Dynamic assignment model of trains and users on a congested urban-rail line

Alexis Poulhès

Dynamic assignment model of trains and users on a congested urban-rail line

Abstract

For the management and planning of urban rail lines, operators can draw upon tools that use train circulation models as well as passenger assignment models. However, these two kinds of simulation models are independent. They do not interact as they do in reality. Yet on certain highly congested lines, high train frequency and large passenger volumes can turn a small incident into a delay on the entire line. Our research presents an integrated model for the simulation of a fixed block urban rail line in interaction with passenger assignment. This operational model introduces new management strategies or rolling stock feature solutions to improve the quality of service on the line. A discrete-event approach simulates the progress of the runs on the line, and the representation of the passengers by origin-destination flow per time step makes it possible to effectively simulate lines with large flows. An application to Line 13 of the Paris metro illustrates the model on a real case of congestion. Sensitivity analyses on the level of demand as well as on service characteristics demonstrate the utility of this integrated approach.

Keywords

Dynamic assignment; transit line; transit assignment; line model; train regulation; train circulation

Introduction

Operationally, passenger assignment models in public transport are used to assess projects for service improvements and new urban rail line infrastructures. In socio-economic assessments, time saved and increased comfort seen in terms of economic gain still receive little consideration in operational studies. Yet in certain big cities with congested networks, the benefits of rolling stock upgrades or of the doubling of certain lines cannot be evaluated in socio-economic assessments unless they include congestion constraints. Similarly, the dynamic management of railway systems emphasizes technical problems and operating constraints that take little account of quality of service as perceived by users.

This research focuses on modeling the interaction between user behavior and service dynamics on a congested, complex rail line. The dynamic modeling of demand under capacity constraints is explained in detail in a previous article (EWGT, Poulhès et al. 2017). Passenger flows arrive in a station on the line and board the trains to their destination, depending on the availability of the service and competition with other users for a train with limited capacity. In that first model, the service is constant over time. It is described exogenously in terms of a scheduled timetable. We will call the demand model in that previous article the hybrid model. The second model presented in this article focuses on the detailed modeling of the progress of trains on a railway line. This article also presents the connection with the passenger assignment model so that both models can be used within an integrated dynamic model.

A transit line is a quasi-closed sub-system on a transit network. Only passengers transferring between lines link the line to the rest of the system. A run refers to the movement of a train from a terminus station to another terminus station with a schedule for the successive stations it serves. Runs traveling

the same route with different departure times constitute a mission. A train can turn back at the terminus to be used by another run. The rail line is divided into blocks with a signal at the entrance to each block. Several systems maintain the movement of the trains, together keeping a minimal interval between consecutive trains.

The modeling of service dynamics centers around 3 rail traffic phenomena that have an impact on passenger traffic (Gentile & Noekel 2016, Leurent 2011):

- The increase in the time spent in the station as a result of longer dwell time, i.e. the time it takes for passengers to board and alight, as well as occasional delays to trains at a station or between stations to simulate an incident.
- The kinematic dynamics of the trains, i.e. the phases of acceleration and deceleration incorporated into the calculation of the travel time on the line. The number of these phases will obviously depend on the number of stations served on a run, but also on the line and signal dynamics. If the line is congested, and depending on the type of signaling system and the operating strategies, travel times can be affected to different degrees.
- Signaling and the interaction between runs. A slowdown in one run can lead to a slowdown in all the runs that follow it. In the case of lines with branches, a delay in the timetable on one branch can also lead to slowdowns on the other branches in order to maintain safety margins or the order of travel on the shared trunk section. Two types of signals are considered in the model. Standard fixed block signals where each train must at all times have at least one empty block behind and in front of it. Mixed signaling, with fixed blocks but where trains can move closer together because each train knows where its predecessor is located on the line.

The aim is to closely model the circulation of the trains and the movements of passengers in their journeys from platform to platform, which include platform waiting time and on board conditions. The goal in this article is not to optimize operations, but only to describe the rail traffic phenomena and the interactions between train traffic and passenger traffic in order to obtain a better idea of the travel times perceived by users, and also to be able to simulate minor incidents and their impacts on travel time. So the purpose is to enrich the hybrid line model to include this dynamic in service provision and therefore to offer a simplified model of the occurrence and propagation of delays in a rail network, consistent with the in-train passenger load and platform capacity. Symmetrically, the travel conditions for users along the line will depend on train delays, with respect to on-board comfort and platform waiting conditions.

The main contributions of this research are: (1) to provide a simple model of run circulation that allows a great deal of freedom in the operational strategies tested, as well as simulating complex lines with several branches and several missions; (2) to construct a single system to simulate the dynamic behaviors of users between an origin station and an exit station, and the dynamics of trains on a railway line under signaling constraints. This will contribute to our understanding of the dynamics involved and help to quantify accurately the users affected by congestion phenomena, so that problems on the line can be better assessed. (3) The scenarios tested may help to suggest original operating solutions both with respect to the line and with respect to the management of passengers in stations or across the network. (4) An original dynamic simulation of passengers and trains on the Paris Line 13 network illustrates the model on a real-world case.

Background

Two fields of research are recruited to construct our simulation model of train progress and passenger assignment on an urban railway line.

On the demand side

On the one hand, a large field of research has focused on modeling passenger assignment in public transport networks. The literature has mainly concentrated on detailing the dynamics of demand in response to an often exogenous service supply.

After a first period of research dedicated to the formulation of hyperpaths (Nguyen & Pallatino, 1989) then of strategies (Spiess and Florian, 1989) on static passenger assignment without capacity constraints, research on incorporating congestion constraints into the models is now beginning to bear fruit (Fu et al. 2012). The first approaches considered a cost function as an exponential function, either on time spent in the vehicle (Lam et al. 1999) or on waiting time (de Cea & Fernandez 1993, Cominetti & Correa 2011) with the notion of effective frequency. Since then, frequency-based static models have been used to simulate very large networks for long-term planning, while incorporating interactions between several capacity constraint factors (Leurent et al. 2014, Verbas et al. 2016). Timetable-based models are a way to achieve greater precision regarding the dynamics of the trains, passenger load and the waiting time for each train (Hamdouch 2010). However, while the influence of irregular service on user behaviors and the impact it has on their generalized cost has often been studied (Szeto et al. 2013, Babaei et al. 2014, Jiang et al. 2016, Leurent et al. 2017), the reciprocal interaction between the users and the service is rarely studied.

In the CapTA model (Leurent et al. 2014), the authors take the view that the dwell time will influence the line's frequency and the upstream travel times, and reciprocally a drop in frequency will increase platform waiting time. However, the model proposes only constant variables in time, so the dynamic knock-on effects were not taken into account, all trains having the same travel time. Cats et al. (2016) propose a dynamic model, Mezzo, around a bus network in which the dwell time of each bus will influence its travel time and therefore will have an impact on wait times in the case of dynamic demand. Hamdouch et al. (2014) propose a timetable-based assignment model that introduces uncertainty into the station to station travel time of each line. This uncertainty may change users' optimum strategy between an origin and a destination. In this way, the authors construct an equilibrium model that takes into account the variance in the generalized cost.

On the rail supply side

Another significant field of research explores the dynamic behavior of railway lines from the perspective of optimizing schedules and service plans. There are many articles that deal with this operational rail problem. Several states of the art have described the main research and issues (Cordeau et al. 1998, Cacchiani et al. 2014, Caimi et al. 2017). Cordeau et al. (1998) concentrate on methods of optimization in rail freight or intercity passenger networks, noting the gap at the time between theoretical research and practice. Cacchiani et al. (2014) list the types of models used in recent research and in particular research that has introduced a passenger component into rail optimization. Caimi et al. (2017) look in particular at the applications of the theoretical models.

Today, for example, there is specialist software that enables operators to prepare schedules on large networks on the basis of the travel constraints of trains or subways (Hastus, 2018, OpenTrack 2018). In 2004, Nash & Huerlimann explained the main principles of the OpenTrack software, which simulates the progress of all the trains running on a section of rail network, taking into account the rolling stock and timetables along with the infrastructure and signaling. Most of the research in this field at the time explored the optimization of timetables in relation to a number of technical constraints in intercity rail networks, which often accommodate several freight or passenger lines on a single track. A 1998 review of the literature showed that the research at the time included just one aspect of optimization. The passenger was still missing. Since then, some scholars have concentrated on urban networks and their specificities (Liebchen, 2008). Some research has introduced a user cost function into the optimization function (Wang et al., 2017). The authors calculate a global cost function which, among other things, includes a passenger-side function by introducing a capacity constraint into the optimization function. The passenger cost function depends on train load as well as on travel time. Another piece of research (H. Niu et al., 2015) proposes a long-term timetable that minimizes total waiting time on a line by formalizing a bottleneck model and residual demand per train. Other research has looked at minimizing the energy consumed by a rail network. Yin et al. (2017) use a single model to combine the minimization of energy consumption and of passenger waiting time on the line. The demand is dynamic and the waiting time takes each vehicle's capacity into account. However, in their model, demand has no impact on the progress of the trains. Sun et al. (2014) present 3 models of waiting time optimization under operational constraint and with or without capacity constraint at train boarding in each station on a line. Applying their models to a 'MRT line' in Singapore, the authors succeed in reducing waiting time by almost 30% and in eliminating 'left-behind' passengers. Farhi et al. (2017) use (Max,+) algebra to describe the progress of trains on a line with turnaround and present the frequency of the line in relation to the number of trains in movement. If there are too many moving trains, safety blocks between two consecutive trains involve congestion on traffic and frequency is then degraded and quality of service therefore deteriorates.

One subcategory of the research deals with the 'rescheduling problem'. This concerns the establishment of a new schedule after a problem on the line. Wagenaar et al. (2017) thus propose a model for reloading a line after an incident, taking into account the behaviors of passengers in the event of a service interruption and limiting the number of trains running empty. Jiang et al. (2016) use automatic fare collection from the Shanghai transit network to explore the boarding-alighting time per train on a simple transit line. The authors point out the heterogeneity in the loading per station and per train during the simulation period. They calculate the total waiting time lost from the limited number of passengers boarding in relation to train capacity and the number of vehicles. The failure-to-board passengers must wait for the next train. The rail dynamics and the interaction between supply and demand is not considered. Li et al. (2017) use an approach similar to ours to construct an optimum control model which takes into account the influence of boarding and alighting passengers on a train's dwell time in a station. A train's dwell time as well as its travel time between 2 stations can be disrupted. The authors therefore introduce a positive time control on these 2 time periods where the goal is to minimize the error between the disruptive situation and the nominal reference situation. Toledo et al. (2010) propose a model based around the Mezzo simulator which simulates bus traffic in a network by having the travel time and dwell time depend on the number of users. The arrival of users at a station follows a Poisson distribution and the travel time of the buses diminishes with the load. Moreover, the buses are represented as agents and resume service at the line terminus. The delay can therefore

accumulate over the day on both of the line's directions of travel. Li and Zhu (2016) propose a dynamic, discrete-event passenger assignment model on a rail network on which a train can experience a delay. The users can then decide whether to remain on the line, to choose another route, to take another transport mode, or to cancel their trip. But passenger flow cannot delay the circulating train due to excess dwell time. They illustrate their model with a simulation of a delay on a line on the Shanghai railway network and observe how the model represents the time absorption of the event.

Specialized operational solutions are becoming compatible, offering new possibilities and opening up to new customers. For example, the Opentrack tool already mentioned can interface with the micro-simulation software Simwalk. The objective is to correlate passenger flows in a station with the smooth running of the trains.

This brief state-of-the-art displays the lack of interchange between the research fields that respectively explore assignment of transit users and train circulation. Some research accounts for exogenous constraints from the other field, but without real interactions.

Method

Despite the existence of a very extensive literature, the quest for a timetable that will be optimum under certain conditions employs mathematical algorithms that are difficult to establish and require strong assumptions (Caimi et al. 2017, p297). This complexity makes modularity difficult, as well as the use of this type of model in a network scale model. To anticipate the use of this line model at network scale and therefore to allow route adjustment in the event of disruption, we will apply the discrete-events simulation method where the events represent the progress of the train. This provides greater freedom for different operating strategies.

On the line modelled, user flows into the station are assigned to the trains as they arrive, but also depend on available on-board capacity.

Assignment of passenger traffic: the hybrid line model

The hybrid model presented in (EWGT, Poulhès et al. 2017) can be used to assign passenger flows on a line with train runs represented on a unit basis over a given time period. The passengers are therefore assigned to a train from station to station according to their time of arrival at the origin station. Train loads will be limited by train capacity, which may oblige passengers to remain on the platform to wait for the next train. One of the limitations of the hybrid model is that it does not take into account the dynamics of the line and in particular of the interaction between passengers boarding and alighting in the station, which impacts the smooth operation of the line at peak hours. The model behaves as if the line studied has a fixed run travel time, with a constant dwell time, and therefore as if headway between trains over the period is fixed from the schedules. The model presented in this article therefore simulates that dynamic.

The service dynamic

Trains move along the railway line from an initial terminus to an end terminus, serving certain stations depending on their schedules. On certain urban lines, demand levels are so high that the operators are tempted to have the maximum possible number of trains running. However, each incident on a train will have repercussions and affect travel times on all the upstream trains. That is why, today, some operators – like RATP for the RER A, the most travelled line on the Île-de-France network – have decided to reduce train frequency at peak times in order to improve regularity, and to increase total capacity. Indeed, train

frequency reaches a threshold limited by the maximum dwell time. If the number of trains is above this value, train circulation creates a bottleneck and the frequency decreases, as explained by the fundamental diagram of traffic flow (Daganzo, 1997).

The simulation of the progress of the trains as far as possible follows the timetable of each initial departure on each run. In order to avoid the possibility of disruption having a retroactive effect on trains downstream from a disruption, the progress of the runs is simulated in chronological order. The runs progress on a discrete event basis. An event corresponds to the transition from the end of one block to the end of the next block. The transition from one block to another will depend on the travel time possible on that block, as well as on the position of the previous train on the line. A train only moves if it is sure that the target block is empty. In the hybrid model, passenger loading takes place in the same temporal and spatial order as the progress of the trains in the service model, defined train by train by the discrete event model.

Application area for the transit operator

The model presented assigns passengers only to their itinerary on the line studied. The possibility of rerouting users to another line in case of a long delay or serious accident is not taken into account. The specific operating procedures or impacts of passenger information are also outside the scope of the model. That is why the scope of simulation is limited to day-to-day congestion or a short incident, which have no impact on individual passenger behavior at line level.

Interactions

The model of flow assignment along a railway line as presented above loads the trains on the line on the basis of their arrival in each station and their dwell time. For each origin-destination link and for each time step chosen, the model calculates a travel time and comfort levels corresponding to the trip conditions simulated. At each time, the conditions of comfort depend on the time spent waiting on the platform for the relevant destination, as well as on the in-train journey. The service model calculates the times of the trains' arrival and departure at each station on the basis of the conditions of passenger exchange on the platform for each train, calculated by the line model. The departure time can be delayed if the movement of the upstream trains is degraded, so that other users who have just arrived in the station are allowed to board the train.

