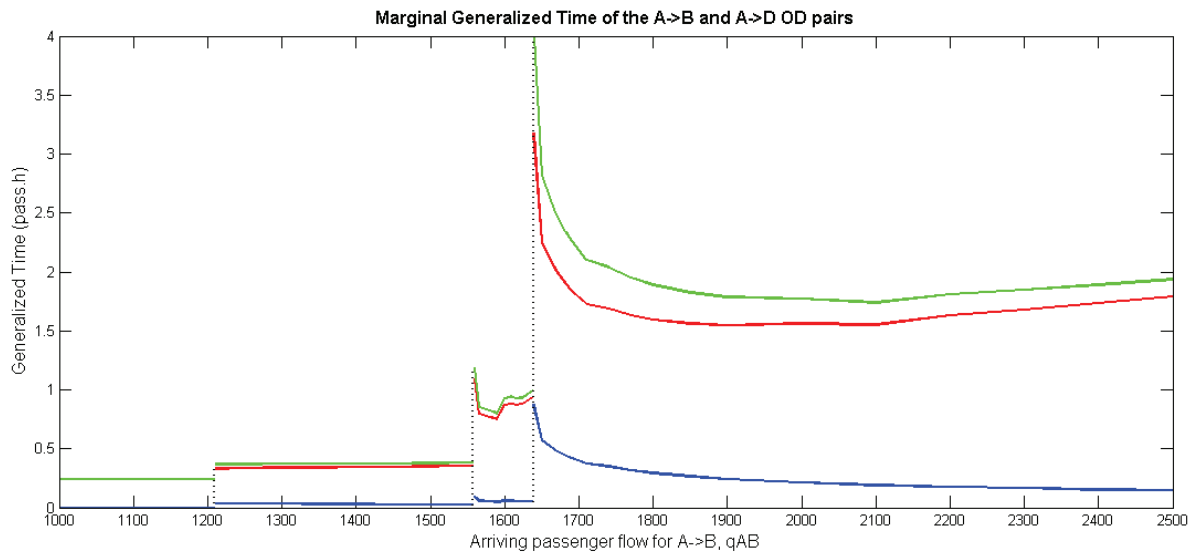


Figure 12: The marginal generalized time of an additional passenger in the AB pair for the AB (red), AD (blue) pairs and the marginal total generalized time (green)

The marginal generalized time for each OD pair and that of the line is illustrated in Figure 12. The red line corresponds to that of the AB pair, $\partial G_{AB} / \partial x_{AB}$, the blue line to the AD pair, $\partial G_{AD} / \partial x_{AB}$, while the green is the marginal generalized total time on the line $\partial G / \partial x_{AB}$. We observe that an increase in the AB flow, x_{AB} , has also an impact to the cost of the AD pair, since they share the same transit services. In addition, discontinuities occur on the graph,

where there is a change in the traffic flow state, since the cost function is non-differentiable (see Figure 9(b)) at these points.

4.5 The Station Platform Occupancy

In addition to the in-vehicle discomfort, crowding and unreliability, felt by the passengers on a high duty line, occasionally the service suffers from operational disfunctionalities provoked by increased dwell time, vehicle bunching and delays, leading to reductions in the service frequency.

That effect is well known to the transit engineers and the scientific community and thoroughly discussed in the Transit Capacity and Quality of Service Manual (TRB, 2003). Lin and Wilson (1992) consider the vehicle's stop time at the station critical for determining the system performance and the quality of service. They identify three direct effects: the dwell time directly affects the vehicle's cycle time; at the station level a stopped vehicle occupies the station line for the dwelling, obstructing the following vehicles; and the dwelling time is believed to be a major factor for travel time variability and vehicle bunching.

Lam et al. (1999) propose a model based on the first effect, relating the fleet size and the cycle time by vehicle to the service frequency. Meschini et al (2007) adopt a similar approach in the setting of a dynamic frequency-based multimodal network. In these models the station dwell time and the section running time of the buses are made flow dependent. However, in the planning horizon the fleet size is easily adjustable and the scarce resource pertains to the second effect linked to the platform availability. Indeed, Harris (2005) and Harris and Anderson (2007) consider the station stopping time and the occupation of the station platform as the crucial factor for determining the performance and the capacity of high duty guided lines (metro and commuter rail).

Although a flow-dependent service operation seems more suitable to macroscopic dynamic assignment models, those proposed so far (e.g. Sumalee et al, 2009) take the service timetable as exogenous, in contradiction with the unreliability that appears under severe congestion. On the other hand, service operations are planned at the line level on the basis of a vehicle running along the service trajectory. The dwelling time and the vehicle load are considered to

influence significantly the transit service (e.g. Vuchic, 2005 and Lai et al, 2011), although a direct connection with the passenger flows is not established.

The restrained frequency model addresses the effect of the dwell time on the transit service performance, thus making the system's performance flow dependent. However, it explains only a part of the system performance, since it is not extended to other aspects of service operation. Aspects, such as rail convergence management and service irregularity are more complex and are not directly imputed to the level of passenger flow. They are not considered in this model.

4.5.1 The Restrained Frequency Model

The Restrained Frequency model tackles the effects of the vehicles' occupation constraints on a station platform. The service frequency is related to the line capacity, defined in the Transit Capacity and Quality of Service Manual (TRB, 2003), where line “*capacity is the maximum number of vehicles (buses, trains, vessels, etc.) that can pass a given location during a given period*”. Therefore, the service frequency designates the number of vehicles of the service that pass a given location during a given period and it is directly confronted with capacity.

At a given section, a passing vehicle blocks the line for a period determined by the vehicle's occupation of the line and a separation time, before the next passing vehicle. The temporal occupation of the line section is defined by the sum of the blocking period of each vehicle. The bigger constraint of the temporal occupation is located at the stops. For a stopping vehicle of service z , the physical occupation of the station platform i depends on the boarding, b_{zi} , and alighting volumes, e_{zi} . The occupation period of the platform is linked to the vehicle's dwelling time $T_{zi}(b_{zi}, e_{zi})$. The separation time – or roll out-roll in time (RORI) – is the period from one vehicle clearing the platform to the following arrival. The shortest time interval at which a vehicle can succeed the previous in a scheduled running without impeding each other is defined as the safe separation time, noted as ω_{zi} (Wendler, 2007). That time includes an operating margin for offering some schedule flexibility, in the context of the planned service to be able to absorb some minor disturbances along the line.

The minimum headway time at a station – the time between two successive vehicle departures from the station – is then composed of the dwell time and the safe separation time:

$$HW_{zi} = T_{zi} + \omega_{zi} \quad (15)$$

The modelling approach is based on guaranteeing that the temporal occupation of a station platform during a given period H doesn't exceed that period. The temporal occupation of a station platform, H' , is the sum of the minimum headways of all the vehicles using the infrastructure of that platform – whether stopping or not – grouped into transit services with similar characteristics. It is defined by the service frequencies and the operation periods.

In a static regime all the vehicles of a transit service have the same characteristics and an average dwell time. Therefore, if φ_{zi} is the frequency of a service $z \in Z_i$ using the platform infrastructure of station i the temporal occupation of the station platform by all the transit services, is:

$$H' = \sum_{z \in i} (T_{zi} + \omega_{zi}) \cdot \varphi_{zi}$$

The restrained frequency model guarantees that the temporal occupation of the station platform H' does not exceed the temporal constraint, the reference period H , such that:

$$H' = \sum_{z \in i} (T_{zi} + \omega_{zi}) \cdot \varphi_{zi} \leq H \quad (16)$$

Enforcing the constraint means physically that if the temporal occupation H' exceeds the reference period, then during the reference period, less vehicles will occupy the station than scheduled. That can be mathematically accomplished by multiplying the left-hand side by a reduction factor $\eta_i \leq 1$. To distinguish the different situations, let us assume φ_{zi}^- the frequency at the arrival of a vehicle of the service z , before the passenger flowing phase at the station i . Then, we impose the reduction factor and φ_{zi}^+ the frequency of the service downstream at the departure. Therefore,

$$\varphi_{zi}^- \equiv \eta_i \cdot \varphi_{zi}^+ \quad (17)$$

Therefore, by definition, we consider:

$$\eta_i = \frac{H}{\sum_{z \in i} (T_{zi} + \omega_{zi}) \cdot \varphi_{zi}^+} \quad (18)$$

The definition of the station occupancy period, H' , includes the vehicles of all the missions that use the station infrastructure, both the stopping and the direct services – for the latter, $T_{zi} = 0$, but $HW_{zi} = \omega_{zi} > 0$.

4.5.2 A Numerical Example

A numerical example is used for the evaluation of the behaviour of a transit line upon enforcing the occupation constraints at a station, according to the previous formulation. Four mass rapid transit lines of the Paris Transit Network are tested: the metro lines, M1, M13 and M14 and the central trunk of the RER A line. Especially for the RER A all routes are considered homogenous. Their characteristics are summarized at Table 5.

Table 5: Operational characteristics of the M1, M13 and M14 metro lines and the RER A

Route	M1	M13	M14	RER A
Nominal frequency per direction, ϕ_{zi} (trains/hour)	34	36	40	30
Standard dwell time, T_{zi}^0 (sec)	20	20	20	40
Safe separation time, ω_{zi} (sec)	85	80	70	80

Figure 13 illustrates the effect of the increase of the average dwell time of the vehicles on the frequency downstream. Since these lines are currently operating almost at capacity levels during the peak hour, we observe that a slight increase in the average dwell time may result in an important decrease in the service frequency. For example, exceeding the scheduled dwell time by 20 seconds implies a reduction of the line frequency downstream by 14% to 18% depending on the line.

In the case of the RER A, the model suggests that an increase of the dwell time from 40 seconds per vehicle to 60 seconds (an average dwell time of 55 seconds is regularly observed during peak hour), reduces the service offered to 26 trains per hour (13% service reduction for 50% dwell time increase). Therefore, an increase in the average dwell time at a station may cause a reduction on the number of vehicles downstream, hence a reduction to the available capacity and an increase in the expected waiting time.

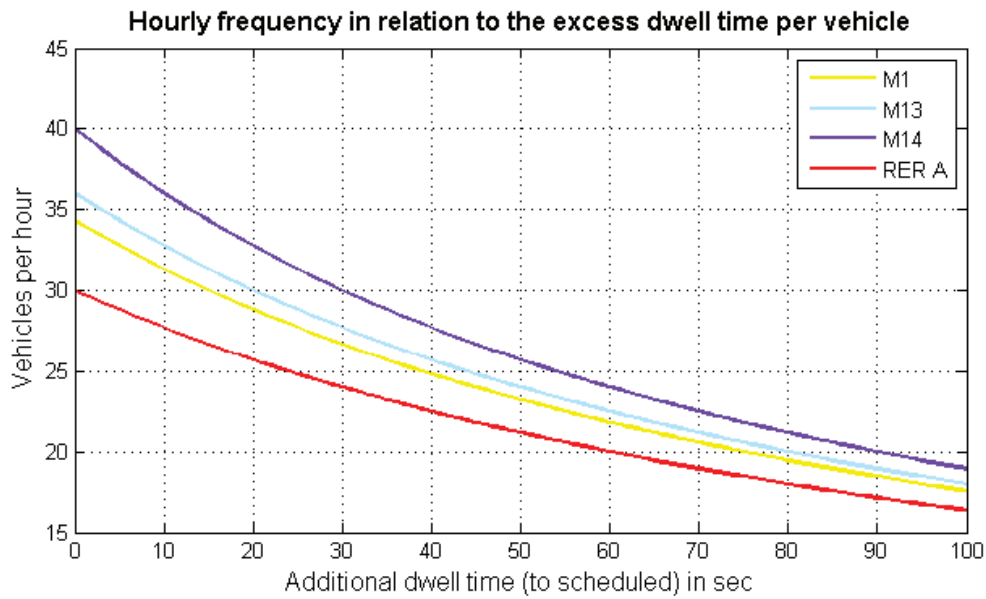


Figure 13: Restrained Frequency at station in relation to the increase on Dwell Time for Metro lines M1, M13, M14 and the RER A

4.5.3 Discussion

The formulation of the restrained frequency model offers an approach taking into account the interaction between the line operation and the passenger flow. Its formulation provides some limitations on the type of effects able to reproduce. Namely, the model is defined within a static transit assignment model and it is affected by the average dwell time of the vehicles of each route. It does not consider the temporal variation of the arriving passengers. In addition, the model tackles at each station the capacity constraints and the frequency propagation downstream, without assuming a spillback effect upstream due to insufficient capacity.

The model described above considers uniquely the effect of the dwell time and station platform occupation on the operation of the transit lines. Nevertheless, it does not take into account more complex operational effects on the line level, such as the convergence of the service branches or the travel time instability due to service reliability.

The model is formulated in a simple way by imposing a reduction factor whenever the temporal occupation of the platform exceeds the reference period. Nonetheless, it has an important effect on transit operation and route choice. It is significant if we consider the secondary effects induced due to the reduction of the frequency and thus the available capacity downstream. Therefore, the path costs and the route choices of the users are

influenced by those secondary effects depend on the vehicle capacity availability, such as total persons' capacity at boarding and seat congestion.

4.6 Combinations of Capacity Effects

Although each capacity effect is treated with specific local models, the secondary effects emerge on the line model, due to the local constraints imposed. Three secondary effects are identified on the line model:

- (I) The restrained frequency on the platform waiting time;
- (II) The restrained frequency on the in-vehicle comfort and
- (III) The transit bottleneck on the in-vehicle comfort

The restrained frequency model treats the effect of the passenger flows, through the vehicle-platform interaction to the transit operation, in particular the service frequency of the transit services. Therefore, restraining the frequency of a service z at the station i affects the intensity of the boarding passengers downstream. That can be approached, either by the available capacity of a vehicle facing a greater exogenous flow (k_{zi}^R opposed to $x_{\ell(i,s)} / \varphi_{zi}^-$) or by the exogenous flow competing for a reduced available capacity of the line ($x_{\ell(i,s)}$ opposed to $\varphi_{zi}^- \cdot \sum_{z \in i} k_{zi}^R$). The impact on the local route proportions and the waiting time can be significant, particularly if a service gets saturated.

Similar impact of the restrained frequency can be identified on the in-vehicle comfort, by means of the available capacity for boarding. An increase of the boarding flows impacts the probability to seat for the on-board and the boarding passengers downstream and, hence, degrades the in-vehicle comfort.

The impact of the transit bottleneck on the in-vehicle comfort is counter-intuitive, since constraining the boarding flows improves the comfort of the passengers on-board. Indeed, in comparison with a model where the boarding flow is unbounded, the transit bottleneck enforces strict capacity constraints on the boarding flow. Therefore, fewer passengers on-board compete for the same number of seats at each competition stage. Furthermore,

enforcing strict capacity constraints on a station implies that the passengers boarding the downstream stations will have a bigger probability to access the less penalizing seating comfort state.

4.7 Conclusion

We presented the concept of the line model and the local models for capacity constraints that compose it: the in-vehicle comfort, the transit bottleneck and the restrained frequency. The line model is located between these local models at the platform level and the network model, acting as an elaborate cost-flow relationship. It allows in the same time to restrict some capacity constraints within the transit line where they are more relevant, and to provide a simple representation of the passengers' access – egress trips within a line, which are treated at the network layer. In other words, the line model acts as a transition between the network and the local flow, while coordinating the line operation and the services.

Into that modular framework, the local capacity models evaluate the effect of the passenger flows and the capacity constraints on the travel cost as well as the interplay between the demand – as passenger flow – and the supply via the service frequency. The framework is essentially systemic and modular: some parts of it may be replaced by more appropriate sub-models, for instance about track occupancy, or the interaction of access and egress flows in station dwelling. Many developments can be thought of: (i) a complimentary model for vehicle alighting and platform access, (ii) door allocation to only one direction of traffic flow, (iii) control of dwell time and restriction of boarding flow, instead of endogenous dwell time, (iv) external traffic on interstation links, (v) featuring the platform layout and its effects on the passenger flows and waiting stock, (vi) the return trip of service vehicles relate the two directions of traffic on a given line, (vii) stochastic features in the formation not only of travel time and generalized cost, but also of flows and stocks and (viii) an elastic vehicle crush capacity, related to the passenger stock failing to board.

Chapter 5:

A Transit Station Model on a Structural Transit Network

5.1 Introduction

In the transit network of large urban areas, it frequently occurs that the transit system is submitted to heavy congestion at the peak hours on working days, especially so at the morning peak in the central part of the urban area. Along with the congestion on the transit services, related to the vehicles' and lines' characteristics, congestion appears at the transit stations. That influences the in-station walking time for the access-egress trips and the transfers of the passengers, hence the quality of service of the transportation system.

The principal design constraint of a transit station is the security evacuation, especially so for underground facilities in the case of fire. Various safety guidelines designate the evacuation flows and determine the passenger behaviour associated with the safety evacuation of a transit station: the NFPA130 (NFPA, 2010) in the United States and the IN1724 (SNCF-GI, 2002) in France. The respect of the safety regulations is one of the reasons of the development of microscopic and macroscopic simulations models, such as PEDROUTE (Buckman and Leather, 1994), Legion (Still, 2000), NOMAD (Hoogendoorn, 2001) and SimPed (Daamen, 2002). These models are often simulation-based. The circulation or the storage of passengers in the transit station is modelled into a very high level of detail.

The effect of the capacity constraints in a transit station is further discussed in the Transit Capacity and Quality of Service Manual (TRB, 2003). It is focused in the nominal operation of the transit system and acts as a guideline for station design. It provides some operational

characteristics of the station elements that act as reference values for station design and suggests a simple framework for estimating the effect of passenger flow.

However, fundamental issues such as the transfer for the transit system and the circulation inside a transit station are not given thorough attention in transit planning, in part due to the absence of efficient transit assignment models. The scientific literature boasts various transit station models for the assessment of the station design (Hoogendoorn et al, 2004) and other operational purposes (Harris, 1991). Other research proposes methods for the assessment of the cost of the transfer in particular stations (Guo and Wilson, 2011). Nevertheless, these approaches treat a transit station separately, independently from the transit system. The in-station path flows are assumed of fixed origin and destination and the network effect is not taken into consideration.

A passenger flow assignment on a transit network focuses on the path choice behaviour in relation to the quality of the transit service. Recent models are capable of efficiently capturing the capacity effects of the transit services on the path choice (such as Cepeda et al., 2006; Kurauchi et al., 2003; and Leurent et al., 2011). Nevertheless, they omit the effect of the in-station passenger flow on the quality of transfers and access-egress trips and the flow assignment on the network.

This chapter is purported to bridge the gap between the microscopic pedestrian simulation and the macroscopic transit assignment models. The modelling approach lays into developing a simple model for the transit station where the effect of the in-station passenger flows on the in-station path travel time and quality is taken into account. That model is integrated to a transit assignment model, such as CapTA.

The rest of the chapter is organised into seven sections. First, a systemic analysis of a transit station is conducted in Section 5.2, where the station elements are described along with station examples. Then, section 5.3 provides the definition of the network topology. The model assumptions and the formulation of the cost flow relationships are outlined in section 5.4. Section 5.5 addresses the articulation of the transit station model to the network model, which stems from CapTA. Section 5.6 deals with an application instance, adapted from the Nation station in the centre of the Paris transit network. Finally, the conclusion summarizes the model characteristics and discusses some points of potential development.

5.2 Systemic Analysis of a Station

The systemic analysis is essential as a lead-in to the modelling process. The boundaries of the system are defined, along with the relations to the external elements and the other systems. Furthermore, the elements of the system are described, along with their relations.

We perform a systemic analysis of the transit station, focusing on the part of the station used by the passengers and the passengers' activities inside. First, an overview allows the reader to place the transit station within the transit system and focus on the functionalities involved and the relations among the elements. Then the station elements are defined, first the horizontal and vertical elements and then the station areas. Finally, the effect of the transit information on the quality of service and the path choice is clarified.

5.2.1 Overview of the Transit Station

The systemic analysis of the transit system performed in Leurent (2011) identifies a transit station as a separate subsystem of the transportation network. The station fulfils two primary functions with respect to the transit users. First, the station acts as an interface between the urban area (city) and the transit network (transit service) providing access to and egress from the transit system. Second, when multiple transit lines serve a transit station, connections between them are available via the transit station allowing a passenger to transfer, hence increasing the connectivity and performance of the system. In relation to these functionalities, straightforward analogies can be found with the role of the airport terminal to the commercial aviation system.

The boundaries of the transit station subsystem are defined in articulation with the multimodal transportation system. Two input and output sides are distinguished. On the city side, a transit user can access to (or egress from) the transit station by private modes: walking, using the pedestrian network, by private car, using the park & ride (as a driver) or the kiss & ride (as a car passenger) facilities available or by cycling, using a bike shelter. On the transit service side, a user boards on (or alights from) a transit vehicle: a bus vehicle at a stop or a rail vehicle at a platform. These correspond to the origins and destinations of the trips within a transit station.

The main functionalities of the intermediate components of a transit station are the accommodation, consolidation and dispersion of the passenger flows, from the origin and destination points. Secondary functionalities of such station components are related with the production of the transit service (eg. fare control barriers and ticket machines) and the non-transit services to passengers (eg. shops) as a way to improve the quality of the service.

The flowchart in Figure 14 illustrates the transit station and its components. The rounded corner rectangles correspond to the external points of the transit station, or the origin and the destination of the pedestrian trips within the station, both city side and transit service side. The rectangles correspond to the areas of the transit station. Their main functionality is the consolidation and dispersion of the passenger flows and the storage of passengers stocks. These areas may include various amenities. Finally, the links represent the linear circulation elements, both horizontal and vertical. They have an analogue function and the use of horizontal or vertical element depends on the station design.

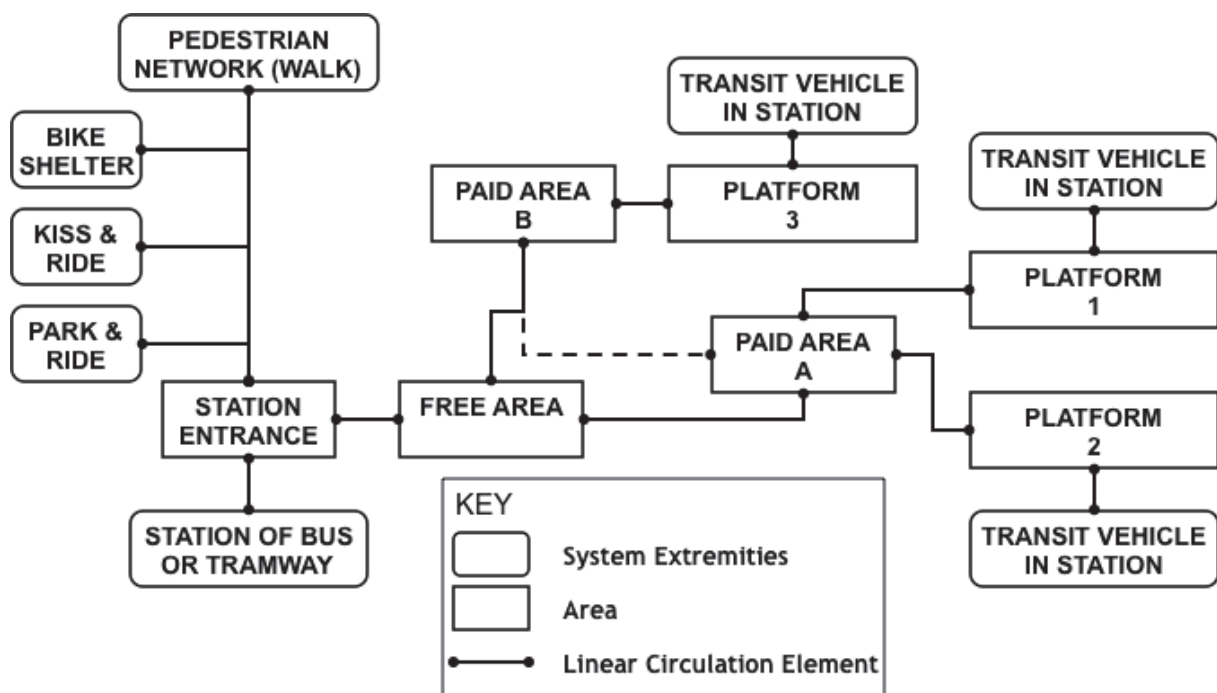


Figure 14: Flowchart of a transit station

The trajectory of a passenger who accesses a transit system originates at the pedestrian network (walking) and the intermodal facilities: park & ride (as a car driver), kiss & ride (as a car passenger) and bike shelter (as a cyclist). Then, he enters the station, circulates through the

station concourse (free and paid area) to reach the platform. Some waiting is needed before a vehicle becomes available for boarding. An egress trip is defined in an inverse way. Similarly, a transferring passenger alighting at the stop or platform circulates at the station concourse and the horizontal and vertical circulation elements to reach the preferred platform. Some waiting may be needed to board the preferred vehicle.

In the following parts we describe the components of the transit station. We first address the linear circulation elements – horizontal and vertical – and then the station areas. The description of the linear circulation elements is related to their functionalities, their geometric characteristics, the passenger flow capacity, the travel time and their particularities. The description of the station areas involves their functionalities, their geometric forms and their position within the operation of the transit system. If not stated otherwise, the values used in the following section stem from the Transport Capacity and Quality of Service Manual (TRB, 2003).

5.2.2 Horizontal Circulation Elements

5.2.2.1 Walkway

A walkway is a linear circulation element connecting two entry/exit points, where usually the length of the elements is far superior from its width. A small ramp may exist, however the vertical displacement is largely inferior to the horizontal movement of the passengers. The walkway may be used for a single or double direction passenger flow, according to the station design.

The capacity of a walkway depends on the pedestrian walking speed, their density and characteristics and the effective width of the walkway at its narrowest point. According to TCQSM, 2003 its capacity (LOS E) is located at 82 p/m/min or equivalent to 4920 p/m/h. Its capacity may be reduced for a bi-directional passenger flow on the walkway according to the relative ratio of these flows (Pushkarev and Zupan, 1975). At free flow conditions a passenger may reach the maximum walking speed on that element, equivalent to 81 m/min or 4,87 km/h. The increase of the passenger flow leads to a reduction of the average speed of the passenger flow and reaches 46 m/min (2,76 km/h) at capacity. When the density of the walking

passengers is greater than the capacity, the speed, as well as the flow, is reduced until reaching a halt.

5.2.2.2 Moving Walkway

A moving walkway resembles the functionality of a walkway for the circulation of passenger flow. It offers mechanical support for the passenger horizontal movement, usually operating at lower than the walking speed, at 1,8-3 km/h. The passenger speed depends on the passenger flow, as suggested for the walkway, increased by the operating speed of the moving walkway. The capacity depends on the entry width, similarly to the escalator. A standard double width moving walkway can transfer up to 5400 p/h. A moving walkway is operated at a single direction. Nevertheless, for long corridors, various layouts exist from a single moving walkway operating at the direction of the dominant flow to one or more walkways for each direction according to the structure of the passenger flows.

5.2.2.3 Fare Control Barriers

Fare control is an important function, linked to the performance of the transit operation. Usually for the rail mode, the fare control takes place inside the transit station. There, the fare control divides the station to the controlled areas (or paid areas), linked to the transit services, and the uncontrolled areas (or free areas), which provide the link to the city. In case of different transit fares among systems, additional fare control barriers may be placed between paid areas inside the transit station. An elementary fare control barrier may have a predetermined usage allowing only the entry or exit of passengers, or alternating the operating direction according to the principal passenger flow. The latter case means that the capacity of a group of fare control barriers may be dynamically allocated during the day according to the characteristics of the flow.

The operating mode of the fare control barriers resembles that of toll plazas on motorways. A service time may be defined by the elementary transaction (validate and cross). That time depends on the passenger characteristics and the type of fare control mechanism and varies from 0,5 to 4 seconds per transaction (TRB, 2003). The capacity of a group of fare control barriers depends on the service time and the number of elementary fare control gates used for each direction.

5.2.2.4 Station Entrance

The station entrance is defined as the limit between the station exterior and the interior of the station and its main functionality is to offer the station concourse a protected environment from external conditions. While for a typical ground transit station, such as a railway station, the station entrance can be easily noticed, corresponding to the entrance of the building, the concept is harder to define for underground transit stations. There, the vertical circulation elements frequently play the role of the station entrance.

In the same way to the fare control barriers, it constitutes an obstacle to the passengers. That has as a result to slow down the passengers. Nevertheless, if the entrance elements operate at predetermined directions, the opposite passenger flows are separated, increasing their fluidity. The capacity of the elements depends on the type of mechanism used for the separation of external and internal area – simple, revolving or sliding door – and the passage characteristics.

5.2.3 Vertical Circulation Elements

5.2.3.1 Stairway

The stairway is the main element for vertical circulation of the passenger flow and functionally resembles a walkway. It is used to connect areas with a height difference, such as different levels of a transit station. Due to the higher effort needed the passengers ascend and descend a stairway in slower speeds than walking. In addition, the ascending speed on the slope varies according to passenger flow from 33 m/min (or 1,98 km/h) to 20 m/min (or 1,2 km/h) and it is usually lower than the descending speed. According to the TCQSM (TRB, 2003), the capacity of a stairway per unit of width, corresponding to the upper limit of the LOS E, is 56 p/m/min (3360 p/m/h). If not considered otherwise, ascending and descending passengers may coincide at a given stairway. The produced friction causes a reduction of the stairway's total capacity.

5.2.3.2 Escalator

The increased effort needed for ascending and descending a stairway makes the escalator an essential element for the vertical circulation. It provides mechanical assistance and in that way increases the quality of the circulation within the station and the passenger flow for the same

width. By equivalence to the moving walkway, it operates in a single direction, and functionally allows the separation of passenger flows.

The operation speed of an escalator varies from 27,4 m/min (1,6 km/h) to 36,6 m/min (2,2 km/h). Therefore, the speed of the passenger flow depends on the operating speed of the escalator and the individual passenger on the escalator. Its capacity depends on the entry width and varies from 2700 p/h to 5400 p/h for a single and a double width escalator respectively.

The station design usually places the escalators adjacent to the stairway. When instant capacity is not sufficient, a queue is created at the base of the escalator. Assuming the passengers wish to reduce their travel time, a passenger will accept a maximum waiting time. This time is related to the difference of travel times between the escalator and the adjacent stairway. The passengers at the end of the queue who perceive their waiting time as greater than that travel time difference will use the adjacent stairway.

5.2.3.3 Elevator

An elevator can be used in two ways: either as a primary vertical circulation element to access levels with an important height difference (such as in deep underground stations) or as a secondary element to facilitate a part of the passenger demand which is excluded from other vertical circulation elements (such as disabled and persons with reduced mobility).



Figure 15: Elevators for primary use in Paris, Auber station (source: author)

The circulation speed and capacity is independent of passenger flow and is linked with the elevator's characteristics. In fact, operationally it can be distinguished from other circulation elements. By its operation characteristics it resembles a transit service with vertical movement. It is characterized by discrete availability and a waiting time until it becomes available, a travel time between stops and a person's capacity of the vehicle (or cage). The elevator used in transit service can be categorized as a fixed route demand responsive transit service.

5.2.4 Station Areas

5.2.4.1 The Station Concourse

The station concourse is defined as an area within the transit station with multiple local origins and destinations of walking trajectories, acting as a junction of different paths. The primary functionalities are the consolidation and dispersion of the passenger flows and the storage of passengers stocks. The secondary functionalities are derived by the amenities located at the station concourses, which offer various services linked to the transit service (ex. ticket sale, information point, waiting area) or not (newspaper stand, etc.). Fare control barriers are usually installed at the station concourse, adding a further distinction between uncontrolled (free) and controlled (paid) area. In addition, it gives access to non-public technical areas.

The station amenities induce a high impact on the passenger flow. On the one hand, these amenities need a minimum space for their operation, limiting the area available for passenger circulation and slowing down the passenger traffic. On the other, their positioning at the station concourse may lead to a more efficient passenger traffic organization with reduced conflicts and hence improve the performance of the facility.

The walking speed is influenced by the passenger density and the friction with conflicting or adjacent passenger flows and can be related to the number of the passengers present per surface area. The passenger flow is usually organized in virtual directed corridors within the station area, whose width fluctuates according to the composition of the flows in the area.

5.2.4.2 Station Platform

The platform is a particular element of the transit system, acting as a pivot point between the transit services and the station. A station platform's primary functionalities lay on the passenger exchange with the transit vehicles (for boarding and alighting), by way of consolidating the passenger flows and boarding, as well as alighting and dispersing at the station. Furthermore, it serves as storage of passengers that wait for the transit service to arrive. Secondary functionalities are related with improving the quality of the waiting with various amenities.

The usual form of the platform resembles a rectangle where the longer sides correspond to the input and output points of that element. The one side – along the tracks in the case of rail traffic – is used for vehicle berthing, while the other offers a connection, by the circulation elements, with the transit station via entry and exit points, as illustrated in Figure 16. When a vehicle is present the passengers board and alight the vehicles. The pedestrians in the station and the passengers on the vehicles are characterized by different flow forms: a continuous flow is associated with the in-station flow, while the on-board passenger are organised in packets, due to the discrete form of the transit service.

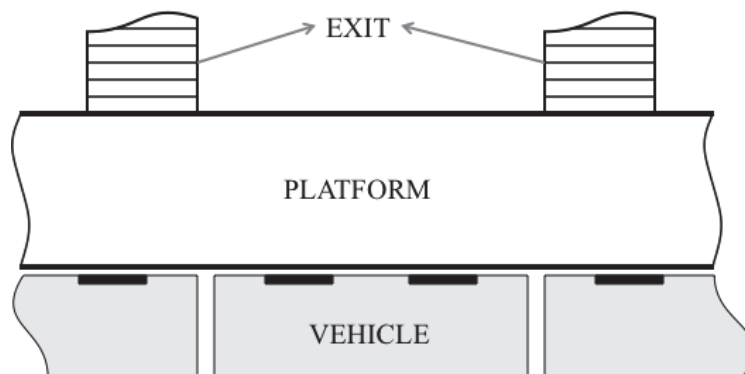


Figure 16: Schematic representation of a station platform

Waiting occurs before boarding, as a result of to the transition between the passenger flow forms. When the number of waiting passengers increases, the passengers choose waiting positions at the edge of the platform. Thus, a walking corridor is liberated at the rear, facilitating the access and egress circulation on the station platform. The waiting at the platform is associated with the storage function: two forms of waiting may be distinguished.

On the one hand, passengers waiting at the track-side of the platform, ensure priority from later arrivals to enter the vehicle and occupy a seat if available. On the other, passengers at a waiting area benefit from the seats available but may face difficulties into ensuring a better comfort state on-board. A waiting trade-off appears, between the waiting comfort state and the in-vehicle comfort state. The waiting choice of an individual passenger depends on the relative time of the waiting time and the in-vehicle travel time as well as the in-vehicle comfort conditions and the probability to access an in-vehicle comfort state.

To summarize the platform functionalities, the boarding passengers use the circulation elements to access the platform, suffer some waiting and when the appropriate vehicle arrives they board. The waiting passengers are not uniformly distributed along the station platform, but certain elements may affect the distribution, such as obstacles, the position of the platform entries and the knowledge of the exact position of the doors when a vehicle has stopped – either by prior knowledge due to a homogenous rolling stock, an information by the operator or platform screen doors. The quality and the performance of the passenger boarding and alighting – linked to the vehicle's dwell time – depends on the relative horizontal and vertical gap between the vehicle and the platform, the number and width of the doors and ease of circulating, due to the density of both the on-board passengers and waiting passengers at platform.

When passengers alight a vehicle, they are in front of the waiting passengers and circulate towards an exit, according to their in-station destination. The exit may be a vertical or horizontal circulation element, depending on the station design. However, the discrete alighting process means that a platform exit faces recurrent packet of passengers arriving. That infers to the need of additional waiting for evacuating the platform. Even when the hourly (average) capacity of the platform exits is sufficient, a bottleneck may be created at the entrance of the circulation element due to the discrete character of the transit services and the passenger arrival rate associated.

A review of the transit systems unveils a diverse typology of platforms where each type of platform has certain particularities due to the associated transit systems. We explicit the main platform types along with their operational characteristics. A central platform (or island platform) is a composite platform formed by two elementary ones. Even though the vehicle berth positions are distinct, they share common entry and exit points, affecting the arrival rate

of the alighting passengers at the circulation element if simultaneous arrivals occur. A perpendicular platform is defined when the single entry/exit is placed at the short sides of the platform. That disposition, an extreme case of uneven entry positioning usually found in rail terminals, complicates the circulation along the platform and favours the access and egress trajectories from certain parts of the vehicles. On the transit operation side, a composite platform is formed when a large quantity of vehicle berths is available (usually at bus terminals) in a common area. While this can be further detailed in the elementary platforms, complex waiting phenomena may occur due to the visual information on the vehicle arrival at the other platforms. Hence queuing passengers are susceptible to overtake or abandon a queue in the presence of passenger congestion or efficient alternative transit services.

5.2.5 The Transit Information in the Station and its Effect on the Route Choice

Intelligent transportation systems are being deployed worldwide as a way to improve the performance of the transit system and the quality of the service offered. That information – produced by information and communication technologies - is highly valued by both existing and potential users (Hickman and Wilson, 1995). Its presence at the various decision points influences the route choices of the passengers. By passenger information we group all the information, static (timetables and service maps) and dynamic (real-time information), that have an effect the passenger's perception of the transit service. According to its location it can be divided into three categories:

- (I) Information prior to the journey
- (II) Information in the station
- (III) Information on-board a vehicle

Various media of information exist, such as paper plans and signs, screens and more recently the cellphones and smartphones. An additional medium of dynamic information is considered the visual information, based on real-time observation. The information distributed usually concerns the time of the next vehicle arrival or the waiting time, but the recent breakthrough in information technology may allow sharing more composite information such as the dynamic comfort state of the next arrivals or the state of the transit network.



Figure 17: An example of real-time information: (a) station screens before the fare control barriers at La Defense station (source: author) – left – and (b) smartphone application (source: SNCF Direct) – right

The dynamic information at a station – usually involving display screens with the next vehicle arrivals or the waiting time – has an important impact on the path choice. Without information on the imminent arrivals at the platforms, a passenger chooses the platform on the basis of minimizing the expected travel time to the destination. Therefore, given the location of the decision point, without any information, the feasible path set is reduced to a single path. On the contrary, when information is available at that decision point, the path choice depends on the relative arrivals of the vehicles on different platforms. That choice falls under the optimal strategies concept (Spiess and Florian, 1989) – where the path probabilities correspond to the relation of the frequencies of the relevant alternatives or bundles, φ_a .

Real-time information at stop is a basic attribute for establishing the stop model: determining the waiting time and route choices. Nevertheless, only few modelling approaches include that attribute and its impact. Gentile et al (2005) redefine the stop model to include the knowledge of the next arrivals a general headway distribution of the transit services. An Erlang distribution (for the vehicle headways) is used to establish a general formulation of passenger waiting. Hence, the algorithm introduced in Gendreau (1984) can be extended to compute the expected waiting time with real-time information.

The impact of information at the stop is a basic issue for establishing a stop model. The waiting time and the route shares depend on them. That has been widely addressed by researchers, often without an efficient algorithm for many alternatives and for a large-scale network. However, Gentile et al (2005) proposed a reformulation of the stop model to include

the impact from the knowledge of the next arrivals. In that case, they adopt the Erlang distribution for the headway of the transit services and develop a general formulation to evaluate the main variables on the stop (waiting time and route share). The exponential and deterministic cases are no other than special cases of the generalized formulation.

Nevertheless, the presence of congestion on the platform alters the characteristics of the waiting time and hence the path probabilities. The waiting time can be split in vehicle waiting time (due to the random arrival distribution of the vehicles) and queuing (due to congestion passengers are unable to board the first vehicle). However, queuing acts as a buffer for waiting of the next available vehicle. Thus, the combined waiting time – when congestion is present – can be no longer considered as a random variable. In that case the combined waiting time has a deterministic value and the line option with the waiting time is reduced to a pedestrian option evaluated at its average generalized time.

Under the effect of information on the service operation and the passenger congestion, the path choice is managed at the dynamic decision points at the station. By that we define the points where real time information (on arrival time and congestion) affects the individual path choices on the network level. The dynamic decision points in the transit station have a great impact on the network path choices. Their effect is perceived individually by each passenger, according to his attributes and the desired trip destination. The trade-off between transfer station, access platform and quality of service can be modelled through these dynamic decision points.

5.2.6 The Transit Station in the Bus Mode

From a simple bus stop to a large bus intercity terminal we can distinguish many different types of bus stations. However, we focus on two types of bus stations linked to urban and regional services: the bus terminal and the urban bus hub.

A typical example of bus transit station is the bus terminal, situated at a business district or the entrance of the dense city centre, such as the Jules Vernes bus terminal in La Défense business district in Paris and the Port Authority bus terminal in Manhattan, New York. It is a multi-storey building or part of a building complex, which serves as terminal to the bus services and offers a connection with the city and urban transit. The in-station trips are bounded by the building. The relation between the in-station transfer trips with the other

transit modes and the access/egress trips depends on the location of each terminal but usually the former dominate over the latter.

The elements of the building are clearly distinguished: horizontal and vertical circulation elements with standard functionalities, while the station areas have a particular form. A clear distinction can be observed between bus-side and passenger-side, and some individual bus platforms are grouped into large waiting areas, like in an airport terminal. Therefore, a passenger may board the most appropriate service to his destination – according to the conditions imposed by the fare system and the type of services grouped.

The most common bus station is the urban bus hub, a set of bus stops around a central location in the city, such as a square, where many bus routes cross, offering an important number of possible transfers, like in Figure 18. Such bus complexes are common in the centre of cities, such as the hub of Chatelet (around Tour St Jacques) at the centre of Paris and that of Trafalgar square in London (Figure 18). Even though they are not bounded by the walls of a building, a group of stops can be defined as a transit station, because they are functionally linked with each other. The organization of a pulse service – especially for low frequency transit systems – improves the coordination of the services and leads to an increase of the number of transfers at these transfer points. The transit station may also give access to other transit modes, such as the metro.



Figure 18: Map of day buses from Trafalgar Square (source: TfL)

In urban bus hubs the access/egress in-station trips are difficult to be observed, as the use of the pedestrian network in all cases makes impossible the division between in-station trip from stop to exit and the rest of the access/egress trip. Anyhow, an in-station transfer trip uses the pedestrian network and its characteristics depend on the general pedestrian traffic present. Due to its exposure on weather and general traffic, the quality of the waiting and transfer depends also on the presence of bus shelters.

While the bus stops feature static information (printed plans etc.) on the neighbouring bus services of the urban hub, the real time information is usually limited to the bus routes serving a given bus stop. No information is given during the transfer trip, except of the visual information consisting of the observation of approaching buses.

5.2.7 The Transit Station in the Rail Mode

The rail transit stations are easier to distinguish from the urban context due to the technical necessity for separate infrastructure. Apart of the simple transit station, the complex rail transit stations destined for urban and regional transit can be classified into three categories:

- Rail terminals: As a result from the historic development of railways, rail stations with suburban and regional dominance terminate at the outskirts of the city centre, such as Saint Lazare station in Paris and Paddington station in London.
- Through-stations: Urban development around existing stations and technical advancement in the building methods for the construction in the city centre (usually underground) result in through-stations where multiple lines are served and high numbers of transfers are effectuated. Such examples are the Chatelet-les-Halles RER station in Paris and the Nuevos Ministerios Cercania station in Madrid, both dedicated to regional rail service.
- Mixed stations: Some stations are characterized by terminal and through service, such as the Milano Garibaldi station with suburban and regional rail service dominance.

These can be easily defined as transit stations, due to their particular technical characteristics. As a result of their increased accessibility (and labour catchment area), it is common that large office complexes and shopping malls are developed around (ex. Gare de Lyon in Paris) or over these rail stations (ex. Cannon St. Station in London). That contributes to the increase

of the access/egress trips on the station and the development of secondary activities. While these activities are generally welcomed by the authorities, because they diversify the revenues of the transit system, the additional pedestrian flow that they generate can have a significant impact on the passenger flows in the station.

The multiplication of the pedestrian and passenger flows together with the increase of the connections with the city and the other transit modes lead to consider a rail transit station, such as a ground rail terminal, as an open area. Since it is a public area, gathering functionalities other than transit, it is impossible to separate the passenger flow from the non-transit pedestrian flow since they are mutually affected. The above implies that an assessment of the travel time and the quality of service of the transfer and access/egress trips needs to take into account the impedance occurring from these secondary flows.

The fare control system installed – result of upstream network planning – constitutes an obstacle to the free passenger flow. However, in these circumstances it can help to separate the passenger flows from non-transit pedestrian flows and to divide the station surface into free and paid area. Further divisions can be made if multiple fare zones or tickets are used. In addition, in large rail stations, being part of large multimodal transit hubs, multiple trajectories exist for each in-station origin-destination trip. The static and dynamic – real time – service information and station signalling and the fare system used have an impact both on the in-station path choice and on the choice of the route (and by extent on the path choice at the network level).

The quality of service on the rail mode – due to its transport capacity and the massive flow it concentrates – is very sensible to the stock of passengers waiting at the station. The storage of passengers is a function located at the concourse level – or the station platforms if they are adequately dimensioned and there is prior information. At the rail terminals, whose platforms are most often not properly dimensioned for accommodating waiting passengers, the waiting area is situated at the main concourse. The dimensioning of such an area depends on the passengers' arrival behaviour and the arrival distribution of the passengers. That can be attributed to the uncertainty on the travel time and the frequency of the transit services (Kroess et al, 2007).

5.3 Network Topology

We define the network topology of the transit station used within a passenger flow assignment model for transit networks. As shown previously, the transit station is composed of various elements with different functionalities, geometrical and operational characteristics.

The topology of the transit infrastructure network is described by a directed graph $G = (N, A)$, with N the set of nodes n of the transit system and $A \subseteq N \times N$ the set of arcs a with endpoints in N .

The demand representation follows some standard principles. The simulation area is divided into zones; each zone includes a set of locations with trip endpoints. Let o and s denote a trip origin and destination respectively, where $o \in O$ and $s \in S$ (not to confuse with the station symbols) and let W be the set of zone pairs from origin to destination. An origin-destination pair on the transit network is defined as $(o, s) \in O \times S$. Passenger demand on the network is modelled as a set of trip makers: each trip maker chooses a network path so as to minimize his own cost of travel between his origin and destination points.

For a given trip-maker, the in-station trajectory corresponds to a path, yielding a generalized cost and a travel time on the basis of local characteristics defined by network element, node or link of the station.

5.3.1 The Lower and Upper Network Layers

The representation of the transit station within a transit assignment model extends the bi-layer network representation of the transit services to the transit station. By that approach we define the flow characteristics of each station element in detail at the lower layer, while keeping the upper layer consistent to the network flow assignment.

On the lower layer of the network each transit station is modelled as a specific sub-network corresponding to the infrastructure network. In other words, each transit station is handled separately and the physical elements of the station (simulating to the supply side) are defined along with their operational characteristics and the associated cost flow functions.

On the upper layer a trip-maker is faced to a service network where each station leg corresponds to a particular origin – destination trajectory in the station. Such leg represents

one or several station physical elements (walkway, stairway, escalator etc.) along the transit station in the infrastructure network. Therefore the upper layer acts as a representation of the demand side, where a trip from an origin to a destination is made up by line leg arcs (corresponding to the access – egress station couples along a line) for the inter-station journey, station leg arcs for the intra-station and the access-egress legs for the respective trips.

There is a twofold relationship between the two layers by station s : top-down, a vector of passenger flows $\mathbf{x}_s = [x_a : a \in A_s]$ by leg a along the station is assigned to the station sub-model, yielding the local passenger flows; bottom-up, the station sub-model yields the vector \mathbf{g}_a of the average generalized cost for $a \in A_s$. In other words, the station model amounts to an elaborate cost-flow relationship in vector form on the upper layer of the network.

5.3.2 The Lower layer representation

The lower layer corresponds to the infrastructure network. A transit station is represented as a specific directed graph $G^{(s)} = (N^{(s)}, A^{(s)})$ made up of a set $N^{(s)}$ of station nodes n , together with a set $A^{(s)}$ of arcs a with endpoints in $N^{(s)}$. We define three subsets of the set $N^{(s)}$ of station nodes: N_E the set of station entry and exit nodes, directly connected with the origin and destination nodes $o \in O$ and $s \in S$ of the transit network, the set N_P of platform edges, acting as an intermediate between the station and the transit service and the set N_S of station nodes. The first two subset act as the origin and destination points for the in-station paths of the station subsystem.

An arc represents a state transition for a trip-maker and the set $A^{(s)}$ is composed of the following subsets of network infrastructure: the subsets A_V of vertical circulation elements, A_H of horizontal circulation elements and A_P of platform arcs.

The infrastructure network arcs correspond to the physical elements used for the passenger circulation and storage within a transit station. Each station element (horizontal and vertical circulation and station area) is characterised by a particular flow-dependent travel time $t_a(y_a)$ and generalized time $g_a(y_a)$, described in the following section.

5.3.3 The Upper layer representation

The upper layer corresponds to the service network, as a representation of the demand side for the network model. As the transit station representation is concerned, the service network is made up of station legs. A station leg corresponds to the average conditions of a passenger circulating from an origin to his destination in the transit station subsystem. It is composed by consecutive physical circulation elements, as defined on the lower layer and its generalized cost and travel time is built upon these elements.

The station leg arcs on the upper layer are denoted by the subset $A_S \equiv A_H \cup A_V \cup A_P$. Overall, the service network is made up of the set $A \equiv A_S \cup A_L \cup A_R$ and the associated node sets. A_L denotes the subset of line legs, representing the average conditions of the transit service of an access – egress station couple on a transit line and A_R the subset of legs for accessing to and egressing from the transit network (see Chapter 6.3).

The transit station is defined as a subsystem of the transit network offering connection between the city and the transit lines and among the transit lines, where the passengers perform access – egress trip and transfers respectively. In the service network, an access itinerary is composed of an access arc and a series of station physical arcs from an origin node $o \in O$ connected to a station entry and exit node, N_E , to a platform edge node N_P , that form a station leg. The inverse is applied for the egress trip. Respectively, a transfer is represented by a station leg $a \in A_S$, connecting two platform edges within the station or a bus stop and a platform edge.

5.4 The Station Model Formulation

In the first part of this section we present the form of passenger flow and their particularities when dealing with capacity constraints. Subsequently, we develop local models addressing the local capacity effects within the station: platform entry and exit flowing, the pedestrian travel time by horizontal and vertical circulation element and the modelling of a station area. Then we sketch the outline of a station model dealing with these effects.

5.4.1 The Passenger Flow Form and their Influence on Capacity

The passenger flow at the transit systems takes two distinct forms: the continuous passenger flow, related to pedestrian walking for transferring and accessing to (or egress from) the transit service and the discrete passenger flow, related to the transit services, where passengers are organized into repeating on-board packets.

A phase transition occurs between these forms at the transit station. On the one hand, the transition from continuous passenger flow to discrete, upon boarding a vehicle, is associated with a waiting time – due to the discrete availability of the transit services. On the other hand, a transition of the discrete passenger flows to continuous takes place at the station upon passenger alighting a transit vehicle. The latter transition impacts the station operation. The circulation elements face recurring packets of arriving passengers and may be subject to congestion. At this point, we define two types of congestion:

- **Instant congestion:** due to the transition between passenger flow forms, the instant discharge rate of the circulation element is inferior to the instant arrival rate of passengers, resulting in the formation of congestion. However, the flow is dissipated before the arrival of the next packet of passengers, meaning that a recurring instant congestion is susceptible with each passenger packet arrival. The capacity during the simulation period is sufficient for the passenger demand;
- **Continuous congestion:** either by a transition phase or the presence of a bottleneck, the congested flow is not dissipated during the simulation period. Passenger congestion is present throughout the entire simulation period and the total capacity during the simulation period is not sufficient for the passenger demand.

The continuous congestion of the station elements has a considerable impact on the quality of the transit service. However, the instant congestion may influence the path choices at the station and at the network level.

The rest of the section addresses the capacity effects of the station physical elements of the lower network layer. A cost flow relationship, along with the model assumptions are explicated for the platform flowing, the pedestrian travel time by horizontal and vertical circulation element and a modelling approach is defined for the station area.

5.4.2 Platform Entry and Exit Flowing

A station platform is an area of the transit station where vehicles of the transit services are made available for boarding or alighting. The edge of the platform has a sufficient length for the vehicle berthing and the vehicle's doors act as an interface between the station and the service. Opposite, the platform's entries or exits are located: vertical or horizontal circulation elements are distributed on the length of the platform and act as a connection between the platform and the rest of the station.

A platform entry flowing consists of the trajectory from the platform's entry to the vehicle's door where the passenger accesses the transit service. That is associated with some waiting, calculated by the line model, via the arrival distribution of the passengers and vehicles and the vehicle boarding capacity effect. However, the presence of a large passenger stock on the platform – waiting for the next vehicle – degrades the circulation conditions. The passenger travel time on the platform is associated with the relation between passenger stock present and the storage capacity of the platform.

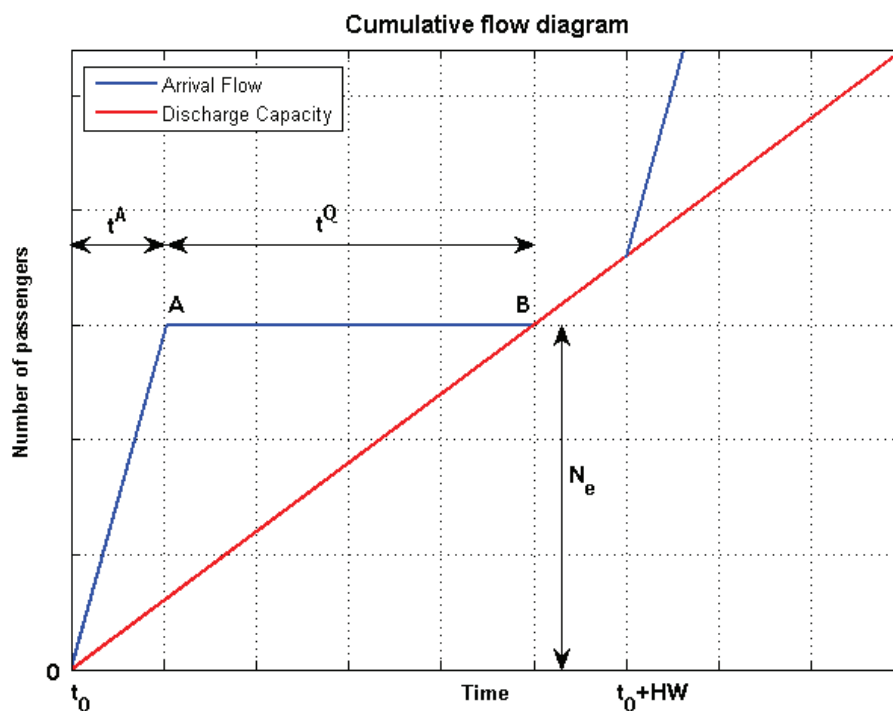


Figure 19: The bottleneck created when exiting a station platform

The platform exit flowing is related to the arrival of a vehicle and the trajectory of the alighting passengers from the doors to the station's exit. In addition to the difficulty of circulation due to the passenger stock, the instant arrival at the exits may cause instant congestion on the entry of the element. In that case an average waiting time is added at the passenger's platform travel time. Figure 19 illustrates the bottleneck produced at the base of single exit element when arriving flow (blue) is confronted to the discharge capacity (red).

Let t^A the walking time on the platform of the most distant passenger to the exit, V_c the walking speed and ℓ^e the catchment area of that exit element. It is straightforward that $t^A = \ell^e / V_c$. Let t^Q be the queuing time of the last passenger. The total evacuation time of the platform, t^E , is calculated through the time needed for the last passenger to evacuate and $t^E = t^A + t^Q$. If N^e denotes the number of alighting passengers on an elementary exit and k^e the nominal capacity of that exit, the evacuation time is $t^E = N^e / k^e$.

