ParisTech
GRANDES ECOLES D'INGENIEURS DE PARIS
PARIS INSTITUTE OF TECHNOLOGY

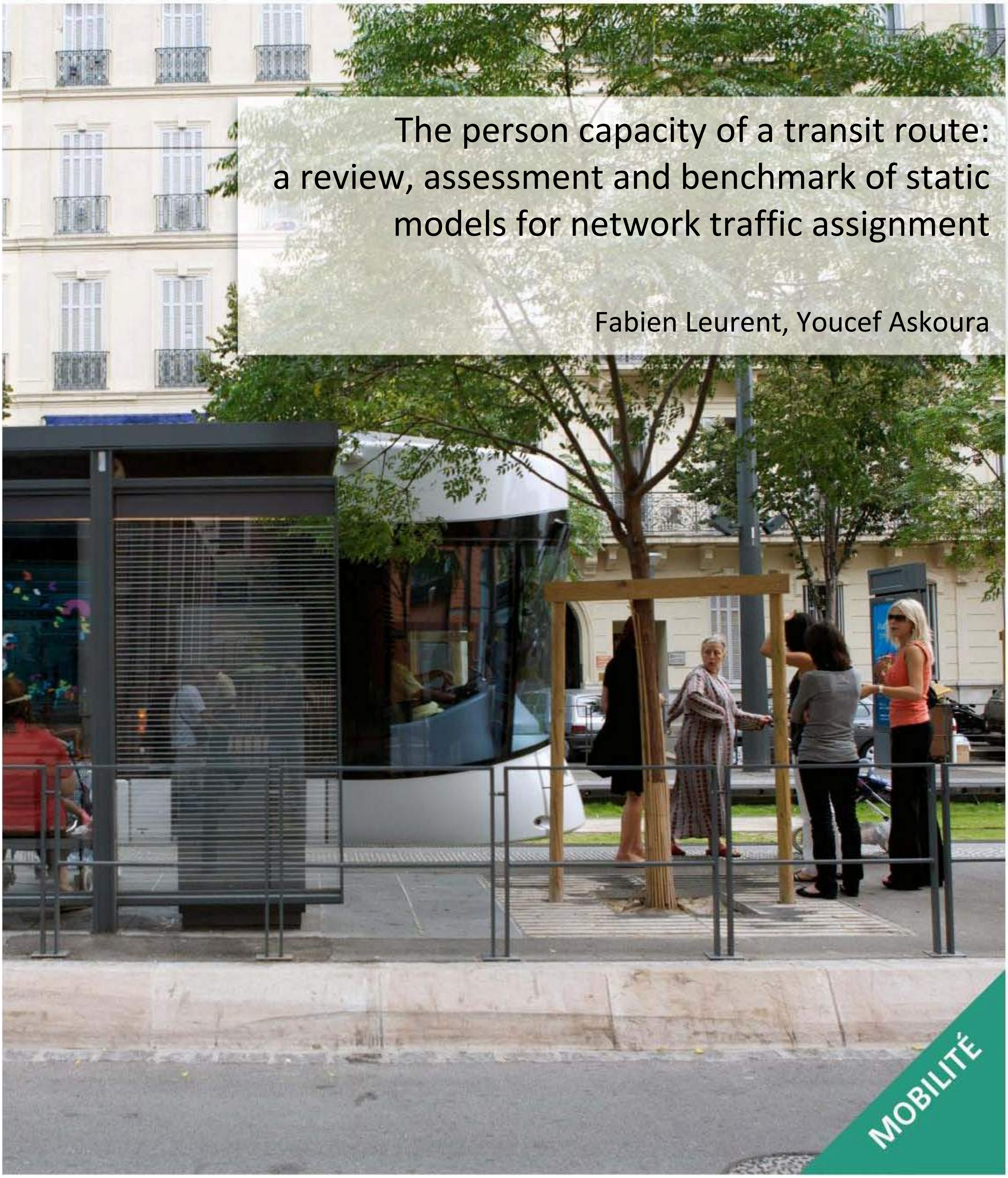
En partenariat avec VINCI

Chaire ParisTech

Eco-conception des ensembles bâtis et des infrastructures

The person capacity of a transit route: a review, assessment and benchmark of static models for network traffic assignment

Fabien Leurent, Youcef Askoura



MOBILITÉ

PAPER ACCEPTED FOR PRESENTATION AT THE 89th TRB MEETING, 2010, AND SUBMITTED FOR PUBLICATION IN TRB JOURNAL

PAPER **TRB # 10-2020**

The person capacity of a transit route: a review, assessment and benchmark of static models for network traffic assignment

AUTHORS' NAMES

Fabien LEURENT, Youcef ASKOURA

SUBMISSION DATE August 1st, 2009

REVISED VERSION November 13, 2009

AUTHOR 1 (CORRESPONDING AUTHOR)

Prof. Fabien LEURENT

Université Paris-Est, Laboratoire Ville Mobilité Transport

19 rue Alfred Nobel, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France

Tel: (33) 164 152 111

Fax: (33) 164 152 140

e-mail: fabien.leurent@enpc.fr

AUTHOR 2

Youcef ASKOURA

Laboratoire Ville Mobilité Transport

19 rue Alfred Nobel, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France

Tel: (33) 164 152 106

Fax: (33) 164 152 140

e-mail: youcef.askoura@enpc.fr

WORD COUNT: 8,000

PLUS 3 FIGURES PLUS 1 TABLE

Fabien Leurent, Y Askoura, LVMT Paris, August 27th 2009

The person capacity of a transit route: a review, assessment and benchmark of static models for network traffic assignment

ABSTRACT

The planning of urban public passenger transport often requires considering the capacity constraints and congestion effects. Modelling the person capacity of a transit route has been the purpose of several recent research works to develop network traffic assignment models along the following tracks: (i) effective frequencies, by De Cea and Fernandez (*1*) and Cepeda et al (*2*); (ii) the capacity by route segment, by Lam et al (*3*); (iii) failure-to-board probability, by Kurauchi et al (*4*); (iv) strategy based on a user preference set, by Hamdouch et al (*5*) ; (v) availability frequency, by Leurent and Askoura (*6*).

Our objective is to describe and compare the models. We characterise and discuss their assumptions and also apply them to a test case which is treated parametrically. We consider specifically: (1) making explicit a capacity constraint; (2) priority for passengers already on board over those boarding; (3) the route's waiting time for a boarding passenger; (4) the distribution of boarding volumes between the attractive routes from a station to a given destination.

1 INTRODUCTION

1.1 Background

Public passenger transport systems have limited capacities for two main reasons: (C1) the number of vehicles and the length of their routes limit the frequency of service whilst (C2) the frequency of service combined with vehicle passenger capacity limits the number of places available to travellers per transit route and per time period. In addition to these basic constraints, there are local constraints with local effects and potential effects by network propagation: (C3) the volume of transit vehicles on a route is limited by the traffic capacity of the given section (on roads or railways) and by the vehicle sojourn capacity at platforms; (C4) the passenger volume loaded in a vehicle is constrained by the length of time it is dwelled at a platform (dwell time) and by the time taken for each passenger to board or exit, which jointly determine the exchange capacity between the service (vehicle) and the platform; (C5) the stock of passengers waiting on a platform is limited by the area of the allocated space; (C6) the circulation of passengers within a station is limited by station access capacity, corridor capacity and other pedestrian traffic factors. All these capacity constraints, involving both vehicles and passengers, are described in detail in the *Transit Capacity and Quality of Service Manual (TCQSM)* from the TRB (*7*), which also deals with even more complicated interactions which reduce capacities. Specifically: (C7) the presence of a stock of passengers at a stop which is not emptied when a vehicle dwells will increase the dwell time thereby slowing the vehicles and reducing service frequency; additionally (C8) local interferences between passenger stocks and flows using different routes constrains their circulation in the station, makes access to the service more difficult and possibly also hinders access to vehicles.

This set (C1 - 8) of capacity effects has to be considered and addressed specifically by the transit network operator and this becomes more important as the volume of passengers and the frequency of services becomes greater - or should be greater if they were not impeded by the congestion.

In addition to making field observations, which are essential for determining the magnitude of these phenomena, the operator has to deal with congestion by adapting its operating methods and means (resources). In order to prepare an operating plan, the operator can simulate projects and operating scenarios in terms of service usage and effects on passengers by using a model to assign traffic to the network's paths.

1.2 Purpose

In a previous article (*8*), we compared the transit system described by the TCQSM with representations from traffic assignment models. We have identified static models that incorporate some capacity effects (C1-3) but not the more complicated effects (C4-8). We also have found dynamic, macroscopic, schedule based models that deal with a single capacity effect: the vehicle's passenger carrying capacity (C2), with methods that are also applicable to (C6). In fact, these dynamic models simply extend the static processing of the (C2) capacity effect, the passenger capacity of a transit route which appears to be the main target for modelling efforts.

