JOURNAL OF ADVANCED TRANSPORTATION
J. Adv. Transp. 2012; **46**:112–138
Published online 15 October 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/atr.146

On seat capacity in traffic assignment to a transit network

Fabien Leurent*

Université Paris-Est, Laboratoire Ville Mobilité Transport, Ecole des Ponts ParisTech, 6-8 avenue Blaise Pascal, Champs sur Marne 77455 Marne la Vallée Cédex 2, France

SUMMARY

Seating or standing make distinct on-board states to a transit rider, yielding distinct discomfort costs, with potential influence on the passenger route choice onto the transit network. The paper provides a transit assignment model that captures the seating capacity and its occupancy along any transit route. The main assumptions pertain to: the seat capacity by service route, selfish user behaviour, a seat allocation process with priority rules among the riders, according to their prior state either on-board or at boarding. To each transit leg from access to egress station is associated a set of 'service modes', among which the riders are assigned in a probabilistic way, conditionally on their priority status and the ratio between the available capacity and the flow of them. Thus the leg cost is a random variable, with mean value to be included in the trip disutility. Computationally efficient algorithms are provided for, respectively, loading the leg flows and evaluating the leg costs along a transit line. At the network level, a hyperpath formulation is provided for supply-demand equilibrium, together with a property of existence and an method of successive averages equilibration algorithm. It is shown that multiple equilibria may arise. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS: transit assignment; sitting behaviour; seated capacity; capacitated assignment; priority rules; line algorithms; route choice; network equilibrium

1. INTRODUCTION

1.1. Setting and literature review

The planning of urban public passenger transport often requires considering the capacity constraints and congestion effects. In the past recent years, transit congestion has been given an increased attention from the research community in the science of transportation and traffic. Based on the Traffic Capacity and Quality of Service Manual (TCQSM [1]), we have identified seven types of congestion effects: (i) the vehicle traffic capacity of the infrastructure; (ii) the operating capacity on a transit route; (iii) vehicle capacity; (iv) the rider capacity of a transit service or route; (v) the passenger capacity of a station; (vi) the transit vehicle capacity of a station; (vii) the access and parking capacity for private vehicles at a station. One type of effect may take several forms: in particular, vehicle capacity can be broken down into seated capacity, rider capacity or boarding and alighting capacity at the doors. Figure 1 depicts the interplay of system components and capacity effects.

Type (i): the vehicle traffic capacity of the infrastructure has been addressed by Spiess and Florian [2] who related the travel time (or cost) T_a of link a to the flow volume x_a by means of a travel time function, $T_a = t_a(x_a)$. This formulation is appropriate for mild congestion, i.e. under capacity, either for the circulation of vehicles on the infrastructure by linking the journey time (cost) to the vehicle flow, or for the travel of riders by linking a discomfort cost to the number of riders on each vehicle.

*Correspondence to: Fabien Leurent, Université Paris-Est, Laboratoire Ville Mobilité Transport, Ecole des Ponts ParisTech, 6-8 avenue Blaise Pascal, Champs sur Marne 77455 Marne la Vallée Cédex 2, France.
E-mail: fabien.leurent@enpc.fr

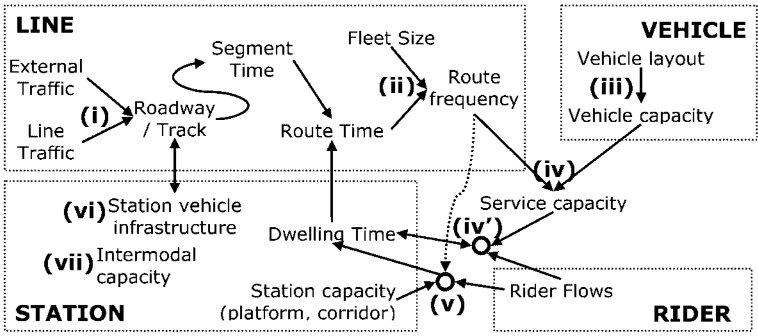


Figure 1. The capacity effects to be modeled in transit assignment.

Type (ii) concerns the operating capacity on a transit route which is limited by the vehicle fleet. Lam et al. [3]. linked the dwelling time of a transit vehicle at a loading platform to boarding and alighting passenger flows: adding the dwell times and the segment run times yields the vehicle journey time along the transit route. The ratio between the fleet size and the journey time determines the line frequency.

Type (iii); the rider capacity of a transit vehicle, is equivalent to Type (iv) in a static model with fixed line frequency and homogeneous vehicle fleet. Vehicle capacity per se stands as the main capacity effect considered in dynamic assignment models. Tong and Wong [4] have proposed a dynamic model with gradual filling of the vehicle but no congestion effect: each rider selects a best route based on the scheduled service times. Nuzzolo et al. (2001) [5] modeled rider discomfort in a crowded vehicle using a discomfort cost which increases with vehicle load. This cost influences the passengers' route choice, and hence passenger flows on the network. But there is no capacity limit and congestion does not affect the time a rider waits for a vehicle. Nguyen et al. [6] have modeled a waiting cost that is interpreted as waiting discomfort based on the difference between boarding passenger volumes and residual vehicle capacity, but which has no influence on physical time. Poon et al. [7] have modeled the boarding capacity limit in a vehicle using a bottleneck model which determines queue time on the basis of the temporal profiles of rider arrivals and departures—with departures being linked to the available vehicle capacity at passage times. The bottleneck is thus freed incrementally by the residual capacities of the vehicles after their passengers have alighted at a transit stop. Meschini et al. [8] have proposed a similar model for a multimodal network.

Type (iv), the rider capacity of a transit service, splits into passenger overall capacity and seat capacity. Passenger overall capacity has been the main capacity effect addressed in static assignment along four alternatives approaches as follows: effective frequency, constraint penalization, failure-to-board and user preference set.

Firstly, De Cea and Fernandez [9] linked the waiting time at boarding to the two flows of on-board and incoming riders, respectively, through a function of 'effective frequency' that yields the reciprocal of the waiting time. Cominetti and Corea [10] grounded this approach on a Markov chain model. Cepeda et al. [11] provided a network assignment model in terms of optimal routing strategy and hyperpaths.

Secondly Lam et al. [12], explicitly constrained the person capacity by line segment: a dual, nonnegative variable is associated to the constraint and takes on a positive value when the flow saturates the capacity, so as to penalize that path and divert the excess flow of incoming riders to the alternative paths. However, the on-board riders are penalized in the same amount as the incoming ones; another drawback is that full capacity is allocated to incoming riders at each station along a line, neglecting the number of on-board passengers at that station.

Thirdly, Kurauchi et al. [13] and Bell and coworkers [14] modelled the probability that an incoming passenger would succeed to board as the ratio between the available capacity and the incoming flow, capped at 1. The complement to one is the failure-to-board probability, which is applied to divert the excess flow to a fictitious 'spill arc'. This diversion, which lacks realism in the static model, is better addressed as a local delay in the dynamic version of the model by Schmöcker et al. [15].

Fourthly Hamdouch et al. [16], evaluated the probability of success-to-board on an attractive line at a station node in the order of user preferences: the available capacity of a route is divided by the residual incoming flow that could not be accommodated by the more attractive lines. Chaining the

DOI: 10.1002/atr

conditional probabilities of failure until success yields a routing probability by attractive line, which is used in a hyperpath framework at the network level. This approach suffers from two drawbacks: the line combination stems from capacity constraints only not from the respective line frequencies; furthermore a passenger has no option to wait for a line to become available to him: this flaw has been corrected in the dynamic version of the model by Hamdouch and Lawphongpanich [17].

In a dynamic setting Tian et al. [18], modelled passengers' interactions along a transit line with egress only at the line terminal (ideally the city centre): the mean travel cost by line segment depends on the number of passengers in the vehicle; every passenger may choose his/her departure time and thereby the vehicle in the service sequence. Tian et al. also considered the seating capacity in a restricted way in which all seats would be occupied from the line origin.

Leurent [19] developed a model of seating capacity along a transit line, with access and egress of passengers at every station and priority rules among the passengers for obtaining a seat: standing passengers on-board have priority over boarding passengers and with an equal probability of obtaining a seat within each group. Assuming that the cost of standing is higher than that of seating by unit of in-vehicle time, the segment cost from the boarding to descending station is a random variable with structural dependency on the seating capacity and the origin–destination (OD) (by segment) matrix of passenger flows. This model of seating capacity was extended to a general network in a static equilibrium framework by Leurent [20], and demonstrated on the Paris area by Leurent and Liu [21].

In a dynamic equilibrium framework Sumalee et al. [22], modelled jointly: the congestion at a stop which results in waiting time prior to boarding, the in-vehicle capacity that causes discomfort to standing riders and the seat capacity, by assuming priority rules of riders with former access on those with later boarding, and greater impetus to get a seat depending on the remaining journey length and the time already spent on-board.

Type (v), the passenger capacity of a station may involve the circulation of pedestrians or the storage of pedestrians at waiting. We have not identified any macroscopic assignment models that deal with these aspects, even though the pedestrian flows in question are just as great, if not greater, than in services. Chapter 4 of Part 7 of the TCQSM outlines a manual method inspired by an assignment model. Also, microscopic simulation models have been developed recently to simulate pedestrians in a complex station but the station is not linked to the rest of the network (e.g. Ref. [23]).

Type (vi), the movement and storage of vehicles in a station: this involves the capacity of the platforms for the passenger interface, the capacity of the corridors leading to the platforms and the vehicle storage capacity in addition to these services. Hamdouch and Lawphongpanich [17] suggested to model a walk arc as a pedestrian bottleneck.

Type (vii), interface capacities with personal modes, in other words, the transit station's capacity for personal vehicles, for riders who combine transit use with use of a two- or four-wheeled personal transport mode. This type includes (a) the road traffic capacity of the station access roads and (b) private vehicle parking capacity. These aspects are briefly described in the TCQSM (Part 7) but have not been addressed in macroscopic transit models.

1.2. Objectives and contribution

The paper has a core objective and a companion objective. The core objective is to provide a basic model at the line level for seat capacity, in which seat capacity and occupancy, comfort states and costs, sitting behaviour and priority rules are captured in an explicit way. The core model addresses a transit line in two steps: first, the *problem of line flow loading* consists in assigning the access–egress trip matrix to the vehicles and the seats; second, the *problem of line leg costing* yields the average cost by transit leg, i.e. by line section from access to egress station. The line flow loading problem is basically an assignment sub-model, whereas the line leg costing problem amounts to a cost–flow relationship at the line level.

The companion objective is to embed the line model in the framework of traffic assignment to a transit network: then passenger comfort can be taken into consideration in route choice. Static assignment is addressed in order to demonstrate that the line treatment does not interfere with the problem of common lines.

The paper makes available in English the line model introduced in a French paper [19] which captures the priority rules in a static setting and provides efficient algorithms to deal with the

DOI: 10.1002/atr

probabilistic features of leg costs. It also provides a static network assignment model of Leurent [20], in a more elaborate way. The static framework is simpler than the dynamic model of Sumalee et al.[22], which captures more traffic phenomena. It can be applied efficiently to large size problems, as demonstrated by Leurent and Liu [21].

1.3. Approach

In the line model for seat capacity, the approach is to identify the residual seat capacity at any stage along the line, and to share it amongst standing riders under priority rules. Two priority rules are assumed in the basic case: first, that standing riders with same level of priority have equal chance of getting a vacant seat; and second, that standing passengers going through a transit stop obtain access to vacant seats prior to riders boarding at that stop. The two priority rules induce the probability to sit either from the previous segment or at boarding, so by successive segments along the transit leg a probability results for each 'service mode' made up of a sequence of m segments at standing followed by n–m segments at seating in a leg of n segments. At the leg level, the service mode is obtained in a somewhat random way (due to first rule), and yields a random leg cost, with distribution characterized by both the segment-state costs and the service mode probabilities. To the leg user, the leg-generalized cost is a random variable that stems from the cost of the service mode which is obtained at random.

Efficient 'line algorithms' are provided to solve the line flow loading problem and the line leg costing problem: the treatment is macroscopic and may be embedded in a transit assignment model, static by line service or dynamic by vehicle service. For simplicity, we have integrated line seat capacity in the macroscopic, static model of hyperpath-based transit assignment of Spiess and Florian [2]. Each transit leg is represented by a network arc: this preserves the basic algorithms to search for a minimal-cost hyperpath and to assign OD flows onto a hyperpath.

1.4. Paper structure

The body of the paper is organized into six sections. Section 2 sets up the core model of a line, with assumptions on comfort states, travel behaviour and priority rules; then the service modes are defined and the leg cost issue is addressed. Section 3 addresses the line-based problems to derive, respectively, the sitting probabilities at line stations by priority status, the service mode probabilities and costs, as well as the mean and variance of the leg cost; efficient algorithms are provided together with a numerical illustration.

In Section 4, the line model is embedded in a network assignment model: each line leg is represented by one network arc. The cost–flow relationship is defined at the arc level for nontransit arcs, and at the line level for leg arcs on the basis of an underlying cost–flow relationship by line segment and comfort state.

Section 5 provides a mathematical analysis of traffic equilibrium in the transit model with seat capacity. The traffic equilibrium is stated as the solution of a nonlinear complementarity problem (NCP) and characterized as the solution to a variational inequality problem (VIP). Based on a regularized cost–flow relationship it is shown that an equilibrium state must exist. A standard method of successive averages (MSA) is provided as solution algorithm. A method is provided to evaluate the duality gap in a simple manner, thus yielding a rigorous criterion for convergence.

In Section 6, a classroom example is designed and dealt with in a parametric analysis to demonstrate that multiple equilibria may arise. Lastly, Section 7 concludes by synthesizing the main outcomes, stating the model outreach and limitations and pointing to potential developments.

2. THE LINE MODEL: ASSUMPTIONS AND BASIC NOTIONS

Here the focus is on a given transit line. Our modelling assumptions pertain to comfort states and their cost (Section 2.1), and also travel behaviour and priority rules across users (Section 2.2). These enable us to define the notion of a service mode (Section 2.3), which is a way of using a line 'leg' from access station to egress station. As the users compete for the residual seat capacity, there is randomness in getting a seat, which makes the leg cost a random variable (Section 2.4). To sum up, the outputs of the line model consist in sitting probabilities and leg costs.

DOI: 10.1002/atr

Table I. Notation for line problems.