Plan

The rest of the article is structured as follows. The first part introduces the elements of the railway line and signaling principles modelled, as well as the construction of the travel time for each block on the basis of the train's service and the signals. The second part describes the movement of the trains with the general algorithm at line level. Part 3 sets out the main assumptions and principles of the hybrid model for assigning demand along the line as well as the construction of waiting time by the bottleneck model. The interaction between the 2 models is made explicit at the platform exchange level. Part 4 details the block by block progress according to the signal type in two cases of signal procedures. Finally, Part 5 illustrates the model on a specific case, Line 13 of the Paris Metro. The final part will set out a number of possible avenues for further research, as well as discussion points and conclusions.

1 The railway line

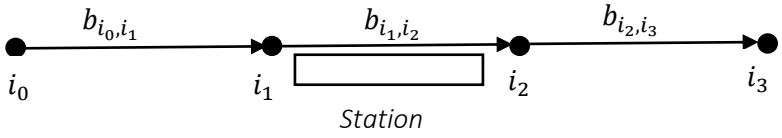
1.1 The infrastructure

A railway line is considered to be two sub-lines each of which represents a direction of travel. In the model, the interactions between the directions of travel is taken into account only through the reversal of trains at the termini. Demand that transfers from one branch to the other going through the first station on the shared section will be considered to be exogenous in the line model. Thus the two sub-lines are connected together only at their termini.

1.1.1 The tracks

The line's rail tracks are represented by a directed graph (N, B) where N is a set of nodes and B a set of arcs.

The arcs represent blocks that are divided into interstation blocks B_I and station blocks B_S, $B = B_I \cup B_S$. The nodes are the points where blocks meet, $N = N_I \cup N_S$ where N_I represents nodes on which the incoming arcs are interstation blocks and N_S, those on which the incoming arcs are station blocks. On complex lines, several arcs can start from a single node, respectively arrive at a single node, but not at the same node. These are the branches that are used to model complex lines.



Only one line can run along each section of track. The Track-Line system is therefore considered to be a closed rail system.

1.1.2 Stations, platforms

All the stations on the line are represented by a single block $b_s \in B_S$ and a single station $n_s \in N_S$ which is the downstream end of the line. In the model, the node n_s will be considered as a station. Two stations are always separated by a block and an interstation node. Between 2 stations, it is possible to have several interstation blocks.

1.1.3 Signals

In the model, 2 types of signals are simulated, which correspond to the systems used in 2018 on most of Île-de-France's urban railway network. Communications-Based Train Control (CBTC), the moving block system (Railsystem, 2015) is not simulated in this article, since it is found on only some metro lines in Île-de-France. We see here the correspondence with the European ETCS category of operating systems, which distinguishes between three levels of signaling. (Berbineau 1999), ERTMS = ETCS + GSM-Railways

(i) For each fixed block, 3 types of signal lights (ERTMS/ETCS level 1)

Let $\sigma_b(i)$ be the current state of the signal at the end of block $b_{\lambda(i),i} \in B$ in which the train is located. $\sigma_b \in \{'g','o','r'\}$.
- green: advance until next signal limited only by the maximum permitted speed
- orange: slow down to 30 km/h before the end of the block. This signal indicates the presence of a train in the 2nd block that the train will pass through.
- red: stop the train because there is a train present in the next block. Progress possible once the block is free.

(ii) SACEM system (ERTMS/ETCS Level 2): operating and driving assistance present in 2018 only on the central trunk of the RER A line. The technical principle is a hybrid between track-based signaling and the transmission of continuous information on the positions of every train on every section of track by radio to the operations center. This simplifies the signaling. If there is a train in the previous block, the train behind it cannot advance into the block, so the signal is red. If the block is empty, the signal is green and the train uses the radio information from the operations center to adjust its speed according to the position of the train ahead.

1.2 The supply

1.2.1 The line and its missions

We assume that a line is a set of missions associated with a single network. The sequence of consecutive blocks through which a mission z passes is defined as a list $B(z) = \{b_1, b_2, \dots, b_n\}$ in which $b_i \in B$. It is associated with a set of stations served, $B_S(z) \subset B_S$. To this is added the constraint that each station in the mission is associated with a single block: $\forall b_s \in B_S(z), \exists! \, b \in B(z) | b = b_s$. If a station is served by several missions on different platforms, a single block is associated with each platform, so there can be several blocks associated with one and the same station on the line.

1.2.2 The train and the run

Trains $k \in K$ circulate on the line L. The rolling stock is the same for all the trains, with identical capacities and internal layouts as well as kinematic characteristics. Runs make up a mission $k_z \in z$. Each run of the mission is associated with a single train k. Conversely, a train can be used for several consecutive runs. There can only be one train on a block, whether stationary or in motion.

Line L on the network (B, N) consists of a set of missions L with trains K.

1.2.3 The timetables

Each run is associated with an initial terminus station n_s and a theoretical departure time $\overline{H}_0$: $k_z(n_s, \overline{H}_0)$. In the model, only the theoretical departure time at the line termini will be considered for all the trains on the line. The times of arrival at the other stations will be calculated by the model on the basis of each train's travel time and the delays due to train congestion.

1.2.4 Travel times

A number of pieces of research have focused on the precise simulation of train trajectories. Kikuchi (1991) uses acceleration ratios and maximum speed to simulate the progress of a train on a railway line. The results are considered conclusive for operational use. Another study calibrates a simple speed calculation model based on the identification of acceleration and deceleration phases, where the traction and resistance forces are calculated on the basis of parameters in the literature and the characteristics of the rolling stock (Wang and Rakha, 2018). In Dullinger et al. (2017) a double level of optimization is presented in order to minimize the energy consumed by optimizing the train's trajectory. A cellular automaton system is used by (SU et al., 2012) to compare the headways between trains on the basis of the signaling system and the speed of the trains. The kinematics remained very simple, despite the fact that the number of cells is very large.

If the line is not automatic, travel time will depend on the behavior of the train engineers (Carre, SNCF). This dispersion around the average travel times is not taken into account in this article. In order to

simplify the model and save computation time, the travel times provided by the operator are used. We assume a known uncongested travel time $t^{g}_{i,\mu(i)}$ as the average travel time in normal operating conditions. In addition, a travel time on a block after an orange signal is defined in advance. Thus, the time $t_{i,\mu(i)}$ for a run $k_z \in z$ to pass through block $b_{i,\mu(i)}$ depends only on the signal color and is constant:

$$\begin{cases} if\ \sigma_i =' g', t_{i,\mu(i)} = t^{g}_{i,\mu(i)} \\ if\ \sigma_i =' o', t_{i,\mu(i)} = t^{o}_{i,\mu(i)} \end{cases}$$

The time can be assumed to be constant as long as the increasing travel time due to congestion is integrated into an added delay time calculated by the model.

The travel times in stations correspond solely to the time taken for the trains to move from the station entry node to the station exit node with a final speed of zero. Stopping time in the station is not included in this time.

2 Train circulation

The aim in this section is to describe the service model in which trains move from their origin in a fixed simulation time interval. This progress takes place from train to train and from block to block served by each run and in the chronological order of events.

The first part describes the general principles of the model and all the assumptions in it. The second part is devoted to the general algorithm with a general diagram of all the procedures and then the general algorithm of the model. The sections that follow detail firstly the assignment model at a station and secondly the movement of a train from one block to another.

2.1 General

2.1.1 Assumptions

The main assumptions in the train circulation simulation model are as follows:

(i) Along most of the urban rail line, automatic regulation procedures should handle disruption to train movement. However, in “normal” operation, i.e. incident-free travel, RATP, the Parisian rail line operator, has only one specific procedure. The only regulatory mechanism is that an exchange time should as far as possible be maintained. Our model therefore includes this only automatic regulation mechanism.

(ii) The effect of engineer behavior on travel times is random. In this research, therefore, we consider that travel time does not depend on this exogenous behavior.

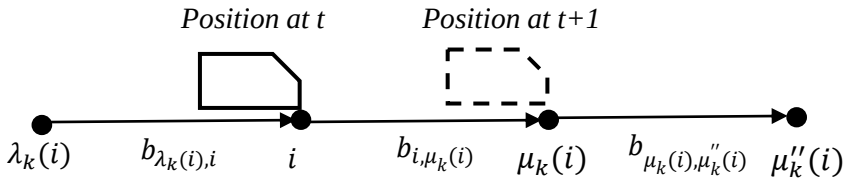
(iii) With no congestion problems, on a given section of line, the travel time is therefore fixed for all the runs belonging to a single mission.

(iv) In this model, a train is considered to be a single point on the track, with the front and back of the train at the same place. The length of the train and its impact on the duration of its presence in a block are not taken into account in the signal’s presence indicator. As soon as the front of the train moves to a new block, the signals change on the block that has just been left and the one before it.

(v) In order to isolate the sub-system line, the number of passengers arriving from outside is estimated exogenously.

2.1.2 Principle

We propose to study the following event (k_z, i): train k_z is located just before node i with a signal that will be green or orange and will allow it to advance to node $\mu_k(i)$, its target node. We know that the signal will be either green or orange, because managing the list in chronological order tells us whether the previous train k on the line left block $b_{i,\mu_k(i)}$ at a time earlier than the time when train k will pass through block $b_{i,\mu_k(i)}$. Moreover, in order to calculate the travel time on the block, we need to know the color of the signal. For this, the following constraint is added to the event: the time $H^+_{\mu''_k(i)}(\beta''_i(k_z))$ at which train $\beta''_i(k_z)$ left the block following $b_{\mu_k(i),\mu''_k(i)}$ must be known at the moment the event is processed.



We only process the instant when the train arrives at the end of the next block. This is a discrete event model. Let $H_i^{-/+}(k_z)$ be the real time of arrival/departure of train k at node i. In this part, we try to calculate the time of arrival of k_z at node $\mu_k(i)$, $H^-_{\mu_k(i)}(k_z)$, the time of departure of k_z from node $\mu_k(i)$, $H^+_{\mu_k(i)}(k_z)$.

The following values are already known at the moment when the event (k_z, i) is processed: $H_i^+(k_z)$ the departure of k_z at node i and $H^-_{\mu_k(i)}(\beta_i(k_z))$; $H^+_{\mu_k(i)}(\beta_i(k_z))$ respectively the arrival and departure of the previous train $\beta_i(k)$ at node $i+1$ named $\mu_k(i)$ on the mission itinerary.

2.2 General algorithm

2.2.1 Lists of events

The list $E_{L,H}$ is the list of events to process at instant H, i.e. the list of trains for which there are no trains in the 2 blocks preceding them. The instant H corresponds to the time of departure from the current node of the first train in the list $E_{L,H}$, $H = H_i^+(k_z)$, $(k,i) = E_{L,H}(1)$. The list $\bar{E}_{L,H}$ is the list of all the other trains that are waiting either for their departure time from the terminus, or for the blocks preceding them to become free.

The list of events for processing, $E_{L,H}$, contains tuples of three variables. Each tuple corresponds to all the trains for which the next 2 target blocks are not occupied by a train at the moment of the simulation. A tuple in $E_{L,H}$ contains:

(i) A train k associated with a node i.
(ii) A block $b_{\lambda(i),i}$ corresponding to the presence of the train k
(iii) A time of departure of train k from block $b_{\lambda(i),i}$, $H_i^+(k_z)$

2.2.2 General diagram

Figure 1 summarizes the links between the steps of the algorithm.

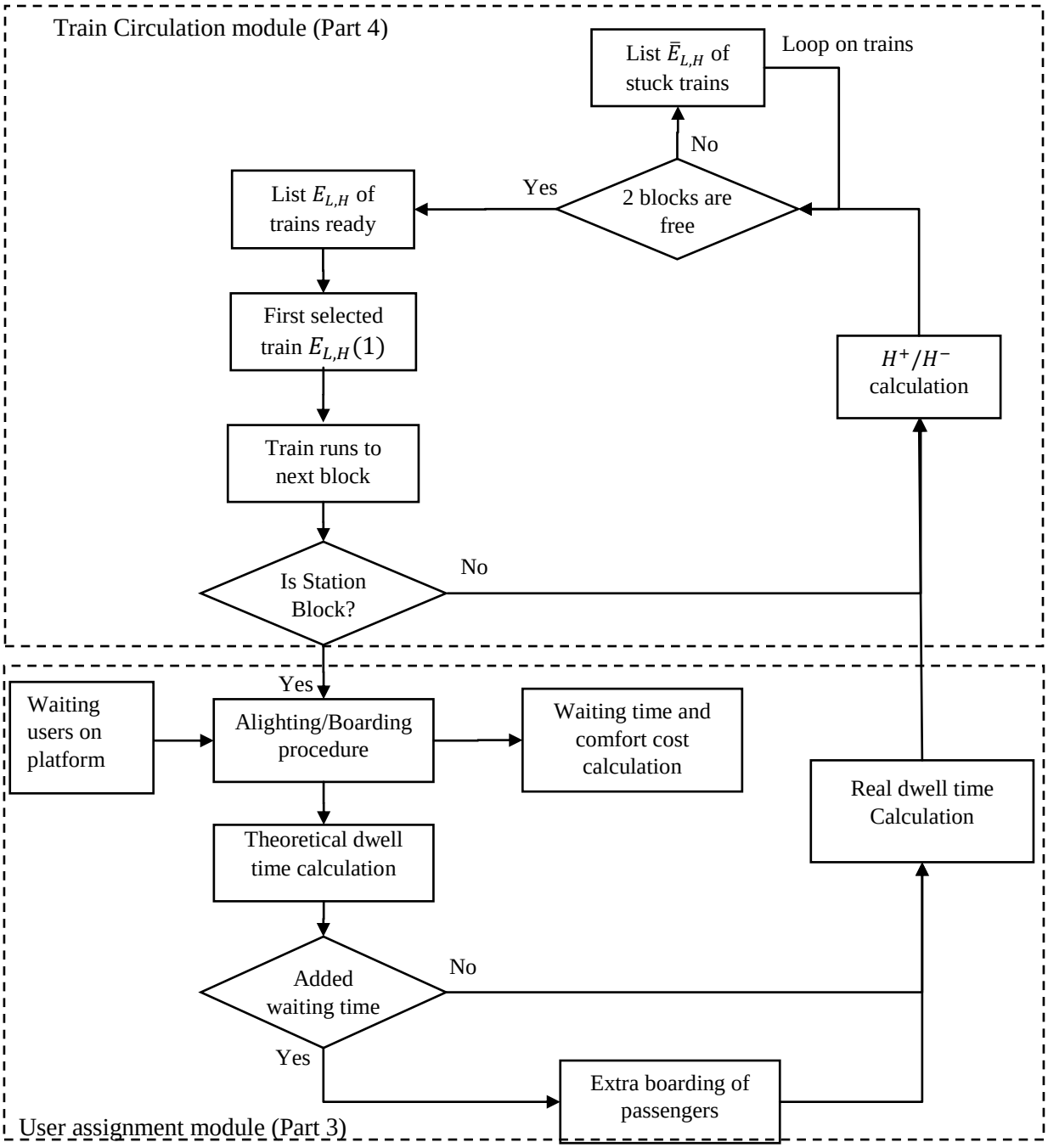


Figure 1: Diagram of the procedure for train progress along the line and passenger assignment.