This article focuses on the passenger capacity of a transit route. Our aim is to present the different models, make explicit and discuss their assumptions and determine their strengths and weaknesses. We will consider specifically the following features in each model:

(1) the explicit or implicit representation of the capacity constraint.

(2) the priority of passengers on board over boarding passengers: a specific aspect of a transit model but analogous to flow priorities at a road junction without traffic lights.

(3) the waiting time for a route for a boarding passenger: the passenger waits for a vehicle to arrive and for a place to be available for him/her, according to the number of waiting passengers.

(4) the distribution of boarding passenger volumes in a station between the attractive routes for a given destination.

Our analysis will consider five distinct models which are:

- (i) effective frequencies, by De Cea and Fernandez (*1*) and Cepeda et al (*2*);
- (ii) capacity by route segment, by Lam et al (*3*);
- (iii) failure-to-board probability, by Kurauchi et al (*4*);
- (iv) strategy based on a user preference set, by Hamdouch et al (*5*);
- (v) availability frequency, by Leurent et Askoura (*6*).

1.3 Method

Our objective mainly concerns the physical and economic outreach of each model. Such a representation involves a set of assumptions that make components of physical or economic significance: the distinction of routes, the explicit representation of vehicle frequency and the passenger loading, the formation of passenger's waiting time, the formation of a stock of waiting passengers and the passenger's economic considerations and behaviour when selecting a service to take them towards their destination. For each model we have considered the basic theory and the mathematical formulae that specify each component. We do not report on the mathematical treatment to deal with a whole network and multiple destinations because this has become standardised with recursive equations to yield the costs for hyper-paths and to distribute volumes, a variational inequation to express the traffic supply-demand equilibrium and, usually, a method of successive averages in order to calculate this equilibrium (e.g.: Spiess et Florian, *9*).

To make the model descriptions more concrete and the comparisons clearer, we have built a test case which involves various components and allows each modelling approach to be characterised. We apply each model to this case by adapting its formulae but keeping their analytical form in order to consider a set of traffic loading situations and determine how the model behaves under different origin-destination volumes.

1.4 Structure

The body of this article consists of three main parts and a conclusion. Section 2 consists of an extensive bibliographic review of transit assignment models that deal with person capacity by service route. Section 3 deals with the test case: having specified the structure, we apply first the model without capacity constraints, namely the optimal strategy model of Spiess and Florian (*9*) which is the standard method for static transit assignment, then the models with capacity constraints. In section 4, we compare the models on numerical grounds and from a theoretical perspective. Section 5 summarises the outcomes and suggests areas for future work.

2 Bibliographic review of models

2.1 Five Generations of Transit Traffic Assignment Models

Models for assigning passenger flows on transit networks started being developed during the '60s. In a primitive generation, models from Dial (*10*), Fearnside and Draper (*11*) and Last and Leak (*12*) dealt pragmatically with two fundamental specificities of transit: (i) the fragmentary availability of the service for a passenger in a station; and (ii) the possibility for a passenger to combine several routes that are apt to bring them nearer to their destination in order to reduce journey time by shortening waiting time.

A second generation of models dealt with these specificities in detail by capturing stochastic phenomena (passenger arrivals and vehicle arrivals) and specific economic behaviours for passenger's choice of itinerary. Chriqui and Robillard (*13*) dealt with line combination at the local level between two stations linked by parallel lines. This was generalized to a network structure by Spiess and Florian (*9*), who introduced the concept of optimal strategy for passenger behaviour and developed efficient algorithms that (i) find the best routing structure with a local routing on the basis of dynamic service opportunities and (ii) load traffic onto the resulting routing structure.

Nguyen and Pallotino (*14*) linked this model to a bigger mathematical framework that they specifically developed: the theory of hyper-paths on a network. A hyper-path is a cycle-free, connected sub-graph which heads towards a given destination. De Cea et al (*1*) accommodated the 'optimal route' model in this framework by restricting the set of hyper-paths to the sub-set of those hyper-paths composed as a sequence of transit stations between a passenger's origin and destination.

There have been three subsequent generations of models: firstly, the development of static macroscopic models to deal with capacity effects; secondly, the development of dynamic macroscopic models to deal with the timing of phenomena and better capture capacity effects; thirdly, the development of microscopic models aimed at dynamic simulations of vehicle and passenger movements on a network with a very fine description and a representation of random aspects. These generations co-exist because static macroscopic models are still the tools that are most commonly used for planning purposes – see Cepeda et al (*2*), Shimamoto et al (*15*), Leurent and Liu (*16*), whilst dynamic macroscopic models are used for more advanced but still limited sub-network studies (Meschini et al (*17*), Hamdouch and Lawphongpanich (*18*)), whilst microscopic simulators are used to study network problems with focus either on vehicle traffic along the network or passenger traffic within a station (Hoogendorn).

2.2 Static Models for Vehicle Passenger Capacity

Let us come to vehicle passenger capacity. In a static model, this capacity is defined at the service route level by adding together the capacities of the vehicle serving the route during the assignment period. If these vehicles are identical and have an identical traffic load, then the residual capacity for a route that is available to board passengers from a station is equal to the residual capacity per vehicle multiplied by the number of vehicles in the period.

Last and Leak (*12*) developed the TRANSEPT multi-modal assignment model, in which the passenger capacity is limited in the vehicles on a route. The passenger loading of the available capacity operates from the origin of the route: at stations where the residual capacity is less than the stock of waiting passengers, only the partial volume compatible with the residual capacity boards whilst the rest of the passenger stock changes to other options. This method complies with the priority of passengers who boarded upstream on the route over those who want to board downstream and relates waiting time to the probability that a passenger will fail to board the first vehicle to arrive. However, the possibility of waiting at the station for the next vehicle is excluded and, consequently, waiting time is under-estimated. The model also fails to consider the distribution of traffic between several attractive routes.

De Cea and Fernandez (*1*) modelled vehicle congestion by linking waiting time before boarding to the volume wanting to board and the volume already aboard. This is why they replaced the route's nominal frequency by the '*effective frequency*' in the route boarding arc, a function that decreases with the two volumes involved. This effective frequency model has also been developed by Wu et al (*19*), Cominetti and Correa (*20*) and Cepeda et al (*2*), and has the advantage of easily incorporating the effect of congestion without perturbing the model's other components. The name of effective frequency could lead to misunderstandings because it is not a frequency that would represent the operational functioning of the system better than the nominal frequency but is instead a virtual, ad-hoc frequency adapted locally to each station in order to simulate the waiting time for the relevant route. The function used by Cepeda et al (*2*) is a mathematical artefact and not a mathematical expression of a physical phenomenon, despite the Markovian model proposed by Cominetti and Correa (*20*). In addition, their functional specification reduces the frequency and capacity offered at all load rates including very low traffic rates which is unrealistic and could result in uncontrolled effects on the distribution of volumes across the network.

Lam et al (*3*) explained the vehicle passenger limits using a quantitative constraint on each section between two consecutive stations on a route: these constraints are added to the system of equations that characterise traffic equilibrium. A dual variable is linked to each constraint, the '*penalty*', which is zero if the limit is not reached or positive otherwise and represents an additional time for the passenger between stations. The benefits of this are an explicit representation of such an additional cost, its determination using the principle of user equilibrium (by comparison of competing itineraries) and its incorporation into the itinerary cost for the passenger. However, the extra cost is applied to every passenger on board, not just the last ones to board; this means it is more a cost of discomfort in the vehicle rather than a cost of the waiting time before boarding.

To better differentiate the time passengers spend in the vehicle or waiting to board, in a similar way to Transept, Kurauchi et al (*4*) related a route's residual capacity after passengers had descended at a station to the volume wanting to board. The ratio capped at 1 is the probability that a waiting passenger will be able to board the first vehicle to arrive whilst the complementary probability corresponds to '*wait for a later vehicle*'. This is the *failure-to-board* probability or p_F. Under a stationary traffic regime, if the stock of waiting passengers is not organised into a queue, then the probability of a passenger boarding each time a vehicle arrives is $1-p_F$.

Kurauchi et al (*4*) initially modelled the diversion of traffic exceeding the capacity on an 'escape arc' leading from the current station to the destination: this method locally constrained the volume to satisfy the capacity constraint but altered the volume on the network for stations situated downstream along the route. Furthermore, the FtB cost was not directly modelled using a physical description of the process: only the probability of failure was included in a function of generalised cost of travel to the destination.