S_ℓ	Number of stations along transit line ℓ
q_{ij}^ℓ	Leg flow from station i to station j along ℓ
κ_ℓ	Seat capacity of the line during the assignment period
$\delta \in \{o, +\}$	Stage prior/posterior to the boarding of incoming riders
κ_i^δ	Residual capacity at station i and stage δ
y_i^δ	Flow of riders candidate to get a seat at station i and stage δ
p_i^0 (resp. p_i^+)	Probability to get a seat at i for on-board (resp. boarding) standing riders
π_{ij}^m	Probability of service mode m from i to j
$\underline{x}$ (resp. $\overline{x}$)	A seated (resp. standing) flow
$x_j^{(i)\delta}$	Flow destined to egress station j on segment $(i, i+1)$ at stage δ
$x_{\geq i}^\delta$	Flow on segment $(i, i+1)$ at stage δ to all egress stations $j > i$
$\underline{c}_a$ (resp. $\overline{c}_a$)	Travel cost at seating (resp. standing) on segment a
c_{ij}^m	Cost from i to j of service mode m (i.e. standing from i to $i+m$ then seating)
c_{ij} (resp. $\hat{c}_{ij}$)	Random cost from i to j (resp. average cost)
γ_{ij}	Average cost from i to j conditional on standing on $(i, i+1)$
v_{ij} (resp.ϖ_{ij})	Variance of cost from i to j (resp. conditional on standing on $(i, i+1)$)

2.1. Comfort states and their cost

In a public transport service, it frequently occurs that several categories of places with distinct attributes are provided to the users. In urban transit by bus, tram, metro or train, the distinction between seating and standing is particularly significant: a seated rider is less inconvenienced by the vehicle's acceleration and deceleration and by the slopes and curves in the vehicle's trajectory; he may invest his travel time into a complementary activity such as reading, listening to music, relaxing or working; moreover, he is much less submitted to crowding.

From observation of traffic in public transit, it appears that riders favour the seating state over the standing state to a large extent: when a crowded vehicle arrives at a station, those riders that stand and stay on-board try to get a seat given away by an outgoing passenger. At the initial station of a crowded metro line, some users prefer to wait for the next train to arrive, rather than to board when there is no vacant seat. Furthermore, on some OD pairs serviced by alternative metro lines, many passengers prefer to use the line with less congestion, even if the travel time is higher.

Let us assume that on a transit line segment from a station to the next, there are two riding states namely seating and standing, with associated costs that reflect the users' preference to have a seat.

Let a denote a line segment, $\underline{c}_a$ its discomfort cost to a rider at seating, and $\overline{c}_a$ its discomfort cost to a rider at standing: the basic assumption is that $\underline{c}_a \leq \overline{c}_a$ whatever the traffic load, i.e. a rider prefers to be seated.

2.2. Travel behaviour and priority rules

It is assumed that every rider is a cost-minimizing individual decision-maker, striving to reduce his travel cost. Then a rider who is standing tries to get a seat as soon as one becomes available. As there is a limited number of available seats, say κ for capacity, and also a number x of riders that would like to sit, it may be the case where $\kappa < x$, meaning that capacity is less than demand.

In this case, only a proportion κ/x of riders may sit. The issue of which riders would get a seat is addressed here in a simple way, assuming that all of them have an equal probability to sit, i.e. neglecting the individual attributes of age and physical need, eagerness-to-sit, planned egress station, etc. In reality, long haul riders may be willing to get a seat more than short haul riders, who may prefer to stand as close to the doors as possible. Our model might be improved in that respect, specifically by weighting the user requests to get a seat by a coefficient based on the time of their leg, either physical time or generalized time including the discomfort of standing.

The time at which seats become available and competition occurs is important: standing riders that stay on board have an advantage over the incoming riders, which is modelled by assuming two successive competitions: the first one among 'through' riders, the other one among incoming riders.

DOI: 10.1002/atr

Thus, in our basic model of seat congestion, two priority rules are assumed: first, that standing riders with same level of priority have equal chance of getting a vacant seat; and second, that standing passengers going through a transit stop obtain access to vacant seats prior to riders boarding at that stop.

2.3. Services modes on a transit leg

On a 'leg' from an access station to an egress station along a transit line, the user is provided service in a particular way, depending on which comfort states he gets on each line segment in the leg.

Let us define a 'service mode' as the sequence of segment comfort states along the leg. In a leg made up of N segments, if there are two comfort states associated to each segment, there could be as much as 2^N service modes associated to the leg, but in fact the users' behaviour reduces the number of alternative service modes to $N+1$: after getting a seat a user is assumed not to release it until arrival at his egress station.

Thus a service mode is fully described by the station $i+m$ at which the user gets a seat, with index $m\in\{0,1,2,\ldots N\}$: by convention, getting a seat on exiting at $i+N$ means standing all over the leg.

The cost of service mode m from access station i to egress station $j=i+N$ is:

$$c_{ij}^{m}=\left[\sum\nolimits_{k=i}^{i+m-1}\overline{c}_{a\approx(k,k+1)}\right]+\left[\sum\nolimits_{k=i+m}^{j-1}\underline{c}_{a\approx(k,k+1)}\right] \tag{1}$$

Under given rider flows by leg along the transit line, the assumption that riders prefer to be seated determines the segment flow of standing riders, say $\overline{x}_a$ on segment a. In the network model of Section 4, it is assumed that value $\overline{c}_a=\overline{c}_a(\mathbf{x})$ is determined as a function $\overline{c}_a$ of the flow vector $\mathbf{x}$, which is a generalization.

The issue of whether a person succeeds in accessing the vehicle, i.e. the in-vehicle person capacity, is a complementary modelling issue that pertains to the platform stage prior to boarding. Thus the related models described in the literature review could be combined to the model of seating capacity with no redundancy nor conflict.

2.4. Leg cost as a random variable

The attribution of a given service mode does not depend solely on the user, because his preference to be seated rather than standing may get into competition with others' preferences, resulting in a collective allocation process of seats rather than in an individual choice of a service mode.

From the modelling assumptions on travel behaviour and priority rules, the seat allocation process at each station along a transit route can be summarized by two sitting probabilities: the first one say p_i^0 for through riders, and the other one say p_i^+ for incoming riders. Here the sitting probabilities are taken as exogenous, so as to derive the distribution of service modes and costs.

Figure 2 depicts the comfort states along a transit route: state transition may occur for a rider from standing aboard at $i-1$ to either seating from i with probability p_i^0 or standing on segment $(i,i+1)$ with probability $1-p_i^0$, or from incoming at i to either seating from i with probability p_i^+ or standing on $(i,i+1)$ with probability $1-p_i^+$.

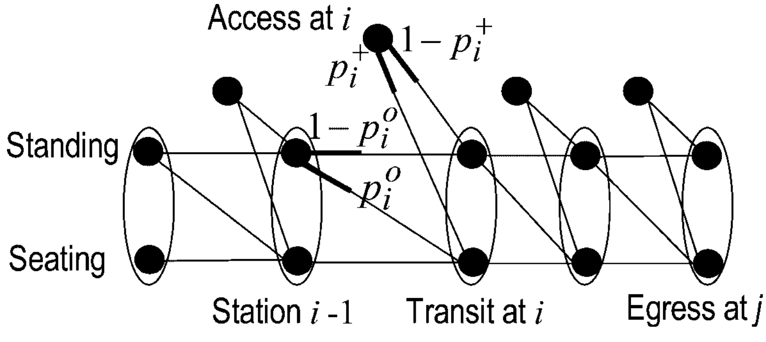


Figure 2. Rider states along a transit line.

The probability of keeping standing from i to $i+N$ is

$$\pi_{i,i+N}^{N}=\left(1-p_i^{+}\right)\left[\prod_{k=1}^{N-1}\left(1-p_{i+k}^{0}\right)\right] \tag{2a}$$

The probability of getting a seat at station $i+\ m$, meaning obtaining service mode m from station i, is:

$$\pi_{i,i+N}^{0}=p_i^{+} \text{ for } m=0 \tag{2b}$$

$$\pi_{i,i+N}^{m}=\left(1-p_i^{+}\right)\left[\prod_{k=1}^{m-1}\left(1-p_{i+k}^{0}\right)\right]p_{i+m}^{0} \text{ for } m\in\{1,2\cdots N-1\} \tag{2c}$$

The probabilities $(\pi_{i,i+N}^{m})_{m=0\cdots N}$ are associated with the service modes and describe the random process of getting a service mode. Service mode m has probability $\pi_{i,i+N}^{m}$ stated in (2) and cost $c_{i,i+N}^{m}$ stated in (1).

Thus, to the user the leg cost is a random variable $c_{i,i+N}$, depending on which service mode is obtained. The mean leg cost $\hat{c}_{i,i+N}\equiv E[c_{i,i+N}]$ is such that:

$$\hat{c}_{i,i+N}=\sum_{m=0}^{N}\pi_{i,i+N}^{m}c_{i,i+N}^{m} \tag{3a}$$

The variance of the leg cost, $v_{i,i+N}\equiv \mathrm{var}[c_{i,i+N}]$, is such that:

$$v_{i,i+N}=\sum_{m=0}^{N}\pi_{i,i+N}^{m}\left(c_{i,i+N}^{m}-\hat{c}_{i,i+N}\right)^2 \tag{3b}$$

The cost variance may be taken into account in the generalized cost of travel along the leg. However, in the network part of the paper the focus is on mean cost only, since cost variability in hyperpath assignment still makes an open issue, which deserves specific research efforts.

3. STATEMENT OF LINE PROBLEMS AND ALGORITHMS

Having provided the assumptions and derived the formulae for the major output variables in the line model, let us now provide efficient algorithms to solve the formulae—and to be included in procedures for network assignment. The algorithms have minimal complexity with respect to the number of distinct variables that are outputted; this efficiency is achieved through a technique of auxiliary variable.

The section roadmap is as follows: first, simple formulae are established for the sitting probabilities along the transit line (Section 3.1). Then, a line flow loading algorithm is provided to load an access–egress trip matrix onto the route segments according to the priority rules and to yield all sitting priorities; letting S denote the number of stops along the transit line, the algorithm has a complexity of $O(S^2)$ hence it is efficient (Section 3.2). Next, recursive formulae are established for the service mode probabilities and costs: the application algorithm has a complexity of $O(S^3)$ and is efficient (Section 3.3). Lastly, a line costing algorithm is provided to compute the mean and variance of the leg costs along the transit route, yielding a reduced computational complexity of $O(S^2)$ which makes it efficient for its reduced purpose (Section 3.4).

3.1. On sitting probabilities

On the route segment $a\approx(i,i+1)$ downstream of station i, let us describe the segment traffic by four flow vectors indexed by egress station j and stage $\delta\in\{0,+\}$ to deal with the sitting of the aboard riders prior to that of the boarding ones: $\underline{x}_j^{(i)\delta}$ (resp. $\overline{x}_j^{(i)\delta}$) is the segment flow of seating (resp. standing) riders destined to j and at stage δ. Let also $\underline{x}_{\geq i}^{\delta}=\sum_{j\geq i}\underline{x}_j^{(i)\delta}$ and $\overline{x}_{\geq i}^{\delta}=\sum_{j\geq i}\overline{x}_j^{(i)\delta}$ be the segment and stage flows of seated and standing riders, respectively. Notation $\{j\geq i\}$ means that station j lies downstream of station i along the line, while $\{j>i\}$ means strictly downstream.

The riders boarding at i are described by a flow vector $\mathbf{q}_i^{\ell}=[q_{ij}^{\ell}]_{j>i}$.

DOI: 10.1002/atr

A seat capacity of κ is assumed for the transit route during the assignment period: as the model is static this is the product of the in-vehicle capacity by the route frequency during the period. Let κ_i^- be the residual seat capacity from the previous station $i-1$: if $\overline{x}^+_{\geq i-1} > 0$ then $\kappa_i^- = 0$. Every seating rider who exits at i releases their seat, which increases the residual seat capacity to

$$\kappa_i^0 = \kappa_i^- + \underline{x}_i^{(i-1)+} \tag{4a}$$

This capacity may be used by the riders standing aboard that continue downstream of i, in number of $y_i^0 = \sum_{j>i} \overline{x}_j^{(i-1)+}$. Their probability to get a seat amounts to:

$$p_i^0 = \min\left\{1, \frac{\kappa_i^0}{y_i^0}\right\}, \text{ which is set to 1 if } y_i^0 = 0 \tag{4b}$$

After their eventual sitting, the residual capacity is decreased to:

$$\kappa_i^+ = \kappa_i^0 - \min\{y_i^0, \kappa_i^0\} \tag{4c}$$

It is available to the incoming riders in number of $y_i^+ = \sum_{j>i} q_{ij}^\ell$, who get a seat with the following probability:

$$p_i^+ = \min\left\{1, \frac{\kappa_i^+}{y_i^+}\right\}, \text{ which is set to 1 if } y_i^+ = 0 \tag{4d}$$

After their eventual sitting, the residual capacity is decreased to:

$$\kappa_{i+1}^- = \kappa_i^+ - \min\{y_i^+, \kappa_i^+\} \tag{4e}$$

Thus formulae (4a)–(4e) enable us to derive the residual capacities and the sitting probabilities at a given station, for riders either aboard or boarding, as functions of the vectors of seating and standing flows by egress station and the access–egress trip matrix.

3.2. *Line flow loading problem and algorithm*

The **line flow-loading problem** is to assign a line access–egress trip matrix to the seating and standing states along the route segments. The outputs consist basically in the sitting probabilities by station node i and priority status (aboard or boarding); the segment flows by stage, comfort state and exit station can also be outputted.

A solution method for the line loading problem has been provided above for a given station: the line loading algorithm consists in applying this method to every station along the line, in turn from the initial station to the final one. Aside from applying (4), an important issue is to obtain the vectors $\underline{x}^{(i)\delta} = \left[\underline{x}_j^{(i)\delta}\right]_{j>i}$ and $\overline{x}^{(i)\delta} = \left[\overline{x}_j^{(i)\delta}\right]_{j>i}$ of seating and standing flows at each stage along the line.

The **line loading algorithm** addresses a route ℓ with S_ℓ stations and seat capacity κ. Input variables also include the access–egress trip matrix $\left[q_{ij}^\ell\right]_{i<j\in\ell}$ and also, if required, the segment and stage flow by comfort state and egress station.

The algorithm is comprised of the following steps:

Initialization. Let $i := 0$; let $\underline{x}_j^{(0)+} := 0$ and $\overline{x}_j^{(0)+} := 0$ $\forall j \in \ell$; let $\underline{x}^+_{\geq 0} := 0$ and $\overline{x}^+_{\geq 0} := 0$.

Termination Test. If $i = S_\ell$ then terminate else let $i := i + 1$ and continue.