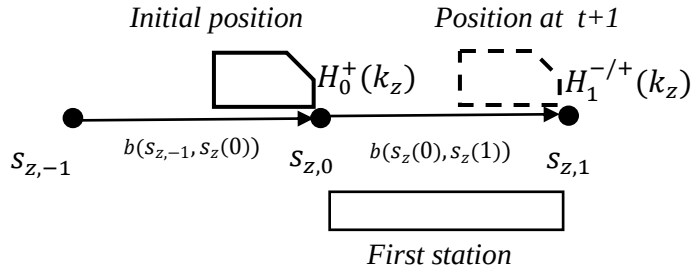
2.2.3 Initialization

We write k_z^0 for the train in mission z which is the first to leave the initial station on the mission $b\big(s_z(0), s_z(1)\big)$ where $s_z(0)$ and $s_z(1)$ are the 2 ends of block b.

The general algorithm calculates the arrival and departure time of the trains at each train target station. The algorithm must be able to hold back trains before the initial station so that they do not have to wait at the end of the very first block they reach. To do this, a virtual block is created before the first station in

each mission. The departures of the trains on the mission are initialized from this block with the scheduled timetable $\widetilde{H}_0(k_z)$:

$$H_0^+(k_z) = \widetilde{H}_0(k_z), k_z \in z$$



The first event for each train will therefore be used to recalculate the real departure from the terminus station $H_1^+(z)$.

We denote by $k_z^0 \in z$ the first train to leave in mission $z \in L$, $\widetilde{H}_{s_{z,0}}(k_z^0) < \widetilde{H}_{s_{z,0}}(k_z), \forall k_z \in z$. If several missions leave from the same initial station, only the train that leaves first out of all these missions should be selected in $E_{L,H}$. It is assumed that all the trains have a first virtual depot block and a second station block. We write $L^0(z) \in L, z \geq 1$ to denote the subset of missions on L that start their itinerary at the same terminus.

We are now going to perform the first initialization of the heap of events to be processed $E_L(H)$. It consists in ordering all the runs belonging to the missions in L^0on the basis of departure time.

Loop on $L^0(z), z \geq 1$

We identify the mission z which has the first train to depart:

$$E_{L,H} \leftarrow (k_z^0, s_{z,0}) \; such \; that \; k_z^0 \in z, \widetilde{H}_{s_{z,0}}(k_z^0) < \widetilde{H}_{s_{z,0}}(k_z), \forall k_{z'}^0, z' \in L^0$$

End loop

Ordering of elements of $E_{L,H}$ in ascending order of departure times such that

$$If \; (k, s_{z,0}) = E_{L,H}(1), \forall k' \in E_{L,H}, \widetilde{H}_{s_{z,0}}(k) < \widetilde{H}_{s_{z,0}}(k')$$

$\bar{E}_{L,H}$: all the other trains with their theoretical departure time.

Lists $E_{L,H}$ and $\bar{E}_{L,H}$ are updated each time an event is processed, as explained below.

2.2.4 General operation

Provided that the set of events $E_{L,H}$ is not empty

- Let $(k_z, i) = E_{L,H}(1)$, the event before the minimum departure time of all the trains in this list.
- Progress of train k_z to station $\mu_k(i)$. Calculation of times $H_{\mu_k(i)}^-(k_z)$, $H_{\mu_k(i)}^+(k_z)$. Release of block $b_{\lambda(i),i}$ and update of the occupancy of the next block $b_{i,\mu(i)}$: $\delta(b_{\lambda(i),i}) = 0 \; and \; \delta(b_{i,\mu(i)}) = 1$. New target blocks for train k: $b^k(1) = b^k(2)$ and $b^k(2) = b_{\mu_k(i),\mu_k''(i)}$

- If the train arrives at a node of a station served by mission z, passengers alight from and board the train in accordance with the principles of the hybrid model. Calculation of time on platform waiting for the train.
- Update of lists $E_{L,H}$ and $\bar{E}_{L,H}$:

-If $\delta\left(b^k(2)\right) = 0$, the two target blocks of k are still free and the train can continue to advance. The pair $(k, \mu_k(i))$ are reintroduced to the set of events for processing $E_{L,H}$ with the corresponding times $H^-_{\mu_k(i)}(k_z), H^+_{\mu_k(i)}(k_z)$, which means that in the list $E_{L,H}$ there are the following inequalities: $H^+_{i_{k1}} < H^+_{\mu_k(i)}(k_z), < H^+_{i_{k2}}$, where $(k1, i_{k1})$ is the element preceding it $(k, \mu_k(i))$ and $(k2, i_{k2})$ is the element following it.

- If $\delta\left(b^k(2)\right) = 1$, the second target block is still occupied. The pair $(k, \mu_k(i))$ is added to the set $\bar{E}_{L,H}$.

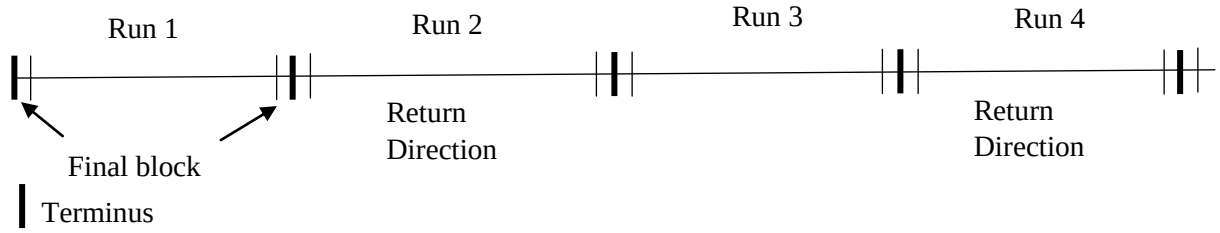
- If $\exists k', b^{k'}(1) = b_{\lambda(i),i}$ *or* $b^{k'}(2) = b_{\lambda(i),i}$ and $\delta\left(b^{k'}(1)\right) = 0$ *and* $\delta\left(b^{k'}(2)\right) = 0$, all the trains that fulfil this condition are denoted K'. There may be several in the case of the convergence of a junction or an initial terminus. The train that corresponds to the departure closest in time is added to the set $E_{L,H}$:

$$E_{L,H} \leftarrow \left(k'_z, i(k'_z)\right) \; such \; that \; \forall k' \in K', \tilde{H}^-_{i(k'_z)}(k'_z) < \tilde{H}^-_{i(k')}(k')$$

End

2.2.5 Train turn-back at the terminus

On a transit line, the rolling stock rotates between the two line directions. The missions divide complex lines into regular loops in which the two directions are symmetrical. The principle adopted is to follow the trains through their successive runs from an initial selected direction. The train itinerary is then no longer defined at the level of the run but at the level of the line with a list of runs to serve. For reasons of simplicity, all trains begin with runs in the same direction and turn back at the end of the run to travel in the reverse direction and so on. The total volume of rolling stock available for the simulation period is either known to the operator or calibrated according to the time taken for the first run to finish the loop. In other words, new trains supply runs until the first train to depart in the simulation returns to its starting point.



3 Assignment of demand

Passengers are introduced into the model through the station directly onto the platforms on the line. Their itineraries are defined as a pair of origin and destination stations on the line. A flow per step time aggregates passengers for common itineraries. It is assumed that the arrival of passenger flows in the station is uniform over the temporal discretization interval. Passengers are included in the model from the moment they start waiting on the platform of origin to board the first train available that serves their destination, until they alight at the destination station. These flows can be calculated by a static

assignment model at the scale of the transit system or extracted directly from ticketing data for the dynamic profiles and from survey results for the average flow, which we will use in the case study for greater reliability in the values.

3.1 Principle

The hybrid line model assigns the flows of passengers on the line on the basis of station to station demand. We will call it the demand model. The supply model will be used as a mechanism to call computation functions from the demand model. On the basis of the real times of arrival in the station and departure from the station, the demand model calculates the number of passengers who board each train. The exchanges between the 2 models is summed up in the diagram below (figure 2):

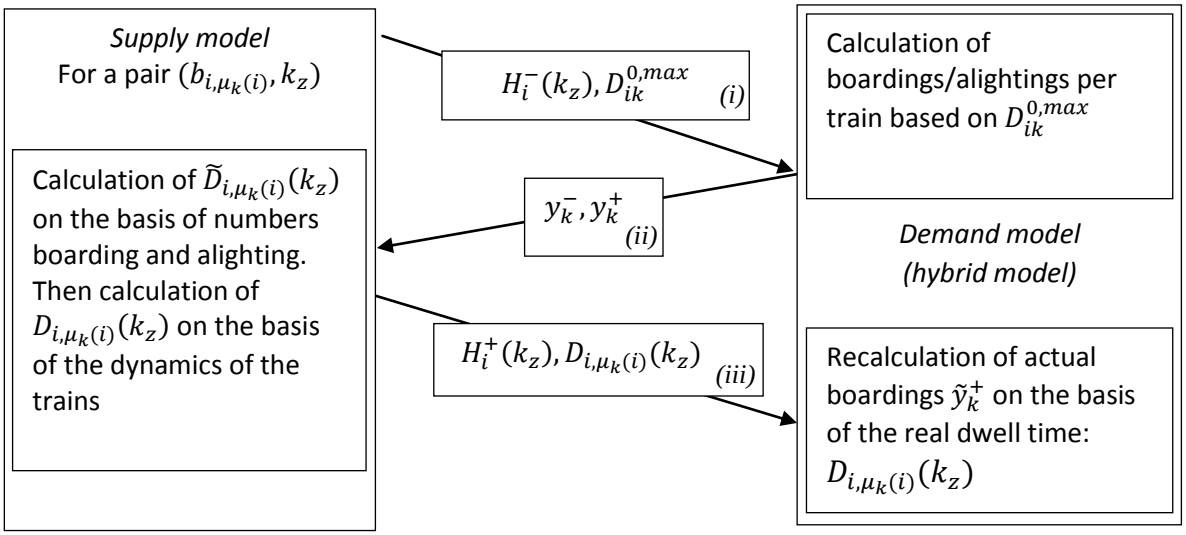


Figure 2: Diagram of the interaction between passenger platform-train exchanges and calculation of the exchange time

Each time a block-run pair $(b_{i,\mu_k(i)}, k_z)$ is processed on a station block, several exchanges between the 2 models are performed.

- *(i)* The first is used by the hybrid model to find out the flows y_{ik}^+ boarding train k in station i. The supply model provides the arrival time of k in the station $H_i^-(k_z)$, as well as the maximal dwelling time $D_{ik}^{0,max}$.
- *(ii)* The second is used to calculate the real dwell time governed by the exchange flows: y_k^- alighting and y_k^+ boarding and the congestion condition.
- *(iii)* The third makes it possible, if residual train capacity κ_{ik} is positive, to continue loading a train that stays at the platform longer than expected.

Part 3 picks up and details the three points (i), (ii) and (iii).

3.2 Calculation of dwell time

Dwell time in the hybrid model is a piece of exogenous data that governs train boarding and residual capacity. In the present model, dwell time is endogenous and varies according to demand and the level of rail traffic. Dwell time will be calculated in several steps. A first step calculates dwell time on the basis

of the numbers of passengers boarding and alighting with a time function that depends on the physical exchange constraints. If the system is running smoothly, the train can leave and this dwell time corresponds to the final dwell time. On the other hand, if the congestion on the trains forces the train to remain in the station for longer, dwell time will increase until the downstream traffic allows the train to start. The additional time for which the doors are open is available to allow passengers to enter the train, provided that there is sufficient capacity. However, dwell time can occasionally be increased because of external factors, such as a passenger blocking a door. An external delay can then be added exogenously. We will see how this is done later on.

3.2.1 State of the art

Most of the literature looks at station exchange time for buses, or for the BRT in China (Li et al., 2012), which is similar to the rail system except that passengers rarely board and alight through the same door. Christoforou et al. (2016) reviewed the state-of-the-art of existing studies and statistical models for urban rail. In particular, there are many models, from the simplest TCQSM (2003), which only considers boarding and alighting at the most critical door, to the most sophisticated – Weston's model verified by Harris and Anderson (2007) on the London and Hong Kong rail networks. Even more recently, one study has shown that the parameter that is most crucial to exchange time is in-train congestion, arguing that this explains much of the increase in dwell time. However, this study is limited to only one station and does not compare several different congestion situations.

3.2.2 Exchange time calculation model

Initially, we calculate the exchange time solely on the basis of the time taken for boarding and alighting. Delays in the movement of the trains are incorporated subsequently, as is explained in the next section.

The calculation is inspired by Suazo et al. (2017), who – on the basis of videos on Line 1 of the Santiago metro in Chile – develop an empirical dwell time calculation model. They use the flow of passengers boarding and alighting and in-train density, but also calculate the dwell time of the passenger flows that have not boarded the train. However, the variable that seems dominant is solely the variable concerning passengers who do not board. We choose instead to keep a term for the density of passengers waiting on the platform for greater continuity between phases with and without congestion.

We propose to use the following formula to calculate our theoretical dwell time. This one takes density into account

$$\begin{cases} D_{ik} = \alpha . \delta_{ik}^{T} . (\delta_{ik}^{P} . D_{ik}^{-} + D_{ik}^{+}) \\ \delta_{ik}^{P} = \exp(\alpha_P . d_i) \\ \delta_{ik}^{T} = \exp(\alpha_T . d_k) \\ D_{ik}^{-} = \theta_k^{-} . y_k^{-} / u_k \\ D_{ik}^{+} = \theta_k^{+} . y_k^{+} / u_k \end{cases} \quad (3.1)$$

where $(\alpha, \alpha_P, \alpha_T)$ is a fixed set of parameters, $(\theta_k^{-}, \theta_k^{+})$ the time for a passenger to board/alight from the run k_z and u_k the number of passenger channels per door, i.e. the number of passengers able to alight/board simultaneously at the same time per door on average. δ_{ik}^{P}, δ_{ik}^{T} are multiplier terms applied to exchange time caused by congestion respectively on the platform and on the train. d_i respectively d_k is the passenger density on the platform/on board the train before passengers alight. D_{ik}^{-}, D_{ik}^{+} are more specifically the terms for the exchange time. High passenger density on the train forces passengers to

alight and re-board, increasing the dwell time exponentially. Furthermore, these passengers interfere with boarding time. Symmetrically, high passenger density on the platform obstructs alighting passengers, but this is independent of boarding time. Finally, D_{ik} is the sum of alighting time and boarding time with classical calculations weighted by congestion terms on board the train and on the platform.

The expected dwell time also depends on the minimum limit D_{ik}^{0} and on the maximum limit $D_{ik}^{0,max}$ as well as the time taken for the doors to open and shut $t^{o+c}(k)$:

$$\widetilde{D}_{i,\mu_k(i)}(k_z) = t^{o+c}(k) + \min\left(D_{ik}^{0,max}, \max\left(D_{ik}^{0}, D_{ik}\right)\right) \qquad (3.2)$$

3.3 Assignment per train

3.3.1 Train alighting

Let $y_{k,s}^{(i)}$ be the flow alighting at i from the origin station s for the run k_z. When this run k_z arrives at a station i, the total flow $y_{ik}^{-} = \sum_{s<is} y_{k,s}^{(i)}$ alights. The expected residual capacity is the total capacity of the rolling stock minus the remaining passengers. Alighting flow refills the residual train capacity from the previous station $\lambda(i)$, $\kappa_{\lambda(i)k}$: $\kappa_{ik} = \kappa_{\lambda(i)k} + y_{ik}^{-}$.