Shimamoto et al (*15*) subsequently included waiting time related to the possibility of failure-to-board in the route cost: this is more realistic and even though there is an excess volume compared to capacity, this volume is reduced. Depending on the time H_z between two vehicles, the passenger has an average waiting time before boarding of d_B such that

$$\begin{aligned} d_B &= (1-p_F).0 + p_F(1-p_F)H_z + p_F^2(1-p_F)(2H_z) + ... + p_F^m(1-p_F)(mH_z)... \\ &= (1-p_F)p_F H_z \sum_{m=0}^{\infty} m p_F^{m-1} = \frac{p_F}{1-p_F} H_z \end{aligned} \quad (1)$$

This waiting time is added to the journey time and waiting time for a route (conditional on boarding the next vehicle). This yields a route cost between the current point and the destination: if several routes are attractive at the current point, this cost has to be calculated taking account of those of the other routes and their frequencies. However, whilst waiting for one route, another route may become available and, if they are attractive and available, a passenger can decide to reduce waiting time, thereby not experiencing all the d_B. Shimamoto et al (*15*) retained the 'no interaction' formula of d_B to evaluate the boarding wait in the composite cost formula, which is only an approximate forecast.

Hamdouch, Marcotte and Nguyen (*5*) proposed a 'strategic' model that we refer to as the User Preference Set (UPS) model because they have their own definition of strategy: each current node is linked with an ordered list of successor nodes towards the destination, a successor node being the head node of an attractive arc chosen from the arcs exiting the current node. It is assumed that the passenger flow is oriented towards attractive arcs as long as there is available capacity then to the next arc in order of preference and so on and so forth. This induces proportions of local assignment that are analogous to the local routing proportions in a hyper-path. Then the strategy costs are derived from the path costs and their respective probabilities. Consequently, the UPS approach is based on the probability of boarding success which makes it similar to the probability of failure-to-board method. However, the model has the following difficulties. Firstly, there is no variability in the route costs unlike the strategy model proposed by Spiess and Florian (*9*): why oblige a passenger to use another itinerary rather than wait and experience a waiting cost? The solution to this weakness is the main addition provided in the dynamic version of the model by Hamdouch et Lawphongpanich (*18*). Secondly, the model is applied to the transit network using only the priority of passengers already on board over those waiting to board: this is certainly a very important effect and has to be represented but the combined treatment fails to deal with the differing discrete availability of routes and the possibility of combining frequencies! This creates a problem especially when there is no saturation because the model is reduced to assigning the optimal path based on average waiting times. Thirdly, the model is applied to an acyclical network: the combined treatment of several destinations is not explained whilst the list of ordered nodes at any node can change with the destination and this causes ambiguities when determining the probability of boarding success.

Tian et al (*21*) 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 (*22*) 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, the segment cost from the boarding to descending station is a random variable with structural dependency on the seating capacity and the origin-destination (by segment) matrix of passenger flows. Leurent (*23*, *16*) extended this model of seating capacity to a general network, in an equilibrium framework. He did not address the service capacity in terms of both standing and seated passengers.

Leurent and Askoura (*6*) modelled availability frequency at a station for boarding a route. The model's principle is as follows: if the volume of people wanting to board exceeds the capacity, not only does the boarding volume saturate the capacity, i.e. equals it, but also a new stock of waiting passengers is created for the route. Each passenger experiences a waiting cost for the next vehicle to arrive if he/she has had to let a full vehicle go by: this is the pass-up situation. This waiting time depends on two variables: the first is the penalty, the dual variable linked to the capacity constraint which is similar to that of Lam et al (*3*) but for a boarding arc rather than an on board arc. The second variable is the frequency reduction, the frequency decline δ_a^s per destination s, which is subtracted from the nominal frequency f_a for the route a in vehicles per period to yield a reduced frequency of $\tilde{f}_a^s \equiv f_a - \delta_a^s$, which is interpreted as the frequency with places available for a passenger who wants to board. When not saturated, the frequency decline is zero and the availability frequency is equal to the nominal frequency which avoids some of the problems found in effective frequency models. The ratio between the frequency decline and the nominal frequency is the failure-to-board probability. This means that the availability frequency model combines characteristic features from various preceding approaches.

3 A TEST CASE FOR USING THE MODELS

To confirm our statements made in the bibliographic review of the models, we built a simple case on which to use and compare the models. We started by specifying the application data, which is fixed in terms of the network and services but variable in terms of demand volumes, in order to analyse the sensitivity to traffic loads (§ 3.1). We then processed the case without capacity constraints (§ 3.2) then apply, in succession, the Lam model with penalties by inter-station segment (§ 3.3), The FF effective frequency model (§ 3.4), the FtB method using failure-to-board probability (§ 3.5), the UPS user preference model (§ 3.6), and the AF model using availability frequency (§ 3.7).

3.1 Case Design and Parameters

A transit network with three stations, A, B and D respectively, served by four transit lines: (1) direct line from A to D, (2) from A to D via B, (3) from B to D and (4) the same.

Each line z has a nominal frequency f_z and total capacity κ_z as specified in figure 2, and a time between stations t_z^{nm} from node n to node m. For each station S, a node S_0 represents a platform where passengers stop or select their next route; a node S_z^- represents descending from route z in station S coming from upstream, whilst a node S_z^+ represents boarding a route z towards a destination. The sojourn time on arc (S_z^-, S_z^+) when the vehicle dwells in the station is ignored. By assumption the routes are operated independently of each other; each one is operated assuming no time memory (i.e. exponential vehicle headway) so the waiting time per passenger is equal to $w_z = 1/f_z$.

The demand consists of two origin-destination (O-D) volumes, one from A to D denoted by q_A and the other from B to D denoted by q_B. Passengers have a homogeneous perception of the travel conditions and we assumed that the generalized cost of travel is reduced to average physical time.

By combining routes z into a sub-set Z', the combined wait is reduced to $w_{Z'} = 1/f_{Z'}$ with $f_{Z'} = \sum_{z \in Z'} f_z$, and by taking the first line to arrive amongst Z' the passenger has an average time of $G_{Z'} = w_{Z'} + \sum_{z \in Z'} \frac{f_z}{f_{Z'}} T_z$, with T_z the time via z from the vehicle's departure to the passenger’s destination.

We have kept the supply variables $(f_z, \kappa_z, t_z^{nm})$ fixed in order to explore:

- for low values of q_A, low then high values of q_B, in order to saturate line 2 from B to D then line 3 and use line 4 as a last resort from B.
- if necessary, we then increase volume q_A in order to saturate line 2 and make line 2 unavailable at B. Line 1, which is not saturated, then becomes the system's 'thermostat' for the origin-destination pair from A to D.

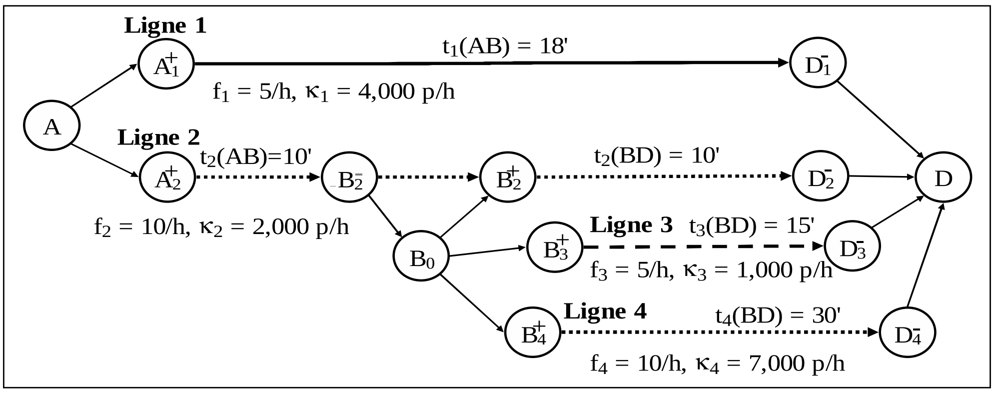


FIGURE 1 Case network

3.2 Model Without Congestion

Lines 2 and 3 are attractive at B: their intervals of minimum and average time $[t_z^{\mathrm{BD}}, t_z^{\mathrm{BD}} + w_z]$ cover each other because they are, respectively, [10’, 16’] for line 2 and [15’, 27’] for line 3. Line 4 is not attractive because of its interval of [30’, 36’]. The combined frequency is $f_{\mathrm{B}} = f_2 + f_3$ = 15/h. The volume q_{B} boards line 2 in a proportion of $h_{\mathrm{B2}} = f_2 / f_B$ of 2/3 and line 3 in a proportion of $h_{\mathrm{B3}} = f_3 / f_B$ of 1/3. The average combined waiting time is $w_{\mathrm{B}} = 1/ f_{\mathrm{B}}$ = 4’ and the average time per passenger from B to D is

$$G_{\mathrm{BD}} = w_{\mathrm{B}} + \frac{f_2}{f_{\mathrm{B}}} t_2^{\mathrm{BD}} + \frac{f_3}{f_{\mathrm{B}}} t_3^{\mathrm{BD}} = 15'\frac{2}{3}.$$

Similarly, at A, lines 1 and 2 have intervals from minimum to average cost of, respectively, [18’, 30’] for line 1 and [20’, 26’] for line 2 making both attractive. The combined frequency is $f_{\mathrm{A}} = f_1 + f_2$ = 15/h. The volume q_{A} boards line 1 in a proportion of 1/3 and line 2 in a proportion of 2/3. The average combined waiting time is w_{A} = 4’ and the average time per passenger from A to D is

$$G_{\mathrm{AD}} = w_{\mathrm{A}} + \frac{f_1}{f_{\mathrm{A}}} t_1^{\mathrm{AD}} + \frac{f_2}{f_{\mathrm{A}}} t_2^{\mathrm{AD}} = 23'\frac{1}{3}.$$

Figure 2 shows the local routing ratios on the network which define the traffic assignment if we ship volumes q_{A} at A and q_{B} at B and transport them to D.