Progression. At station i:

- let first $\kappa_i^0 := \left(\kappa - \underline{x}^+_{\geq i-1} + \underline{x}_i^{(i-1)+}\right)^+$ then $y_i^0 := \overline{x}^+_{\geq i-1} - \overline{x}_i^{(i-1)+}$ and $p_i^0 := \min\{1, \kappa_i^0 / y_i^0\}$.
- let $\underline{x}_j^{(i)0} := \underline{x}_j^{(i-1)+} + p_i^0 \overline{x}_j^{(i-1)+}$ and $\overline{x}_j^{(i)0} := (1 - p_i^0)\overline{x}_j^{(i-1)+}$ $\forall j > i$.
- let $\underline{x}^0_{\geq i} := \underline{x}^+_{\geq i-1} - \underline{x}_i^{(i-1)+} + p_i^0 y_i^0$ and $\overline{x}^0_{\geq i} := (1 - p_i^0) y_i^0$.
- let $\kappa_i^+ := \left(\kappa_i^0 - p_i^0 y_i^0\right)^+$, then $y_i^+ := \sum_{j>i} q_{ij}^\ell$ and $p_i^+ := \min\{1, \kappa_i^+ / y_i^+\}$.

DOI: 10.1002/atr

- let $\underline{x}_j^{(i)+} := \underline{x}_j^{(i)0} + p_i^+ q_{ij}^\ell$ and $\overline{x}_j^{(i)+} := \overline{x}_j^{(i)0} + (1-p_i^+) q_{ij}^\ell \ \forall j > i$.
- let $\underline{x}_{\geq i}^+ := \underline{x}_{\geq i}^0 + p_i^+ y_i^+$ and $\overline{x}_{\geq i}^+ := \overline{x}_{\geq i}^0 + (1-p_i^+) y_i^+$.
- Go to *Termination Test*.

This algorithm may be streamlined further on, notably so by dropping the (i) and δ superscripts in the segment and stage flow variables: then the algorithm amounts to update, at each segment and stage, the working variables of segment flow by egress station. In its current form, the treatment of a current station requires a number of operations proportional to $S_\ell - i$, which induces a computational complexity of $O(S_\ell^2)$—a still modest number for typical transit routes. The complexity is minimal with respect to the number of distinct variables that can be outputted, since there are $O(S_\ell^2)$ segment and stage flows by current segment, comfort state and egress station.

A circular line would require a more involved treatment, which is outlined in Appendix A.

Instance 1. A transit line with 4 stations, seat capacity of 100 passengers per hour, given segment costs by comfort state and given leg trip matrix is depicted in Figure 3. The trip flows by route leg are as follows:

- At $i = 1$ there are $y_1^+ = 120$ boarding riders, among whom $q_{12} = 50$ are destined to station 2, $q_{13} = 30$ to station 3 and $q_{14} = 40$ to station 4.
- At $i = 2$ there are $y_2^+ = 90$ boarding riders, among whom $q_{23} = 60$ and $q_{24} = 30$.
- At $s = 3$ there are $y_3^+ = 50 = q_{34}$ boarding riders.

The line-loading algorithm yields the following selected results:

- At $i = 1$, $\kappa_1^0 = 100 = \kappa_1^+$, $p_1^0 = 1$, $y_1^+ = 120$ hence $p_1^+ = \frac{5}{6}$, $\underline{x}_{\geq 1}^+ = 100$ and $\overline{x}_{\geq 1}^+ = 20$, $\underline{x}_2^{(1)+} = 41.7$, $\underline{x}_3^{(1)+} = 25$, $\underline{x}_4^{(1)+} = 33.3$ whereas $\overline{x}_2^{(1)+} = 8.3$, $\overline{x}_3^{(1)+} = 5$ and $\overline{x}_4^{(1)+} = 6.7$.
- At $i = 2$, $\kappa_2^0 = 41.7$, $p_2^0 = 1$ hence all standing riders that go through get a seat, $\kappa_2^+ = 30$, $y_2^+ = 90$ hence $p_2^+ = 1/3$, $\underline{x}_{\geq 2}^+ = 100$ and $\overline{x}_{\geq 2}^+ = 60$, $\underline{x}_3^{(2)+} = 50$, $\underline{x}_4^{(2)+} = 50$ whereas, $\overline{x}_3^{(2)+} = 40$ and $\overline{x}_4^{(2)+} = 20$.
- At $i = 3$, $\kappa_3^0 = 50$, $p_3^0 = 1$ hence all standing riders that go through get a seat, $\kappa_3^+ = 30$, $y_3^+ = 50$ hence $p_3^+ = 3/5$, $\underline{x}_4^{(3)+} = 100$ and $\overline{x}_4^{(3)+} = 20$.
- At $i = 4$, all riders come out.

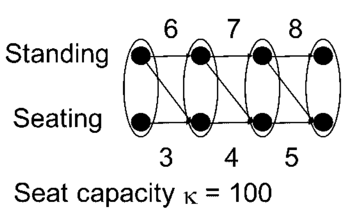
Instance 2. Let us adapt the trip matrix of Instance 1 by replacing its first line with $q_{1\bullet} = [150\ 30\ 140]$. The line loading algorithm yields sitting probabilities as follows:

$\mathbf{p}_i^+ = [0.3125\ 0\ 0\ 0]$ and $\mathbf{p}_i^o = [1\ 0.401\ 0.2013\ 0]$

3.3. *Cost and flow share of service modes*

The **service mode problem** consists in deriving the probability and cost by service mode, as stated in (1) and (2), from the sitting probabilities (taken here as inputs).

Let us first provide an algorithm to evaluate the flow share of all service modes along legs with egress at a given station j, by dealing with the access stations i in backward order from downstream to upstream. We shall use a sequence of auxiliary variables $\left[\rho_k^m : k \leq j, m = 0, 1, \ldots, j-k\right]$ in order to save computational effort.



Access \ Egress	2	3	4
1	50	30	40
2		60	30
3			50

Figure 3. Segment costs by comfort state and line trip matrix.

DOI: 10.1002/atr

The service mode algorithm is made up of the following steps:

Initialization. Let $\rho_j^0 := 1$ and $i := j$.

Progression. Let $i := i-1$. If $i = 0$ then terminate, else do:

- let $\rho_i^0 := p_i^o$.
- let $\rho_i^m := (1-p_i^o)\rho_{i+1}^{m-1}$ for $m \in \{1, 2, \ldots, j-i\}$.
- Go to *Progression*.

Then $\pi_{ij}^m = p_i^+$ if $m = 0$, or $(1-p_i^+)\rho_{i+1}^{m-1}$ if $m \geq 1$.

The computation of service mode costs is still easier, on the basis of:

$$c_{i,j}^m = \overline{c}_{i,i+1} + c_{i+1,j}^{m-1} \text{ for } m \geq 1, i < j$$

$$c_{i,j}^0 = \underline{c}_{i,i+1} + c_{i+1,j}^0 \text{ if } i < j$$

The computational complexity of both algorithms is in $O\left(S_\ell^3\right)$ since it amounts to the number of egress stations, times the number of access stations, times the number of service modes from access to egress. This complexity is minimal since there are $O\left(S_\ell^3\right)$ output variables π_{ij}^m: thus the service mode algorithm is efficient.

Instance 2 (continued). Let us evaluate the cost and flow share of the service modes for the trips destined to station 4. By entry station i and service mode m, the costs and flow shares are, respectively:

$$[C_{i4}^m] = \begin{bmatrix} 12 & 15 & 18 & 21 \\ 9 & 12 & 15 & \\ 5 & 8 & & \end{bmatrix} \text{ and } [\pi_{i4}^m] = \begin{bmatrix} 0.3125 & 0.2757 & 0.0829 & 0.3289 \\ 0 & 0.2013 & 0.7987 & \\ 0 & 1 & & \end{bmatrix}$$

3.4. Line leg costing algorithm

To save even more on computational effort, let us now provide an algorithm to evaluate the mean and variance of leg cost. Auxiliary variables are still useful to average the cost of the downstream sub-path conditional on the upstream state of comfort.

To obtain the mean leg cost $\hat{c}_{ij}$ from access station i to egress station j, let us associate two costs to that leg, namely:

- the seating cost $c_{ij}^0 = \sum_{k=i}^{j-1} \underline{c}_{a\approx(k,k+1)}$,
- an auxiliary mean cost γ_{ij} which is a mean cost from i to j conditional on standing on segment $(i, i+1)$.

These costs satisfy the following recursive equations, due to the law of total probability:

$$c_{i,j}^0 = \underline{c}_{a\approx(i,i+1)} + c_{i+1,j}^0 \tag{5a}$$

$$\gamma_{i,j} = \overline{c}_{a\approx(i,i+1)} + p_{i+1}^0 c_{i+1,j}^0 + \left(1-p_{i+1}^0\right)\gamma_{i+1,j} \tag{5b}$$

$$\hat{c}_{i,j} = p_i^+ c_{i,j}^0 + \left(1-p_i^+\right)\gamma_{i,j} \tag{5c}$$

The line costing algorithm addresses a given egress station j by dealing with the access stations $i \leq j$ in backward order from downstream to upstream. The initial conditions are $c_{i,j}^0 = 0$ and $\gamma_{i,j} = 0$ at $i = j$. As the treatment of i as an access station amounts to computing six additions and four products, the mean costing of all legs with egress at j has a computational complexity of $O(j)$, making the mean costing of all legs along the line a $O\left(S_\ell^2\right)$ burden. Thus the mean leg-costing algorithm is of minimal complexity, since there are $O\left(S_\ell^2\right)$ output variables $\hat{c}_{ij}$.

The variance of the leg cost can be obtained in a similar way, on the basis of an auxiliary variance $\varpi_{i,j}$ defined as the variance of the leg cost from i to j conditional on standing on segment $(i, i+1)$. The

DOI: 10.1002/atr

conditional variance $\varpi_{i,j}$ and the variance v_{ij} satisfy the following recursive formulae:

$$\varpi_{i,j} = \left(1-p^0_{i+1}\right)\left[\varpi_{i+1,j} + p^0_{i+1}\left(\gamma_{i+1,j}-c^0_{i+1,j}\right)^2\right] \tag{6a}$$

$$v_{ij} = \left(1-p^+_i\right)[\varpi_{i,j} + p^+_i\left(\gamma_{i,j}-c^0_{i,j}\right)^2] \tag{6b}$$

Equation (6b) stems from the theorem of total variance: a leg from i to j is either a leg beginning at standing or an all-seating leg, with respective probabilities $1-p$ and p. The interclass variance of the leg cost amounts to

$p(1-p)(\Delta C)^2$

with $\Delta C = \gamma - c^0$ the difference in class costs. Within the class of all-seating trips the cost variance is zero. Within the class of legs that begin at standing, the cost variance is $\varpi_{i,j}$ by definition. To sum up, the intraclass variance amounts to $(1-p)\varpi_{i,j}$.

A similar proof applies to (6a) by considering that a leg from i to j that begins at standing has its second segment either at standing or at seating.

Instance 1 (continued). Let us apply the line costing algorithm to station 4 as egress, on assuming the following segment costs: $\underline{c}_{12} = 3$, $\overline{c}_{12} = 6$, $\underline{c}_{23} = 4$, $\overline{c}_{23} = 7$, $\underline{c}_{34} = 5$ and $\overline{c}_{34} = 8$:

At $s = 4$, $c^0_{44} = \gamma_{44} = 0 = \hat{c}_{44}$ and $v_{44} = \varpi_{44} = 0$.

At $s = 3$, $c^0_{34} = 5$, $\gamma_{34} = 8$, $\hat{c}_{34} = 6.2$, $\varpi_{34} = 0$ and $v_{34} = 2.16$.

At $s = 2$, $c^0_{24} = 9$, $\gamma_{24} = 12$, $\hat{c}_{24} = 11$, $\varpi_{24} = 0$ (all riders sit at 3), $v_{24} = 2$.

At $s = 1$, $c^0_{14} = 12$, $\gamma_{14} = 15$, $\hat{c}_{14} = 12.5$, $\varpi_{14} = 0$ (all riders sit at 2), $v_{14} = 1.25$.

In this example, the cost variance takes on moderate values; it increases with respect to the leg number of segments in a less than proportional way since the more segments in a leg, the more opportunities to get a seat.

Instance 2 (continued). Let us apply the line-costing algorithm to the modified trip matrix, hence to the modified sitting probabilities. By entry station i from 1 to 3 and exit station j from 2 to 4, the mean costs and cost variances are, respectively:

$$[\hat{C}_{ij}] = \begin{bmatrix} 5.06 & 10.30 & 16.28 \\ & 7 & 14.40 \\ & & 8 \end{bmatrix} \text{ and } [v_{ij}] = \begin{bmatrix} 1.03 & 6.43 & 13.75 \\ & 0 & 1.44 \\ & & 0 \end{bmatrix}$$

Under increased entry-exit flows it turns out that the cost variance may take much larger values.

4. NETWORK REPRESENTATION AND COST–FLOW RELATIONSHIP

Let us now turn to our second objective: to combine seat capacity and passenger comfort states with route choice in a model of traffic assignment to a transit network. This involves the following three issues: firstly, to link the stations of line access and egress to the network paths from origin to destination nodes; secondly, to represent the line legs as network sub-paths and include the leg average cost in the path cost; thirdly, to determine the line matrix of leg flows on the basis of the OD flows and their network routes.

To address these issues, we shall first provide a network representation that accommodates transit routes on the basis of one arc per leg of entry–exit stations, i.e. the leg-as-arc format. Then, we shall define the model variables, from the variables of flow and cost by network element to the path and hyperpath variables. Lastly, we shall define the cost–flow relationship in accordance with the line model and provide a continuity property.

The hyperpath setting has been the standard framework for static models of traffic assignment to a transit network since the fundamental contributions of Spiess [24], Spiess and Florian [2] and Nguyen and Pallotino [25]. Loosely speaking, a transit hyperpath is a bundle of paths from a set of origin nodes to a single destination node, with routing options at nodes of access to transit lines, and associated routing proportions that stem from the lines' cost and service frequency. The set of hyperpaths may be

DOI: 10.1002/atr

subject to restrictions, of which the main instance is a bundle of paths that comply with a given sequence of transit stations for access, transfer and egress: this is the 'optimal route' model of De Cea and Fernandez [26].

Table II. Notation for network problems.