3.3.2 Capacity available for boarding

Boarding capacity could be restricted by the residual capacity and by the maximum expected dwell time $D_{ik}^{0,max}$. For the second restriction, the dwell time formulation can be used to calculate the associated maximum boarding flow, in which the dwell time calculated is equal to $D_{ik}^{0,max}$. In consequence, the capacity available for boarding candidates $\bar{\kappa}_{ik}$ is

$$\bar{\kappa}_{ik} = \min\left\{\kappa_{ik}, \left(\frac{D_{ik}^{0,max}}{\alpha.\delta_{ik}^{T}} - \delta_{ik}^{P} D_{ik}^{-}\right).\frac{\mu_k}{\theta_k^{+}}\right\} \qquad (3.3)$$

3.3.3 Boarding candidates

For the boarding procedure, the number of boarding candidates depends on the stations served by the run. If the residual capacity is insufficient, we apply a FIFO principle, where the first passengers to arrive on the platform have priority. The remaining boarding candidates form the platform stock, which is stored for each run to order the candidates in a queue. For the service station s, successfully boarded candidates $\hat{y}_{k,ii}^{(s)}$, who are limited by the available capacity $\bar{\kappa}_{ik}$, form the total number of boarding passengers $y_k^{+} = \sum_{s>i} \hat{y}_{k,i}^{(s)}$.

A waiting time could be deduced for each train and each destination by resolving a bottleneck model. For a detailed description of the assignment procedure and the waiting time model, readers are referred to Poulhès et al. (2018).

Drawing on the model explained in Leurent (2012), an average time spent sitting or standing per origin-destination pair is used to calculate a generalized in-train time on the basis of each train's load.

3.3.4 Additional boarding due to train delay

In case of additional delay as train departure time is upper than $H_i^+(k_z) > H_i^-(k_z) + D_{i,\mu_k(i)}(k_z)$ users can continue to board the train. If we define the additional flow as $\tilde{y}_k^+$, according to $q_{is}([H_i^-(k_z) + D_{i,\mu_k(i)}(k_z), H_i^+(k_z)])$ the exogenous flow arriving at the station i to s during the additional interval time:

$$\tilde{y}_k^+ = \min\left(\kappa_{ik}, \sum_{s>i} q_{is}\left([H_i^-(k_z) + D_{i,\mu_k(i)}(k_z), H_i^+(k_z)]\right)\right) \quad (3.4)$$

4 Block crossing

This section describes the calculation of the times $H^-_{\mu_k(i)}(k_z)$ and $H^+_{\mu_k(i)}(k_z)$ of run k_z in terms of all the interactions included in the model: the progress of the previous trains, potential delays caused by operations or passengers, in particular the increase in platform exchange time. We will begin by presenting the case of a fixed block, and then an ETCS Level 2 type block.

4.1 Fixed block case

4.1.1 Principle

By construction, the trains move forward by one block which means that at the moment the train leaves node i we necessarily know that block $b_{i,\mu_k(i)}$ is free, since that is what is required with this type of signaling system. On the other hand, it is possible for a train to be present on block $b_{\mu_k(i),\mu_k''(i)}$. We call this train $\beta_i(k)$ – it can change along the line and depends on the location of i on the line, because of the existence of junctions. In the case of a fixed block line, the signal at a node i depends on whether or not a train is present in block $b_{\mu_k(i),\mu_k''(i)}$ between $\mu_k(i)$ and $\mu_k''(i)$.

4.1.2 For instance

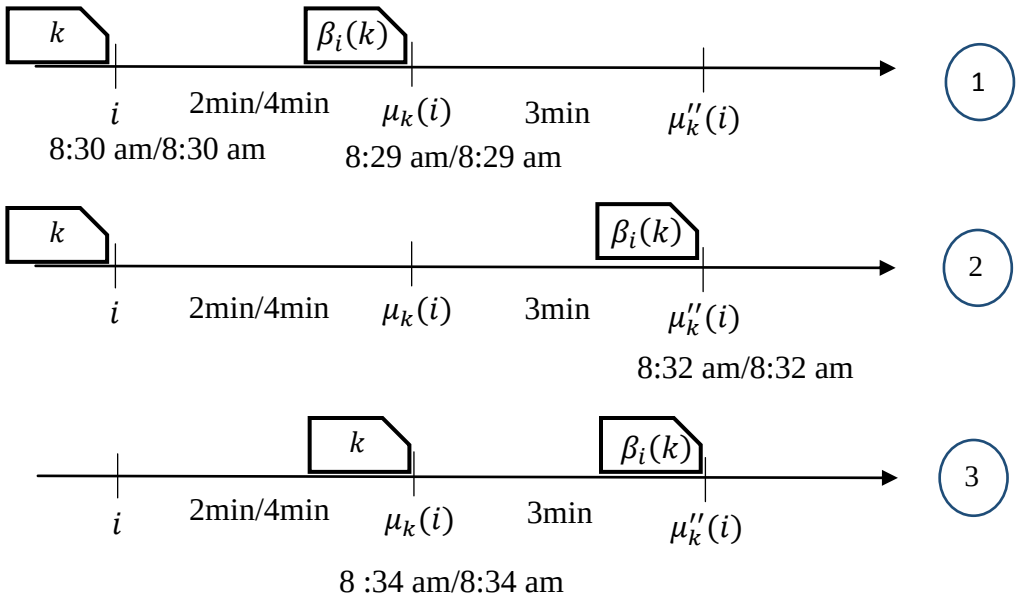


Figure 3: Simple case of block crossing

In figure 3, the initial diagram (1) depicts the case where a first train releases space for the next train. The first step (2) is the previous train's progress $\beta_i(k)$. This releases block $b_{i,\mu_k(i)}$ and thus the train k switches from the list $\bar{E}_{L,H}$ to the list $E_{L,H}$, the list of events to be processed. When the train k becomes the one with the closest departure time, it can move on to the next block (3). The exit time of train $\beta_i(k)$ is compared to the entry time k. The arrival time of k is earlier than the departure time of $\beta_i(k)$. The signal for $b_{i,\mu_k(i)}$ is then orange. The travel time is also 4 minutes and the arrival time in $\mu_k(i)$ is 8:34 a.m. The departure time from $\mu_k(i)$ depends on the departure time of $\beta_i(k)$ from $\mu_k''(i)$. In this instance, the gap is positive, the signal for the train is then green and the departure time from $\mu_k(i)$ is 8:34 a.m.

4.1.3 Calculation

An event corresponds to the movement of the train belonging to the first element in the list $E_{L,H}$, upon which that element is removed from the list. A new element is created with this train and added in $E_{L,H}$ or $\bar{E}_{L,H}$according to the conditions of circulation. If necessary, other elements move from list $\bar{E}_{L,H}$ to $E_{L,H}$.

The first step is to know the color of the signal at i. This enables us to calculate the speed of k on section $b_{i,\mu_k(i)}$. For this, we need to compare the departure time of the last train on the second target block of k_z, $b_{\mu_k(i),\mu_k''(i)}$ with the arrival time of k_z on target block $b_{i,\mu_k(i)}$.

If $H_i^+(k_z) < H_{\mu_k(i)}^+(\beta_i(k_z))$ means that $\beta_i(k_z)$ is on $b_{\mu_k(i),\mu_k''(i)}$ at the time $H_i^+(k_z)$.

From this we deduce:

$$\begin{cases} \sigma_i\left(H_i^+(k_z)\right) =' o' \\ t_{i,\mu_k(i)}(k_z) = t_{i,\mu_k(i)}^o(k_z) \end{cases}$$

If $H_i^+(k_z) > H_{\mu_k(i)}^+\left(\beta_i(k_z)\right)$ this means that $b_{\mu_k(i),\mu_k''(i)}$ is free at time $H_i^+(k_z)$.

Hence,

$$\begin{cases} \sigma_i\left(H_i^+(k_z)\right) = 'v' \\ t_{i,\mu_k(i)}(k_z) = t_{i,\mu_k(i)}^v(k_z) \end{cases}$$

The time of arrival at node $\mu_k(i)$ will depend on the nature of block $b_{i,\mu_k(i)}$. If it is an interstation block, only the travel time $t_{i,\mu_k(i),}$ will determine $H_{\mu_k(i)}^-(k_z)$. On the other hand, if $b_{i,\mu_k(i)}$ is a station block, the stop time must be added. This stop time will depend on the exchange time of passengers boarding and alighting (see equation (4.1)) but will also depend on whether or not there is a train present in the next block (equation (4.2)). If a train $\beta_i(k_z)$ is present, the signal at the end of the station will be red and the train will have to wait until the train has left block $b_{\mu_k(i),\mu_k''(i)}$.

(i) Case $b_{i,\mu_k(i)}$ interstation arc or station not served by run k_z.

For both types of signals, the arrival and departure times of run k_z can be calculated at node $\mu_k(i)$:

$$\begin{cases} H_{\mu_k(i)}^-(k_z) = H_i^+(k_z) + t_{i,\mu_k(i)}(k_z) \\ H_{\mu_k(i)}^+(k_z) = \max(H_{\mu_k(i)}^-(k_z), H_{\mu_k''(i)}^+\left(\beta_i(k_z)\right)) \end{cases} \quad (4.3)$$

(ii) Case $b_{i,\mu_k(i)}$ station block where k_z stops, $b_{i,\mu_k(i)} \in B_S(z)$.

$$\begin{cases} H^-_{\mu_k(i)}(k_z) = H^+_i(k_z) + t_{i,\mu_k(i)}(k_z) \\ H^+_{\mu_k(i)}(k_z) = \max(H^+_i(k_z) + \widetilde{D}_{i,\mu_k(i)}(k_z), H^+_{\mu''_k(i)}(\beta_i(k_z))) \\ D_{i,\mu_k(i)}(k_z) = \max\left(\widetilde{D}_{i,\mu_k(i)}(k_z), H^+_{\mu_k(i)}(k_z) - H^-_{\mu_k(i)}(k_z)\right) \end{cases} \quad (4.4)$$

4.2 Case of an ETCS Level 2 type block (RER A)

In the case of a fixed block line and a radio link between the trains and the operations center, the signal is simplified to 2 colors, red and green. The train at i only needs to know that block $b_{i,\mu_k(i)}$ through which it has to travel is empty up to $\mu_k(i)$. Depending on the future occupancy of block $b_{\mu_k(i),\mu''_k(i)}$, the train will adopt its behavior in block $b_{i,\mu_k(i)}$. The train is likely to slow down, but only to stop if the safety margin with the previous train $\beta_i(k_z)$ makes it necessary. To simplify the approach, this time will be counted as the difference between the arrival time at node $\mu_k(i), H^-_{\mu_k(i)}(k_z)$ and the departure time from $\mu_k(i)$, $H^+_{\mu_k(i)}(k_z)$.

The specificity of this type of block in the model will be that the trains will continue to advance from block end to block end, but their progress will no longer depend only on the signal, but also on the safety margin with the previous train. In reality, if a previous train is occupying the second block $b_{\mu_k(i),\mu''_k(i)}$, the next train will decelerate in block $b_{i,\mu_k(i)}$ without necessarily stopping at $\mu_k(i)$. In the model, the next train rolls during the free time $t^g_{i,\mu_k(i)}$ and stops as required at the next signal.

The progress of events will therefore be slightly different than for a standard fixed block. The target block just has to be free for train k to be able to move forward. The list $E_{L,H}$ will therefore contain all the runs for which the target block is free. In addition, the expected departure time from node i, $H^+_i(k)$ will depend on the previous train's time of departure from node $\mu_k(i)$ because of the operating strategies. In the list $E_{L,H}$, therefore, an expected departure time from node i is assumed, $\widetilde{H}^+_i(k'_{z'})$ calculated at the preceding event of a train $k'_{z'}$.

For each event, therefore, $H^+_i(k_z)$ and $H^-_{\mu_k(i)}(k_z)$ are calculated.

4.2.1 Departure from an interstation block

If $i \in N_I$, i.e. if node i is not a station node:

The expected time for run k_z to reach node $\mu_k(i)$ is

$$\widetilde{H}^-_{\mu_k(i)}(k_z) = H^+_i(k_z) + t^v_{i,\mu_k(i)}(k_z) \quad (4.5)$$

This will be the time achieved if the time gap between the departure from $\mu_k(i)$ of the previous train $\beta_i(k_z)$ and train k_z is sufficient, i.e. greater than the safety margin between trains ω_i.

We write $H^{o-}_{\mu_k(i)}(k_z)$ for the operating time needed for run k_z *to* arrive at i with a sufficient time gap from train $\beta_i(k_z)$ which precedes it:

$$H^{o-}_{\mu_k(i)}(k_z) = H^+_{\mu_k(i)}(\beta_i(k_z)) + \omega_i \quad (4.6)$$

(i) If the safety margin between the trains is sufficient

Train k_z does not accumulate any delay from the previous trains if $\widetilde{H}^{-}_{\mu_k(i)}(k_z) \geq H^{o-}_{\mu_k(i)}(k_z)$. In this case, we have:

$$\begin{cases} H_i^+(k_z) = H_i^-(k_z) \\ H^-_{\mu_k(i)}(k_z) = H_i^-(k_z) + t^v_{i,\mu_k(i)} \end{cases} \quad (4.7)$$

(ii) If the margin is insufficient, the train must slow down or stop

If $\widetilde{H}^{-}_{\mu_k(i)}(k_z) < H^{o-}_{\mu_k(i)}(k_z)$ the run k_z will not be able to reach the target station at the expected time. Since all the delays are attributed in waiting time at the end of the previous block, a departure time from i will be calculated, $H_i^+(k_z)$, which assigns all the downstream delays to be attributed. The minimum departure time from station i corresponds to an equality between $\widetilde{H}^{-}_{\mu_k(i)}(k_z)$ and $H^{o-}_{\mu_k(i)}(k_z)$. This gives us:

$$H_i^+(k_z) = H^+_{\mu_k(i)}\big(\beta_i(k_z)\big) + \omega_i - t^v_{i,\mu_k(i)} \quad (4.8)$$

We therefore have the pair:

$$\begin{cases} H_i^+(k_z) = H^+_{\mu_k(i)}\big(\beta_i(k_z)\big) + \omega_i - t^v_{i,\mu_k(i)} \\ H^-_{\mu_k(i)}(k_z) = H_i^+(k_z) + t^v_{i,\mu_k(i)} \end{cases} \quad (4.9)$$

4.2.2 Departure from a station block

(i) If the safety margin between trains is sufficient

Taking $\widetilde{D}_{\lambda(i),i}(k_z)$ as the expected dwell time of train k at station $b_{\lambda(i),i}$, the time for train k to reach the station $\mu(i)$ is

$$\begin{cases} H^-_{\mu_k(i)}(k_z) = H_i^+(k_z) + D_{\lambda(i),i}(k_z) + t^v_{i,\mu_k(i)}(k_z) \\ with\ H_i^+(k_z) = H_i^-(k_z)\ and\ D_{\lambda(i),i}(k_z) = \widetilde{D}_{\lambda(i),i}(k_z) \end{cases} \quad (4.10)$$