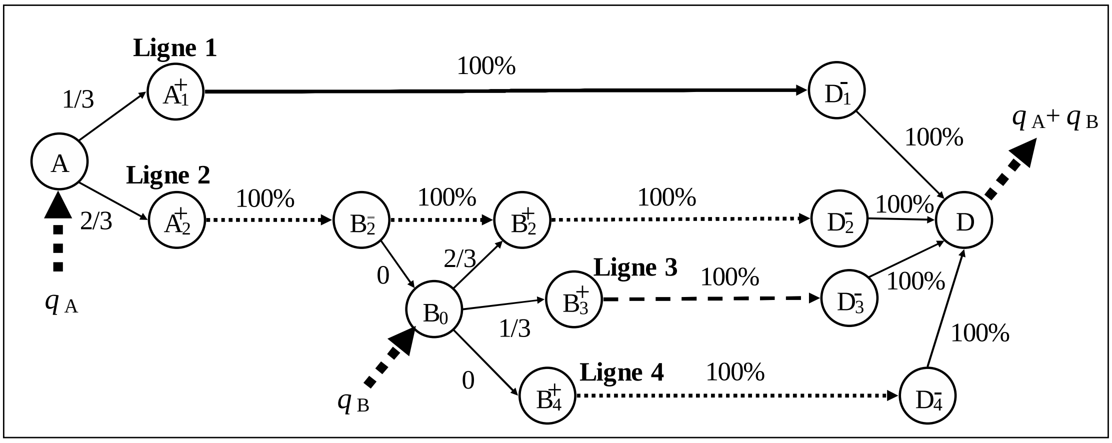


FIGURE 2 Local routing proportions, model without congestion

3.3 “Lam” Model With Penalties by Segment Between Stations

Lam et al (*3*) linked a capacity constraint to vehicle passenger capacity for each inter-station arc and not for the route boarding arcs. In our case, the volume from A on line 2 between B and D is subject to a capacity constraint as is that which boards at B, in other words, the passengers that are already on board would not have precedence over boarding passengers.

Let us derive the consequences of this in our case, denoting by γ_2^{BD} the penalty associated to constraint $q_{\mathrm{A}} \frac{f_2}{f_{\mathrm{A}}} + q_{\mathrm{B}} \frac{f_2}{f_{\mathrm{B}}} \leq \kappa_2$. When the constraint is saturated, the cost interval for line 2 becomes $t_2^{\mathrm{BD}}+\gamma_2^{\mathrm{BD}}+[0,w_2]$ at B where route 2 is only attractive if its minimum cost stays below or equal to the average cost of route 3, $w_3+t_3^{\mathrm{BD}}$.

The maximum penalty which maintains the attractiveness of line 2 at B is

$$\tilde{\gamma}_2^{\mathrm{BD}} \approx w_3+t_3^{\mathrm{BD}}-t_2^{\mathrm{BD}} = 17'.$$

But if we assume that the volume from A is subject to a local penalty, the overal costs of line 2 becomes $(w_2+)t_2^{\mathrm{AB}}+t_2^{\mathrm{BD}}+\gamma_2^{\mathrm{BD}}$ with a cost interval of γ_2^{BD}+[20', 26'] which is not more competitive when γ_2^{BD} = 17' compared to line 1 which is assumed not to be saturated. In this event the volume from A would not use line 2, as if it was yielding priority to the volume from B: but it should hold the priority! The limit value $\hat{\gamma}_2^{\mathrm{A}}$ for maintaining line 2 attractiveness at A is imposed by the average time of line 1 at A: $G_1^{\mathrm{AD}}=w_1+t_1^{\mathrm{AD}}$ = 30', through $\hat{\gamma}_2^{\mathrm{A}}=G_1^{\mathrm{AD}}-t_2^{\mathrm{AD}}$ = 10'. Beyond this value, the model is no longer consistent with the priorities between passenger volumes.

By setting $q_{\mathrm{A}} =$ 1,500 p/h, we can study the state of the system with respect to q_{B}. When route 2 is not saturated, q_{A} is assigned at 1/3 on line 1 and 2/3 on line 2, making $v_{2\mathrm{A}} =$ 1,000 p/h and leaving a residual capacity of $\kappa'_2 =$ 1,000 p/h on this line. This capacity is sufficient to allow all q_{B} to be distributed between lines 2 and 3 up to a value of

$$\tilde{q}_{\mathrm{B}}^1=\frac{f_2+f_3}{f_2}\kappa'_2 = 1{,}500 \text{ p/h}.$$

Beyond $\tilde{q}_{\mathrm{B}}^1$, line 2 should be saturated between B and D. If $\gamma_2^{\mathrm{BD}} \geq 10$ then all the capacity κ_2 is available for q_{B}, but this flow can only saturate capacity if

$$q_{\mathrm{B}} \geq \tilde{q}_{\mathrm{B}}^2 \equiv \frac{f_2+f_3}{f_2}\kappa_2 = 3{,}000 \text{ p/h}.$$

On $[\tilde{q}_{\mathrm{B}}^1,\tilde{q}_{\mathrm{B}}^2]$ the volume from B can not board line 2, which requires $\gamma_2^{\mathrm{BD}} \geq 17$ in order to make it unattractive but this also would prevent q_{A} from using the line: The model is inconsistent for this interval.

For $q_{\mathrm{B}} \geq \tilde{q}_{\mathrm{B}}^2$, q_{B} saturates line 2 with penalty $\gamma_2^{\mathrm{BD}} \geq 0$ and also line 3 with penalty $\gamma_3^{\mathrm{BD}} \geq 0$, because $\tilde{q}_{\mathrm{B}}^2 \frac{f_3}{f_2+f_3}=\kappa_3$. The requirements about penalties are that: $\gamma_2^{\mathrm{BD}} \geq 10$ so as to exclude q_A, and $\gamma_2^{\mathrm{BD}} \leq 17-\gamma_3^{\mathrm{BD}}$. This model is under-determined but as the penalties are experienced by users on the origin-destination pair from B to D we can assume that the collective behaviour minimises them and retain $\gamma_2^{\mathrm{BD}}=10$ and $\gamma_3^{\mathrm{BD}}=0$. This fixes

$$G_2^{\mathrm{BD}}=\frac{1+f_2(t_2+\gamma_2)+f_3t_3}{f_2+f_3} = 22'\frac{1}{3} \text{ at point } q_{\mathrm{B}}=\tilde{q}_{\mathrm{B}}^2.$$

For $q_B > \tilde{q}_B^2$, line 4 has to be made competitive at B which requires penalties γ_2 and γ_3 such that

$$\frac{1+f_2(t_2+\gamma_2)+f_3(t_3+\gamma_3)}{f_2+f_3} \geq t_4\,,$$

in addition to the preceding constraints. The minimum numerical values are $\gamma_2 = 17'\frac{2}{3}$ and $\gamma_3 = 7'\frac{2}{3}$ with an average cost $G_2^{BD} = t_4 =$ 30'.

The $\{2, 3, 4\}$ combination of lines can accommodate q_B by saturating line 2 if $q_B \geq \tilde{q}_B^3 \equiv \frac{f_2+f_3+f_4}{f_2}\kappa_2$ has a value of $\tilde{q}_B^3 =$ 5,000 p/h. On $]\tilde{q}_B^2, \tilde{q}_B^3]$, the volume is distributed between the combinations {2,3} and {2,3,4}, which conserves the link volumes v_{2B} and v_{3B}.

Beyond $\tilde{q}_B^3$, as lines 2 and 3 are saturated, q_B has to be spread between {2,3,4} and {4}, which requires

$$\frac{1+f_2(t_2+\gamma_2)+f_3(t_3+\gamma_3)+f_4t_4}{f_2+f_3+f_4} \geq t_4 + \frac{1}{f_4}\,,$$

in addition to the preceding constraints. The minimum numerical values are $\gamma_2 = 26'$ and $\gamma_3 =$ 21', for an average cost of $G_2^{BD} = t_4 + \frac{1}{f_4} =$ 36'.