N	Set of nodes n
A	Set of nodes $a \approx (n_a^+, n_a^-)$ with tail node n_a^+ and head node n_a^- in N
A_n^+	$\equiv \{a : n_a^+ = n\}$ subset of arcs tailed at n
A_n^-	$\equiv \{a : n_a^- = n\}$ subset of arcs headed to n
$n_{\ell,i}^+$ (resp. $n_{\ell,i}^-$)	Access (resp. Egress) node at station i along line ℓ
A_ℓ	Arc subset of line legs $a \approx (n_{\ell,i}^+, n_{\ell,j}^-)$ with $i < j$
c_a	Travel cost along arc a
f_a	Frequency associated to arc a
κ_ℓ	Seat capacity of line ℓ during the assignment period
f_ℓ	Service frequency of line ℓ by unit time during the assignment period
S	Set of destination nodes s
W^s	Set of origin–destination pairs destined to s
q_{ns}	OD flow from n to s
$\mathbf{x}_{AS}$	Flow vector by arc and destination
$\mathbf{x}_A$	Flow vector by arc
$h = (\hbar, \hat{h})$	Hyperpath with arc set $\hbar$ and routing field $\hat{h}$
$\hat{h}_r$	Proportion of hyperpath flow on path r along h
$R_{ns}(h)$	Set of elementary paths r along hyperpath h from n to s
F_m^h	Combined frequency at node m for the boarding arcs in $A_m^+ \cap h$
w_m^h	Waiting delay at node m along h
$C_{ns}(h, \mathbf{x}_{AS})$	Travel cost from n to s along h with respect to flow state $\mathbf{x}_{AS}$
$\mathbf{X}_{NS}$	$\equiv [q_{ns}^h]_{n,s,h}$ Hyperpath flow vector
μ_{ns}	Dual variable of minimum cost from n to s
$\chi_{NS}(\mathbf{X}_{NS})$	Vector function of hyperpath costs by OD pair
$\chi_{N+A}(\mathbf{Z}_{NS} : \mathbf{X}_{NS})$	$\equiv \mathbf{Z}_{NS} \cdot \chi_{NS}(\mathbf{X}_{NS})$ with node-based part χ_N and arc-based part χ_A

4.1. Network representation

In a traffic assignment model, the network representation is purported to describe the routes that are available to the network users in terms of path topology and of physical and economic attributes such as travel time and money cost. It is also purported to model the users' route choices and to aggregate the resulting chosen paths into path flows and link flows—in other words the traffic loads.

Let us consider a network $G = [N, A]$ that is comprised of a set N of nodes n and a set A of arcs $a \approx (m, n)$ with tail node $n_a^+ = m$ and head node $n_a^- = n$ in N. The subset of the arcs that go out (resp. come in) node n is denoted as A_n^+ (resp. A_n^-).

A transit line ℓ in L, the set of lines (or transit routes), has S_ℓ stations i numbered from 1 to S_ℓ and $S_\ell - 1$ segments $(i, i+1)$. A line station is modelled by two nodes: $n_{\ell,i}^+$ for access and $n_{\ell,i}^-$ for egress, except at the line endpoints where only access or egress is permitted. The set of network nodes required to model the line is $N_\ell = N_\ell^+ \cup N_\ell^-$ with $N_\ell^+ = \{n_{\ell,i}^+ : i = 1 \cdots S_\ell - 1\}$ the subset of access nodes and $N_\ell^- = \{n_{\ell,i}^- : i = 2 \cdots S_\ell\}$ that of egress nodes. The line segments are not represented by network arcs but give rise to access–egress pairs of stations, (i, j) with $1 \leq i < j \leq S_\ell$, each of which is represented by a leg arc $a \approx (n_{\ell,i}^+, n_{\ell,j}^-)$. Let $A_\ell = \{a \approx (n_{\ell,i}^+, n_{\ell,j}^-) : 1 \leq i < j \leq S_\ell\}$ denote the line subset of leg arcs, in number of $|A_\ell| = \frac{1}{2}(S_\ell - 1)(S_\ell - 2)$. Thus line ℓ is represented by $2(S_\ell - 1)$ nodes and $(S_\ell - 1)(S_\ell - 2)/2$ arcs; these network elements are called the line nodes and the line arcs, respectively.

There are also network nodes to represent the trip end centroids (i.e. the zones of origin and destination), the endpoints of walk arcs, and the junctions where a rider may choose a line option: a such node m is connected to a line access node $n_{\ell,i}^+$ by a boarding arc $a \approx (m, n_{\ell,i}^+)$. Symmetrically, there are alighting arcs $a \approx (n_{\ell,j}^-, m)$. The network arcs include the line leg arcs, the boarding arcs, the

DOI: 10.1002/atr

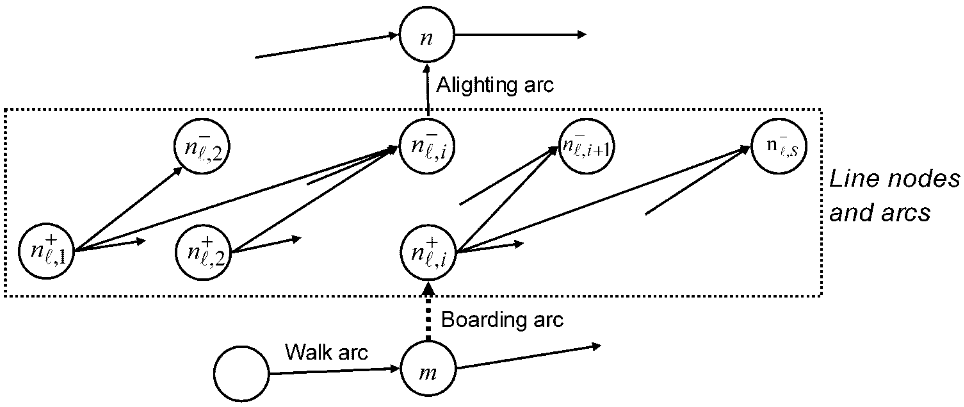


Figure 4. Representation of transit line within assignment network.

alighting arcs and the arcs of other transportation modes such as walking, used for transfer or connection to and from the centroid nodes. Figure 4 depicts the kinds of nodes and arcs associated to a transit route ℓ, be it by inclusion or by connection.

In large transit networks the representation of line legs as network arcs yields a very large assignment network. Taking as instance the Paris multimodal transit network that involves 120 train routes and 1320 bus routes, the assignment network involves about 570 000 arcs, among which 354 000 leg arcs, 131 000 transfer arcs and 84 000 other walking arcs. A segment-based representation would involve about 242 000 arcs, among which 26 500 segment arcs only [21].

The leg-as-arc representation involves an important assumption about route choice and user behaviour: that each rider who chooses a leg will travel along it up to its exit station, whatever the service mode he gets. Yet, in the basic transit assignment model of optimal strategy, the rider is assumed to be cost-minimizing under dynamic information limited to his perception. Indeed, a rider standing on-board that could not get a seat does perceive that he is standing; eventually he would consider whether to continue aboard or to exit at the current stop where he failed to get a seat. Such within-leg rerouting behaviour is prohibited under the leg-as-arc representation.

4.2. Model variables and hyperpath issues

A transit line ℓ has attributes (f_ℓ, κ_ℓ) of vehicle frequency and seat capacity, respectively, during the assignment period.

Any arc a has an average traversal cost, $\hat{c}_a$, and a service frequency, f_a, which is set to either line frequency f_ℓ when a is a boarding arc to line ℓ or infinity otherwise.

The set of destinations is denoted by S – for sink nodes. To a destination s is associated a set W^s of OD pairs $w = (o, s)$ with OD flow of q_{os}. To the set $W = \cup_{s \in S} W^s$ is associated the vector of OD flows $\mathbf{q} = [q_w]_{w \in W}$, which is the OD trip matrix.

At the arc level, let x_{as} denote a flow of users on arc a destined to s.

A vector of arc-destination flows $\mathbf{x}_{AS} = [x_{as} : a \in A, s \in S]$ is called a **network flow state**. Its restriction to the legs of a transit line ℓ is denoted as $\mathbf{x}_{\ell S} = [x_{as} : a \in A_\ell, s \in S]$.

A **hyperpath** $h = (\hbar, \hat{h})$ with destination node s is a pair of an arc set $\hbar$ and a routing field $\hat{h}$. The hyperpath arc set $\hbar$ contains no oriented cycle and is such that each arc $a \in \hbar$ belongs to a positive path within $\hbar$ towards s. The routing field $\hat{h}$ is a mapping of A onto $[0, 1]$ such that $\hat{h}_a = 0$ if $a \notin \hbar$ and $\sum_{a \in A_m^+} \hat{h}_a = 1$ for all nodes m lying between n and s along $\hbar$ (except for s), A_m^+ being the set of arcs that go out of m.

From a given node n to a destination node s, let H_{ns} denote the set of hyperpath arc sets $\hbar$ from n to s *via* the network, eventually submitted to some more restrictions (e.g. to obtain an optimal-route model in the sense of De Cea and Fernandez, 1988).

In transit assignment, the routing field $\hat{h}$ associated to a hyperpath arc set $\hbar$ is further constrained at each node m that is connected to a line access node $n^+_{\ell,i}$ by a boarding arc on the basis of the following condition of flow share:

$$\hat{h}_a = \frac{1_{\{a \in \hbar\}} f_a}{F^{\hbar}_m} \tag{7}$$

wherein $F^{\hbar}_m = \sum_{a \in A^+_m} 1_{\{a \in \hbar\}} f_a$ denotes the combined frequency of the active lines that can be accessed from this node.

We further constrain the arc set $\hbar$ of hyperpath h by imposing two node-based conditions as follows:

(i) if $A^+_m \cap \hbar$ contains boarding arcs then it cannot contain a nonboarding arc, and
(ii) if $A^+_m \cap \hbar$ does not contain boarding arcs then it is limited to at most one arc, for which $\hat{h}_a = 1$.

Under these conditions the routing field $\hat{h}$ is uniquely determined by the arc set $\hbar$, which enables us to assimilate hyperpath $h = (\hbar, \hat{h})$ to its arc set $\hbar$.

Then, given a network flow state $\mathbf{x}_{AS}$, the travel cost from node n to destination s along hyperpath $h \in H_{ns}$ is defined as the average travel cost along its paths:

$$C_{ns}(h, \mathbf{x}_{AS}) = \sum_{r \in R_{ns}(h)} h_r \left[\left(\sum_{m \in r} w^h_m \right) + \sum_{a \in r} c_a(\mathbf{x}_{AS}) \right] \tag{8}$$

wherein:

- $R_{ns}(h)$ is the set of elementary paths along h from n to s,
- $h_r = \prod_{a \in r} h_a$ is the path flow proportion,
- $c_a(\mathbf{x}_{AS})$ is the travel cost of arc a conditional on flow state $\mathbf{x}_{AS}$,
- $w^h_m \equiv \alpha_m / F^h_m$ if m has outgoing boarding arcs or 0 otherwise: it stands for the waiting delay at node m.

The waiting weighting factor α_m is equal to one if the transit lines accessible from m are serviced by vehicles with interarrival times that obey to a negative exponential distribution; under other service assumptions it may take a different value, so that w^h_m would in all way stand for the average waiting time at m until the arrival of the next vehicle from among the transit lines that are accessible (i.e. their boarding arc $a \in h$).

4.3. Cost–flow relationship

The arc costs are assumed to depend on the network flow state, $\mathbf{x}_{AS}$, on the basis of the following cost–flow function:

$$\mathbf{x}_{AS} \mapsto \mathbf{c}_A(\mathbf{x}_{AS}) = [c_a(\mathbf{x}_{AS}) : a \in A] \tag{9}$$

This framework encompasses a range of congestion effects:

- Local congestion induced by the local arc flow $x_a = \sum_{s \in S} x_{as}$, by restricting $c_a(\mathbf{x}_{AS}) = c_a(x_a)$.
- Seat congestion along a transit leg $a \in A_\ell$, by restricting $c_a(\mathbf{x}_{AS}) = c_a(\mathbf{x}_\ell)$ in which $\mathbf{x}_\ell = [x_a : a \in A_\ell]$.

More precisely, the effect of the seat congestion model on the leg cost can be stated in three steps. Firstly, the line-loading algorithm amounts to a pair of functions:

$$\mathbf{p}^0_\ell = \mathbf{P}^0_\ell(\mathbf{x}_\ell) \tag{10a}$$

$$\mathbf{p}^+_\ell = \mathbf{P}^+_\ell(\mathbf{x}_\ell) \tag{10b}$$

for the sitting probabilities along the line stations, either from on-board or at entry. These depend on the line trip matrix, $\mathbf{x}_\ell$.

DOI: 10.1002/atr

Secondly, the line segment costs $\underline{c}_{i,i+1}$ at seating and $\overline{c}_{i,i+1}$ at standing can be modelled as functions of the line trip matrix:

$$\underline{c}_{i,i+1} = \underline{C}_{i,i+1}(\mathbf{x}_\ell) \tag{11a}$$

$$\overline{c}_{i,i+1} = \overline{C}_{i,i+1}(\mathbf{x}_\ell) \tag{11b}$$

The only constraint is that $\underline{C}_{i,i+1} \leq \overline{C}_{i,i+1}$ at each point $\mathbf{x}_\ell$.

Thirdly, the line leg costing algorithm derives the mean leg costs from the sitting probabilities $p^0_{\ell,i}$ and $p^+_{\ell,i}$ together with the segment costs $\underline{c}_{i,i+1}$ and $\overline{c}_{i,i+1}$:

$$\hat{c}_{i,j} = \hat{C}_{i,j}(\mathbf{p}^0_\ell, \mathbf{p}^+_\ell, [\underline{c}_{i,i+1}]_{i\in\ell}, [\overline{c}_{i,i+1}]_{i\in\ell}) \tag{12}$$

Thus, by the composition of functions, leg $a \approx (i,j)$ along line ℓ is associated with

$$c_a = \hat{C}_{i,j}\big(\mathbf{P}^0_\ell(\mathbf{x}_\ell), \mathbf{P}^+_\ell(\mathbf{x}_\ell), [\underline{C}_{i,i+1}(\mathbf{x}_\ell)]_{i\in\ell}, [\overline{C}_{i,i+1}(\mathbf{x}_\ell)]_{i\in\ell}\big) \tag{13}$$

which is hereafter denoted by $c_a = c_a(\mathbf{x}_{AS})$, knowing that a is a leg arc.

To ensure the existence of a traffic equilibrium, we require a property of regularity (i.e. at least continuity) for the cost–flow relationship. It turns out that under the basic definition (13) some discontinuities may arise in the line loading algorithm at points $\mathbf{x}_\ell$ such that $\kappa^\delta_i = 0$ and $y^\delta_i = 0$ at a given station i and stage δ.