By making the appropriate algebraic transformations, the average waiting time, w^e , at the exit element due to the instant bottleneck, is given by the following formula:

$$w^e = \frac{1}{2} \left(\frac{N^e}{k^e} - \frac{\ell^e}{V_c} \right) \quad (19)$$

If the distribution of the passengers e_{zi} alighting from a vehicle of the service z at station i is homogenous and the station exits are equally distributed, the number of alighting passenger at each station amounts to $N^e = e_{zi} / E$, where E is the number of exits available. Also, the catchment area of an exit element will be $\ell^e = 0,5 \cdot \ell^P / E$, where ℓ^P is the platform length. The previous formula takes the form:

$$w^e = \begin{cases} 0 & \text{if } e_{zi} \leq (\ell^P k^e / 2 \cdot V_c) \\ \frac{1}{2} \left(\frac{e_{zi}}{E \cdot k^e} - \frac{\ell^P}{2 \cdot E \cdot V_c} \right) & \text{if } e_{zi} > (\ell^P k^e / 2 \cdot V_c) \end{cases} \quad (20)$$

The unbounded bottleneck presented previously, assumes that a passenger who presented himself at the base of the escalator remains indefinitely, until evacuating. However, usually the station design includes an escalator adjacent to a stairway, the latter having sufficient capacity to absorb the excess demand. In such cases, when the waiting time of the passenger

exceeds a certain value, t_{\max}^Q , we assume a deterministic behaviour of the passengers at the back who use the adjacent stairway instead of queuing. The maximum waiting time at the escalator is subject to the estimated difference of the travel time between the escalator and the stairway. An equivalent decision case would be the choice between a discontinuous bus service to the next stop or walking. The passengers giving up queuing by using the stairway suffer an additional time of t_{\max}^Q . The average waiting time for the passengers at the exit is:

$$w_e' = \begin{cases} 0 & \text{if } e_{zi} \leq \frac{\ell^P k^e}{2 \cdot V_c} \\ \frac{1}{2} \left(\frac{e_{zi}}{E \cdot k^e} - \frac{\ell^P}{2 \cdot E \cdot V_c} \right) & \text{if } \frac{\ell^P k^e}{2 \cdot V_c} < e_{zi} \leq \frac{\ell^P k^e}{2 \cdot V_c} + E \cdot k^e t_{\max}^Q \\ e_{zi} \cdot t_{\max}^Q - \frac{t_{\max}^Q{}^2 \cdot E \cdot k^e}{2} - \frac{\ell^P \cdot k^e \cdot t_{\max}^Q}{2 \cdot V_c} & \text{if } e_{zi} > \frac{\ell^P k^e}{2 \cdot V_c} + E \cdot k^e t_{\max}^Q \end{cases} \quad (21)$$

Figure 20 illustrates the average waiting time of the passengers on a RER platform with 4 exits and an escalator in each exit, for the bounded and unbounded bottleneck. We assume that the passengers who would wait longer than $t_{\max}^Q = 1 \text{ min}$ use the adjacent stairway. Although at that case we do not assume any capacity constraints for the stairway, an additional bottleneck could be created if an important passenger flow is present.

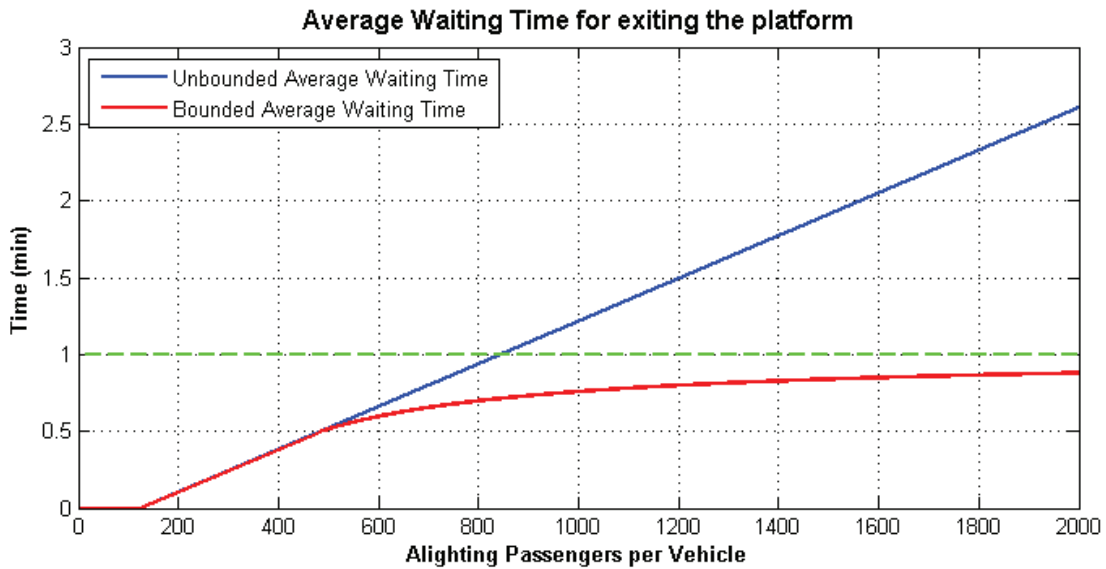


Figure 20: Average Additional Time of an alighting passenger to exit a station platform

Three platform exit phases are designated in Figure 20. For a small passenger flow, the arrival rate at the exit is inferior to the discharge rate, making the exit immediately available, without any additional time due to waiting. The second traffic state is limited to the arrival rate being higher than the discharge rate. Then a bottleneck is created at the base of the exit. If the alighting passengers are distributed homogeneously, the additional time due to waiting at the bottleneck increases linearly. However, in the case of the bounded model an additional traffic state can be distinguished. When the queuing time of the last arriving passengers exceeds their maximum admissible queuing time then, the last arriving passengers use the adjacent stairway, setting an upper bound to the waiting time.

5.4.3 Pedestrian Travel Time by Circulation Element

The design guidelines (TRB, 2003) suggest that the density of the circulating passengers affects the walking speed. In linear elements, such as walkways and stairways, the relation between passenger flows and passenger density is straightforward. Figure 21 illustrates the speed flow diagram of the elementary circulation elements in a transit station, a walkway and a stairway (ascending), according to the TCQSM (TRB, 2003) guidelines.

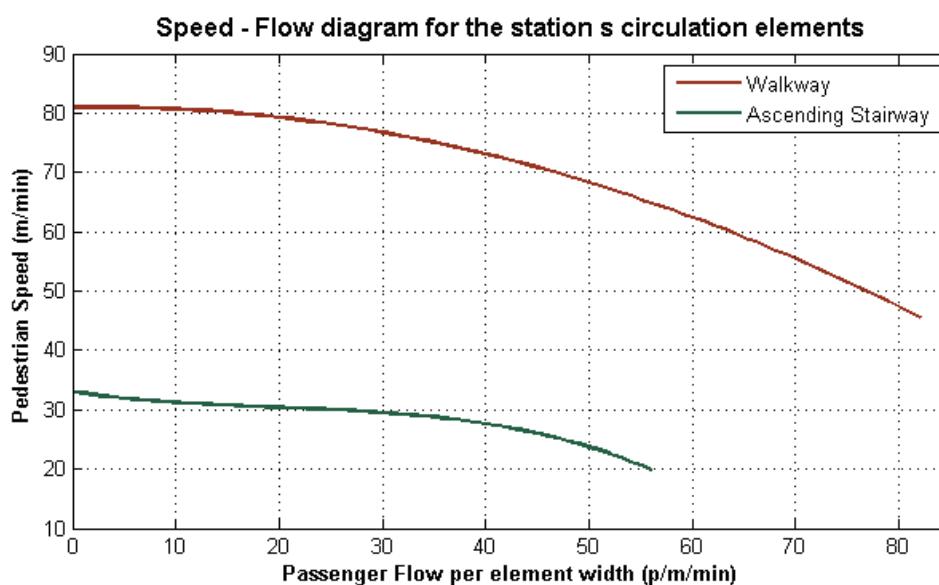


Figure 21: Speed - flow diagram for the station's circulation elements

A reduction of the walking speed is observed, from a free-flow pedestrian speed until the element's saturation. Higher passenger densities result in lower speed and flow and are not taken into account in the model. The walking speed of the passengers in any circulation

element is flow dependent. If the free-flow travel time of that element is available, an additional travel time can be evaluated as a function of the flow-to-capacity ratio on the element, since a variation in the walking speed – given the constant geometric characteristics – influences the travel time.

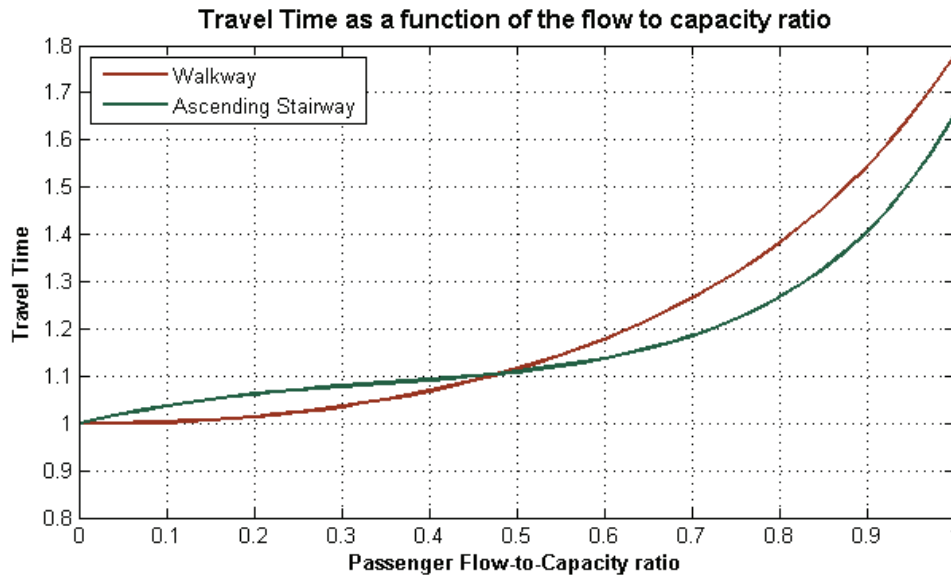


Figure 22: Travel time multiplier diagram in function of the flow-to-capacity ratio

A path is formed by a series of horizontal and vertical circulation elements with different attributes, such as capacity and free-flow travel time. The passenger flow on these elements is used to deduce the additional travel time according to Figure 22.

The station's circulation elements are rarely subject to instant congestion, since the scarce capacity at the platform exit leads to the downstream homogenisation of the passenger flows. However, the elements may be subject to continuous congestion if the available capacity during the simulation period is not sufficient.

A simple modelling approach is used, based on queuing theory for calculating the additional time due to continuous congestion. During the simulation period H , if y_a is the in-station passenger flow arriving at the element's entry, $N_d = y_a \cdot H$ is the number of passenger demanding the passage. Let k_a the maximum discharge rate (capacity) of the circulation element and H' the time needed for cooling the demand. It exists, $N_d = y_a \cdot H = k_a \cdot H'$ and therefore an additional time w_a is calculated for the passengers, where:

$$w_a = \begin{cases} \frac{y_a - k_a}{k_a} & \text{if } y_a > k_a \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

5.4.4 Pedestrian Travel Time by Area Element

A station area, such as the station concourse, is an element with various converging and diverging passenger flows. Nevertheless, the pressure of the passenger flow results in the formation of virtual walkways. Therefore, neglecting the idle passengers using the amenities, the passenger flows in a station area are modelled by directed arcs and a virtual capacity may be defined. A station area is reduced to a set of horizontal circulation arcs with predefined capacities.

5.4.5 A Transit Station Model Algorithm

The conditions external to the station pertain to the passenger load vector by station leg, $\mathbf{x}_S = [q_{uv} : u, v \in V_S]$, where $V_S \equiv N_E \cup N_P$ denotes the set of origin and destinations nodes of the in-station paths. By platform element, we denote φ_i^P the frequency of the services that serve the station, inherited from the line model, E_i , the number of platform exits and ℓ_i^P the platform length. We associate by circulation element a , a free flow travel time t_a^0 and an element capacity k_a .

A network equilibrium model with fixed demand, satisfying Wardrop's user optimality can be defined for each transit station subsystem. The nodes $i \in N^{(s)}$ represent the origin, destination and intersections of the physical arcs, $a \in A^{(s)}$. The origin and destination nodes u, v coincide with the endpoints of the station legs and we define an origin destination pair $(u, v) \in V_S^2$. These create arc flows y_a and the cost of travelling on an arc is flow-depended, $g_a(y_a)$, $a \in A^{(s)}$, as described in the previous sections. The passenger flow in the transit station is described by the in-station flow vector $\mathbf{y}_s = [y_a : a \in A^{(s)}]$.

The network equilibrium model for the transit station can be formulated as a fixed point problem. Each station $s \in S$ is treated separately by an iterative algorithm, based on the Method of Successive Averages (MSA). The traffic equilibrium problem of station

assignment can be solved, using a decreasing sequence of positive numbers $(\lambda_\mu)_{\mu \geq 0}$ with $\lambda_0 = 1$, as follows:

- *Initialisation:* Set $\mathbf{y}_s = 0$ and $\mu = 0$;
- *Cost Formation:* by station element based on \mathbf{y}_s . We treat each physical element on the transit station s .
- *Auxiliary State:* Shortest path search for all the origin-destination pairs and load the OD flows along them, yielding an auxiliary traffic state $\hat{\mathbf{y}}_s$. By destination, a shortest path is built on the station network from each node recursively, by a standard graph-search algorithm (Dijkstra, 1959).
- *Convex Combination:* Let $\mathbf{y}'_s = (1 - \lambda_k) \cdot \mathbf{y}_s + \lambda_k \cdot \hat{\mathbf{y}}_s$
- Evaluate *convergence criterion* between \mathbf{y}_s and \mathbf{y}'_s . If it is sufficiently small then stop with solution \mathbf{y}'_s , evaluate the cost of the station legs A_s . Else increment μ , replace \mathbf{y}_s by \mathbf{y}'_s and go to Cost Formation.

A convergence criterion can be based on the flow gap of the arcs between two consecutive traffic states, $\|\mathbf{y}'_s - \mathbf{y}_s\|$.

5.5 The Articulation with the Network Model

The transit station model is used as an elaborate cost-flow relationship within a network assignment model, in order to evaluate the generalized cost of the station legs for each transit station. In that form, it can be integrated in the CapTA network model. Further details are given in Chapter 6.

5.5.1 Upper Layer Route Choice

On the upper layer, station legs a have generalized cost g_a , but they are characterized by a waiting time w_a and a revised frequency $\hat{\phi}_a$, related to that of the line leg at the head of the

station leg. Let us define the revised frequency of a station leg, as a function of the conditions inherited by the line leg at the station leg's endpoint and the presence of in-station information. In that way, the station leg includes the dynamic decision points crossed by the in-station path, whose real-time information is susceptible to influence the path choice on the network layer. That choice is treated at the upper layer, in accordance with the characteristics of the line legs and the alternative station legs, using the revised frequency as follows.

In CapTA the treatment of the revised frequency is standard when $w_a = \alpha / \varphi_a$. If $w_a > \alpha / \varphi_a$ due to crowding congestion on the line platform, the passenger may not be able to board in the first incoming vehicle. In this case the leg option is akin to a pedestrian option evaluated at its average cost only. To integrate continuous (pedestrian) and discrete (uncongested transit) availability, let us define a “discontinuity attenuation function” denoted ψ that is continuous and decreases from 1 at 0 to $\psi(x) = 0$ whatever $x \geq \varepsilon$ a small positive parameter. The average waiting w_a and effective operating frequency φ_a delivered by the line model yield a revised frequency of

$$\begin{aligned}\hat{\varphi}_a &= \varphi_a / \beta_a \text{ if } \beta_a \equiv \psi(w_a - \alpha / \varphi_a) > 0, \text{ or} \\ \hat{\varphi}_a &= \infty \text{ if } \beta_a = 0\end{aligned}\tag{23}$$

In the presence of information, line combination will occur at the station with the formation of a line bundle, as considered in the basic model of line combination (Chriqui and Robillard, 1975), due to the random cost of waiting for a line leg $a \in A_L$. In that case the average waiting cost of a line bundled is evaluated by the combined revised frequency of the line legs of the bundle.

Then, line bundling on the upper layer is based on the revised frequencies and proceeds in the classical way, yielding route-based hyperpaths as in (De Cea and Fernandez, 1993). The presence of congestion will modify the revised frequencies of the station legs. The discontinuity attenuation function is in fact a development of the line model, as it involves only results of that model. It is further coordinated with the station model, based on the leg representation of the network layer.

5.6 An Application Instance of the Station Model

A transit station of the Greater Paris transit network is chosen as an application instance for the transit station model. The Nation station (Figure 23) is an intermodal transit terminal, located at the dense zone of the transit system. Apart from the RER A – for which is the first station in Paris from the East suburbs – it is served by 4 metro lines (2 of them as a terminal station), and 4 bus lines.

The input data are extracted by a network assignment on the Greater Paris network for the morning hyper-peak hour with an increase by 30% of the average peak flow (see simulation results of Chapter 9). While there is a significant number of access-egress trips (8250 passengers, whose one endpoint is Nation), there are largely inferior to the number of transfer passengers, who make the 83% of the in-station passenger flow.



Figure 23: A schematic map of the transit system at Nation station (source: RATP)

The upper layer network of the Nation station is composed of 14 station leg endpoints, corresponding to the origin and destination points inside the station: 10 line platform edges, 2 bus stops and 2 entries/exits, as an articulation with the access/egress trips. The paths within the station RER are represented by 134 station legs, connecting each leg endpoint (except the entries/exits with each other).

The exit flow on the platform induces a bottleneck at the base of the circulation elements. Table 6 illustrates some selected platforms. The A/R indicates the directions of the transit lines.

Table 6: Characteristics of station platforms and the additional waiting time of the platform exit flow

Platform	Exit Flow (pass/h)	Nb of Exits	Frequency (veh/h)	Platform Length (m)	Combined Exit Capacity (pass/min)	Additional Waiting Time (in min)
RERA A	10 174	4	30	224	136	0,45
RERA R	10 164	4	24	224	136	0,61
M1 R	6 013	2	34	90	124	0,56
M2 R	6 024	3	34	75	186	0,40
M6 R	7 638	2	34	75	124	0,79
M9 A	10 653	4	34	75	248	0,57

The exogenous passenger flow induces an increase in the travel time of the circulation elements from 8% to 41%, depending on their capacity, type and passenger flow. The travel time of the station legs, connecting the station input/output nodes, faces an increase of 15% to 79%. However a large part of that increase is attributed to the bottleneck created during the platform exit flowing. Here are some results for the transfer between lines:

- From RERA to M9, the travel time passes from 2,90 to 3,94 min (of which 0,45 min for platform exit flowing), an increase by 36%
- From M1 to M6, an increase by 66% to 1,99 min, of which 0,56 min of platform exit flowing
- From M2 to RERA, the transfer time increases from 4,00 to 5,12 min (+28%, where 0,40 min of platform exit flowing).

These results indicate a significant increase of the travel times within the station. It certainly does not have the same order of magnitude as the waiting time and the in-vehicle travel time. However, its impact on the path choices of the passengers and the path flows on the transit network remain to be evaluated. In the case of the Greater Paris transit network, the transfer

and access/egress trips on the network amount approximately to 30% of the average generalized time of a trip.

5.7 Conclusion

The transit station is a complex system, composed by a variety of elements, each one with distinct geometrical characteristics, functionalities and operational attributes. Furthermore, it is an important subsystem of the transit network, acting as an intermediary for the transfers between transit lines and for the access/egress trips to the transit services.

The station model constitutes a framework to represent the features and phenomena, of physical and microeconomic nature, that take place in the station and impact the transit operations and passenger path choices. With the bi-layer representation the model handles the physical interaction of the passengers in the station on the lower layer and the economic appraisal of the in-station trips on the upper layer.

The application at Nation intermodal station shows the effect of the passenger flows on the in-station pedestrian trips. The bi-layer representation leads to an increase, in the first place, of the upper layer station arcs, but the degree of detail chosen for the lower layer depends on the station particularities. Even though the influence of the in-station travel time on the total travel time is not such as that of the in-vehicle travel time, the more realistic representation of the capacity effects within a station is expected to have an impact on the passenger path choices, mainly when they involve transfers through congested transit stations.

Certain capacity effects are included in the current model such as the recurring bottleneck at the platform exits when a vehicle arrives, or the impact of the passenger flow to the travel time. Nonetheless, the framework is essentially systemic and modular: some parts may be replaced by more appropriate sub-models and other effects may be added. Many developments can be thought of: additional capacity effects may be included, such as the impact of the passenger stock in the travel time in the station areas, or the operation of the elevator as a principal circulation element.

PART III: THE NETWORK MODEL – THE CapTA SIMULATOR

Chapter 6:

A Systemic Representation of the CapTA Model

6.1 Introduction

The previous chapters describe the modelling principles and the behaviour of the models restrained to a sub-system of the transit network. They address specific capacity constraints in detail for each sub-system and evaluate the impact of passenger flows on the quality of service. These are calculated in the context of a network traffic assignment model. That is essential for the cost evaluation, in order to ensure that the network externalities are taken into consideration in a coherent way for the entire network. The equilibrium between transport demand and supply of transit services has also to be addressed in a coherent way. The objective of this chapter is to present an overview of the Capacitated Transit Assignment network model; hereby referred to as CapTA model. That is structured in a systemic way.

The chapter is structured in seven sections. Section 6.2 addresses a systemic analysis of the transit assignment models and details the components and the state variables of such models. Section 6.3 focuses on a fundamental characteristic of the CapTA model, the bi-layer representation of the supply and demand. Sections 6.4 and 6.5 describe the characteristics of the components on the demand and supply side of a transit assignment model. Section 6.6 summarizes the hierarchical relation between the model levels. Finally, section 6.7 makes a comparison between the CapTA model and other transit assignment models, while suggesting some development perspectives.

6.2 Systemic Analysis of the Transit Assignment Models

Similar to the transit system, a network traffic assignment model can be described in a systemic way. As such, a systemic representation of its components and relationships follows, based in Leurent (2003) and illustrated in Figure 24.

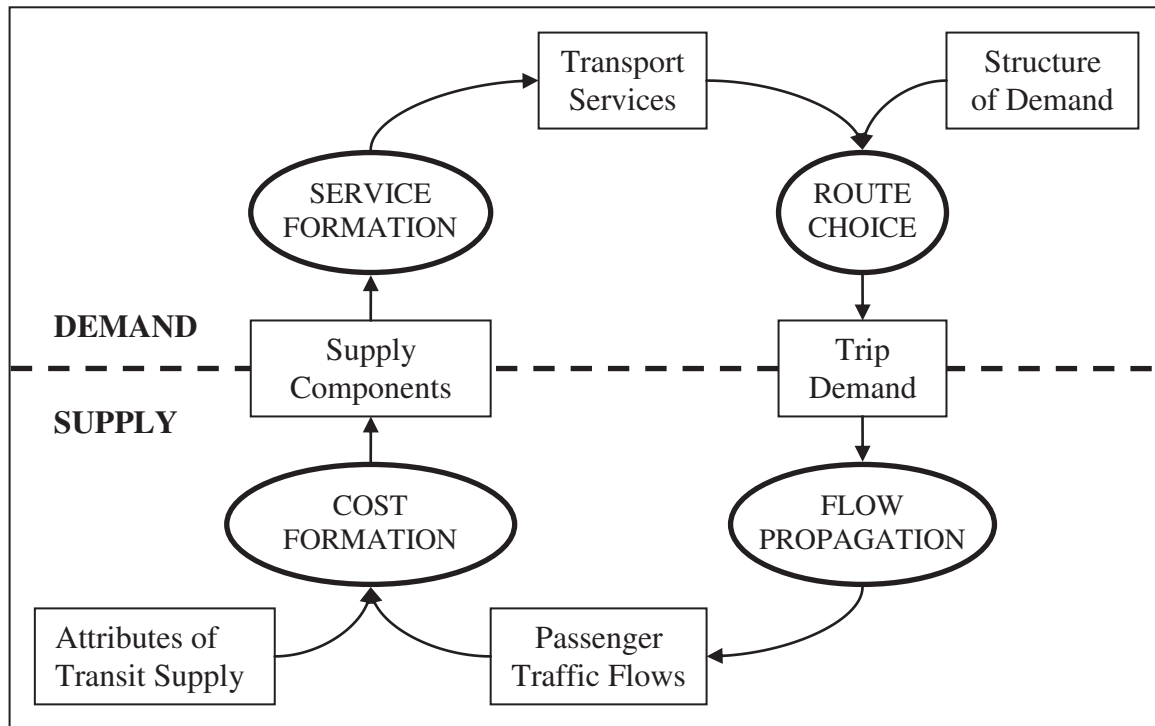


Figure 24: Systemic Representation of the Variables (boxes) and Relationships (ellipses) of a Network Traffic Assignment Model

The boxes represent the main variables of the traffic assignment model. The following six variables are distinguished:

- *Attributes of Transit Supply:* They are defined as the characteristics of each component of the transit supply. These are mainly the travel time and capacity of each element, along with its frequency, in the case of a frequency-based approach.
- *Structure of Demand:* It corresponds to the volume of trips demanded for each Origin – Destination (OD) pair.
- *Supply Components:* They refer to the quality of service, the perceived travel times and the service frequencies, under the influence of the passenger traffic flows.

- *Transport Services:* They include the choice-set available to the transit users and the attributes (cost, etc.) of the available paths.
- *Trip Demand:* It includes the passenger traffic flow assigned to the supply network for each origin – destination pair.
- *Passenger Traffic Flows:* They are the aggregated passenger traffic flows on each element of the transit network.

The ellipses in Figure 24 correspond to the models manipulated at each stage of the traffic assignment. Each model represents a causal relationship between the input and the output variables. Four relationships are distinguished. They are defined as follows:

- *Service Formation:* For each transit user, a service is made up of a route, or a set of routes, with a given quality of service on the basis of the service characteristics of the arcs and nodes the service is composed of.
- *Route Choice:* Each user is considered as a rational economic agent who chooses the optimal service in order to minimize his perceived travel time. For each user, the route choice is modelled as an economic trade-off between quality of service and travel time. The output of the model consists in assigning the users to specific transport services.
- *Flow Propagation:* The passenger traffic flows are assigned to the supply network elements.
- *Cost Formation:* For each network element, the volume of passengers who wishes to traverse it faces the capacity constraints. That yields a quality of service and the operation characteristics of the element, on the basis of the traffic flows.

These models are placed within a loop: each model modifies certain state variables, which serve as input to the following model. The network traffic assignment stops when traffic equilibrium is reached. In Figure 24, we observe that the dotted line separates the equilibrium loop in two sides, in economic terms: they are expressed as the demand-side and the supply-side of the traffic assignment. On the demand-side, the demand for transport is assigned to the transport services on the basis of a fixed quality of service and given operational

characteristics. If \mathbf{g} is the vector of perceived travel time, φ that of the service frequencies and \mathbf{x} the vector flows of the elements, the demand-side corresponds, in the CapTA model, to the mapping $(\mathbf{g}, \varphi) \rightarrow \mathbf{x}$. On the supply side, the perceived cost and service frequencies of the arcs are calculated on the basis of the flow vector, as the inverse mapping $\mathbf{x} \rightarrow (\mathbf{g}, \varphi)$. An equilibrium, expressed as a fixed point problem is reached.

6.3 The Bi-layer Representation of Supply and Demand in CapTA

A fundamental characteristic of the CapTA model is the bi-layer representation of the transit network. Two superposed layers are defined, the upper and lower layer; each one with the appropriate representation of the supply of transit services and of the demand for transport.

On the one hand, the upper layer is used by the demand-side components, for the estimation of the passenger flows on the entire network on the basis of the travel conditions (quality of service and operational characteristics).

On the other hand, the lower layer expresses the supply-side. There the supply network is represented in a higher detail in order to realistically capture the effect of the capacity constraints to the passenger flows and the travel conditions. That allows adapting the lower layer network representation to the needs of the local models used. Thereafter, the supply and demand representations of the upper and lower network are further detailed.

6.3.1 The Representation of the Transit Supply

The upper layer network $G = (N, A)$ is comprised of a set N of nodes i and a set A of arcs $a \approx (i, j)$. In the case of CapTA, the arcs on the upper layer are called network legs. They correspond to specific trip segments, within each one of the network's subsystems. Hereby three types of network legs are distinguished, per sub-system: a line leg, a station leg and an access-egress leg. They reflect different segments of a trip of an OD pair. Their cost corresponds to the average travel conditions faced by a passenger.

The line leg is the basic network element used in the CapTA model. It expresses a single trip, from boarding to alighting in a transit line. Its main attributes are related to the average

generalized cost and the combined frequency, as evaluated by the line model, described in Chapter 4. In CapTA, the set of line legs is noted A_L .

The station leg describes a user's trip inside a transit station from an entrance to a station platform or vice versa, or from platform to platform when a passenger is transferring. There the passenger is subject to discomfort and delay due to the capacity constraints inside the station. These are considered in the station model, described in Chapter 5. The set of station legs of the upper layer is noted A_S .

The access-egress leg represents the trip to access to or egress from the transit network. It corresponds to the trip from the origin until entering the transit network – a station or a stop – and the opposite trip to reach one's destination, by private mode (pedestrian, car as driver, car as passenger, bicycle, etc). An access-egress model, currently not developed in the CapTA model, can address more realistically the travel conditions and the capacity constraints related to these modes (e.g. road traffic) and the intermodal facilities (such as the capacity and pricing of park and ride facilities and bike shelters). The set of access-egress legs is A_R .

On the upper layer, the service network is $A \equiv A_L \cup A_S \cup A_R$. Therefore, a trip from an origin to a destination is composed of a succession of network legs: line legs, station legs and access-egress legs, as illustrated in Figure 25.

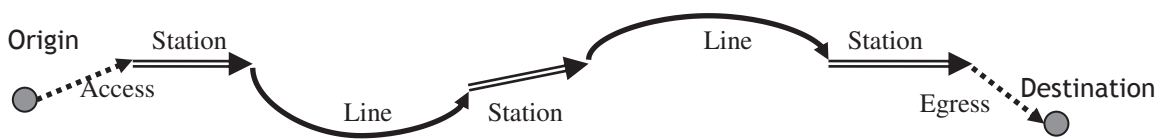


Figure 25: An origin - destination trip on the upper layer

On the lower layer, the representation of the supply is much more detailed. Various attributes are included. They are related to the travel time, the capacity (seat and total capacity of a vehicle, etc.). That representation is specific to each system sub-model and independent from the other models. For instance, a line is represented as an independent sub-network, composed of boarding, alighting, interstation and sojourn arcs. A station is represented by platform arcs, horizontal and vertical circulation elements etc. The supply representation of the lower layer of the line and station models is thoroughly described in chapters 4 and 5 respectively.

6.3.2 The Representation of the Transport Demand

On the upper layer the demand is represented by journeys effectuated from an origin to a destination node. The set of destinations s is noted S and that of the origins o is noted O . We associate to each destination a set W^s of origin – destination (OD) pairs $w = (o, s)$ with an OD flow q_{os} . Therefore a vector of OD flows $\mathbf{q} = [q_w]_{w \in W}$ is associated to the set $W = \bigcup_{s \in S} W^s$. The vector \mathbf{q} corresponds to the OD matrix and is considered fixed, or the results of a travel demand model.

On the lower layer, the demand is represented in a similar way, with trip origins and destinations. For each specific sub-network, the origin and destination nodes correspond to its boundaries: the platforms for the line model; station entries and exits and platforms for the station models. However the specific OD matrix is not fixed, but it is rather a result of the passenger flows of the specific network legs. That top-down relationship, where the flows of the upper layer form the origin-destination matrix of each independent sub-network of the lower layer, guarantees the coherence between the sub-networks of the transit network.

6.4 Demand-Side Model Components

The demand-side model components focus on the behaviour of the passengers. On the basis of the costs of the trip elements, various choices are made available to the passengers. In the case of the transit network, these choices concern the line or set of lines to board, the access and transfer stations, with respect to the passenger's knowledge of the transit network, the information that is available to him and his travel preferences. Hereby, the hyperpath concept that serve as a basis for the route choice on the upper layer is presented.

6.4.1 Strategies and Hyperpaths

A trip from an origin to a destination is represented as a path, r , on the transit network $G = (N, A)$. Each available path yields a generalized cost g_r , built sequentially according to the local generalized cost of the network element; node or link or service. A passenger is modelled as a rational economic agent with cost minimization behaviour. Therefore, at each decision point there is a set of rules, which by applying them he can get to the destination.

This corresponds to the concept of strategy, introduced in Spiess and Florian (1989). The optimal strategy is defined as the strategy of minimum cost. The random arrival of the vehicles at the station implies that a passenger's choice is not confined to a single service. Rather he makes a selection of a set of potential services, called the attractive set, so that the expected travel time is reduced. The mechanism suggests that once in a station, a passenger fixes a set of attractive services and boards the first arriving vehicle among them. Despite of the randomness of the transit supply, the route choice behaviour is assumed to be deterministic. Although some existing stochastic path choice models are used, they do not necessarily correspond to behavioural attributes.

Since the transit network is expressed as a directed graph $G = (N, A)$, the optimal strategies on the upper layer can be also expressed by a graph theoretic approach, such as the hyperpaths introduced in Ngyuen and Pallotino (1988). A hyperpath h from a node n to a destination node s is an acyclic subgraph $\hat{h} \subseteq A$ with at least one arc. The behavioural assumption for the hyperpath concept is limited to the pre-trip choice on behalf of the passengers of the attractive lines at each node.

The hyperpath $h = (\hat{h}, \hat{h})$ from a node i to a destination s – where node s does not have any successors – is defined by a pair, the arc set $\hat{h} \subseteq A$ and the routing field \hat{h} . The arc set \hat{h} is composed in such way that each arc $a \in \hat{h}$ belongs to a positive path of the set towards s , while no oriented cycles are included. The routing field \hat{h} is a mapping of the set A on $[0,1]$. Let \hat{h}_a be the flow proportion of the tail node i of an outgoing arc, $a \approx (i, j)$, of the hyperpath h . If A_i^+ is the forward star of node i , it holds that $\hat{h}_a = 0$ if $a \notin \hat{h}$ and $\sum_{a \in A_i^+} \hat{h}_a = 1$ for all nodes m of the hyperpath between n and s , except for the destination node s .

Let us define as $R_{i(s)}(h)$ the set of elementary paths of the hyperpath from i to s and $r \in R_{i(s)}(h)$ an elementary path, composed of a series of arcs between i and s . Therefore, the proportion of the flow \hat{h}_r of the route r from i to s will be the product of the flow proportions \hat{h}_a of the arcs $a \in r$, $\hat{h}_r = \prod_{a \in r} \hat{h}_a$. Let $1_{\{a \in r\}}$ an indicative function equal to 1 if

the arc a traverses the path r and 0 otherwise. The path flow proportion can be transformed into:

$$\hat{h}_r = \prod_{a \in A} \hat{h}_a 1_{\{a \in r\}} \quad (24)$$

Therefore a hyperpath can be interpreted as a set of paths whose flow proportions can be calculated directly from the routing field of the hyperpath. The flow of a hyperpath can be easily translated into an arc flow vector. That is fundamental characteristic, since that relationship between the path and the hyperpath implies the direct transformation of path-based models and algorithms to hyperpath-based ones.

The cost of a hyperpath, $\tau_{i(s)}^h$, is calculated according to the optimal strategies (Spiess and Florian, 1989), on the basis of the cost of the downstream arcs and nodes. Its mathematical formulation follows at Chapter 7.4.3. Nonetheless, in that case, the cost formation requires an additive cost structure. The non-additive costs of some congestion phenomena (such as in-vehicle comfort) can be lifted with the adoption of a leg-based network representation, such as the one in the CapTA model.

6.4.2 Service Formation and Route Choice

The service formation and route choice steps of the assignment model are used for modelling the passenger behaviour described previously. In the case of traffic assignment models with explicit enumeration of the available paths, such as some schedule-based models, a distinct service formation model is essential for the evaluation of the costs of the available paths before the route choice model. However, in frequency-based models, such as in CapTA, the hyperpaths are implicitly enumerated within the route choice model. The result is a hypertree, which is the combination of the hyperpaths for all the origins of the network to a single destination.

The CapTA model follows the hyperpath framework developed previously and addresses route choice in both network layers. On the lower layer, the line model handles the route choice, in a simple way. Within the line, a passenger at an access station i boards the first vehicle with available vehicle capacity that serves his egress station s . Thus the service routes are bundled together. That is realistic, especially so when the vehicles do not overtake one another.

The route choice on the upper layer is associated with the optimal strategy and the route choice can be expressed as a hyperpath. That happens because in CapTA the hyperpath formation is sensible to the fictive frequencies on the line legs. The fictive frequency $\hat{\phi}_a$ is a mathematic artefact (see Chapter 7.4.2) that allows passing from a stochastic waiting at uncongested traffic due to the random arrivals of the transit vehicles, to the deterministic waiting of a passenger stock if the service is saturated. The former traffic state assumes availability related to the frequency of vehicle arrivals. The latter is reduced to an arc with continuous availability, similar to a pedestrian arc. Furthermore, the fictive frequency of an unsaturated line leg depends on the passenger flow vector \mathbf{x} , via the restrained frequency and transit bottleneck models intervening locally at the line model. A corollary result concerns the complications to address the traffic equilibrium since the feasible set of the hyperpaths is not a finite topological space.

The mathematical treatment of the strategies and the characterization of the equilibrium are further developed in Chapters 7.3 and 7.4. The route choice model composes the shortest hypertree by calling an adapted optimal strategy algorithm, described in Chapter 8.4.1.

6.5 Supply-Side Model Components

The supply-side addresses the assignment of the passenger flows to the elementary components of the transit network, with respect to the passengers' route choice and the formation of the cost of these components, on the basis of the passenger flows.

6.5.1 Volume Loading and Flow Propagation

In CapTA, the route choice model assigns the passenger flows to the shortest hyperpath per destination with a relaxation of the capacity constraints. The volume loading and flow propagation steps assign the passenger flows specifically to the network elements. This step is embedded to the hyperpath framework, introduced previously. Indeed, the route choice model considers each destination separately and builds the shortest hypertree, in other words the combination of the hyperpaths from all the origins $o \in O$ to a single destination s . Therefore, the passenger flows are assigned to the upper network elements, on the basis of the origin-destination flows and the flow proportions of the arcs that compose the hyperpath. That is

achieved by a specific algorithm, described in chapter 8.4.2. An auxiliary arc flow state $\hat{\mathbf{x}}_A$ of the assignment comes from the accumulation of the arc flows x_{as} , for all destinations $s \in S$.

The Method of Successive Averages is used to solve the traffic equilibrium problem. That method implies that the arc flow vector \mathbf{x}_A^μ of an iteration μ stems from the convex combination of the arc flow vector $\mathbf{x}_A^{\mu-1}$ of the previous iteration and the auxiliary arc flow vector $\hat{\mathbf{x}}_A$. The combination step comes from a predefined decreasing sequence of positive numbers $(\lambda_\mu)_{\mu \geq 0}$, with initial value $\lambda_0 = 1$.

6.5.2 Cost Formation

The cost formation is a central element of the CapTA model. Specific sub-models are used for each type of sub-network. A twofold relationship is established by these specific models between the two network layers. Top-down, the passenger flows of the network legs determine the OD flows of the sub-networks. Local models are applied for the evaluation of the physical conditions of the traffic flows under strict capacity constraints. Bottom-up the cost of the network legs is evaluated with respect to the cost of the lower layer service elements. In other words, these specific models amount to elaborate cost-flow relationships for a set of network legs. In addition, a network leg cost according to the preferences of the classes of transit users (especially when pedestrian circulation and perception of comfort are addressed) can be evaluated.

The modularity of the modelling framework of CapTA allows the development of more than one variants of the system sub-model, applied in different circumstances. Indeed, the CapTA network model calls two types of line models. The main variant of the line model addresses the capacity constraints with particular local models (the in-vehicle comfort, the transit bottleneck and the restrained frequency model) for suburban rail, metro, light rail services and segregated bus services, as described in Chapter 4. Nevertheless, an additional variant tackles specifically the simple bus services. Their cost evaluation is limited to the in-vehicle comfort. That does not imply a loss in the model's realism, since in the Greater Paris transit network the bus services are generally complementary to the structural network and only occasionally passengers face difficulties at boarding due to insufficient capacity. On the contrary, it allows addressing each specific mode in the appropriate context. The bus services, mixed in general

traffic, would require an additional model able to address the influence of the general traffic on the operational characteristics, different from the one developed in the dissertation for rail-services with exclusive right-of-way.

6.6 The Hierarchy of Models in CapTA

In CapTA we deal with the specific stages of the transit assignment by calling various models. These are structured in three hierarchical levels: the network level, the specific system level, and the local level. They have different functions, as summarized in Table 7.

Table 7: Model hierarchy and their characteristics

MODELLING ELEMENT	NETWORK REPRESENTATION	TASKS AND DESCRIPTION OF THE MODEL
Transit Network	(Upper Layer) Network Legs: {Line leg; station leg; access/egress leg}	<ul style="list-style-type: none"> - Network flow assignment with relaxation of capacity constraints - Strategies composed by network legs
Network Sub-system	<ul style="list-style-type: none"> - Specific Network legs of the sub-system - (Lower Layer) Service links of the specific system 	<ul style="list-style-type: none"> - Manages the transition between the network model and the local models - Coordinates the local flowing and the cost evaluation models; - (line model) Uses the chronological ordering of the operations
Local Elements	Service Links of the lower-layer network	<ul style="list-style-type: none"> - Local flowing with capacity constraints - Evaluates the impact of local route share and operational characteristics on the cost of network elements

The network model addresses the transit system as a single entity. It provides the general framework for the network traffic equilibrium and focuses on the behavioural characteristics of the passenger demand for transit services. On that level a flow assignment is performed with relaxation of the capacity constraints. The route choices of the passengers are based on the node strategies, with respect to the travel conditions and the operational characteristics of the network legs (line, station and access-egress legs). The graph representation corresponds

to the hyperpath framework. This level assures the coherence of the results of the specific system sub-models, which are evaluated independently.

The system sub-models act as elaborate cost-flow relationships of the network legs and take a central part on the CapTA model. They treat each sub-system separately and manage the transition between the upper and lower layer. In addition, a system sub-model stands as a coordinator of the local models, according to each system's particularities. In the case of the line model a transit line is addressed in a disaggregated level by transit route. These are coordinated at each transit stop.

Finally, the local models address specific phenomena at the local level. For that purpose the service representation is very detailed. There, the local flow is faced to strict capacity constraints. They are developed so as to apprehend the impact of the passenger flows on the local service elements and the route share: the line model at a given station calls the in-vehicle comfort, the transit bottleneck and the restrained frequency models depending on the model variant used.

6.7 The Modelling Differences of the CapTA Model and Some Development Perspectives

The CapTA model was briefly described in the previous sections. Many differences can be identified from other transit assignment models on the scientific literature, as discussed in Chapters 2 and 3. More particularly, this model includes many innovations, linked to the modelling of the capacity constraints on the transit system. A fundamental difference of the CapTA model and other research models is the representation of the transit network. The transit assignment models generally adopt a single layer network representation for the cost formation and the flow assignment. On the contrary, a bi-layer representation is introduced in this model with the appropriate rules for the transposition between the upper and lower layers. The upper layer corresponds to a leg-based representation, as in De Cea and Fernandez (1989). The leg-based representation is not limited to the transit lines, but it can be extended to the in-station circulation and the access-egress trips. Therefore, at the network level, the CapTA model yields passenger flows, on the basis of flow dependent legs throughout the phases of the transit trip from origin to destination.

The physical conditions of a trip are handled on a detailed lower layer. There, the line model handles the combined effect of local route choice and waiting time due to congestion. Contrary to the effective frequencies function in Cepeda et al (2006), an increase in the waiting time takes place only when at least one transit service reaches capacity. The increase in the waiting time is related to the traffic regime, as described in Chapter 4.4.2.1.

However, the passenger flow greatly influences the discrete temporal availability of the transit services, especially so during congested regimes. The CapTA model allows adjusting the service frequencies (operating or fictive) on both layers. These adaptations modify the established hyperpath framework and the concept of optimal strategies and are further discussed in Chapter 7. The model does not make use of the standard formulation of the hyperpaths. Indeed, at the decision points the model makes use of the fictive frequencies coupled with the frequency attenuation factor (β_a). This treatment complies with the passenger behaviour at the decision points (stations) where congestion replaces waiting with queuing. Further influence of the real-time information on the route choice can be included in the station model.

Moreover, the CapTA model assigns passenger flows on both layers. On the one hand, on the lower layer, strict capacity constraints are applied and the flow is conserved on the vehicle level, since passengers interfere with service frequency. On the other hand, the upper layer deals with passenger flows aggregated on the line level and an average generalized cost, with a relaxation of the capacity constraints. In these terms, the lower layer yields an average cost of a trip for an individual passenger and the upper layer flow assignment is reduced to an unconstrained one with respect to the cost of the lower layer. The choice of the relaxation of the capacity constraints on the flow assignment of the upper layer and the monotone cost function, as a result of the line model, facilitates the definition of an optimization problem and the characterization of the traffic equilibrium. Although this approach guarantees a mathematical coherency, the formation of a bottleneck will keep a part of the passenger flows from being propagated downstream.

The bi-layer network representation combined with the selected modelling structure and the frequency adaptations are significant innovations of the CapTA model, compared to existing models. While not all the issues on the research community are addressed, that approach opens new perspectives on transit assignment modelling under capacity constraints. It is

possible to include specific models for in-station circulation (the station model developed in Chapter 5) and the access/egress journey. The mathematical properties of the cost-flow relationship resulting from the station model seem to be obvious, especially since a Dijkstra shortest path algorithm is selected with monotone local cost functions. However, some issues of continuity and differentiability have to be dealt with specifically. Equivalent treatment has to be addressed when developing an access/egress systemic model.

Spiess and Florian (1989) introduced the optimal strategies with specific hypothesis on passenger behaviour, based on unconstraint conditions. Indeed, without congestion the model addresses the discrete availability of the transit services by assuming that a passenger boards the first arriving vehicle among the attractive lines. The presence of congestion and queuing modifies the availability of certain services and influences the local route choice and the waiting time. More importantly, mingling on the platform creates a buffer time for accessing the transit services. This issue is handled in CapTA by adopting a fictive frequency $\hat{\phi}$ on the upper layer line legs, yielding a continuous service with the presence of congestion.

Although the passengers are generally assumed to have a perfect knowledge of the travel conditions, this is rarely the case within transit systems, especially so when they are characterized by a significant travel time variability. The passengers' behaviour depends on the presence of real-time information at the decision points and on mobile devices. With the explosion of intelligent transportation systems it is essential to handle the effect of travel information in transit assignment models, especially in conjunction with transit congestion and service reliability. That is treated partly in the station model by explicitly modelling the decision points and in the network model by some adaptations on the route choice.

Equally, the variability of the travel time and the travel conditions can be dealt from the CapTA model, as suggested above. In addition to the service reliability and the stochastic aspect of the line operation, the presence of different comfort states during the trip (in-vehicle, on-platform etc.) has a significant influence on the model. That can be typically interpreted by the problem of transfer between parallel services. There, the passengers with the same origin destination trip make a choice of transfer station on the basis of their actual and expected future travelling conditions.

Chapter 7:

Mathematic Characterization of the CapTA Model

7.1 Introduction

Although particular methods have been developed for transit, a flow assignment model remains an optimization problem subject to a number of physical, economic and mathematical constraints. This chapter is dedicated to the mathematical characterization of the CapTA model, presented in Chapter 6. Contrary to other studies, this research is focused on the systemic aspect of the transit network, together with the physical and economic analysis of the passenger flow problem.

A mathematical treatment is pursued hereby, for establishing the mathematical properties of the optimization problem. Rather than redefining an optimization problem and demonstrating the existence and uniqueness of the equilibrium solution, we develop the relevant mathematic formulation for the CapTA model and provide the connexion with already proven existence and uniqueness theorems.

The mathematical treatment reflects the physical conditions and the economic relations addressed in the flow assignment. The priority of the CapTA model is given in the exhaustive representation of the public transportation system without restraining its representation. That requires a thorough mathematical treatment of the leg cost-flow relationship, to ensure the necessary mathematical properties. Furthermore, on the basis of the optimal strategies (Spiess and Florian, 1989) the traffic equilibrium is formulated as a twofold vector of arc flows and strategy shares and it is characterized as a variational inequality problem

The chapter is divided into 4 sections. First, a detailed review of the optimization models and of the characterization of the equilibrium for various transit assignment models is presented (section 7.2). In section 7.3, the interactions between supply and demand in the CapTA model are meticulously established and more precisely the dependencies of the state variables throughout the model. Further mathematical treatment of these interactions and the formulation of the traffic equilibrium are pursued in section 7.4.

7.2 A Literature Review on the Mathematic Formulation and the Characterization of Traffic Equilibrium

The literature review is concentrated in specific models and issues relevant to the characterization of the network equilibrium. In particular, we address the optimal strategy concept, the hyperpath framework and selected congestion-based variable frequency and stochastic transit assignment models. The formulation of the network equilibrium is analysed in detail, faced to the local route choice problem and the presence of capacity effects. Some necessary algorithms are included.

7.2.1 The Optimal Strategies Approach

The Optimal Strategies approach is introduced in Spiess (1984) and formally presented in Spiess and Florian (1989) as a method to link the effect of the waiting phenomena to the selection process of routes on a transit network (see Chapter 2.3.2). According to the authors, a strategy is defined as a set of rules that, when applied, allow the transit user to reach his or her destination. The cost of each arc $a \in A$ of the network $G = (N, A)$ is decomposed on a nonnegative constant travel time t_a and the cumulative density function V_a for the waiting time. The latter depends on the relation between the distribution of the vehicle headways for the arc a and the distribution of passenger arrivals on node i .

7.2.1.1 The mathematical formulation of the local route choice

The concept of strategy suggests that from each node i on the network we can specify a nonempty set of attractive lines. Therefore, if A_i^+ is the forward star of node i , the local

strategy suggests that there is a subset $\bar{A}_i^+ \subseteq A_i^+$ of outgoing arcs that take part on strategy \bar{A} . The subset \bar{A}_i^+ corresponds to the set of attractive lines and implies that a passenger at i will board the first vehicle of any of these lines.

With the construction of the attractive set of outgoing arcs at i , \bar{A}_i^+ , we can establish the combined expected waiting time w_i and the arc proportion $\hat{h}_a, \forall a \in A_i^+$. By definition, $\hat{h}_a = 0$, if $a \notin \bar{A}_i^+$. The formulations of the expected combined waiting time $w_i(\bar{A}_i^+)$ and the probability that arc a is served first, $\hat{h}_a(\bar{A}_i^+)$ can be therefore expressed via the waiting time distribution V_a . For matters of simplicity, let us associate the waiting time with the frequency φ_a of the services and a positive parameter α , whose value is related with the waiting discipline and passenger arrival distribution and the distribution of vehicle headways. Therefore, the expression takes the form:

$$w_i(\bar{A}_i^+) = \alpha / \sum_{a \in \bar{A}_i^+} \varphi_a \quad (25)$$

$$\hat{h}_a(\bar{A}_i^+) = \varphi_a / \sum_{a \in \bar{A}_i^+} \varphi_a, \forall a \in \bar{A}_i^+ \quad (26)$$

We denote x_i the sum of the volumes of all incoming arcs and the demand at that node to destination s , $x_i = \sum_{a \in A_i^-} x_a + q_{is}$. The node volume is distributed to the outgoing arcs, according to the arc proportions:

$$x_a = \hat{h}_a(\bar{A}_i^+) \cdot x_i \quad (27)$$

The algorithmic treatment of the local route choice – by extension the attractive subset of lines – is similar to the greedy algorithm proposed in Chriqui and Robillard (1975), adapted to reflect the optimal strategies characteristics. We denote $\tau_{i(s)}$ the cost from the node i to destination s and g_a the cost for traversing the arc a . For a node i , we examine the outgoing arcs $a \approx (i, j)$, where $a \in A_i^+$, ordered by increasing cost to destination, $g_a + \tau_{j(s)}$. If that cost is lower than the current label, $\tau_{i(s)} \geq g_a + \tau_{j(s)}$ we add the arc to the attractive subset of outgoing arcs, $\bar{A}_i^+ \leftarrow \bar{A}_i^+ \cup \{a\}$, while updating the node cost-to-destination in respect to the arc frequencies, $\tau_{i(s)} \leftarrow [\varphi_i \tau_{i(s)} + \varphi_a (g_a + \tau_{j(s)})] / (\varphi_i + \varphi_a)$ and the node

frequency of outgoing arcs, $\varphi_i \leftarrow \varphi_i + \varphi_a$. The algorithm updates the node label a maximum of $|A_i^+|$ times, where $|\cdot|$ denotes the cardinal of the set. In Spiess and Florian (1989) the local routing algorithm is integrated to the optimal strategy algorithm. It examines all the nodes and evaluated the cost to a single destination, based on the Dijkstra shortest path algorithm.

7.2.1.2 The formulation of the network equilibrium model

In their influential paper, Spiess and Florian (1989) formulate an optimization problem as an extremal problem, on the basis of constant arc costs, without capacity effects. The objective function developed includes the constant travel time on the arcs and the expected combined waiting time expressed on the nodes. Let δ_a be a dual variable of the optimal strategy and x_i the exit flow from node i , the objective function (28) is subject to non-negativity constraints (28).c, conservation of flow (28).b and local route choice by node (28).a. The minimization problem takes the following form:

$$\min \sum_{a \in A} g_a x_a + \sum_{i \in N} \frac{x_i}{\sum_{a \in A_i^+} \varphi_a \delta_a} \quad (\text{P}) \quad (28)$$

Subject to:

$$x_a = \frac{\delta_a \varphi_a}{\sum_{a' \in A_i^+} \varphi_{a'} \delta_{a'}} \cdot X_i, \quad a \in A_i^+, \quad i \in N \quad (28).a$$

$$X_i = \sum_{a \in A_i^-} x_a + q_{is}, \quad i \in N \quad (28).b$$

$$X_i \geq 0, \quad i \in N \quad (28).c$$

$$\delta_a = \begin{cases} 1 & \text{if } a \in \bar{A} \\ 0 & \text{otherwise} \end{cases} \quad (28).d$$

We observe that this minimization problem has a non-linear objective function and non-linear constraints. It is possible to reduce this problem to a linear programming problem (LP), via some simple transformations. Indeed by setting $\omega_i = x_i / \sum_{a \in A_i^+} \varphi_a \delta_a$, it can be expressed as the following linear programming problem:

$$\min \sum_{a \in A} g_a x_a + \sum_{i \in N} \omega_i \quad (\text{LP}) \quad (29)$$

Subject to:

$$\sum_{a \in A_i^+} x_a - \sum_{a \in A_i^-} x_a = q_{is}, \quad i \in N \quad (29).a$$

$$x_a \leq \varphi_a \omega_i, \quad a \in A_i^+, \quad i \in N \quad (29).b$$

$$x_a \geq 0, \quad a \in A \quad (29).c$$

The previous formulation corresponds to the extremal formulation of the primal problem. In Spiess and Florian (1989) the authors examine the dual problem. Let μ_a a variable associated with the expected combined waiting time. The dual problem (D) can be formulated as:

$$\max \sum_{i \in N} q_{is} \tau_{i(s)} \quad (D) \quad (30)$$

Subject to:

$$\tau_{j(s)} + g_a + \mu_a \geq \tau_{i(s)}, \quad a \in A \quad (30).a$$

$$\sum_{a \in A_i^+} \varphi_a \mu_a = 1, \quad i \in N \quad (30).b$$

$$\mu_a \geq 0, \quad a \in A \quad (30).c$$

The authors use three algorithmic proofs, based on the solution algorithm of the section 7.2.1.3 to demonstrate that $\bar{\tau}$ is a feasible solution of the dual problem (D), that \bar{x} is a feasible solution of the primal problem (LP) and finally, that the couple $(\bar{\tau}, \bar{x})$ is the optimal solution, since the weak complementarity slackness conditions of the primal-dual problem are satisfied.

In Spiess and Florian (1989) two solutions are proposed to address specifically the walking arcs, since they are not subject to the waiting phenomenon. The first approach lies in assuming that a walking arc is identical to the other arcs with waiting though its frequency is equal to a sufficiently large number that produces a combined waiting time close to zero. The second approach treats the walking arcs as complementary elements without waiting, resulting in some necessary adaptations on the optimization problem and the proofs.

7.2.1.3 The solution algorithm of the network equilibrium

In Spiess and Florian (1989) two algorithms are described for determining the optimal strategies and the flow loading per destination. The optimal strategies algorithm is a Dijkstra shortest path algorithm, adapted to integrate the local route choice at the stop level. It begins in a reverse order, from the destination s and examines all the network arcs in increasing

cost-to-destination $g_{(i,j)} + \tau_{j(s)}$ order. The attractive set consists then of an ordered list of the arcs by ascending order of cost to destination.

The flow-loading algorithm is based on the flow conservation at the nodes and the previous labelling procedure. For the ordered attractive arc set, each arc is treated in decreasing order. The node flow x_i is assigned to each outgoing arc of the attractive subset $a \in \bar{A}_i^+$, in respect to the arc frequencies.