In summary, capacity constraints on inter-station arcs result in paradoxes and unrealistic effects on volume distribution and the allocation of capacity by erroneously ignoring the volume priorities between those passengers on board and those boarding.

3.4 "FF" Effective Frequency Model

In the effective frequency method, a volume wanting to board is faced to the boarding capacity which is equal to the difference between nominal capacity and the volume that remains on board (which has priority to continue its journey). If the residual capacity is insufficient, then boarding passengers have a boarding delay that is modelled by

$$d_{B2}^e = 1/\mathrm{F}_{2B}$$

depending on the effective frequency $\mathrm{F}_{2B} \leq f_2$, where d_{B2}^e includes the service's average waiting time.

De Cea and Fernandez (*1*) suggested a two argument function: the boarding volume v_a^+ and transit volume v_a^o using the formula

$$\mathrm{F}_a(v_a^+, v_a^o) = f_a[1+\beta(\frac{v_a^+}{\kappa_a - v_a^o})^\alpha]^{-1} \text{ with } \alpha, \beta > 0 \text{ for } v_a^+ < \kappa_a - v_a^o \text{ and}$$

$$\mathrm{F}_a = 0 \text{ for } v_a^+ \geq \kappa_a - v_a^o .$$

This formula involves every flow volume which is a source of two major inconveniences: not only does it effect low volumes, which is unrealistic, but its activation does not require that volumes are approaching the residual capacity. This provokes the propagation of unrealistic volumes upstream and of unrealistic costs downstream in network assignments. The same comment also applies to the function suggested by Cepeda et al (*2*): zero for $v_a^+ \geq \kappa_a - v_a^o$ and for $v_a^+ < \kappa_a - v_a^o$, $\mathrm{F}_a(v_a^+, v_a^o) = f_a[1-(\frac{v_a^+}{\kappa_a - v_a^o})^\alpha]$ with $\alpha > 0$.

To correct these twin problems, we suggest limiting the frequency reduction to one step $[\pi_a, 1]$ in the ratio between the volume and the residual capacity, with π_a at a level of 80% or 90%. Here is a formula of the function with an exponent of $\alpha \geq 1$:

$$\mathrm{F}_a(v_a^+, v_a^o) = f_a[1-\{(\tfrac{\min\{\rho,1\}-\pi_a}{1-\pi_a})^+\}^{\alpha_a}] \text{ with } (y)^+ = \max\{y,0\}$$

$$\text{and } \rho = \frac{v_a^+}{\kappa_a - v_a^o} \text{ or } \rho = \frac{v_a^+ + v_a^o}{\kappa_a}.$$

This specification ensures that the frequency remains unchanged as long as $\rho \leq \pi$. Consequently, the frequency reduction only becomes active when volumes exceed a proportion π of the residual capacity. This linkage between the respective states of frequency reduction and load rate ρ makes the time effect on waiting delay analogous to a boarding penalty.

Let us consider for the purpose of the illustration that $\rho = v_a^+ / \kappa'_a$ with $\kappa'_a \equiv \kappa_a - v_a^o$ and set $\pi = 90\%$ and $\alpha = 1$. For q_A = 1,500 p/h, $v_{2\mathrm{B}}^o = \frac{2}{3} q_\mathrm{A}$ = 1,000 p/h, which leaves a residual capacity for boarding at B on line 2 $\kappa'_2 \equiv \kappa_2 - v_{2\mathrm{B}}^o$ = 1,000 p/h.

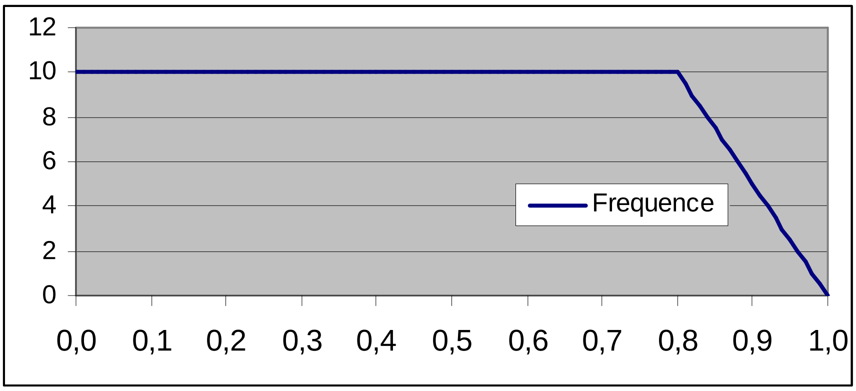


FIGURE 3 Graph of an effective frequency function.

Let us provide an overview of how the volume assignment changes in relation to the demand volume q_B. Lines {2, 3} are attractive from 0 to q_B^3 when line 4 becomes attractive then three lines {2, 3, 4} are attractive. The interval $[0, q_B^3]$ is divided into three segments: firstly $[0, q_B^1]$ where the frequencies for lines 2 and 3 are unchanged, secondly $[q_B^1, q_B^2]$ where frequency F_2 is reduced but not F_3, and thirdly $[q_B^2, q_B^3]$ where F_2 and F_3 are reduced. Beyond q_B^3, we find a segment $[q_B^3, q_B^4]$ where two strategies co-exist {2, 3} and {2,3,4}, then for $q_B \geq q_B^4$ only the full strategy {2,3,4} is used without reducing F_4 on $[q_B^4, q_B^5]$ then reducing it on $[q_B^5, q_B^6]$, where $q_B^6 = \kappa'_2 + \kappa'_3 + \kappa'_4$.

Let us determine q_B^1, the maximum volume for which lines 2 and 3 are attractive without a frequency reduction. Then $q_B \frac{f_2}{f_2+f_3} = \pi_2 \kappa'_2$, yielding q_B^1 = 1,500 p/h.

Beyond q_B^1, congestion reduces the frequency of line 2: the split with line 3 operates according to

$$\frac{v_2}{F_2} = \frac{v_3}{F_3} = y\,.$$

Under our assumptions for the range of values considered it holds that $F_2 = f_2(1 - \frac{v_2/\kappa'_2 - \pi_2}{1-\pi_2})$, thus $v_2 / F_2 = y$ is equivalent to $v_2 = \kappa'_2 \frac{y}{\kappa'_2(1-\pi_2)/f_2 + y}$.

Line 3 frequency is unchanged up to a value $v_3 = \pi_3 \kappa'_3$, thus $y = \pi_3 \kappa'_3 / f_3$ which corresponds to $q_B^2 \equiv \pi_3 \kappa'_3 + \kappa'_2 \frac{\pi_3 \kappa'_3 / f_3}{\kappa'_2(1-\pi_2)/f_2 + \pi_3 \kappa'_3 / f_3}$. In figures, $q_B^2 =$ 1,847 p/h.

On $[q_B^1, q_B^2]$, we determine the volume distribution by replacing y by $(q - v_2)/f_3$, which induces a second degree equation in $v = v_2$: were $\kappa'' = (1-\pi_2)\kappa'_2 f_3 / f_2$, the equation is $v^2 - (\kappa'_2 + \kappa'' + q)v + q\kappa'_2 = 0$.

Beyond q_B^2, both lines are subject to frequency reductions. As $v_3 = q_B - v_2$, the volume is distributed in accordance with

$$v\, f_3 (1 - \frac{(q-v)/\kappa'_3 - \pi_3}{1-\pi_3}) = (q-v)\, f_2 (1 - \frac{v_2/\kappa'_2 - \pi_2}{1-\pi_2}),$$

which is still $v\phi(1 - \frac{q-v}{\kappa'_3}) = (q-v)(1 - \frac{v}{\kappa'_2})$, where $\phi = \frac{f_3}{1-\pi_3}\frac{1-\pi_2}{f_2}$, from which we get the following equation for deducing $v \equiv v_2$ from $q \equiv q_B$:

$$v^2[\frac{\phi}{\kappa'_3} - \frac{1}{\kappa'_2}] + v[1 + \phi - \frac{\phi q}{\kappa'_3} + \frac{q}{\kappa'_2}] - q = 0$$

In this regime, as $v_2 \in \kappa_2'[\pi_2,1]$ and $v_3 \in \kappa_3'[\pi_3,1]$, the ratio $\frac{v_2}{v_3} \in \frac{\kappa_2'}{\kappa_3'}[\pi_2,\frac{1}{\pi_3}]$: the linkage effect on each route induces reasonable consequences for the ratio.