This leads us to define a continuous approximation of the cost functions, on the basis of a positive parameter ε—to be chosen close to zero—and the derived variables κ^δ_i and y^δ_i of residual capacity and standing riders, respectively:

$$c^0_{i,j} = \underline{c}_{a\approx(i,i+1)} + c^0_{i+1,j} \tag{14a}$$

$$z^{\varepsilon 0}_i = \min\{y^0_i, \max\{\varepsilon, \kappa^0_i\}\} \tag{14b}$$

$$\gamma^\varepsilon_{i,j} = \overline{c}_{i,i+1} + \frac{z^{\varepsilon 0}_{i+1}}{y^{\varepsilon 0}_{i+1}} c^0_{i+1,j} + \left(1 - \frac{z^{\varepsilon 0}_{i+1}}{y^{\varepsilon 0}_{i+1}}\right)\gamma^\varepsilon_{i+1,j} \quad \text{if} \quad y^{\varepsilon 0}_{i+1} > 0 \quad \text{or} \quad \overline{c}_{i,i+1} + c^0_{i+1,j} \quad \text{if} \quad y^{\varepsilon 0}_{i+1} = 0 \tag{14c}$$

$$z^{\varepsilon +}_i = \min\{y^+_i, \max\{\varepsilon, \kappa^+_i\}\} \tag{14d}$$

$$\overset{\frown}{c}{}^{\varepsilon}_{i,j} = \frac{z^{\varepsilon +}_i}{y^{\varepsilon +}_i} c^0_{i,j} + \left(1 - \frac{z^{\varepsilon +}_i}{y^{\varepsilon +}_i}\right)\gamma^\varepsilon_{i,j} \quad \text{if} \quad y^{\varepsilon +}_i > 0 \quad \text{or} \quad c^0_{i,j} \text{ if } y^{\varepsilon +}_i = 0 \tag{14e}$$

On comparing (14) to (5), if $\kappa^0_{i+1} \geq \varepsilon$ then $z^{\varepsilon 0}_i = \min\{y^0_i, \kappa^0_i\}$ so that if $y^0_{i+1} > 0$ then (14c) is equivalent to (5b). Similarly, if $\kappa^+_i \geq \varepsilon$ and $y^+_i > 0$ then (14e) is equivalent to (5c). Parameter ε makes a lower bound on any residual capacity, so as to ensure the regularity of the composite cost function (conditional or average). Letting $\varepsilon \to 0^+$ makes the approximate costs in (14) as close as required to the basic costs in (5), except at the points of discontinuity with $(\kappa^\delta_i, y^\delta_i) = (0,0)$.

Let us, then, replace (13) with the following (15):

$$c^\varepsilon_a = \hat{C}^\varepsilon_{i,j}(\mathbf{P}^0_\ell(\mathbf{x}_\ell), \mathbf{P}^+_\ell(\mathbf{x}_\ell), [\underline{C}_{i,i+1}(\mathbf{x}_\ell)]_{i\in\ell}, [\overline{C}_{i,i+1}(\mathbf{x}_\ell)]_{i\in\ell}) \tag{15}$$

which is hereafter denoted by $c^\varepsilon_a = c^\varepsilon_a(\mathbf{x}_{AS})$, knowing that a is a leg arc.

Theorem 1 continuity of leg cost–flow relationship

Assume that transit line ℓ has segment seated and standing costs $\underline{C}_{i,i+1}$ and $\overline{C}_{i,i+1}$ that are functions of leg trip matrix, $\mathbf{x}_\ell \geq 0$. If the segment cost functions are continuous (resp. sub-differentiable) with respect to $\mathbf{x}_\ell$, so are the approximate average leg cost functions $\hat{C}^\varepsilon_{i,j}$ defined in (15), and so are the average leg cost functions $\hat{C}_{i,j}$ defined in (13) except at points where $\kappa^\delta_i = 0$ and $y^\delta_i = 0$.

To demonstrate the theorem we shall use three lemmas.

Lemma 1 regularity of sitting probability

DOI: 10.1002/atr

For $(\kappa, y) \geq 0$ *let* $p(\kappa, y) \equiv \min\{1, \kappa/y\}$ *if* $\kappa > 0$ *and* $y > 0$, *else* $p(\kappa, y) \equiv 0$ *if* $\kappa = 0$, *else* $p(\kappa, y) \equiv 1$ *if* $\kappa > 0$ *and* $y = 0$. *Then function p is continuous with respect to* (κ, y), *except at* $(0, 0)$. *It is continuously differentiable except along* $\{\kappa = y\}$.

Proof The property is obvious if $\kappa > 0$, $y > 0$ and $\kappa \neq y$ since then either $\kappa > y$ yielding $p = 1$, or $y > \kappa$ hence $p = \kappa/y$ which is continuously differentiable on the restricted domain. The line $\{\kappa = y\}$ separates the two sub-domains of differentiability. If $\kappa > 0$ then p is right continuous at $y = 0$ since $\kappa/y \to \infty$ as $y \to 0^+$ hence $p = 1$. If $\kappa = 0$ and $y > 0$ then $p(0, y) \to 0$ as $y \to 0^+$ meaning that p is continuous in y at $(0, 0)$.

Lemma 2 continuity throughout line loading

At every station i and stage δ *along line* ℓ, *the functions* κ_i^δ *and* y_i^δ *are continuous and sub-differentiable with respect to* $\mathbf{x}_\ell$. *This regularity property also holds for every derived function* $\underline{x}_j^{(i)\delta}$ *and* $\overline{x}_j^{(i)\delta}$. *The sitting probability* p_i^δ *is regular except perhaps if* $y_i^\delta = 0$.

The *proof* proceeds by induction from the origin station $i = 1$ with $\kappa_1^- = \kappa_\ell$ and variables $\underline{x}_j^{(0)+}$, $\overline{x}_j^{(0)+}$, $\underline{x}_{\geq 0}^+$ and $\overline{x}_{\geq 0}^+$ that are nil. This implies that $\kappa_1^0 = \kappa_\ell$ and $y_1^0 = 0$, which are regular w.r.t. $\mathbf{x}_\ell$. Let the induction assumption be that the lemma property holds at station i and stage δ. Let us consider the case where the next station and stage pair is station $i + 1$ at stage '0' where the riders standing on-board try to get a seat. From the induction assumption, the number of seated passengers that exit at $i + 1$, $\underline{x}_{i+1}^{(i)+}$, is regular: the same applies to $\kappa_{i+1}^0 = \kappa_{i+1}^- + \underline{x}_{i+1}^{(i)+}$, to $y_{i+1}^0 = \overline{x}_{\geq i}^+ - \overline{x}_{i+1}^{(i)+}$ and their minimum function

$$\left[p_{i+1}^0 y_{i+1}^0\right] \equiv \min\left\{\kappa_{i+1}^0, y_{i+1}^0\right\}$$

because the operators of addition, subtraction and minimization preserve the regularity property. If $y_{i+1}^0 > 0$ then $p_{i+1}^0 = \left[p_{i+1}^0 y_{i+1}^0\right]/y_{i+1}^0$ is regular as well, and so are the products $p_{i+1}^0 \overline{x}_j^{(i)+}$ for $j > i$ and the derived functions $\overline{x}_j^{(i+1)0}$ and $\underline{x}_j^{(i+1)0}$. If $y_{i+1}^0 \to 0$ then $\left[p_{i+1}^0 y_{i+1}^0\right] \to 0$ as do the products $p_{i+1}^0 \overline{x}_j^{(i)+}$, hence regularity also holds.

This demonstrates that the induction assumption holds for the next stage of $(i, +)$. If $(i, \delta) = (i, 0)$ then the next stage is $(i, +)$: the proof of the lemma property is similar, because the simple operators of minimization, addition, subtraction, product and division by a positive function maintain the regularity w.r.t. $\mathbf{x}_\ell$. Caution is only required when $y_j^\delta = 0$ since the function of sitting probability may not be continuous there.

Lemma 3 continuity throughout line costing

If the segment cost functions $\underline{C}_{i,i+1}$ and $\overline{C}_{i,i+1}$ by comfort state are continuous (resp. sub-differentiable) w.r.t. $\mathbf{x}_\ell$ between stations k and j, then so are the approximate functions of seated cost, average cost, conditional cost, cost variance and conditional variance for the line leg (k, j).

The *proof* proceeds by induction from the egress station j to the upstream stations k in backward order. To initialize the Induction process, let us consider a 'fictitious' leg (j, j) where all the cost functions are null. Then, assuming that the lemma property holds at every station i located between $k + 1$ and j, let us establish that it also holds at station k. From (5a) the basic seated cost function is regular since regularity is maintained through the operator of addition. From Lemma 2 and the conservation of regularity through maximization and minimization, the $z_i^{\varepsilon\delta}$ functions in (14b) and (14d) are regular. From Lemma 1 and the truncation of the residual capacity from below at level ε, the approximate sitting probabilities $z_i^{\varepsilon\delta}/y_i^\delta$ are regular, even at $y_i^\delta = 0$ where this ratio is equal to 1. Thus the composite costs in (14c) and (14e) are regular.

The same lines of proof ensure the regularity of the approximate functions of cost variance that are adapted from (6) by replacing the true sitting probabilities with the z/y ratios. This ends up the proof that the induction assumption holds at station k.

The same proof also applies to the basic cost functions $\hat{C}_{ij}$ defined in (13) at $\mathbf{x}_\ell$ such that $(\kappa_i^\delta, y_i^\delta) \neq (0, 0)$ for every station $i \in \ell$ and stage $\delta \in \{0, +\}$.

Overall, Lemmas 1–3 make Theorem 1 hold true.

The approximate cost function is lower than the basic cost, because the approximate residual capacity is larger than the true residual capacity: this enables to (fictively) increase the sitting

DOI: 10.1002/atr

probability and the proportion of riders who incur seated cost instead of standing cost. Thus the approximate cost function tends to make the leg more attractive than it would be under the basic cost: the effect on route choice, however, is limited since both the seated cost and the sitting probabilities are maintained.

5. EQUILIBRIUM ANALYSIS

Having laid the foundation, we are now ready to define a traffic equilibrium and to carry out the relevant mathematical analysis. We shall first set up the feasible set of flow states and their hyperpath representation, then define a traffic equilibrium in the form of a NCP. After providing an equivalent characterization in the form of a VIP, we shall demonstrate the existence of an equilibrium state. The MSA makes a heuristic algorithm to compute an equilibrium state: its convergence can be assessed in a rigorous way on the basis of a duality gap criterion, for which we shall provide a simple formula.

5.1. *On feasible flow states and hyperpath representation*

Definition 1, Feasible network flow state. *A network flow state* $\mathbf{x}_{AS} = [x_{as}]_{a\in A, s\in S}$ *is feasible if it is nonnegative and if it satisfies the node conservation of flow by destination:*

$$x_{as} \geq 0 \ \forall a\in A, s\in S \tag{16a}$$

$$\sum\nolimits_{a\in A_m^+} x_{as} = q_{ms} + \sum\nolimits_{a\in A_m^-} x_{as} \ \forall s\in S, m\in N, m \neq s \tag{16b}$$

in which A_m^+ [resp. A_m^-] denotes the subset of arcs that go out [resp. come in] node m and q_{ms} is a given OD flow from node m to destination node s.

Let $\mathbf{E_x}$ be the set of feasible network flow states.

Recalling the sets H_{ns} of hyperpaths from n to s, a **hyperpath flow state** $\mathbf{X}_{NS}$ is defined as a linear combination of elementary flows along hyperpaths h with coefficients q_{ns}^h:

$$\mathbf{X}_{NS} = [q_{ns}^h : s\in S, n\in N\backslash\{s\}, h\in H_{ns}] \tag{17}$$

Definition 2. Feasible hyperpath flow state. *Given the* OD *trip matrix* $\mathbf{q}_{NS} = [q_{ns}]_{n\in N, s\in S}$, *a hyperpath flow state* $\mathbf{X}_{NS}$ *is feasible if it is nonnegative and it satisfies the conservation of flow by* OD *pair:*

$$q_{ns}^h \geq 0 \ \forall s\in S, n\in N\backslash\{s\}, h\in H_{ns} \tag{18a}$$

$$\sum\nolimits_{h\in H_{ns}} q_{ns}^h = q_{ns} \forall s\in S, n\in N, n \neq s \tag{18b}$$

A hyperpath flow state, $\mathbf{X}_{NS}$, induces a network flow state, $\mathbf{x}_{AS} = \mathbf{A}(\mathbf{X}_{NS})$, in the following way:

$$x_{as} = \sum_{n\in N} \sum_{h\in H_{ns}} q_{ns}^h \sum_{r\in R_{ns}(h)} \hat{h}_r 1_{\{a\in r\}} \tag{19}$$

in which $R_{ns}(h)$ denotes the set of elementary paths within h from n to s (positive paths with no node repetition), $1_{\{a\in r\}}$ is equal to 1 if $a\in r$ or 0 otherwise, and $\hat{h}_r = \prod_{a\in r} \hat{h}_a$ is the proportion of flow carried out from n to s *via* route r on h.

From the basic properties of hyperpaths [25], a feasible hyperpath flow state induces a feasible network flow sate. Nonnegativity (16a) comes from (18a), (19) and the nonnegativity of $\hat{h}_a$ hence of $\hat{h}_r$. The conservation of the destination flow at node m in (16b) is derived as follows:

$$\sum\nolimits_{a\in A_m^+} x_{as} - \sum\nolimits_{a\in A_m^-} x_{as} = \sum\nolimits_{n\in N} \sum\nolimits_{h\in H_{ns}} q_{ns}^h \sum\nolimits_{r\in R_{ns}(h)} \hat{h}_r \Delta_{rm}$$

in which $\Delta_{rm} = \sum_{a\in A_m^+} 1_{\{a\in r\}} - \sum_{a\in A_m^-} 1_{\{a\in r\}}$. Now, if m is not incident to r then all the terms in the sum are zero, yielding $\Delta_{rm} = 0$. Then, if m is incident to r but not an endpoint node n or s of r, then $\Delta_{rm} = 0$ since one and only one arc in A_m^+ has $1_{\{a\in r\}} = 1$ and the same applies to A_m^-. Lastly, if $m = n$

DOI: 10.1002/atr

then $\Delta_{rm} = 1$ since $1_{\{a \in r\}} = 0$ for all a in A_m^-. Thus

$\sum_{a \in A_m^+} x_{as} - \sum_{a \in A_m^-} x_{as} = \sum_{h \in H_{ms}} q_{ms}^h \sum_{r \in R_{ms}(h)} \hat{h}_r = \sum_{h \in H_{ms}} q_{ms}^h = q_{ms}$

from the assumption (18b) on $\mathbf{X}_{NS}$.