Spiess and Florian (1989) demonstrate that the complexity of these algorithms is subject to the optimal strategy search. If $|A|$ is the cardinal of the set A , the complexity per destination s amounts to $O(|A|^2)$ and can be simplified to $O(|A|\log|A|)$ if a heap is used.

7.2.1.4 Optimal Strategies with flow-dependent arc costs

As an extension of the extremal formulation of the optimization problem with constant arc costs, Spiess and Florian (1989) examine an alternative non-linear optimization problem with flow dependent arc costs, given by a continuous function, $g_a(\mathbf{x})$, $\forall a \in A$, based on Wardrop's first principle (Wardrop, 1952). The previous formulation of the arc cost, leads to two remarks: first, when we consider an additive cost structure on the network, where $g_a(\mathbf{x}) = g_a(x_a)$, $\forall a \in A$, there exists an equivalent convex combination problem; and second, since the arc cost is dependent to the cumulative arc flow for all destinations, it is essential to consider the optimization problem for all destinations combined.

The previous optimization problem (P) may be formulated as a variational inequality problem (VIP). If the arc cost function $g_a(\mathbf{x})$ is not strictly monotone, the uniqueness conditions of the equilibrium flows are not satisfied. Conversely, if the arc cost is defined by a continuous increasing function, the minimization problem (P) can be formulated as an extremal optimization problem, separable by destination. An adapted algorithm is proposed, which involves an additional step before the research of the optimal strategy for updating the flow-dependent arc costs.

7.2.1.5 Application of the Optimal Strategies on a leg-based network

De Cea and Fernandez (1989) develop a different network representation based on route sections. A route section is defined as the part of a transit route between two consecutive transfers, similar to the leg used in CapTA. A route section is characterized by an in-vehicle travel time corresponding to the weighted average of the lines composing it, and a waiting time. The model is extended in De Cea and Fernandez (1993) in order to include a flow-dependent leg cost, sensible to route congestion.

Various modifications on the transit assignment model are imputed to the leg-based network representations. The equilibrium flows are calculated on the basis of the route section. Therefore, the passenger flows of the transit lines are handled in the local level similarly to the common lines problem in Chriqui and Robillard (1975). The passenger flows on a route section boarding a set of lines compete with the other boarding flows and the flows on-board for the available capacity.

The network representation allows for a simplified route choice algorithm compared to the one from Spiess and Florian (1989), with flow dependent route sections. The user equilibrium problem is transformed to an equivalent variational inequality problem (VIP), and corresponds to an asymmetric network assignment problem. De Cea and Fernandez (1993) propose a diagonalization method for the transformation of the vector cost function. Then, the symmetric assignment problem is solved by the Frank-Wolfe algorithm. The initialization step of the algorithm consists of constructing the attractive line set for each section.

7.2.2 The Hyperpath Framework

The variety of network representations and the need to efficiently capture the common lines problem incited Nguyen and Pallottino (1988) to develop a generic modelling framework for large-scale transit networks using graph theory. That graph-theoretic framework can be used for a number of assignment models, such as deterministic, stochastic, etc. giving some flexibility to the modelling process. In that framework, the trip from an origin to a destination can be described by a hyperpath. The mathematic formalization of the hyperpath concept is described in detail in chapter 6.4. However, the initial definition concerned a subgraph from an origin to the destination. That is extended in Cominetti and Correa (2001) for addressing any node $i \in N$ of the network to a destination.

7.2.2.1 The network optimization problem

Given the cost formulations of the hyperpath, the description of the user equilibrium follows. In Nguyen and Pallottino (1988) the user equilibrium is formulated as a variational inequality problem (VIP) on the hyperpath space, where the main constraints are the conservation of flow on the hyperpath and the non-negativity of the flows. This VIP can be transformed into an equivalent optimization problem in the space of paths, based on the arc costs and the waiting time of the hyperpath. Indeed, it resembles the optimization problem developed in Spiess and Florian (1989) for the flow-dependent arc costs.

We define the hyperpath flow state \mathbf{x}_{OS}^H and the arc flow state $\mathbf{x}_A = [x_a : a \in A]$. In addition the waiting time vector will be $\mathbf{w}_I^{Hos} = [w_i^h : i \in R_{OS}, h \in H_{OS}]$. The optimization problem in the space of the feasible hyperpath flows can be transformed into the following inequality:

$$(\mathbf{x}_A - \mathbf{x}_A^*)' \cdot g_A(\mathbf{x}_A^*) + (\mathbf{x}_{OS}^H - \mathbf{x}_{OS}^{H*})' \cdot \mathbf{w}_I^{Hos} \geq 0 \quad (31)$$

The most interesting result of the previous formulation, as noted in Nguyen and Pallottino (1988) is that the solution of the shortest hyperpath optimization problem can be identified by the arc flow state and the corresponding total waiting time. Furthermore, in the case of optimal strategies, the waiting time can be calculated directly from the hyperpath arc set \hat{h} .

7.2.2.2 Solution Algorithms

Based on the optimal strategies concept of Spiess and Florian (1989) – defining in accordance the routing proportions and waiting time in relation to line frequencies – two algorithms are described in Nguyen and Pallottino (1989) for determining the shortest hyperpaths to one destination, the Dijkstra algorithm and the Bellman-Ford algorithm.

Dijkstra Shortest Hypertree Algorithm

The authors suggest a label setting Dijkstra algorithm for determining the shortest hyperpaths from the directed graph's origins to a destination s . It is an arc-selection based algorithm, adapted to the hyperpath concept, following the Generalized Bellman's equations. After the initialization, the algorithm begins at the destination and examines the incoming arcs $(i, j) \in A_j^-$ at an increasing cost-to-destination order $g_{(i,j)} + \tau_{j(s)}$. If that complies with the

inclusion condition, $\tau_{i(s)} > g_{(i,j)} + \tau_{j(s)}$, the arc is added to the hypertree and the node labels are updated. In the case of a stop node, multiple outgoing arcs are possible and the node labels are then updated. If the arc list is implemented as a heap, the algorithms complexity for each destination is equal to $O(|A|\log|A|)$.

Shortest Hypertree Algorithm

An alternative to the Dijkstra algorithm is described in Nguyen and Pallottino (1989), based on the label correcting Bellman-Ford algorithm. After the initialization, the algorithm begins from the destination: at each node i , the node is removed from the selection list and the outgoing arcs $(i, j) \in A_i^+$ are examined at an increasing arc cost order, $g_{(i,j)}$. Hence, the optimal outgoing arc set is established on the basis of the current node costs. The algorithm examines then the incoming arcs $(k, i) \in A_i^-$. The nodes k (tail of arc (k, i)) are added to the selection on a cost descending order.

Therefore, the algorithm revisits each node until it produces the minimum cost to the destination. Based on the non-negativity of arc costs, at every pass at least one node will be updated to its minimum cost. The algorithm's complexity for one destination is $O(|A||N|\log|N|)$. In Nguyen and Pallotino (1988) the authors estimate that the label correcting algorithm is more flexible than the Dijkstra shortest hypertree algorithms but lacks in calculation speed.

7.2.3 Congestion-Based Variable Frequency Models

Beyond the equilibrium assignment and the solution algorithms based on constant arc costs, further research focused on the local route choice and characterization of the network equilibrium of variable frequency arcs. In this line of research where both local routing probabilities and the waiting time depend on the flow-dependent line frequencies, Cominetti and Correa (2001) provide extensive mathematical treatment for the local route choice and the characterization of the network equilibrium based on flow-dependent frequencies. In addition, they propose a Dijkstra-based solution algorithm. On the basis of the previous work Cepeda et al (2006) reformulate the local route choice and define a duality gap. Moreover, they develop a consistent solution algorithm for passenger flow assignment on large-scale networks.

7.2.3.1 The consequences of the frequency variation

As a way to establish a direct link between transit congestion and travel conditions on public transportation, the frequency of the arcs can be defined as a function of the vector of the passenger flows, $\varphi(\mathbf{x})$. That can have some important consequences, on the physical and economic grounds and on the mathematical treatment of the equilibrium.

Earlier research (Chriqui and Robillard, 1975; Spiess and Florian, 1989; Ngyuen and Pallottino, 1988) established the arc frequency as the determinant of the local route assignment and the combined waiting time. Therefore, a flow-dependent frequency characterized by a decreasing function, would alter the local route choices, reducing the flow on the most congested lines. Furthermore, it entails increased waiting time for the line and increased cost on the local strategy.

The consequences of a variable arc frequency on the mathematical characterization of the network equilibrium are considerable, especially for a hyperpath-based supply representation. Indeed, the extremal formulation of the optimization problem in Spiess and Florian (1989) guarantees the existence of an equilibrium, which is unique, when the arc frequencies and the arc cost (or link travel time) are considered constant. When variable frequency and arc cost are considered, different mathematical tools are used for the characterization of the equilibrium, mainly by transforming the optimization problem to a Variational Inequality Problem (VIP) or a Fixed Point Problem (FPP).

7.2.3.2 The selection of the local choice set

The mathematical characterization of the attractive set of routes is an important contribution of Cominetti and Correa (2001). The effect of congestion is taken under consideration by the introduction of a flow-dependent frequency – named effective frequency – and defined by the following strictly decreasing function $\varphi_a : [0, \kappa'_a] \mapsto (0, \infty)$ with $\varphi_a(x_a) \rightarrow 0$ when $x_a \rightarrow \kappa'_a$, approaching the available capacity – or saturation flow. They further demonstrate that the arc flow can be expressed as a function of the hyperpath flows and the arc frequencies. That facilitates the optimization problem by substituting the space of hyperpath flows by the simpler space of the arc flows.

In order to overcome the frequency-based formulation of the local route choice, they introduce a variable α and a function $v(\alpha)$ as the inverse of the function $x_a \mapsto x_a / \varphi_a(x_a)$. Although variable α is interpreted as a waiting time in Cepeda et al (2006), the definition proposed implies it to make reference to the number of candidate passengers for a vehicle on line a , similar to the passenger stock n_{zi} of the transit bottleneck model (see chapter 4.3)

The total waiting time of a strategy in Cominetti and Correa (2001) is expressed as a function $w(\alpha)$ of the number of candidate passengers. An extremal nonlinear optimization problem is formulated where $f(\mathbf{x}, \alpha) = \sum_{a \in A_i^+} x_a t_a + w(\alpha)$ is the objective function and they demonstrate the existence of an optimal solution (\mathbf{x}, α) which satisfies the Karush-Kuhn-Tucker conditions. An algorithmic proof is given for demonstrating the optimality of the generation of the local strategy and the expression of the optimal hyperpath flows on the arc flow vector. The boundary conditions revealed in Cominetti and Correa (2001) should be highlighted. It is noted that the uniqueness of the solution is not assured in the case of equality of the cost-to-destination $g_{a'} + \tau_{j'(s)}$ of an arc $a' \approx (i, j')$ with the cost of the strategy $\bar{\tau}_{i(s)}$. In that case, the flow can be assigned to two strategies, although the cost is constant. Indeed, that is a distinct local route choice problem, also mentioned in Cepeda (2002), where the previous optimization problems fail to guarantee the uniqueness of the equilibrium.

7.2.3.3 The network equilibrium and the Hyperpath-Dijkstra algorithm

Cominetti and Correa (2001) express the travel time from a node i to a destination s , $\tau_{i(s)}$ as the unique solution of the generalized Bellman equations, similarly to prior research (Nguyen and Pallotino, 1988; Spiess and Florian, 1989). In addition, they define the potential cost of arc $a \approx (i, j)$ to destination s , as $t_{a(s)} = t_a(\mathbf{x}) + \tau_{j(s)}(\varphi(\mathbf{x}), t(\mathbf{x}))$, where the travel time and the frequency of the arc is sensitive to the vector of network flows. If the node exit flow per destination s is inferior of the available capacity, $x_{i(s)} < \sum_{a \in A_i^+} \kappa_{a(s)}$, we define the optimal set of the local routing problem, $E_{i(s)}(\mathbf{x}) = \{x_{a(s)}, a \in A_i^+\}$. The optimal solution of the network is $E(\mathbf{x}) = \prod_{s \in S, i \in N} E_{i(s)}(\mathbf{x})$.

An important point towards the simplification of the network representation is the formulation of the strategies per node indifferently of their origin. That is also presented in Cepeda (2002) under the name *node-strategy* and allows considering every node of the network $n \in N$ as a potential origin node. The solution of a complex network may be expressed as a superposition of the local route choices problems. A hyperpath h_{is} can be defined for any network node $i \in N - \{S\}$ – destination pair on the network. The characteristics of the hyperpaths presented in chapter 6.4.1 correspond to this general description.

A Fixed Point Problem is formulated for the characterization of the equilibrium. The authors use a set of infinite arcs (or pedestrian) in order to provide an upper bound on the problem and assure the compactness of the feasible set. If the network capacity is sufficient, these arcs are unlikely to be used. The existence of the network equilibrium is demonstrated by the Kakutani's fixed point theorem.

An alternative solution algorithm is proposed in Cominetti and Correa (2001) for building the optimal hyperpath. The Hyperpath-Dijkstra Algorithm is based on a Dijkstra shortest path algorithm, adapted for the hyperpath case. The algorithm begins from the destination s and scans all the nodes on the network. When an arc $a \approx (i, j)$ improves the cost to destination, it is added to the attractive set. However, since the arcs are not selected in an increasing time-to-destination order, a loop is included for removing the arcs with a superior travel time $t_{a(s)}$ from the updated time to destination $\tau_{i(s)}$. If $d = \max(|A_i^+| : i \in N)$, the algorithm per destination has a maximum complexity of $O(|N| \ln |N| + |A| \cdot \ln d)$.

7.2.3.4 The duality gap

Cepeda et al (2006) revisit the local route choice problem, presented by Cominetti and Correa (2001) in order to propose a duality gap function adapted to the optimal strategies concept. They interpret the variable α introduced in Cominetti and Correa (2001) as the waiting time of the local strategy and they introduce a direct link between that and the passenger flows. They use an alternative characterization of the equilibrium, based on the following equation.

$$\frac{x_a}{\varphi_a(\mathbf{x})} \begin{cases} = \alpha & \text{if } t_a < \tau(\mathbf{x}) \\ \leq \alpha & \text{if } t_a = \tau(\mathbf{x}) \\ = 0 & \text{if } t_a > \tau(\mathbf{x}) \end{cases} \quad (32)$$

Although the conditions for the inequalities can be easily proved, the case of equality $t_a = \tau(\mathbf{x})$ is not sufficiently addressed, despite the proofs provided in Cepeda (2002) and Cepeda et al (2006). In that there is not a one-to-one mapping between flow and cost. Nevertheless, by adopting the generalized Bellman equations, they extend the local route choice problem is extended to the network. For each node $i \in N \setminus \{S\}$ it holds:

$$\sum_{a \in A_i^+} [t_a(\mathbf{x}) + \tau_{j(s)}(\mathbf{x})] \cdot x_a^s + \max_{a \in A_i^+} \frac{x_a^s}{\varphi_a(\mathbf{x})} \geq \tau_{i(s)} \sum_{a \in A_i^+} x_a^s \quad (33)$$

An algorithmic proof is provided for demonstrating that the equation (33) becomes equality if the optimal strategy flows are considered. By rearranging the variables, they provide a duality gap function which is null when the optimal strategy is reached. Although that duality gap function is a useful corollary of the local route choice problem, its major difference with previous approaches is that it doesn't guarantee flow stability at equilibrium, but rather a convergence of the strategy costs.

7.2.3.5 The MSA network algorithm

A solution algorithm for the flow assignment on large-scale network is provided in Cepeda et al (2006). Cepeda (2002) analyzes the alternative solutions that were pursued. A projected generalized gradient method developed in Constantin (1993) was tested but it was left aside for several reasons: it is hard to develop an analytical formulation for the problem due to its high complexity; the proof of convergence in Constantin (1993) is based in some particularities inherent to the problem, not found in the current optimization problem; and it provided extremely slow convergence in the small network tested.

A different approach by Cepeda (2002) was to formulate the optimization problem as a quasi variational inequality problem (qVIP). In that case, the space of feasible solutions depends on the optimal arc flow vector \mathbf{x}^* . However, a solution algorithm cannot be provided, since the feasible flow state is not a polytope, nor convex in the general case.

The authors provide in Cepeda et al (2006) a simple heuristic minimization method for solving the network equilibrium, the method of successive averages (MSA). The algorithm calculates for each iteration an auxiliary state by solving the optimal strategy problem based on the network flows of the previous state. A convex combination of the previous and the auxiliary state with a predefined step guarantee the convergence. The difference of the updated state from the optimal solution is measured by the duality gap function presented previously in 7.2.3.4. An initial state can be calculated as an all-or-nothing assignment.

The heuristic algorithm developed leads intuitively to convergence, especially so since the Method of Successive Average on the vector of arc flows is used. Nevertheless, it seems that the authors do not provide a formal proof of the feasibility of the solution, but rather some indications of that. Indeed, the use of strict capacity constraints limited to the effective frequency function transforms the assignment to an unconstrained flow assignment, where it is possible for some arc flows to occasionally exceed the arc capacity. The use of the proposed function of effective frequency, or the truncated frequencies – as an alternative to the use of additional pedestrian arcs – and guarantee the existence of a feasible solution in the case of a network with sufficient overall capacity.

Additional methodological concerns arise from the assignment model. The authors provide various approaches for determining a solution algorithm, from an extremal formulation and a quasi VIP. Although they use an algorithm based on the Method of Successive Averages (MSA), they do not formulate it as a part of a Fixed Point Problem (FPP). The model developed in Cepeda et al (2006) lacks a thorough mathematical characterization of the equilibrium. However, the main objective of the authors is to provide a functional, easy to implement, assignment algorithm, which is clearly the case.

7.2.4 Stochastic Transit Assignment with Capacity Constraints

Despite of the extended mathematical treatment of the previous research, alternative approaches have been proposed as means to make the assignment models more realistic. In particular Lam et al (1999, 2002) explore the transit assignment focusing around the questions of the Stochastic User Equilibrium (SUE), the enforcement of strict capacity constraints and a variable frequency of the transit routes.

To overcome shortcomings of previous research, they adopt a network representation based on route sections, such as in De Cea and Fernandez (1993). A route section $\ell_{(i,j)}$, with direct similarities to the line leg used in CapTA, is defined between two transfer nodes and is composed of all the direct attractive lines z which satisfy the optimization problem in Chriqui and Robillard (1975). Therefore, the local route proportions within the route section correspond to the proportion of their frequencies and the waiting time is calculated by means of the combined frequencies, $w_\ell = \alpha / \sum_{z \in \ell} \varphi_z$. A path from an origin $o \in O$ to a destination $s \in S$ is composed of a sequence of route sections.

7.2.4.1 Stochastic User Equilibrium and multipath assignment

Lam et al (1999, 2002) develop a Stochastic User Equilibrium assignment model based on the principle that there exist a variation in the perception of the travel times (or costs) from the transit users. Therefore, a random variable is included in the path cost function for each of the in-vehicle travel time t_ℓ and waiting time w_ℓ . These error terms can be jointly expressed by the random error term ξ . If $\tilde{g}_{r(o,s)} = g_{r(o,s)} + \xi_{r(o,s)}$ is the passenger perceived total travel time on path $r(o,s) \in R_{os}$, it is assumed that $E[\xi_{r(o,s)}] = 0$ and $E[\tilde{g}_{r(o,s)}] = g_{r(o,s)}$.

A logit model is selected for the assignment of the origin-destination flows to the appropriate paths, according to the relative path costs. A parameter θ is used for considering the different degree of passengers' knowledge of the path travel cost. When $\theta \rightarrow \infty$, the SUE resembles a DUE. If the passenger flows on a given path $r(o,s) \in R_{os}$ is noted $x_{r(o,s)}$, the assignment of the passenger flows between alternative routes will be given by the following equation.

$$\ln(x_{r(o,s)} / x_{r'(o,s)}) = -\theta(g_{r(o,s)} - g_{r'(o,s)}) \quad (34)$$

However, the cost of the path includes an additional term $d_{r(o,s)}$, the passenger overload delay. That represents the effect of enforcing strict capacity constraints.

$$d_{r(o,s)} = \begin{cases} = 0 & \text{if } x_{r(o,s)} \leq \kappa_{r(o,s)} \\ > 0 & \text{otherwise} \end{cases} \quad (35)$$

The value of the previous penalty, $d_{r(o,s)}$, is defined implicitly for every path by the appropriate transit assignment problem with bottleneck. Indeed, contrary to other models that

define the arc flows based on the costs of the arcs on every given step of the model, Lam et al (1999) develops a model that estimates the cost of congestion, with strict capacity constraints on the passenger flows on the paths. The penalty $d_{r(o,s)}$ is defined at every step from the Lagrangian multipliers of the stochastic user equilibrium on the space of the path flows (Lam et al, 2002):

$$\text{Min} \sum_{\ell \in \bar{A}} (t_{\ell} + w_{\ell}) + \frac{1}{\theta} \sum_{o \in O, s \in S} \sum_{r(o,s) \in R_{os}} x_{r(o,s)} (\ln x_{r(o,s)} - 1) \quad (36)$$

The associated constraints are related to the conservation of flow, the link between arc flows and path flows, the strict capacity constraints on the route sections and finally the non-negativity of the path flows. The passenger overload delay is assigned to the routes using an arc section incidence vector. The solution algorithm is based on the algorithm for solving SUE problems with explicit constraints proposed by Bell (1995). That algorithm needs a complete enumeration of the paths on the network or a Dijkstra shortest path algorithm in the case of large networks. The stopping criterion of the algorithm is based on the stability of the Lagrangian multiplier.

Although the approach proposed in Lam et al (1999) leads to a fully functional algorithm, some alternatives may be considered. Indeed, it seems that the use of an augmented Lagrangian allows faster convergence to the optimal solution, without tampering with the fundamentals of the optimization problem. That alternative can provide an enhanced label update procedure for the Lagrange multipliers, one of which is linked to the passenger overload delay.

7.2.4.2 Interaction between service frequency and passenger flow

In Lam et al (2002) the authors further considered the interaction between the boarding and alighting passenger flows and the service frequency of the routes. A dwell time function of the boarding and alighting flows – therefore the passenger flow vector – is defined in Lam et al (1998). They estimate a total vehicle service time T_z of a vehicle on the transit route, composed of the dwell time T_i , the travel time between stops $t_{(i,j)}$ and the necessary turning time in a terminal, t^T . If the fleet size is constant, the vehicle service time and the number of vehicles are associated with the route frequency.

The SUE transit assignment problem presented previously is modified by expressing waiting time and capacity of the route sections as functions of the service frequencies. Therefore it can be transformed to a Fixed Point Problem (FPP) where the optimal solution (\mathbf{x}, φ) is reached with constant arc costs. The solution algorithm of the FPP lies in solving the SUE transit assignment problem based on constant frequency and then updating the service frequency based on the updated path flows and the dwell time function.

The proposed algorithm does not follow the obvious approach for this kind of problem. In fact, the use of such embedded algorithms, where a stationary state is reached for the flow assignment for every frequency update, seems computationally expensive. An alternative approach of an MSA algorithm consists of a joint calculation of the auxiliary state, first of the flow vector, based on the arc costs, and then of the line frequencies, based on the updated flow vector. That approach should reduce the complexity of the solution algorithm and provide a faster solution.

In Lam et al (2002) the frequency of the service routes is calculated explicitly for each SUE assignment. They further remark that since the boarding and alighting flows of the transit vehicles are bounded by the vehicle capacity, there exist upper and lower bounds on the dwell times and the frequencies of the transit routes.

7.3 Systemic Characterization of the CapTA Model

The CapTA model consists in a set of variables that are involved in a set of relationships, where each variable depends on some other ones. Demand-side variables depend on supply-side variables and vice-versa. In this section, we gather the dependencies throughout the model and provide the overall logical structure of the model, laying the emphasis on the upper layer arc cost-flow functions.

The section is organized into six parts: the first three deal with the upper layer demand and supply functions and their interaction, while the following two address the traffic flowing on the lower layer and the evaluation of the line leg cost. Then the logical structure is illustrated in an influence diagram.

7.3.1 Demand Functions on the Upper Layer

At the network layer, the demand for travel is expressed by the origin destination trip matrix $\mathbf{Q}_{NS} = [q_{is}]_{i \in N, s \in S}$. A passenger travelling on the network experiences a travel cost linked to the travel cost of the links and nodes he traverses. At node i a passenger disposes of a set of local strategies (Spiess and Florian, 1989) $d_{iS} \in D(i)$ which allow him to reach destination s . d_{iS} will be noted as d for simplicity. A local strategy corresponds to a path from node i to destination s made up of a sequence of links. The cost of a local strategy, $\tau_{i(s)}^d$ stems from the travel costs of the downstream links and nodes. A passenger is assumed to be a rational economic agent who seeks to minimize his travel cost. Consequently, the passenger flows are assigned to the strategies of minimum cost.

Let $\xi_{i(s)}^d$ be the share of a strategy $d \in D(i)$, $\xi_{i(s)}^d \geq 0$ and $\sum_{d \in D(i)} \xi_{i(s)}^d = 1$. Then, at node i the flow on a strategy is defined as $x_{i(s)}^d = \xi_{i(s)}^d (\sum_{a \in A_i^-} x_{as} + q_{is})$. The network flow state $\mathbf{x}_{AS} = \mathbf{A}_{AS}(\mathbf{x}_{NS}^D, \mathbf{p}_A^D)$ stems from the flows on the strategies \mathbf{x}_{NS}^D and the route shares of the outgoing arcs, $\rho_A^D = [\rho_a^d : a(i, j) \in A, d \in D(i)]$. This corresponds to a traffic equilibrium if there exists a cost matrix $\bar{\tau}_{NS} = [\bar{\tau}_{i(s)} : s \in S, i \in N - \{s\}]$, such that for all $s \in S$ and $i \in N - \{s\}$ the following conditions apply:

$$\xi_{i(s)}^d \geq 0, \forall d \in D(i) \quad (\text{S.1}).a$$

$$\sum_{d \in D(i)} x_{is}^d = q_{is} + \sum_{a \in A_i^-} x_{as} \quad (\text{S.1}).b$$

$$\tau_{i(s)}^d - \bar{\tau}_{i(s)} \geq 0, d \in D(i) \quad (\text{S.1}).c$$

$$\xi_{i(s)}^d (\tau_{i(s)}^d - \bar{\tau}_{i(s)}) = 0, \forall d \in D(i) \quad (\text{S.1}).d$$

7.3.2 Traffic Functions on the Upper Layer

Let $\rho_a^d \geq 0$ note the route shares of outgoing arc $a \in A_i^+$ on strategy $d \in D(i)$, with $\rho_a^d = 0$ if arc $a \notin d$ and $\sum_{a \in A_i^+} \rho_a^d = 1$. Then the network flow state \mathbf{x}_{AS} is linked to the node exit flows \mathbf{x}_{NS}^D of the strategies: $x_{as} = \sum_{d \in D(i)} x_{i(s)}^d \rho_a^d$. Therefore this relation can be stated as:

$$\mathbf{x}_{AS} = \mathbf{A}_{AS}(\mathbf{x}_{NS}^D, \mathbf{p}_A^D) \quad (\text{S. 2})$$

The vector of flows by link on the upper layer \mathbf{x}_A , $a \in A$, is linked to the vector of flows by destination since, $x_a = \sum_{s \in S} x_{as}$. Since $A_L \subseteq A$ the line leg flow state, \mathbf{x}_L , is related to the network flow state :

$$\mathbf{x}_L = \mathbf{A}_L(\mathbf{x}_{AS}) \quad (\text{S. 3})$$

The leg flow vector is used in the line model for the calculation of the leg cost vector. By line ℓ , the dependencies can be structured into two parts, first the physical conditions are determined in the basis of the traffic flows along the line (section 7.3.4) and second the leg-cost is evaluated according to these conditions (section 7.3.5).

7.3.3 Upper Layer Interaction of Demand and Supply

For a given network flow state, \mathbf{x}_L , the average generalized cost of the upper layer arcs is a function detailed in section 7.4.3. It can be stated synthetically by,

$$\mathbf{g}_L = \mathbf{G}_L(\mathbf{x}_L) \quad (\text{S. 4})$$

The cost $\tau_{i(s)}^d$ of a strategy d from a node i to destination s depends on the travel cost of the networks legs and the route shares of the outgoing arcs. The route shares depend on the associated fictive frequencies. So,

$$\boldsymbol{\tau}_{NS}^D = \mathbf{T}_{NS}^D(\mathbf{g}_L, \hat{\phi}_L) \quad (\text{S. 5})$$

At node i , the minimum travel cost to destination s , $\bar{\tau}_{i(s)}$, stems from the local cost of the network elements, \mathbf{g}_A , and the shares of the outgoing arcs, \mathbf{p}_A^D . Let $\xi_{i(s)}^d$ be the share of a strategy $d \in D(i)$, $\xi_{i(s)}^d \geq 0$ and $\sum_{d \in D(i)} \xi_{i(s)}^d = 1$. We can define a multi-valued mapping of the twofold vector $\mathbf{X} = (\mathbf{x}, \boldsymbol{\xi})$:

$$(\mathbf{x}_{AS}, \boldsymbol{\xi}_{NS}^D) \in X_{NS}(\mathbf{Q}_{NS}, \boldsymbol{\tau}_{NS}^D, \hat{\phi}_L) \quad (\text{S. 6})$$

7.3.4 Line Functions of Traffic Flowing

At station i , the demand for a destination depends on the partial stock σ_{ij} by egress station j . It is calculated jointly for all transit services $z \in \ell$ and all egress stations $j \in N_\ell, j > i$, by the transit bottleneck model. It is related to the leg flows, $x_{\ell(i,j)}$, and by service z , to the available capacity k'_{zi} and the service frequency inherited from upstream, φ_{zi}^- . In Leurent and Chandakas (2012), the partial stock for all the destinations at an access station i is defined as the minimum of a potential function $\tilde{f}(\sigma)$. For simplicity we omit the access station index. Then,

$$\sigma_N = \arg \min_{\sigma} \tilde{f}(\sigma) = B_N(\mathbf{x}_L, \mathbf{k}'_{ZN}, \varphi_{ZN}^-) \quad (\text{S. 7})$$

The passenger stock n_{zi} waiting for transit service z , depends on the partial stock per destination of the stations served.

$$\mathbf{n}_{ZN} = N_{ZN}(\sigma_N) \quad (\text{S. 8})$$

By service z , the probability of immediate boarding depends on the available capacity k'_{zi} and the demand, expressed by the passenger stock n_{zi} .

$$\pi_{ZN} = \Pi_{ZN}(\mathbf{n}_{ZN}, \mathbf{k}'_{ZN}) \quad (\text{S. 9})$$

The number of boarding passengers by service is related to the passenger stock and the probability of immediate boarding, $y_{zi} = n_{zi} \cdot \pi_{zi}$, so

$$\mathbf{y}_{ZN} = Y_{ZN}(\mathbf{n}_{ZN}, \pi_{ZN}) \quad (\text{S. 10})$$

By service z , the available frequency is related to the restrained frequency inherited from upstream and the probabilities of immediate boarding:

$$(\varphi\pi)_{ZN} = \text{AF}_{ZN}(\varphi_{ZN}^-, \pi_{ZN}) \quad (\text{S. 11})$$

By service z , the probabilities of sitting if on-board p_{zi}^o and sitting at boarding p_{zi}^+ depend on the vehicle loads and the seat available capacity k_{zi}^r . In addition, the density of the

standing passengers is related to the number of standees and the available surface available, considered constant. Therefore,

$$\mathbf{p}_{ZN}^+ = \mathbf{P}_{ZN}^+(\mathbf{y}_{ZN}, \mathbf{k}_{ZN}^r) \quad (\text{S. 12. a})$$

$$\mathbf{p}_{ZN}^o = \mathbf{P}_{ZN}^o(\mathbf{y}_{ZN}, \mathbf{k}_{ZN}^r) \quad (\text{S. 12. b})$$

$$\mathbf{d}_{ZN} = \mathbf{D}_{ZN}(\mathbf{y}_{ZN}^r) \quad (\text{S. 13})$$

By service z the dwell time is a function of the boarding and alighting flows. Therefore,

$$\mathbf{T}_{ZN} = \mathbf{DT}_{ZN}(\mathbf{y}_{ZN}) \quad (\text{S. 14})$$

By service z , the restrained frequency φ_{zi}^- depends on the temporal occupancy of the station track and thus on the dwell time of all the services $z \in Z_{\ell(i)}$ using the track at station i . So,

$$\varphi_{ZN}^- = \mathbf{RF}_{ZN}(\varphi_{ZN}^-, \mathbf{T}_{ZN}) \quad (\text{S. 15})$$

By service z , an increase in the vehicle run time is related to the dwell time and the restrained frequencies. It yields a delay propagated upstream,

$$\mathbf{t}'_A = \mathbf{T}_A(\mathbf{T}_{ZN}, \varphi_{ZN}^-) \quad (\text{S. 16})$$

7.3.5 Cost Functions

By service leg $z(i, j)$, a sitting cost by section is built in relation to the updated travel time, t'_{za} . The intermediate auxiliary cost is evaluated by backward accumulation along the service leg and depends on the on-board sitting probabilities p_{zi}^o , the in-vehicle density d_{zi} and the sitting cost. Therefore:

$$\mathbf{t}_{ZL} = \mathbf{T}_{ZL}(\mathbf{t}'_A) \quad (\text{S. 17})$$

$$\gamma_{ZL} = \Gamma_{ZL}(\mathbf{t}'_A, \mathbf{t}_{ZL}, \mathbf{p}_{ZN}^o, \mathbf{d}_{ZN}) \quad (\text{S. 18})$$

The in-vehicle generalized travel time is a function of the sitting probabilities p_{zi}^+ , the auxiliary cost and the actual travel time:

$$\mathbf{g}_{ZL} = \mathbf{G}_{ZL}(\mathbf{p}_{ZN}^+, \mathbf{t}_{ZL}, \gamma_{ZL}) \quad (\text{S. 19})$$

By line leg $\ell(i, j)$, the average in-vehicle generalized cost $\tilde{g}_{\ell(i, j)}$ is related to that of the service legs $g_{z(i, j)}$, and the available frequency, $(\varphi\pi)_{zi}$.

$$\tilde{\mathbf{g}}_L = \tilde{\mathbf{G}}_L(\mathbf{g}_{ZL}, (\varphi\pi)_{ZN}) \quad (\text{S. 20})$$

The fictive frequency on the leg, $\hat{\varphi}_{\ell(i, j)}$ and the waiting time depend on the service legs and the available frequency, $(\varphi\pi)_{zi}$, so

$$\hat{\varphi}_L = \text{FF}_L((\varphi\pi)_{ZN}) \quad (\text{S. 21})$$

$$\mathbf{w}_L = \mathbf{W}_L((\varphi\pi)_{ZN}) \quad (\text{S. 22})$$

By line leg $\ell(i, j)$ the average generalized cost $g_{\ell(i, j)}$ is a function of the in-vehicle generalized cost $\tilde{g}_{\ell(i, j)}$ and the available frequency, $(\varphi\pi)_{zi}$, so

$$\mathbf{g}_L = \mathbf{G}_L(\tilde{\mathbf{g}}_L, (\varphi\pi)_{ZN}) \quad (\text{S. 23})$$

7.3.6 Logical Structure

The equations presented previously portray an interconnected system of dependencies between the model variables in CapTA. Figure 26 serves as an overview of the logical structure. Thus, the influences between the variables can be traced.

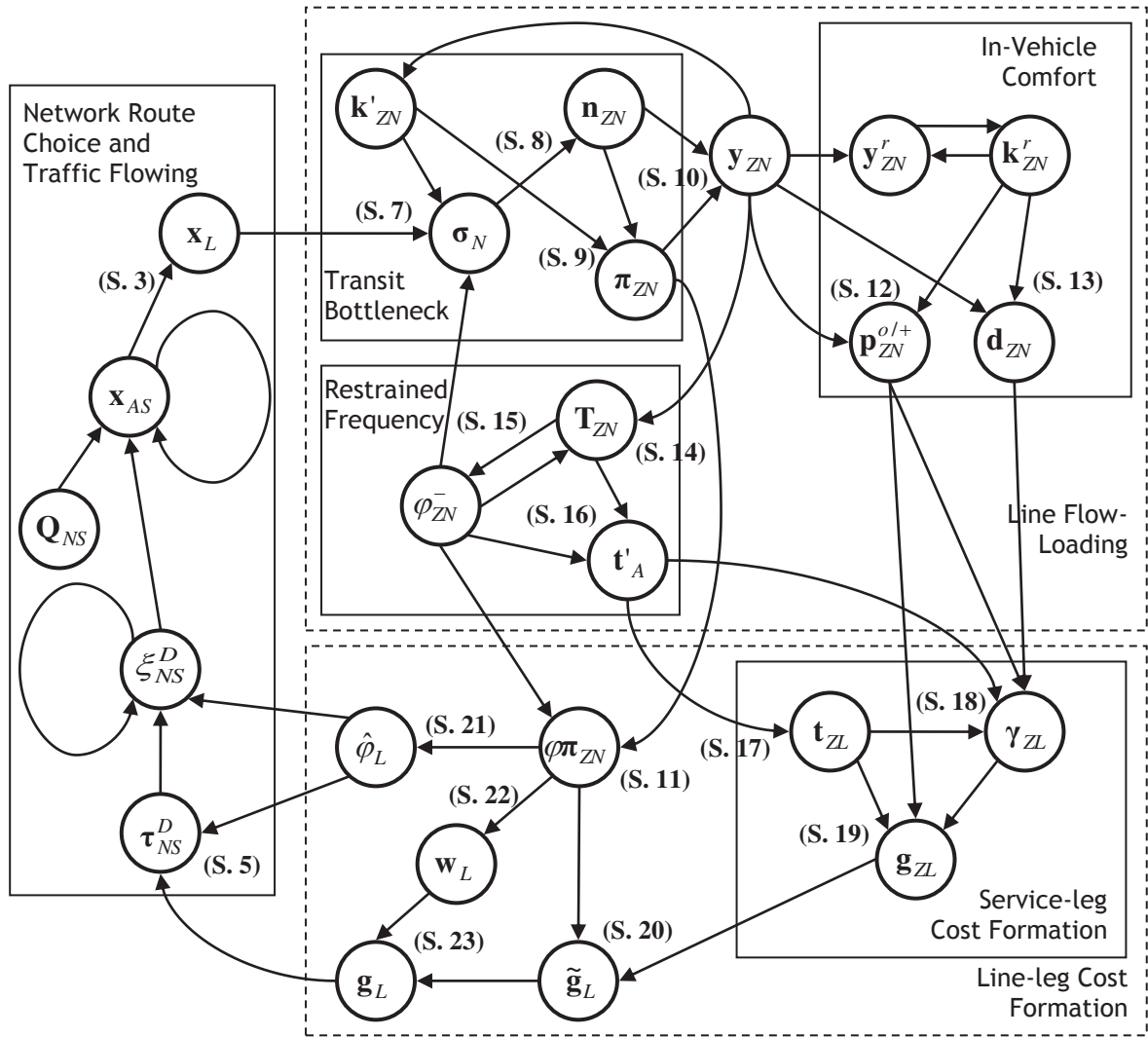


Figure 26: Structure of dependencies between the model variables in CapTA

In addition, the following sub-systems can be identified:

- A route choice and network traffic flowing model in which the traffic equilibrium is addressed as a twofold vector of the main state variables, \mathbf{x}_{AS} and ξ_{NS}^D . These are determined on the basis of the costs of the strategies and the travel demand. The leg flow vector stems from the convex combination of the auxiliary arc flow vector and the is a convex combination of the arc flow vectors of the auxiliary problem and of the previous network state,
- A line model for flow loading with the following sub-systems intervening at each station:

- The transit bottleneck model yielding passenger stocks and probabilities of immediate boarding,
- An in-vehicle comfort model for the calculation of the sitting at each stage,
- A restrained frequency model yielding restrained frequencies and in-vehicle travel times on the basis of vehicle loads from the transit bottleneck.
- A line model for cost evaluation to determine the average generalized cost of the line legs and the respective fictive frequencies. It contains a sub-system for the evaluation of the average generalized cost of a service leg.

The structure of the CapTA model is controlled, since the route choice, the line flowing and the line cost formation sub-models are straightforward, with no internal feedbacks. Nevertheless, there are some feedbacks between the models, namely the dependency between the transit bottleneck and the restrained frequency models: the former determines the vehicle loads in a station i that influence the restrained frequency in the station $i + 1$. In that way we avoid a direct feedback between frequencies and flows. The logical flow is illustrated in Figure 27.

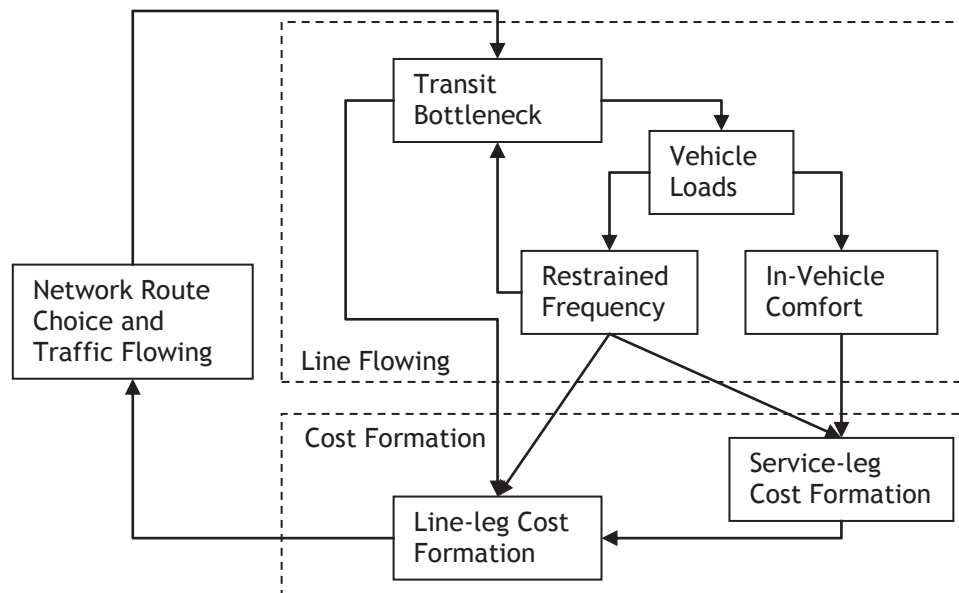


Figure 27: The sub-models and the logical flow between them

7.4 Mathematical Analysis of the CapTA Model

The previous section describes the line cost formation process used in the line model. We further establish the regularity and the mathematical properties of the previous dependency functions. Sections 7.4.1 and 7.4.2 are devoted in the line flow loading and line leg costing functions. Then, sections 7.4.3 and 7.4.4 address the formulation of the traffic equilibrium as a variational inequality problem. Finally, section 7.4.5 sketches as equilibrium algorithm and defines a rigorous convergence criterion.

The mathematical analysis of the CapTA model is dealt in two parts. First, we address some particularities which arise from the exhaustive representation of the transit network and the capacity effects considered. The main issues concern the frequency adaptations along the model and the formulation of the arc flow-cost function. Then we formulate a traffic equilibrium with respect to a twofold vector $\mathbf{X} = (\mathbf{x}, \xi)$ of arc flow per destination and share of strategy assignment per node and destination. We characterize an equilibrium state as the solution of specific variational inequality problem (VIP) for which we provide an existence theorem and a computational method.

7.4.1 Mathematical Properties of Line Flow Loading Functions

In section 7.3.2, the dependencies of the state variables between demand and supply were defined. The mathematical properties of these functions are posed in this section. Six Lemmas are developed to demonstrate the regularity of the functions stated in the previous section. This section does not follow the order of the functions in section 7.3.

First we examine the properties of the functions used in the restrained frequency model; namely the dwell time function, the track occupancy, the reduction factor and the physical travel time function. Then, the functions linked to the boarding process are addressed: the partial stock and stock per transit service and the probability of immediate boarding. Finally, we develop the lemmas for the continuity of the loading process and the regularity of the sitting probabilities function.

The regularity of the dwell time function, DT_{ZN} , and of the restrained frequency one, RF_{ZN} , are examined first. If we consider the station platform as a scarce resource, there is a direct effect on the service frequency at the station. The modelling process of the dwell time and the

adaptation of the restrained frequency, φ^- , are thoroughly described in Chapter 4.5. At a station i , the restrained frequency at departure is calculated by the frequency at arrival φ_{zi}^+ of service z and the station track occupancy, $H' = \sum_{z' \cap i} (T_{z'i} + \omega_{z'i}) \varphi_{z'i}^+$. The reduction factor η_i is calculated from the following equation:

$$\eta_{zi} \equiv \min\left\{1, \frac{H}{\sum_{z' \cap i} (T_{z'i} + \omega_{z'i}) \varphi_{z'i}^+}\right\} \quad (37)$$

Where T_{zi} is given by the function DT_{ZN} of boarding and alighting flows per vehicle of the service z . $\mathbf{y} = [y_{ij}^z : i, j \in \ell, i < j, z \in \ell]$ is the vehicle loads matrix of the services on the line, calculated by the transit bottleneck model. In the unconstraint case, it would be $\mathbf{y} = \mathbf{x} / \varphi$. The reduction factor is defined as $\eta = \eta(\mathbf{y})$ and respectively, the restrained frequency is $\varphi^- = RF(\mathbf{y})$. The separation time ω is a positive constant, related to the type of services.

If e_{zi} and b_{zi} are the alighting and boarding volumes of a service, given by the service volume vector \mathbf{y}_z , the dwell time function in Leurent et al (2011) is defined as a linear function of these volumes, $DT_{ZN}(\mathbf{e}_{ZN}, \mathbf{b}_{ZN})$. This linearity implies that there is a maximum when $e_{zi} = b_{zi} = k_z$. Anyhow, for any type of function, there exists an upper bound $DT_{zi}(\mathbf{y}_z) \leq M$, since it is reasonable to consider that the operational protocol of the service dictates an maximum dwelling: a vehicle cannot dwell indefinitely at a station.

Lemma 1 – Regularity of track occupancy and of reduction factor

The station track occupancy is a function of the in-vehicle volume matrix \mathbf{y} and the service frequencies φ_z of the services of the line, denoted by, $H'_i = H_N(\mathbf{y}, \varphi_z)$. The platform temporal occupancy has a lower bound for scheduled time $H' \geq H_{\min}$ and it is strictly increasing. Let the reduction factor be $\eta_i = \min\{1, H_i / H'_i\}$. It is therefore continuous with respect to (\mathbf{y}, φ) and continuously differentiable except along $\{H = H'\}$. The same is valid for $RF_{zi}(\varphi_{zi}^-, \mathbf{T}_{zi})$.

Proof. By definition, the dwell time function, $DT_{zi}(\mathbf{y}_z)$, is continuous and strictly increasing in respect to the vehicle loads and has an upper bound, $DT_{zi}(\mathbf{y}_z) \leq M$. Furthermore, the station track occupancy H' is continuous, strictly increasing and has an upper bound as a sum of strictly increasing bounded functions. The regularity of the reduction factor stands for $H \neq H'$, since the function yields $\eta = 1$ for $H > H'$ or $\eta = H/H'$ for $H < H'$, which is continuously differentiable on their restricted domain. The line $\{H = H'\}$ separates the two sub-domains of differentiability. The reduction factor is strictly decreasing and has a lower bound, $\eta \geq \varepsilon$. The regularity property and the lower bound will also hold for the restrained frequency function $RF_{zi}(\varphi_{zi}^-, \mathbf{T}_{zi})$, since by definition $\varphi_{zi}^- = \eta_i \cdot \varphi_{zi}$.

Lemma 2 – Regularity of physical travel time function T_A

The physical travel time of a lower-layer arc t_a is a function of the dwell time T_{zi} and of the reduction factor η_i . It is continuous with respect to (y, φ) and continuously differentiable except along $\{H = H'\}$.

Proof. The arc travel time is subject to a combination of functions. First, the travel time increases in respect to the dwell time. By definition the dwell time function $DT_{zi}(\mathbf{y}_z)$, is continuous and strictly increasing in respect to the vehicle loads and so it is the arc travel time. Secondly, the restrained frequencies produce a vehicle flowing delay reflected travel time increase on the vehicle travel propagated upstream. It depends on the regularity factor and by Lemma 1 travel time is continuous and differentiable except along $\{H = H'\}$.

Enforcing the vehicle's total capacity constraints means comparing the demand for transport, expressed by the partial stock per destination with the transit supply. The transit bottleneck model estimates the partial stock σ_{ij} on the basis of the exogenous flows $\mathbf{x}_{\ell(i)} = [x_{\ell(i,j)}]_{j>i}$, the available capacities $[k'_{zi}]_{z \in Z_{\ell(i)}}$ and the service frequencies $[\varphi_{zi}^-]_{z \in Z_{\ell(i)}}$ of the services stopping at node i .

In Leurent and Chandakas (2012b), it is demonstrated that the partial stock σ_{ij} is the solution of the equation $(2/HN_{ij})\sigma_{ij} = x_{\ell(i,j)}/\sigma_{ij} - (\varphi\pi)_{ij}$. We define the associated mapping $\tilde{F}_{ij}(\sigma_{ij}) = (2/HN_{ij})\sigma_{ij} - x_{\ell(i,j)}/\sigma_{ij} + (\varphi\pi)_{ij}$, which yields $\tilde{\mathbf{F}} = [\tilde{F}_{ij} : j > i, j, i \in N_\ell]$. Then $\boldsymbol{\sigma} = [\sigma_{ij} : j > i, i, j \in N_\ell]$ is the vector of partial stocks from a given access station i . A potential function is constructed so as $\nabla \tilde{f}(\boldsymbol{\sigma}) = [\tilde{F}_{ij} : j > i, j, i \in N_\ell]$. Therefore, the solution corresponds to the value of $\boldsymbol{\sigma}$ that minimizes the potential function: $\boldsymbol{\sigma} = \arg \min_{\boldsymbol{\sigma}} \tilde{f}(\boldsymbol{\sigma})$. In addition, the solution σ_{ij} belongs to the feasible set $V = \{\sigma_{ij} \geq 0 : \forall j > i, \sigma_{ij} \leq N_{ij}\}$.

Lemma 3 – Regularity of partial stock and stock per service

If the potential function $\tilde{f}(\boldsymbol{\sigma}_N, \mathbf{x}_L, \mathbf{k}'_{ZN}, \varphi_{ZN}^-)$ is differentiable with respect to its parameters, \mathbf{x}_L , \mathbf{k}'_{ZN} , φ_{ZN}^- , the same stands for the functions $\boldsymbol{\sigma}_N = \mathbf{B}_N(\mathbf{x}_L, \mathbf{k}'_{ZN}, \varphi_{ZN}^-)$ and $\mathbf{n}_{ZN} = \mathbf{N}_{ZN}(\boldsymbol{\sigma}_N)$.

Proof. The stock per service is the sum of the partial stocks per destination, for the destinations served, $n_{zi} = \sum_{j>i, j \in N_z} \sigma_{ij}$. If the partial stock is continuous, the continuity is maintained by the operator of addition.

By construction, the potential function is defined from a reference point $\boldsymbol{\sigma}_0$, such that $\tilde{f}(\boldsymbol{\sigma}) = \sum_{j>i \in N_\ell} \int_{\sigma_0}^{\sigma} \tilde{F}_{ij}(\boldsymbol{\sigma}) d\sigma_{ij}$. The potential function is continuously differentiable for any value of \mathbf{x}_L and φ_{ZN}^- . It is furthermore continuously differentiable with respect to \mathbf{k}'_Z except at $\{k'_{zi} = n_{zi}\}$. The partial stock σ_{ij} is the solution of the minimization program $\min_{\boldsymbol{\sigma}} \tilde{f}(\boldsymbol{\sigma})$ subject to $\boldsymbol{\sigma} \geq 0$ and $\boldsymbol{\sigma} \leq [N_{ij}]_{j>i}$. According to the envelope theorem (Milgrom and Segal, 2002) if the potential function is differential almost everywhere and has directional derivatives everywhere, the potential function is differentiable on the parameters. Since the potential function is differentiable, by applying the envelope theorem, the partial stock, and by extension the stock per service, is also continuously differentiable with respect to its parameters, except at $\{k'_{zi} = n_{zi}\}$.

Lemma 4 – Regularity of probability of immediate boarding

It stands $n_{zi} = N_{zi}(\sigma_N)$ and $\sigma_i = B_i(\mathbf{x}_L, \mathbf{k}'_{zi}, \varphi_{zi}^-)$. These functions yield $n_{zi} = 0$ for $\mathbf{x}_{\ell(i)} = 0$ and $n_{zi} > 0$ for $\mathbf{x}_{\ell(i)} > 0$. Let the function $\Pi(\mathbf{x}_L, \mathbf{k}'_{zi})$ be $\Pi(\mathbf{x}_{\ell(i)}, \mathbf{k}'_{zi}) \equiv \min\{1, k'_{zi} / n_{zi}\}$, for $x_\ell > 0$ and $k'_{zi} > 0$, else $\Pi(\mathbf{x}_L, \mathbf{k}'_{zi}) \equiv 0$ if $k'_z = 0$, else $\Pi(\mathbf{x}_L, \mathbf{k}'_{zi}) \equiv 1$ if $k'_z > 0$ and $x_\ell = 0$. Then, the function Π is continuous with respect to $(\mathbf{x}_L, \mathbf{k}'_{zi})$, except at $(0,0)$. It is continuously differentiable except along $\{k'_z = n_z(\mathbf{x}, \varphi^-, k'_z)\}$. The same applies for the function of available frequency $(\varphi\pi)_{ZN} = AF_{ZN}(\varphi_{ZN}^-, \pi_{ZN})$.

Proof. The probability of immediate boarding function is obviously continuous for $k'_{zi} > 0$, $x_\ell > 0$ and $k'_z \neq n_z(\mathbf{x}, \varphi^-, k'_z)$ since the function will yield either $\pi = 1$ if $k'_z > n_z(\mathbf{x}, \varphi^-, k'_z)$, or $\pi_z = k'_z / n_z(\mathbf{x}, \varphi^-, k'_z)$ if $k'_z < n_z(\mathbf{x}, \varphi^-, k'_z)$. It is differentiable on each restricted domain and by Lemma 3 the stock per service function is also continuous. The two sub-domains of differentiability are separated by the line $\{k'_z = n_z\}$. If $k'_z > 0$, then the function Π is right continuous at $x_\ell = 0$, since $k'_z / n_z \rightarrow \infty$ as $x_\ell^+ \rightarrow 0^+$ and therefore $\pi = 1$. On the other hand, if $k'_z = 0$ and $x_\ell > 0$, then it holds $\Pi_{zi}(x_\ell, 0) \rightarrow 0$ as $x_\ell^+ \rightarrow 0^+$ and therefore function Π_{zi} is continuous in the vicinity of $(0,0)$. Furthermore, since the probability of immediate boarding function is regular, the same exists for the function of available frequency $(\varphi\pi)_{ZN} = AF_{ZN}(\varphi_{ZN}^-, \pi_{ZN})$, since the product of Π_{zi} by RF_{zi} is continuous.

Although the function $\Pi(\mathbf{x}_L, \mathbf{k}'_{zi})$ is continuous close to $(0,0)$, a discontinuity arises at that point. In order to ensure the existence of equilibrium, the property of regularity must hold for all values. The function can be made continuous, by enforcing on each transit service and station a strictly positive, though arbitrarily small minimal residual capacity, ε . This seems also realistic since a passenger would prefer to board a very crowded vehicle, rather than wait indefinitely long. We define a continuous approximation of the probability of immediate boarding function in respect to the positive parameter ε and the derived variable k'_{zi} :

$$\pi_{zi}^\varepsilon = \min\left\{1, \frac{\max\{k'_{zi}, \varepsilon\}}{n_{zi}}\right\} \quad (38)$$

In other words, the parameter ε constitutes a lower bound of any residual capacity in order to ensure the regularity of the function. If we pose $\varepsilon \rightarrow 0^+$, we approach as close as required to the basic function, except of the point of discontinuity at $(\mathbf{x}_L, \mathbf{k}'_{zi}) = (0,0)$. Moreover, we can define the same approximation for the available frequency function $\text{RF}_{zi}^\varepsilon$.

We develop a lemma for the vehicle loading throughout a transit service.

Lemma 5 – Continuity throughout line flow loading

Along a transit service $z \in \ell$ of a transit line, at every station i , the functions K_{zi} and Y_{zi} are continuous and sub-differentiable with respect to the vector \mathbf{x}_L . That also applies for the comfort loading functions, K_{zi}^r and Y_{zi}^r , and the derived functions.