On $[q_{\mathrm{B}}^3, q_{\mathrm{B}}^4]$ the solution is a convex combination $(1-\eta)S_{2+3}+\eta S_{2+3+4}$ of two strategies, {2,3} and {2,3,4}, with a coefficient η that varies from 0 to 1. This means that line 4 is progressively loaded thanks to the second strategy whilst the volumes boarding lines 2 and 3 at B remain constant.

At $\eta=0$, volume q_{B}^3 and its distribution between lines 2 and 3 are determined by three conditions:

$v_2+v_3=q_{\mathrm{B}}^3$,

$\frac{v_2}{\mathrm{F}_2}=\frac{v_3}{\mathrm{F}_3}=y$, which in this case is equivalent to $\mathrm{F}_a=\frac{\kappa_a'}{\kappa_a'(1-\pi_a)/f_a+y}$ and

$1+\mathrm{F}_2 t_2+\mathrm{F}_3 t_3=t_4(\mathrm{F}_2+\mathrm{F}_3)$ which is equivalent to $(t_4-t_2)\mathrm{F}_2+(t_4-t_3)\mathrm{F}_3=1$.

From this we deduce the following equation in y:

$$\frac{\kappa_2'}{\kappa_2'(1-\pi_2)/f_2+y}+\frac{\kappa_3'}{\kappa_3'(1-\pi_3)/f_3+y}=1.$$

The solution y^* determines v_2 and v_3, thus $q_{\mathrm{B}}^3=v_2+v_3$. In figures, y^* = 570 p/veh thus $v_2=$ 983 p/h, $v_3=$ 966 p/h and $q_{\mathrm{B}}^3=$ 1,949 p/h.

The value q_{B}^4 being defined by $\frac{v_4}{\mathrm{F}_4}=y^*$, we get $v_4=$5,700 p/h and $q_{\mathrm{B}}^4\equiv q_{\mathrm{B}}^3+v_4=$ 7,640 p/h.

Beyond q_{B}^4, we determine q_{B}^5 where $\hat{v}_4=\pi_4\kappa_4'$, thus $\hat{y}=$ 630 p/veh and $v_2=\mathrm{V}_2(\hat{y})$, $v_3=\mathrm{V}_3(\hat{y})$. In figures, we find that q_{B}^5 = 8,254 p/h.

3.5 "FtB" Failure-to-Board Probability Model

Kurauchi et al (*4*) addressed the capacity constraint on passenger capacity in vehicles on a transit line on the basis of the failure-to-board probability model: for a route that attracts a volume of passengers that is greater than the residual capacity, the excess volume is removed from the network. It is assigned to a fictitious arc from the current station to the destination, in other words an escape arc.

The ratio between the residual capacity and the attracted volume, capped at 1, is the probability of successful boarding: its complement at 1 is the probability of failure-to-board, or p_F. In our reference case, when $v_2=q_{\mathrm{B}}\frac{f_2}{f_2+f_3}$ exceeds κ_2' at node B, then $p_F^{\mathrm{B},2}=1-\frac{\kappa_2'}{v_2}$. The excess volume $v_2 p_F^{\mathrm{B},2}$ is evacuated onto an escape arc for values of q_{B} over $q_{\mathrm{B}}^{(1)}=\kappa_2'\frac{f_2+f_3}{f_2}$.

Additionally, when $v_3 = q_{\mathrm{B}} \frac{f_3}{f_2+f_3}$ exceeds κ'_3 then $p_F^{\mathrm{B},3} = 1 - \frac{\kappa'_3}{v_3}$ and $v_3 p_F^{\mathrm{B},3}$ is artificially evacuated.

In this model, the maximum capacity of a route is locally respected but to the detriment of the continuity of volumes on the network: the excess volume is truncated instead of being transferred spatially towards other means of transport or, temporally, into a stock of waiting passengers. The dynamic version of the model (Schmöcker et al, *24*) includes the temporal transfer which is very necessary in practice. In the static model, the absence of the transfer of the excess volume has consequences for the simulated downstream volumes that have a greater impact as the line's capacity increases. To try and partly remedy this, the model's authors penalise the failure-to-board in the generalised cost of travel in order to ensure that the propagation of costs upstream from the station orients the volume towards other parts of the network. Their penalty formula is

$$d_F^{\mathrm{B},2} = -\theta.\ln(1 - p_F^{\mathrm{B},2}) , \tag{2}$$

with a penalty coefficient θ that is a calibration parameter. The penalty is zero when capacity is not saturated: it increases with the failure-to-board probability. However, it does not depend on the path that has to be covered on the network to reach the destination which is not very realistic in a static system.

By fixing $q_{\mathrm{A}} =$ 1,500 p/h, thus $\kappa'_2 =$ 1,000 p/h residual capacity on line 2 at B let us study the state of the system with respect to q_{B}. We also fix $\theta = 10'$.

Up to $\tilde{q}_{\mathrm{B}}^1 = \frac{f_2+f_3}{f_2}\kappa'_2 =$ 1,500 p/h, the system is not saturated, $p_F^{\mathrm{B},2} = p_F^{\mathrm{B},3} = 0$. Beyond $\tilde{q}_{\mathrm{B}}^1$, line 2 is saturated. Line 3 is saturated from $\tilde{q}_{\mathrm{B}}^2 = \frac{f_2+f_3}{f_3}\kappa'_3 =$ 3,000 p/h.

On $[\tilde{q}_{\mathrm{B}}^1, \tilde{q}_{\mathrm{B}}^2]$ the generalised cost is

$$\begin{aligned} G_{\mathrm{BD}} &= \frac{1 + f_2 t_2 + f_3 t_3}{f_2 + f_3} - \theta \ln(1 - p_F^{\mathrm{B},2}) \\ &= \frac{1 + f_2 t_2 + f_3 t_3}{f_2 + f_3} + \theta \ln \frac{q_{\mathrm{B}} f_2}{\kappa'_2 (f_2 + f_3)} \end{aligned}$$

Beyond $\tilde{q}_{\mathrm{B}}^2$, and up to $\tilde{q}_{\mathrm{B}}^3$ when line 4 becomes attractive at B, the generalised cost is

$$\begin{aligned} G_{\mathrm{BD}} &= \frac{1 + f_2 t_2 + f_3 t_3}{f_2 + f_3} - \theta \ln(1 - p_F^{\mathrm{B},2}) - \theta \ln(1 - p_F^{\mathrm{B},3}) \\ &= \frac{1 + f_2 t_2 + f_3 t_3}{f_2 + f_3} + \theta \ln \frac{q_{\mathrm{B}}^2 f_2 f_3}{\kappa'_2 \kappa'_3 (f_2 + f_3)^2} \end{aligned}$$

In $\tilde{q}_{\mathrm{B}}^3$, $G_{\mathrm{BD}} = t_4$ which defines

$$\tilde{q}_{\mathrm{B}}^3 = (f_2 + f_3) \sqrt{\frac{\kappa'_2 \kappa'_3}{f_2 f_3}} \exp\{\tfrac{1}{2\theta}[t_4 - \tfrac{1 + f_2 t_2 + f_3 t_3}{f_2 + f_3}]\} .$$

In figures $\tilde{q}_{\mathrm{B}}^{3} =$ 4,344 p/h.

Beyond $\tilde{q}_{\mathrm{B}}^{3}$, the probability of failure-to-board line a is $p_F^{\mathrm{B},a} = (1 - \frac{f_2+f_3+f_4}{f_a}\kappa'_a / q_{\mathrm{B}})^{+}$ and the generalised cost is

$$G_{\mathrm{BD}} = \frac{1 + f_2 t_2 + f_3 t_3 + f_4 t_4}{f_2 + f_3 + f_4} - \theta \ln[(1 - p_F^{\mathrm{B},2}).(1 - p_F^{\mathrm{B},3}).(1 - p_F^{\mathrm{B},4})].$$

Including line 4 in the attractive combination mechanically reduces the volume assigned to the other lines and thus their failure-to-board probability. In Figure 4 at § 4.2 we can see a discontinuity in the generalised cost just after $\tilde{q}_{\mathrm{B}}^{3}$, which is not very realistic.

3.6 "UPS" User Preference Set model

In the user preference set model, the passengers that board line 2 at A have priority over those wanting to board at B. As line 2 has a non-saturated average cost that is minimal between B and D, passengers at A have at least three competitive strategies: the ordered lists of preferences are, in order of rising average costs [2], [2,1] and [1].

If $q_{\mathrm{A}} \le \kappa_2$ then all the volume is assigned to line 2 which has the lowest average cost. [2] is the optimal strategy between A and D without combining it with line 1 even though this has a minimum cost that is less than line 2's minimum cost!