(ii) If the safety margin is insufficient, $H^-_{\mu_k(i)}(k_z) + \omega_i < H^+_{\mu_k(i)}\big(\beta_i(k_z)\big)$

a. If $H_i^-(k_z) + \widetilde{D}_{\lambda(i),i}(k_z) < H^+_{\mu_k(i)}\big(\beta_i(k_z)\big)$ train k knows without having left station i that it will not arrive at the next station at the expected time. The train will therefore be able to wait for the necessary time at station i, and we simply take $H_i^+(k_z)$ the departure time from station i as the maximum between the time $H_i^{D+}(k_z)$ potentially degraded by an increase in the exchange time, $H_i^{D+}(k_z) = H_i^-(k_z) + \widetilde{D}_{\lambda(i),i}(k_z)$ and $\widetilde{H}_i^+(k_z)$, the departure time considering knock-on delay from the previous trains, $\widetilde{H}_i^+(k_z) = H_i^-(k_z) + \widetilde{D}_{\lambda(i),i}(k_z) + t^v_{i,\mu_k(i)}(k_z)$.

$$\begin{cases} H_i^+(k_z) = \max\left(\widetilde{H}_i^+(k_z), H_i^{D+}(k_z)\right) \\ H_{\mu_k(i)}^-(k_z) = H_i^+(k_z) + t_{i,\mu_k(i)}^v \\ D_{\lambda(i),i}(k_z) = H_i^+(k_z) - H_i^-(k_z) \end{cases} \quad (4.11)$$

b. Otherwise train k leaves station i without knowing that it will not be able to reach the target block at the expected time. In this case, the departure time from i: $H_i^+(k_z)$ is the arrival time plus the expected dwell time and the arrival time in $\mu_k(i)$ depends on the departure time of the previous train. If the safety margin is sufficient to reach the next node, arrival time $H_{\mu_k(i)}^-(k_z)$ is the sum of $H_i^+(k_z)$ and travel time, although $H_{\mu_k(i)}^-(k_z)$ corresponds to the safety margin with the previous train $H_{\mu_k(i)}^+\left(\beta_i(k_z)\right) + \omega_i$.

$$\begin{cases} H_i^+(k_z) = H_i^-(k) + \widetilde{D}_{\lambda(i),i}(k_z) \\ H_{\mu_k(i)}^-(k_z) = \max\left(H_i^+(k_z) + t_{i,\mu_k(i)}^v, H_{\mu_k(i)}^+\left(\beta_i(k_z)\right) + \omega_i\right) \\ D_{\lambda(i),i}(k_z) = \widetilde{D}_{\lambda(i),i}(k_z) \end{cases} \quad (4.12)$$

4.2.3 Disruptions

Only 2 types of disruption to the progress of the trains are considered in our model. The first is the type of delay on the current train that does not depend on the previous trains. The second is the delay propagated by the previous trains on the line, which occurs for example on lines with high train frequencies or poor scheduling.

Primary delay

Primary delay $R_i'(k_z)$ is delay experienced by the current train k_z at node i, which does not depend on the previous trains on the track. In our model, the primary delay is obtained in only 2 ways:

(i) Endogenously with an increase in station dwell time, in interface with the hybrid model which calculates the expected boarding and alighting flows: $If\ \widetilde{D}_{i,\mu_k(i)}(k_z) > D_{ik}^0,\ R_i'(k_z) = \widetilde{D}_{i,\mu_k(i)}(k_z) - D_{ik}^0$

(ii) Exogenously by introducing a one-off delay caused by a problem on the line: $R_i'(k_z) = t_{i,\mu(i)}^{ex}$ in which $t_{i,\mu(i)}^{ex}$is an additional exogenous time. This fixed delay can be added to a train-block pair in the simulation to simulate an incident on the line.

Secondary delay

Secondary delay, $R_i''(k_z)$, is the delay passed on to the current train k_z at station i by the previous train $\beta_i(k_z)$ on the track. In this delay, therefore, only delays to the preceding trains which have not been transmitted to the nodes previous to i on the line are considered:

$$R_i''(k_z) = H_{\mu_k(i)}^+\left(\beta_i(k_z)\right) + \omega_i - H_i^+(k_z) + t_{i,\mu_k(i)}^v \quad (4.13)$$

4.3 Train circulation on a branch in the case of a junction

The aim of this paragraph is to define an operational strategy for the movement of the trains on a line with several branches. In the absence of incidents or congestion on the line, the missions follow each other as set out in the line plan. Otherwise, several operational methods can be implemented. The method chosen is to deal with congestion upstream. This method can be used to slow down the trains

on the branches or to have them leave their terminus later in order to anticipate train congestion at the junction. We will give a detailed description of the model for a line with this type of operation.

The principle is as follows: at the events associated with its progress, a train recalculates the projected time difference of arrival at the convergence relative to its successor on the shared section. If this time difference is sufficient to advance, it continues. Otherwise, the knock-on delay is attributed and the projected time is updated. And so on, step-by-step. This forward calculation is carried out for all branch convergences encountered by the train.

Let us define N_J as the set of convergence points on the line between 2 or more branches. For each mission, we define the successive convergence points $N_J(z) \in N_J$ and for each run $k_z \in z$, the expected antecedent $j(k_z)$ at junction $n_j(z) \in N_J(z)$ i.e. the train that will be the $\beta_i(k_z)$ of k_z when k_z reaches the node $n_j(z)$.

Let us take a train $k_z \in z$ waiting before junction $n_j(z)$ at node i at time $H_i^+(k_z)$. The projected departure time $H_{n_j(z)}^+(k_z)$ at junction $n_j(z)$ of train k_z is calculated from the time $H_i^+(k_z)$ at current station i and the travel times up to the junction:

$$H_{n_j(z)}^+(k_z) = H_i^+(k_z) + \sum_{i<j,j+1<n_J(z)} t_{j,j+1} \qquad (4.14)$$

In addition, the projected time $H_{n_j(z)}^+(j(k_z))$ of $j(k_z)$ is calculated. If $H_{n_j(z)}^+(k_z) < H_{n_j(z)}^+(j(k_z))$, the train k_z will arrive before $j(k_z)$ in $n_j(z)$. A delay is then applied to k_z corresponding to the difference: $H_{n_j(z)}^+(j(k_z)) - H_{n_j(z)}^+(k_z)$.

In generalizing to all junctions, the gap from the previous trains is calculated at each junction and the maximum gap is allocated to train k_z if its value is positive:

$$\Delta H_{N_J(z)}^+(k_z) = \max_{n_j(z) \in N_J(z)} \left(H_{n_j(z)}^+(k_z) - H_{n_j(z)}^+(j(k_z)) \right) \qquad (4.15)$$

In equations (4.3) and (4.4), the term $H_{\mu_k(i)}^+(k_z)$ is updated directly from the conditions of the previous trains on the line. In the case of a junction where the method of handling employed is congestion anticipation, it also depends on what is happening further up the line, at the successive convergences:

$$H_{\mu_k(i)}^+(k_z) = \max\left(H_{\mu_k(i)}^-(k_z), H_{\mu_k''(i)}^+(\beta_i(k_z)) \right) + \Delta H_{N_J(z)}^+(k_z) \qquad (4.16)$$

5 Case study: line 13 of the Paris metro

The model is coded with the Matlab software without using any specific library. A five-minute simulation with Matlab is run on Line 13, in which 3 hours of morning peak are simulated. Calculating the waiting time is the most time-consuming procedure.

5.1 Situation

Line 13 of the Paris metro crosses the city from North to South with, unlike most of the other lines on the network, a significant part of it outside the city of Paris. The line passes through Montparnasse Station

and Saint-Lazare Station. It also has a junction that divides the line in the North into 2 branches with 6 and 8 stations. This line is familiar to Parisians, since it is one of the most congested and therefore one of the least popular of the lines. Though recently renovated, the rolling stock dates back to the late 1970s. In order to limit problems of cascading congestion, sliding gates have been installed on certain congested platforms along the line, pending the automation of the line in 2025. The signaling on the line is a fixed block system, with one specificity, which is that a train cannot enter the 2 blocks following an occupied block. The model was adapted to this specificity. Like all the lines on the Paris Metro, it is operated by RATP. The line is mapped on figure 4 with blocks and stations.

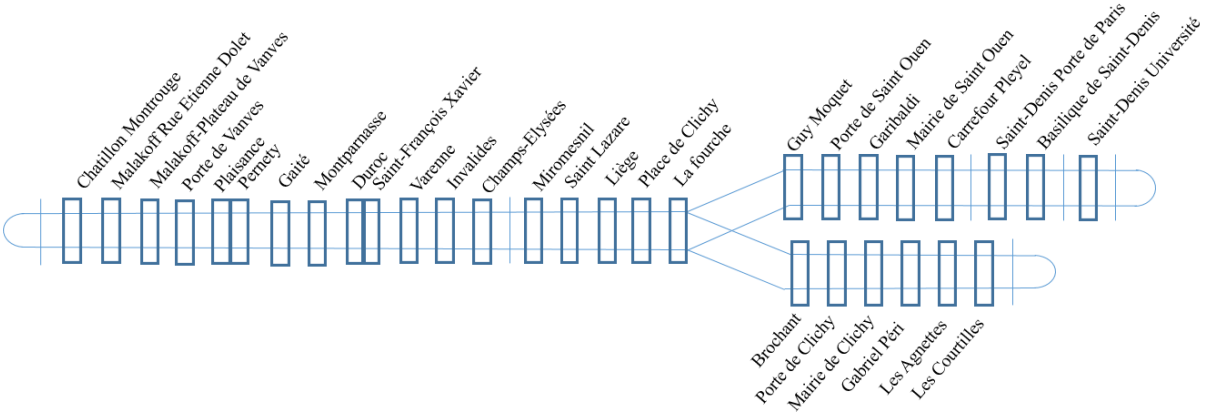


Figure 4: Both directions of line 13 with station and block details

5.2 The input data

In order to be applied, the dynamic line model in this article requires a number of input data. First, the routes of the line in each direction, with all the blocks, the distance between them and the theoretical travel time on each of them, as well as the programmed departure schedule for each run from its terminus. Second, the demand data are constructed from Navigo travel card data and from the TJRF (daily rail traffic) survey conducted by RATP, the operator of the Paris metro lines.

5.2.1 The representation of the line and the theoretical timetables

The central trunk consists of 32 blocks plus 17 on the Eastern branch, 11 on the Western branch, with 21 stations. The total distance on the central trunk is 10 km, 5.9 km on the Western branch and 7.5 km on the Eastern branch.

Two types of missions serve the line with the same rolling stock, one mission per branch alternately, providing service to all the stations visited. The frequency of the 2 missions is therefore theoretically the same.

RATP provided the signaling plan with the theoretical times per block as well as the theoretical departure times of all the runs at the terminus of each mission on the line. We will simulate the progress of the trains in the South-North direction over the period 7 a.m.-9 a.m., which corresponds to the morning peak period defined by the Île-de-France transport organizing authority, Île de France Mobilité (IDFM). 70 runs are scheduled during this period with an average expected headway of 1 minute 45 seconds.

The capacity of each train is calculated on the basis of the number of seats in each piece of rolling stock and its effective surface area for people standing, multiplied by a factor of 4 people per square meter. This corresponds to a total capacity per train of $188 + 96.5 * 4 = 574$.

In order to make the results easier to understand and to test some complex behaviors, the simulations presented correspond only to the south to north runs. Indeed, after a loop of trains, the results are difficult to analyze, since variation increases as the simulation progresses. What is the real influence of the variable? How to be sure that the results are consistent? The sensitivity analysis seek to understand the interaction between the circulation of users and trains without the historical situation of congestion on the line. That is why the results presented do not include train turn-back.

5.2.2 The users

The TJRF (Trafic Journalier Réseau Ferré – daily rail traffic) survey, based on surveys conducted in all the stations on the network, provides information on the origin and destination stations of every user per one-hour period on every railway and bus line in the RATP network. We will use the 2016 TJRF provided by IDFM for line 13 to obtain the value of the flows on each pair of stations on the line for the morning peak period (mpp). According to these data, 140,000 users travel on the line in the South-North direction during the mpp. The benefit of this kind of data is that they provide a good representation of total flow. By contrast with the automatic-fare-collection (AFC) data collected by IDFM and named NAVIGO travel card data, the survey includes ticket using passengers. Line 13 serves two interurban train stations with many visitor flows and transfer flows from other metro lines, who do not have to swipe their transit pass. For this reason, using AFC data to count total origin-destination flow would not be accurate. We also assume that users are restricted to the line and cannot decide to change their itinerary.

However, the survey data are static, so AFC data are used to obtain a representation of the dynamics of the flows on the line. AFC, where users swipe their travel cards when they enter the metro network represents a very significant proportion of mpp travelers. However, except on part of the RER, the regional express rail network, users do not swipe their cards when exiting. IDFM, which manages and supplies the data, reconstructed some of the journeys and routes on the basis of successive individual AFC data on the way in, since an entry may correspond to the exit from a previous trip. We use this reconstructed trip database to obtain an approximation of the dynamics of the flows on the line. Users who buy single tickets, who cheat or who make multiple connections, will not be included in the trips. Interurban flows coming from the stations will therefore not be included, as well as a not insignificant proportion of the big connection stations. That is why we will only use these data to divide up the flow of TJRF data in the mpp time interval. AFC data at the entrance to stations are timed to the second, but the card readers are generally located less than one minute from the platform. We will therefore consider the swipe time as rounded up to the minute, as with the time of arrival on the platform. The chosen time step is 5 minutes. The number of swipes is summed per step time and per station pair. They provide the probability of falling within each step time in the peak period. Final flow is this probability multiplied by the flow from the TJRF survey. Figure 5 represents these probabilities for 4 major stations on the line, in the North-South direction. The flows double between the busiest interval and the intervals at either end of the rush-hour.