Proof. We use an inductive process for the proof. From the origin station $i=1$, we have $k'_1 = k_z$ for a given service and $y_{zi} = 0$. Let the Induction Assumption that the Lemma property holds for a station i and the transit service z . Let us consider the next station $i+1$. If k_{zi}^- is the available capacity at the departure from station i , the number of passengers alighting at $i+1$, $y_{zi+1}^- = \sum_{j < i+1} y_{z(j,i+1)}$ and the available capacity $k'_{zi+1} = k_{zi}^- - y_{zi+1}^-$ are regular. Therefore, these functions are continuous and sub-differentiable.

Similarly, by using an Induction Assumption, we prove that the functions of flow loading per comfort state $r \in \{\bar{r}, r\}$ are continuous and sub-differentiable.

Lemma 6 – Regularity of density and sitting probabilities functions

For $(y_{zi}, k_{zi}^r) \geq 0$: let $p_{zi}^{+lo} = P_{zi}(y_{zi}, k_{zi}) \equiv \min\{1, k_{zi} / y_{zi}\}$ if $y_{zi} > 0$ and $k_{zi}^r > 0$, else $P_{zi}(y_{zi}, k_{zi}) \equiv 0$ if $k_{zi}^r = 0$, else $P_{zi}(y_{zi}, k_{zi}) \equiv 1$ if $y_{zi} = 0$ and $k_{zi}^r > 0$. Then function P_{zi} is continuous with respect to (y_{zi}, k_{zi}^r) , except at $(0,0)$. It is continuously differentiable except

along $\{k_{zi}^r = y_{zi}\}$. The function $d_{zi} = D_{zi}(y_{zi}^{\bar{r}})$, where $D_{zi}(y_{zi}^{\bar{r}}) \equiv y_{zi}^{\bar{r}} / A^r$ and A^r is a constant related to the type of vehicle, is continuously differentiable.

Proof. The density function $D_{zi}(y_{zi}^{\bar{r}})$ is by definition continuously differentiable on its domain. The sitting probabilities functions are continuous for $k_{zi}^r > 0$, $y_{zi} > 0$ and $k_{zi}^r \neq y_{zi}$, since function $P_{zi}(y_{zi}, k_{zi})$ yields either $P_{zi}(y_{zi}, k_{zi}) = 1$ if $k_{zi}^r > y_{zi}$ or $P_{zi}(y_{zi}, k_{zi}) = k_{zi} / y_{zi}$ if $k_{zi}^r < y_{zi}$. They are differentiable in their restricted domain. In addition, the line $\{k_{zi}^r = y_{zi}\}$ separates two sub-domains of differentiability. If $k_{zi}^r > 0$, then P_{zi} is right continuous at $y_{zi} = 0$, since $k_{zi}^r / y_{zi} \rightarrow \infty$ as $y_{zi} \rightarrow 0^+$, therefore, $p_{zi} = 1$. Respectively, if $k_{zi}^r = 0$ and $y_{zi} > 0$, then $p_{zi} \rightarrow 0$ as $y_{zi} \rightarrow 0^+$. This point of discontinuity can be lifted if we use an approximation of the available capacity.

Similarly to Lemma 4 and the approximation of the function of the probability of immediate boarding, we enforce at each transit service and station a strictly positive, though arbitrarily small minimal residual capacity, ε . We define a continuous approximation of the sitting probability function in respect to the positive parameter ε and the derived variable k_{zi}^r :

$$p_{zi}^{(+/\circ)\varepsilon} = \min\left\{1, \frac{\max\{k_{zi}^r, \varepsilon\}}{y_{zi}}\right\} \quad (39)$$

In other words, the parameter ε constitutes a lower bound of any residual capacity in order to ensure the regularity of the function. If we pose $\varepsilon \rightarrow 0^+$, we approach as close as required to the basic function, except of the point of discontinuity at $(y_{zi}, k_{zi}^r) = (0, 0)$.

7.4.2 Mathematical Properties of Line Costing Functions

This section includes three lemmas for the demonstration of the regularity of the line leg cost-flow relationship on the basis of the physical conditions developed previously. Similarly to the description in section 7.3.4, the Lemmas concern the continuity through the service and line leg costing and the regularity of the fictive frequency function.

Lemma 7 – Continuity through line costing on the service legs

If, between stations i and j on a service z , the segment cost functions $t_{k,k+1}^r$ by comfort state $r = \{\underline{r}, \bar{r}\}$ are continuous with respect to \mathbf{y}_z , then the approximate functions of seated cost $T_{\text{ZL}}^\varepsilon$, of conditional cost $\Gamma_{\text{ZL}}^\varepsilon$ and of in-vehicle generalized travel time $G_{\text{ZL}}^\varepsilon$ on the service leg $z(i, j)$ are also continuous.

Proof. We proceed by induction from the egress station j to the access station i in reverse topological order. For the initialization, the cost functions of a fictive leg $z(j, j)$ are null. Let the Induction Assumption that the property holds at every station k , located between $i+1$ and j . At station i , since by Lemma 2 the physical time functions are continuous, the continuity is also maintained for their sum, which produced the seated cost function. From Lemma 6 and the approximation approach, the sitting probability functions are continuous even at $y_{zi}^r = 0$, where it takes the value $p_{zi}^r = 1$. Therefore, the composite approximate functions of the conditional cost, $\Gamma_{\text{ZL}}^\varepsilon$ and the in-vehicle average generalized time $G_{\text{ZL}}^\varepsilon$ maintain the continuity through the operator of addition. The continuity is maintained at station i and by the Induction Assumption it holds for all the stations.

Furthermore, the in-vehicle average generalized time $G_{\text{ZL}}^\varepsilon$ has lower and upper bounds. The lower bound, $G_{\text{ZL}}^\varepsilon \geq m$, obviously stems from a trip where no constraints are enforced, such as the case where only a single user travels. On the opposite side, the upper bound, $G_{\text{ZL}}^\varepsilon \leq M$, corresponds to the worst case trip, where a user stands for the whole duration of the trip at maximum standing density. This existence of the upper is linked to the existence of strict capacity constraints at boarding.

Lemma 8 – Regularity through line leg costing

If the available frequency function is continuous, with respect to the line flows \mathbf{x}_L , then the approximate functions of the average in-vehicle generalized cost, \bar{G}_L^ε , of the waiting time W_L^ε and of the average generalized cost G_L^ε are also continuous.

Proof. According to Lemma 7, the approximate function of the average in-vehicle generalized cost of a service leg $z(i, j)$ is continuous. Furthermore, a truncation of the available vehicle capacity, yields a continuous approximation of the available frequency function AF_{zi}^ε , according to Lemma 4. The probability of immediate boarding yields $\pi_{zi}^\varepsilon = 1$ for $x_\ell = 0$ and by the capacity truncation, it stands $\pi_{zi}^\varepsilon \neq 0$. We denote $(\varphi\pi)_{ij} = \sum_{z \in (i, j)} \varphi_{zi} \pi_{zi}^\varepsilon$ the combined available frequency. From Lemma 1, φ_{zi} is continuous and has a lower positive bound. Therefore, their sum yields also a continuous function where $(\varphi\pi)_{ij} \neq 0$. The approximate function of the average generalized in-vehicle cost \bar{G}_ℓ^ε of the line leg $\ell(i, j)$ will also be continuous as a linear combination of continuous function.

The waiting time of the line leg is defined as $w_\ell = \alpha / (\varphi\pi)_\ell$, which is continuous since $(\varphi\pi)_{ij} \neq 0$ everywhere. Finally, the continuity is maintained for the approximate function of the average generalized function G_ℓ^ε from the operator of addition. Finally, since $(\varphi\pi)_\ell > 0$ and by Lemma 6 function \bar{G}_ℓ^ε is bounded, the average generalized function has an upper bound, $G_\ell^\varepsilon \leq M < \infty$

Lemma 9 – Regularity of fictive frequency and leg waiting functions

The discontinuity attenuation β_a is a function of the excess waiting time, w'_a on a line leg. It stands $\psi(0) = 1$ and $\psi(w') = 0$, for every $w' > \varepsilon$, a small parameter. If $\beta_a = \psi(w')$, let $\hat{\varphi}_a = \varphi_a / \beta_a$ if $w' < \varepsilon$ and $\hat{\varphi}_a = \infty$ for every $w' > \varepsilon$. If $\psi(w')$ is a continuous function, so does the function of fictive frequency, $FF_a((\varphi\pi)_a)$. The same applies for the leg waiting time w_a .

Proof. The leg waiting is the inverse of the leg fictive frequency, $w_a = \alpha / (\varphi\pi)_a$, where α is the waiting weighting factor and depends on the queuing discipline, the vehicle and passenger arrival distribution. By definition and Lemma 4, the function is continuous, since the property is conserved with the operator of multiplication, if it is nowhere zero.

The discontinuity function $\psi(w')$ is by definition a continuous strictly decreasing function. If w' is the excess waiting time where $w'_a = (w_a - 1/\varphi_a)$ and $\psi(w') = 0$ for every $w' > \varepsilon$, it is continuous everywhere. Therefore, the fictive frequency function $\text{FF}_a((\varphi\pi)_a)$ is continuous for $w' < \varepsilon$, since $\hat{\varphi}_a = \varphi_a / \beta_a$, where $\beta_a \equiv \psi(w_a - \alpha / \varphi_a) > 0$ is strictly positive.

Nevertheless, the previous definition suggests a discontinuity when $\hat{\varphi}_a = \infty$ (for $\beta_a = 0$). That implies that for a queuing time exceeding a small parameter ε , the line leg $a \in A_\ell$ is reduced to a pedestrian arc with continuous availability and infinite frequency. The regularity of the functions is addressed by accepting an upper bound on the leg frequency $\hat{\varphi}_a = \min\{\varphi_\infty, \varphi_a / \psi(w_a - \alpha / \varphi_a)\}$ (see appendix A4 in Leurent et al., 2012b). The fictive frequency function is continuous for all the values and differentiable on the two sub-domains, except along $\{(w_a - \alpha / \varphi_a) = \varepsilon\}$.

The effects of the local models on the line level are expressed in sections 7.4.1 and 7.4.2 where the dependencies are declared among the state variables of the line model. The continuity of the line leg cost-flow relationship is demonstrated in Theorem 1.

Theorem 1 – Continuity of leg cost-flow relationship

Let the service leg costs $G_{z(i,j)}^\varepsilon$ of the transit services $z \in \ell$ be functions of the leg flow vector with $\mathbf{x}_L \geq 0$. Respectively, let the expected waiting time for a line leg to be a function of the leg flow vector. If the travel cost functions and the waiting cost function are continuous with respect to \mathbf{x}_L , so it is the approximate average generalized line leg cost function $G_{\ell(i,j)}^\varepsilon$, except at the points where $k'_z = 0$ and $x_\ell^+ = 0$.

The Lemmas 1 to 9 that have been developed in the previous sections make the Theorem 1 hold true.

7.4.3 Formulation of Traffic Equilibrium

As defined previously the line model amounts to an elaborate three-fold function $[\mathbf{g}_L, \mathbf{w}_L, \hat{\phi}_L]$ of the line leg vector $\mathbf{x}_L \subseteq \mathbf{x}_A$. Let $G = [0, \infty)^A$, $W = [0, \infty)^A$ and $\hat{\Phi} = (0, \infty)^A$ denote the vector of feasible sets of the vectors \mathbf{g}_A , \mathbf{w}_A and $\hat{\phi}_A$. The line model can be formulated as a function of the arc flow per destination matrix, \mathbf{x}_{AS} , hence:

$$[\mathbf{g}_L, \mathbf{w}_L, \hat{\phi}_L] = F_L(\mathbf{x}_{AS}) \quad (40)$$

Theorem 1 proves the continuity of the line leg cost with respect to the line leg flow vector. Equivalently, the combination of Lemmas 4 and 9 suggests the continuity of the line leg waiting time and the fictive frequency. By extent these are also continuous with respect to the arc flow per destination matrix, \mathbf{x}_{AS} .

We further determine the cost structure by destination $s \in S$. For each node $i \in N \setminus \{s\}$ the local route choice to the destination is given by a local travel strategy $d \subseteq A_i^+$, where A_i^+ is the set of outgoing arcs from node i . For each node and strategy, we associate a cost $\tau_{i(s)}^d$ to destination s . Let $\bar{\tau}_{i(s)}$ denote the cost of the strategy of minimum cost from node i to destination s and $\hat{\phi}_i^d$ the combined revised frequency, $\hat{\phi}_i^d \equiv \sum_{a \in d} \hat{\phi}_a$. The set of local strategies at i is expressed as $D(i) \equiv \{d \subseteq A_i^+, d \neq \emptyset\}$. The local travel strategy yields a travel cost:

$$\tau_{i(s)}^d = \frac{\alpha + \sum_{a \in d} \hat{\phi}_a (g_a + \bar{\tau}_{j(s)})}{\hat{\phi}_i^d} \text{ where } a \approx (i, j) \quad (41)$$

Let us demonstrate the continuity of the travel cost to destination with respect to $(\mathbf{g}_A, \hat{\phi}_A)$.

Lemma 10 – Continuity of the travel cost to destination function

If the arc cost-flow G_L and fictive frequency functions FF_L are continuous, so does the function of travel cost to destination $\tau_{NS}^D = T_{NS}^D(\mathbf{g}_A, \hat{\phi}_A)$.

Proof. Theorem 1 and Lemma 9 imply that the arc cost-flow function and the fictive frequency function are continuous. In addition, by combination of Lemma 1 of section 7.4.1 and Lemma 9, the fictive frequency function has lower and upper bounds. The continuity is maintained through the operations of addition and division, as long as the denominator in the division is not zero.

From any current node $i \in N \setminus \{s\}$ we consider that within the directed graph $G = (N, A)$ there is a direct path to destination s composed of arcs $a \in A$. The travel cost to destination is the solution of the generalized Bellman equations (Nguyen and Pallotino, 1988; Spiess and Florian, 1989). Therefore:

$$\bar{\tau}_{i(s)} = \begin{cases} 0 & \text{if } i = s \\ \min_{d \in D(i)} \frac{\alpha + \sum_{a \in d} \hat{\phi}_a (g_a + \bar{\tau}_{j(s)})}{\hat{\phi}_i^d} & \text{if } i \neq s \end{cases} \quad (42)$$

The recursive building of optimal hyperpaths on the basis of the leg conditions $(\mathbf{g}_A, \hat{\phi}_A)$ on the upper layer links yields minimum hyperpath costs. In their Proposition 1, Cominetti and Correa (2001) demonstrate that for each couple $(\mathbf{g}_A, \hat{\phi}_A) \in G \times \hat{\Phi}$ there is a unique solution of the generalized Bellman equations (42) that yields the minimum travel costs to destination $\bar{\tau}_{Ns}$. Furthermore, they prove that the solution is continuous with respect to $(\mathbf{g}_A, \hat{\phi}_A)$.

The authors define a mapping $M : [0, \infty]^N \rightarrow [0, \infty]^N$ with $M_i(\bar{\tau}_{Ns}) = 0$ if $i = s$ and $M_i(\bar{\tau}_{Ns}) = \min_{d \in D(i)} [\alpha + \sum_{a \in d} \hat{\phi}_a (g_a + \bar{\tau}_{j(s)})] / \hat{\phi}_i^d$ if $i \neq s$. The mapping is monotone with respect to $\bar{\tau}_{Ns}$. Therefore, the equations (42) are formulated as a Fixed Point Problem $\bar{\tau}_{Ns} = M_N(\bar{\tau}_{Ns})$. The continuity of the mapping and the compactness of its feasible set assure the existence and uniqueness of a solution. These conditions are also satisfied in CapTA since every node i can be connected to s by a path with at most $|N| - 1$ arcs. That implies that the vector $\bar{\tau}_{Ns}$ has an upper bound, hence it is finite.

Proposition 1 of Cominetti and Correa (2001) is valid in the case of the CapTA model. Therefore, the minimum travel costs to destination $\bar{\tau}_{Ns}$ can be expressed as a function $\bar{\tau}_{Ns} = \bar{T}_{Ns}(\mathbf{g}_A, \hat{\phi}_A)$ which is continuous with respect to $(\mathbf{g}_A, \hat{\phi}_A)$.

Let $\xi_{i(s)}^d$ denote the flow share of each local strategy $d \subseteq D(i)$. We can therefore define the vector $\xi_{i(s)}^D \equiv [\xi_{i(s)}^d : d \in D(i)]$, where $\xi_{i(s)}^d \geq 0$ and $\sum_{d \in D} \xi_{i(s)}^d = 1$. From the previous definition of the minimum hyperpath cost, a strategy assignment $\xi_{i(s)}^D$ at a current node $i \in N \setminus \{s\}$ is optimal if and only if any strategy of strictly positive share is of minimum cost:

$$\xi_{i(s)}^d (\tau_{i(s)}^d - \bar{\tau}_{i(s)}) = 0, \forall d \in D(i) \quad (43)$$

Let ρ_a^d denote the route share of outgoing arc $a \in A_i^+$ on strategy d . There would be $\rho_a^d = 0$ if arc $a \notin d$ and $\sum_{a \in A_i^+} \rho_a^d = 1$. In the case of the strategies described in Spiess and Florian (1989) the route share of an outgoing arc in strategy d stems from the frequencies in a proportional way. In CapTA model the route share stems from the fictive frequencies, $\hat{\phi}_a$. Therefore, denoting by $1_{\{a \in d\}}$ the indicator function and assuming that strategy d is a non-empty set, the route share is derived as:

$$\rho_a^d = \hat{\phi}_a 1_{\{a \in d\}} / \hat{\phi}_i^d, \text{ where } a \in A_i^+ \quad (44)$$

The strategy assignment and the local route share of each strategy determine the arc share h_{as} per destination in the following way:

$$h_{as} \equiv \sum_{d \in D(i)} \rho_a^d \cdot \xi_{i(s)}^d \quad (45)$$

Lemma 11 – Continuity of the arc share function

The arc share \mathbf{h}_{AS} is a continuous function of \mathbf{x}_{AS} and of ξ_{NS}^D .

Proof. For a given strategy, the route share is a continuous with respect to \mathbf{x}_{AS} . Indeed, Lemma 9 implies that the fictive frequency function is continuous. In combination with Lemma 1 the fictive frequency function FF_A has upper and lower bounds and therefore the route share function is always defined. Therefore the continuity is preserved through the operations of addition and division, since the operands are not zero. By construct, the arc share is also continuous with respect to the vector of strategy assignment.

By destination s , the total arc flow transiting at node i is conserved, according to the equation:

$$\sum_{a \in A_i^+} x_{as} = q_{is} + \sum_{a \in A_i^-} x_{as}$$

Since the outgoing flow per strategy d is $x_{i(s)}^d = \xi_{i(s)}^d \cdot (q_{is} + \sum_{a \in A_i^-} x_{as})$ and the flow of an arc is $x_{as} = \sum_{d \in D(i)} x_{i(s)}^d \rho_a^d$, the arc flow can be expressed on the basis of the strategy assignment:

$$x_{as} = h_{as} (q_{is} + \sum_{b \in A_i^-} x_{bs}), \quad \forall a, b \in A_i^+ \quad (46)$$

In order to make use of equation (46) we put it under matrix form. Let us define the node-arc incidence function $e_{ia}^{+(-)} \equiv 1$ if $a \in A_i^+$ (resp. $a \in A_i^-$) or 0 otherwise. Furthermore, if $|\cdot|$ denotes the cardinal of a set, let $\mathbf{e}^{+(-)} \equiv [e_{ia}^{+(-)} : i \in N, a \in A]$ be a $|N| \times |A|$ matrix related to the directed graph $G = (N, A)$. Finally, let us define the diagonal matrix $\mathbf{h}_s \equiv \text{Diag}(h_{as} : a \in A)$. Equation (46) in vector form amounts to:

$$\mathbf{x}_{As} = \mathbf{h}_s \cdot \mathbf{e}^{+T} (\mathbf{q}_{Ns} + \mathbf{e}^- \mathbf{x}_{As}) \quad (47)$$

Definition 1 – Traffic Equilibrium

A twofold vector $\mathbf{X} = (\mathbf{x}, \xi)$ is a traffic equilibrium if and only if it satisfies the system of conditions described by $\mathbf{x}_A = \sum_{s \in S} \mathbf{x}_{As}$ and the equations (40), (41), (43), (44) and (45).

7.4.4 Equilibrium Characterization as a Variational Inequality Problem

Here we define a Variational Inequality Problem that characterizes a traffic equilibrium. Given vector \mathbf{X} , let us define a quadratic function on the basis of equation (47):

$$J_s(\tilde{\mathbf{x}}_{As}, \mathbf{X}) = \frac{1}{2} \left\| \tilde{\mathbf{x}}_{As} - \mathbf{h}_s \cdot \mathbf{e}^{+T} (\mathbf{q}_{Ns} + \mathbf{e}^- \tilde{\mathbf{x}}_{As}) \right\|^2 \quad (48)$$

By construct if $\tilde{\mathbf{x}}_{As}$ satisfies equation (47) then it must minimize function J_s . The converse also holds true. Define $E_s \equiv \{\tilde{\mathbf{x}}_{As} \geq 0 : \sum_{a \in A_i^+} \tilde{x}_{as} \leq \sum_{i \in N} q_{is}\}$ as the feasible set of $\tilde{\mathbf{x}}_{As}$ and $B_{i(s)}^D \equiv \{(\xi_{i(s)}^d)_{d \in D(i)} : \xi_{i(s)}^d \geq 0 \text{ and } \sum_{d \in D(i)} \xi_{i(s)}^d = 1\}$ that of the strategy assignment $\xi_{i(s)}^D$.

Lemma 12 – Compactness of the feasible sets

The feasible sets (a) E_s and (b) $B_{i(s)}^D$ are closed and compact.

Proof. (a) The set of arc flow per destination has lower and upper bounds defined respectively by the non-negativity constraint ($x_{as} \geq 0$) and the total flow on the network ($x_{as} \leq \sum_{i \in N} q_{is}$). Since it contains its boundaries, it is a closed set. In addition, it is a nonempty set, since it exists at least one path with sufficient capacity ($k > \sum_{i \in N} q_{is}$). Furthermore, the graph contains a finite number of arcs, hence a finite number of paths for each origin destination couple, which guarantees its finite dimension. Therefore, it is a nonempty, convex polytope of finite dimension, hence a compact set.

(b) The same properties apply to the strategy assignment since by definition for any $\xi_{i(s)}^d \in \xi_{i(s)}^D$, we have $\xi_{i(s)}^d \geq 0$ and $\sum_{d \in D(i)} \xi_{i(s)}^d = 1$. Thus, the set $B_{i(s)}^D$ is closed and compact.

Let $\tilde{\nabla} J_s(\tilde{\mathbf{x}}_{As}, \mathbf{X})$ be the gradient of the quadratic function (48) along $\tilde{\mathbf{x}}_{As}$. The minimization of J_s with respect to $\tilde{\mathbf{x}}_{As}$ over E_s at $\tilde{\mathbf{x}}_{As}^*$ is a convex minimization program which can be expressed as a variational inequality:

$$\forall \tilde{\mathbf{x}}_{As} \in E_s, \tilde{\nabla} J_s(\tilde{\mathbf{x}}_{As}^*, \mathbf{X}) \cdot (\tilde{\mathbf{x}}_{As} - \tilde{\mathbf{x}}_{As}^*) \geq 0 \quad (49)$$

This formulation allows us to formulate a variational inequality problem with respect to the vector $\mathbf{X} = (\mathbf{x}, \xi)$. Let \mathbf{F} denote the cost function of \mathbf{X} : it has components T_{NS}^D along $\xi_{i(s)}^d$ and $\partial J_s(\tilde{\mathbf{x}}_{As}, \mathbf{X}) / \partial \tilde{x}_{as}$ along x_{as} .

Definition 2 – Variational Inequality Problem (VIP)

If $\mathfrak{R} \equiv \prod_{s \in S} E_s \times \prod_{s \in S, i \in N} B_{i(s)}^D$, find $\mathbf{X}^* \in \mathfrak{R}$ such that $F(\mathbf{X}^*) \cdot (\mathbf{X} - \mathbf{X}^*) \geq 0$, $\forall \mathbf{X} \in \mathfrak{R}$.

Theorem 2 – Characterization of the VIP

(a) If \mathbf{X} is a traffic equilibrium, then it solves the Variational Inequality Problem (Def.2).

(b) Any solution of the VIP is a traffic equilibrium.

Proof. (a) The first part of the demonstration lies in showing that a traffic equilibrium $\mathbf{X} = (\mathbf{x}, \xi)$ must belong to the admissible set. Indeed, this is true for the part ξ_{NS}^D by the definition of a strategy assignment. Furthermore, from the formation of the travel costs of the strategy, equation (43) states that only the strategies that locally minimize the travel cost have a strictly positive strategy share $\xi_{i(s)}^d$. Equation (41) implies that at node $i \in N$ any local strategy must comply with a network lattice made up of links $a \approx (i, j)$ such that $\bar{\tau}_{i(s)} > \bar{\tau}_{j(s)}$. Finally, the node flow conservation equation (46) guarantees that at any arc $a \in A$ the arc flow per destination x_{as} stems from the path flows from origin nodes i upstream of a via the lattice to destination s with inflows of q_{is} on the local network. Therefore, $x_{as} \leq \sum_{i \in N} q_{is}$, so the arc flow per destination is bounded. This implies that $\mathbf{X} \in \mathfrak{R}$. Since by construction the Variational Inequality Problem conditions reflect the equilibrium conditions, the twofold vector $\mathbf{X} = (\mathbf{x}, \xi)$ solves the VIP.

(b) Let $\mathbf{X}^* \equiv (\mathbf{x}_{AS}^*, \xi_{NS}^{D*})$ be a solution of the Variational Inequality Problem. The construction of function $\mathbf{F}(\mathbf{X})$ implies that $\mathbf{x}_A = \sum_{s \in S} \mathbf{x}_{As}$, the equation (41) of the cost of a local strategy and the generalized Bellman equations (42) hold true. Furthermore, the conditions for \mathbf{x}_{AS}^* to be a solution of the VIP corresponds to the equation (41) for ξ_{NS}^D . Therefore the lattice condition of acyclic chaining of strategies holds true. The flow conservation on the node suggests that equation (47) must have a solution, noted $\hat{\mathbf{x}}_{AS}$. In that case, $\hat{\mathbf{x}}_{AS}$ must also be a solution of equation (48), thus it should belong to the

admissible set E_s and it should yield $J_s(\hat{\mathbf{x}}_{AS}, \mathbf{X}^*) = 0$. Indeed, since $\mathbf{X}^* \equiv (\mathbf{x}_{AS}^*, \xi_{NS}^{D*})$ is the solution of the VIP over E_s it minimizes function J_s , hence $J_s(\mathbf{x}_{AS}^*, \mathbf{X}^*) = 0$. Therefore, there exists a \mathbf{x}_{AS}^* that satisfies equation (47). That completes the proof that a solution $\mathbf{X}^* \equiv (\mathbf{x}_{AS}^*, \xi_{NS}^{D*})$ of the VI problem is a traffic equilibrium.

Theorem 3 – Existence of traffic equilibrium

The Variational Inequality Problem has at least one solution in \mathfrak{R} which corresponds to a traffic equilibrium.

Proof. Lemma 12 states that the sets E_s and $B_{i(s)}^D$ are nonempty convex polytopes. Hence the admissible set of the VI problem, $\mathfrak{R} \equiv \prod_{s \in S} E_s \times \prod_{s \in S, i \in N} B_{i(s)}^D$ is also a nonempty compact set. Furthermore, the cost function $F(\mathbf{X})$ is continuous: the component $\partial J_s(\mathbf{x}_{AS}, \mathbf{X}) / \partial x_{as}$ is continuous with respect to \mathbf{X} and \mathbf{x}_{AS} , and the component T_{NS}^D is continuous according to Lemma 10. However, this is true under the appropriate approximations developed in Lemmas 4, 6, 8 and 9. Since the admissible set of the VI problem is a nonempty compact set and the cost function is continuous, it has a solution. Theorem 2 shows that this solution corresponds to a traffic equilibrium of the CapTA model.

7.4.5 Equilibration Algorithm

A solution algorithm is further proposed. Since the cost function $F(\mathbf{X})$ of the VI problem is continuous, the solution can be reached by successive approximations. This corresponds to the Auxiliary Problem Principle (Cohen, 1988): an iterative algorithm is specified of which the elementary step amounts to solving an auxiliary problem, in which a simpler optimization problem that yields an auxiliary state which is used to enhance the current state. In traffic assignment the Auxiliary Problem Principle is implemented in the two most widely used algorithms: the Frank-Wolfe algorithm (Frank and Wolfe, 1956) and the Method of Successive Averages, MSA (Beckmann et al, 1956).

The Auxiliary Problem Principle consists of a typical step at each iteration μ : given the current state $\mathbf{X}^{\mu-1} = (\mathbf{x}^{\mu-1}, \xi^{\mu-1})$, we evaluate $F(\mathbf{X}^{\mu-1})$ and solve the VI problem. That

amounts to minimize the duality gap function $F(\mathbf{X}^{\mu-1}) \cdot (\hat{\mathbf{X}} - \mathbf{X}^{\mu-1})$ which yields an auxiliary state $\hat{\mathbf{X}} = (\hat{\mathbf{x}}, \hat{\xi})$. Let $(\lambda_\mu)_{\mu \geq 0}$ be a decreasing sequence of positive numbers. The next current state $\mathbf{X}^\mu = (\mathbf{x}^\mu, \xi^\mu)$ is derived by the convex combination $\mathbf{X}^\mu \equiv \mathbf{X}^{\mu-1} + \lambda_\mu (\hat{\mathbf{X}} - \mathbf{X}^{\mu-1})$. However, the algorithm should revise ξ^μ to avoid the existence of cyclic paths. In order to ensure the existence of a solution, the arcs $a \approx (i, j)$ such that $\bar{\tau}_{i(s)} \leq \bar{\tau}_{j(s)}$ should be discarded on the basis of the costs of the current state. In such algorithm, the duality gap $F(\mathbf{X}^\mu) \cdot (\hat{\mathbf{X}} - \mathbf{X}^\mu)$ provides a rigorous convergence criterion.

Nevertheless, dealing explicitly with the strategy assignment vector is likely to be computationally costly. An alternative can be a heuristic algorithm, such in Cepeda et al (2006), which focuses on the arc vector \mathbf{x} only and derives vector ξ only on special occasion. To do so, they propose a specific algorithm that starts from a given destination and proceeds backwards to deal with every node. Given vector \mathbf{x}_{AS} , algorithm A of the proof of Theorem 3.2 (Cepeda et al, 2006) decomposes the local vector of arc flows as a positive linear combination of strategy assignments. This is done by progressive subtraction of the projection of its residual flows $x_{i(s)}^d$ along the strategy, as follows:

Initialization: Set $x_{i(s)}^d \leftarrow 0, \forall d \in D(i)$

Main: $\forall d \in D(i), d \neq \emptyset$, do:

$$x_{i(s)}^d \leftarrow \min_{a \in d} \{x_{as} \hat{\phi}_i^d / \hat{\phi}_a\}, \quad \xi_{i(s)}^d \equiv x_{i(s)}^d / \sum_{a \in A_i^+} x_{as}$$

$$\text{and } \forall a \in d \text{ let } x_{as} \leftarrow x_{as} - h_{i(s)}^d \hat{\phi}_a / \hat{\phi}_i^d$$

This algorithm uses the vector of arc flows \mathbf{x}_{AS} in order to evaluate the flows on each local strategy \mathbf{x}_{NS}^D and the local strategy share ξ_{NS}^D . It proceeds in reverse order, compared to the flow assignment algorithm of the auxiliary state. The algorithm A on \mathbf{x}_{AS} enables the evaluation of the duality gap $F(\mathbf{X}^\mu) \cdot (\hat{\mathbf{X}} - \mathbf{X}^\mu)$.

7.5 Conclusion

The chapter discusses the mathematical treatment of some transit assignment models. The first section shows the particular treatments of various approaches of static transit assignment models. There, the conditions of local and global traffic equilibriums are detailed. That mathematic characterization of the traffic equilibrium comes along with the description of efficient algorithms.

Furthermore, a detailed representation of the mathematical composition of the CapTA model follows. Although the line and station models address each model component in detail and with respect to each sub-system characteristics, the consistency is maintained at the system as a whole. The structure of dependencies of the state variables is useful for distinguishing the relationship among them and especially the distinction between the upper and lower network layers. In total 12 Lemmas and 3 Theorems are defined, in order to explicit the connections of these variables and discuss their mathematical properties.

Finally, the characterization of the traffic equilibrium on the upper network layer is pursued. The leg-based representation of the upper layer makes the demonstration of the continuity of the leg cost-flow relationship developed on the lower layer necessary. In addition, the replacement of the leg frequencies with the fictive frequencies demands to further specify the conditions for traffic equilibrium. Thus, a variational inequality problem is formulated on the basis of a twofold vector $\mathbf{X} = (\mathbf{x}, \xi)$ of arc per destination flows and node strategy shares. It is shown that the solution of the VI problem corresponds to a traffic equilibrium. An equilibration algorithm is sketched. Although a heuristic algorithm can be used to reduce the computational effort, a rigorous convergence criterion can be determined in relation with the VI problem formulated.

Chapter 8:

Algorithmic Implementation and Computational Aspects of the CapTA Model

8.1 Introduction

While the mathematic characterization of the traffic equilibrium guarantees a sound modelling framework, it is fundamental to develop or combine efficient algorithms to build a transit simulator. This chapter discusses the new algorithms that accompany the development of the conceptual models and the adaptations of existing algorithms to the particularities of the CapTA model. Three main characteristics are highlighted by these algorithms.

First, the algorithms that address the route choice at the upper layer and the flow assignment are clearly inspired by those proposed in Spiess and Florian (1989) and Ngyuen and Pallottino (1988). However, the research of the optimal strategy and the flow assignment on the hyperpaths carry some necessary adaptations. These modifications stem from the complexity of the route choice model and the frequency adaptations connected to the service availability and the presence of congestion, as described in chapter 6 and 7.

Second, along with the formulation of a line model in chapter 4, an appropriate line algorithm is developed. While the network flows and operational characteristics are external inputs, the model addresses the effect of the passenger flow and congestion on the transit operation and on the cost of the individual trip of a passenger within the line. The line model, as it is developed for the guided transit lines, is based on a treatment of the stations on a topological ordering. It is executed by two core line algorithms: a line flow loading ZIP algorithm for loading passenger flows into routes and vehicles according to the capacity constraints and a

leg costing UNZIP algorithm for the economic evaluation of the actual and generalized time of each access – egress station couple. The chapter describes the line model variant used for the guided transit. The bus routes are treated on the same basis by a simpler line model, which includes only the in-vehicle comfort.

Thirdly, within the line model, we apply the local constraints at each station through a series of local models: the exit model, the in-vehicle model, the platform model and the track model for restrained frequencies. These are implemented in the ZIP algorithm where certain algorithms are called. The PASS_BOTTLE algorithm calculates the passenger stock per egress station and the probability of immediate boarding of that stock; the COMF_ALC estimates the probabilities of a change in the comfort state, particularly the probability of standing on-board passengers to sit and the probability of boarding passengers to sit; and the RSTR_FRQ algorithm calculates the dwell time of each service and the total track occupation at the platform for appropriately restraining the service frequencies. The economic evaluation of the transit trips is made in a reverse topological order. Within the UNZIP algorithm we call the INV_COST algorithm for the evaluation of the cost of the route-legs.

The chapter is divided into five parts and a conclusion. After a description of the structure of the algorithm, the transit network algorithm is presented (section 8.2). Section 8.3 proposes the main algorithms for the line model: the ZIP and UNZIP algorithms. In addition, we explicit the algorithms of the local models: the PASS_BOTTLE, the COMF_ALC, the INV_COST and the RSTR_FRQ. Then, the network algorithms on the upper layer for the composition of the auxiliary state of the network are described (section 8.4). Finally, some elements of the architecture of the software built for the transit assignment model are presented (section 8.5).

Table 8: Basic Notations of Chapter 8

Variable	Definition
N	Set of nodes of the service network, $G = (N, A)$
A	Set of the arcs of service network, $A \equiv (A_S \cup A_R) \cup A_L$
ℓ	Transit line, $\ell \in L$ the set of lines of the transit network
A_ℓ	Set of the arcs of the lower layer on line ℓ
z	Service route z

N_z	Set of stations of line ℓ serviced along the route z
φ_{z0}	Nominal frequency of route z
k_{z0}^R	Total capacity of a vehicle of the route z
k_{z0}^r	Vehicle capacity in passengers with comfort state r of the route z
x_{is}	Passenger flow from the station couple $i \rightarrow s$ on line ℓ
H	Reference assignment period $H = 1$
σ_{is}	Passenger stock from station i to egress station s , $\sigma_i = [\sigma_{is} : s > i]$
n_{zi}	Stock of passengers willing to board route z at station i , $n_{zi} = \sum_{s \in N_z} \sigma_{is}$
π_{zi}	Probability of immediate boarding on route z of the stock n_{zi}
$g_{\ell(i,s)}$	Generalized time of the trip $i \rightarrow s$ on line ℓ
x_A	Current traffic state
\hat{x}_A	Auxiliary traffic state
$[q_{os}]_{o \in O_s, s \in S}$	OD trip matrix

8.2 The Transit Network Model

8.2.1 The Structure of the Network Model

The CapTA model is characterised by a bi-layer network representation. On the lower level, each transit line is modelled as a specific sub-network, with vehicle links (for sojourn and line section between adjacent stations) and pedestrian links (for boarding and alighting). The upper layer is composed of line legs (reflecting the cost of an individual trip within a line) and pedestrian links.

On the upper layer network, a stationary user equilibrium of passenger traffic is defined as a conjunction of passenger choice of minimum cost strategies, passenger assignment to upper layer link, flow conservation in the upper layer nodes and the dependency of generalized time, wait time and frequency on the vector of upper layer link flows. The traffic equilibrium is computed by a Method of Successive Averages (MSA), adapted from Leurent (2012a). The line model of seat capacity can be replaced with the more comprehensive CapTA line model.

The initial flow state corresponds to an empty network ($\mathbf{x}_A \leftarrow 0$). In order to calculate the cost of the upper level network, the line model is called. By using the vector of passenger flows of the upper layer $\mathbf{x}_L = [x_a : a \in A_L]$ of the line legs, the line model yields three vectors that characterise the line leg costs: the average minimum generalized cost, \mathbf{g}_L , the waiting time \mathbf{w}_L and the fictive frequency $\hat{\phi}_L$. These vectors are calculated by the ZIP and UNZIP algorithm of the line model algorithm, described in section 8.3.

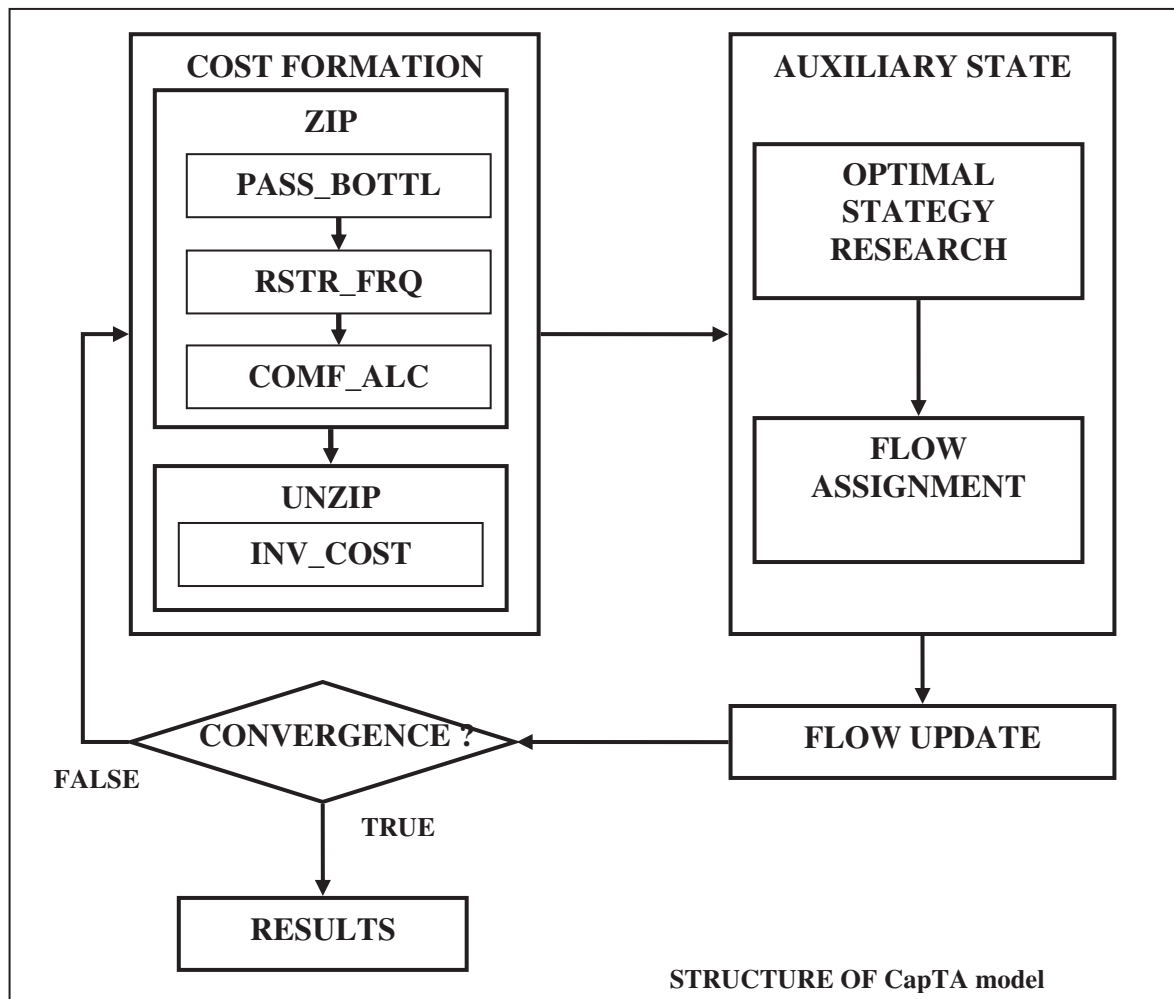


Figure 28: The structure of the CapTA model

The revised cost $g_A(\mathbf{x}_A)$ and the fictive frequency $\hat{\phi}_L$ of the line legs are used for the calculation of the auxiliary flow vector $\hat{\mathbf{x}}_A$. The route-based hyperpaths and the flows are calculated in a classical way as in De Cea et al (1989). The flow assignment procedure assigns the passenger origin-destination flows on the elements of the optimal sub-set, therefore

creating an auxiliary flow state $\hat{\mathbf{x}}_A$. That procedure, modified to account for the bundling of the line legs at the station, is further described in section 8.8.4 .

At the flow update stage, the arc flow vector of the auxiliary state $\hat{\mathbf{x}}_A$ is combined with that of the previous traffic state \mathbf{x}_A and thus creating a new traffic state with a flow vector \mathbf{x}'_A . That procedure corresponds to a convex combination with an adjustable step.

Finally a test is made to evaluate the convergence of the traffic states. If the stopping criteria are met, the loop is broken and the stationary state corresponds to a passenger flow vector \mathbf{x}'_A and the associated costs $\mathbf{g}_A(\mathbf{x}_A)$. Otherwise, the updated flow vector is used as an input for the cost formation procedure. Figure 28 outlines the structure of the CapTA network model, and the local and line algorithms are explicitly illustrated.

8.2.2 The Transit Network Assignment Algorithm

A solution algorithm is proposed for the implementation of the CapTA model. It follows the directions described previously, with four well identified steps. The first network state is acquired by a network flow assignment on the basis of the leg cost of an empty network. Then three steps (cost formation, auxiliary state and flow update) are repeated until the flows have converged to a steady state solution. The convergence is guaranteed by the Method of Successive Averages (MSA) and the existence of a traffic equilibrium (see Chapter 7.4.4). A decreasing sequence of positive numbers $(\lambda_\mu)_{\mu \geq 0}$ with $\lambda_0 = 1$ is used for the convex combination of the arc flows.

Steps 1 to 3 of the following algorithm are repeated until the stopping criteria have been met; the gap between two successive states is sufficiently small. This gap is calculated through the weighted sum of the relative variation of the arc flow vector on the upper layer, according to the following equation:

$$GAP(\mu) = \sum_{a \in A, s \in S} \left(\frac{|\hat{x}_{as}^\mu - x_{as}^\mu|}{x_{as}^\mu} \right) \quad (50)$$

The description of the transit network algorithm follows.

Algorithm MAIN:

Input:

$G = (N, A)$: the service network
 for each arc $a \in A$: $g_a = G_a(0)$ the cost of an arc without congestion, $\hat{\phi}_a$ the fictive frequency of an arc and k_{z0}^r the arc total capacity
 for each service route z
 $[q_{os}]_{o \in O_s, s \in S}$: the OD trip matrix
 $(\lambda_\mu)_{\mu \geq 0}$: a decreasing sequence of positive numbers, with $\lambda_0 = 1$
 ε : the tolerance on the convergence level
 μ : the counter of iterations
 μ_I^{lim} : the maximum number of iterations

Output:

for each arc $a \in A$: \mathbf{x}_A the arc flow vector, \mathbf{g}_A the arc costs, $\hat{\phi}_A$ the fictive frequencies.

Initialization:

Set $x_A = 0$ and $x_{AS} = 0$
 Let $\mu = 0$

Main:

Step 1 – Formation of Costs:

For each line $\ell \in L$, evaluate the cost of the transit legs:

Execute $\text{ZIP}(N_\ell, x_a, \varphi_{z0}, k_{z0}^R, k_{z0}^r, N_z, Z_i) \rightarrow (\pi_{zs}, \varphi_{za})$ **and**
 $\text{UNZIP}(N_\ell, N_z, \pi_{zs}, \varphi_{za}) \rightarrow \bar{g}_{\ell(i,s)}$

Step 2 – Auxiliary State:

Let $\hat{\mathbf{x}}_{AS} = 0$. For every destination node $s \in S$, do:

- Find the optimal hyperpath destined to s under the updated arc costs \mathbf{g}_A
- Load the OD flows $[q_{os}]_{o \in O_s, s \in S}$ on the optimal hyperpaths to s , and calculate the auxiliary arc flows per destination, $\hat{\mathbf{x}}_{AS}$ and
- Let $\hat{\mathbf{x}}_A = \hat{\mathbf{x}}_A + \hat{\mathbf{x}}_{AS}$ for all $a \in A$

Step 3 – Flow Update:

Let $\mathbf{x}_{AS} = (1 - \lambda_\mu) \cdot \mathbf{x}_{AS} + \lambda_\mu \cdot \hat{\mathbf{x}}_{AS}$ and

$\mathbf{x}_A = (1 - \lambda_\mu) \cdot \mathbf{x}_A + \lambda_\mu \cdot \hat{\mathbf{x}}_A$

Step 4 – Stopping Criterion:

If $GAP \leq \varepsilon$ or $\mu \geq \mu_1^{\lim}$ then terminate, else let $\mu \leftarrow \mu + 1$ and go to Step 1

END

8.3 The Line Model Algorithm

The line model algorithm acts as an elaborate cost flow relationship as described in Chapter 4. Each line is treated independently and the effect of the passenger flows on the in-vehicle and on-platform quality of service and the line operations within a transit line is evaluated. A simpler variant of the line model algorithm is used for the bus mode, where the capacity constraints are limited to the in-vehicle comfort.

For each line $\ell \in L$, two algorithms are applied: the ZIP – line flow loading – algorithm and the UNZIP – leg costing – algorithm. The former addresses every station $i \in N_\ell$, where N_ℓ is the set of stations of the line ℓ , in a topological order, by imposing the local capacity constraints. The latter treats every egress station $i \in N_\ell$ in a reverse topological order (from the terminal nodes and upstream) in order to evaluate the cost of the service legs. The following section presents the two core algorithms and the local capacity constraints algorithms.

8.3.1 The Line Flowing Problem and the ZIP Algorithm

The line flow-loading problem consists of assigning the line access – egress trip matrix inherited from the network assignment to each route segment. In addition, it enforces the local capacity constraints, such as the total person's capacity and seat capacity of a vehicle and the station platform temporal occupancy by calling the relevant local capacity models. The algorithm's outputs are the passenger stock waiting at a station, coupled with the probability of immediate boarding on each route, the sitting probabilities at each stage of the trip and the restrained frequency for all transit routes using the station infrastructure – whether stopping or not.

The Line Flow Loading Algorithm proceeds in the direction of traffic along the line, by handling the track links or equivalently the line stations in forward topological order. It consists in one initialization step followed by the sequential treatment of stations. To initialize the algorithm, the residual capacities and the current service frequencies are set at their respective nominal values. The on-board flows $y_{z,a}^{s,r}$ are set to zero. The arc notations are illustrated in Figure 29.

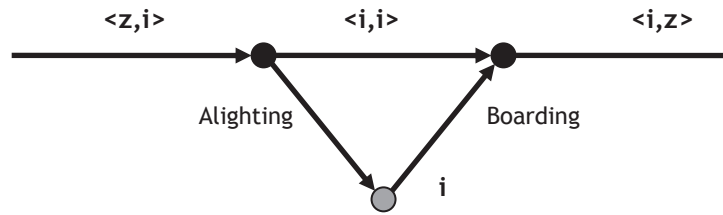


Figure 29: The representation of the vehicle links of a service with reference to the station platform i

The treatment of each station i involves successive steps as indicated below, in which $\langle z, i \rangle$ denotes the interstation link prior to i for service z , $a = \langle i, i \rangle$ the sojourn link and $\langle i, z \rangle$ the next link for that service:

- By service z such that $i \in N_z$, apply the Exit model to yield the exit load e_{zi} and update the residual vehicle capacity by comfort state, $k_{z,a}^r$. Apply the In-Vehicle Comfort model to the riders that remain on board and obtain updated $y_{z,a}^{s,r}$ and $k_{z,a}^r$.
- Apply the Transit Bottleneck model at i to the incoming flows x_{is} for $s > i$ and the residual overall capacities $k_{z,a}^R$, so as to get the boarding flows b_{zis} by destination s and by vehicle of any service z stopping at i and serving s .
- By service z stopping at i and destination s served by z , apply the In-Vehicle Comfort model to get the $y_{z,\langle z,i \rangle}^{s,r}$ for all stations $s > i$ in N_z , together with the residual capacities $k_{z,\langle i,z \rangle}^r$.
- Apply the Track occupancy and Restrained Frequency model to all services passing at i - i.e. using link $\langle i, i \rangle$ or $\langle i, z \rangle$ - so as to determine the service frequency $\varphi_{z,\langle i,z \rangle}$.

The algorithm is structured as follows:

Algorithm ZIP:

Input:

N_ℓ : the set of stations in a topological order of the line ℓ

x_{is} : the passenger flow from an access station $i \in N_\ell$ to egress station $s \in N_\ell : s > i$

$Z_{\ell(i)}$: the set of routes of line ℓ passing from station i

$z \in Z_\ell$: the service routes composing line ℓ and for each route:

φ_{z0} : the nominal frequency of the route

k_{z0}^R : the vehicle capacity in passengers for the service route

k_{z0}^r : the vehicle capacity in passengers with comfort state r

N_z : the ordered set of stations serviced along route z

Output:

π_{zi} : the probability to board of the passenger stock waiting the route $z \in Z_\ell$ at station $i \in N_z$

$\varphi_{z<z,i>}$: the frequency of route $z \in Z_\ell$ at arrival at station $i \in N_z$

Initialization:

Set $H = 1$

For each service route, $z \in Z_\ell$

Let $[y_{zs}^{ar}]_{i \in z} \leftarrow 0$ for all $r \in R$ and $[y_{zs}^{aR}]_{i \in z} \leftarrow 0$

Set $k_{za}^r = k_{z0}^r$ for all $r \in R$ and $k_{za}^R = k_{z0}^R$

Main:

For each station node $i \in N_\ell$ in a topological order:

For each service route $z : i \in N_z$, serving node i , **do**:

If i **the origin station** of the route, **set** $\varphi_{z< i, z >} \leftarrow \varphi_{z0}$ and $k_{z< i, z >}^R = k_{z0}^R$

else set $\varphi_{z< i, z >} \leftarrow \varphi_{z< z, i >}$ the local frequency at station i and apply the Exit model:

Calculate the exit volume $e_{zi} = \sum_{r \in R} y_{zi}^{< z, i > r}$ and update the residual capacity $k_{z< i, z >}^R = k_{z< z, i >}^R + e_{zi}$

For each destination $s \in N_\ell : s > i$,

Set $\varphi_s^B \leftarrow 0$

For each route $z \in Z_\ell$, if $s \in N_z$, **do** $\varphi_s^B \leftarrow \varphi_s^B + \varphi_{z< z, i >}$

```

Execute PASS_BOTTTL( $k_{z<i,i>}^R, \varphi_{z<i,i>}, \varphi_s^B, x_{is}, H$ )  $\rightarrow (\sigma_i, n_{zi})$ 
For each service route  $z \in Z_\ell : i \in N_z$ , do:
    Calculate  $n_{zi} = \sum_{s \in N_z} \sigma_{is}$ 
    If  $n_{zi} = 0$ , then  $b_{zis} = 0$ 
    If  $n_{zi} > 0$ , then
         $\pi_{zi} \equiv \min\{1, \frac{k_{z<i,i>}^R}{n_{zi}}\}$ 
         $b_{zi} = \sum_{s \in N_z} \pi_{zi} \sigma_{is}$  where  $b_{zis} = \pi_{zi} \sigma_{is}$ 
    Let  $y_{zi} = y_{zi-1} - y_{zi}^- + y_{zi}^+$ 
    Execute COMF_ALC( $y_{ij}^z, [y_{za}^{sr}]_{s \in N_z}, k_z'^S, A_z$ )  $\rightarrow ([y_{za}^{sr}]_{s \in N_z}, k_z'^S, p_{zi}^o, p_{zi}^+, d_{zi})$ 
    Execute RSTR_FRQ( $\varphi_{z<i,i>}, T_{zi}^0, T_{zi}^1, \omega_{zi}, e_{zi}, \theta_{zi}^-, b_{zi}, \theta_{zi}^+, H$ )  $\rightarrow (T_{zi}, \varphi_{z<i,z>})$ 
END

```

In order to assess the complexity of the ZIP algorithm, it is broken down to the elementary processes. The initialization process will need $O(|N_\ell||Z_\ell|)$ elementary operations, where $|\cdot|$ denotes the cardinal of the set. Then, by station we require by elementary process:

- The exit model requires $O(|Z_\ell|)$ for alighting and the evaluating the residual capacities and $O(|N_\ell||Z_\ell|)$ for updating the load by service and destination;
- The platform model requires $O(|N_\ell||Z_\ell|) + O(|N_\ell|^2)$ for solving the fixed point problem by a fixed number of iterations of the Newton-Raphson algorithm (PASS_BOTTTL);
- The comfort model requires $O(|R|^2|N_\ell||Z_\ell|)$ for updating passenger loads and residual capacities at each stage either from on-board or boarding;
- The restrained frequency model is of minimal complexity, requiring at most $O(|Z_\ell|)$ elementary operations.

By aggregating the elementary process within the ZIP model, the treatment of a station requires $O(|R|^2|N_\ell||Z_\ell|) + O(|N_\ell|^2)$ elementary operations, of which the dominant term is

$O(|N_\ell|^2)$. The overall complexity of the ZIP algorithm for treating all the station in line ℓ , amounts to $O(|N_\ell|^3)$. It is reduced to $O(|N_\ell|^2)$ for a single-service line.

8.3.2 The Leg Costing Problem and the UNZIP Algorithm

The line leg-costing problem evaluates the individual cost of a passenger for a trip between a couple of access – egress stations. The travel conditions include the generalized in-vehicle time as well as the time spent on the platform – queuing or waiting at an eligible place for the arrival of the next vehicle. These conditions are adapted to integrate the perceived discomfort of each state.

The line flow model provides the physical conditions along a trip: the probabilities of immediate boarding, the in-vehicle seating conditions and the restrained frequency of the service routes at each station. The generalized time stems from multiplying the actual time with the discomfort coefficients, χ_{za}^r , of each travel state.

Then the UNZIP algorithm treats the access stations $i \in N_\ell$ in a reverse topological order from downstream to upstream while updating the in-vehicle generalized cost of the access – egress station couples upstream. The local generalized time is calculated by network element according to the physical time and the passenger flow, in respect with the in-vehicle comfort. The F_DELAY function propagates upstream a vehicle delay due to the restrained frequencies.

The complexity of the algorithm is calculated by breaking it up to its elementary processes. Therefore, the initialization requires $O(|R||N_\ell||Z_\ell|)$ elementary operations. The treatment of each egress station requires $O(|R|^2|N_\ell||Z_\ell|)$. The complexity of the UNZIP algorithm over the entire set of station on the line amounts to $O(|R|^2|N_\ell|^2|Z_\ell|)$.