5.2. Definition of traffic equilibrium using an nonlinear complementarity problem

Definition 3. Traffic Equilibrium. *Given the OD trip matrix* $\mathbf{q}_{NS} = [q_{ns}]_{n \in N, s \in S}$ *and the cost–flow relationship defined by* (8), (9) *and* (15)*, a hyperpath flow state* $\mathbf{X}_{NS} = [q_{ns}^h : s \in S, n \in N \backslash \{s\}, h \in H_{ns}]$ *with* $\mathbf{x}_{AS} = \mathbf{A}(\mathbf{X}_{NS})$ *is a traffic equilibrium if there exists a matrix* $\mu_{NS} = [\mu_{ns} : s \in S, n \in N \backslash \{s\}]$ *such that, for all* $s \in S, n \in N \backslash \{s\}$*:*

$$q_{ns}^h \geq 0 \quad \forall h \in H_{ns} \tag{20a}$$

$$\sum_{h \in H_{ns}} q_{ns}^h = q_{ns} \tag{20b}$$

$$C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS}) - \mu_{ns} \geq 0 \quad \forall h \in H_{ns} \tag{20c}$$

$$q_{ns}^h [C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS}) - \mu_{ns}] = 0 \quad \forall h \in H_{ns} \tag{20d}$$

The interpretation is as follows: to each destination s and from each node n, a hyperpath cost $C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS})$ cannot be less than μ_{ns} and only a hyperpath with cost $C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS}) = \mu_{ns}$ may carry a positive flow q_{ns}^h, which implies that under equilibrium the dual variable is a minimum hyperpath cost for the OD pair (n, s). This coincides with Wardrop's definition of user equilibrium in traffic assignment, according to which each user makes his routing choice so as to minimize his own travel cost.

The set of conditions (20) is a NCP in the variable $(\mathbf{X}_{NS}, \mu_{NS})$, with associated cost function as follows:

$(\mathbf{X}_{NS}, \mu_{NS}) \mapsto [C_{ns}^{\varepsilon}(h, \mathbf{A}(\mathbf{X}_{NS})) - \mu_{ns} : s \in S,\ n \in N \backslash \{s\},\ h \in H_{ns}]$

5.3. Equilibrium characterization using a variational inequality problem

Theorem 2 characterization of traffic equilibrium

A hyperpath flow state $\mathbf{X}_{NS} = [q_{ns}^h : s \in S,\ n \in N \backslash \{s\},\ h \in H_{ns}]$ *that is feasible for the OD trip matrix* $\mathbf{q}_{NS} = [q_{ns}]_{n \in N, s \in S}$ *is a traffic equilibrium if and only if, for any feasible hyperpath flow state* $\mathbf{Y}_{NS} = [\eta_{ns}^h : s \in S, n \in N \backslash \{s\}, h \in H_{ns}]$, *it holds that*:

$$\chi_{NS}(\mathbf{X}_{NS}) \cdot (\mathbf{Y}_{NS} - \mathbf{X}_{NS}) \geq 0 \tag{21}$$

in which $\chi_{NS}(\mathbf{X}_{NS}) = [C_{ns}^{\varepsilon}(h, \mathbf{A}(\mathbf{X}_{NS})) : s \in S, n \in N \backslash \{s\}, h \in H_{ns}]$.

Proof Assume first that $\mathbf{X}_{NS}$ is an equilibrium state. Letting μ_{NS} be the associated matrix of dual variables as of (20), for any $\mathbf{Y}_{NS} = [\eta_{ns}^h]$ it stems from (20c) that

$\eta_{ns}^h [C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS}) - \mu_{ns}] \geq 0 \quad \forall h \in H_{ns}$

Summation over $h \in H_{ns}$ yields that $\sum_{h \in H_{ns}} \eta_{ns}^h C_{ns}^{\varepsilon h} \geq \sum_{h \in H_{ns}} \eta_{ns}^h \mu_{ns} = q_{ns} \mu_{ns}$

From (20d) we get that $\sum_{h \in H_{ns}} q_{ns}^h C_{ns}^{\varepsilon h} = \sum_{h \in H_{ns}} \eta_{ns}^h \mu_{ns} = q_{ns} \mu_{ns}$, hence

$\sum_{h \in H_{ns}} C_{ns}^{\varepsilon h} (\eta_{ns}^h - q_{ns}^h) \geq 0$

This yields (21) after summing over $s \in S$ and $n \in N$.

Conversely, assume that (21) holds at $\mathbf{X}_{NS}$ and take $\mathbf{Y}_{NS} = [\eta_{ns}^h]$ equal to $\mathbf{X}_{NS}$ except perhaps on node–destination pair (n, s): if there is only one hyperpath h in H_{ns} then letting $\mu_{ns} = C_{ns}(h, \mathbf{x}_{AS})$ obviously satisfies (20c and 20d). If there are two or more hyperpaths in H_{ns}, for any h with $q_{ns}^h > 0$ let

DOI: 10.1002/atr

us define $\eta^h_{ns} = q^h_{ns} - \theta$ and $\eta^{h'}_{ns} = q^{h'}_{ns} + \theta$ for a small positive number θ: then (21) yields that

$$\left(C^{\varepsilon h'}_{ns} - C^{\varepsilon h}_{ns}\right)\theta \geq 0$$

which implies that the cost of any hyperpath h with positive flow q^h_{ns} in $\mathbf{X}_{NS}$ is minimal on H_{ns}: on defining $\mu_{ns} = \min_{h \in H_{ns}} C^{\varepsilon}_{ns}(h, \mathbf{x}_{AS})$, then (20c and 20d) follows.

5.4. Equilibrium properties

Theorem 3 existence of traffic equilibrium

Assuming that the cost functions C_a are continuous, there exists an equilibrium state for the model of transit assignment with seat capacity.

Proof Based on the VIP formulation and Theorem 1, the assumption ensures that the cost functions $\mathbf{x}_{AS} \mapsto C^{\varepsilon}_{ns}(h, \mathbf{x}_{AS})$ are continuous with respect to the network flow state $\mathbf{x}_{AS}$. As this state stems from a continuous function $\mathbf{A}$ of variable $\mathbf{X}_{NS}$ (a combination of hyperpath flows), the composed function $\mathbf{X}_{NS} \mapsto C^{\varepsilon}_{ns}(h, \mathbf{A}(\mathbf{X}_{NS}))$ is continuous, and so is function χ_{NS}. As the domain set $\mathbf{E}_{\mathbf{X}}$ of feasible hyperpath flows is convex and compact, this ensures that the variational inequality admits at least one solution.

The uniqueness of an equilibrium does not hold in general, as will be shown in Section 6.

5.5. A criterion of duality gap for convergence to equilibrium

The VIP formulation is useful to establish the existence of an equilibrium state and also to design an equilibration algorithm. As our model satisfies the assumptions in Ref. [27] about the mapping of arc costs, the two methods that these authors showed to be globally convergent are applicable to search for equilibrium: namely the linearized Jacobi method and the projection method. However, for the sake of simplicity we prefer to use the well-known MSA: although we have not demonstrated this method to be globally convergent on theoretical grounds, it is possible to assess its convergence in any application by evaluating the duality gap at the current flow state. Indeed, when the duality gap comes close to zero it is guaranteed that the current flow state is close to a traffic equilibrium, since the duality gap is continuous when the VIP function χ_{NS} is continuous.

Then the remaining issue is to evaluate the duality gap at each current flow state of a given iteration in the MSA. It would be most cumbersome to compute the duality gap in a straightforward way, since this would require not only to store all the used hyperpaths in the computer memory, but also to perform a hyperpath costing along each of them to evaluate their cost under the current traffic conditions. Hereafter a simple method is provided, of which the only requirement is to update two real variables by iteration along the MSA. It is applicable to any hyperpath-based transit assignment model in which the hyperpath cost depends on the flow state only through the arc costs, not the node costs (thus the line frequency cannot be related to the traffic flows).

Let us consider hyperpath flow state $\mathbf{X}_{NS}$ [resp. $\mathbf{Z}_{NS}$] with associated network flow state $\mathbf{x}_{AS}$ [resp. $\mathbf{z}_{AS}$] and coordinate q^h_{ns} [resp. η^h_{ns}] on hyperpath h from node n to destination s. Our aim is to evaluate the cost to carry the flow $\mathbf{Z}_{NS}$ under the travel conditions associated to $\mathbf{X}_{NS}$: this cost is defined as

$$\chi_{N+A}(\mathbf{Z}_{NS} : \mathbf{X}_{NS}) \equiv \mathbf{Z}_{NS} . \chi_{NS}(\mathbf{X}_{NS}) = \sum\nolimits_{s \in S} \sum\nolimits_{n \in N} \sum\nolimits_{h \in H_{ns}} \eta^h_{ns} C^{\varepsilon}_{ns}(h, \mathbf{x}_{AS}) \tag{22}$$

Let us split this cost into an arc-based part, χ_A, and a node-based part, χ_N. The arc-based part is

$$\begin{aligned} \chi_A(\mathbf{Z}_{NS} : \mathbf{X}_{NS}) &= \sum\nolimits_{s \in S} \sum\nolimits_{n \in N} \sum\nolimits_{h \in H_{ns}} \eta^h_{ns} \sum\nolimits_{r \in R_{ns}(h)} \hat{h}_r \sum\nolimits_{a \in r} c_a(\mathbf{x}_{AS}) \\ &= \sum\nolimits_{a \in A} c_a(\mathbf{x}_{AS}) \sum\nolimits_{s \in S} \sum\nolimits_{n \in N} \sum\nolimits_{h \in H_{ns}} \eta^h_{ns} \sum\nolimits_{r \in R_{ns}(h)} \hat{h}_r 1_{\{a \in r\}} \end{aligned}$$

hence

$$\chi_A(\mathbf{Z}_{NS} : \mathbf{X}_{NS}) = \sum\nolimits_{a \in A} c_a(\mathbf{x}_{AS}) z_a \tag{23}$$

which is easy to evaluate whatever the $\mathbf{Z}_{NS}$ state.

About the node-based part,

$\chi_N \equiv \sum_{s\in S} \sum_{n\in N} \sum_{h\in H_{ns}} \eta^h_{ns} \sum_{r\in R_{ns}(h)} \hat{h}_r \sum_{m\in r} w^h_m$

let us notice that w^h_m does not depend on $\mathbf{x}_{AS}$ (when no congestion effect is involved in line frequency) and that any term $\sum_{r\in R_{ns}(h)} \hat{h}_r \sum_{m\in r} w^h_m$ is a function of h and $R_{ns}(h)$ only, from here denoted as ρ^h_{ns}. Thus the node-based part is a linear function of $\mathbf{Z}_{NS}$ only, on the basis of its coordinates η^h_{ns} in the space of hyperpaths:

$$\chi_N(\mathbf{Z}_{NS}) = \sum_{s\in S} \sum_{n\in N} \sum_{h\in H_{ns}} \eta^h_{ns}\rho^h_{ns} \tag{24}$$

Consider now an auxiliary hyperpath flow state $\mathbf{Y}^{(k)}_{NS}$ that arises at the kth iteration in an MSA application with sequence of step sizes $(\zeta_k)_{k\geq 0}$ such that $\zeta_0 = 1$. The global cost $\mathbf{Y}^{(k)}_{NS} \cdot \chi_{NS}(\mathbf{X}^{(k)}_{NS})$ of that auxiliary state under the traffic conditions induced by the current flow state $\mathbf{X}^{(k)}_{NS}$ is merely

$\mathbf{Y}^{(k)}_{NS} \cdot \chi_{NS}\left(\mathbf{X}^{(k)}_{NS}\right) = \sum_{s\in S} \sum_{n\in N} q_{ns} u^{(k)}_{ns}$

in which $u^{(k)}_{ns}$ is the travel cost by unit flow from node n to destination s under flow state $\mathbf{X}^{(k)}_{NS}$: this stems from the definition of the auxiliary state by the assignment of all OD flows to their hyperpath of minimal cost under the current traffic conditions. Based on (23) it is straightforward to recover

$$\chi_N\left(\mathbf{Y}^{(k)}_{NS}\right) = \mathbf{Y}^{(k)}_{NS} \cdot \chi_{NS}\left(\mathbf{X}^{(k)}_{NS}\right) - \chi_A\left(\mathbf{Y}^{(k)}_{NS} : \mathbf{X}^{(k)}_{NS}\right) = \left[\sum_{s\in S} \sum_{n\in N} q_{ns} u^{(k)}_{ns}\right] - \sum_{a\in A} c_a(\mathbf{x}_{AS}) y^{(k)}_a \tag{25}$$

Let us now turn our attention to the current state $\mathbf{X}^{(k)}_{NS}$, which is constructed as the weighted average of the previous auxiliary states $\mathbf{Y}^{(j)}_{NS}$, for $j < k$, with weight coefficients $\zeta^{(k)}_j = \zeta_j/\Gamma_{k-1}$ where $\Gamma_k \equiv \sum_{j=0}^{k} \zeta_j$. The arc-based transport cost is evaluated straightforwardly. Being a linear function, the node-based transport cost can be derived from the linear decomposition of $\mathbf{X}^{(k)}_{NS}$:

$$\chi_N\left(\mathbf{X}^{(k)}_{NS}\right) = \chi_N\left(\sum_{j=0}^{k-1} \zeta^{(k)}_j \mathbf{Y}^{(j)}_{NS}\right) = \sum_{j=0}^{k-1} \zeta^{(k)}_j \chi_N\left(\mathbf{Y}^{(j)}_{NS}\right) \tag{26}$$

Lastly, let us define $\beta_k \equiv \sum_{j=0}^{k} \zeta_j \chi_N\left(\mathbf{Y}^{(j)}_{NS}\right)$ so that the ratio β_{k-1}/Γ_{k-1} amounts to the node-based costs of the current flow state under its own traffic conditions:

$$\chi_N\left(\mathbf{X}^{(k)}_{NS}\right) = \frac{\beta_{k-1}}{\Gamma_{k-1}} \tag{27}$$

To sum up, at iteration k in the MSA algorithm the duality gap is

$$DG_k = \left(\mathbf{X}^{(k)}_{NS} - \mathbf{Y}^{(k)}_{NS}\right) \cdot \chi_{NS}\left(\mathbf{X}^{(k)}_{NS}\right) = \frac{\beta_{k-1}}{\Gamma_{k-1}} + \left[\sum_{a\in A} c_a(\mathbf{x}^{(k)}_{AS}) x^{(k)}_a\right] - \left[\sum_{s\in S} \sum_{n\in N} q_{ns} u^{(k)}_{ns}\right] \tag{28}$$

This duality gap formula applies to any hyperpath-based transit assignment model in which the cost–flow relationship pertains to the arcs—as in (9) and (15)—and the node-based waiting costs are independent of the flow.