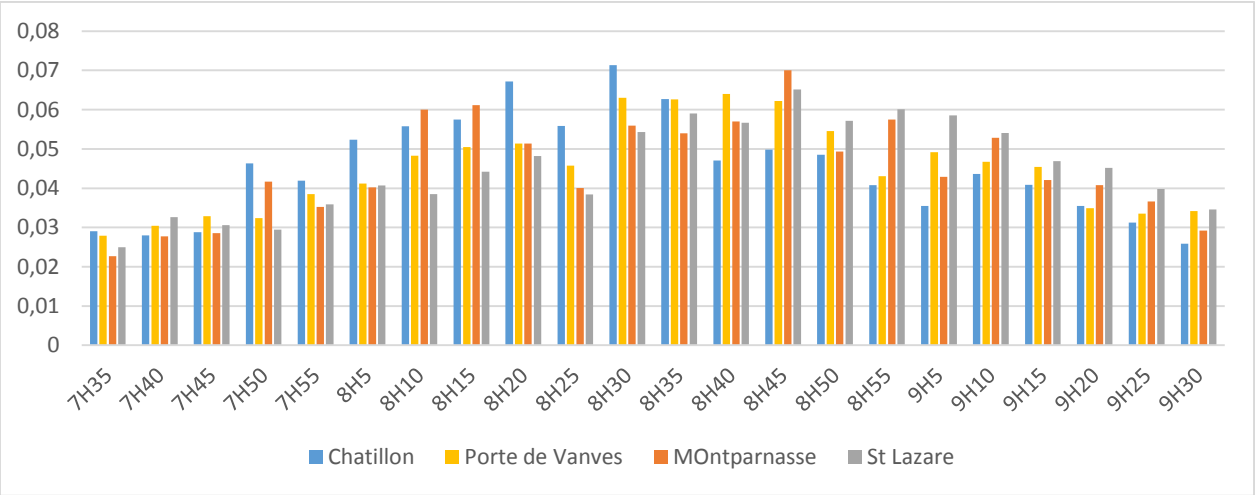


Figure 5: Travel card swipes per 5-minute time step for 4 stations on the South-North line

5.2.3 Calculating the exchange time

We choose the following parameters in the dwell time formula (Eq (3.1)) $(\alpha, \alpha_T, \alpha_P) = (4; 0.174; 0.174)$. These values come from Suazo et al. (2017). For density on the platform, we choose to take the surface area corresponding to half the length of the train multiplied by a depth of 2 meters, i.e. $\frac{68}{2} * 2 = 68\, m^2$. These platform density parameters are drawn from observed passenger behavior: passengers tend to aggregate where the doors open, emptying the rest of the platform area. The in-train density is calculated from the surface area used by standing passengers, on the basis of the characteristics of the rolling stock on Line 13 ($96.5\, m^2$). Finally, the number of channels per door is 2 channels for boarding or alighting, the minimum and maximum dwell time values are fixed for all the stations on the line and all the trains at respectively 10s and 50s. The door opening and closing time is set at 5s. Figure 6 confirms the logical response of dwell time value to the parameters of train and platform user densities for total boarding and alighting flows. An example of reading the graph: for an average on board density of 3p/m² and an average platform density of 3p/m², since the alighting flow is equal to the boarding flow (50 passengers/train), the corresponding dwell time is 31s. The dwell time formula is linear for boarding and alighting flows but exponential for platform and train densities. In figure 6, the dwell time curves are a positive function of the flows, therefore linear and increasing. A crowded platform and train increase the steepness of the dwell time curve, since it is difficult for exchange flows to move in the event of congestion. The default maximum dwell time is set at 50s in the reference situation.

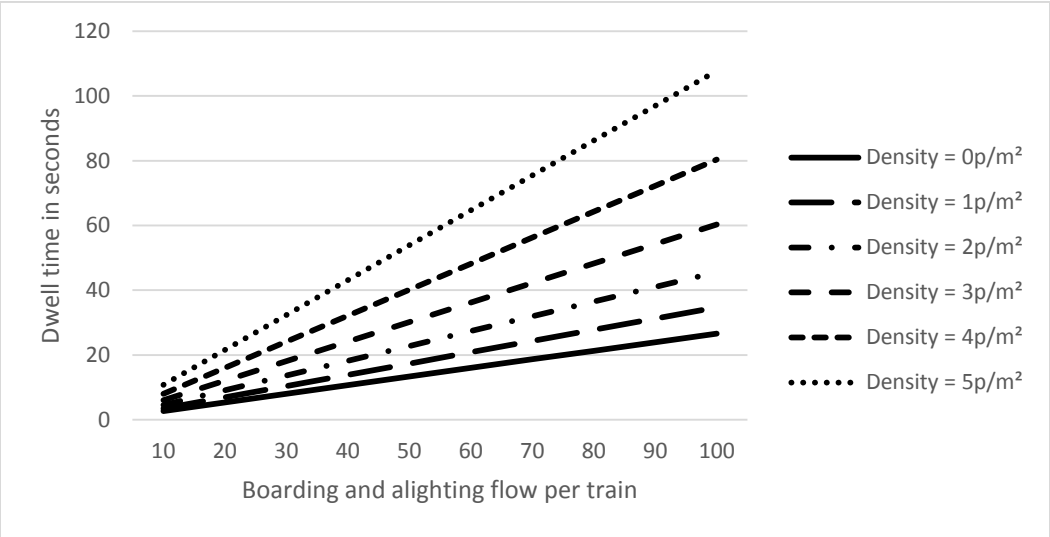


Figure 6: Dwell time values based on boarding and alighting flows per train and average on-board and platform density

5.3 Aggregated result

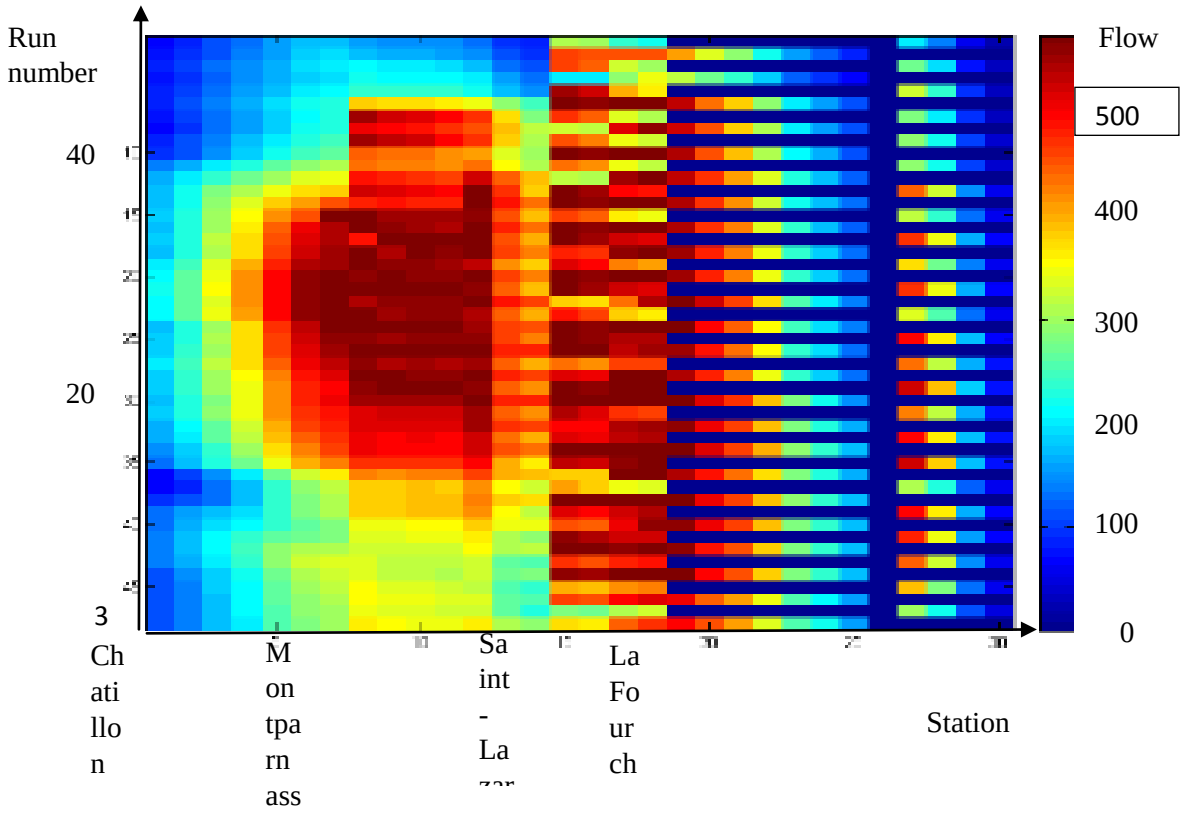


Figure 7: Morning peak-hour load of each train in the mpp for the south-north direction

Figure 7 depicts the total boarding flow per train limited by the train capacity of 574 passengers. The west branch is displayed first. This standard space-time representation reflects the peak hour, shown in red, and the spread of congestion. The intercity rail stations (Montparnasse, Saint-Lazare) are the main initial passenger distributors. The western branch also seems to be used more, but only its first three stations.

5.4 The quality of service indicators

We chose to assess the operation of Line 13 through 6 quality-of-service indicators. The first three describe the travel conditions for users and the last three the performance of the line operator.

(i) Average waiting time per origin station for all the trains in the simulation period.
(ii) The sum total of line users who were not able to board each train.
(iii) Average density of standing passengers for each train along the line.
(iv) The frequency per station on the line over the period 8 a.m.-9 a.m..
(v) The time of each run from terminus station to terminus station. This time does not include the delay between the theoretical departure from the terminus and the actual departure, which does not affect passenger travel time.
(vi) The dwell time calculated from the boarding and alighting of passengers, defined by equation (3.1).

In the three series of graphs in the next paragraph, the first line presents respectively indicators (i), (ii) and (iii) and the second indicators (iv), (v) and (vi). On the dwell time indicator (vi), we add a line that represents the maximum dwell time, which limits passenger boarding.

Indicators (i), (iv) and (vi) are calculated as an average per train for each station on the central trunk, i.e. between Chatillon-Montrouge and La Fourche. These indicators are averages in the mpp 8 a.m. to 9 a.m., the middle of the simulation time window. The most congested stations are stations 8 (Montparnasse) and 15 (Saint Lazare), which are the busiest stations. The other indicators (ii), (iii) and (vi) are averages per station for each train in the total simulation time period: 7:30 a.m. to 9:30 a.m.

Passengers who miss their trains differ according to the train's mission. Missions in fact alternate on the line, with one out of two trains going onto the western branch, the other onto the eastern branch. All the trains are omnibus trains.

5.5 Sensitivity analysis

The 6 indicators described above are assessed with 3 sensitivity analyses:

(i) Total variation in demand. A homogeneous demand ratio is developed for all the pairs of stations on the line. The ratio values are 0.9 / 1 / 1.1 / 1.2 / 1.3. This ratio is a multiplier value of all demand volumes.
(ii) The number of channels per unit. The goal is to analyze the impact of increasing door size on the performance of the line. The values tested are 1.6 / 1.8 / 2 / 2.2 / 2.4.
(iii) The maximum dwell time, i.e. the maximum length of time chosen by the engineers to close the doors of their train in the station, regardless of the exchanges on the platform. Drivers tend to let as many passengers as possible board their trains to the detriment of overall line performance. The maximum dwell time ranges through the values 40s /50s /60s /70s /80s.

First of all, a brief explanation of the 6 graphs. The first one represents the average waiting time per user at each station on the central trunk. With a frequency of 20 trains, a waiting time of 15 minutes means that users fail to board on 4 trains. In the event of high congestion, the first stations on the line are often boarding stations and thus competition to board quickly increases waiting time as shown in the first graph in figure 8. The second graph reflects the waiting time of users through the total number of missing users per train. Obviously, as long waiting times indicate users missing several trains, such users are often counted on several consecutive trains until they succeed in boarding. Figures 8 and 9 show that waiting times along the line are disparate and users who miss trains are temporally disparate, with a peak of total train missing passengers in the middle of the simulation period (8:40 a.m.). The third graph is the temporal change in average standing density, which is considered homogeneous along the train and strictly less than 4 passengers/m². The fourth is the sum of the trains which serve each station along the central trunk. The fifth is the difference between the arrival time of the train at the last station on the central trunk and the departure time from the first station on the line for each train in the simulation. Thus, the curve is the time change in travel time on the central trunk, including dwell times and the traffic congestion. The last graph shows the average dwell time calculated on the basis of the user congestion and exchange flows at each station along the line. On figure 8, a horizontal line limits the value of the real dwell time.

Some general results: heavy congestion on the line reduces the planned morning peak hour frequency of 34 trains between 8 and 9 a.m. This applies not only after the critical congested stations but all along the line. The results show that frequency can increase along the line, depending on the dynamic of space-time congestion. Contrary to the assumptions of static models, the frequency changes significantly along the line, and contrary to the assumptions of the CapTA model (Leurent et al. 2014), frequency does not necessarily decrease with congestion.

Then some general remarks: the difference in flow and in the number of stations between the two branches is reflected in the big difference in the number of missing passengers between two consecutive trains with different destinations. Heterogeneity in the 2-hour morning peak hours is significant. There is a clear peak hour from 8 to 9 a.m. with an average of 20% longer travel time or a doubling of average on-board passenger density. The dwell time calculated for some simulations is limited by the fixed maximum dwell time in many stations, reflecting high levels of congestion. Between 8 and 8:30 a.m., the trains are heavily loaded, with an average density across the whole line reaching 3.5 passengers/m², as against a limit of s 4 passengers/m².

5.5.1 Demand sensitivity

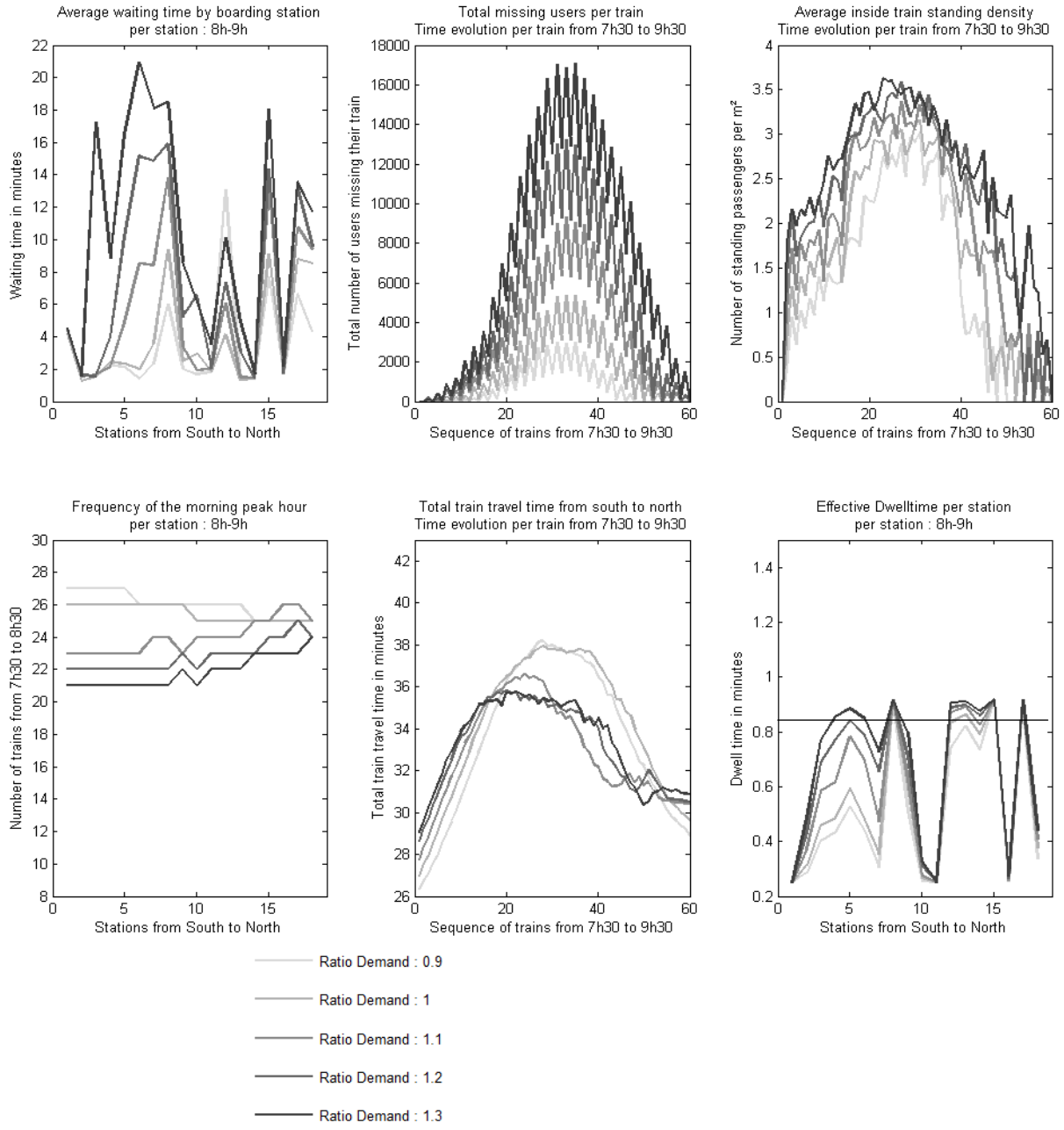


Figure 8: Results when the total demand ratio multiplier is modified (values tested: 0.9 / 1 / 1.1 / 1.2 / 1.3)

Increasing demand by 30% multiplies the number of passengers who miss a train by 30% and doubles waiting time (as shown in figure 8). On the other hand, frequency falls by an average of only 15%. This low frequency, combined with a fairly stable exchange time, close to the maximum value for all the simulations, results in reduced travel time on the line for the whole time period, on average 2 minutes less for the hyper peak period from 8 a.m.-9 a.m. In time dynamics, waiting time is increased in particular at the start of the line, up to Montparnasse Station (8). After that, only the very busy stations have travelers who experience long waiting times, regardless of demand. An interesting result is the non-

linearity of missing passengers with respect to demand ratio: at the maximum number of missing passengers, reducing initial demand by 10% cuts the number of missing passengers by half. Conversely, increasing demand by 10% doubles the number of missing passengers, and the same number of missing passengers are added with each 10% increase in demand. Complex phenomena appear with this dynamic. The travel time of the train is lower for high demand in the high peak period. This contradiction is explained by the lower frequency for high demand at the departure of the runs, which allows train traffic to flow more freely.