The algorithm is structured as follows:

Algorithm UNZIP:

Input:

N_ℓ : the set of stations of line ℓ , in a topological order

$z \in Z_\ell$: the service routes composing line ℓ

N_z : the set of stations served by route z

π_{zi} : the probability of immediate boarding on a vehicle of route z of the passenger stock

$\varphi_{z< z,i>}$: the frequency of route z at arrival

Output:

$\bar{t}_{\ell(i,s)}$: the average travel time of the line leg (i,s)

$\bar{g}_{\ell(i,s)}$: the average generalized time of the line leg (i,s)

$\hat{\varphi}_{\ell(i,s)}$: the fictive frequency of the line leg (i,s)

$w'_{\ell(i,s)}$: the average platform waiting time

Main:

For each station node $s \in N_\ell$ in a reverse topological order:

For each service route $z : s \in N_z$, do:

Execute $F_DELAY(t_{za}, \varphi_{za}) \rightarrow (t'_{za})$

Execute $INV_COST(t'_{za}, \chi_{za}^r, \mathbf{p}_z^o, \mathbf{p}_z^+) \rightarrow (\bar{t}_{z(i,s)}, \bar{g}_{z(i,s)})$

For each access station, $i \in N_\ell : i < s$, leading to s ,

Set $(\varphi\pi)_{is} \leftarrow 0$, $\varphi_s^B \leftarrow 0$, $\bar{t}_{\ell(i,s)} \leftarrow 0$ and $\bar{g}_{\ell(i,s)} \leftarrow 0$

For each service route when $i \in N_z$, do:

$(\varphi\pi)_{is} \leftarrow (\varphi\pi)_{is} + \varphi_{z< i,i>} \pi_{zi}$ and $\varphi_s^B \leftarrow \varphi_s^B + \varphi_{z< i,i>}$

For each service route when $i \in N_z$, do:

$\bar{t}_{\ell(i,s)} \leftarrow \bar{t}_{\ell(i,s)} + \frac{\varphi_{z< i,i>} \pi_{zi}}{(\varphi\pi)_{is}} \cdot \bar{t}_{z(i,s)}$ and

$\bar{g}_{\ell(i,s)} \leftarrow \bar{g}_{\ell(i,s)} + \frac{\varphi_{z< i,i>} \pi_{zi}}{(\varphi\pi)_{is}} \cdot \bar{g}_{z(i,s)}$

Calculate the **line leg cost**:

$$\beta_{\ell(i,s)} = \begin{cases} 1 - \left(\frac{1}{(\varphi\pi)_{is}} - \frac{1}{\varphi_s^B} \right) \cdot \varphi_s^B & \text{if } \left(\frac{1}{(\varphi\pi)_{is}} - \frac{1}{\varphi_s^B} \right) < \frac{1}{\varphi_s^B} \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\varphi}_{\ell(i,s)} = \begin{cases} \varphi_s^B / \beta_{\ell(i,s)} & \text{if } \beta_{\ell(i,s)} > 0 \\ \varphi_\infty & \text{otherwise} \end{cases}$$

$$g_{\ell(i,s)} = \bar{g}_{\ell(i,s)} + \frac{\alpha}{(\varphi\pi)_{is}} \text{ and } w'_{\ell(i,s)} = \beta_{\ell(i,s)} \cdot \alpha / \varphi_s^B$$

END

8.3.3 The Local Constraints Algorithms

The ZIP and UNZIP algorithms provide the framework for the separate treatment of the transit lines in the context of the cost formation. Nevertheless, in order to assess the capacity effects, the ZIP algorithm calls specific algorithms, which calculate the effects of the capacity constraints; the waiting time of the passengers at the station, the perceived in-vehicle journey time and the station platform temporal occupancy by the vehicles of the line.

The PASS_BOTTL algorithm is used by the platform model to calculate the partial stock of passengers for each egress station downstream. Given the available capacity at a station, the flow of passengers for each OD station couple and the service routes' characteristics, the algorithm calculates the partial stock of passengers at the station for each egress station, $s > i$, and the stock of passengers willing to board at any service route $z : i \in N_z$ at station i .

The in-vehicle comfort is treated by two algorithms, the COMF_ALC and the INV_COST. These are adapted from Leurent (2012a) to the architecture of CapTA. The COMF_ALC algorithm, used in ZIP, estimates the probabilities of a passenger to change comfort state, given the alighting, on-board and boarding flows at a station. The INV_COST algorithm is called by the UNZIP algorithm and calculates the perceived average in-vehicle travel time of the service legs.

Finally, the RSTR_FRQ algorithm addresses the restrained frequency problem at a transit station. Considering the operating characteristics of each service route $z \in Z_{\ell(i)}$ sharing the infrastructure of line ℓ and the boarding and alighting flows for the stopping services, the algorithm calculates the dwell time of a vehicle of each service route and assesses its impact on the frequency of the services.

8.3.3.1 The Transit Bottleneck Algorithm

The Transit Bottleneck algorithm is structured as a Fixed Point Problem, which is solved through a Newton-Raphson algorithm. An initial state of the partial stock vector is calculated in relation to the flow and frequency of each line leg. We approach the solution vector by a

simultaneous evaluation of the partial stock for every destination. The solution is reached when the solution of the $F(\sigma)$ function is sufficiently close to zero.

Algorithm PASS_BOTTL:

Input:

$k_{z<i,i>}^R$: the relative capacity of the vehicles of route z at station i

$\varphi_{z<i,i>}$: the frequency of the mission z at station i

φ_s^B : the frequency of the services from station i to egress station s

x_{is} : the total flow of the station couple (i, s)

H : the reference period

ε : the tolerance on the convergence level

μ_N^{\lim} : the maximum number of iterations

Output:

σ_{is} : the partial stock of passengers from access station i to egress station s

n_{zi} : the stock of mission $z \in Z_\ell$ at station i

Initialization:

Set $\sigma^o \leftarrow 0$ and $\sigma \leftarrow 0$ and $\mu_N \leftarrow 0$

For each egress station s , do:

$$\sigma^o(s) = \frac{H \cdot x_{is}}{\varphi_s^B} \text{ and } \sigma \leftarrow \sigma_o$$

Set $GAP = 100 \gg \varepsilon$

Main – Newton Algorithm:

While $GAP > \varepsilon$ **and** $\mu_N \leq \mu_N^{\lim}$

$$\text{Calculate } F(\sigma_{is}) = \frac{2\sigma_{is}}{x_{is}H^2} - \frac{x_{is}}{\sigma_{is}} + (\varphi\pi)_{is}, \text{ for all } s > i$$

with available frequency, $(\varphi\pi)_{is} \equiv \sum_{z:s \in N_z} \varphi_{z<i,i>} \pi_{zi}(\sigma)$ and the probability of

immediate boarding, $\pi_{zi} \equiv \min\{1, \frac{k_{zi}^R}{n_{zi}}\}$

Calculate $J(F(\sigma))$ with :

$$\text{the diagonal terms } \frac{\partial F_s}{\partial \sigma_{is}} = \frac{2}{x_{is} \cdot H^2} + \frac{x_{is}}{(\sigma_{is})^2} - \sum_{z:s \in N_z} \frac{\varphi_{z<i,i>}}{n_{zi}} \pi_{zi}(\sigma)$$

$$\text{and the non-diagonal terms } \frac{\partial F_s}{\partial \sigma_{is}} = - \sum_{z:s \in N_z} \frac{\varphi_{z<i,i>}}{n_{zi}} \pi_{zi}(\sigma)$$

Set $\sigma \leftarrow (v - (J(F(\sigma)))^{-1} \cdot F(\sigma))$

Let $GAP \leftarrow F(\sigma)$ and $\mu_N \leftarrow \mu_N + 1$

END

8.3.3.2 The In-vehicle Comfort Algorithms

As mentioned previously, the in-vehicle comfort is estimated by adapting the line model for seat capacity in Leurent (2012a) to the architecture of the CapTA model. Indeed, the model is composed of two algorithms, at the image of the line model in CapTA.

The COMF_ALC algorithm uses the exit flows, in-vehicle load per comfort state and boarding flows in order to calculate the probability of a passenger to change comfort state at a station, namely, the probability of the on-board standing passengers to sit, p_{zi}^o , once seats are liberated by alighting passengers and the probability of the boarding passengers to sit p_{zi}^+ . In addition, the COMF_ALC algorithm includes the calculation of the average density of the standing passengers per trip segment. These calculations are made on the basis of the passenger flow per vehicle of a service z . It is executed at each station during the ZIP algorithm and is structured as follows:

Algorithm COMF_ALC:

Input:

b_{zis} : the passenger flow of the (i, j) access – egress station couple, $b_{zis} = \sigma_{is} \pi_{zi}$

$[y_{za}^{sr}]_{s \in N_z}$: the vehicle passenger flow with comfort state r per egress station $s \in N_z$ on an arc a of a route $z \in Z_\ell$

k_{za}^r : the available vehicle capacity for comfort state r on an arc a of route z . The sitting state is given by \underline{r} (or simply r), while the standing state is \bar{r}

A_z : the surface available at the vehicle for the standing passengers

Output:

- $[y_{za}^{sr}]_{s \in N_z}$: the updated vehicle flow with comfort state r per egress station.
 k_{za}^r : the available vehicle capacity for comfort state r on an arc a of route z
 p_{zi}^o : the probability of occupying a seat if standing on-board at i
 p_{zi}^+ : the probability of occupying a seat if boarding at i
 d_{zi} : the density of standing passengers at i

Main:

Step 1 – Seat liberation: Residual seat capacity after seated passengers alight:

$$k_{z<i,i>}^r = k_{z<z,i>}^r + y_{z<z,i>}^{i,r}$$

Step 2 – On-board competition:

$$y_{z<i,i>}^{\bar{r}} = \sum_{s>i} y_{z<i,i>}^{s,\bar{r}} \text{ for on-board standing passengers}$$

$$p_{zi}^o = \min\{1; \frac{k_{z<i,i>}^r}{y_{z<i,i>}^{\bar{r}}}\} \text{ for probability of sitting at } i, \text{ or } p_{zi}^o = 1 \text{ if } y_{z<i,i>}^{\bar{r}} = 0$$

$$k_{z<i,i>}^r = k_{z<i,i>}^r - p_{zi}^o \cdot y_{z<i,i>}^{\bar{r}}, \text{ updating residual capacity}$$

$$\text{for } s > i, y_{z<i,i>}^{s,r} = y_{z<i,i>}^{s,r} + p_{zi}^o \cdot y_{z<i,i>}^{s,\bar{r}} \text{ and } y_{z<i,i>}^{s,\bar{r}} = (1 - p_{zi}^o) \cdot y_{z<i,i>}^{s,\bar{r}} \text{ for egress flows}$$

Step 3 – Boarding competition:

$$b_{zi} = \sum_{s>i} b_{zis} \text{ for boarding passengers}$$

$$p_{zi}^+ = \min\{1; \frac{k_{z<i,i>}^r}{b_{zi}}\} \text{ for probability of sitting at } i, \text{ or } p_{zi}^+ = 1 \text{ if } b_{zi} = 0$$

$$k_{z<i,z>}^r = k_{z<i,i>}^r - p_{zi}^+ \cdot b_{zi}, \text{ updating residual capacity}$$

$$\text{for } s > i, y_{z<i,z>}^{s,r} = y_{z<i,i>}^{s,r} + p_{zi}^+ \cdot b_{zis} \text{ and } y_{z<i,z>}^{s,\bar{r}} = y_{z<i,i>}^{s,\bar{r}} + (1 - p_{zi}^+) \cdot b_{zis} \text{ for egress flows}$$

Step 4 – Density of standing passengers:

$$d_{zi} = \frac{\sum_{s>i} y_{z<i,z>}^{s,\bar{r}}}{A_z}, \text{ the density of the standing passengers.}$$

END

The INV_COST algorithm is called by the UNZIP algorithm. Its purpose is to calculate the in-vehicle cost of the service legs ending to the station. By egress station of a service route the cost from each access station is constructed in a recursive way by using the sitting probabilities (on-board and boarding) calculated with the COMF_ALC algorithm.

Algorithm INV_COST:

Input:

- t_{za} : the physical time of the arcs of the service z
 χ_{za}^r : the discomfort factor related to state r
 $[p_{zi}^o]_{i \in N_z}$: the probability of occupying a seat if standing on-board at i
 $[p_{zi}^+]_{i \in N_z}$: the probability of occupying a seat if boarding at i

Output:

- $\bar{t}_{z(i,s)}$: the average travel time by station couple $(i,s) : s < i$, at i for the service z
 $\bar{g}_{z(i,s)}$: the average generalized time of the service leg between a station couple $(i,s) : s < i$

Main:

For each egress station $s < i$ in a reverse topological order of the service:

$\bar{t}_{z(i,s)} = \bar{t}_{z(i,j)} + \bar{t}_{z(j,s)}$, the physical time for $a \approx (i,j)$, where $\bar{t}_{z(i,j)} = t_{za}$

$g_{z(i,s)}^r = g_{z(i,j)}^r + g_{z(j,s)}^r$, the leg time for comfort state r (seated)

$\gamma_{z(i,s)} = t_{z(i,i+1)} \chi_{z(i,i+1)}^r + p_{zi+1}^o \cdot g_{z(i+1,s)}^r + (1 - p_{zi+1}^o) \cdot \gamma_{z(i+1,s)}$, the auxiliary variable and

$\bar{g}_{z(i,s)} = p_{zi}^+ g_{z(i,s)}^r + (1 - p_{zi}^+) \cdot \gamma_{z(i,s)}$, the average generalized time of the service leg

END

8.3.3.3 The Restrained Frequency Algorithm

The RSTR_FRQ algorithm is divided into two steps. In the first, the dwelling time of all the vehicles of all the transit services that stop at the station platform are calculated. That dwelling time is used for the evaluation of the actual temporal occupation period of the station platform, H' . All the transit services sharing the line infrastructure $z \in Z_{\ell(i)}$ are included in that calculation, both the stopping and the express services.

In the second step the modelling period is confronted with the actual temporal occupation of the platform. If the latter is greater than the former, a reduction factor of the platform occupancy, η_i , is calculated. It is applied to reduce the nominal frequency for all the transit services $z \in Z_{\ell(i)}$ using the line infrastructure at station i . Hence the frequency is propagated to the downstream stations.

The algorithm is of minimal complexity. At each station we consider the dwell time of the stopping service routes $z : i \in N_z$. Therefore the complexity amounts to $O(|Z_\ell|)$. The algorithm is structured as follows.

Algorithm RSTR_FRQ:

Input:

For a station $i \in N_\ell$ and

for the transit services $z \in Z_{\ell(i)}$ using the track by the platform, we distinguish, (a) the services dwelling at i , $z : i \in N_z$ and (b) the direct services, that use the infrastructure without stopping, $Z_{\ell(i)} - [z : i \in N_z]$.

$\phi_{z< i, z >}$: the frequency of a service z at arrival in the station

T_{zi}^0 : the planned acceptable dwell time ($T_{zi}^0 = T_{zi}^D$, a small value for direct services)

ω_{zi} : the sum of safe separation time and operating margin of $z \in Z_i$ at i

T_{zi}^1 : the operating time of the vehicle of service z at the station i

e_{zi} : the number of passengers alighting from the vehicle of the service z at the stop i

$\theta_{zi}^- \equiv t_{zi}^- / X_z$: a ratio of the time needed for alighting per user t_{zi}^- which is a joint characteristic related to vehicle and station design and X_z the number of doors used.

b_{zi} : respectively, the number of passengers boarding a vehicle

$\theta_{zi}^+ \equiv t_{zi}^+ / X_z$: respectively the ratio of time boarding per user t_{zi}^+ and X_z

H : the reference period used for the frequency calculation

Output:

For a station $i \in N_\ell$ on line ℓ , for each transit service $z \in Z_{\ell(i)}$

T_{zi} : the dwell time of a vehicle of transit service $z \in Z_{\ell(i)}$ at the station i , if $i \in N_z$

$\phi_{z< i, z >}^-$: the frequency of a service z at departure from station i

Initialization:

Let $H' \leftarrow 0$

Main:

Step 1 – Calculate H' :

for each transit service $z \in Z_{\ell(i)}$ **using the line infrastructure, do:**

Add **Safe Separation Time:** $H' \leftarrow H' + \omega_{zi} \cdot \varphi_{z<z,i>}$

Calculate **Dwell Time:**

If z **direct, set** $T_{zi} = T_{zi}^D$

else, $T_{zi} = \max(T_{zi}^0, T_{zi}^1 + \theta_{zi}^- \cdot e_{zi} + \theta_{zi}^+ \cdot b_{zi})$

Add **Dwell Time:** $H' = H_{real} + T_{zi} \cdot \varphi_{z<z,i>}$

Step 2 – Enforce the Local Constraints and Propagate Frequency:

If $H' > H$, **then** $\eta_i = \frac{H}{H'}$

else, $\eta_i = 1$

for each $z \in Z_{\ell(i)}$, **let** $\varphi_{z<i,z>} = \eta_i \cdot \varphi_{z<z,i>}$

and execute $ADD_DT(T_{zi}, T_{zi}^o) \rightarrow (t_{za})$

END

8.4 The Auxiliary Network State

The transit network assignment algorithm suggests that an auxiliary network state is composed of a flow vector, calculated for each destination on the basis of an arc cost vector in two steps. These stem from the algorithms proposed in Spiess and Florian (1989) for calculating the optimal strategies, while adapting them to the particularities of the CapTA model.

In fact, the service network $G = (N, A)$ consists of the set of transit nodes and the arc set $A \equiv A_M \cup A_L$, where A_L states the set of the line legs and $A_M = A_R \cup A_S$ the set of pedestrian links. Indeed, in this case, the CapTA model is limited to the applying the line model. Therefore, the network representation does not use the more consistent station legs

(described in Chapter 5). To get around that, current algorithm considers the presence of platform links $a \in A_P \subseteq A_M$: an arc connecting the main station with the specific platform line. Hence, at the platform level, the passenger chooses the line leg – composed of the transit services between an access and an egress platform – which minimizes his travel time. Nonetheless, at the station level, the presence of real-time information conditions the choice of the platform, in relation with the expected waiting time. The flow will be assigned to the optimal strategy sub-set of a destination s , creating an auxiliary flow vector per destination \hat{x}_{AS} .

8.4.1 The Adapted Optimal Strategy Algorithm

An algorithm similar to the one in Spiess and Florian (1989) is used for the research for the optimal strategies and the composition of the hypertree per destination s . Nevertheless, the current AD_OPTSTR algorithm is adapted to the particular passenger behaviour considered in this model and to the modified representation of the transit network.

First, the route choice hierarchy made by the transit users is discussed. We assume that a passenger at the station concourse level chooses the platform according to the information provided and the knowledge on the operation characteristics. Once on the platform, he boards the first vehicle with available capacity, which stops at his desired egress station, according to the assumptions of the line model. Therefore, that particular two-level route choice implies that from a decision point at the station level, the choice is related to the random arrival of traffic vehicles. The attractive set and the associated choice behaviour is represented by a hyperpath. That is further influenced by the congestion of the line legs, where the use of a mathematic artefact, such as the attenuator of discontinuity β_a of the transit service reduces a congested line leg to a link with instant availability (see Chapter 7.4.3 – Lemma 9).

Second, the algorithm is adapted to abide by the specific leg-based network description of CapTA. At the platform, a user chooses a single line leg of minimum cost to destination. However, a line combination (bundle) is formed at the decision points within the station, based on the type of information available and the presence of congestion. The waiting time is adapted to the line combination and the hyperpath, at the station level.

Algorithm AD_OPTSTR:

Input:

$G = (N, A)$, the network with $A \equiv A_M \cup A_L$

g_a : cost for every arc in A

Output:

\bar{A} : ordered set of the arc in the optimal strategy for each destination

τ_i, w_i : total time and waiting time for each node $i \in N$

Initialization:

Set $\tau_i = \infty, \forall i \in N - \{s\}, \tau_s = 0$ where s is a destination

Set $w_i = 0, i \in N$

Set $A' = A_s$ and $\bar{A} = \emptyset$

Main:

Step 1 – Advance to next link:

If $A' = \emptyset$, stop

else

find $a = (i, j) \in A'$ which satisfies:

$$\tau_j + g_{\ell(i,j)} \leq \tau_j' + g_{\ell(i',j')}, \text{ where } a' = (i', j') \in A'$$

$$A' = A' - \{a\}$$

Step 2 – Update node label:

If $\tau_i \geq g_{\ell(i,j)} + \tau_j$, then

- If $a = (i, j) \in A_L$,

then $\tau_i = \tau_j + g_{\ell(i,j)} - w_{\ell(i,j)}'$ and $w_i' = w_{\ell(i,j)}', \hat{\phi}_i = \hat{\phi}_{\ell(i,j)}$

- else if $a = (i, j) \in A_P$, a platform then

if $w_j' > 0$ do

$$\tau_i = \tau_i \cdot \frac{\hat{\phi}_i}{\hat{\phi}_i + \hat{\phi}_j} + (\tau_j + g_{(i,j)}) \cdot \frac{\hat{\phi}_j}{\hat{\phi}_i + \hat{\phi}_j} - \frac{\alpha}{\hat{\phi}_i + \hat{\phi}_j},$$

$$w_i = \frac{1}{\hat{\phi}_i + \hat{\phi}_j} \text{ and}$$

$$\hat{\phi}_i = \hat{\phi}_i + \hat{\phi}_j$$

else

$$\tau_i = \tau_j + g_{(i,j)}, w_i' = 0 \text{ and } \hat{\phi}_i = \phi_\infty$$

- else

$$\tau_i = \tau_j + g_{(i,j)}$$

$$\bar{A} = \bar{A} + \{a\}$$

Go to Step 1

END

8.4.2 The Adapted Flow Assignment Algorithm

The AD_OPTSTR algorithm defines the attractive set of arcs by an ordered set of links with increasing cost to the destination. Then, the AD_FLASS algorithm simply sweeps the arcs of the attractive set at a reverse order (in decreasing cost to the destination) and assigns the corresponding passenger flow with respect to certain rules.

A simple adaptation of the algorithm in Spiess and Florian (1989) is needed at the station level, so as to integrate the effects modelled. While at the platform level the flow is assigned entirely to the attractive line leg, at the station the passenger is assigned to the hyperpaths on the basis of the fictive frequencies of the line legs, $\hat{\phi}_a$. Hence, with the presence of congestion, this choice is reduced to a pedestrian link.

Algorithm AD_FLASS:

Input:

\bar{A} : ordered set of the arc in the optimal strategy for each destination

$\hat{\phi}_i$: the fictive frequency for each node $i \in N$

$\hat{\phi}_{\ell(i,j)}$: the fictive frequency of the arcs $a \approx (i, j)$

$w'_{\ell(i,j)}$ the waiting times of the arcs $a \approx (i, j)$

$g_{(i,j)}$: cost for every arc in A_s

Output:

x_a : passenger flow for each arc

Initialization:

Set $x_a = 0, \forall a \in A$

Set $X_i = q_{is}, i \in N$, where $q_{is} \geq 0$ for zone origin

Main – Loading:

Do for every arc $a \in A$, in decreasing order of $(\tau_j + g_{(i,j)})$:

If $a \in \bar{A}$, then:

- If $a \in A_p$ and $w'_{\ell(i,j)} > 0$, then:

$$x_a = x_a + \frac{\hat{\phi}_{\ell(i,j)}}{\hat{\phi}_i} \cdot X_i \text{ and } X_i = X_j + \frac{\hat{\phi}_{\ell(i,j)}}{\hat{\phi}_i} \cdot X_i$$

- else, $x_a = x_a + X_i$ and $X_j = X_j + X_i$

otherwise, $x_a = 0$

END

8.5 Software Architecture

The algorithm presented is implemented in a software created for that particular purpose. The CapTA software is programmed in C++ language and built around five executables and libraries. The diagram in Figure 30 shows the dependencies of each component. The CapTA software is composed of 3 executables and 2 libraries. It is thoroughly described in Poulhes and Chandakas (2011).

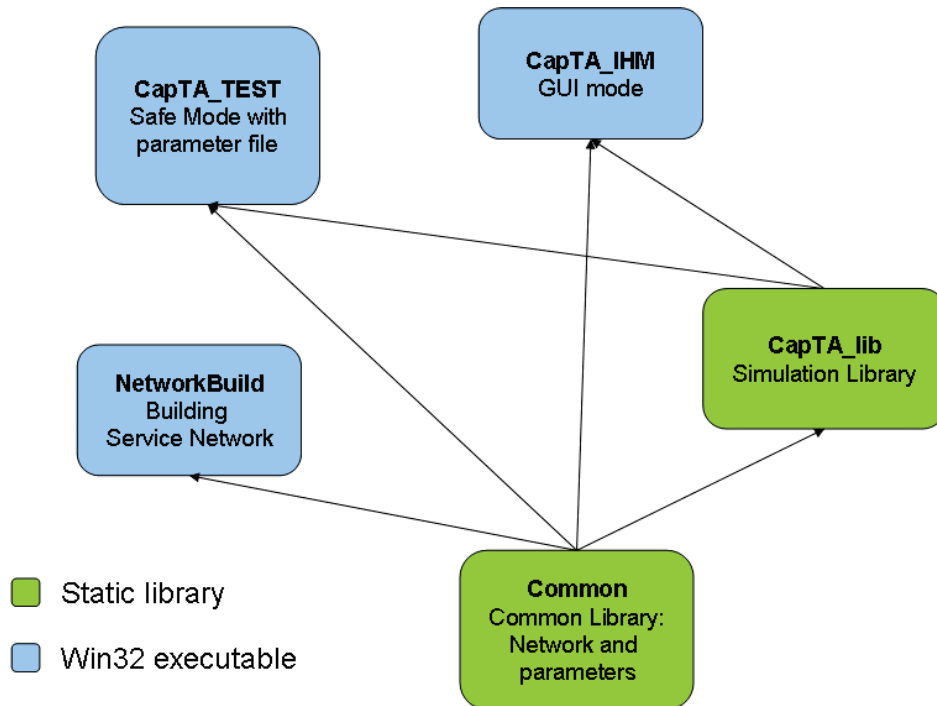
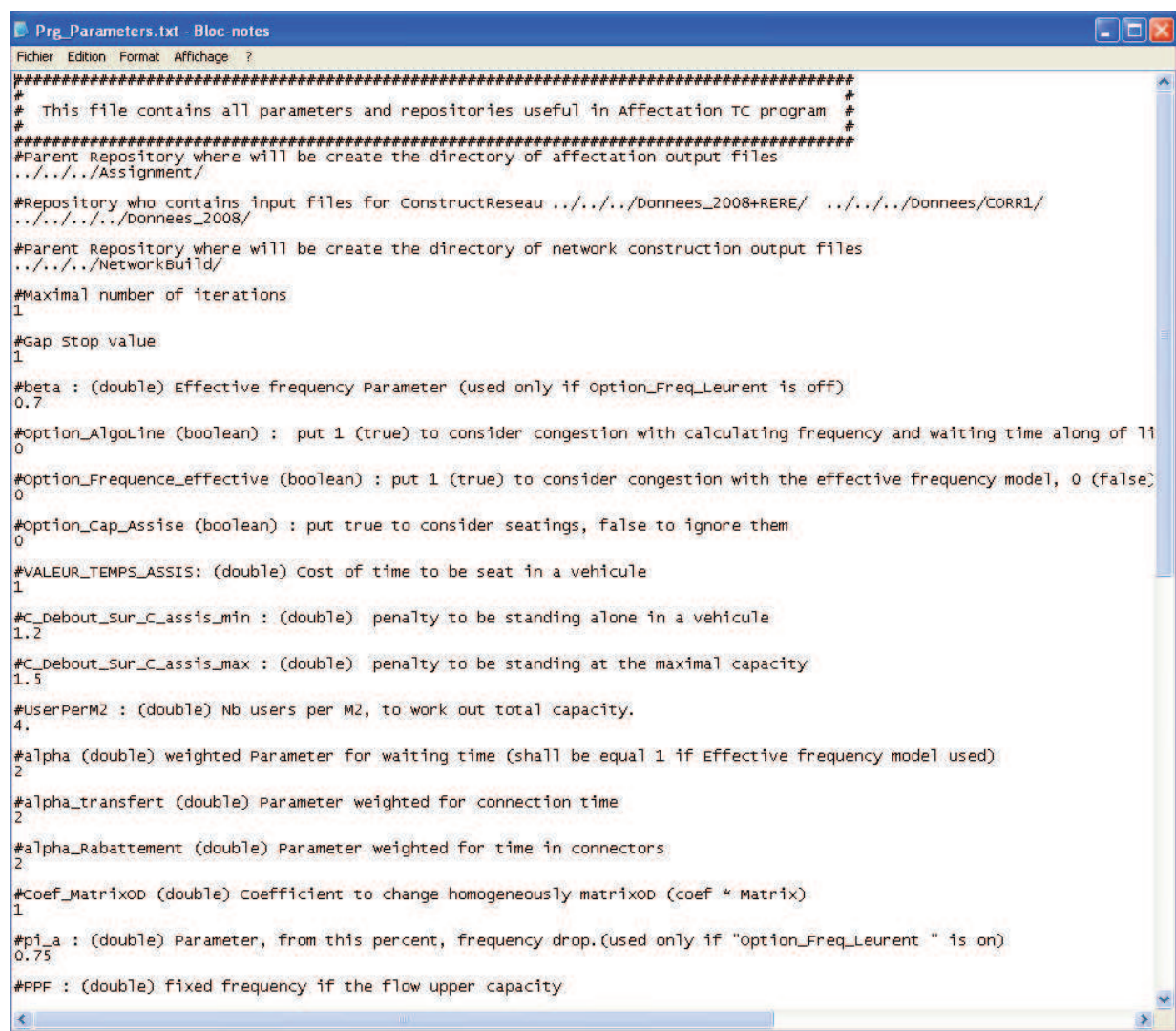


Figure 30: The dependencies between libraries of the software (Poulhes and Chandakas, 2011)

8.5.1 Network Build

The Network Build is an executable, which serves to transform the standard network, as input with a common format, to the service network used later by the assignment executables. It creates the platform access and egress arcs, if they are not present, and associates the physical transit arcs with the composite service legs and line legs. It is an essential part for preparing the network for the transit assignment. It uses the functions stored at the common library.

8.5.2 CapTA Assignment



```
#####
# This file contains all parameters and repositories useful in Affectation TC program #
#####
#Parent Repository where will be create the directory of affectation output files
../../../../Assignment/

#Repository who contains input files for ConstructReseau ../../../../Donnees_2008+RERE/ ../../../../Donnees/CORR1/
../../../../Donnees_2008/

#Parent Repository where will be create the directory of network construction output files
../../../../NetworkBuild/

#Maximal number of iterations
1

#Gap Stop value
1

#beta : (double) Effective frequency Parameter (used only if option_Freq_Leurent is off)
0.7

#Option_AlgoLine (boolean) : put 1 (true) to consider congestion with calculating frequency and waiting time along of 11
0

#Option_Frequence_effective (boolean) : put 1 (true) to consider congestion with the effective frequency model, 0 (false)
0

#Option_Cap_Assise (boolean) : put true to consider seatings, false to ignore them
0

#VALEUR_TEMPS_ASSIS: (double) Cost of time to be seat in a vehicule
1

#C_debout_sur_C_assis_min : (double) penalty to be standing alone in a vehicule
1.2

#C_debout_sur_C_assis_max : (double) penalty to be standing at the maximal capacity
1.5

#UserPerm2 : (double) Nb users per M2, to work out total capacity.
4.

#alpha (double) weighted Parameter for waiting time (shall be equal 1 if Effective frequency model used)
2

#alpha_transfert (double) Parameter weighted for connection time
2

#alpha_Rabattement (double) Parameter weighted for time in connectors
2

#Coef_Matrixod (double) coefficient to change homogeneously matrixod (coef * Matrix)
1

#pi_a : (double) Parameter, from this percent, frequency drop.(used only if "option_Freq_Leurent " is on)
0.75

#PPF : (double) fixed frequency if the flow upper capacity
```

Figure 31: An example of the text file with the model parameters

The CapTA_IHM and CapTA_TEST are used to make a complete assignment on the service network with the CapTA model. The CapTA_IHM uses a graphic interface to introduce the

inputs and the parameters required by the software. A window illustrates graphically the evolution of the assignment procedure. The CapTA_TEST uses a text file for the same purposes, acting as a safe mode for executing the assignment software. They both make use of the functions described in the CapTA_lib assignment and common library.

8.5.3 The Libraries

The objects of the CapTA software are concentrated in the two static libraries, the Common and CapTA_lib. The Common is used for the objects, which are called by both the assignment and the network building phase. It stores the parameters and the network classes (arcs, legs etc.). It also contains some general utility functions.

The CapTA software is polyvalent, offering the possibility to program and run a transit assignment model, other than the CapTA model presented here. Capta_lib, as shown in Figure 32, contains the basic algorithms for running a transit assignment with a different network representation, arcs, service legs and line legs. From that starting point, various models may be included, such as the Effective Frequencies model described in Cepeda et al (2006).

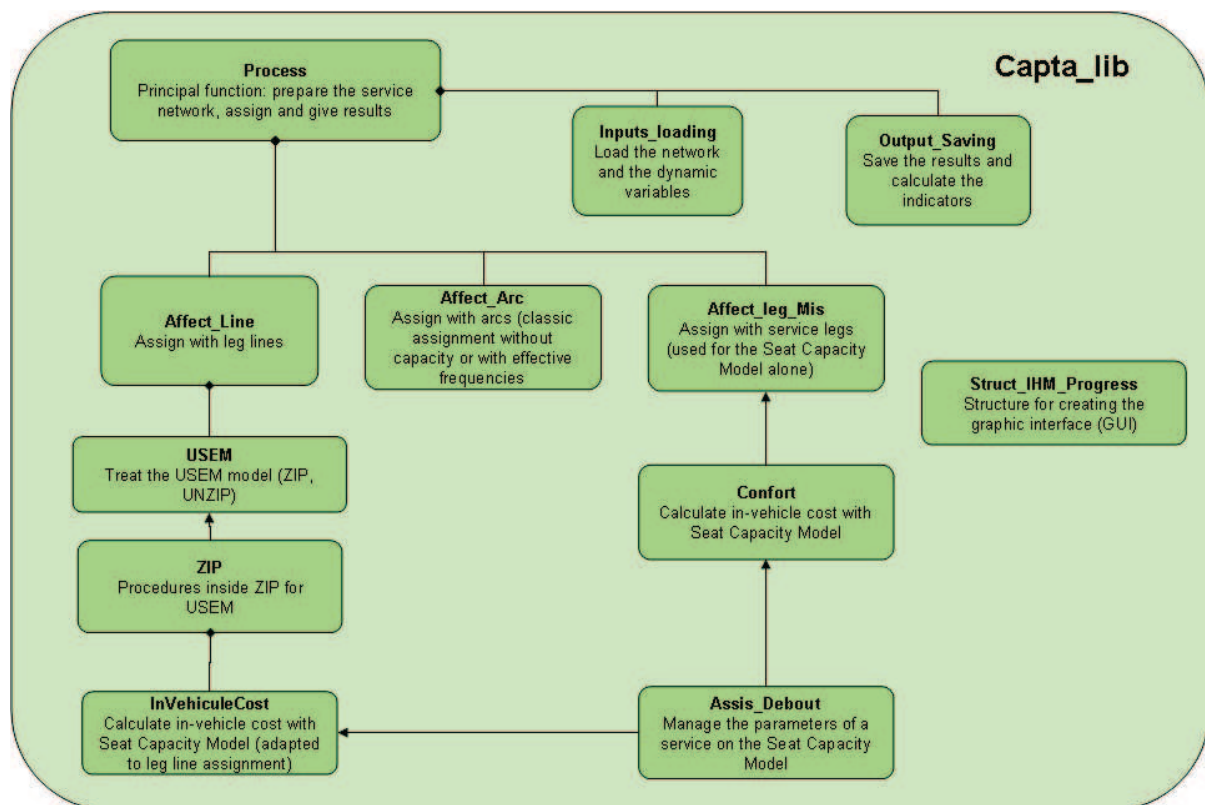


Figure 32: The VTOC of the Capta_lib architecture (Poulhes and Chandakas, 2011)

For a typical transit assignment with CapTA, the program will call the process, which handles the entire assignment procedure. It is supposed to call the appropriate procedures to load the attributes (Inputs_loading), execute the assignment phase using a line leg network description (Affect_line) and then calculate some indicators and export the results (Output_Saving). The Affect_line procedure runs the assignment procedure by calling the USEM for the cost formation phase. That in its turn uses the ZIP and InVehicleCost processes to estimate the effects of the capacity constraints. A software architecture based on 5 libraries is used to distinguish the functions used only occasionally and mutualise in one (CapTA_lib) those used more commonly.

8.6 Conclusion

The algorithm and the software implementation of the CapTA model are presented in this chapter. Its algorithmic implementation is structured at three levels – the network, the line and locally at the station – each one endowed with the appropriate set of algorithms. At the network level, we treat the line legs and the private arcs using a set of algorithms similar to those described in Spiess and Florian (1989). While some adaptations to the specific network description and route choice of the model are needed, such classic framework allows us to focus on the other aspects of transit assignment under capacity constraints.

The main breakthroughs concern the cost formation algorithm at the line level. Two algorithms are called: the ZIP algorithm for the passenger flow loading and the enforcement of the capacity constraints and the UNZIP algorithm for the evaluation of the perceived generalized cost of the line legs. Although the double passage on each line may increase slightly the overall calculation time, two advantages are highlighted. First, the transit operation is internalized and the line interfaces are clearly defined; that is consistent with the characteristics of the line system. Second, the ZIP – UNZIP algorithm couple is a coherent approach which addresses the capacity effects in an endogenous way inside a transit line. The line sub-model is extensible; the addition of specific local constraint models is possible if it does not affect significantly the model architecture.

The architecture of the model allows then a structured treatment of the line's local constraints at the station. Once the inputs and outputs of each local capacity model are determined, these

local models are called in the chronological order of the stopping process, as described in the line model. The capacity models are independent from one another and it is possible to isolate the effects, giving the visibility and the ability to control the modelling process.

The CapTA model is built as a structured algorithm where the boundaries of each operation are well identified. Thus, a solution algorithm with minimal complexity and great flexibility is provided. On the downside, there is a slight increase in the running time for the entire assignment procedure, which can be treated with the parallelization of some processes, such as the cost formation per transit line and the flow assignment per destination.

PART IV: APPLICATION OF THE CapTA SIMULATOR

Chapter 9:

A Simulation of the Transit Network of the Paris Metropolitan Region

9.1 Introduction

Once the transit assignment model is described and its behaviour is tested in simple networks, it is desirable to apply it to a large-scale network. Paris, capital of France, and the Ile-de-France region cover an area of 12 012 km² and accommodate a population of 11,4 million (OMNIL, 2012), making it an ideal field of application.

From the opening of the first metro line in 1900, an extensive public transportation system has been built, with a diversity of competing and complementary transit modes: boat, coach, buses, light rail, metro, RER and commuter rail. With 41 million daily trips by all motorized and non-motorized transport modes, the adequacy of the transportation system is crucial. In addition, the population boom and transport policies favourable to public transportation have put the transit system under pressure and the saturation of some of its elements make the travel conditions intolerable (Le Figaro, 2010).

The analysis of the results does not pretend to be a diagnosis of the state of the Greater Paris transit network. It rather acts as an assessment of the simulation capabilities of the CapTA simulator on a large-scale network and as a means to further investigate the model's behaviour. Moreover, these results serve as the basis for the elaboration of specific indicators mirroring the travel conditions on the network.

The chapter is articulated in 8 parts and a conclusion. First, the transport demand (section 9.2) and the transit supply (section 9.3) of the application field, the Paris Metropolitan Region are

presented. In section 9.4 the model variants as well as the simulation parameters and characteristics are described. Thereafter, sections 9.5-9.8 present the results from the simulations and the comparison among the model variants. The passenger flow and the traffic state are dealt in section 9.5. The effect of the passenger flow on the transit operation is treated in section 9.6. Finally, sections 9.7 and 9.8 deal with the impact of the capacity effects on the passengers. We conclude by an assessment of the capability of the CapTA model to handle a large-scale transit network and discuss the potential developments.

9.2 The Transport Demand in the Paris Metropolitan Region

Along with its big population, the Greater Paris accommodates massive transport flows. According to the last global travel survey – the EGT 2010 – (OMNIL, 2012), in 2010 we observed 41 million daily trips by all transport modes (not-motorized and motorized) of which 39% were pedestrian trips. The public transportation was involved in 34% of the motorized trips, with significant spatial diversity; its modal share was up to 73% if the trip has an origin or destination at the centre of the agglomeration.

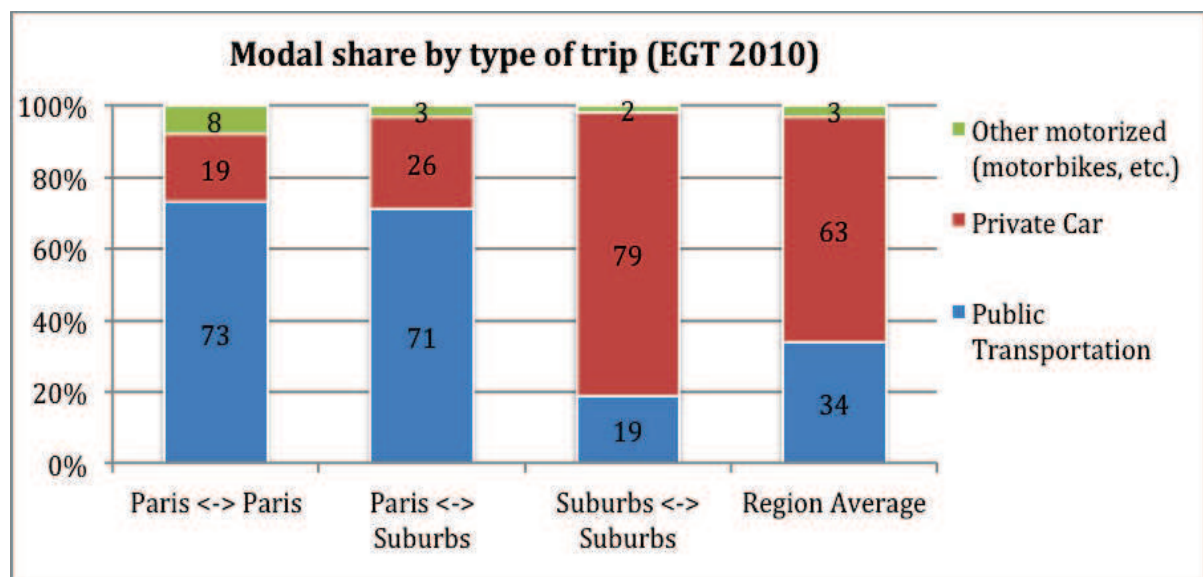


Figure 33: Modal share of the motorized trips in Greater Paris by geographic relation (source: OMNIL, 2012)

Figure 33 illustrates the modal shares of the motorized trips, grouped by geographic relation. While the car captures nearly 80% of the motorized trips inside the suburbs, within Paris (*intra-muros*) and for the radial trips – where Paris is at one end – public transportation captures over 70% of the volume, due to a dense network and its greater efficiency compared to car.

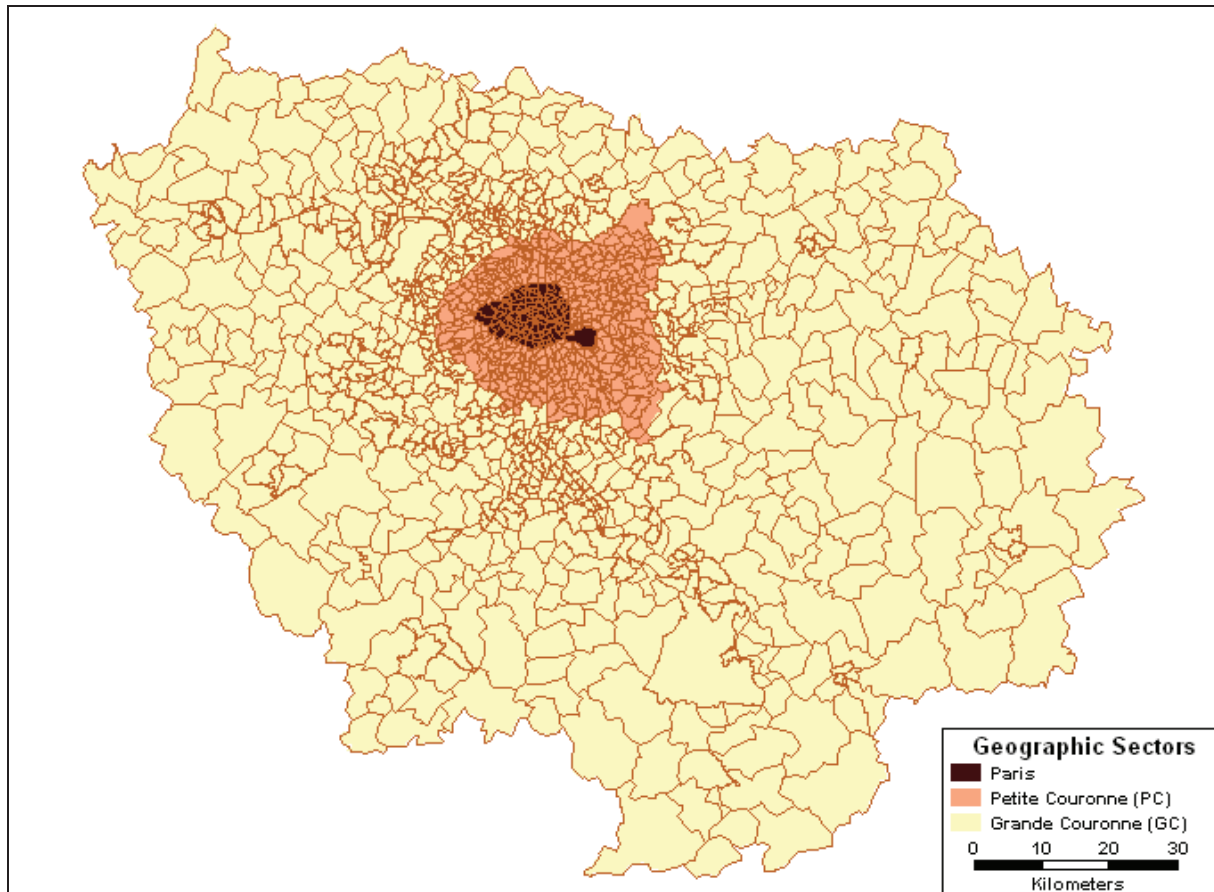


Figure 34: The map of the demand zones of the Paris Metropolitan Region by geographic sector

The simulation of the Paris Metropolitan Region presented in this chapter concerns exclusively the public transportation modes. The input consists of the data distributed by the DRIEA (the Regional Directorate for Infrastructure and Development). In particular, the Greater Paris is divided into 1305 emission and reception zones with the same number of zone centroids. The 1305x1305 Origin – Destination (OD) matrix concerns the morning peak hour trips by public transportation in 2008. By excluding 84 367 intra-zone trips (with the same origin and destination), the typical hour consists of 1 146 976 public transportation trips.

Despite the fact that the Origin-Destination matrix corresponds to the public transportation trips made in the most loaded period of the day, the morning peak hour, preliminary assignments made with CapTA exhibited only minor congestion issues on the 2008 reference. In order to test the ability of the CapTA model to represent situations of critical saturation the initial OD Matrix was homogeneously multiplied by a 1,3 factor. It can be interpreted as a representation of the hyperpeak flow within the morning peak, but its main purpose is to test CapTA for high transport flows.

9.3 The Transit Supply in the Paris Metropolitan Region

The Greater Paris is characterised by a dense transit system and a great diversity of modes and types of services. From the metro- and bus-dominated central area, passing from the RER which cuts across the agglomeration to connect the suburbs between them and with Paris, to the standard commuter rail connecting the 6 central Paris train stations to the near and distant suburbs, we observe a significant variability in the frequency and the commercial speed of the services. Table 9 summarizes the main characteristics of the Greater Paris transit system by transport mode.

Table 9: The characteristics of the Greater Paris transit system

Transit Mode	Nb of Lines	Station or stops	Total length (km)
Rail	13	443	1 401
Metro	16	301	215
Tram	4	71	42
Bus	1 372	30 100	23 263
VAL	2	8	11

The transit supply used in the CapTA model is based on the data provided by the DRIEA. Some transformations were necessary. The service timetable from the initial data was used in order to estimate the hourly frequency of each transit service. The in-vehicle travel time of the transit arcs was calculated on the basis of the actual length of the transit links within the network and an in-vehicle travel time ration per transit route provided. Further modifications were made to in order to adapt the travel time of the transit services to the actual station-to-station travel times.

The transit network is characterised by a multitude of different types of rolling stocks. There are at least 17 types of trains with different compositions and interior design characteristics at the rail network. The same occurs in the metro, with 7 types of vehicle on rail or tyre. As long as the buses and the coaches are concerned it is not simple to define the actual models circulating. We have defined 7 categories of road vehicles with average characteristics (Annex A and Chandakas, 2012). We assigned a type of rolling stock on each one of the transit routes of the Greater Paris network. An additional table was created with the characteristics of the transit services: seat and total capacity, headway, dwell time and number of flow streams for the passenger exchange at stations.

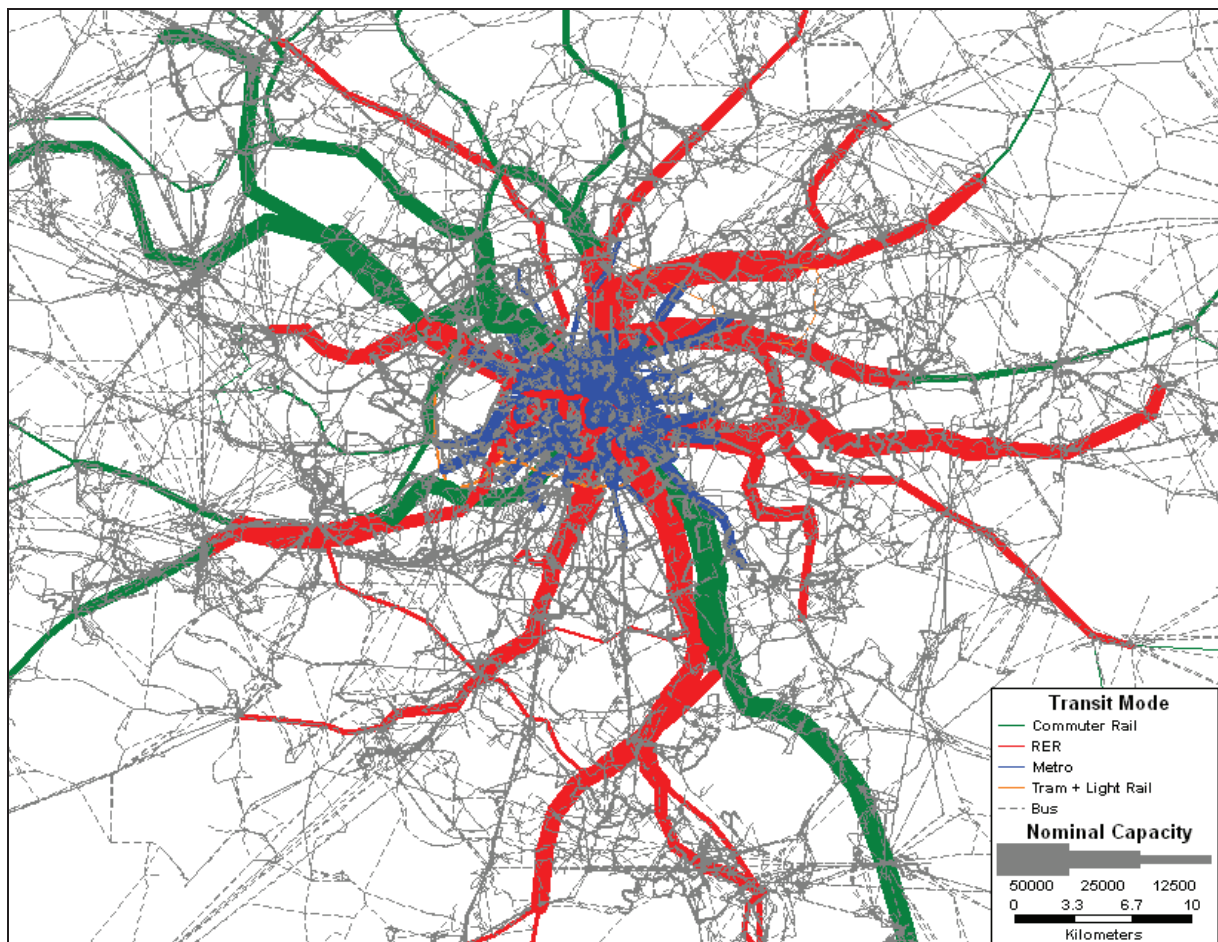


Figure 35: The total capacity of the transit lines, by transport mode (central sector)

In Figure 35 the line capacity per transit mode is shown on the central section of the Greater Paris transit network. The green and red lines designate the commuter rail and RER, while the blue line corresponds to the metro lines. The light rail (and tramway) is coloured in orange,

but it is masked by the other lines with greater capacity. The bus services are indicated in grey.

The final service network is transformed into the calculation network, which will be used in the transit assignment. The modifications mainly concern the guided transport modes (rail, metro, tramway and VAL) and their stations. The transit lines are identified (Chapter 4) together with the station platforms for each direction. These are distinguished by adding a platform walking arc between them and the station concourse level, for each direction as described in Chapter 8.4. When the simulation includes in-vehicle comfort on the buses, service legs are added to the calculation network.

Table 10: The main characteristics of the modelled transit network

Transit Supply	
Directional Transit Lines	95
Transit Services (all modes)	4 742
of which affected to transit lines	259
Station Platforms	1 889
Zones	1 305

Then, the bi-layer representation of the transit services stems from the initial network, by adding the relevant network legs. A line leg is added between each station couple within the line.

Table 11: The graph elements of the service network

Service Network Elements	
Initial Nodes	17 413
Service Network Nodes	159 447
Initial Arcs	64 558
Standard Arcs Representation	296 525
Line Legs	30 729
Total Legs and Arcs (for upper layer assignment)	307 694

Table 11 summarizes the network used as input for the network assignment model and the line models. We observe that by substituting the line legs to the service arcs on the transit lines, there is only a slight increase in the number of link elements (+3,8%). In other words, some 19 560 transit arcs (boarding, in-vehicle sojourn and interstation and alighting arcs) are

replaced by 30 729 line legs (+57%) that connect the station couples within each transit line. The number of line legs depends only on the number of stations on a transit line, whereas the arcs also depend on the number of alternative transit services within a transit line.

9.4 Simulation Characteristics and Model Variants

The Greater Paris transit network was simulated by 4 variants of the CapTA model. The purpose is to investigate how the various effects contribute to the passengers' path choice and to the traffic state of the transit network. Hence, the model variants are compared with a reference situation: the unbounded model (UC). Hereby, the model variants are defined as follows:

- **UC:** It corresponds to the unbounded model, where no capacity constraints apply. The passenger flow under these free-flow conditions select the best strategies, independent of congestion;
- **CNC:** This variant includes the total capacity and platform occupancy constraints, modelled by the transit bottleneck and the restrained frequency models. Simple applications in Chapter 4.4 suggest that the passengers make a trade-off between platform waiting (due to insufficient vehicle capacity) and in-vehicle comfort;
- **CWCF:** It includes all the capacity constraints included in CapTA, namely, the total capacity and restrained frequency for guided transit (rail, metro, tramway) and the in-vehicle comfort for all the transit modes (guided transit and buses). In-vehicle comfort only depends on whether the passenger is seated or standing and the comfort multiplier is fixed;
- **CWCV:** Equivalent to the previous variant, except that the in-vehicle comfort factor is variable, depending linearly on the density of the standing passengers in the vehicles (which cannot exceed the maximum passenger density). It is the most comprehensive simulation. If not stated otherwise, the results of the bounded model correspond to the CWCV model variant.

The main global parameters in a CapTA simulation are the penalty factors to be considered in the generalized cost function. Indeed, the generalized cost of a path is the sum of the actual travel times for each arc in the path weighted by a penalty factor which depends on the importance of that element to the user's trip. In the simulation we distinguish four different types of cost components: access – egress, transfer, waiting and in-vehicle. Each one has a different penalty factor, according to Table 12.