If $q_{\mathrm{A}} > \kappa_2$ then the optimal strategy becomes [2, 1]: strategy [2] can not be achieved because demand pressure makes obtaining a place on line 2 a matter of chance, i.e. a stochastic process. The volume from A boards line 2 in a proportion of $p_{\mathrm{A2}} = \kappa_2 / q_{\mathrm{A}}$, and line 1 in a proportion of $p_{\mathrm{A1}} = 1 - p_{\mathrm{A2}}$, which yields an average cost per passenger of

$$G_{\mathrm{AD}} = p_{\mathrm{A1}}(w_1 + t_1^{\mathrm{AD}}) + p_{\mathrm{A2}}(w_2 + t_2^{\mathrm{AD}}).$$

For passengers coming from B, if $q_{\mathrm{A}} < \kappa_2$ then there is still available capacity $\kappa'_2 \equiv \kappa_2 - q_{\mathrm{A}}$. The possible strategies by order of increasing average cost are [2], [2,3], [2,3,4], [3], [3,4], [4]. If $q_{\mathrm{B}} \le \kappa'_2$ then [2] is the optimal strategy If $q_{\mathrm{B}} > \kappa'_2$ then path 2 on its own can not make a strategy because the saturation makes obtaining a place on this line a matter of chance. The volume from B accesses line 2 in a proportion of $p_{\mathrm{B2}} = \kappa'_2 / q_{\mathrm{B}}$, line 3 in a volume of $v_3 = \min\{\kappa_3, q_B - \kappa'_2\}$ thus a proportion of $p_{\mathrm{B3}} = v_3 / q_{\mathrm{B}}$, and finally line 4 if $(1 - p_{\mathrm{B2}})q_B > \kappa_3$, in a proportion of $p_{\mathrm{B4}} = 1 - p_{\mathrm{B2}} - p_{\mathrm{B3}}$.

If line 2 is saturated at node A, then it is unavailable at B, the volume q_{B} is distributed to lines 3 and 4 using the same formulae as before, replacing p_{B2} with zero.

The average cost per passenger from B is

$$G_{\mathrm{BD}} = p_{\mathrm{B2}}(w_2 + t_2^{\mathrm{BD}}) + p_{\mathrm{B3}}(w_3 + t_3^{\mathrm{BD}}) + p_{\mathrm{B4}}(w_4 + t_4^{\mathrm{BD}}).$$

3.7 "AF" Availability Frequency Model

The traffic assignment in the availability frequency model is the same as in the model without saturation when there is no saturation. In our case, this applies for the volume coming from B up to a value of

$$q_{\mathrm{B}}^{[1]} \equiv \frac{f_2+f_3}{f_2}\kappa'_2 \text{, with } \kappa'_2 \equiv \kappa_2 - \frac{f_2}{f_{\mathrm{A}}} q_{\mathrm{A}} .$$

Between $q_{\mathrm{B}}^{[1]}$ and $q_{\mathrm{B}}^{[2]} \equiv \kappa'_2 + \kappa'_3$, there is only one traffic regime where $v_2 = \kappa'_2$ and $v_3 = q_{\mathrm{B}} - v_2$. In this regime, as route 2 is saturated, a marginal passenger can only use route 3 and experiences a (marginal) cost $\mu = t_3 + 1/f_3$.

In the static framework the marginal cost μ establishes a reference for saturated routes: $\mu = t_2 + \gamma_2$ with γ_2 a boarding penalty for a saturated route. Both lines are attractive thus for both of them we impose that

$$\frac{v_2}{\tilde{f}_2^{\mathrm{B}}} = \frac{v_3}{f_3} = y ,$$

with y an auxiliary variable and $\tilde{f}_2^{\mathrm{B}}$ the frequency of route 2 with capacity available at B: as for $v_2 = \kappa'_2$, it becomes $v_3 = q_{\mathrm{B}} - \kappa'_2$ which determines $y = (q_{\mathrm{B}} - \kappa'_2)/f_3$ thus

$$\tilde{f}_2^{\mathrm{B}} = \frac{\kappa'_2}{y} = f_3 \frac{\kappa'_2}{q_{\mathrm{B}} - \kappa'_2} ,$$

and consequently, the frequency decline due to the capacity constraint:

$$\delta_2^{\mathrm{B}} \equiv f_2 - \tilde{f}_2^{\mathrm{B}} = \frac{q_{\mathrm{B}} f_2 - \kappa'_2 (f_2 + f_3)}{q_{\mathrm{B}} - \kappa'_2} .$$

The frequency decline for lack of available capacity is interpreted as a failure-to-board for a passenger in a vehicle serving the route with the probability $\tilde{p}_F^{\mathrm{B},2} = \delta_2^{\mathrm{B}} / f_2$.

When q_{B} increases, the frequencies are progressively reduced. Line 4 becomes attractive from $q_{\mathrm{B}}^{[2]}$, which increases the average in-vehicle time. The average cost μ also increases as a marginal cost per passenger. Specifically, if $q_{\mathrm{B}} \geq \kappa'_2 + \kappa'_3$ then $v_2 = \kappa'_2$ and $v_3 = \kappa'_3$ thus $v_4 = q_{\mathrm{B}} - \kappa'_2 - \kappa'_3$. This fixes $y = v_4 / f_4$ thus again $\tilde{f}_2^{\mathrm{B}} = \kappa'_2 / y$ and $\tilde{f}_3^{\mathrm{B}} = \kappa'_3 / y$. Furthermore, penalty γ_a stems from $\gamma_a = \mu - t_a$, with $\mu = t_4 + \frac{1}{f_4}$.

Formally, let us define $\tilde{f}_\Sigma = \sum_a \tilde{f}_a$ and set $w \equiv (H + \sum_a \tilde{f}_a \gamma_a) / \tilde{f}_\Sigma$, H being a base time in accordance to the definition of the line frequencies. The term w is interpreted as the average waiting time per passenger. Then

$$\mu = w + \frac{\sum_a \tilde{f}_a t_a}{\tilde{f}_\Sigma}$$

expresses the average cost per passenger produced by the combination of attractive routes.

4 SYNTHESIS AND COMPARISON OF THE MODELS

Having discussed the principles and used each model on a test case, we can summarise the specificities of each model (§ 4.1) and compare the models from numeric (§ 4.2) and theoretical (§ 4.3) standpoint.

4.1 Qualitative Summary by Model

We observed that the Lam model has certain inconsistencies in its interpretation with an application shifted spatially for the capacity penalty and a lack of priority between passenger volumes.

The UPS model is very close to a road assignment except that passengers on board have priority over boarding passengers. The constraint of spatial transfer from a saturated route, rather than waiting in place, is poorly representative of a passenger's behaviour.

The FF model is grounded on sound physical and economic principles. It includes priorities between volumes, increased waiting times when loads approach capacities, combinations of attractive missions including waiting time and the distribution of the volume between attractive routes according to reasonable proportions.

The AF model is easy to use; a route can be loaded up to its nominal capacity and then only the local frequency is adjusted to represent the shortage of available capacity. The average time depends on unsaturated attractive routes and fixes penalties for waiting in the station for all passengers at boarding.

4.2 Numerical Comparison

Table 1 summarises the application of the different models to the test case, for q_{A} =1,500 p/h and $q_{\mathrm{B}} =$ 1,800 p/h. The Lam model is not well-defined at these values. We noted that the case parameters yield values that are quite similar for each model in terms of both volumes and travel times.

Figure 4 shows the change in G_{BD} against q_{B}, still using q_{A} =1,500 p/h. In general, the function increases in each model with the exception of a local accident in model FtB (cf. § 3.5). The UPS model behaves smoothly, the FtB is more jerky, the FF model advance by stages of varying slope and is unchanged over large ranges whilst the Lam and AF models increase by jumps.