5.5.2 Maximal dwell time sensitivity

Figure 9 confirms that increasing door size logically leads to an increase in the speed of exchange on the platform, and therefore has positive repercussions for the performance of the line. The result of the simulations does indeed show a reduction in the exchange time at the least busy stations on the line, which leads to an increase in frequency and minimum waiting time. However, the benefits seem to come only at the beginning of the peak period and the increase in the number of passengers on the trains along the line degrades the quality of service, in particular travel time in the second part of the peak period. The curves on the graphs for missed trains reverse at the end of the peak period. The 1.6 channels/door provides poor quality of service, but all higher values have similar results.

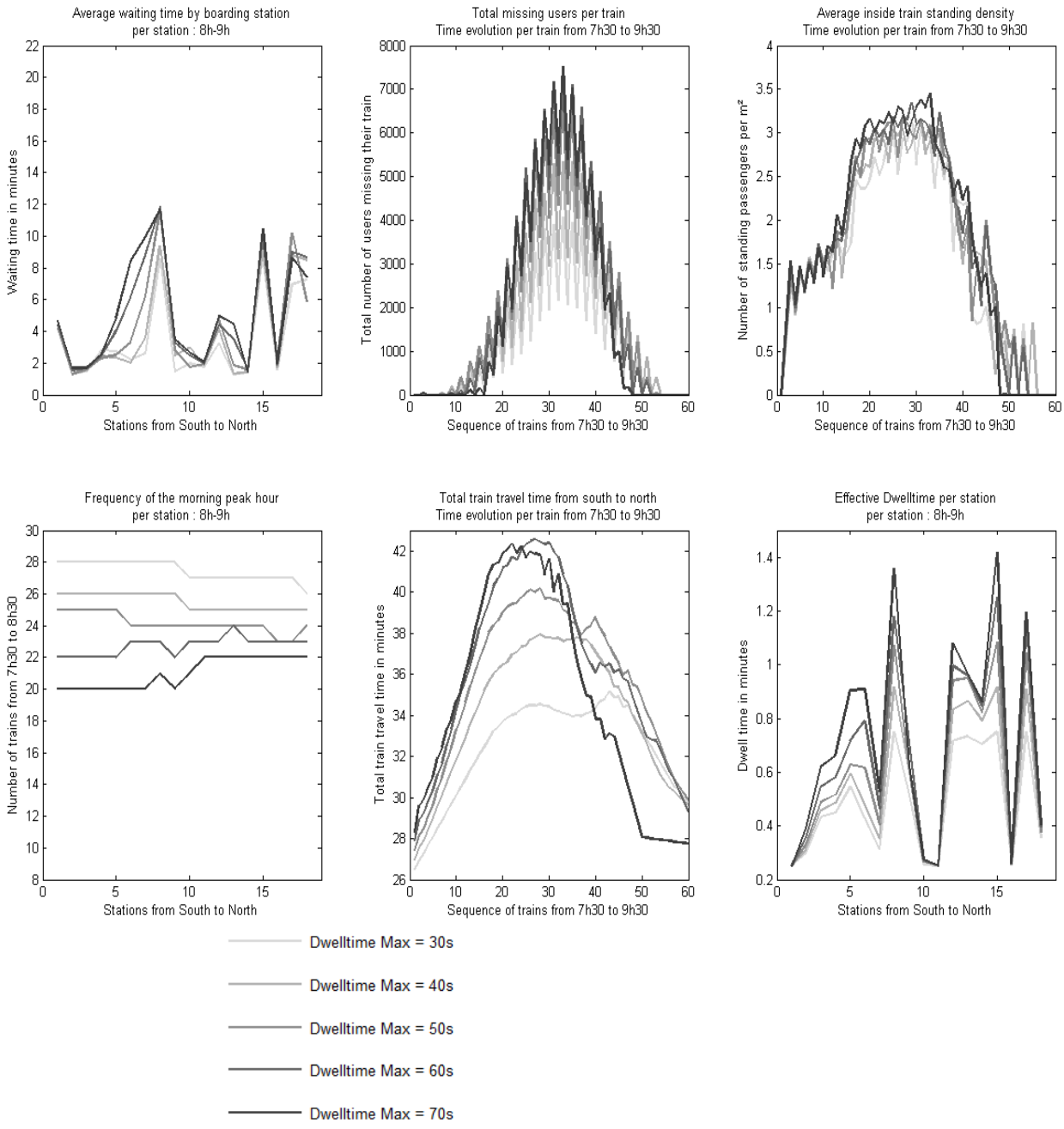


Figure 9: Results when the maximum dwell time value is modified (values tested: 40s /50s /60s /70s /80s)

Figure 9 illustrates that dwell time max is a central instrument of regulation and that quality of service deteriorates with an increase in dwell time. The peak period is more spread out when dwell time is limited. The frequency is initially reduced by more than 30% from the theoretical value with a maximum dwell time of 70s but improves along the central trunk. Yet limiting maximum dwell time leads to deteriorating conditions along the line. Similarly, at the end of the peak period, train travel time is shorter with a maximum dwell time of 70s. Another interesting result is the correlation between the dwell time calculated from boarding and alighting flows and the maximum dwell time. For the lowest

maximum dwell times, high frequency and low boarding and waiting flows limit the actual dwell time. A virtuous circle is created, limiting the impact of the maximum dwell time constraint on waiting flows. This simulation confirms an important result for regulation: limiting the dwell time max by forcing the doors to close improves the quality of service. However, at the train level, conductors intuitively feel that keeping the doors open benefit passengers. This is correct at the level of individual trains, but not at the line level where it leads to a deterioration in the average quality of service.

These original findings above all highlight time-related results that vary a great deal over the peak period on a congested rail line, which reinforces the need not to be limited to static passenger assignment models. Another result that emerges is the reality of peak times for users of the line, with an accumulation of passengers unable to board that rises to a maximum in the middle of the morning peak (8:10 a.m. on the line) and then diminishes symmetrically. The in-train passenger density indicator shows the same peak time phenomenon. The average number of standing passengers per train, which is close to 3 people per square meter at peak times, indicates a homogeneous load on the line (the maximum capacity is set at 4 people per square meter). Nonetheless, the dispersal of waiting time and dwell time on the line would suggest the opposite.

The scheduled frequency between 8 a.m. and 9 a.m. is, according to the theoretical timetables, 34 trains per hour in the South-North direction. This frequency is never achieved in the simulations, regardless of the values chosen. In all the simulations, when the frequency is low enough at the start of the line following congestion at the beginning of the peak period, it always propagates to the rest of the line. This confirms that the performance of the line is more robust when frequency is limited and the system can deal with one-off events such as a significant increase in dwell time at one station. However, this result can seem counterintuitive if the line is simply considered as a fluid with a constant initial rate of movement with a bottleneck at a congested station.

5.6 Operational benefit

5.6.1 General discussion

On congested urban rail lines, operators need specific tools to improve the quality of service provided to passengers. Our model is a first step towards integrating passenger constraints into train circulation models. Dwell time is central to the interaction between service and demand. Thus, parameters that influence dwell time such as door size, number of doors, rolling stock capacity or platform area can be changed to offer operators potential solutions for network congestion. Benefits are evaluated on the basis of passenger travel time broken down into two terms: waiting time and time spent on board in comfortable conditions. The second operational benefit lies in the ability to simulate operational strategies. For instance, maximum dwell time is mostly set by train conductors, and from their point of view allowing additional passengers on board is beneficial. But for the line system, we have shown that this has the effect of degrading average travel conditions.

5.6.2 Example of operational use: simulation of a short incident

To complete the discussions of the results, a short incident is simulated. A 5-minute delay at the 5th station is applied to the 25th train in the simulation. The spatiotemporal graph of the progress of the trains (figure 10) shows that the resulting delay has a heavy impact on up to 3 upstream trains, before

being gradually absorbed. Downstream from the disruption, the delay relative to the previous train is absorbed 20 blocks, i.e. 10 stations, away. The delay disappears on the branches, represented one after the other on the graph. On figure 11, for the 5th station where the incident occurs, the waiting time increases from less than one minute to 5 minutes on average for passengers boarding the 25th train. On-board travel time from the first to the 6th station confirms the mitigation of the impact of the incident 3 trains after the first delayed train.

The train congestion slows travel time along the line, as shown by the increasing gap after the first train, continuing until the 30th station. An additional delay at the beginning of the line allows free flow downstream of the delayed train. A fixed maximum dwell time is also essential in order to limit the impact of the increase in passenger congestion. These two combined effects absorb the one-off external delay.

As seen in this instance with high congestion, a short incident can be quite quickly absorbed. Keeping a maximum dwell time is therefore a primary method of regulation for the operator. If a reasonable maximum dwell time is not maintained, a short incident can get out of control and produce systemic congestion. Sensitivity tests on the maximum dwell time value confirm that on a congested line where conductors do not enforce door closures to limit dwell time, line performance is degraded. More studies are needed to evaluate the link between congestion and maximum dwell time value. Calibrating a maximum dwell time for platform flow density and in-train density could lead to a better understanding of the spread of space-time congestion in the event of an incident.

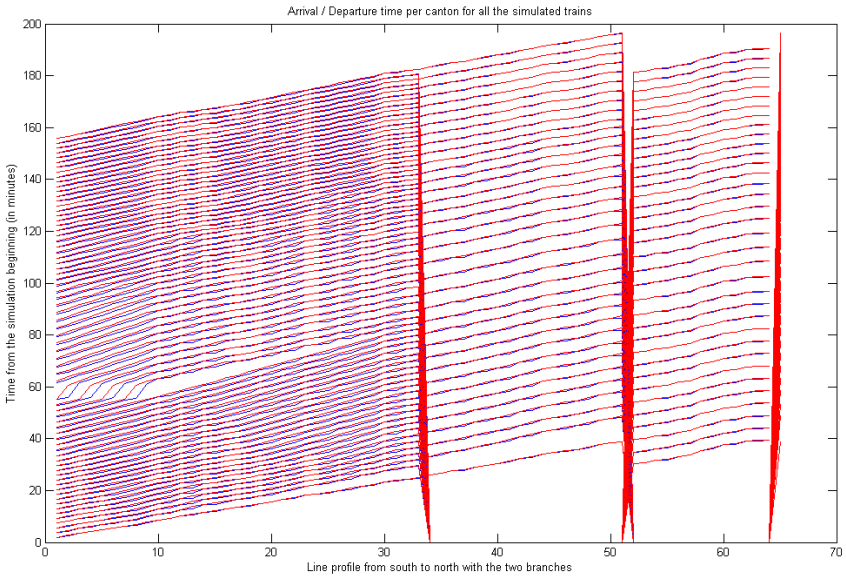


Figure 10: Spatiotemporal profile of the progress of the trains with application of a five-minute delay

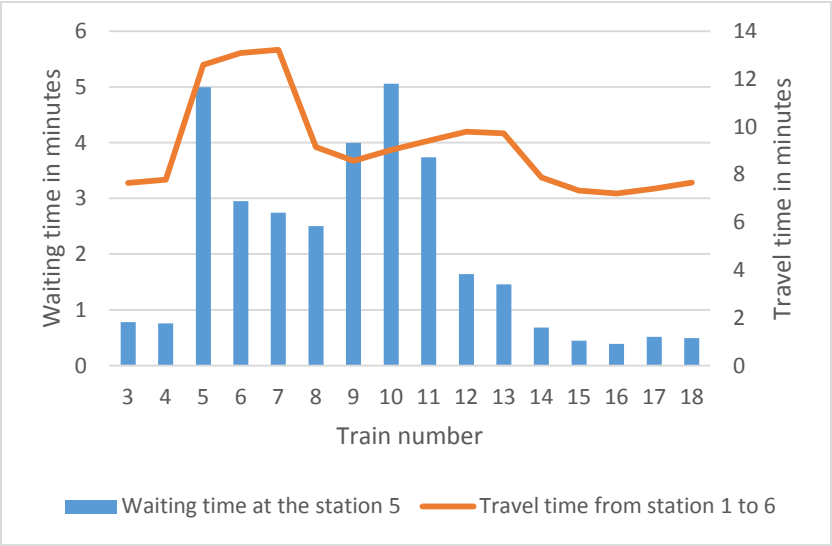


Figure 11: Waiting time at the 5th station and travel time from first station to the 6th station

Conclusion

In our research, the dynamic simulation of the operation of a railway line and of passenger assignment are handled interactively. We provide a discrete-event model for the progress of trains which can very easily be applied operationally. Our model for the dynamic assignment of passenger flows on a line uses the operator's passenger counts as well as card swiping data, in order to obtain accurate passenger numbers. The interaction between supply and demand takes place through a lengthening of the dwell time per train and per station, which affects the performance of the line and the travel conditions for passengers. An application of the model to the highly congested Line 13 of the Paris metro shows the main dynamic phenomena observable on congested lines: a drop in the frequency of the line with a maximum possible frequency at a certain level of demand, a significant deterioration in the quality of service with the increase in demand on the line, or a drop in frequency and a rise in congestion with the increase in maximum dwell time. These simulations also reveal finer dynamic phenomena along the line or in the peak period, notably with an increase in frequency along the line when congestion is high and a concentration of the peak with a high maximum dwell time, or the fact that lower frequency increases travel time. The behavior of the model is consistent with the operator's observations, but the values obtained should be confirmed with field measurements, such as train loads or observed frequencies. Platform waiting phenomena also seem to be exacerbated in congested stations, but very limited in most of the other stations. In our model, the average on-board density on trains is limited by a theoretical threshold, but highly congested lines exceed this mean value. Indeed, this could be explained by the fluctuation of the capacity threshold, whereby passengers force boarding and thus create capacity in the station while the available boarding capacity is close to zero.