Table 12: Multipliers of physical time

Penalty factors of cost components	
Access and Egress	2
Transfers	2
Waiting in platform	2
In-vehicle comfort	Fixed 1,8 or Variable 1,2-2

The in-vehicle comfort model is able to include a variety of penalty structures: a constant penalty for standing; a fixed or variable multiplier of the physical travel time; or a combination of those. In addition, the generalized cost for seating passengers can also be made to depend on the density of standing passengers. That has not been applied in current simulation. The fixed comfort multiplier is assumed to be equal to 1,8 of the in-vehicle seating time, as suggested in Leurent and Liu (2009).

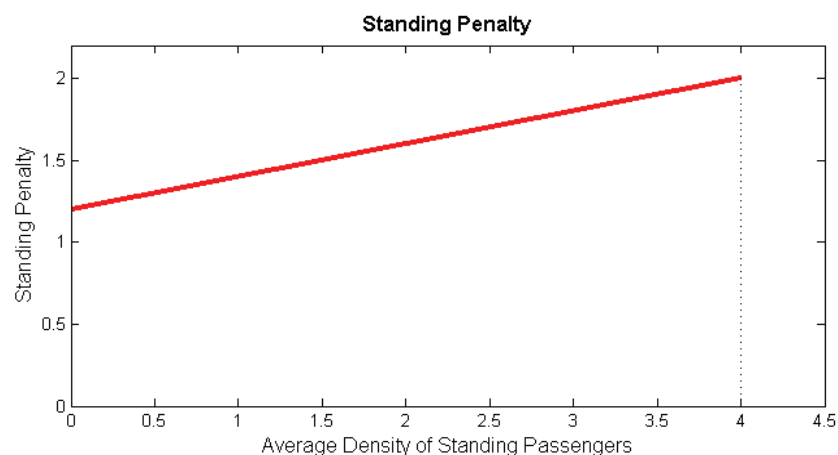


Figure 36: The standing comfort penalty as a function of the density of standing passengers

Some researches (TRB, 2003; Kroess et al, 2007) suggest that the comfort multipliers depend on the density of standing passengers. That is the reason why CWCV is the main variant of

the model. A linear function of the standing penalty factor $\chi_{za}^r(d_a)$ is considered, where d_a is the density of standing passengers for a certain trip segment a . The CWCV variant adopts a linear standing multiplier, varying from 1,2 (for $d_a = 0$) to a maximum of 2,0 (for $d_a = 4p/m^2$), as shown in Figure 36.

Furthermore, an OD matrix multiplier is added as a simulation parameter. It allows multiplying all the cells of the OD Matrix by a given coefficient. As noted above, this coefficient is set at 1,3 in all the simulations in this chapter. All the parameters are included in a simple text file (see Chapter 8.5) and can be easily accessed and modified.

The model was coded in C++ using an object-oriented approach. The simulation runs on a 2,66 GHz Intel PC with 4 GB of RAM. The average run time per iteration amounts to 8 minutes for the unbounded model (UC). Each iteration in the capacitated model (CWCF and CWCV) – extending the in-vehicle comfort also to the bus services – requires about 23 minutes. Within an iteration, the in-vehicle comfort for the buses takes additional 6 minutes and for the guided transport, 7 minutes: the treatment of the in-vehicle comfort alone takes 13 minutes per iteration.

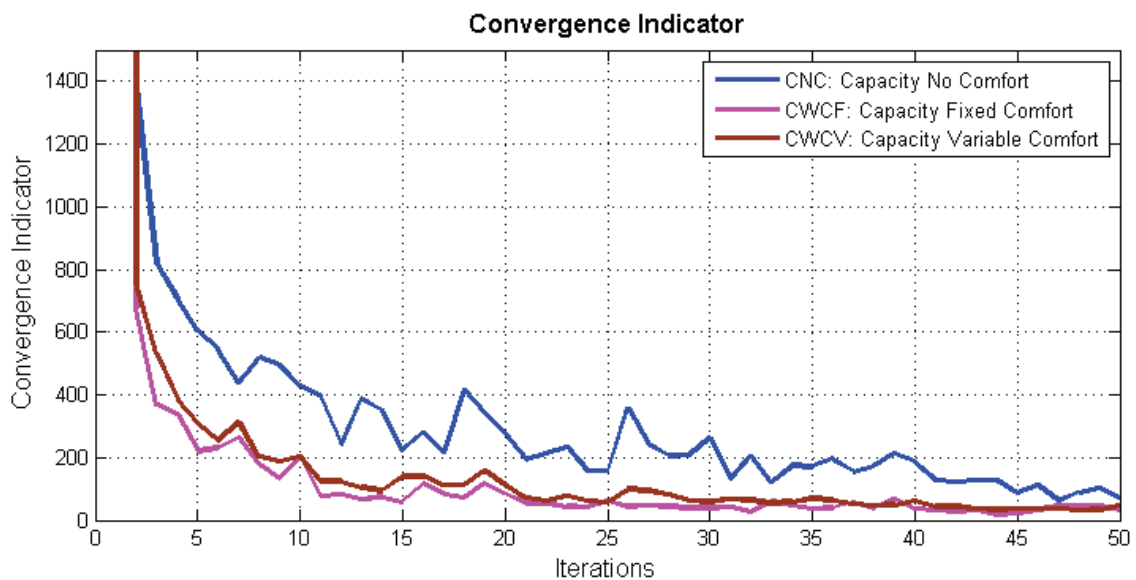


Figure 37: Convergence Indicator of the bounded model variants

The convergence is calculated by the average gap of the passenger flows on each arc (Chapter 8.2.2). Figure 37 illustrates the convergence of the model variants, CNC, CWCF and CWCV.

An acceptable level was reached after 50 iterations with a gap reduced to 1‰ of the initial value. We observe that the inclusion of comfort leads to a quicker convergence. That can be attributed to the most attractive itineraries being more penalized. Therefore, the passengers are dispersed to alternative itineraries.

9.5 Passenger Flows on the Transit Network

The passenger flow is the main output of the transit assignment model. The unbounded variant (UC) corresponds to the itineraries, without congestion, under free-flow conditions. Some structural lines, such as the main commuter lines, RER A (east-west) and RER B (north-south) are particularly loaded, with flow to capacity ratios that can reach 1,7 (or 100 000 passengers per direction per hour, for a capacity of 58 000) for the central section of the RER A. There are 42 000 passengers in excess at the most loaded hour compared to nominal capacity: this, together with the fact that the passengers of the RER A often do not have satisfying alternatives, explains why the RER A suffers from considerable congestion.

For the bounded variants (CNC, CWCF, CWCV), the route choice takes into consideration the effect of the passenger flows on the generalized cost and the path choice due to the local capacity constraints: the vehicle and route total capacity, the seat occupation and the vehicle capacity of the station track.

Figure 38 illustrates the passenger flows on the network (line width) and the flow to capacity ratio (colour) on the guided mode lines for the CWCV variant. The line segments with a flow up to 75% of total operative capacity are in light and dark green. The yellow (resp. orange) arcs designate a flow to capacity ratio comprised between 75% and 90% (resp. 90% to 100%) of the line's capacity respectively. In purple we illustrate the links where the hourly passenger flow exceeds the hourly capacity. It should be noted that for the CWCV variant these segments are limited (9 out of 1750) and the highest flow exceeds line capacity by only 3,75%. It is apparent that the capacity constraints induce the passengers faced with increased waiting time to choose to transfer to alternative paths with lower perceived costs.

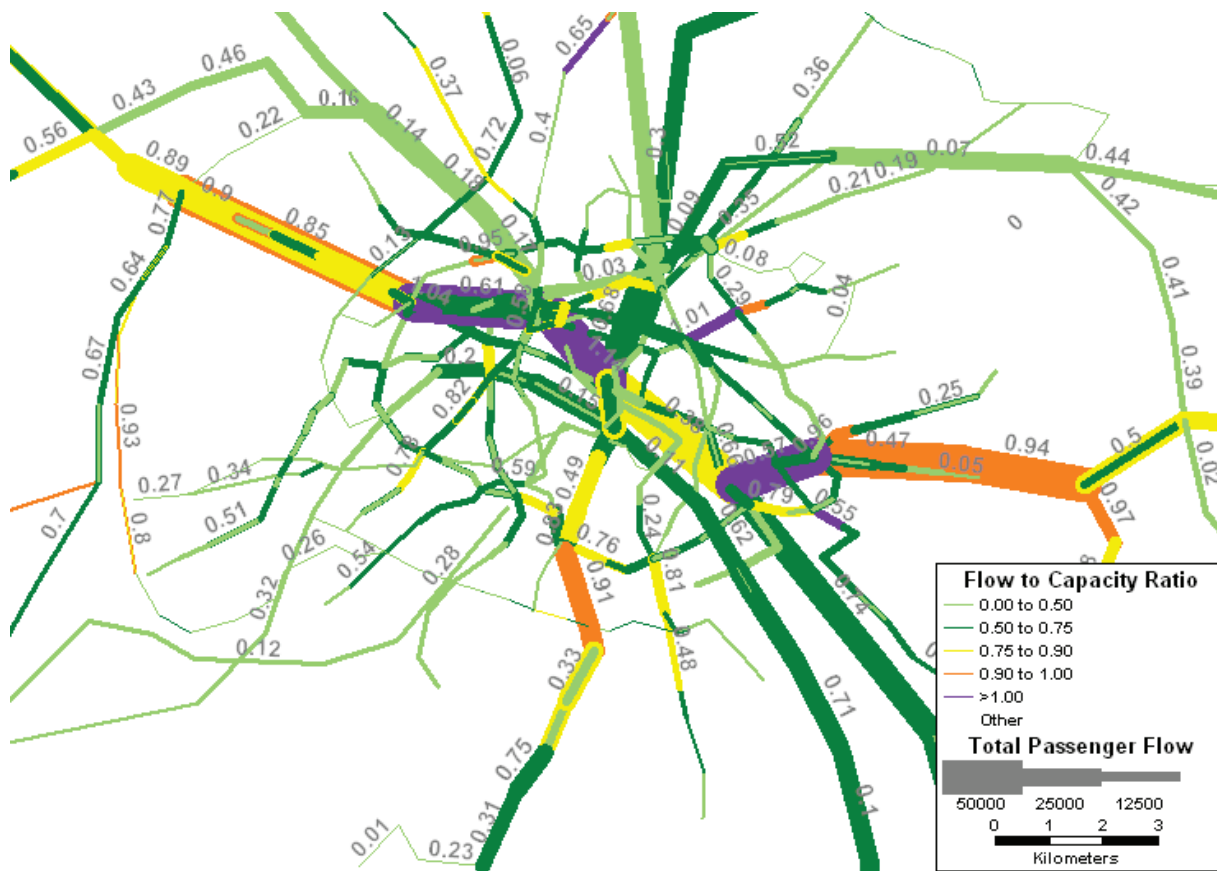


Figure 38: Passenger flows and the ratio of flow to operative capacity for the bounded model

Figure 39 illustrates the distribution of the volume-to-capacity ratios of the rail arcs (light rail, metro and commuter rail) for the four model variants. We observe that the relaxation of the capacity constraints at the network flow assignment for the bounded models (CNC, CWCF and CWCV) results in some arcs being loaded with a passenger flow exceeding their nominal capacity. For the CNC model the distribution of the volume-to-capacity ratios is quasi-identical to the unbounded variant up to 80% flow-to-capacity ratio and a significant number of arcs exceed nominal capacity (48 out of 1750 with the most loaded arc exceeding capacity by 8%). For the unbounded model there are more oversaturated arcs (58) and their volume-to-capacity ratios are more severe, with a maximum of 1,78 of the nominal capacity. The models with in-vehicle comfort show fewer arcs that exceed line capacity (18 and 9 for the CWCF and CWCV respectively), demonstrating that comfort plays an important role in spreading passengers to alternative routes.

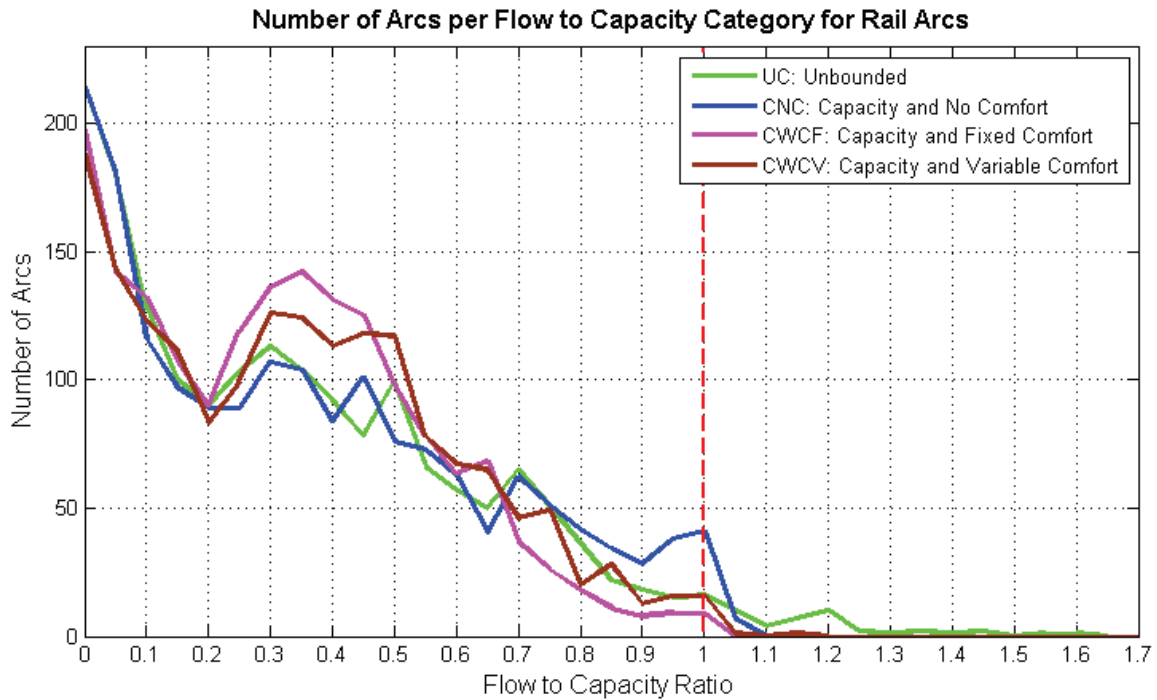


Figure 39: Number of Arcs in relation to flow to capacity ratio

The CNC model unveils the impact of the transit bottleneck model (addressing vehicle capacity) by reducing the flow of oversaturated arcs, just below their nominal capacity. Indeed, at the CNC model variant, a large number of arcs – compared to all other variants – lies between 80% and 100% of the operative capacity. As long as the comfort variants (CWCF and CWCV) are concerned, the distribution of the arcs reaches a peak around 35% of the capacity. That reflects the vehicles' seat capacity. Indeed, the seat capacity roughly corresponds to 40-60% of total vehicle's capacity for rail, 20% for metro and 30% for tramway. Comparing the two comfort variants, we observe a higher concentration of arcs close to the seat capacity for the CWCF, combined with a lower concentration at high flow-to-capacity ratio. That stems from the fixed penalty discomfort of the CWCF variant, penalizing the standing position uniformly, whereas the density is taken into account in CWCV.

9.6 The Operation of Transit Routes Under Capacity Constraints

The track occupancy and the restrained frequency model (Chapter 4) evaluates the impact of passenger flows on the vehicle flowing and route performance. According to the model, the operative frequency, φ_z , of a transit route at a station depends on the dwell time (and therefore the boarding and alighting flows) and the safe separation time between two following vehicles and that for all vehicles of the transit routes sharing the track infrastructure.

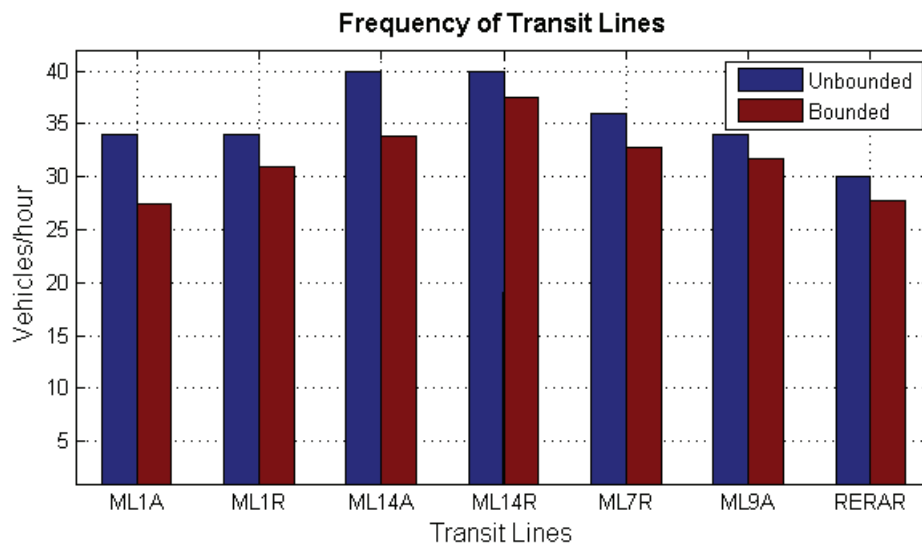


Figure 40: The initial and restrained frequency at the terminal for some rail lines in the Greater Paris transit network.

The simulation of the Greater Paris transit network allows investigating the behaviour of the bounded variants. We can make two observations with concordance to the theoretic model:

- The effects change significantly whether the model is applied to an unbounded model (UC) or a bounded one (CWCV). Indeed, taking into account the capacity effects disperses the passenger flows from the most loaded lines, modifying the structure of the network flows. Therefore, the critical points of the variants (UC, CWCV) are not located at the same stations and the network externalities influence the localization and the magnitude of their effects.

- The restrained frequency effect is not triggered at the most loaded sections, but rather at the stations where the passenger exchange exceeds the exchange capacity. Thus, the restrained frequency can occur upstream of the loaded sections, reducing the line capacity where it is needed the most.

The simulation on the Greater Paris transit network demonstrates that a heavy-duty network with massive passenger flows sees frequencies drop along some of its transit routes. Figure 40 illustrates the drop of the hourly frequency (compared to the nominal frequency) on some structural routes on the bounded CWCV variant. Line M1 (east-west) of the Paris Metro shows a reduction of its hourly frequency from 34 veh/hour to 27,4 veh/hour for the eastbound and 31 veh/hour for the westbound service, while line M14 (automatic, east-west) faces a reduction by 6-15%, from a nominal frequency of 40 veh/hour to 33,9 veh/h for the westbound and 37,4 veh/h for the eastbound service.

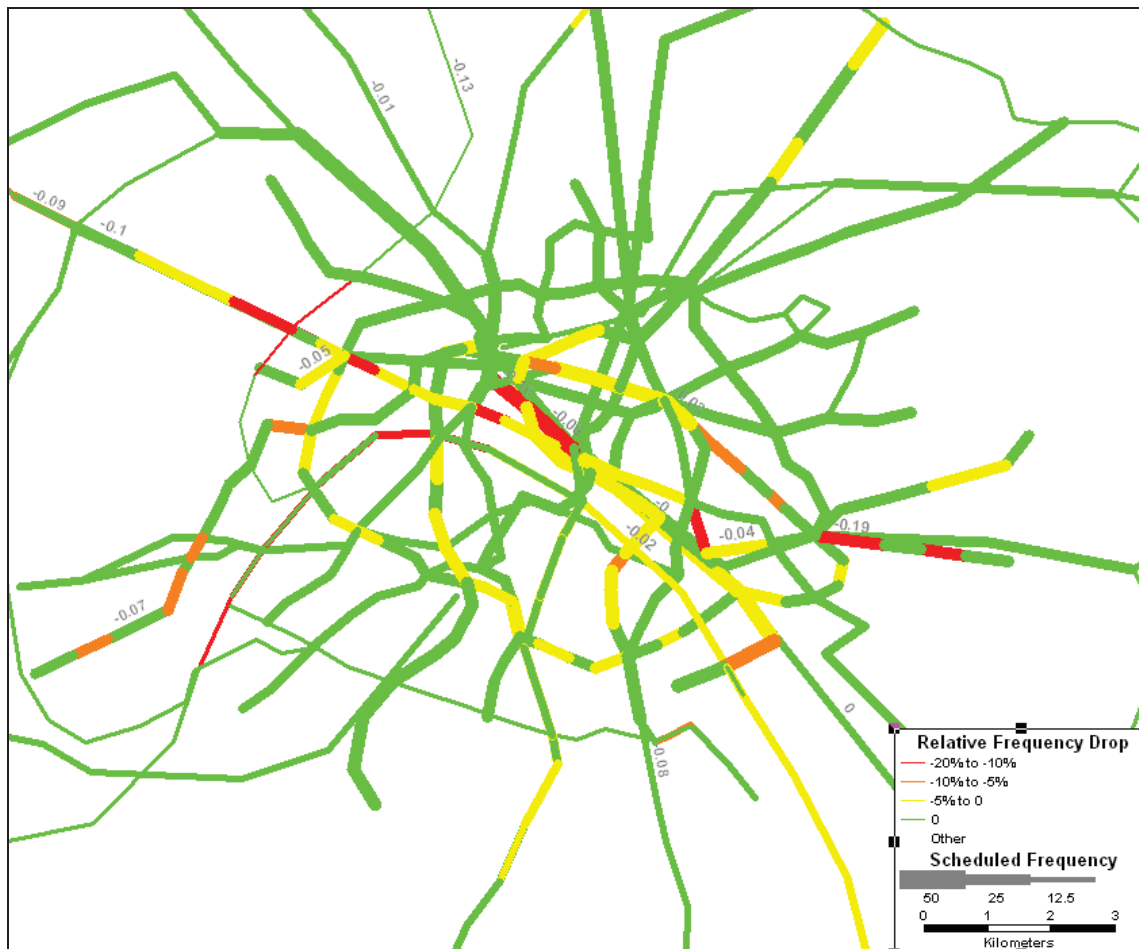


Figure 41: The relative frequency drop due to the restrained frequency in the central sector of the Paris Metropolitan Region

Figure 41 demonstrates the relative frequency drop in the central sector of the Greater Paris transit network. While the great majority of the structural transit lines are not affected, a reduction up to 5% of their initial frequency can be observed in lines such as Line M6, M7 and M9 (depending on direction) and the RER C.

The RER A westbound service faces a reduction of its nominal frequency by 9,3% from 30 veh/h to 27,2 veh/h. Consequently, the service capacity on the RER A westbound service is reduced by 9,3% west of Etoile or equivalently by 5 360 passengers and 1 620 seats per hour. That corresponds to the capacity of 2,1 duplex high-capacity trainsets.

Currently the dwell time function of the restrained frequency model is identical for all the types of modes and platform dispositions. However, more detailed functions can be developed, addressing the interface between platform and vehicle, in-vehicle and on-platform passenger density and the presence of sliding doors at the platform. Thus, it will be able to apprehend more complex effects linked to the line performance.

9.7 Impacts on Users

The average generalized cost on the transit network consists of the waiting, the in-vehicle time and the walking time needed for transfer and access-egress. Table 13 summarizes the average generalized time for the model variants, as well as the details of its components. The composition of the average optimal generalized time (GT) for the CWCV model variant is analyzed. The waiting time component consists of both the initial waiting time and any further waiting time due to potential transfer. Waiting time amounts to 29,1% of the total GT. The most important element is the in-vehicle travel time with 41,5% of the total GT, while the transfer is only 5,6% and the access – egress time forms the 23,8% of the total GT.

Table 13: Average generalized time (in minutes) on the Greater Paris transit network

Model Variant	Optimal GT	Actual TT	Perceived WT	Perceived IVTT	Perceived Transfer T	Perceived Access-Egress	Nb of Transfers
UC	61,56	40,63	18,79	23,10	3,96	15,71	1,42
CNC	62,73	41,40	19,44	23,65	3,98	15,67	1,44
CWCF	69,96	41,77	20,02	29,02	3,97	16,95	1,35
CWCV	68,45	41,70	19,90	28,40	3,88	16,27	1,35
%diff CWCV-UC	11,2%	2,63%	5,9%	23%	-2,04%	3,6%	-5,13%

We observe that in the CWCV variant the average generalized time for a journey on the network increases by 11,2% in comparison to the unbounded model (UC). Most of this change results from the increase of the average in-vehicle travel time (26% and 23% for the CWCF and CWCV variants) in comparison to the UC (and the CNC) variant, due to the inclusion of in-vehicle discomfort penalties. On the other hand, the increase in the waiting time is relatively modest (it amounts to approximately 14 800 additional waiting hours during the morning peak hour). In Chapter 4.4.2.1 we analysed the behaviour of the transit bottleneck model under various traffic states. When the services of a line are saturated the waiting time increases disproportionately to the passenger flow. The combination of the moderate increase in the waiting time and the presence of few arcs whose flow exceed line capacity imply that the overall capacity of the network is sufficient. Note that the average costs discussed here also include the bus routes where only the in-vehicle comfort model is applied.

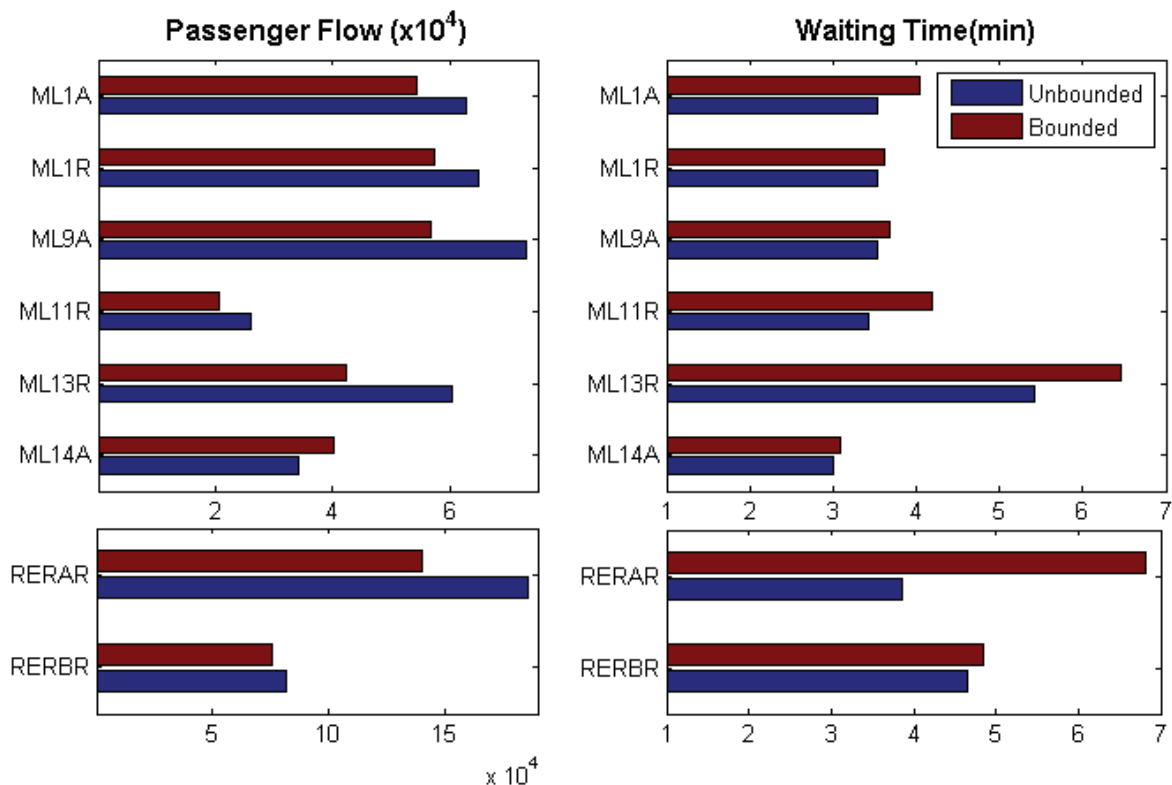


Figure 42: Passenger Flows and Average Perceived Waiting Time on selected lines for the unbounded (UC) and bounded (CWCV) model

The increased waiting time in some lines makes other transit lines attractive to certain passengers and disperses the passenger flows. Congestion is concentrated on a number of

structural lines, inducing a considerable increase in the waiting time of the passengers. The average waiting time of the eastbound service of line M1 increases by 14% (mainly due to the reduced capacity from the restrained frequency), while for the southbound line M11 increases by 22% and both directions of line M14 face an increase of roughly 3%. These are significant effects, as they correspond to an additional waiting time of 230h for the users of eastbound line M1 and 130h for the users of southbound line M11, during the morning peak hour.

Applying the capacity constraints to a simulation modifies significantly the distribution of the passenger flows on the transit lines. Figure 42 (left hand side) illustrates the total number of passengers travelling on each line during the simulation period for the unbounded (UC, in blue) and the bounded model (CWCV, in brown). We observe a reduction of the passenger flows on the second case. In addition, the effect on the waiting time under strict capacity constraints is illustrated at the right hand side of the figure. The increase in the average waiting time is moderate (up to 5%), except for the eastbound RER A (+76%), the southbound M13 (+19%), the M11 (+22%) and the eastbound M1 lines (+14%). This observation implies that a transit assignment model with capacity constraints, such as CapTA, succeeds in both representing passengers choosing alternative routes to avoid the effects of congestion and the impact of public transport level of service of the passengers who do not change their itinerary.

9.8 The Impact of In-Vehicle Comfort on the Users

In the previous section, we observed that the in-vehicle travel time is the most significant component of the total travel time. In addition, the increase of the generalized time between the UC and the CWCV variants is attributed by 3/4 to the increase of the in-vehicle travel time. The in-vehicle comfort model (see Chapter 4.3), which impacts all the transit modes (including buses) allocates the passengers to the seats available and evaluates the cost of a trip, with respect to the sitting probabilities and the standing densities.

The analysis focuses on two elements, the in-vehicle cost and the average standing time. Standing passengers can be found in the buses, the tramlines, the metro, the RER and the commuter rail (CR) lines. However, due to the linear discomfort function adopted in CWCV, the additional cost depends on the density of the standing passengers. The additional cost of

an average trip (sitting and standing passengers combined) on the metro lines varies from 27% of the average actual travel time for line M10 to 68% for the heavy-loaded line M14.

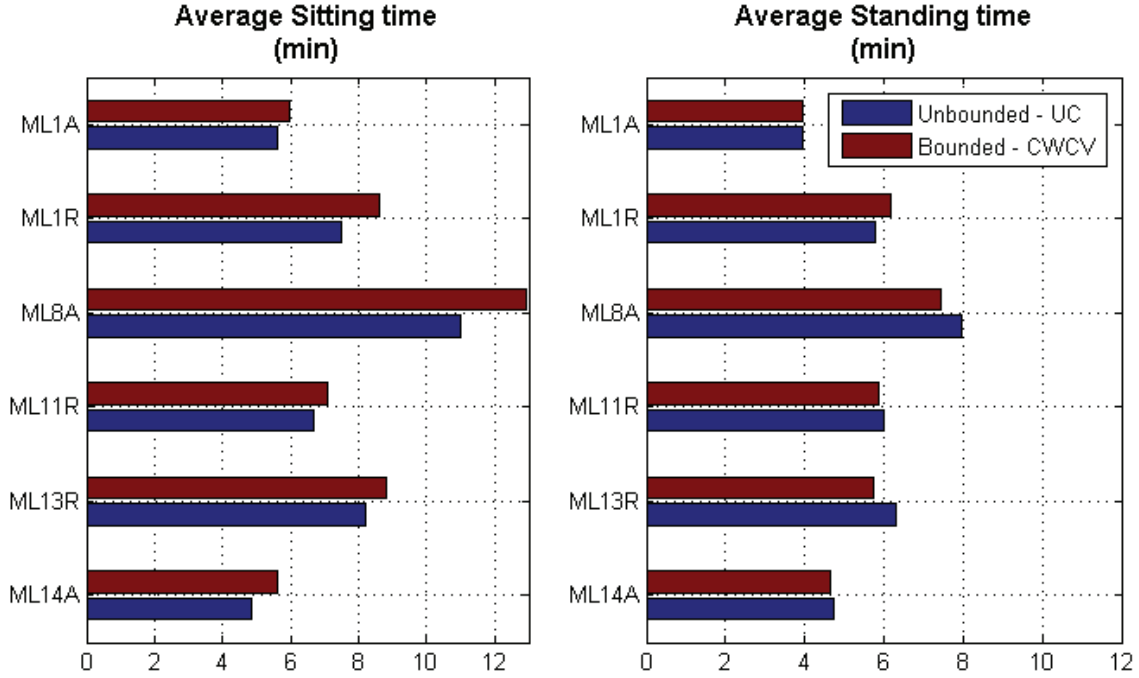


Figure 43: Average Sitting and Standing time on selected lines for the unbounded (UC) and the bounded model (CWCV)

Denote $b_{zi}^{\bar{r}}$ the number of passengers that do not find a seat when boarding and $y_a^{\bar{r}}$ the standing passengers at arc a . Then, we can evaluate an average standing time $\bar{t}_\ell^{\bar{r}}$ per line ℓ as the ratio of the total standing time and the total number of passengers who have been standing during a trip, as follows:

$$\bar{t}_\ell^{\bar{r}} = \frac{\sum_{a \in z, z \in \ell} y_a^{\bar{r}} \cdot t_a}{\sum_{i \in z, z \in \ell} b_{zi}^{\bar{r}}} \quad (51)$$

It expresses the average time a passenger will stand if he does not find a seat at boarding. Figure 43 illustrates the average sitting and standing time for a trip for the UC and CWCV model variants. In the case of the M14 the average standing time is evaluated at 2,5 and 4,5 minutes according to the direction. Nevertheless, the impact of standing is more notable on longer lines, such as M8. There, a passenger who fails to seat at boarding will stand for an average of 7,5 minutes (CWCV model) before sitting or alighting. The comparison shows that

in the case of the bounded model the average sitting time increases significantly (from 10,5 to 12,5 minutes for M8).

To summarize, two observations are highlighted following the analysis:

- The intensity and the length are not always correlated and should not be confused with each other. Although discomfort leads to an increase of the cost of the trip up to 68% for M14, the average standing time amounts only to 4,5 minutes. On the other hand in M8 a passenger stands longer (7,5 minutes) but in better conditions;
- The in-vehicle comfort significantly alters the structure of passenger flows. The comparison of the UC and CWCV variants (Figure 43) reveals an increase of the average sitting time and a decrease of the average standing time of CWCV compared to UC (except for line M1). That, combined with the reduction of the number of transfers (-5%), implies that indeed, if seated, the passengers prefer longer trips rather than frequent transfers (which are associated with the risk of standing afterwards).

9.9 Conclusion

In order to evaluate the capability of the CapTA model to simulate large-scale networks, the model is tested on the Paris Metropolitan Region. Indeed, this transit network, with 4 tramlines, 14 metro lines and 13 commuter rail lines and a central core with line redundancies, offers an ideal field for applying a transit assignment model with capacity constraints. Massive passenger flows are concentrated on main north-south lines, the RER B and metro line M13 and east-west axes such as the RER A and metro line M1 and M14. A number of scenarios were tested (UC, CNC, CWCF, CWCV) in order to detect the behaviour of each one faced to high passenger flows. The main comparisons between the unbounded model variant (UC) and the bounded one (CWCV) reveal the adequacy of the CapTA model for simulating such a network.

The high passenger flow impacts the operation of the transit routes and the frequency is restrained in most of the metro lines of the Paris network. The frequency is reduced up to 15 and 19% on the metro lines M14 and M1 respectively, which induced secondary effects along the lines due to the reduced downstream capacity and the increased passenger stock on each

vehicle arrival. On the passenger side, these effects contribute to the increase of the average waiting time up to 14% and 22% for the line M1 and M11 respectively. Although, the increase seems moderate, it corresponds, only for the eastbound direction of line M1 to additional 230 actual hours waiting during the morning peak hour. In addition to longer waiting times, the passengers face in-vehicle discomfort when they travel standing. The average travel time a passenger stands depends on the topology and on the congestion of the line. Thus, lines without excessive passenger flows may have the longest standing time, such as line M8 where it takes values from 5,5 to 7,5 minutes depending on the direction.

The demand scenario is fictive and we cannot make an extensive analysis of the state of the transit network of the Paris Metropolitan Region. However, the model accomplishes its main objectives:

- It addresses a variety of capacity constraints whose effects can be clearly distinguished and prioritized. In this network, the most significant effect is the in-vehicle comfort, since the overall network capacity is sufficient;
- Although a relaxation of capacity effects intervene on the network assignment, the line model achieves to disperse the flows when they exceed line capacity. On the CWCV variant the maximum flow-to-capacity ratio exceeds capacity by 3,7%;
- The calculation time of an iteration amounts to 23 minutes, mostly for the steps of cost formation and of the optimal strategy search.

The computation time can be reduced if we treat in parallel order the multiple independent processes treated sequentially. Other than the computation efficiency, the model, being modular, can be further developed. A more detailed interaction may be considered between the boarding and alighting flows, the in-vehicle passenger load and passenger stock at the platform as well the vehicle's and platform's architecture and how these components influence the dwell time. Further microscopic phenomena can be added, such as a willingness to board an overcrowded vehicle, if the available capacity is insufficient and the passenger stock is too important, thus affecting the in-vehicle discomfort of on-board passengers. Finally, additional models may be added, first, within the line model, by considering the effect of general traffic conditions to the vehicle's journey time and second, by developing a

station model to capture the effect of passenger flows on the quality and travel time of transfers and access-egress trips.

Chapter 10:

The results of the CapTA model on a restrained network

10.1 Introduction

This chapter provides an application for the CapTA model at the line level, which serves as a showcase of the capabilities of the model. It further demonstrates the level of detail of the modelling approach and investigates the behaviour of the line model. The RER A – the busiest commuter rail line in the Paris Metropolitan Area transit network – is chosen as an application instance for the model. The transit supply stems from the 2008 data of the DRIEA, while the demand on the line is estimated by an assignment of the same model on the Paris Metropolitan Area (Chapter 9).

The CapTA model aims to capture the capacity phenomena related with the vehicle seat capacity, the total capacity and the interplay of passenger flows at access and egress with the dwell time and the service frequency. These phenomena are dealt with by line of operations on the basis of a set of local models yielding specific flows and costs. The line model is described in detail in Chapter 4.

The rest of the chapter is structured in seven parts and one conclusion. First, we present the initial and extracted transit network, the demand used and the services present in the RER A line, before focusing on the passenger flows on the network level. Then, we relate the dwell time and the operating frequency with the passenger exchange flows. In the fifth part, a transit service is used as an example to demonstrate the relation between the residual capacity of the service and the boarding flows and inquire on the in-vehicle comfort. Part six focuses on the

formation and the behaviour of the passenger stock of different services. More precisely, the partial stock to La Defense station is extracted and related to the exogenous flow. Then the effects relative to the generalized cost are illustrated. Finally, we assess the effect of the replacement of the lower capacity one-level trainsets with more spacious duplex ones.

10.2 Transit Network, Trip Demand and Supply of Services

10.2.1 The Initial Network

The CapTA model is first applied to the transit network of the Paris Metropolitan Region (Chapter 9). The transit network is composed of 95 directional railway lines (for the guided transit modes – Train, RER, Metro and Light Rail) yielding 259 guided transit services, and 4 483 bus services. Demand at the morning peak hour involves 1,23 million trips between 1 305 travel demand zones. The network contains about 159 800 nodes and 307 700 arcs and line legs. An acceptable level of convergence was reached after 50 iterations (with a gap reduced to 1‰ of the initial value).

10.2.2 The Restrained Network - RER A Commuter Rail Line

The results are limited, for clarity, to the RER A, the busiest commuter rail line on the network, with a daily patronage often exceeding the one million passenger threshold. The RER A line is composed of 46 stations spread on its five branches and the central trunk. Two branches on the east (Marne-la-Vallée on the northeast and Boissy-St-Leger on the southeast) and three branches on the west (Cergy-le-Haut and Poissy on the northwest and St-Germain-en-Laye on the southwest) converge to the central trunk. Therefore, the east and west suburbs are connected to the centre of Paris and La Defense business district. The restrained network contains 1 200 service network nodes, while the 182 initial arcs are transformed into 1 702 line legs.

The demand on the RER A was extracted by an assignment on the entire network as described previously. An OD multiplier equal to 1,3 and 50 iterations on the bounded model yield the passenger flows used as input on the restrained network. The simulations (variant CWCV of the model) share the same penalty coefficients for the trip components, (see Chapters 9.4 and

10.2.4). The results show about 107 000 passengers circulating eastbound, while 141 000 passengers board on the westbound direction which connects the centre of Paris with La Defense business district on the east. The RER A line totals around 248 000 passengers during a morning hyper-peak hour on both directions.

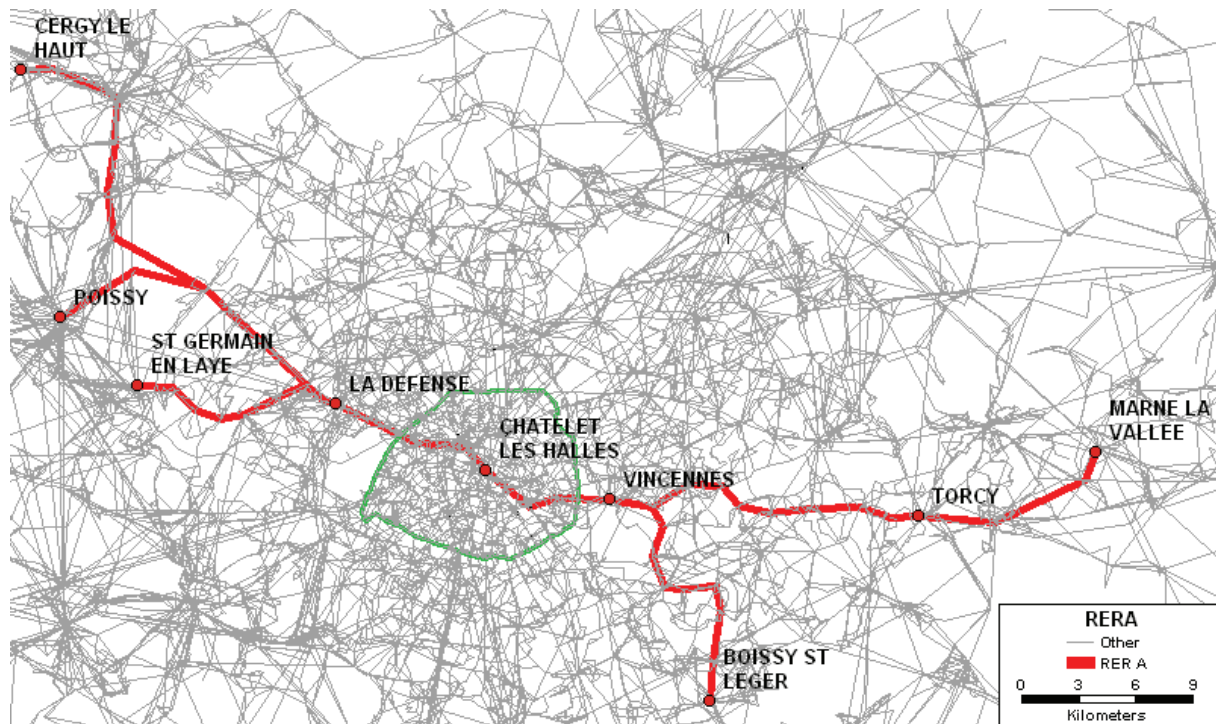


Figure 44: The RER A (red) and the rest of the transit network (grey)

10.2.3 The Westbound Transit Supply

During the morning hyper-peak hour, the westbound transit supply is composed of 18 trains from Marne-la-Vallée (MLV) and 12 trains from the Boissy-St-Leger (BOI) branch that converge at Vincennes station (see Figure 45). Thereafter, 30 trains per hour run on the central trunk until La Defense, before diverging to the three branches, in the number of 5 for Poissy (POI), 5 for Cergy (CRG) and 16 for St Germain-en-Laye (STG) branch (4 vehicles per hour terminate at La Defense).

These trains are associated with 15 different transit services, each one characterized by a scheduled frequency, a stopping policy (express, local, etc.) and a type of rolling stock, detailed in Table 14. This level of detail of the service variation can be captured sufficiently by the line model in CapTA.

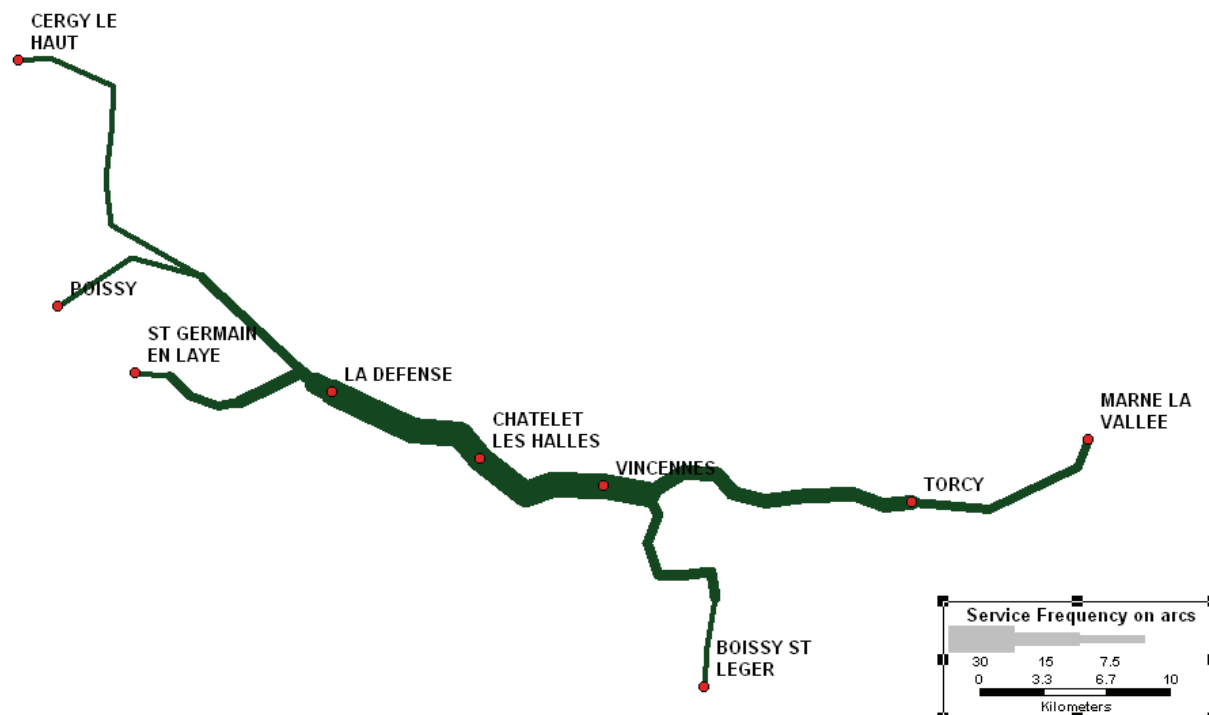


Figure 45: The frequency at departure on the westbound RER A

Three types of rolling stock run on the tracks of the RER A. One-level trains, such as the MS61 with 600 seats and a total capacity of 1 888 passengers per train set and the MI84 with 432 seats and a total capacity of 1 760 passengers coexist with more spacious duplex trains; the MI2N offers 1 056 seats and a total capacity per train set of 2 580 passengers.

The boarding and alighting passengers are evenly distributed on flow streams. A flow stream is defined as a lane at the door of a vehicle. Its width is equal to 0,65m (EgisRail, 2009) and can accommodate exactly one passenger boarding or alighting simultaneously. The number of flow streams indicates the exchange capacity: the maximum boarding and alighting rate of passengers. Per vehicle side, we can find 64 flow streams on the MS61, 72 on the MI84 and 90 on the MI2N (see Table 15).

Table 14: The detailed stopping policy for the westbound transit services

Section	Transit Service Type of Vehicle Frequency (vehicles/hour)	BABY BIPE BROU			TATI TEDY THEO TOUL				UCLA UGIN UNIR UPAC				YRIS XUTI ZARA ZEBU			
		MI84	MI84	MI84	MI84	MI84	MI84	MI84	MI84	MI2N	MI2N	MI2N	MI84	MS61	MS61	MS61
		1	1	2	2	1	1	1	1	1	2	1	4	5	6	1
MLV BRANCH	MARNE LA VALLEE CHES	●	●		●	●		●	●	●	●	●				
	Val d'Europe	●	●		●	●		●	●	●	●	●				
	BUSSY ST GEORGES	●	●		●	●		●	●	●	●	●				
	TORCY	●	●	●	●	●	●	●	●	●	●	●	●			
	LOGNES	●	●	●	●	●	●	●	●	●	●	●	●			
	NOISIEL	●	●	●	●	●	●	●	●	●	●	●	●			
	NOISY CHAMPS	●	●	●	●	●	●	●	●	●	●	●	●			
	NOISY LE GRAND MONT	●	●	●	●	●	●	●	●	●	●	●	●			
	BRY SUR MARNE	●	●	●	●	●	●	●	●	●	●	●	●			
	NEUILLY PLAISANCE	●	●	●	●	●	●	●	●	●	●	●	●			
BOI BRANCH	VAL DE FONTENAY	●	●	●	●	●	●	●	●	●	●	●	●			
	BOISSY ST LEGER													●		●
	SUCY BONNEUIL													●		●
	LA VARENNE CHENNEVIERES													●	●	●
	CHAMPIGNY													●	●	●
	LE PARC DE ST MAUR													●	●	●
	ST MAUR CRETEIL													●	●	●
	JOINVILLE LE PONT													●	●	●
	NOGENT SUR MARNE													●	●	●
	FONTENAY SOUS BOIS													●	●	●
CENTRAL TRUNK	VINCENNES	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	NATION	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	GARE DE LYON	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	CHATELET LES HALLES	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	AUBER	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	CHARLES DE GAULLE ETOILE	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	GRANDE ARCHE DE LA DEFENSE	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	NANTERRE PREFECTURE	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
STG BRANCH	NANTERRE UNIVERSITE												●	●	●	●
	NANTERRE VILLE												●	●	●	●
	RUEIL MALMAISON												●	●	●	●
	CHATOU CROISSY												●	●	●	●
	LE VESINET CENTRE												●	●	●	●
	LE VESINET LE PECQ												●	●	●	●
	ST GERMAIN EN LAYE												●	●	●	●
CRG BRANCH	HOUILLES CARRIERES S				●	●	●	●	●	●	●	●				
	SARTROUVILLE				●	●	●	●	●	●	●	●				
	MAISONS LAFFITTE				●	●	●	●	●	●	●	●				
	ACHERES VILLE								●	●	●	●				
	CONFLANS FIN D'OISE								●	●	●	●				
	NEUVILLE UNIVERSITE								●	●	●	●				
	CERGY PREFECTURE								●	●	●	●				
	CERGY ST CHRISTOPHE								●	●	●	●				
POI	CERGY LE HAUT								●	●	●	●				
	ACHERES GRAND CORMIE				●	●	●	●								
	POISSY															

10.2.4 Simulation Parameters and Convergence

The generalized time is evaluated from the actual travel time of the trip components as described in Chapter 9.4. Each component is subject to a penalty coefficient according to its importance to the passenger's trip. The assignment on the RER A restrained network uses the parameters of the central variant of the CapTA model (CWCV, see Chapter 9.4). Hence, the access and egress trips and the transfer time (without the additional waiting) are multiplied by 2, as happens with the waiting time. The in-vehicle comfort makes reference to the obvious differentiation between two comfort states; travelling standing or sitting. A linear function of

the standing penalty factor $\chi_{za}^r(d_a)$ where d_a is the density of standing passengers for a certain trip section a . The standing penalty coefficient varies from 1,2 (for $d_a = 0$) to 2 (for $d_a = 4p/m^2$), while the multiplier for travelling seated is 1.

An acceptable level of convergence was reached after 30 iterations, with a convergence indicator reduced to 1% of the initial value.

10.3 Passenger Flows on the Network

10.3.1 Passenger Flows on the Line

The passenger flow throughout the restricted network is quite similar with the flows on the RER A of the initial network. Figure 46 illustrates the network passenger flows and the flow-to-capacity ration on the line. The width of the lines represents the passenger flow, while the colours (from light green to red) correspond to the ratio of the total flow on the line to the capacity of the transit services. Additional attention must be given due to the overlapping of the westbound and eastbound arcs.

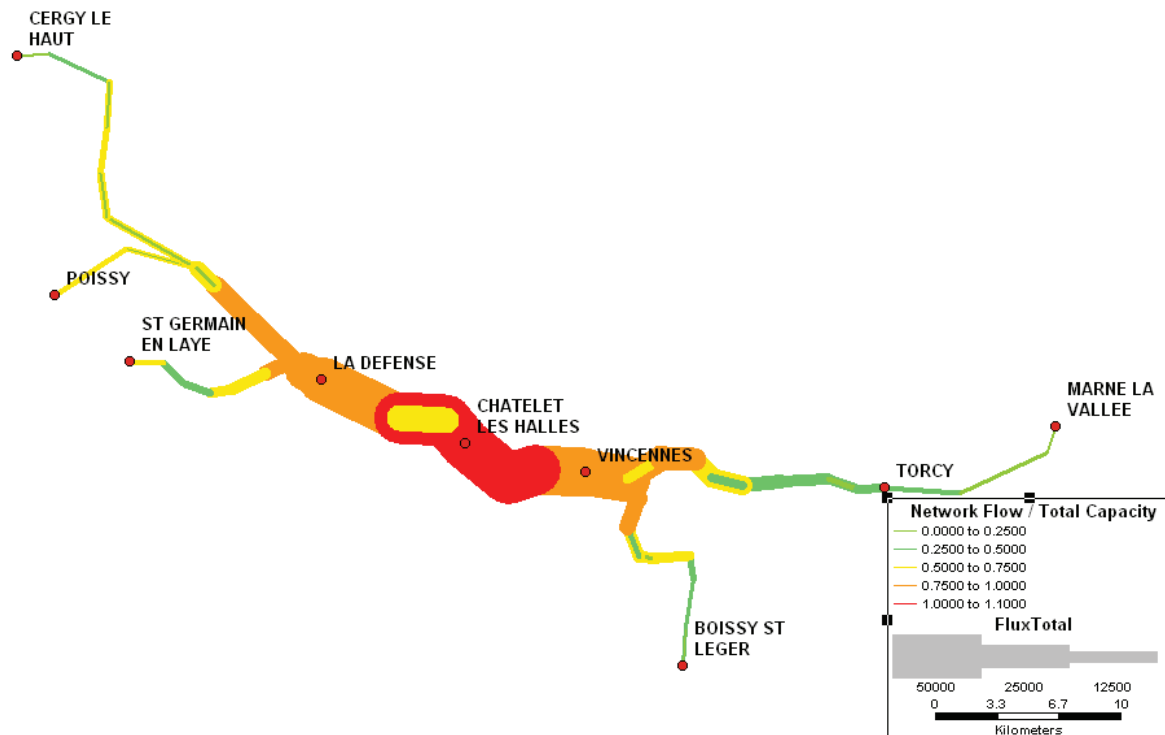


Figure 46: The passenger flow throughout the restrained RER A network

As expected, the passenger flow and the used capacity are low at the extremities of the RER A on both directions, while increasing significantly before the various branches converge to the central trunk. Indeed the flow-to-capacity ratio exceeds 75% before the convergence of the branches to the central trunk. The maximum load corresponds to the westbound interstation between the Chatelet-les-Halles and Auber stations, with 59 319 passengers. The flow exceeds slightly the nominal capacity from Nation to Etoile at the central trunk, to reach a peak on the Chatelet-les-Halles–Auber section with 1,03 of the nominal line capacity.

10.3.2 Boarding and Alighting Passengers Flows

We focus our analysis on the westbound section. In Figure 47, we show the boarding and alighting hourly flows on the busiest line section, from Noisy-le-Grand (MLV branch) and St-Maur Creteil (BOI branch) stations on the east to the Maison Lafitte (POI-CRG branch) and Reuil Malmaison (STG branch) stations on the west.