5.6. *Assignment algorithm: a method of successive averages*

Traffic assignment to a transit network with seat congestion can be performed by means of the following equilibration algorithm, which makes use of two network flow states $\mathbf{x}_{AS}$ for a current state and $\mathbf{y}_{AS}$ for an auxiliary state, two related overall arc flow vectors $\mathbf{x}_A$ and $\mathbf{y}_A$, one matrix of node potentials by destination $\mathbf{u} = [u_{ns} : s\in S, n\in N]$, an iteration counter k, real variables β, Γ, U, W and Z. Input variables consist in G, L, arc costs $\underline{\mathbf{c}} = [\underline{c}_a]_{a\in A}$ and $\overline{\mathbf{c}} = [\overline{c}_a]_{a\in A}$, line attributes $\mathbf{f} = [f_\ell]_{\ell\in L}$ and $\kappa = [\kappa_\ell]_{\ell\in L}$, OD trip matrix $\mathbf{q} = [q_{os}]_{o\in O_s, s\in S}$, a tolerance η on the convergence level, and a sequence of decreasing positive numbers $(\zeta_k)_{k\geq 0}$ with $\zeta_0 = 1$.

The equilibration algorithm is made up of five steps:

Initialization. Set $\mathbf{x}_{AS} := 0$ and $\mathbf{x}_A := 0$. Let $k := 0$, $\beta := 0$ and $\Gamma := 0$.

Cost–Flow Relationship. Evaluate the arc costs $c_a = C_a(\mathbf{x}_{AS})$ for all $a \in A$ as in Section 4.3.

Network Costing and Flow Loading. Let $U := 0$ and $\mathbf{y}_A := 0$. For every destination node $s \in S$:

- Find the optimal hyperpath destined to s under the current arc costs, yielding node potentials u_{ns}.
- Load the OD flows $[q_{ns} : n \in N]$ on the currently optimal hyperpaths destined to s, yielding arc flows y_{as}.
- Let $U := U + \sum_{n \in N} q_{ns} u_{ns}$. Let $y_a := y_a + y_{as}$ for all $a \in A$.

Flow Update. Let $W := U - \sum_{a \in A} c_a y_a$. If $k > 0$ let $Z := \frac{\beta}{\Gamma} - U + \sum_{a \in A} c_a x_a$. Then let $\Gamma := \Gamma + \zeta_k$, $\beta := \beta + \zeta_k W$, $\mathbf{x}_{AS} := \mathbf{x}_{AS} + \zeta_k(\mathbf{y}_{AS} - \mathbf{x}_{AS})$ and $\mathbf{x}_A := \mathbf{x}_A + \zeta_k(\mathbf{y}_A - \mathbf{x}_A)$.

Convergence Test. If $k > 0$ and $Z \leq \eta$ then terminate, else let $k := k + 1$ and go to step *Cost Flow Relationship*.

This is a mere MSA, in which the auxiliary flow state $\mathbf{y}_{AS}$ is a user-optimized assignment of all the OD flows on the basis of the costs induced by the current flow state $\mathbf{x}_{AS}$. The convergence criterion Z is the duality gap of the previous section.

In the application of the algorithm, working variables u_{ns} need not be saved by destination, which allows for using only one vector $[u_n : n \in N]$ instead of a matrix $[u_{ns} : s \in S, n \in N \backslash S]$. Furthermore, in the standard case where the arc costs depend on $\mathbf{x}_{AS}$ through $\mathbf{x}_A$ only, then it is not necessary to save the arc flows by destination. Thus the memory requirements of the algorithm exceed those of the uncapacitated model of Spiess and Florian [2] only by the representation of one arc per leg.

The network assignment algorithm was implemented in cooperation with the RATP (Paris metro operator), with some refinements that pertain to the issues of line sub-services and of generalized cost including mean and variance of leg cost: results are available in two related papers [21,28]. In short, the equilibrium results appear to be stable in terms of arc flows and elementary costs (including access–egress costs by service mode), though variable in terms of selected hyperpaths. The comparison between the seat-congestion model and the previous model, a standard hyperpath transit assignment model, showed that at the morning peak the transit flows could vary by as much as 30% along the metro lines that carry about 5000–20 000 passengers per hour and per direction. Along the regional train lines the relative change is limited to a small percentage, because there are few alternative transit routes and their base flow is as high as 20 000–60 000 passengers per hour and per direction. A number of 30 iterations was sufficient to achieve a satisfactory level of convergence, meaning that including seat capacity in static transit assignment is similar to modelling link travel time functions in static roadway assignment as concerns the computational costs.

6. A SIMPLE CASE OF HYPERPATH CHOICE

To illustrate the seat congestion effect in conjunction with hyperpath choice in the context of traffic assignment, let us consider a simple case inspired from the transit network in the Paris area. At the morning peak period, the most crowded transit line in Paris is a heavy metro line called line A of RER (the Réseau Express Régional, in fact a regional train with very high frequency), particularly so in the westbound direction because there is a major Business Centre at station La Défense, and at the Châtelet station which allows for transfer between three RER lines and five metro lines. Trip-makers willing to board line A at Châtelet to La Défense have to wait for the second or third train to come, because of a queue. Thus it is highly likely that on-board they will stand within high crowding. To avoid congestion, some trip-makers prefer to use an alternative route: those coming from the Belleville metro station may board line A either at Châtelet after using line 11, or at station Nation after using line 2. Boarding line A at Nation is easier as there is no queue.

6.1. Case data and test parameters

Figure 5 illustrates the network and transport data for a trip from Belleville to La Défense. On line A, the leg from Châtelet to La Défense costs either 11 minutes at seating or 20 minutes at standing, on the basis of Stated Preference data [28]. The leg from Nation to Châtelet costs 6 minutes when seated,

DOI: 10.1002/atr

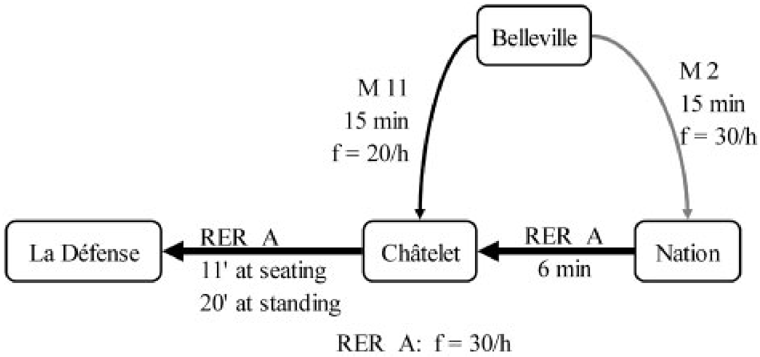


Figure 5. Transit map for the Belleville–La Défense relationship.

which is assumed here. The leg from Belleville to Châtelet *via* line 11 takes 15 minutes including transfer, as takes the leg from Belleville to Nation *via* line 2 also including transfer.

Service frequencies are about 20 veh/hour on line 11, 30 veh/hour on line 2 and 30 veh/hour on line A. The waiting weighting factor α is set equal to 1. The traffic flows and capacity are described on the basis of three parameters q, ξ and κ: q is the OD flow from Belleville to La Défense; ξ is the flow from other origins boarding line A at Châtelet; κ is the seat capacity available on line A at the Nation station.

On the Belleville-La Defense relationship, Path 1 with access to line A at Châtelet and Path 2 with access to line A at Nation make up three hyperpaths: hyperpath 1 made up of path 1 alone; hyperpath 2 made up of path 2 alone; and hyperpath 3 that combines path 1 and path 2 at the Belleville station.

Figure 6 depicts the assignment network under the leg-as-arc format. For numerical illustration, let us fix OD flows $q_{BD} \equiv q = 8000$ passenger per hour and $q_{CD} \equiv \xi = 7000$ per hour. The latter flow is assigned to its single path at every iteration in the MSA, whereas the former is assigned to its Hyperpath 1 (simple path boarding line A at node C) at the Initialization Step, then to its hyperpath 3 (combination) at every iteration with $k > 0$.

At equilibrium, the sitting probability along line A is 1 at Nation and 51% at Châtelet. The evolution of the duality gap along the iterations in the MSA is displayed on Figure 7, with step size set to

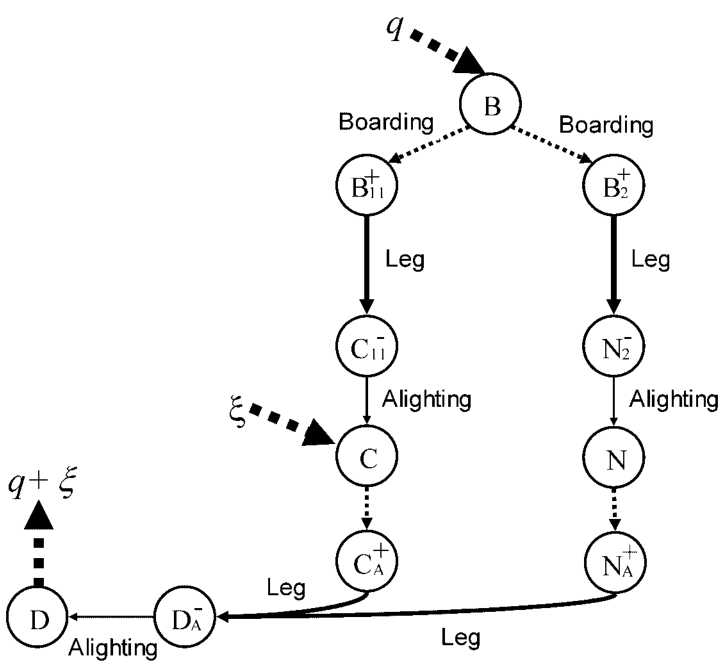


Figure 6. Assignment network for the Belleville–La Défense relationship.

DOI: 10.1002/atr

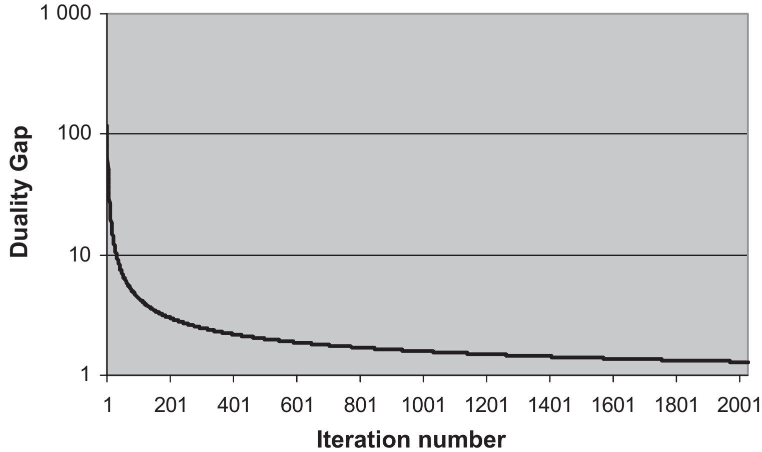


Figure 7. Duality gap against iteration number in the MSA.

$\zeta_k = 1/(1 + k/50)$ in the convex combination at iteration $k > 0$. For reference, the total travel cost at equilibrium is 6640 passenger.hour.

If seat capacity was neglected, then from Belleville only Path 1 would be attractive, yielding a total travel cost of $8000 \times 31' + 7,000 \times 14' = 5933$ passenger.hour. Thus the generalized travel costs would be underestimated by 10%, whereas the traffic loads onto the links to the Chatelet and Nation stations would be even more significantly different.

6.2. *On paths hyperpaths and costs*

On the Belleville-La Defense relationship the respective hyperpath costs are the following functions of the probability p to take a seat on boarding at the Châtelet station:

$$c_1^{\min} = 11p + 20(1-p) + \tfrac{60}{30} + 15 = 37-9p$$

$$\hat{c}_1 = c_1^{\min} + \tfrac{\alpha}{f(\text{line}11)} = 40-9p$$

$$c_2^{\min} = 11 + 6 + \tfrac{60}{30} + 15 = 34$$

$$\hat{c}_2 = c_2^{\min} + \tfrac{\alpha}{f(\text{line}2)} = 36$$

$$\hat{c}_3 = \frac{\alpha + f_{11}c_1^{\min} + f_2 c_2^{\min}}{f_{11} + f_2} = 36.4-3.6p$$

6.3. *Equilibrium analysis and nonuniqueness*

Let us first assess the domain where path 1 has cost less than path 2:

$$\hat{c}_1 \leq \hat{c}_2 \Leftrightarrow p \geq \tfrac{4}{9}$$

Then let us assess the domain where the combination, hyperpath 3, is better than any single path: the conditions are that

$$\begin{cases} c_1^{\min} \leq \hat{c}_2 \\ c_2^{\min} \leq \hat{c}_1 \end{cases} \Leftrightarrow p \in [\tfrac{1}{9}, \tfrac{2}{3}]$$

Thus the supply-demand equilibrium takes on the following states, depending on the value of parameter p:

$p < 1/9$: only path 2 is used.
$p = 1/9$: both path 2 and hyperpath 3 may be used.
$p \in](1/9), (2/3)[$: only hyperpath 3 is efficient, not path 1 or path 2 on a single basis.

DOI: 10.1002/atr

$p = 2/3$: both path 1 and hyperpath 3 may be used.
$p > 2/3$: only path 1 is used.

Knowing which hyperpaths are used, we may relate p to parameters κ, ξ and q:

When path 1 alone is used, then $p = \min\{1, \kappa/(\xi + q)\}$, so the requirement that $p > 2/3$ yields that $q < (3/2)\kappa - \xi$.

When path 2 alone is used, then $p = \max\{(\kappa - q)/\xi, 0\}$, so the requirement that $p < 1/9$ yields that $q > (\kappa - \xi)/9$.

If hyperpath 3 alone is used, then $p = (\kappa - \pi_2 q)/(\xi + \pi_1 q)$ where $\pi_2 = f_2/(f_2 + f_{11})$ and $\pi_1 = 1 - \pi_2$: the requirement that $1/9 < p < 2/3$ yields that

$$\frac{\kappa - \xi/9}{\pi_2 + \pi_1/9} > q > \frac{\kappa - 2\xi/3}{\pi_2 + 2\pi_1/3}$$

At states where several hyperpaths are used, the OD flow q splits into hyperpath flows x_i so that $p = (\kappa - \pi_2 x_3 - x_2)/(\xi + \pi_1 x_3 + x_1)$, which yields

$$p x_1 + x_2 + (p\pi_1 + \pi_2) x_3 = \kappa - p\xi$$

At $p = 2/3$, $x_2 = 0$ so $(2/3)x_1 + (\pi_2 + (2/3)\pi_1)x_3 = \kappa - (2/3)\xi$: the smallest q complying to this requirement is $(\kappa - (2/3)\xi)/(\pi_2 + (2/3)\pi_1)$ if all OD flow is carried by x_3 and none by x_1, and the largest is $(3/2)\kappa - \xi$ all carried by x_1.

At $p = 1/9$, $x_1 = 0$ so $x_2 + (\pi_2 + (1/9)\pi_1)x_3 = \kappa - (1/9)\xi$: the smallest q complying to this requirement is $\kappa - (1/9)\xi$ if all OD flows are carried by x_2 and none by x_3, and the largest is $(\kappa - (1/9)\xi)/(\pi_2 + (1/9)\pi_1)$ all carried by x_3.