Our model is not a precise model of a transit line but the demonstration of the usefulness of an integrated simulation approach. There are a number of ways in which our work could be improved, but here are just a few.

In the dynamic demand assignment model, users are represented by flows per time step and are assigned per train. The representation of the congestion points between users is spatially aggregated

over all the platforms and in all the compartments of each train. Dwell time calculation could therefore be improved by a fine-grained representation of the exchanges per train, as in this model, and per door, with a spatial distribution model that would take into account the effect of the critical door on dwell time (TCQSM, 2003). Similarly, the costs of congestion are averaged per train and not per compartment or even the space in each compartment. The model proposed in this article relies solely on the expected interactions between passenger flows and exchange times. The model needs more refined parameters. In particular, some of the lines on the Île-de-France urban rail network (SNCF Lines H and L) are now fully fitted with sensors that count the number of passengers boarding and alighting per door. These data, combined with the NAVIGO swipe data and the TJRF counts would help to calibrate our dwell time calculation model.

In the models of the movement of trains on the line, a number of assumptions made are open to discussion.

- Maximum dwell time is set at the same value for all the trains, independent of congestion. However, as seen in the simulation results, this parameter is central to the specific line. The link between congestion, passenger behavior (blocking the doors) and conductor behavior (leaving the doors open) needs further study. Further work could also compare consistency in maximal dwell time and its impact on the quality of service perceived by the users.
- The on-board density is set at a constant theoretical value of 4 passengers per meter. Yet this threshold is exceeded on highly congested rail lines. The observed phenomena are still valid, sensitivity to density is similar to that of the number of channels per door.
- The travel time of the trains remained fixed in our simulations. With additional data on the kinetics of the rolling stock, the travel times can be calculated more accurately. Moreover, a random term could be added and calibrated in future research in order to better simulate the real behavior of users and engineers and to test their importance to the performance of the line.
- The model in this research simulates the operation of a line with fixed blocks, so it cannot simulate certain service modification scenarios, such as the automation of a metro line, though this is being introduced on a number of such lines. The simulation of mobile blocks (CBTC, permanent communication between each train and the operations center) requires more detailed line discretization than fixed block simulation. Computation will then take much longer.
- The model is disaggregated by train in order to be able to simulate regulation strategies in response to incidents that deliberately hold back trains downstream from the disruption. These forms of strategies used at RATP in order not to leave periods of time with no service on the line could be tested with our model. The management strategies on the branches described in paragraph 3.5 would also be an important topic of study for the regulation of a line.

For an urban rail operator, this type of model can work alongside standard timetable and rolling stock optimization models in order to help anticipate the impact of train and passenger congestion phenomena. It can also be used to test operating strategy scenarios, such as limiting maximum dwell time or managing branches. In terms of planning, mobile blocks could be modelled in order to quantify the benefits to users of line automation. The establishment of an objective function based on quality of service for users of the line could form the foundation for a set of original operational initiatives.

Acknowledgements

The author would particularly like to thank Fabien Leurent, the instigator of this research, and Chaire Île de France Mobilité, which is funding it. My thanks also go to RATP for making these data on Line 13 available and to Florian Schanzerbächer for his advice and his experience of rail operations.

Bibliography

Abed, S.K. (2010). European Rail Traffic Management System – An Overview. In: 1st International Conference on Energy, Power and Control, 173-180

Babaei M., Schmöcker J.-D., Shariat-Mohaymany A. (2014), *The impact of irregular headways on seat availability,* Transportmetrica A: Transport Science, 10:6, 483-501, DOI: 10.1080/23249935.2013.795198

Berbineau M., *Les systèmes de transmission sol-train état de l'art and perspectives*, Recherche-transport-sécurité, janvier-mars 1999 P 35-46, Vol 62

Biagi,M. , Carnevali L., Paolieri M., Vicario E.(2017), *Performability evaluation of the ERTMS/ETCS – Level 3*, Vol 82, Septembre 2017,p314-336.

Cacchiani V., Huisman D., Kidd M., Kroon L., Toth P., Veelenturf L., Wagenaar J., (2014), An overview of recovery models and algorithms for real-time railway rescheduling, Transportation Research Part B, 63: p 15-37

Caimi G., Kroon L., Liebchen C., (2017) Models for railway timetable optimization: Applicability and applications in practice, Journal of Rail Transport Planning & Management, V6, p 285-312

Carnevali L. , Flammini F. , Paolieri M., Vicario E., (2015), *Non-markovian performability evaluation of ERTMS/ETCS level 3* European Workshop on Computer Performance Engineering, pp. 47-62, 10.1007/978-3-319-23267-6_4

Cats O., West J., Eliasson J., (2016) *A dynamic stochastic model for evaluating congestion and crowding effects in transit systems*, Transportation Research Part C V89, p43-57

Christoforou Z., Chandakas E., Kaparias I., (2016), An analysis of dwell time and reliability in urban light rail system, TRB 2016 Annual Meeting

Cominetti R., Correa J., (2001), Common-lines and passenger assignment in congested transit networks. Transportation Science 35, 250–267

Cordeau J-F, Toth P., Vigo D., (1998), A Survey of Optimization Models for Train Routing and Scheduling, Transportation Science 32(4): 380-404

Cury J., Gomide, F., & Mendes, M. J. (1980). A methodology for generation of optimal schedules for an underground railway systems. IEEE Transactions on Automatic Control, 25, 217–222

Daganzo C.F. (1997) Fundamentals of transportation and traffic operations, Elsevier Science Ltd., Oxford

de Cea, Fernandez, (1993), Transit assignment for congested public transport system: an equilibrium model. Transportation Science 27 (2), 133–147

Dullinger C., Struckl W., Kozek M., (2017), *Simulation-based multi-objective system optimization of train traction systems,* Simulation Modelling Practive and Theory, V72 p104-117, 2017

ERTMS. (2013). URL:http://www.ertms.net/ertms/ertmsbenefits.aspx. (Reached on: 02.04.2013).

Farhi N., Nguyen C., Haj-Salem H., Lebacque J-P, (2017), *Traffic modeling and real-time control for metro lines, the effect of passengers demand on the traffic phases.* In proceedings of american Control Conference

Fu Q., Liu R. Hess S., (2012), *A review on transit assignment modelling approaches to congested networks: a new perspective,* Social and Behavioral Science Procedia V54,1145-1155

Gentile G., Noekel K., (2016), Modelling public transport passenger flows in the era of intelligent transport systems, COST action TU1004, Springer,

Giro, Hastus software for Bus, Subway, Tram, http://www.giro.ca/en/solutions/bus-metro-tram , view in sept 2018

Hamdouch Y., Ho H.W., Sumalee A., Wang G., (2011), *Schedule-based transit assignment model with vehicle capacity and seat availability*, Transportation Research Part B V45, p1805-1830

Hamdouch Y., Szeto W.Y., Jaing, Y., (2014), *A new schedule-based transit assignment model with travel strategies and supply uncertainties*, Transportation Research Part B

Harris NG., and Anderson RJ., (2007), An international comparison of urban rail boarding and alighting rates. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, Vol. 221, pp. 521-526

Jiang Y., Szeto W.Y., (2016), *Reliability-based stochastic transit assignment: Formulations and capacity paradox*, Transportation Research Part B

Jiang Z., Hsu C., Zhang D. Zou X., (2016), *Evaluating rail travel timetable using big passengers' data*, Journal of Computer and system sciences, V82, p144-155

Kikuchi S, (1991), A simulation model of train travel on a rail transit line, Journal of advanced transportation, V25 N2 p211-224,

La vie du Rail, *CBTC, le système qui fait passer un train toutes les 90 secondes*, 9 nov 2012, RailPassion

Lam W., Cheung C, Poon Y, (1998), A Study of Train Dwelling Time at the Houn Kong Mass Transit Railway System, Journal of Advenced Transportation, Vol 32, No 3, p 285-296

Lam W.H.K., Gao, Z.Y., Chan, K.S., Yang, H., (1999), A stochastic user equilibrium assignment model for congested transit networks. Transportation Research Part B 33 (5), 351–368

Les Inrocks, https://www.lesinrocks.com/2018/04/25/actualite/pourquoi-la-ligne-13-du-metro-est-consideree-comme-la-ligne-de-lenfer-111075039/

Leurent F (2012) On Seat Capacity in Traffic Assignment to a Transit Network. Journal of Advanced Transportation, 46/2: 112-138

Leurent F. Chandakas E., Poulhes A., (2014), *A traffic assignment model for transit on a capacitated network: bi-layer framework, line sub-models and large-scale application.* Transportation Research Part C V47,3-27

Leurent F., Pivano C., Leurent F., (2017), *On passenger Traffic along a transit line: stochastic model of station waiting and in-vehicle crowding under distributed headways*, European Working Group Conference

Leurent F., (2011), *Transport capacity constraints on the mass transit system: a systemic analysis*, Eur Transp Res Rev V3, p11-21

Li F., Duan Z., Yang D., (2012), Dwell time estimation models for bus rapid transit stations, Journal of Modern Transportation, V20, N3, p168-177

Li S., Dessouky M., Ynag L., Gao Z., (2017), Joint optimal train regulation and passenger flow control strategy for high-frequency metro lines, Transportation Research Part B, 99: p113-137

Li W., Zhu W., (2016), A dynamic simulation model of passenger flow distribution on schedule-based rail transit networds with train delays, Journal of Traffic and Transportation Engineering 3: p364-373

Liebchen, C. (2008), The first optimized railway timetable in practice. Transportation Science, 42, 420–435.

Nash A. , Huerlimann D., (2004), *Railroad simulation using OpenTrack*, Computers in Railways IX, 2004 Witpress

Nguyen , S., Pallottino, S., Malucelli, F., (2001), A modeling framework for passenger assignment on a transport network with timetables. Transportation Science 35 (3), 238–249

Niu H., Zhou X., Gao R., (2015), Train scheduling for minimizing passenger waiting time with time-dependent demand skip-stop patterns: nonlinear integer programming models with linear constraints, Transportation Research Part B, 76, pp. 117-135

Opentrack, Railway technology, Home page of OpenTrack, http://www.opentrack.ch/opentrack/opentrack_e/opentrack_e.html, view in sept 2018

Poulhes A., Pivano C., Leurent F., (2017), Hybrid *Modeling of Passenger and Vehicle Traffic along a Transit Line: a sub-model ready for inclusion in a model of traffic assignment to a capacitated transit network*,Transportation Research Procedia V27, p164-171

Railsystem, http://www.railsystem.net/communications-based-train-control-cbtc/, 2015

Simwalk*, Combining Railway Network Simulation with Passenger Modeling*, http://www.simwalk.com/downloads/simwalk_whitepaper_transport_open.pdf

SimWalk, Opentrack, Combining Railway Network Simulation with Passenger Modeling, http://www.simwalk.com/downloads/simwalk_whitepaper_transport_open.pdf, view in sept 2018

Spiess H., Florian, M., (1989), Optimal strategies: a new assignment model for transit networks. Transportation Research Part B 23 (2), 83–102.

Su H, Cai J ., Huang S, (2012), Simuation System Enginnering for Train Operation Based on Cellular Automaton, Systems Engineering Procedia V3 13-21

Suazo-Vecino G., Dragicevic M., Munoz J-C., (2017), Holding boarding passengers to improve train operation based on an econometric dwell time model, TRB 2017 Annual Meeting

Szeto W.Y., Jiang Y., Wong K.I., Solayappan M., (2013), *Reliability-based stochastic transit assignment with capacity constraints: formulation and solution method*, Transportation Research Part C

Toledo T., Cats O., Buoghout W., Koutsopoulos H., (2010), *Mesoscopic simulation for transit operations*, Transportation Research Part C, V18, p896-908

Transportation Research Board. Transit Capacity and Quality of Service Manual. 2003

Verbas O., Mahmassani H., Hyland M. , (2016), *Gap-based transit assignment algorithm with vehicle capacity constraints: Simulation-based implementation and large-scale application.* Transportation Research Part B V93, p1-16

Ville Rail Transports, https://www.ville-rail-transports.com/lettre-confidentielle/ligne-voie-lautomatisation/

Wang J., Rakha H., (2018), *Longitudinal train dynamics model for a rail transit simulation system*, Transportation Research Part C, V89, p111-123, 2018

Wang Y., Laio Z., Tang T., Ning B., (2017) *Train scheduling and circulation planning in urban rail transit lines,* Control Engineering Practice

Yin J., Yang L., Tang T., Gao Z., Ran B., (2017), Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches, Transportation Research Part B, 97: p182-213

Zimmermann A., Hommel G., (2003), *A train control system case study in model-based real time system design*, International Parallel and Distributed Processing Symposium, IEEE, pp. 118-126, 10.1109/IPDPS.2003.1213234

Appendix 1: Variables

Model input data

$\mu_k(i)$	Next train target node currently being processed which is located at node i
$\beta_i(k_z)$	Train that passes through block $b_{i,\mu_k(i)}$ just before k_z.
$\lambda_k(i)$	Block that train k has just passed through
$\mu_k''(i)$	2nd next node that train k is going to pass through
ω_i	Minimum operator time between 2 consecutive trains at node i
$t_{i,\mu(i)}$	Minimum travel time between 2 consecutive stations served by a train. We will call this the theoretical travel time, i.e. the time provided by the operator associated with a run. Theoretical station dwell time will not be considered to be included in this time.
D_{ik}^{o}	Theoretical dwell time set by the operator per train k_z and per station $i \in N_S$. It does not include the time taken for the doors to open and shut.
$D_{i,\mu_k(i)}^{0,max}(k_z)$	Maximum exchange time
$\sigma_i(k_z)$	Signaling at node i at the time when rune k_z arrives at i. $\sigma_i(k_z) \in \{'g','o','r'\}$
$\bar{R}_i(H)$	Exogenous one-off delay at station i at time H

Main state variables

$y_{k,r}^{(i)}$	Passengers on train k who alight at station i and who boarded at a previous station $r < i$, y_k^- sum total
$\hat{y}_{is}^{(k)}$	Passenger flows boarding train k at station i for destination $s > i$. y_k^+ sum total
$\tilde{y}_{is}^{(k)}$	Passenger flows that actually board train k at station i taking into account all the delays.
$H_i^{-/+}(k)$	Arrival/departure time of train k at station i.
$\tilde{H}_{\mu(i)}^{-/+}(k)$	Expected forecast time of arrival/departure of train k at $\mu(i)$ considering no delays.
$H_i^{o-/+}(k)$	Predicted arrival/departure time of train k from i potentially degraded by the exchange time.
$E_{L,H}$, $\bar{E}_{L,H}$	Dynamic set of trains, the first containing trains with signals that allow them to advance, the second all the other trains.
$\bar{R}_i(H)$	Exogenous or endogenous delay at station i at time H