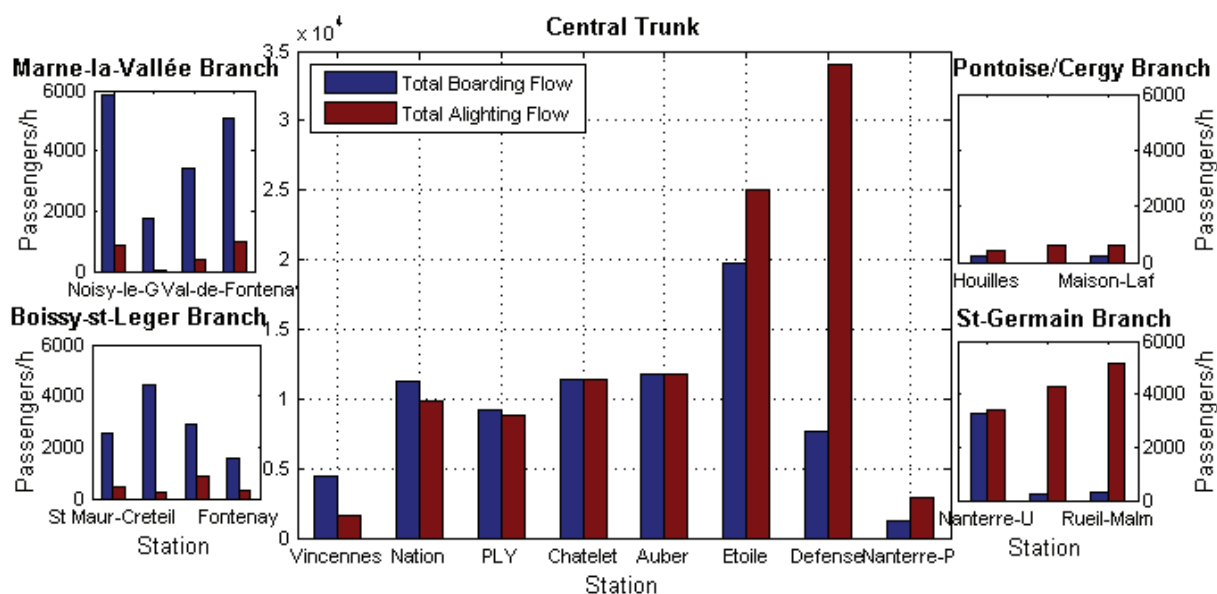


Figure 47: Boarding and Alighting Flows on the busiest sections of the westbound service

The relation between the boarding (blue) and alighting (brown) flows in Figure 47 allows distinguishing three sections with specific flow characteristics:

- Access section: From the east suburbs, until Vincennes, the boarding flows clearly dominate, on an up to seven times ratio, the alighting flow. That segment acts as a feeder from the east suburbs to the centre of Paris and La Defense business district;

- **Transfer section:** At the central segment, from Nation until Etoile stations, the boarding and alighting flows are equal and the boarding is subject to the limited available capacity. The stations of the central trunk of the RER A act as transfer poles for the structural transit network of the Paris Metropolitan Region;
- **Egress section:** The westbound service from La Defense station to the western stations is characterized by an alighting flow exceeding the boarding one. That abides by the line topology, since the on-board passengers naturally alight when reaching the end of the line.

In addition, we inquire on the size of the passenger exchange flow. At the branches the flow is limited and the boarding or alighting flows never exceed 6 000 passenger/hour. However, that is not the case for the central trunk, where the boarding and alighting flows are significant and their combined flow varies from 18 000 – 45 000 passengers/hour. Indeed, as shown at section 10.4, the passenger exchange flow saturates the capacity offered and the increased dwell times of the vehicles of the transit services on the line result in a reduction of the downstream frequency.

10.4 Dwell Times and Service Frequency

The dwell time of a vehicle is related to the passenger exchange flow per vehicle and the number of flow streams available. The boarding and alighting flows are assumed to be evenly distributed on the available flow streams. Thus, in the line model, the dwell time function is linear, depending on the elementary boarding (or alighting) time per passenger (equal to 1,55 seconds/passenger). That makes the dwell time of a vehicle insensible to the congestion on-board and the density of the passengers present on the platform.

The calculated dwell time is composed of the minimum time for the passenger exchange and the operation time – the period without passenger movement, necessary for the operation of the doors and the safe departure of the vehicle. Hence, the actual dwell time is the maximum of the calculated and the scheduled dwell time. That implies the existence of a critical passenger exchange flow. It is defined as the passenger flow for which the calculated dwell

time is equal to the scheduled one, amounting to a little less than one half of the total capacity for the rolling stock circulating on the RER A.

Table 15: The characteristics of the rolling stock circulating on the RER A

	Seat capacity (per vehicle)	Total person capacity	Nb doors	Nb of flow streams	Critical passenger flow (w reduced flow streams)
MI84	432	1760	32	64	832
MS61	600	1888	36	72	929
MI2N	1056	2580	30	90	1161

The suboptimal use of the flow streams is addressed by reducing their number by 1/3. Therefore, 43, 46 and 60 simultaneously available flow streams per vehicle side are considered for the MI84, MS61 and MI2N vehicles respectively.

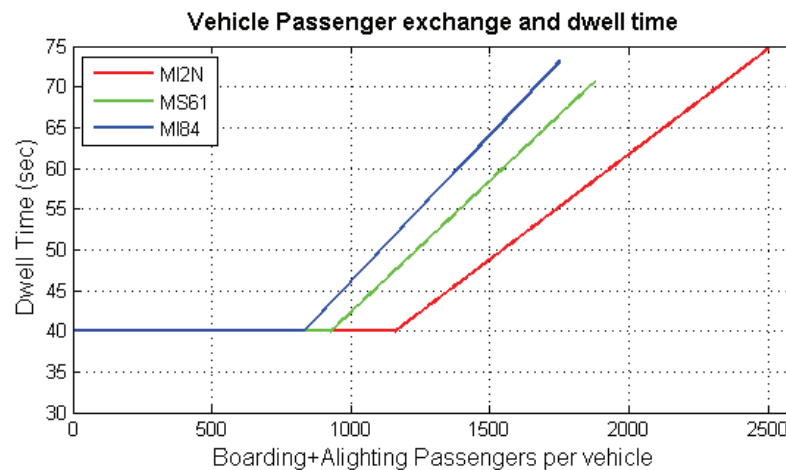


Figure 48: Dwell Time in relation to the exchange volume on the rolling stock used in the RER A

In the case of the westbound RER A services, the total passenger exchange flow at the Etoile station amounts to 45 000 passengers, leading to a frequency restraining downstream. Depending on their characteristics, the dwell time of each service varies from 49,8 to 61,4 seconds/vehicle, higher than the 40 seconds/vehicle of the schedule, as illustrated in Figure 49. The differences in dwell time stem from the variability of the boarding and alighting flows among services and the differences of the rolling stock assigned to each service. The total exchange flow of the westbound service in Etoile is 44 804 passengers/hour with 44% alighting and the remainder boarding.

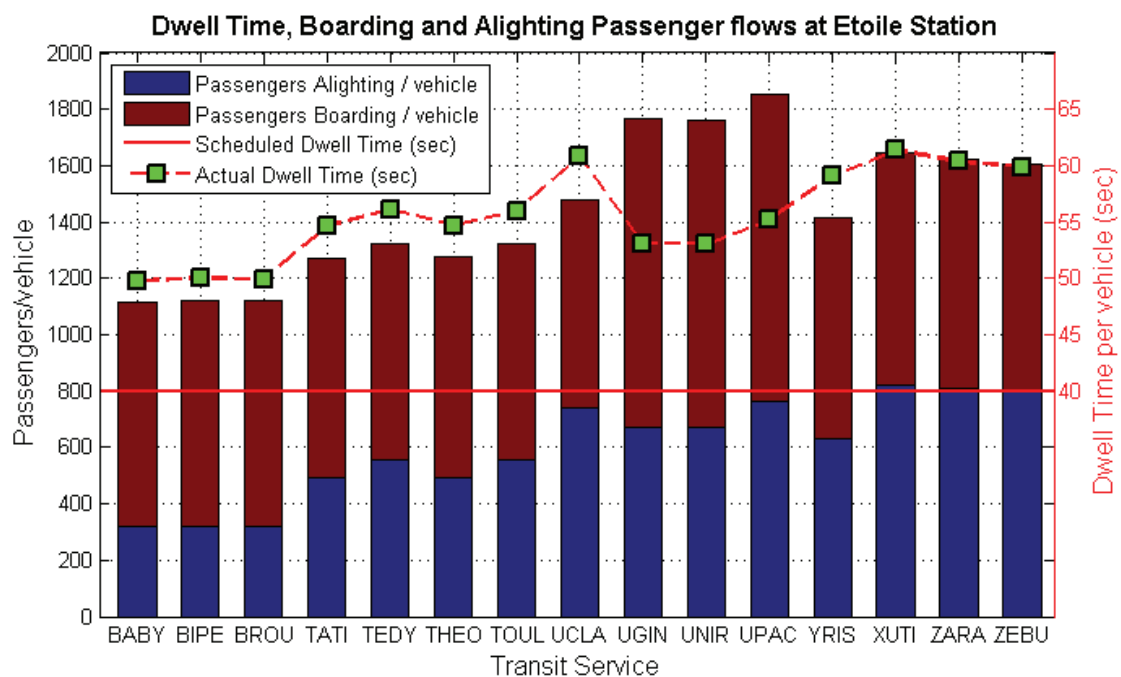


Figure 49: Boarding, alighting passenger flows and dwell time per vehicle of the westbound services at Etoile station

As shown in Figure 50, the combined frequency downstream is reduced in Etoile station by 13,6% from 30 to 25,91 veh/h, as does the capacity. Although at La Defense station the dwell time varies from 40 to 65 seconds, the frequency is not restrained, since the platform occupation time doesn't exceed the reference period.

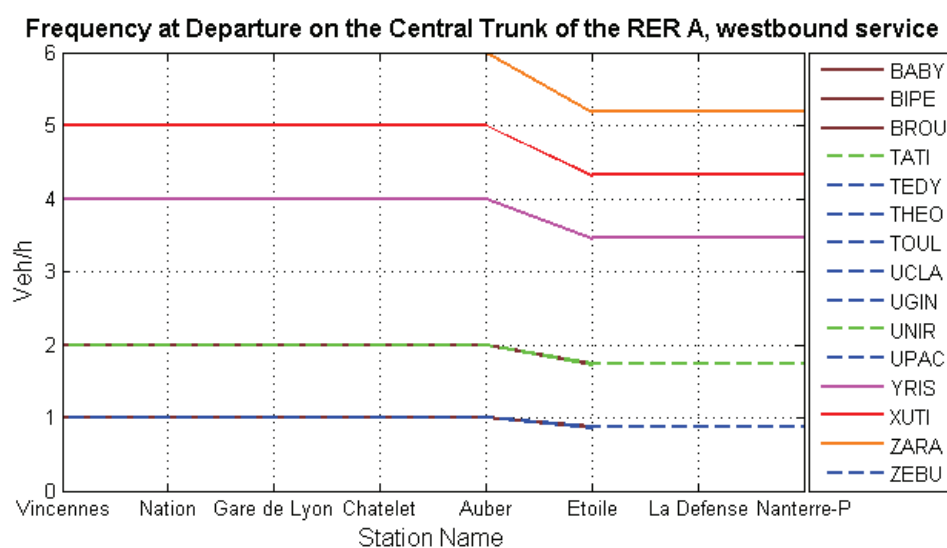


Figure 50: Frequency at departure for the westbound transit services at the Central Trunk

10.5 Residual Capacity and Boarding Flows on a Vehicle

The CapTA model is characterised by strict capacity constraints at the line level (and their relaxation at the network level). Besides the competition among services, the boarding flow depends on the in-vehicle capacity available. Figure 51 illustrates the boarding and alighting process for the TEDY service (Marne-la-Vallée to Poissy).

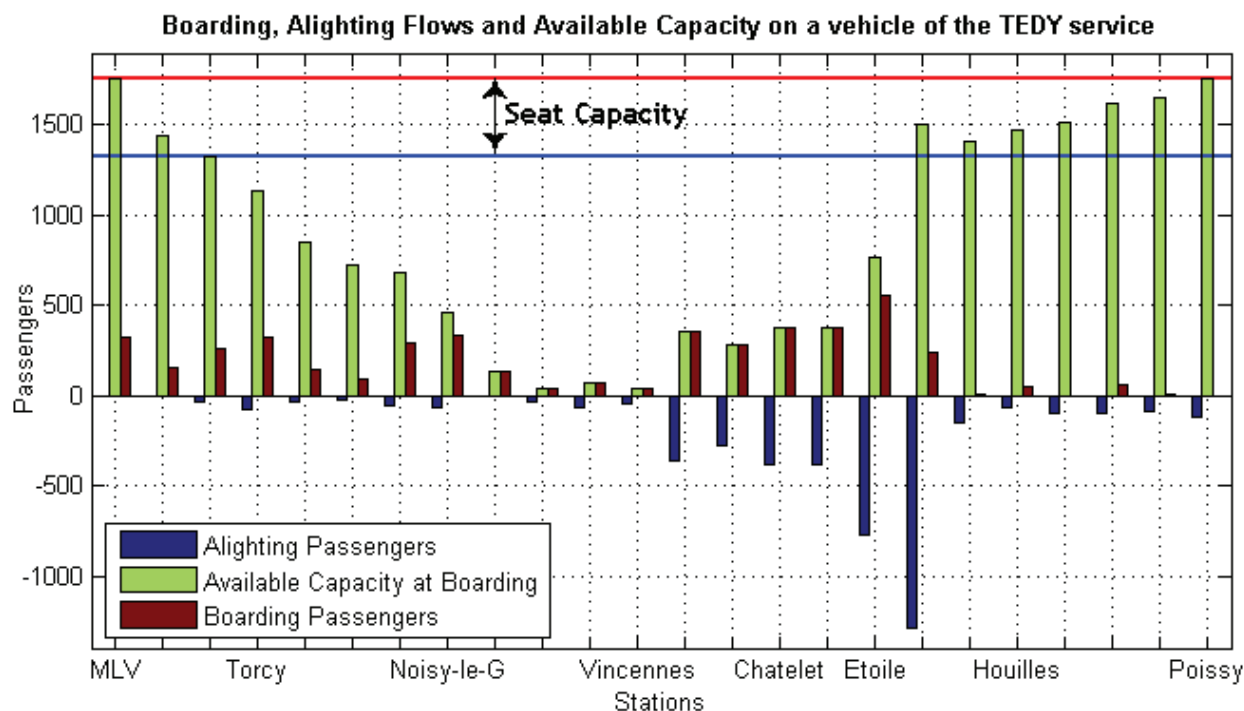


Figure 51: The boarding and alighting process on a vehicle of the TEDY service

We apprehend the relation between the available capacity and the boarding flow. At the beginning of the service, the available capacity is sufficient for the boarding flow. However, at the busiest section any available capacity comes from the alighting passengers who liberate space for the passengers waiting on the platform. At these stations the alighting flow, the available capacity and the boarding flow are equal. Therefore, at the line level, when the available capacity is not sufficient, only a share of the waiting passengers can board. That induces an increase in the average waiting time and contributes to the formation of a passenger stock.

Furthermore, the area between the red and blue lines corresponds to the vehicle's seat capacity. The available capacity being inside that area means that some boarding passengers

will find directly a seat. In the case of the TEDY service (as well as the other services on the MLV branch), the seats are occupied very early on the trip and no seats are available at boarding between Bussy-St-Georges (see Table 14 for stopping policy) and Etoile. To summarize, in 14 out of 24 stations served by TEDY the available seats are only those liberated by the alighting passengers and occupied by the on-board passengers, who have priority. Hence, the boarding passengers are sure to make a part of the trip standing. The probability to find a seat downstream depends on the intensity of the on-board standing and alighting sitting passengers.

10.6 Passenger Stock and Probabilities of Immediate Boarding

10.6.1 *The Passenger Stock on the Central Trunk of the RER A*

The transit bottleneck model is used to apprehend the effect of total capacity by explicitly describing a passenger stock. Indeed, from a given access station i , a passenger stock, σ_{is} , per egress station s denotes the number of passengers wishing to board any of the transit services directly connecting that couple of stations. It depends on the exogenous flow x_{is} by destination station and the available boarding capacity k_{zi} by vehicle. The passenger stock of a transit service n_{zi} is the sum of the partial stocks of all the stations called by that service z . Hence, at a given station two services with identical stopping policies downstream will have the same passenger stock, whatever their load. The vehicle inflows depend on the probability of immediate boarding π_{zi} .

Figure 52 illustrates the passenger stock at boarding of the westbound transit services on the central trunk of the RER A. The services are grouped by destination according to their colour. Two observations are made in relation with the behaviour of the model:

- At each station, we compare the passenger stocks of each service. On the first section, from Vincennes to Auber, the boarding passengers are destined to the stations of the central trunk (with an important part alighting at La Defense), where the stations are served by all the services. Thus, the passenger stock is similar. However, at Etoile and

La Defense stations, before the divergence to the branches, the passenger stock of the transit services varies according to their stopping policies downstream and the partial passenger stock per destination. For example, the passenger stock of XUTI (red), ZARA (orange) and ZEBU (brown) differs according to their stopping policy, even though all three serve the St-Germain-en-Laye branch (southwest).

- Along the line, the value of the partial stock depends on the level of saturation of the transit services. When the combined capacity of the services is sufficient but at least one service is at capacity, the passenger stock slightly increases compared to $x_{is} / \sum_{z \in (i,s)} \phi_{zis}$. That is attributed to the failure of some passenger to board the first vehicle and is extended to the average waiting time for the station couple. For example, at Nation, the waiting time for a trip to La Defense is equal to 3,06 minutes (compared to a reference time of 2 minutes). Furthermore, when the combined capacity is insufficient, the passenger flow cannot be evacuated during the simulation period and the passenger stock increased significantly. In Chatelet, the stock reaches a value of 2 300 and the waiting time to La Defense is 12,09 minutes – a six fold increase from the reference situation.

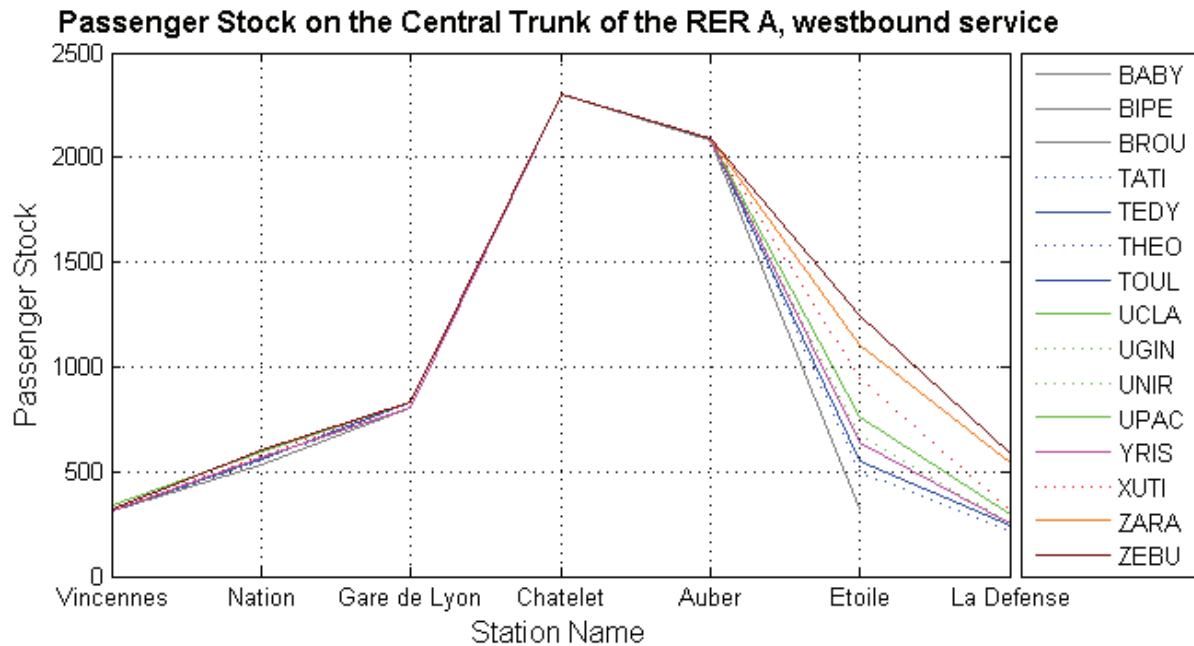


Figure 52: The Passenger Stock at the central section

10.6.2 The Partial Passenger Stock to La Defense Station

We isolate the westbound trips ending at La Defense and focus on the relation between the exogenous flow x_{is} and the discharge rate q_{is}^- . When no congestion occurs, these values should be equal. That is true only approximately, since some simplifying assumptions lead to a slight underestimation of the discharge rate, even though insignificant. Indeed, it varies around 1-5% of the exogenous flow according to the combined frequency of the services. On the other hand, when the combined capacity is not sufficient, as is the case in Chatelet and Auber, the difference between the exogenous flow and the discharge rate is significant, as shown on Figure 53.

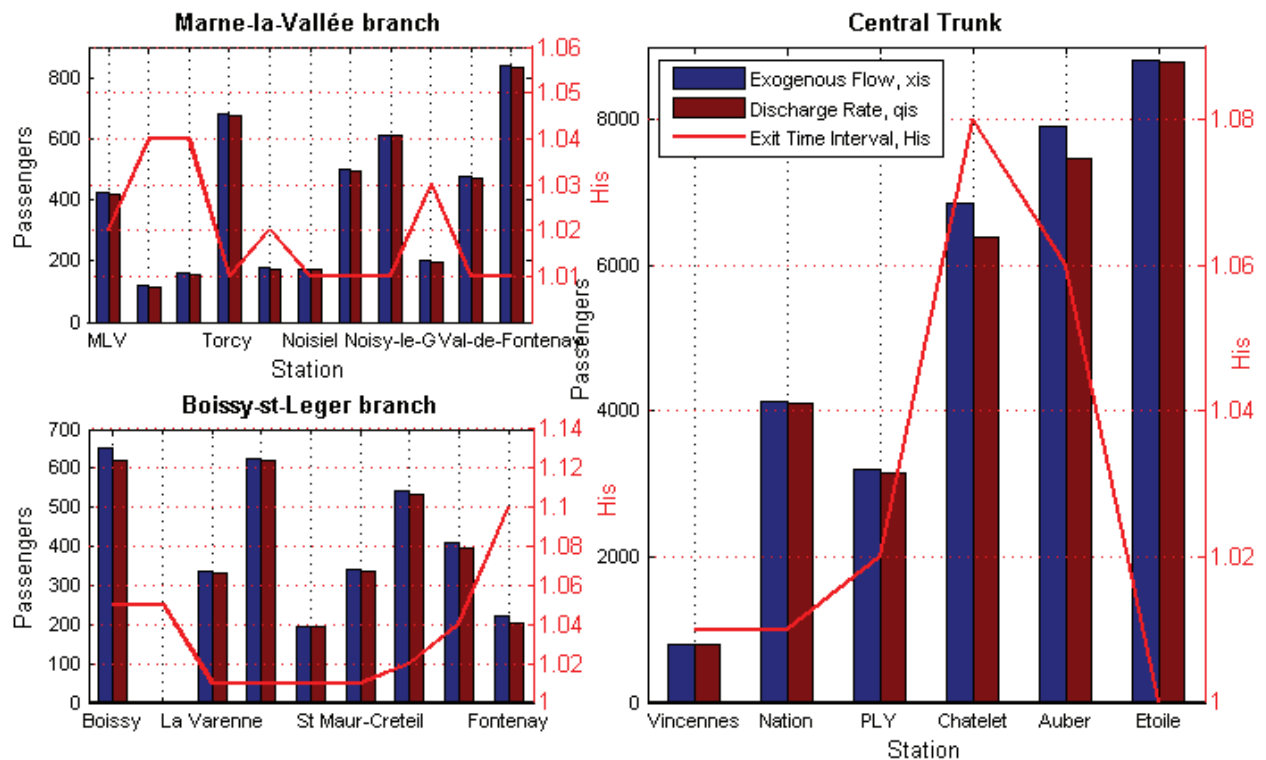


Figure 53: The exogenous flow versus the discharge rate and the exit time interval alighting at La Defense

The difference between exogenous flow and discharge rate can be expressed by the exit time interval, H_{is} ; the time needed to evacuate the passenger flow waiting to board during the simulation period, expressed by the ratio x_{is} over q_{is}^- (Chapter 4.4.1). In the uncongested conditions it is slightly higher than 1, due to the non-equality between x_{is} and q_{is}^- , explained previously. Nonetheless, if congestion occurs, the exit time interval reflects the pressure

exercised by the passenger stock at boarding. In fact, we observe that in Figure 53 the highest value of H_{is} is not found at the central trunk, but rather at Fontenay station ($H_{is} = 1,10$ at Boissy-St-Leger branch) with a low exogenous flow. The limited frequencies and the low available capacity offered at Fontenay station make the evacuation of the passengers present take longer.

10.6.3 Boarding Flows and Route Proportions

The flow boarding a certain vehicle of a transit service depends on the vehicle's residual capacity after alighting and the passenger stock of that service. In addition, the probability of immediate boarding ($\pi_{iz} = \max\{1; k_{zi}/n_{zi}\}$) acts as the confrontation of the history of the service – which defines the vehicle load at arrival and the available capacity k_{zi} – and the downstream stopping policy – which designates the passenger stock, n_{zi} .

Table 16: Boarding Flows for the unbounded and bounded model variants

Transit Service	Frequency (veh/h)	Route Proportions for Bounded	Route Proportions for Unbounded	% difference
BABY	1	3,20%	3,08%	4%
BIPE	1	2,93%	3,08%	-5%
BROU	2	6,68%	6,16%	8%
TATI	2	5,86%	6,37%	-8%
TEDY	1	3,16%	3,24%	-2%
THEO	1	3,31%	3,18%	4%
TOUL	1	3,13%	3,24%	-3%
UCLA	1	2,93%	3,56%	-18%
UGIN	1	5,31%	3,54%	50%
UNIR	2	10,61%	7,09%	50%
UPAC	1	4,14%	3,56%	16%
YRIS	4	16,73%	13,14%	27%
XUTI	5	12,65%	16,59%	-24%
ZARA	6	16,88%	20,71%	-18%
ZEBU	1	2,47%	3,45%	-28%
TOTAL	30	100%	100%	

Although the transit services with identical stopping policies downstream have the same passenger stock n_{zi} , their boarding flow, and in extent their routing proportions, depends also on that available capacity, k_{zi} . By comparing the route proportions of the bounded and the unbounded variants at Nation station, we observe differences, which range from -28% to

+50%, according to the services' characteristics. Indeed, the UGIN and UNIR services increase their share on the boarding flow from 3,54% and 7,09% to 5,31% 10,61%, respectively, since they have sufficient capacity to accommodate the flow failing to board other services (see Figure 54) and a probability of immediate boarding equal to one.

Having identified the partial stock of the Nation – La Defense station couple, we can disaggregate the boarding flows of the services per destination, illustrated in Figure 54. The boarding flow destined to La Defense (in blue) and the miscellaneous (in brown) compose the total boarding flow, opposed to the service's available capacity (in green). The red squares correspond to the stock's probability of immediate boarding (right axis). Due to the vehicle's intrinsic characteristics and the destination competition, we observe that the boarding flow to La Defense is not evenly distributed along the services. It varies according to probability of immediate boarding and the frequency of the services.

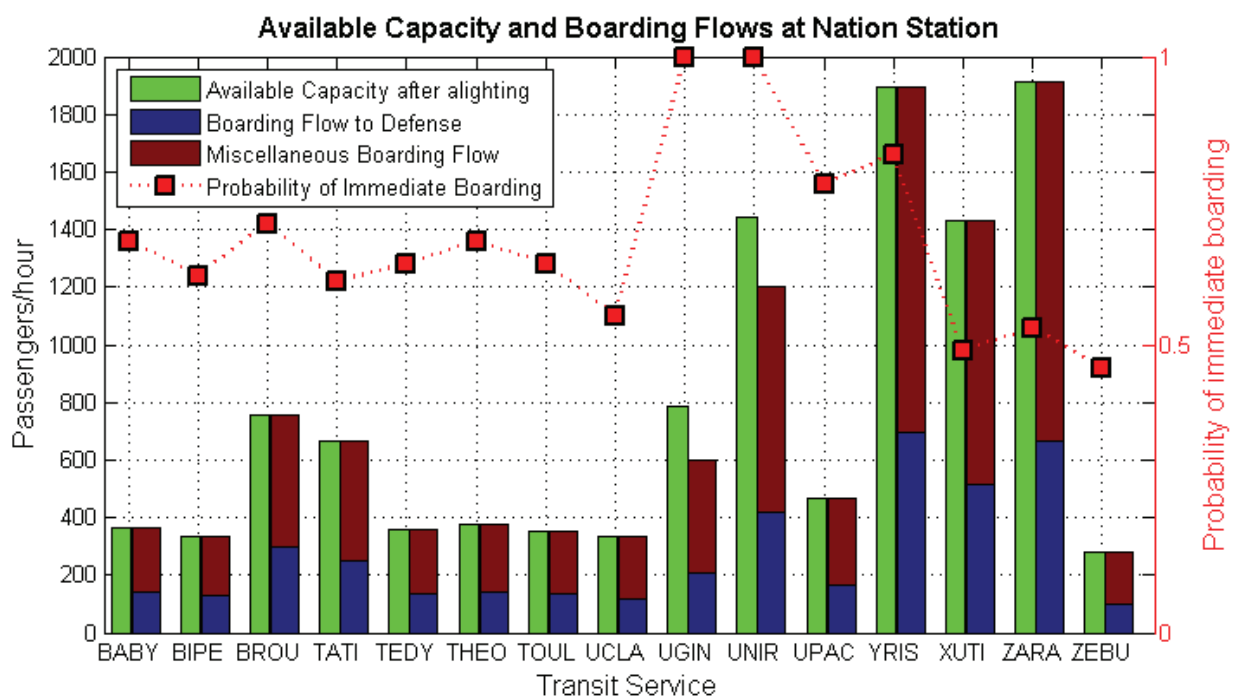


Figure 54: Route Proportions at boarding and Probability of Immediate Boarding at Nation

10.7 Average Generalized Cost

The average generalized cost for both directions on the RER A line includes the in-vehicle travel time (depending on the comfort level) and the platform waiting. We compare the total

travel time and its components of the unbounded variant with those of the bounded one – including the three types of capacity constraints simulated in the CapTA model. The average generalized travel time of a trip is 34,05 minutes for the bounded variant, an increase by 41% from the unbounded benchmark (24,16 minutes). That increase can be attributed for 1/3 to the increase of the perceived waiting time and for 2/3 to the increase on the perceived in-vehicle travel time due to the less comfortable states suffered by the passengers on-board. In addition, the difference of the actual travel time between these models is attributed to the waiting time, since in the restrained RER A network, there are no alternative routes.

The line model takes into account the local capacity constraints for the evaluation of the cost of the line legs along a line. Figure 55 illustrates the actual (continuous line) and the perceived (dotted line) travel time of the westbound trips ending at la Defense from the station on the Marne-la-Vallée branch, the Boissy-St-Leger branch and the central trunk. The in-vehicle travel time (IVTT) is green and the waiting time (WT) is red. The total travel time (TT) is the sum of the previous components and is depicted in blue.

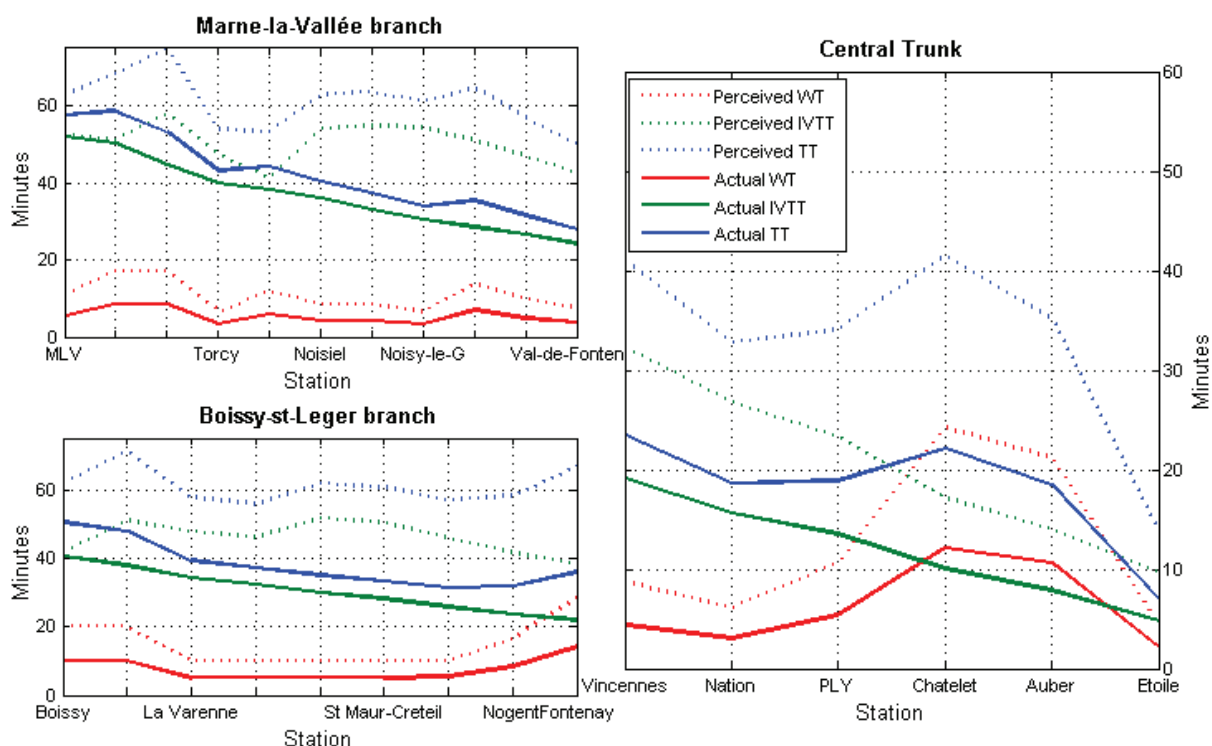


Figure 55: The average generalized time of the trips ending at La Defense station

We observe that naturally the actual IVTT decreases for the trips beginning closer to La Defense, similarly to the distance between the boarding and alighting stations. Furthermore, the gap between actual and perceived IVTT increases as the on-board volume increases, expressing the difficulty to find a seat and the density of standing passengers on-board. Due to the chosen parameters, that gap amount to 40-70% of the actual IVTT. We make three relevant observations concerning the travel time along a radial line:

- Although the length of the trips changes significantly, we observe that the perceived TT on the branches oscillates around 60 minutes. In other words, there are not any notable differences between trips with different access station heading to La Defense. The low frequencies and the length of the trip from the furthest stations are compensated by the easy access to the vehicle's seats. The trips that start closer to Paris are shorter but face degraded travel conditions;
- A comparison of a trip beginning at Noisiel and another from an upstream station (Lognes, between Torcy and Noisiel) reveals that the second is more attractive, with a lower perceived TT. That is attributed to the availability of vehicle's seats, since a passenger downstream will travel standing;
- Equivalent to the previous case, the actual TT from Chatelet (TT of 21 minutes) is greater than the one from upstream stations, such as Nation (18 minutes). The main reason is the increased waiting due to the presence of an important passenger stock.

10.8 The Impact of a New Rolling Stock

As a means to deal with the saturation of the RER A, the transport authority (STIF) and the transit operator (RATP) of the Paris metropolitan area decided to invest 2 billion euros for the complete replacement of the one-level trains on the RER A, MS61 and MI84 by 130 duplex trainsets, the MI09 (Le Monde, 2012). The new train is very similar to the actual duplex, MI2N, and offers 948 seats for a total capacity of 2 600 passengers/vehicle (RATP, 2011). Therefore, the line will see a significant increase on its nominal capacity by 30% to be able to carry about 77 000 passengers per hour and direction. The model makes it possible to assess the effect of additional capacity to the level of service, without any service modification.

Applying the CapTA model at the line level shows the effects of the replacement of all one-level trainsets with the new duplex MI09. The capacity offered is sufficient to carry all the transport demand during the morning peak hour. However, that does not apply for the passenger capacity exchange since the boarding and alighting flows in Etoile lead to a reduction of the combined frequency by 3% (as opposed to -13.6% of the benchmark). The average travel time per trip on both directions is reduced by 17% to 28,2 minutes. Globally, the reduction of the total travel time for all the passengers during the morning peak hour roughly amounts to 24 150 hours. By multiplying the peak hour gain by 5 to deduce the entire day and by 250 for a year, we estimate a conservative annual gain of 30,2 million hours per year. With a value of time at 10€/hour, the replacement of the one-level trains will bring a 302 million € socio-economic benefit.

Although that estimation unveils the evident socio-economic benefit of the project, these indications should be taken with extreme caution since some mutually excluding effects are not considered. First the current assignment does not take into account the network effect; the new travel conditions on the RER A may attract new passenger flows, inducing additional discomfort on the line, but easing the congestion on other lines of the network. In addition, these values correspond to the 2008 situation and not to a more saturated transport demand in medium term, if no investment is made. Finally, the operational benefits are not taken into consideration, although the new trainsets boast a reduction by 31-55% of the energy consumption per passenger from the rolling stock it will replace.

10.9 Conclusion

This chapter offered a simplified application instance of the CapTA model, restraining the simulation to the RER A line, the busiest commuter rail line in the Greater Paris transit network, with a patronage exceeding one million passengers daily. It completes the simulation on the entire transit network and focuses on particular elements of the line model.

Analysing the westbound service of the RER A, we produce a series of remarks related to the capacity effects. We observe that although the Chatelet – Auber is the busiest section, with 59 300 passengers/hour on board, it is at Etoile station where the passengers exchange seems insufficient. Indeed, the vehicles' dwell time greatly exceed the scheduled 40 seconds,

resulting in a restrained frequency and a loss of capacity downstream by 13,6%. The seats are occupied early on a vehicle's trip and some saturation appears before the convergence of the branches to the central trunk. As a result of the heterogeneous stopping policies and rolling stock of the transit services, the flow is not uniformly distributed among the services, some being saturated while other offering available capacity, which in turn affects the waiting time of the passengers on the platform. Taking La Defense business district as a destination, the previous effects lead to a generalized trip time which oscillates around 60 minutes at the branches – whatever the access station is – and to an actual travel time which may even increase as we approach the destination, if we compare a trip started from Nation (18 minutes) to one from Chatelet (21 minutes).

By integrating these capacity effects to the CapTA model, we may proceed to conservative estimations of a project, which would offer greater capacity, without changing the service timetable. Indeed, in the case of the replacement of the actual old one-level trainsets on the RER A by more spacious duplex ones, the model brings sufficient evidence to back up a cost benefit analysis and implies a yearly gain of 30,2 million hours.

Some modelling issues appear for further research. In the current assignment model, we adopted a comfort multiplier, which ranges from 1,2 to 2, according to the average density of the standing passengers. We agreed to the hypothesis of a gradual increase of the comfort penalty due to the on-board conditions. However we do not differentiate between the waiting time linked with the random arrival of the next vehicle and the quasi-deterministic waiting in a saturated platform. Furthermore, when the capacity available is insufficient at boarding, the passenger stock willing to board explodes along with the expected waiting time while the probability of immediate boarding plummets. That situation could designate a sort of pressure exercised at the boarding stock to board the vehicle. In other words, that could lead into an overcrowding acceptance, where the boarding flow exceeds the available capacity at boarding and the vehicle load exceeds its nominal capacity (considering 4 standing persons per available m^2). An increased passenger load will worsen the level of comfort perceived by the passengers already on-board.

Chapter 11:

General Conclusion

11.1 Introduction

The research project associated with this dissertation makes an effort to answer to the issues raised by the scientific community of transit assignment modelling and thus provide some novel elements to this research sector. Namely, we sought to develop new approaches for modelling the capacity effects of various elements of the transit system and formalize a coherent framework that can jointly address them. These efforts resulted in the elaboration of a transit assignment model, CapTA, for Capacitated Transit Assignment, able to simulate large-scale networks.

The general conclusion is structured in two sections. First, we address the main results of this work and some conclusions for the CapTA model (section 11.2). Second, the main elements that emerged during this work are discussed (section 11.3) in connection with future research perspectives that can result in the elaboration of more realistic transit assignment models.

11.2 Main Results and Conclusions

A set of selected issues has been examined in this work. In addition to these results, some general conclusions can be drawn. These conclusions are outlined here.

11.2.1 Representation of the Transit Network

A fundamental component of the CapTA model is the bi-layer representation of the transit network. That stems from the motivation to provide a network representation adapted to the

needs of each modelling stage. Two superposed network layers are defined: the upper and lower network layers, with respect to the demand and supply side of the modelling process (Chapter 6).

The upper layer focuses on a thorough description of the passenger behaviour on a transit network. The network elements of the upper layer network, which represent the trips of an individual on each sub-system, are called network legs. Three types of network legs have been identified with respect to the nature of the subsystems: the line leg, the station leg and the access-egress leg. Their cost is defined from the average travel conditions and quality of service perceived by an individual.

On the lower layer, the service representation is more detailed and emphasizes on the effects of the capacity constraints. Each sub-system is represented by an isolated sub-graph. Particular local models evaluate the effect of the capacity constraints on the passenger flow and the local route choice, the quality of service and the operational characteristics. That implies that the lower layer representation of each system is specific to that system and the local modelling approach.

The use of that bi-layer representation has an impact on the transit assignment model. The use of an additional network layer (the upper leg-based network) increases the size of the service network. In the case of the Greater Paris transit network (Chapter 9.3) the set of line legs is 57% larger than that of the line service arcs it replaces. However, that corresponds to an increase of only 3,8% on the link elements of the entire upper layer network. Furthermore, that distinction requires the elaboration of specific procedures for the modelling of the top-down and bottom-up relationships of the upper and lower network layers.

Nevertheless, the main advantage of the bi-layer representation is that each step of a traffic assignment model can be addressed with the appropriate network. Indeed, that distinction allows the elaboration of a modular modelling framework where the local models on the lower layer are not subject to the constraints of the architecture of the upper layer network assignment. Therefore, local models can be added or modified and the lower layer representation can gain in detail where it is required; all that without any significant impact on the modelling process. A corollary effect of this bi-layer representation is the distinction between the flow loading (and evaluation of physical conditions) and the cost evaluation

processes. That allows the evaluation of distinct travel costs for each class of passengers, on the basis of his preferences.

11.2.2 Modular Framework and Specific Models

The use of a bi-layer network allows the development of a modular framework. That is elaborated around three hierarchical modelling levels, described in Chapter 6. The set of models contains a mixture of novel models that bring significant innovations and existing models with the necessary adaptations and enrichment. These models were the object of a consistent algorithmic development with objective to simulate large-scale networks; in the case of this study, the Greater Paris transit system.

The CapTA model contains local models that address the issues at the stop level and a network model for the management of the assignment process. However, the specific bi-layer representation requires an intermediate level of system sub-models. These system sub-models act as a coordinator of the local effects and assure the transition between the upper and lower layer. Namely, the line and station models developed here offer the flexibility to the modeller to address all relevant capacity constraints at the station and the line level, such as the in-vehicle quality of service, the platform waiting and boarding process and the vehicle temporal occupancy of the platform.

The in-vehicle quality of service is dealt with the in-vehicle comfort model. That is an adaptation of an existing model described in Leurent (2012), which distinguishes the sitting and standing comfort states. The initial model carries two adaptations. First, the standing penalty becomes sensible not only to the travel time but also the density of the standing passengers. A similar penalty can be introduced for the sitting passengers. Second, a modification is suggested on the algorithm (see Annex B) in order to address the use of folding seats within the vehicle. These adaptations increase the realism of the in-vehicle comfort model.

The platform waiting and boarding process required the elaboration of the transit bottleneck model. Its main characteristic is the identification of a passenger stock per destination. That allowed modelling the competition of a passenger stock for a given destination at boarding with passengers stocks of other destinations and for all the services available at the platform. The main advantage of that approach is the joint assessment of the route share to all available

routes and of the combined waiting time separately for each access-egress couple on the line. In addition, that leads to the specification of three traffic states (chapter 4.4.2), by order of increasing passenger flow:

- The unsaturated state, where the route share and waiting time are consistent with a unbounded assignment;
- The semi-congested traffic state, where insufficient capacity at one or more services results increased route shares for the remaining services, in addition to a small increase of the combined waiting time, and
- The congested state, where all transit services for a given destination are congested. There, the passengers face a significant increase in the waiting time on that access – egress couple.

Finally the in-vehicle and on-platform passenger flows influence the service operations along a transit line, captured by the restrained frequency model. There an increase of the dwelling time of the vehicles may restrain the service frequency downstream on the basis of a frequency-based modelling approach. That frequency reduction affects directly the waiting time. But its indirect effects in the waiting time and the quality of service are more significant, since it leads to a reduction of the downstream available capacity. For example the frequency reduction on the westbound service of the RER A (chapter 9.6) results in a reduction of the hourly capacity, equivalent to 2,1 high-capacity trainsets.

In addition to the novel line and station models, the use of a leg-based upper layer network representation permits to tamper with the unbounded network traffic assignment. The main adaptation concerns the replacement of the route frequencies with the fictive frequencies on the line legs. That has a two-fold effect on the assignment process:

- First, it disconnects the frequency of the line legs with that of the services forming it and allows the bundling of the available options at the station level, with respect to the presence of real-time information, and
- Second, it eases the distinction between an unsaturated discontinuous line leg and a congested one. In the latter case, the service becomes continuous and that leg is

reduced to a pedestrian option with a fixed waiting time due to the mingling of the passengers at the platform.

Some important adaptations were made on the mathematical formulation of the upper layer network model. There, the transition between discontinuous and continuous service is achieved by a discontinuity attenuator, which reduces the fictive frequency once congestions exists. A traffic equilibrium is defined as a twofold vector of arc flows and strategy share by node. That is characterized as a variational inequality problem. Although the equilibration algorithm is likely to be computationally costly, it can provide a rigorous convergence criterion.

11.2.3 Congestion Effects and Evaluation

A model such as CapTA is an ideal tool for examining the effects of the capacity constraints on the transit network. The transit assignment on the large-scale network produces two straightforward observations. In Chapter 9 we compared different variants of the CapTA model, with respect to the simulation parameters and the capacity constraints included. We observed that among the variants considered, those that included the in-vehicle comfort (CWCF and CWCV) lead to a faster and smoother convergence to the static state, than the variant where only total capacity (CNC) is considered. That is attributed to the nature of the in-vehicle comfort model, since the most loaded lines are much more penalized and that from a lower flow-to-capacity ratio. That facilitates the spread of high passenger flows to alternative itineraries, as long as the network has sufficient capacity.

Similar observation is made on the distribution of the arcs with respect to their flow-to-capacity ratio (Chapter 9.5). The traffic assignment on the upper layer takes place with a relaxation of the capacity constraints. Nevertheless, the use of capacity constraints achieves in reducing the number of arcs that exceed their capacity. The most efficient model variant is the CWCV bounded one, where very few arcs exceed the line capacity and that by no more than 3,7% of the nominal capacity.

A transit assignment model with capacity constraints achieves in simulating certain aspects of the passenger behaviour. These are revealed when comparing the bounded model with the results of an unbounded transit assignment. In the case of the Paris Metropolitan Region, we focus on the results of the CNC variant where only the transit bottleneck and restrained

frequency models are considered. We observe that the effects of congestion are not concentrated in the waiting time, but rather the increase of the travel time, compared to the unbounded model is distributed to the waiting time and the in-vehicle travel time. Imposing a capacity constraint on boarding does not affect only waiting. Especially in a network with strong connectivity where alternative itineraries are available, the expected trade-off of the passengers between waiting time, in-vehicle travel time and the other components of a trip is revealed. Whatsoever, the in-vehicle comfort stands for the 3/4 of the increase in generalized travel time, compared to the unbounded conditions. If we limit our analysis on selected lines, we observe that an increase of the average sitting time per trip is combined with a simultaneous decrease in the average standing time per trip. That indicates that a passenger prefers to make a longer trip seated, and avoids standing and frequent transferring. However, the latter depends on the saturation of the transit network and the possibility that the passenger improves his travel conditions.

A significant result of the analysis on the restrained network (Chapter 10) has to do with the evolution of a trip travel time along a transit line. We focus on the average generalized travel time of each station of the westbound service of the RER A to La Defense business district (Chapter 10.7). We found out that the average generalized time of a station upstream can be lower to that of a station closer to the destination. That is explained by the opportunity of a passenger at each station to travel seated. Once the seats on a radial network are occupied, the next passenger boarding travels in more onerous conditions standing; and in the case of the RER A at high densities. The same occurs at the central trunk of the RER A, where the intensity of the boarding flows at one station makes it less attractive than the stations upstream (even if compared to the actual travel time to the destination).

A significant element of CapTA – compared to other models – is its ability to evaluate the effect of a change in the rolling stock, even without a change on the planned service. In that case, the travel time gained comes from the improvement of the service operations and of the in-vehicle comfort. For example, in the case of the RER A, the replacement of current one-level trainsets with high-capacity duplex ones, increases the peak-hour capacity of the line by 30% and facilitates the passenger exchange at dwelling. These result to the service frequency being less restrained and to an improvement of the in-vehicle comfort. They are evaluated as a actual gain of 24 150 passenger hours for the morning peak hour only, under certain flow

assumptions. Taking into account these effects has certainly a positive impact in the socio-economic analysis of projects that seek to improve the productivity and the offered capacity of existing infrastructure.

11.3 Discussion and Perspectives

A number of issues that occurred during the development of the CapTA model are discussed. These concern the mathematic characterization of the equilibrium and the validation process. Finally, some future perspectives for the elaboration of more realistic models are suggested.

11.3.1 Issues Relevant to the Mathematic Characterization of the Equilibrium

The mathematic characterization of the existence and uniqueness of the traffic equilibrium is the centrepiece of each network traffic assignment model. In fact, a continuity of the arc cost with respect to the traffic flows is required on the feasible set of the solution. That is generally maintained, except at certain points. Namely, a discontinuity arises when on a node i , the cost $\tau_{i(s)}^d$ to reach a destination s by the strategy d coincides with the minimum cost of an alternative option. That discontinuity is curbed if the flow is assigned to two simultaneous strategies. In order to curb the non-uniqueness of that problem, various mathematical approaches have been suggested.

The previous case showed that it is possible that on certain conditions the uniqueness of an equilibrium solution is not assured. That is also the case when locally a discontinuous service (with frequency and waiting time) is compared with a pedestrian choice with continuous availability. At the critical point the pedestrian link yields identical and minimal cost to destination: then either one would yield an optimal path. To curb that non-uniqueness issue, an approximation is generally used. The private links are associated with a very high frequency, and are transformed into discontinuous services. Despite of the convenience of this adaptation, it seems essential to include a more coherent distinction between continuous and discontinuous service, carried over to the traffic assignment algorithms.

11.3.2 Confront the Models to Observations

The development of any models requires a validation step. There we confront the model results – in the case of a transit assignment, the passenger flows – to the field observations in order to verify that the model can reproduce correctly the observed flows. That is essential when building specific regional models. Since a model is by definition a simplified representation of the reality, errors can be present at various stages: at the zone representation and the quality and quantity of connectors; at the description of the service network (particularly for an extensive and complicated network) and at the travel demand model used for the elaboration of the modal origin-destination matrix.

In addition to the inaccuracy of the modelling data, the model and its very basic assumptions need to be validated against observed data sets. The calibration process concerns the parameters used throughout the model, such as the in-vehicle comfort and waiting time penalties. However, a main difficulty consists in the verification of the assumptions relevant to the passenger behaviour. Basic assumptions, such as the complete rationality of the economic agents and the use of hyperpaths have to be validated, or at least the conditions for which the validity is guaranteed have to be identified.

The validation process of the regional models requires extensive origin destination surveys and passenger flow counts, which are quite onerous. The validation of the modelling assumptions is usually the object of particular stated preference surveys. However, the development of intelligent transport services and the use of intelligent devices is a major source of data. The use of the so-called “Big Data”, from numerous intelligent devices, such as cellphones, e-ticketing, GPS devices – that produce a large quantity of information – seems necessary for the validation of the basic behavioural assumptions of the assignment models.

11.3.3 Improve and Enrich the Transit Assignment Models

Some issues linked to the development of the flow assignment models are finally addressed and some future research themes are suggested. A fundamental issue of any transit traffic assignment model has to do with the computational time needed. In the case of the CapTA model, a single iteration of the bounded model (CWCV) on a large-scale network such as the Paris Metropolitan Transit network takes up to 23 minutes in an efficient personal computer. A speeding-up of the simulator can be easily achieved, especially so since the algorithm

includes many independent processes that can be parallelized. This parallelization process can easily reduce the calculation time, although there exists a minimum bound. The calculation time is an important issue mainly in schedule-based assignment models that require extensive calculations over the reference period. The reduction of the computation time can come from more efficient algorithms and the improvement of the convergence for reducing the number of iterations needed to reach a satisfactory solution.

The modular framework is an important element of the CapTA model. The existing models can be modified and additional models can be added for modelling more detailed interactions between the system components. Such interactions can include the overcapacity effect produced when a boarding passenger faces a long waiting time and a very loaded. In fact, the boarding passengers are willing to accept – and make the others suffer – higher standing densities than the nominal capacity to compensate for a significant waiting time. Therefore the nominal capacity seems a strict but flexible variable, sensible to the waiting time of the boarding passengers.

An additional interaction to be modelled concerns the vehicle dwelling. This complex phenomenon concerns the boarding and alighting passenger flows and is influenced by the general on-platform and in-vehicle conditions. Degraded conditions, attributed to the interaction among boarding flows, alighting flows, in-vehicle load, on-platform passenger stock and vehicle and platform characteristics may result in longer dwelling times and increased travel time variability.

In these lines, we can sketch other complex interactions. Modelling accurately the in-vehicle and on-platform waiting conditions should reveal the passengers' trade-offs relevant to the waiting conditions. These conditions – seating or standing – depend on the availability of seats on the platform and the in-vehicle comfort states. In case in-vehicle sitting is difficultly available, a passenger may wait standing to improve his priority to board and his probability to get a seat. The decisions relate the perceived in-vehicle travel time and the perceived waiting time and depend on the passengers' characteristics.

An additional level of realism will be gained with the modelling of the reliability of the transit network: the recurrent small and occasional big delays of the transit services. Although that is a random phenomenon, it highly influences the passengers' perception on the quality of

service and the perceived travel time. That can be modelled exogenously (as a penalty) or endogenously with respect to certain conditions: mixture of heterogeneous services, vehicle load, and variability of dwell time. Anyhow, it is important for the evaluation of infrastructure projects that seek to improve the reliability of the services.

In addition to the modelling of particular passenger behaviour, the CapTA model allows the modelling of the service operations and the development of particular system sub-models. The existing line model includes some elements of the service operation with the modelling of the vehicle dwelling. However, the operational model can be extended to account for the delays caused by the convergence of rail branches and by the heterogeneity of the service. As long as the bus services are concerned, a specific line model can be developed for taking into account the influence of general traffic and traffic lights on the travel time and the journey variability. Furthermore, an access-egress model can be created for modelling in detail the quality of the interface between the intermodal facilities and the private modes (park and ride facilities, bike shelters and neighbouring road and pedestrian network) and the modes used for the access to and egress from the transit network.

The development of the intelligent transportation system and the explosion of the use of the intelligent devices and of the information available modify profoundly the behaviour and the choices of the passengers. A passenger decision is not limited to certain decision points, but the multiplication of the intelligent devices allows him to make his choice at any moment. It is essential to adapt the existing assignment models and develop new ones to account for the way the presence and the type of the real-time information available influences the passengers' choices.

References

- [1] Aguilera V., Allio S., Benezech V., Combes F., Millon C. (2012), *Estimating the quality of service of underground transit systems with cellular network data*, Paper presented in the 4th Transportation Research Arena Conference, Athens
- [2] Andreasson I. (1976), *A method for the analysis of transit networks*, Proceeding of the 2nd European Congress on operations Research (Edited by Mark Roubens), North Holland, Amsterdam
- [3] Añez J., De La Barra T., Pérez B. (1996), *Dual graph representation of transportation networks*, Transportation Research Part B, Vol. 30, pp 209-216
- [4] Banos A., Charpentier A. (2010), *Simulating pedestrian movement in dynamic environments*, Cybergeog: European Journal of Geography: Systèmes, Modélisation, Géostatiques, article 499
- [5] Beckmann M., McGuire C.B., Winsten C.B. (1955), *Studies In The Economics of Transportation*, Yale University Press, New Haven, Connecticut
- [6] Bell M.G.H. (1995), *Alternatives to Dial's logit assignment algorithm*, Transportation Research Part B, Vol. 29 (4), pp 287-295
- [7] Benezech V. (2013), *Passengers' experience of quality of service in urban transport networks : A Stochastic Approach*, PhD Thesis, Ecole Doctorale "Ville, Transports et Territoires", Université Paris-Est
- [8] Benezech V., Combes F., Leurent F. (2013), *Analysis of the variability of travel conditions along a transit line*, Paper presented at TRB Annual Meeting 2013, Washington D.C.