TABLE 1 Comparison of Models on a Test Case

Variable	Without constraint	Lam	FF	FtB	UPS	AF	
vA_1	500	0	500	500	500	500	
vA_2	1000	1500	1000	1000	1000	1000	
vB_2	1200	undefined	945	1000	500	1000	
vB_3	600	undefined	855	600	1000	800	
vB_4	0	undefined	0	0	300	0	
G_AD	23.3	*23.3*	23.3	23.3	26.0	23.3	
G_BD	15.7	undefined	18.0	17.5	25.5	27.0	
		γ_2^{BD} 17.0		p_F^{B2} 16.7%	p_{2B} 27.8%	$\tilde{f}_2^B$ 6.3	$(\delta/\tilde{f})_2^B$ 0.375
		not included in cost from A		p_F^{B3} 0.0%	p_{3B} 55.6%	$\tilde{f}_3^B$ 5.0	$(\delta/\tilde{f})_3^B$ 0
				p_F^{B4}	p_{4B} 16.7%	$\tilde{f}_4^B$ 10.0	$(\delta/\tilde{f})_4^B$ 0

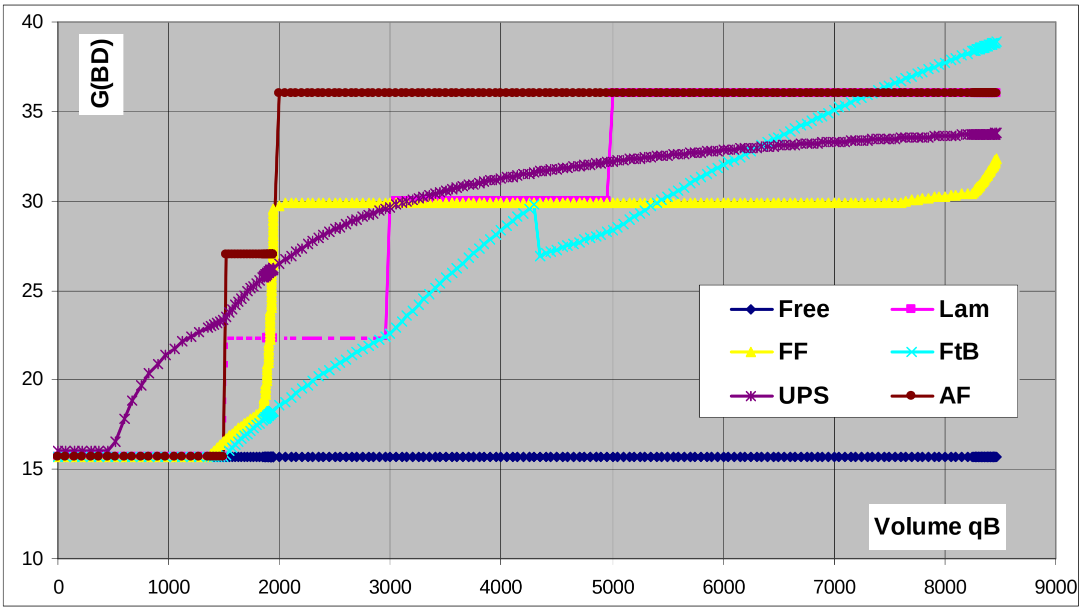


FIGURE 4 Origin-destination cost from B to D against O-D volume qB.

4.3 Theoretical Comparison

The person capacity of a transit route is captured by a constraint that limits the volume on the basis an effective frequency function in the FF model or directly in the other models. The constraint effects the volume on a route segment in the Lam model or, more realistically, the boarding volume in the UPS, AF and FtB models. This latter model applies the constraint ex-post rather than ex-ante to the detriment of flow continuity. In the FF and AF models, the frequency is locally reduced when the volume approaches (FF) or saturates (AF) the capacity.

The priorities between passenger volumes, those on boards over those wanting to board, are explicitly captured in the FF, AF, FtB and UPS models. It is ignored in the Lam model with potentially some counter-intuitive effects.

In respect of estimating waiting time and the overall cost per passenger, a reduced frequency or a saturation penalty can increase the waiting time when full capacity is reached. The AF model combines both effects whilst the FF model uses only reduced frequency; the FtB model penalises a failure-to-board. The Lam model penalises capacity but by route segment and not at boarding, which offsets the place it is applied. The UPS model does not include penalties directly because the excess volume when capacity is saturated is transferred to another option with a higher cost.

For the distribution of volumes between routes and the passenger's choice of route, each model includes a composition of attractive routes. In the Lam, FF, AF and FtB models, route attractiveness has a standard definition: a route is attractive when its minimum cost is lower than the estimated average costs of any routes that are unavailable at that time. The notion of attractiveness is unique and debateable in the UPS model (see § 2.2).

5 Conclusion

5.1 Recapitulation

Based on a critical review of the principles in the models and their application to a test case, we defined the behaviour and scope of a model for passenger capacity of a transit route incorporated in a traffic assignment model. We jointly assessed five capacity models, which allowed us to compare them. The FF and AF models appear to be relevant and robust. The Lam model suffers from an offset between the boarding arc and the route segment. The FtB and UPS models do not involve waiting time, or only slightly, whilst this variable plays a fundamental role in the capacity effect under review; this drawback is corrected in the dynamic versions of these models.

5.2 Future Work

Even in a static framework, the explicit representation of the volume of passengers involved in a waiting situation allows the stock of passengers present in a limited space, such as a boarding platform, to be identified. This identification may enable one to model the effect (C5) of the capacity to stock passengers in a confined space.

Another subject for research is evaluating the capacity of a route. Demand variability over time and space and also the specificities of individual travellers make the passenger capacity of a transit route a macroscopic variable that is subject to variations: its variations and their effects on volumes could be the topic of a stochastic model.

6 REFERENCES

1. De Cea J. and E. Fernández, Transit assignment for congested public transport systems: an equilibrium model, *Transportation science,* Vol. 27, 1993, pp. 133-147.
2. Cepeda M., R. Cominetti and M. Florian, A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria, *Transportation Research Part B*, Vol. 40, 2006, pp. 437–459.
3. Lam W. H. K., Z. Y. Gao, K. S. Chan and H. Yang, A stochastic user equilibrium assignment model for congested transit networks, *Transportation Research Part B*, Vol. 33, 1999, pp. 351-368.
4. Kurauchi F., M. G. H. Bell and J.D. Schmöcker, Capacity constrained transit assignment with common lines, *Journal of Mathematical Modelling and Algorithms*, Vol. 2, 2003, pp. 309-327.
5. Hamdouch Y., P. Marcotte and S. Nguyen, Capacitated transit assignment with loading priorities, *Mathematical Programming B*, Vol. 101, 2004, pp. 205–230.
6. Leurent F. and Y. Askoura "*Availability frequency: a novel model for traffic assignment to a capacitated transit network*". Working paper available on request from the authors, 2009.
7. *Transit Capacity and Quality of Service Manual (TCQSM)*, 2nd Edition, Transportation Research Board of the National Academies, Washington, D.C., 2003.
8. Leurent F. 2009, On Capacity and Congestion in Transit Systems, *Bonn EURO Conference*, July 8, 2009.
9. Spiess H. and M. Florian, Optimal strategies : a new assignment model for transit networks, *Transportation Research Part B,* Vol. 23, 1989, pp. 83–102.
10. Dial R.B., Transit pathfinder algorithms, *Highway Research Record*, Vol. 205, 1967, pp. 67–85.
11. Fearnside K. and D.P. Draper, Public transport assignment—a new approach, *Traffic Engineering and Control*, Vol. 12, 1971, pp. 298–299.
12. Last A. and S. E. Leak, Transept: A bus model, *Traffic Engineering and Control*, Vol. 17, 1976, pp. 14–20.
13. Chriqui C. and P. Robillard, Common bus lines, *Transportation Science*, Vol. 9, 1975, pp. 115–121.
14. Nguyen S. and S. Pallotino, Equilibrium traffic assignment in large scale transit networks, *European Journal of Operational Research*, Vol. 37, 1988, pp. 176–186.
15. Shimamoto H., Y. Iida, J. D. Schmöcker, F. Kurauchi and M. G. H. Bell, Evaluating public transit congestion mitigation measures using a passenger assignment model, *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 6, 2005, pp. 2076-2091.
16. Leurent F. and K. Liu, On seat congestion, passenger comfort and route choice in urban transit: a network equilibrium assignment model with application to Paris, TRB Meeting 2009, N° #09-1784.

17. Meschini L., G. Gentile and N. Papola, A frequency-based transit model for dynamic traffic assignment to multimodal networks, In: *Presented at 17th International Symposium on Transportation and Traffic Theory*, London, July 2007.

18. Hamdouch Y. and S. Lawphongpanich, Schedule-based transit assignment model with travel strategies and capacity constraints, *Transportation Research Part B,* Vol. 42, 2008, pp. 663–684.

19. Wu J., M. Florian and P. Marcotte, Transit equilibrium assignment : a model and solution algorithms. *Transportation Science*, Vol. 28, 1994, pp. 193–203.

20. Cominetti R. and J. Correa, Common-lines and passenger assignment in congested transit networks, *Transportation Science*, Vol. 35, 2001, pp. 250–267.

21. Tian Q., H.J. Huang and H. Yang, Equilibrium properties of the morning peak-period commuting in a many-to-one mass transit system, *Transportation Research Part B*, Vol. 41, 2007, pp. 616-631.

22. Leurent F., Confort et qualité de service en transport collectif urbain de voyageurs : analyse, modélisation et évaluation, *Actes INRETS*, n°118, 2006, pp. 111-134.

23. Leurent F., Modelling seat congestion in transit assignment, *Submitted for publication to Transportation research Part B,* 2008.

24. Schmöcker J. D., M.G.H. Bell and F. Kurauchi, A quasi-dynamic capacity constrained frequency-based transit assignment model, *Transportation Research Part B*, Vol. 42, 2008, pp. 925-945.