At a given OD flow q, from one up to five traffic equilibria may exist:

- on $[0, (\kappa - 2\xi/3)/(\pi_2 + 2\pi_1/3)]$ there is one equilibrium state with $p > 2/3$;
- on $[(\kappa - 2\xi/3)/(\pi_2 + 2\pi_1/3), (3/2)\kappa - \xi]$ there are three equilibria, first with $p > 2/3$, second with $p = 2/3$ and last with $p \in](1/9), (2/3)[$;
- on $[(3/2)\kappa - \xi, \kappa - (1/9)\xi]$ there is one equilibrium with $p \in](1/9), (2/3)[$ (assuming a nonempty interval, i.e. $9\kappa < 16\xi$);
- on $[\kappa - (1/9)\xi, (\kappa - \xi/9)/(\pi_2 + \pi_1/9)]$ there are two equilibria with either $p \in](1/9), (2/3)[$ or $p = (1/9)$;
- on $[(\kappa - \xi/9)/(\pi_2 + \pi_1/9), \kappa]$ one equilibrium state with $p < 1/9$ (assuming a nonempty interval);
- the case where $[(\kappa - 2\xi/3)/(\pi_2 + 2\pi_1/3), (3/2)\kappa - \xi] \cap [\kappa - (1/9)\xi, (\kappa - \xi/9)/(\pi_2 + \pi_1/9)] \neq \varnothing$ might also arise, leading to five equilibria at any value of q within the intersection.

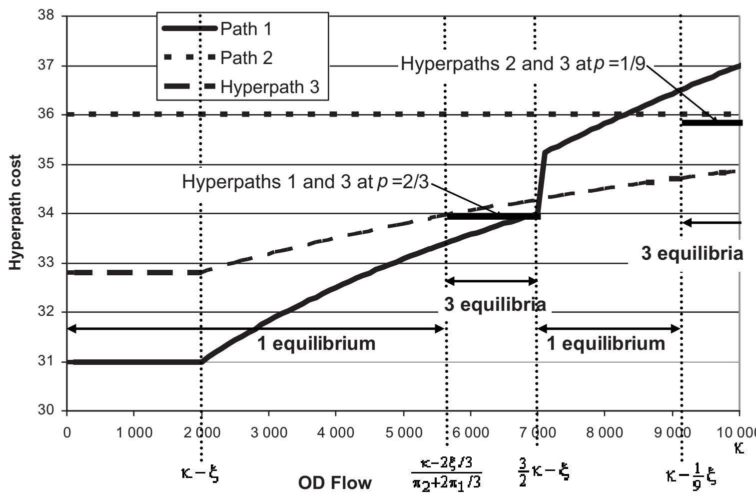


Figure 8. Equilibrium states in a cost-flow diagram.

DOI: 10.1002/atr

For $q > \kappa$ our study should be modified to take into account the seat congestion on the Nation—Châtelet segment.

Figure 8 depicts the hyperpath costs with respect to OD flow q, for parameter values of $\kappa = 10\,000$ and $\xi = 8000$. As $\pi_1 = 0.4$ and $\pi_2 = 0.6$, $(\kappa - \xi/9)/(\pi_2 + \pi_1/9) > \kappa$ so the regime with $p < 1/9$ falls out of the range $[0, \kappa]$ for q. From 0 to $(\kappa - 2\xi/3)/(\pi_2 + 2\pi_1/3)$ only path 1 is used. From $(\kappa - 2\xi/3)/(\pi_2 + 2\pi_1/3)$ to $(3/2)\kappa - \xi$, either path 1 alone, or hyperpath 3 alone, or both can support equilibrium. From $(3/2)\kappa - \xi$ to $\kappa - \xi/9$ only hyperpath 3 is used. From $\kappa - \xi/9$ to κ, either path 2 alone, or hyperpath 3 alone, or both can support equilibrium.

7. CONCLUSION

We have represented seat capacity in the basic framework for static traffic assignment to a transit network: namely the hyperpath model with line combination. The occupancy of the seat capacity influences the quality of service to the rider: by line segment the cost of being seated is less than that of standing. The competition of the riders to get a seat makes the seat allocation process a *random process* under priority rules, with sitting probabilities at each stage along the line. Each rider gets a *service mode* along his leg, with deterministic cost but allocated in a probabilistic way. We have defined line problems of, first, flow loading to yield the sitting probabilities and, second, leg costing to yield the leg average cost and cost variance. Each problem has been associated with a computationally efficient algorithm, with complexity of $O(S_\ell^2)$ to yield the results for all of the legs. The line algorithms amount to a complex cost–flow relationship which derives the average cost by leg from the entry–exit matrix of leg flows.

In a network setting, the leg costs influence the riders' route choice. Each leg has been represented by one network arc that can be included in a hyperpath to a given destination. Traffic equilibrium has been defined as a nonlinear complementary problem and cast into a VIP. Using a regularized cost–flow relationship, the existence of an equilibrium has been demonstrated. Uniqueness does not hold in general. A method of successive average has been provided as a heuristic equilibration algorithm, together with a theoretically sound convergence criterion on the basis of a duality gap. A simple formula has been given for the duality gap, which is generic for a class of static transit assignment models.

Further work on the seat congestion model might focus on the following issues:

- Identification of more comfort states, e.g. several types of seats or comfort classes.
- In-vehicle choice of egress station depending on the current user's comfort state.
- A discomfort function may be associated to each comfort state. Provided that seating is still more valuable than standing, the loading algorithm is unchanged and in the costing algorithm it is required to evaluate the discomfort function only once by segment, stage and comfort state.
- Inclusion of 'transaction' costs because of effort to get a seat; in conjunction with modelling other capacity constraints in transit such as access–egress capacity at vehicle dwelling in a station.
- Inclusion of costs which are nonlinear functions of the number of segments in leg or the leg travel time.
- Taste differentiation among the riders, so as to model heterogeneous trade-offs between state discomfort, travel time and also the fare.
- Discrete choice model to compete or not for a seat, so that the sitting probability might be varied among candidate riders. This would enable to model a higher share for longer legs.
- Discrete choice model of the user's willingness to obtain a given comfort state by associating a 'choice' probability to each comfort state.
- Priority rules for social or commercial purpose.

In the author's opinion, the discomfort function, taste differentiation and discrete choice extensions could yield some properties of equilibrium uniqueness.

ACKNOWLEDGEMENTS

Thanks go to Kaisheng Liu for useful comments about the network algorithms and the implementation work that he achieved in his doctoral thesis. The editorial suggestions of two anonymous referees have been very useful to improve clarity.

REFERENCES

1. TRB. *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://gulliver.trb.org/publications/tcrp/tcrp_webdoc_6-a.pdf First edition 1999. 2003.
2. Spiess H, Florian M. Optimal strategies: a new assignment model for transit networks. *Transportation Research B* 1989; **23**: 83–102.
3. Lam WHK, Zhou J, Sheng ZH. A capacity restraint transit assignment with elastic line frequency. *Transportation Research Part B* 2002; **36**: 919–938.
4. Tong CO, Wong SC. A stochastic transit assignment model using a dynamic schedule-based network. *Transportation Research B* 1999; **33**: 107–121.
5. Nuzzolo A, Russo F, Crisalli U, (2001) A Doubly Dynamic Schedule-based Assignment Model for Transit Networks. *Transportation Science* **35**: 268–285. DOI: 10.1287/trsc.35.3.268.10149
6. Nguyen S, Pallottino S, Malucelli F. A Modeling Framework for passenger assignment on a transport network with timetables. *Transportation Science* 2001; **35**: 238–249.
7. Poon MH, Wong SC, Tong CO. A dynamic schedule-based model for congested transit networks. *Transportation Research Part B* 2004; **38**: 343–368.
8. Meschini L, Gentile G, Papola N. A frequency-based transit model for dynamic traffic assignment to multimodal networks, Presented at the *17th International Symposium on Transportation and Traffic Theory*, London, July. 2007.
9. De Cea J, Fernandez E. Transit assignment for congested public transport systems: an equilibrium model. *Transportation Science* 1993; **27**: 133–147.
10. Cominetti R, Correa J. Common-lines and passenger assignment in congested transit networks. *Transportation Science* 2001; **35**: 250–267.
11. Cepeda M, Cominetti R, Florian M. A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. *Transportation Research Part B* 2006; **40**: 437–459.
12. Lam WHK, Gao ZY, Chan KS, Yang H. A stochastic user equilibrium assignment model for congested transit networks. *Transportation Research Part B* 1999; **33**: 351–368.
13. Kurauchi F, Bell MGH, Schmöcker JD. Capacity constrained transit assignment with common lines. *Journal of Mathematical Modelling and Algorithms* 2003; **2**: 309–327.
14. Shimamoto H, Iida Y, Schmöcker JD, Kurauchi F, Bell MGH. Evaluating public transit congestion mitigation measures using a passenger assignment model. *Journal of the Eastern Asia Society for Transportation Studies* 2005; **6**: 2076–2091. Available online at http://www.easts.info/on-line/journal_06/2076.pdf.
15. Schmöcker JD, Bell MGH, Kurauchi F. A quasi-dynamic capacity constrained frequency-based transit assignment model. *Transportation Research Part B* 2008; **42**: 925–945.
16. Hamdouch Y, Marcotte P, Nguyen S. Capacitated transit assignment with loading priorities. *Mathematical Programming B* 2004; **101**: 205–230.
17. Hamdouch Y, Lawphongpanich S. Schedule-based transit assignment model with travel strategies and capacity constraints. *Transportation Research Part B* 2008; **42**: 663–684.
18. Tian Q, Huang HJ, Yang H. Equilibrium properties of the morning peak-period commuting in a many-to-one mass transit system. *Transportation Research Part B* 2007; **41**: 616–631.
19. Leurent F. Confort et qualité de service en transport collectif urbain de voyageurs: Analyse, modélisation et évaluation. *Proceedings of the ATEC Congress, CD Rom edition*. ATEC: Paris, France, 2006a.
20. Leurent F. Modelling seat congestion in transit assignment, Working document available at: http://hal.archives-ouvertes.fr/hal-00348409/fr/. 2008.
21. Leurent F, Liu K. On seat congestion, passenger comfort and route choice in urban transit: a network equilibrium assignment model with application to Paris, TRB Meeting 2009, No. # 09–1784. 2009.
22. Sumalee A, Tan Z, Lam WHK. Dynamic stochastic transit assignment with explicit seat allocation model. *Transportation Research Part B* 2009; **43**: 895–912.
23. Hoogendoorn SP, Daamen W. Applying microscopic pedestrian flow simulation to railway design evaluation in Lisbon. *Transportation research record* 2005; **1878**: 83–94.
24. Spiess H. Contributions à la Théorie et aux Outils de Planification des Réseaux de Transport Urbain. Publication du Centre de Recherche sur les Transports, Université de Montréal. 1985.
25. Nguyen S, Pallotino S. Equilibrium traffic assignment in large scale transit networks. *European Journal of Operational Research* 1988; **37**: 176–186.
26. De Cea J, Fernandez E. Transit assignment to minimal routes: an efficient new algorithm. *Traffic Engineering and Control* 1989; **30**: 491–494.
27. Wu HJ, Florian M, Marcotte P. Transit equilibrium assignment: a model and solution algorithms. *Transportation Science* 1994; **28**: 193–203.
28. Debrincat L, Goldberg J, Duchateau H, Kroes E, Kouwenhowen M. Valorisation de la régularité des radiales ferrées en Ile de France. *Proceedings of the ATEC Congress, CD Rom edition*. ATEC: Paris, France. 2006.

APPENDIX A. HOW TO DEAL WITH A CIRCULAR LINE

A circular transit line ℓ is such that the vehicles serve each station 1 to S_ℓ along the line then again station 1 and so on. It cannot be assumed that there are no passengers in a vehicle when it comes to dwell at station 1, which makes the line loading a much more involved problem, whereas the line costing problem is not more difficult. Indeed, the line costing requires only one adaptation from the basic treatment in Section 3.4: when costing the legs (i, j) for all access stations i upstream of j, we have now to consider a list $\{j+1, \cdots S_\ell, 1, \cdots j-1\}$ of access stations.

To adapt the line loading algorithm and determine the sitting probabilities, let us denote by $\chi^\ell_{\geq i}$the total passenger flow on segment $(i, i+1)$ where $i+1$ denotes station 1 if $i = S_\ell$. As no passenger has interest in travelling the line in its entirety, we can safely assume that

$$\chi^\ell_{\geq i} = \sum\nolimits_{j,k} q^\ell_{jk} \delta_{j \leq i < k}$$

wherein $\delta_{j \leq i < k} = 1$ if i lies between j and k downstream of j and strictly upstream of k in the direction of the stream, or $\delta_{j \leq i < k} = 0$ otherwise.

If there is one station at which $\chi^\ell_{\geq i} \leq \kappa_\ell$, then station $i+1$ can be chosen as the initial station to begin the line loading algorithm, with Initialization stage replaced by

Initialization. Let $\underline{x}^0_{\geq i} := \chi(i)$, $\overline{x}^0_{\geq i} := 0$, $\underline{x}^{(i)0}_j := \sum_k q^\ell_{kj} \delta_{k \leq i < j}$ and $\overline{x}^{(i)0}_j := 0$.

If there is no station with vacant seat capacity left for incoming riders, then $p^+_i = 0$ for every i and all of the p^0_i satisfy that $p^0_i y^0_i = \kappa^0_i$.

The seated riders who exit at station i vacate a residual capacity equal to their flow which is

$$\kappa^0_i = \sum_j q^\ell_{ji} \prod_{k=j+1}^{i-1} (1 - p^0_k).$$

As all of this is occupied with sitting probability p^0_i by the standing riders who remain on board, it holds that

$$p^0_i \sum_{k<i<j} q^\ell_{kj} \prod_{m=k+1}^{i-1} (1 - p^0_m) = \sum_j q^\ell_{ji} \prod_{k=j+1}^{i-1} (1 - p^0_k).$$

This makes a system of S_ℓ nonlinear equations in S_ℓ variables p^0_i. It can be solved as a fixed-point problem

$\mathbf{p}^0 = \mathbf{F}(\mathbf{p}^0)$, in which

$$\mathbf{F}_i(\mathbf{p}^0) = \frac{\sum_j q^\ell_{ji} \prod_{k=j+1}^{i-1} \left(1 - p^0_k\right)}{\sum_{k<i<j} q^\ell_{kj} \prod_{m=k+1}^{i-1} \left(1 - p^0_m\right)}$$

Solving this problem yields the sitting probabilities p^0_i.

DOI: 10.1002/atr