- [9] Blue V.J. and Adler J.L. (2001), *Cellular automata microsimulation for modelling bi-directional pedestrian walkways*, Transportation Research Part B, vol. 35, pp 293 – 312
- [10] Blum J.R. (1954), *Multidimensional Stochastic Approximation Methods*, Annals of Mathematical Statistics, Vol. 25, No. 4, pp 737-744
- [11] Bouzaïene-Ayari B., Gendreau M., Nguyen S. (1995), *An equilibrium-fixed point model for passenger assignment in congested transit networks*, Technical Report CRT-95-57, U. de Montréal
- [12] Bouzaïene-Ayari B., Gendreau M., Nguyen S. (2001), *Modelling Bus Stops in Transit Networks : A survey and New Formulations*, Transportation Science, vol. 35, No. 3, pp. 304-321
- [13] Buckman L.T. and Leather J.A. (1994) *Modeling station congestion the PEDROUTE way*, Traffic Engineering and Control. Vol. 35, pp 373-377
- [14] Bureau of Public Roads (1964) *Traffic Assignment Manual*, US Department of Commerce, Urban Public Division, Washington D.C.
- [15] Cepeda M. (2002), *Modèle d'équilibre dans les réseaux de transport en commun: le cas des capacités explicites des services*. PhD Thesis, Département d'Informatique et Recherche Opérationnelle, Publication 2002-43, CRT, Université de Montréal
- [16] 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*, Transport Research Part B, pp. 437-459
- [17] Chabini I. (1998), *Discrete dynamic shortest path problems in transportation applications: complexity and algorithms with optimal running time*, Transportation Research Record Vol. 1645, pp. 170-175
- [18] Chandakas E. (2009), *La capacité des transports ferroviaires de l'Ile-de-France face à la hausse du trafic à long terme*, Master Thesis, Master « Cité et Mobilité », Ecole Nationale des Ponts et Chaussées

- [19] Chandakas E. (2012a), *Note sur l'Ajustement des données de temps de trajet du réseau DRIEA 2008*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est (in french)
- [20] Chandakas E. (2012b), *Note sur la Capacité du Matériel Roulant et son Affectation sur le Réseau Francilien*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est (in french)
- [21] Chandakas E. and Leurent F. (2013), *A Transit Station Model for Passenger Traffic Assignment with Capacity Constraints for a Transit Network*, Paper #1993 presented in the 13th WCTR Conference, 2013, Rio de Janeiro, Brazil
- [22] Chandakas E., Leurent F., Poulhes A. (2013a), *A Transit Assignment Model with Capacity Constraints: A Large-Scale Implementation*, Paper #1440 presented in the 13th WCTR Conference, 2013, Rio de Janeiro, Brazil
- [23] Chandakas E., Leurent F. and Poulhes A. (2013b), *Un modèle de trafic de voyageurs en transports collectifs, sensible aux contraintes de capacité*, In Peuportier B. (ed) *Eco-conception des ensembles bâtis et des infrastructures*, pp 167-188 (in french)
- [24] Chandakas E., Leurent F., Poulhès A. (2014), *Les contraintes de capacité dans les trafics de voyageurs et de véhicules en transport collectif : un modèle de simulation et son application au Grand Paris*, Paper presented at the ATEC ITS France Conference, Paris, France, January 2014 (in french)
- [25] Cohen G. (1988), *Auxiliary Problem Principle extended in Variational Inequalities*, Journal of Optimization Theory and Applications, Vol. 59, No. 2, pp 325-333
- [26] Combes F., Benezech V., Chandakas E., Leurent F., Samadzad M. and Windisch E. (2011), *Revue Critique du modèle Antonin2*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est (in french)
- [27] Chriqui C. and Robillard P. (1975), *Common Bus Lines*, Transportation Science 9, pp. 115-121

- [28] Cominetti R. and Correa J. (2001), *Common-Lines and Passenger Assignment in Congested Transit Networks*, Transport Science, Vol. 35, No.3, pp 250-267
- [29] Daamen W. (2002), *SimPed : a Pedestrian Simulation Tool for Large Pedestrian Areas*, conference proceedings EuroSIW, 24-26 June 2002
- [30] Daly P.N., McGrath F., Annesley T.J. (1991), *Pedestrian speed/flow relationships for underground stations*, Traffic Engineering and Control, Vol. 32, Issue 2, 75-78
- [31] De Cea J. and Fernandez E. (1989), *Transit Assignment to Minimal Routes: An efficient new Algorithm*, Traffic Engineering and Control, pp 491-494
- [32] De Cea J. and Fernandez E. (1993), *Transit Assignment for Congested Public Transport Systems: An equilibrium Model*, Transportation Science vol.27, No2, pp. 133-147
- [33] De Cea J., Bunster J.P., Zubieta L. and Florian M. (1988), *Optimal Strategies and Optimal routes in public transit assignment models: an empirical comparison*, Traffic Engineering and Control, pp, 520-526
- [34] Debrincat L., Goldberg J., Duchateau H., Kross E., Kouwenhoven M. (2006), *Valorisation de la régularité des radiales ferrées en Ile-de-France*. Proceedings of the ATEC Congres 2006, CD Rom edition.
- [35] Dial R.B. (1967), *Transit Pathfinder Algorithm*, Highway Research Record 205, HRB, National Research Council, Washington DC, pp. 67-85
- [36] Dijkstra E.W. (1959), *A note on two problems in connexion with graphs*, Numerische Mathematik, Vol. 1, 269-271
- [37] EgisRail (2009), *Mission d'assistance à l'élaboration du Schéma Directeur du Matériel Roulant Ferroviaire Ile-de-France*, April 24, 2009
http://www.stif.info/IMG/pdf/Deliberation_no2009-0576_relative_au_schema_directeur_du_materiel_roulant.pdf (Accessed April, 2012)

- [38] Fearnside K. and Drapper D.P. (1971), *Public Transport Assignment – A new Approach*, Traffic Engineering and Control vol. 13, pp 298-299.
- [39] Fiegel J., Banos A., Bertelle C. (2009), *Modelling and Simulation of pedestrian behaviours in transport areas: the specific case of platform/train exchanges*, Proceedings of ICCSA 2009, June 29 – July 2, Le Havre
- [40] Frank M. and Wolfe P. (1956), *An Algorithm for Quadratic Programming*, Naval Research Logistics Quarterly, Vol. 3, pp 95-110
- [41] Gallo G., Longo G., Pallottino S., Nguyen S. (1993), *Directed Hypergraphs and Applications*, Discrete Applied Mathematics, Vol. 42, Issue 2-3, pp. 177-201
- [42] Gendreau M. (1984), *Etude Approfondie d'un modèle d'équilibre pour l'affectation des Passagers dans les Réseaux de Transport en Commun*, Pub. 384, Centre de Recherche Sur les Transports, Université de Montréal
- [43] Gentile G., Nguyen S., Pallottino S. (2005), *Route choice on transit networks with online information at stops*, Transport Science Vol.39, pp. 289-297
- [44] Guo Z. and Wilson N.H.M. (2011), *Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground*, Transportation Research Part A, Vol. 45, 91-104
- [45] Hamdouch Y. and Lawphongpanich S. (2008), *Schedule-based transit assignment model with travel strategies and capacity constraints*, Transportation Research Part B, Vol 42, pp 663 – 684
- [46] 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, Vol. 45, pp 1805-1830
- [47] Hamdouch Y., Marcotte P., Nguyen S. (2004), *A Strategic Model for Dynamic Traffic Assignment*, Networks and Spatial Economics, Vol. 4, pp 291-315
- [48] Harris N.G. (1989), *Capacity Restrained Simulation in Public Transport Environment*, Traffic Engineering and Control, Vol 30-6

- [49] Harris N.G. (1991), *Modelling walk link congestion and the prioritisation of congestion relief*, Traffic Engineering and Control, Vol 32, Issue 2, 78-80
- [50] Harris N.G. (2005), *Train Boarding and Alighting Rates at High Passenger Loads*, Journal of Advanced Transportation Vol. 40, No.3, pp 249-263
- [51] Harris N.G. and Anderson R.J. (2007), *An international comparison of urban rail boarding and alighting rates*, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, Vol.221, Issue 4, pp 521-526
- [52] Hasselström D. (1989), *Public Transportation Planning – A Mathematical Programming Approach*, PhD Thesis, Department of Business Administration, University of Gothenburg, Sweden
- [53] Helbing D. and Molnar P. (1995), *Social Force model for pedestrian dynamics*, Phys. Rev. E, Vol. 51, Issue 5, pp 4282-4286
- [54] Hickman M.D. and Wilson N.H.M. (1995), *Passenger Travel Time and Path Choice Implications of Real-Time Transit Information*, Transportation Research Part C, Vol. 3, No. 4, pp 211-226
- [55] Hoogendoorn S.P. (2001), *Normative Pedestrian Flow Behaviour: Theory and Applications*. Research Report Vk2001.02, Transportation and Traffic Engineering Section, Delft University of Technology
- [56] Hoogendoorn S.P. and Bovy PHL (2004), *Pedestrian route-choice and activity scheduling theory and models*, Transportation Research Part B, vol 38, pp 169-190
- [57] Hoogendoorn S.P. and Daamen W (2005), *Pedestrian Behaviour at Bottlenecks*, Transportation Science, Vol 39, No.2, May 2005, pp 147-159
- [58] Hoogendoorn S.P., Hauser M., Rodrigeus N. (2004), *Applying microscopic pedestrian flow simulation to railway station design evaluation in Lisbon, Portugal*, Proceedings of the 83rd Annual Meeting of Transportation Research Board, Washington D.C., pp 83-94

- [59] International Union of Railways, UIC (2010), *Influence of the European Train Control System on the capacity of nodes*, May 2010
- [60] Klüpfel H., Schreckenberg M. Meyer-König T. (2005), *Models for Crowd Movement and Egress Simulation*, In: Traffic and Granular Flow '03, Hoogendoorn S.P., Luding S, Bovy PHL, Schreckenberg M, Wolf DE, Ed. Springer Berlin Heidelberg, 2005, pp 357-372
- [61] Kroes E., Kouwenhoven M., Duchateau H., Debrincat L., Goldberg J., (2007), *Value of Punctuality on Suburban Trains to and from Paris*. Transportation Research Record: Journal of the Transportation Research Board, 2006, pp 67-75
- [62] Kurauchi F., Bell M.G.H., Schmocker J.-D. (2003), *Capacity Constrained Transit Assignment with Common Lines*, Journal of Mathematical Modelling and Algorithms, Vol. 2, pp 309-327
- [63] Lai Y.-C., Wang S.-H., Jong J.-C. (2011), *Development of analytical capacity models for commuter rail operations with advanced signaling systems*, Paper #11-3424 presented at the 90th Annual Transportation Research Board Meeting, Washington DC, January 2011
- [64] Lam W.H.K. and Bell M.G.H. (2003), *Advanced Modelling for Transit Operations and Service Planning*, Emerald Group Publishing, 349p.
- [65] Lam W.H.K., Cheung C.-Y., Lam C.F. (1999a), *A study of crowding effects at the Hong Kong light rail transit stations*, Transportation Research Part A, Vol 33, pp 401-415
- [66] Lam W.H.K., Gao Z.Y., Chan K.S., Yang H. (1999b), *A stochastic user equilibrium assignment model for congested transit networks*, Transportation Research Part B, Vol. 33, pp 351-368
- [67] Lam W.H.K., Zhou J., Sheng Z.-H. (2002), *A capacity restraint transit assignment with elastic line frequency*, Transportation Research Part B, Vol 36, pp 919-938

- [68] Last A. and Leak S.E. (1976), *Transept: A bus model*, Traffic Engineering and Control, pp 14-20
- [69] Le Clercq F. (1972), *A public Transport Assignment Method*, Traffic Engineering and Control 14, pp 91-96
- [70] Le Figaro (2010), *Les franciliens souffrent de leurs transports en commun*, Le Figaro report on 02/03/2010 <<http://www.lefigaro.fr/entreprise/2010/03/02/05011-20100302ARTFIG00516-les-franciliens-souffrent-de-leurs-transports-en-commun-.php>> (accessed on June 2012)
- [71] Le Monde (2012), *La RATP achète 70 nouvelles rames pour renouveler celles du RER A*. Le Monde report on June 29, 2012. http://www.lemonde.fr/economie/article/2012/06/29/la-ratp-achete-70-nouvelles-rames-pour-renouveler-celles-du-rer-a_1727035_3234.html (Accessed July 30, 2012)
- [72] Leurent F. (2003), *On network assignment and supply-demand equilibrium: an analysis framework and a simple dynamic model*. Paper presented at the 2003 European Transport Conference (CD Rom edition)
- [73] Leurent F. (2006), *Confort et qualité de service en transport collectif urbain de voyageurs: analyse, modélisation et évaluation*. Proceedings of the ATEC Congress 2006, CD-Rom edition (in french)
- [74] Leurent F. (2010a), *modèle PSAT*, linéaire, 1a, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est., jan 2010
- [75] Leurent F. (2010b), *modèle PSAT*, multiclasse 0a, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est., fév 2010
- [76] Leurent F. (2010c), *Un modèle à Stocks de Passagers et Seuils d'Attractivité pour l'affectation du trafic sur un réseau de transport collectif sous contraintes de capacité des véhicules*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est, mai 2010
- [77] Leurent F. (2011a), *Transport capacity constraints on the mass transit system: a systemic analysis*, European Transportation Research Review, 2011

- [78] Leurent F. (2011b), *The Transit Bottleneck Model*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est.
- [79] Leurent F. (2012a), *On Seat Capacity in Traffic Assignment to a Transit Network*, Journal of Advanced Transportation, April 2012, Vol 46, Issue 2, pp 112-138, doi: 10.1002/atr.146
- [80] Leurent F. (2012b), *A regularization process to yield smooth upper layer link flows with respect to link costs and frequencies in the CapTA model*, Working Document, LVMT.
- [81] Leurent F. and Askoura Y. (2010), *The person capacity of a transit route: a review, assessment and benchmark of static models for network traffic assignment*, Paper Presented at TRB Meeting 2010
- [82] Leurent F. and Benezech V. (2011), *The Passenger Stock and Attractivity Threshold Model for Traffic Assignment on a Public Transit Network with Capacity Constraint*, Paper Presented at TRB Annual Meeting 2011, Washington D.C.
- [83] Leurent F. and Chandakas E. (2012a), *Capacité des véhicules, attente à quai et choix d'itinéraire en transport collectif urbain – modélisation pour la planification du réseau TC*, In Hégron G. and Prévost T (eds) *Modélisation Urbaine : de la Représentation au Projet*. Collection Références du Commissariat Général au développement durable, pp 192-195 (in french)
- [84] Leurent F. and Chandakas E. (2012b), *The Transit Bottleneck Model*. Procedia - Social and Behavioral Sciences, Vol. 54, pp. 822-833.
- [85] Leurent F. and Liu K. (2009), *On seat congestion, passenger comfort and route choice in urban transit: a network equilibrium assignment model with application to Paris*, Paper Presented at TRB Annual Meeting 2009, Washington D.C.
- [86] Leurent F., Chandakas E. and Poulhes A. (2011), *User and service equilibrium in a structural model of traffic assignment to a transit network*, in Zak J. (ed) *The State of the Art in the European Quantitative Oriented Transportation and Logistics Research – 14th Euro Working Group on Transportation & 26th Mini Euro*

- Conference & 1st European Scientific Conference on Air Transport. Elsevier Procedia Social and Behavioral Sciences, 20: 495-505.
- [87] Leurent F., Chandakas E. and Poulhes A. (2012a), *A passenger traffic assignment model with capacity constraints for transit networks*. Procedia - Social and Behavioral Sciences, Vol. 54, pp. 772-784.
- [88] Leurent F., Chandakas E. and Poulhes A. (2012b), *Bi-layer equilibrium of user and service in a capacitated model of traffic assignment to a transit network*. Working paper submitted for publication to European Transport.
- [89] Lin T.-M. and Wilson N.H.M (1992), *Dwell Time Relationships for Light Rail Systems*, Transportation Research Record, Vol. 1361, pp.287-295
- [90] Liu K. (2009), *Modélisation du choix d'itinéraire plurimodal en transport urbain de voyageurs*, PhD Thesis, Ecole Doctorale « Ville et Environnement », Université Paris Est (in french)
- [91] Marcotte P. and Nguyen S. (1998), *Hyperpath Formulations of Traffic Assignment Problems*, In: Equilibrium and Advanced Transportation Modelling, P. Marcotte, S. Nguyen, Kluwer Academic Publisher, pp 175-183
- [92] Marcotte P., Nguyen S., Schoeb A. (2004), *A Strategic flow model of Traffic Assignment in Static Capacitated Networks*, Operations Research, Vol. 52, No. 2, March-April 2004, pp 191-212
- [93] Marguier P.H.J. and Ceder A. (1984), *Passenger Waiting Strategies for Overlapping Bus Routes*, Transportation Science 18, pp 207 – 230
- [94] Merlin P. (1991), *Géographie, économie et planification des transports*, Presses Universitaire de France, 480p
- [95] Meschini L., Gentile G., Papola N. (2007), *A frequency-based transit model for dynamic traffic assignment to multimodal networks*, In: Presented at 17th International Symposium on Transportation and Traffic Theory, London, July, 2007

- [96] Milgrom P. and Segal I. (2002), *Envelope Theorems for Arbitrary Choice Sets*, *Econometrica*, Vol. 70, N o. 2, pp 583-601
- [97] Moller and Pedersen J., (1999), *Assignment model of timetable based systems (TPSCHEDULE)*, Proceedings of 27th European Transportation Forum, Seminar F, Cambridge, England, pp 159-168
- [98] National Fire Protection Association (2010), *NFPA 130: Standard for Fixed Guideway Transit and Passenger Rail Systems*, 2010 Edition
- [99] Nguyen S. and Pallottino S. (1988), *Equilibrium Traffic Assignment for large scale transit Networks*, *European Journal Operational Research* 37, 1988, pp 176-186
- [100] Nguyen S., Pallottino S., Malucelli F., (2001), *A modelling Framework for Passenger Assignment on a Transport Network with Timetables*, *Transportation Science*, Vol 35, No. 3, August 2001, pp 238 – 249
- [101] Nuzzolo A., Crisalli U., Rosati L. (2007), *Congested Transit Networks: A schedule-based dynamic assignment model with explicit vehicle capacity constraints*, In Proceeding of European Transport Conference, Leeuwenhorst, The Netherlands
- [102] Nuzzolo A., Crisalli U., Rosati L. (2012), *A schedule-based assignment model with explicit capacity constraints for congested networks*, *Transportation Research Part C*, Vol. 20, pp 16-33
- [103] Nuzzolo A., Russo F., Crisalli U., (2001), *A Doubly Dynamic Schedule-based Assignment Model for Transit Networks*, *Transport Science* Vol. 35, No. 3, pp 268 – 285
- [104] OMNIL (2012), *Synthèse de Principaux Résultats de l'EGT 2010*, Edition of the Observatoire de la mobilité en Ile-de-France, accessed November 2012 (in French)
- [105] Ortuzar J.D. and Willumsen L.G. (2001), *Modelling Transport 3rd edition*, Wiley, London
- [106] Oxford Economic Forecasting (2003), *The Economic Effects of Transport Delays on the City of London*, Corporation of London, July 2003

- [107] Paris S., Donikian S., Bonvalet N. (2006), *Environmental abstraction and path planning techniques for realistic crowd simulation*, Computer Animation and Virtual Worlds, Vol. 17, pp 325-335
- [108] Poulhes A. and Chandakas E. (2011), *Simulateur de Flux de Voyageurs dans les Transports Collectifs Urbains, Document de Conception*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est (in french)
- [109] Pushkarev B. and Zupan J. (1975), *Urban Space for Pedestrians*, MIT Press, Cambridge, Massachusetts
- [110] RATP (2011), *Mise en service du nouveau train à deux niveaux*. Press release, http://www.ratp.fr/fr/upload/docs/application/pdf/2011-12/dp_mise_en_service_mi09.pdf (Accessed July 30, 2012)
- [111] Robillard P. (1971), *(0,1) Hyperbolic Programming Problem*, Naval Research Log Quart 18, No1
- [112] Schmöcker J.-D., Bell M.G.H., Kurauchi F. (2008), *A quasi-dynamic capacity constrained frequency-based transit assignment model*, Transportation Research Part B, Vol. 42, pp 925 – 945
- [113] Schmöcker J.-D., Fonzone A., Bell M.G.H., Kurauchi F., Shimamoto H. (2011), *Frequency-based transit assignment considering seat capacities*, Transportation Research Part B, Vol. 45, pp 392 – 408
- [114] Shimamoto H., Kurauchi F., Iida Y., Schmocker J.-D., Bell M.G.H. (2005), *Evaluation of public transport congestion mitigation measures using a passenger assignment model*, Journal of the Eastern Asia Transportation Studies, Vol. 6, pp 2076 – 2091
- [115] SNCF-GI, (2002), *IN1724 : Sécurité du public dans les gares, à la traversée de voies et sur les quais*, Direction Déléguée Systèmes d'Exploitation et Sécurité (in french)
- [116] Spiess H. (1983), *On Optimum Route Choice Strategies in Transit Networks*, Pub 286, Centre de Recherche sur les Transports, Université de Montréal

- [117] Spiess H. (1984), *Contributions à la théorie et aux outils de planification des réseaux de transport urbain*, PhD Thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal, Quebec (in french)
- [118] Spiess H., Florian M. (1989), *Optimal Strategies: A New Assignment Model For Transit Networks*, Transportation Research Part B, pp 83-121
- [119] Still G.K. (2000), *Crowd Dynamics*, PhD Thesis, University of Warwick, 2000
- [120] Sumalee A., Zhijia T., Lam W.H.K. (2009), *Dynamic Stochastic transit assignment with explicit seat allocation model*, Transportation Research Part B, Vol. 43, pp 895 – 912
- [121] Sumi T., Matsumoto Y., Miyaki Y. (1990), *Departure time and route choice of commuters on mass transit systems*, Transportation Research Part B, Vol. 24, Issue 4, pp. 247 – 262
- [122] Teklu F. (2007), *A Stochastic Process Approach for Frequency-based Transit Assignment with Strict Capacity Constraints*, Networks and Spatial Economics, Vol. 8, No. 2-3, pp 225 – 240
- [123] Teknomo K. (2006), *Application of microscopic pedestrian simulation model*, Transportation Research Part F, Vol. 9, pp 15 – 27
- [124] Thomas R. (1991), *Traffic Assignment Techniques*, Avebury Technical Editions, 392p
- [125] Tian Q., Huang H.-J. and Yang H. (2007), *Equilibrium properties of morning peak-period commuting in a many-to-one mass transit system*, Transportation Research Part B, Vol. 41, no. 6, pp 616-631
- [126] Tong C.O., Wong S.C. (1999), *A stochastic transit assignment model using a dynamic schedule-based network*, Transportation Research Part B, Vol. 33, pp 107 – 121

- [127] Tong C.O., Wong S.C., Poon M.H., Tan M.C., (2001), *A schedule based dynamic transit network model – recent advances and prospective future research*, Journal of Advanced Transportation vol. 35, Iss. 2, pp 175 – 195
- [128] Transportation Research Board (2003), *Transit Capacity and Quality of Service Manual*, On-line report prepared for the Transit Cooperative Research Program, available on-line at the following website address: http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_webdoc_6-a.pdf . First edition, 1999
- [129] Trozzi V., Hoosseinloo S.H., Gentile G., Bell M.G.H. (2010), *Dynamic hyperpaths: the stop model*, In Proceeding of Models and Technologies for Intelligent Transportation Systems, International Conference Rome, ed. Fusco, Aracne, Rome, pp. 334-343
- [130] Trozzi V., Kaparias I., Bell M.G.H., Gentile G., (2012), *A dynamic route choice model for public transport networks with boarding queues*, Transportation Planning and Technology, Vol. 36, Issue 1, pp. 44-61
- [131] Trozzi V., Gentile G., Bell M.G.H., Kaparias I. (2013), *Dynamic User Equilibrium in public transport networks with passenger congestion and hyperpaths*, Transportation Research Part B, Vol. 57, pp. 266-285
- [132] Vuchic V.R. (2005), *Urban Transit: Operations, Planning and Economics*. Wiley Editions, 665 p.
- [133] Wardrop J.G. (1952), *Some theoretical aspects of road traffic research*, Proceedings of the Institute of Civil Engineers, Part II, pp 325-378
- [134] Wong S.C., Tong C.O. (1998), *Estimation of time-dependent origin-destination matrices for transit networks*, Transportation Research Part B, Vol 32, pp 35 – 48
- [135] Wu J.H., Florian M., Marcotte P. (1994), *Transit equilibrium assignment: A model and solution algorithms*, Transportation Science, Vol 28 (3), pp 193-203

ANNEXES

ANNEX A:

Vehicle Capacity on the Greater Paris Transit Network

Introduction

This annex acts as a census of the characteristics of the rolling stock that circulate on the transit network of the Paris Metropolitan Region. All the transit modes are included, from the buses and coaches to the regional trains. The characteristics of the vehicles are grouped in two parts: the passenger capacity and the exchange capacity of the vehicles. The annex contains only the vehicle characteristics that are relevant to the CapTA model. The rolling stock described is allocated to transit services used in the simulation in Chapter 9 and 10. The asterisk denotes that the value of the attributed is not confirmed.

Passenger Capacity

The passenger capacity is defined with respect to the seats available and the total passenger capacity of the rolling stock on the network. As long as the seat capacity is concerned, we distinguish the vehicle's seat capacity with and without the use of the folding seats. Furthermore, the data contain the surface available for the standing passengers (if the folding seats are not used).

The vehicle's total passenger capacity is defined as the sum of the seats (without the folding seats) and maximum number of standing passengers admitted, which is the product of the surface available for the standing passengers and the maximum density (4 standing passengers / m²).

The rolling stock is distinguished by transit mode. Table 17 contains the characteristics for the Parisian metro. The data are limited to the situation in 2008. The new data of the renovated vehicles after 2008 are not included.

Table 17: The passenger capacity of the metro

Rolling Stock	Nb of carriages	Line	Seats (without folding seats)	Seats (with folding seats)	Surface (m ²)	Total Capacity
MP 89 CC	6	1	132	242	147	720
MF 67	5	2	120	264*	114	575
MF 67	5	3	120	264*	114	575
MF 67	3	3bis	72	156*	67	341
MP 59	6	4	144	320	139	700
MF 67	5	5	120	264*	114	575
MF 73	5	6	120	266	114	575
MF 77	5	7	128	240	112	574
MF 88	3	7bis	64	64	72	351
MF 77	5	8	128	240	112	574
MF 67	5	9	120	264*	114	575
MF 67	5	10	120	264*	114	575
MF 59	4	11	96	212	92	464
MF 67	5	12	120	264*	114	575
MF 77	5	13	128	240	112	574
MP 89 CA	6	14	144	216	145	722

* estimation

Table 18 contains the characteristics of the rolling stock used in the RER and the Commuter Rail by the rail operators, RATP and SNCF. Some trainsets are used by both operators and have a double name. These are the MI79 (RATP) or Z8100 (SNCF) and the MI84 (RATP) or Z8400 (SNCF). In addition, the Z22500 trainset of the SNCF and the MI2N of the RATP are very similar in their characteristics, but must be distinguished in matters of their interior design and passenger capacity.

Furthermore, in Table 18, the US makes reference to the elementary train that can run on the rail network. The UM_x denotes a trainset composed of x elementary trains. The xV represents the number of carriages destined to passengers (for the train composition with an engine carriage) and the xC symbolizes the number of carriages to compose an elementary train. The lines on which these trains are assigned are given on an indicative basis.

Table 18: The passenger capacity on the RER and commuter rail

Rolling Stock	Line	Seats (without folding seats)	Seats (with folding seats)	Surface (m ²)	Total Capacity
Z5300.US	D, R, N	390	469	101	794
Z5300.UM2	D, R, N	780	938	202	1588
Z5600.4C.UM2	C, R	1102	1160	194	1878
Z5600.6C.US	C	883	940	155	1503
Z6100.UM3	K, H	819	849	198	1611
Z6300.UM3		534	729	195	1314
Z6400.UM2	L	730	904	184	1466
Z6400.GCO.US	L	272	272*	100	673
Z8100.UM2	B	630	856	264	1686
Z8800.UM2	C, U	1070	1128	186	1814
Z20500.4C.UM2	C, D, P, L, H K	1200	1200	202	2002
Z20500.5C.UM2		1608	1608	252	2610
Z.20900.4C.UM2	C, H, K	968	1000	194	1744
Z22500.UM2	E	1100	1100	366	2564
Corail.7V		545	545	113	994
Corail.8V		623	623	284	1759
Corail.9V		700	700	145	1278
Corail.10V		760	760	-	760
RIB/RIO.8C		880	880	142	1446
Z20500.4C+5C		1404	1404	227	2306
X72500.3C.UM2		438	456	-	456
VB2N.6V	J	888	888	155	1510
VB2N.7V	J, H	1045	1045	184	1778
VO2N.8V	PSL	980	980	-	980
MS61.UM3	A	600	876	322	1887
Z2N RATP.UM2	A	1056	1056	381	2580
MI84.UM2	A	432	*	332	1760
NAT.7C	L	318	392	109	754
NAT.8C	J, H	394	486	130	915

* estimation

Table 19 summarizes the passenger characteristics of the light rail vehicles. However, they contain the bus vehicles that are used in the TVM bus rapid transit line. Furthermore, the light rail train Siemens Avanto S70 is usually known with its SNCF name of U25500.

Table 19: The passenger capacity of the light rail and bus rapid transit vehicles

Rolling Stock	Line	Seats (without folding seats)	Seats (with folding seats)	Surface (m ²)	Total Capacity
TFS	T1	54	58	30	174
Citadis 302.US	-	32	48	45	212
Citadis.302.UM2	T2	64	96	90	424
Citadis 402.US	T3	78	78*	57	304
Avanto S70	T4	80	86	41	242
Irisbus Agora L	TVM	45	45*	26	149

* estimation

Finally, the passenger capacity of the buses and coaches used in the Paris Metropolitan Region are addressed. Nevertheless, the diversity of the rolling stock used and the absence of precise information make the allocation of vehicles to transit routes a hard task. The passenger capacity of these vehicles is a product of the information of the Optile (concession of private operators) and the RATP to the STIF (the Metropolitan Transport Authority). Table 20 and Table 21 report the composition of the fleet and its passenger characteristics of the buses and coaches respectively.

Table 20: The bus fleet composition and its characteristics (source: STIF from Optile and RATP)

	Optile			RATP		
	Nb of Vehicles	Nb of Seats	Total Capacity	Nb of Vehicles	Nb of Seats	Total Capacity
Standard Bus	2000	21-35	80-125	3660	26-37	91-115
Articulated Bus	331	35-61	136-198	476	38-58	133-167
Minibus	44	9-5	17-37	85	9-14	22-32
Midibus	194	16-25	63-90	72	14-21	55-73

Table 21: The coach fleet composition and its characteristics (source: STIF from Optile and RATP)

	Optile			RATP		
	Nb of Vehicles	Nb of Seats	Total Capacity	Nb of Vehicles	Nb of Seats	Total Capacity
Standard Coach	2011	49-64	49- 95	0	0	0
Minicoach	48	7-23	7-23	4	16	16
Midicoach	30	27- 32	27- 38	2	35- 39	35- 39

We observe that there is a big diversity of bus and coaches running in the network. For the simulation, we consider a standard bus and coach vehicles by category, whatever their

operator. The characteristics of the vehicles included in the simulation of the Greater Paris transit network are given in Table 22.

Table 22: The passenger capacity of the buses and coaches

Vehicle	Seats (without folding seats)	Seats (with folding seats)	Surface (m ²)	Total Capacity
Standard Bus	31	34	18	103
Standard Articulate Bus	48	54	26	152
Standard Minibus	12	12	4	28
Standard Midibus	21	21	28	77
Standard Coach	49	49	0	49
Standard Minicoach	16	16	0	16
Standard Midicoach	32	32	0	32

Exchange Capacity

The exchange capacity is defined as the maximum number of passengers that can cross a given section. Namely, it is limited to the number and the width of the doors and more precisely to the number of flow streams available per vehicle side. A flow stream is defined as the width at the door occupied by a single passenger when boarding or alighting a vehicle. The standard width of a flow stream is assumed to be 650 mm (EgisRail, 2009).

For the calculation of the number of flow streams it suffices to divide the width of a vehicle door by the elementary width of a flow stream. The number of flow streams per door is the result of the division rounded down to the closest integer. The total number of flow streams is the sum of the flow streams of all the doors per side of the rolling stock.

Table 23: Exchange capacity on the metro

Rolling Stock	Line	Nb of carriages per trainset	Nb of doors per side	Flow streams per side	Typical door width (mm)
MP 89 CC	1	6	18	36	1650
MF 67	2	5	20	40	1300
MF 67	3	5	20	40	1300
MF 67	3bis	3	12	24	1300
MP 59	4	6	24	48	1300
MF 67	5	5	20	40	1300
MF 73	6	5	20	40	1300
MF 77	7	5	15	30	1575
MF 88	7bis	3	9	18	1586

Rolling Stock	Line	Nb of carriages per trainset	Nb of doors per side	Flow streams per side	Typical door width (mm)
MF 77	8	5	15	30	1575
MF 67	9	5	20	40	1300
MF 67	10	5	20	40	1300
MF 59	11	4	24	48	1300
MF 67	12	5	20	40	1300
MF 77	13	5	15	30	1575
MP 89 CA	14	6	18	36	1650

Table 24 resumes the exchange capacity of the rolling stock on the RER and the Commuter Rail. However, it is common to a number of trainsets to have doors with different width. Namely, for the Z5600 and the similar rolling stock there is:

- Z5600.4C: composition (ZB+ZRB+ZRAB+ZB). It has 4 doors with a width of 1300mm and 4 doors with a width of 1800mm
- Z5600.6C: composition (ZB+3*ZRB+ZRAB+ZB). It has 4 doors with a width of 1300mm and 8 doors with a width of 1800mm

On the RER C a variety of rolling stock runs in various compositions (Z8800.4C.UM2, Z20500.4C.UM2 and Z20900.4C.UM2). However, they have the same characteristics with respect to the exchange capacity.

Table 24: Exchange capacity on the RER and Commuter Rail

Rolling Stock	Line	Nb of carriages per trainset	Total nb of doors per side	Flow streams per side	Typical door width (mm)
Z5300.US	D, R, N	4	12	24	1250
Z5300.UM2	D, R, N	8	24	48	1250
Z5600.4C.UM2	C, R	8	16	32	1300 – 1800 (ZR)
Z5600.6C.US	C	6	12	24	1300 – 1800 (ZR)
Z6100.UM3	K, H	9	27	54	1700
Z6300.UM3		9	27	54	1700
Z6400.UM2	L	8	24	48	1300
Z8100.UM2	B	8	32	64	1300
Z8800.UM2	C, D, P, L, H, K	8	16	32	1300 – 1800 (ZR)
Z20500.4C.UM2		8	16	32	1300 – 1800

Rolling Stock	Line	Nb of carriages per trainset	Total nb of doors per side	Flow streams per side	Typical door width (mm)
					(ZR)
Z20500.5C.UM2	C, H, K	10	20	40	1300 – 1800 (ZR)
Z22500.UM2	E	10	30	90	2000
Corail.7V		7	14	14	650
Corail.8V		8	16	16	650
Corail.9V		9	18	18	650
Corail.10V		10	20	20	650
RIB/RIO.8C		8	24	48	1300
Z20500.4C+5C		9	18	36	1300 – 1800 (ZR)
CAR		1	1	1	650
X72500.3C.UM2		6	6	6	650
VB2N.6V	J	6	12	24	1800
VB2N.7V	J, H	7	14	28	1800
VO2N.8V	PSL	8	16	32	1800
Z.20900.4C.UM2	C	8	16	32	1300 – 1800 (ZR)
MS61.UM3	A	9	36	72	1300
Z2N RATP.UM2	A	10	30	90	2000
MI84.UM2	A	8	32	64	1300
NAT.7C	L	7	7	21	1950
NAT.8C	J, H	8	8	24	1950

Table 25 groups the exchange characteristics of the light rail trains and the vehicles used in the TVM bus rapid transit line.

Table 25: Exchange capacity on the light rail and bus rapid transit

Rolling Stock	Line	Nb of carriages per elementary unit	Total nb of doors per side	Flow streams per side	Typical door width (mm)
TFS	T1	3	4	8*	-
Citadis 302.US	-	5	6	10*	-
Citadis.302.UM2	T2	10	12	20*	-
Citadis 402.US	T3	7	8	14*	-
Avanto S70	T4	5	5	10*	-
Irisbus Agora L	TVM	1	3	6*	-

* estimation

ANNEX B:

On the In-Vehicle Comfort Model with Folding Seats

Introduction

The in-vehicle comfort model is developed in Chapter 4 and the algorithms are detailed in Chapter 8.3.3.2. Furthermore, the folding seats and their activation mechanism are described in Chandakas (2009) and developed in Chapter 4.3.3. Annex B presents a variant of the in-vehicle comfort model, including the effect of the folding seats. Moreover, in chapter 4.3.3.2 we describe the in-vehicle comfort states available when taking into account the folding seats.

Passenger's Travel Behaviour and Priority Rules

Every passenger seeks to minimize his perceived travel cost. Therefore, a passenger who is standing has the motivation to seat on a normal seat and if he fails on a folding seat. In addition, a passenger who is seated on a folding seat wishes to occupy a normal seat, in order to reduce the probability to stand later, if the folding seats are deactivated due to the downstream density of standing passengers. Therefore, if the number of available seats is not sufficient for the standees, only a proportion will be seated.

Once a seat becomes available, there is competition. At that time, we will have standing and “folding” passengers that stay on-board with a priority over boarding passengers. We assume that folding and standing passengers have equal probability to occupy a normal seat. Therefore we have four successive competitions: first, among standing and folding passengers for normal seats. Then, for standing passengers to occupy folding seats. Later for boarding passengers to occupy normal seats and if failing, folding seats.

The Folding Seat Mechanism

We define the folding seats as a different in-vehicle comfort state, due to their activation/deactivation particularity. Namely, at low levels they can be used to increase in-vehicle comfort, but at high levels of passenger flow, the passengers are requested to liberate their seats to make more room for standing passengers and therefore to improve their standing conditions. In that case they join them in standing. On the contrary, when the density of standing passengers is reduced, some standing passenger will activate the folding seats. However, the precariousness of sitting makes the passengers sitting on folding seats to be equally motivated with the standing passengers to occupy normal seats.

We further describe the conditions for activating and deactivating folding seats. When at boarding the standing passengers exceed a given standing density, the passengers on the folding seats will be forced to give up their seats (deactivate folding seats) for liberating standing space and join them in standing. In contrast, when the folding seats are not used and the density of standing passengers fall to a certain level, the standing passengers will use the folding seats (activating). The density levels for activating and deactivating folding seats don't have to be equal.

Service Leg Cost

The attribution of a given service mode does not depend on the passenger, since he will always try to improve his travelling conditions and occupy a less onerous comfort state. The seat allocation process includes three comfort states and two user classes, with transition probabilities between each state and for each class of passengers. We define the following state transitions: $p_{zi}^{(o/+),r}$ for through and boarding passengers that occupy a normal seat and $p_{zi}^{(o/+),\tilde{r}}$ for through and boarding passengers that occupy a folding seat. Finally, the dual variable δ_{zi}^f is used to state whether the folding seats comfort state is activated ($\delta_{zi}^f = 1$) or not ($\delta_{zi}^f = 0$). On the basis of these probabilities and the binary variable, the generalized time of the access-egress service leg is calculated in a recursive way.

Service Flow Loading Model: Implementation Algorithm

The line flow loading algorithms seeks to assign the line's access-egress matrix to the seating, folding seats and standing states, along the line segments and therefore to establish sitting and folding sitting probabilities at station i for standing passengers, either boarding or on-board. We use the access – egress trip matrix (of route legs). At a station we have $k_{z<z,i>}^r$ the available capacity at arrival for each comfort state \underline{r} , \tilde{r} . The algorithm calculates the probabilities for occupying a comfort state when on-board, $p_{zi}^{o,r}$, or boarding $p_{zi}^{+,r}$. The line algorithm treats each station at a topographic order from upstream to downstream. The algorithm is built in seven steps, as follows.

Algorithm FS_COMF_ALC:

Input:

- b_{zis} : the passenger flow of the (i, j) access – egress station couple, $b_{zis} = \sigma_{is} \pi_{zi}$
- $[y_{za}^{sr}]_{s \in N_z}$: the vehicle passenger flow with comfort state r per egress station $s \in N_z$ on an arc a of a route $z \in Z_\ell$
- k_{za}^r : the available vehicle capacity for comfort state r on an arc a of route z . The sitting state is given by \underline{r} , the “folding” state by \tilde{r} and the standing state is \bar{r}
- A_z^a : the surface available on the vehicle for the standing passengers when the folding seats are in use
- A_z^d : the surface available on the vehicle for the standing passengers when folding seats are not in use
- d^a : the critical density of the standing passengers for the activation of the folding seats
- d^d : the critical density of the standing passengers for the deactivation of the folding seats

Output:

- $[y_{za}^{sr}]_{s \in N_z}$: the updated vehicle flow with comfort state r per egress station.
- k_{za}^r : the available vehicle capacity for comfort state r on an arc a of route z
- $p_{zi}^{o,r}$: the probability of occupying a seat if standing on-board at i
- $p_{zi}^{+,r}$: the probability of occupying a seat if boarding at i
- d_{zi} : the density of standing passengers at i
- δ_{zi}^f : the state of folding seats at i

Initialization:

Folding seat activation and deactivation vehicle critical volumes:

$$y^a = k_z^r + d^a \cdot A_z^d \text{ for the activation of the folding seats}$$

$$y^d = k_z^r + k_z^{\tilde{r}} + d^d \cdot A_z^a \text{ for the deactivation of the folding seats}$$

$$\delta_{zi}^f \leftarrow 1$$

Main:

Step 1 – Seat liberation: Residual capacity per comfort state (seats and folding seats) after seated passengers alight:

$$k_{z<i,i>}^r = k_{z<z,i>}^r + y_{z<z,i>}^{i,r} \text{ for seat capacity}$$

$$k_{z<i,i>}^{\tilde{r}} = k_{z<z,i>}^{\tilde{r}} + y_{z<z,i>}^{i,\tilde{r}} \text{ for folding seat capacity}$$

Step 2- Folding Seat Status:

Vehicle volume on-board at departure: $y_{z<i,z>} = \sum_{s>i} \sum_r y_{z<i,z>}^r$

If $\delta_{zi-1}^f = 1$ (folding seats are used) and $y_{z<i,z>} > y^d$, then

$$\delta_{zi}^f \leftarrow 0 \text{ and } k_z^{\tilde{r}} \leftarrow 0$$

$$\text{for } s > i, y_{z<i,i>}^{s,\tilde{r}} \leftarrow y_{z<i,i>}^{s,\tilde{r}} + y_{z<i,i>}^{s,\tilde{r}} \text{ and } y_{z<i,i>}^{s,\tilde{r}} \leftarrow 0$$

else if $\delta_{zi-1}^f = 0$ (folding seats are not used) and $y_{z<i,z>} < y^a$, then

$$\delta_{zi}^f \leftarrow 1 \text{ and } k_z^{\tilde{r}} \leftarrow k_{zo}^{\tilde{r}}$$

else if $\delta_{zi-1}^f = 0$ (folding seats are not used) and $y_{z<i,z>} > y^a$, then

$$p_{zi}^{o,\tilde{r}} = 0 \text{ and } p_{zi}^{+,\tilde{r}} = 0$$

Step 3 – On-board competition for seats:

$$- y_{z<i,i>}^{\tilde{r}} = \sum_{s>i} y_{z<i,i>}^{s,\tilde{r}} \text{ for on-board standing passengers and}$$

$$- y_{z<i,i>}^{\tilde{r}} = \sum_{s>i} y_{z<i,i>}^{s,\tilde{r}} \text{ for on-board passengers on folding seats}$$

$$- p_{zi}^{o,r} = \min\left\{1; \frac{k_{z<i,i>}^r}{y_{z<i,i>}^{\tilde{r}} + y_{z<i,i>}^{\tilde{r}}}\right\} \text{ for probability of sitting at } i, \text{ or } p_{zi}^{o,r} = 1 \text{ if the vehicle}$$

$$\text{flows are } y_{z<i,i>}^{\tilde{r}} = y_{z<i,i>}^{\tilde{r}} = 0$$

$$- k_{z<i,i>}^r = k_{z<i,i>}^r - p_{zi}^{o,r} \cdot (y_{z<i,i>}^{\tilde{r}} + y_{z<i,i>}^{\tilde{r}}), \text{ updating residual seat capacity and}$$

$$\text{respectively for the folding seats } k_{z<i,i>}^{\tilde{r}} = k_{z<i,i>}^{\tilde{r}} + p_{zi}^{o,r} \cdot y_{z<i,i>}^{\tilde{r}}$$

$$- \text{for } s > i, \text{ “sitting” } y_{z<i,i>}^{s,r} = y_{z<i,i>}^{s,r} + p_{zi}^{o,r} \cdot y_{z<i,i>}^{s,r}, \text{ “folding” } y_{z<i,i>}^{s,\tilde{r}} = (1 - p_{zi}^{o,r}) \cdot y_{z<i,i>}^{s,\tilde{r}}$$

$$\text{and standing } y_{z<i,i>}^{s,\tilde{r}} = (1 - p_{zi}^{o,r}) \cdot y_{z<i,i>}^{s,\tilde{r}} \text{ for egress flows}$$

Step 4 – On-board competition for folding seats:

if $\delta_{zi}^f = 1$, on-board competition for folding seats, then

- $p_{zi}^{o,\tilde{r}} = \min\{1, \frac{k_{z<i,i>}^{\tilde{r}}}{y_{z<i,i>}^{\tilde{r}}}\}$ at i , or $p_{zi}^{o,\tilde{r}} = 1$ if $y_{z<i,i>}^{\tilde{r}} = 0$
- $k_{z<i,i>}^{\tilde{r}} = k_{z<i,i>}^{\tilde{r}} - p_{zi}^{o,\tilde{r}} \cdot y_{z<i,i>}^{\tilde{r}}$, updating residual capacity
- for $s > i$, $y_{z<i,i>}^{s,\tilde{r}} = y_{z<i,i>}^{s,\tilde{r}} + p_{zi}^{o,\tilde{r}} \cdot y_{z<i,i>}^{s,\tilde{r}}$ and $y_{z<i,i>}^{s,\tilde{r}} = (1 - p_{zi}^{o,\tilde{r}}) \cdot y_{z<i,i>}^{s,\tilde{r}}$

Step 5 – Boarding competition for seats:

$b_{zi} = \sum_{s>i} b_{zis}$ for boarding passengers

$p_{zi}^{+,r} = \min\{1, \frac{k_{z<i,i>}^r}{b_{zi}}\}$ for probability of sitting at i , or $p_{zi}^{+,r} = 1$ if $b_{zi} = 0$

$k_{z<i,i>}^r = k_{z<i,i>}^r - p_{zi}^{+,r} \cdot b_{zi}$, updating residual capacity

for $s > i$, $y_{z<i,i>}^{s,r} = y_{z<i,i>}^{s,r} + p_{zi}^{+,r} \cdot b_{zis}$ and $y_{z<i,i>}^{s,r} = y_{z<i,i>}^{s,r} + (1 - p_{zi}^{+,r}) \cdot b_{zis}$ for egress flows

Step 6 – Boarding competition for folding seats:

if $\delta_{zi}^f = 1$, boarding competition for folding seats, then

- $p_{zi}^{+, \tilde{r}} = \min\{1, \frac{k_{z<i,i>}^{\tilde{r}}}{y_{z<i,i>}^{\tilde{r}}}\}$ at i , or $p_{zi}^{+, \tilde{r}} = 1$ if $y_{z<i,i>}^{\tilde{r}} = 0$
- $k_{z<i,i>}^{\tilde{r}} = k_{z<i,i>}^{\tilde{r}} - p_{zi}^{+, \tilde{r}} \cdot y_{z<i,i>}^{\tilde{r}}$, updating residual capacity
- for $s > i$, $y_{z<i,i>}^{s, \tilde{r}} = y_{z<i,i>}^{s, \tilde{r}} + p_{zi}^{+, \tilde{r}} \cdot y_{z<i,i>}^{s, \tilde{r}}$ and $y_{z<i,i>}^{s, \tilde{r}} = (1 - p_{zi}^{+, \tilde{r}}) \cdot y_{z<i,i>}^{s, \tilde{r}}$

Step 7 – Density of standing passengers:

$$d_{zi} = \frac{\sum_{s>i} y_{z<i,i>}^{s,r}}{\delta_{zi}^f \cdot A_z^a + (1 - \delta_{zi}^f) \cdot A_z^d}, \text{ the density of the standing passengers at station } i$$

END

Service Leg Costing Model: Implementation Algorithm

The FS_INV_COST algorithm is called by the UNZIP algorithm (see Chapter 8.3.2). Similarly to the INV_COST algorithm in Chapter 8.3.3.2, by egress station of the service route the cost from each access station is constructed in a recursive way by using the sitting probabilities calculated by the FS_COMF_ALC algorithm. The most significant modifications concern the use of three auxiliary variables for the calculation of the cost:

- $\bar{\gamma}_{z(i,s)}^r$ is the auxiliary cost if a passenger is standing at i . Downstream he may improve his comfort by occupying a seat or a folding seat;
- $\tilde{\gamma}_{z(i,s)}^r$ is the auxiliary cost if a passenger occupies a folding seat at i . Downstream he may improve his comfort by occupying a seat, or degrade and stand it if the folding seats are deactivated;
- $\gamma_{z(i,s)}$ is the auxiliary cost if a passenger does not occupy a seat at i . This cost depends on the probability of occupying a folding seat state or not at station i , or find a seat at a downstream station $i+1$.

The service leg cost formation is outlined hereby:

Algorithm FS_INV_COST:

Input:

t_{za} : the physical time of the arcs of the service z

χ_{za}^r : the discomfort factor related to state r

$[p_{zi}^{o,r}]_{i \in N_z}$: the probability of occupying comfort state r at i

$[p_{zi}^{+,r}]_{i \in N_z}$: the probability of occupying comfort state r at i

$[\delta_{zi}^f]_{i \in N_z}$: the state of folding seats at i

Output:

$\bar{t}_{z(i,s)}$: the average travel time by station couple $(i,s) : s < i$, at i for the service z

$\bar{g}_{z(i,s)}$: the average generalized time of the service leg between a station couple $(i,s) : s < i$

Main:

For each egress station $s < i$ in a reverse topological order of the service:

- $\bar{t}_{z(i,s)} = \bar{t}_{z(i,j)} + t_{z(j,s)}$, the physical time for $a \approx (i,j)$
- $\bar{g}_{z(i,s)}^r = \bar{g}_{z(i,j)}^r + \bar{g}_{z(j,s)}^r$, the leg time for comfort state r (seated)
- $\bar{\gamma}_{z(i,s)}^r = t_{z(i,i+1)} \chi_{z(i,i+1)}^r + (1 - p_{zi+1}^{o,r}) [(p_{zi+1}^{o,\tilde{r}} \cdot \tilde{\gamma}_{z(i+1,s)}^r + (1 - p_{zi+1}^{o,\tilde{r}}) \cdot \bar{\gamma}_{z(i+1,s)}^r)]$, the auxiliary variable if standing at i ,
- $\tilde{\gamma}_{z(i,s)}^r = t_{z(i,i+1)} \chi_{z(i,i+1)}^{\tilde{r}} + (1 - p_{zi+1}^{o,r}) [(\delta_{i+1}^f \cdot \tilde{\gamma}_{z(i+1,s)}^r + (1 - \delta_{i+1}^f) \cdot \bar{\gamma}_{z(i+1,s)}^r)]$, the auxiliary variable if occupying a folding seat at i ,

- $\gamma_{z(i,s)} = p_{zi}^{+,\tilde{r}} \cdot \gamma_{z(i,s)}^{\tilde{r}} + (1 - p_{zi}^{+,\tilde{r}}) \cdot \gamma_{z(i,s)}^{\bar{r}} + p_{zi+1}^{o,r} \cdot g_{z(i+1,s)}^r$, the auxiliary variable if not seated at i , and

- $\bar{g}_{z(i,s)} = p_{zi}^+ g_{z(i,s)}^r + (1 - p_{zi}^+) \cdot \gamma_{z(i,s)}$, the average generalized time of the service leg

END

Conclusion

The algorithms presented in the annex address the inclusion of the folding seats and their activation mechanism in the calculation of the in-vehicle comfort. The added complexity is not significant and the general mathematic framework remains valid.

ANNEX C:

Research Activity

Working Documents

- Combes F., Benezech V., Chandakas E., Leurent F., Samadzad M. and Windisch E. (2011), *Revue Critique du modèle Antonin2*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est (in french)
- Poulhes A. and Chandakas E. (2011), *Simulateur de Flux de Voyageurs dans les Transports Collectifs Urbains, Document de Conception*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est (in french)
- Chandakas E. (2012a), *Note sur l'Ajustement des données de temps de trajet du réseau DRIEA 2008*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est (in french)
- Chandakas E. (2012b), *Note sur la Capacité du Matériel Roulant et son Affectation sur le Réseau Francilien*, Working Document, Ecole Nationale des Ponts et Chaussées, Université Paris Est (in french)

Paper

- Leurent F., Chandakas E. and Poulhes A. (2012b), *Bi-layer equilibrium of user and service in a capacitated model of traffic assignment to a transit network*. Working paper submitted for publication to European Transport

Book Chapter

- Leurent F. and Chandakas E. (2012), *Capacité des véhicules, attente à quai et choix d'itinéraire en transport collectif urbain – modélisation pour la planification du réseau TC*, In Hégron G. and Prévost T (eds) *Modélisation Urbaine : de la Représentation au Projet*. Collection Références du Commissariat Général au développement durable, pp 192-195 (in french)
- Chandakas E., Leurent F. and Poulhes A. (2013b), *Un modèle de trafic de voyageurs en transports collectifs, sensible aux contraintes de capacité*, In Peuportier B. (ed) *Eco-conception des ensembles bâtis et des infrastructures*, pp 167-188 (in french)

Participation in Conferences with Proceedings

- Leurent F., Chandakas E. and Poulhes A. (2011), *User and service equilibrium in a structural model of traffic assignment to a transit network*, Paper presented at the XIV Euro Working Group on Transportation & 26th Mini Euro Conference & 1st European Scientific Conference on Air Transportation, Poznan, Pologne, September 2011
- Leurent F., Chandakas E. and Poulhes A. (2012), *Intégrer des contraintes de capacité dans un modèle d'affectation du trafic de voyageurs sur un réseau de transports collectifs*, Paper presented at the ATEC ITS France Conference, Versailles, France, February 2012 (in french)
- Leurent F., Chandakas E. and Poulhes A. (2012), *A passenger traffic assignment model with capacity constraints for transit networks*, Paper presented at the XV Euro Working Group on Transportation, Paris, France, September 2012
- Leurent F. and Chandakas E. (2012) *The transit bottleneck model*, Paper presented at the XV Euro Working Group on Transportation, Paris, France, September 2012
- Leurent F. and Chandakas E. (2013), *The transit bottleneck model and its application to a transit line in Paris*, Paper presented at the Transportation Research Board 92rd Annual Meeting, Washington D.C., January 2013

- Chandakas E. and Leurent F. (2013), *A Transit Station Model for Passenger Traffic Assignment with Capacity Constraints for a Transit Network*, Paper #1993 presented in the 13th WCTR Conference, 2013, Rio de Janeiro
- Chandakas E., Leurent F., Poulhes A. (2013a), *A Transit Assignment Model with Capacity Constraints: A Large-Scale Implementation*, Paper #1440 presented in the 13th WCTR Conference, 2013, Rio de Janeiro
- Chandakas E., Leurent F., Poulhès A. (2014), *Les contraintes de capacité dans les trafics de voyageurs et de véhicules en transport collectif : un modèle de simulation et son application au Grand Paris*, Paper presented at the ATEC ITS France Conference, Paris, France, January 2014 (in french)

Participation in Conferences without Proceedings

- Leurent F., Chandakas E. (2011), *Sur la contrainte de capacité dans les véhicules de transport collectif*, 1st Conference of the Chair “The socio-economic and modelling of urban public transport” : « Contraintes de capacité et congestion dans les réseaux de transport en commun », Marne-la-Vallée, France, January 2011
- Leurent F., Chandakas E. (2012), *Modélisation intégrée des effets capacitifs dans les systèmes de transport en commun : le modèle CapTA*, 2nd Conference of the Chair “The socio-economic and modelling of urban public transport”: « Contraintes de capacité dans les réseaux de transport en commun : Modélisation et conception », Marne-la-Vallée, France, Mars 2012
- Leurent F., Chandakas E. and Poulhes A. (2012), *A capacity constrained transit assignment model and its application to the Paris network*, International Conference on Reliability and Statistics in Transportation and Communication (RelStat’12), Riga, Latvia, October 2012