

Optimization of a passenger railway transportation plan considering mobility flows and service quality

Lucile Brethomé ^{a,c,1}, Cécile Bouvet-Agnelli ^a, Rémy Chevrier ^a,
Niels van Oort ^b, Joaquin Rodriguez ^c

^a SNCF Direction Innovation & Research

40, avenue des Terroirs de France, F-75611 Paris Cedex 12, France

¹ E-mail: lucile.brethome@sncf.fr, Phone: +33(0) 1 57 23 62 76

^b Faculty of Civil Engineering and Geosciences, Delft University of Technology
P.O. Box 5048, 11 2600 GA Delft, The Netherlands

^cIFSTTAR – ESTAS

20, Rue Élisée Reclus, F-59666 Villeneuve d'Ascq, Lille, France

Abstract

This research focuses on designing transportation plan for SNCF Transilien (French railway operator for the Parisian suburban mass transit). The objective is to develop methods and decision support tools to propose a timetable adapted to the passenger demand in the Parisian mass transit system, including comfort and reliability criterias.

This paper aims to present the first step of this research. We propose a graph theoretic ILP formulation for the Line Planning Problem, minimizing both travelers travel time and operating cost. We furthermore develop a multi-objective method to solve this problem. This method offers a pool of solutions in order to let the final designer choose the solution. We report computational results on real world instances provided from SNCF Transilien.

Keywords

Line Planning, Linear Programming, Multi-objective optimization.

1 Introduction

For historical reasons, the design of the SNCF railway services is, as elsewhere, mainly driven by incremental optimization of production resources (and primarily the available train-paths, human and material resources): timetable, infrastructure maintenance, rolling stock and crew. Moreover, each year, marginal modifications are executed on the previous timetable, but a totally new timetable is rarely designed.

To answer the need for mobility, it becomes important to have decision support tools on the design of an optimized transportation plan with regard to passenger flows and service quality. This design breaks up with the historical paradigm of optimization of the production resources. Proposing an offer consistent with real needs is crucial for SNCF. The traffic increased by 30% over the last decade, reaching 3 million passengers a day on the Parisian suburban area network. An offer more adapted to passenger demand will give improvements concerning average travel time, service reliability, connections and comfort.

Line planning is studied at the second stage of the transportation plan conception, which is composed of five consecutive steps: analysis of demand, line planning, timetabling,

rolling stock and crew management. We define a train by an origin, a destination, a route, timings and planned days. A line is a set of trains which differ only by their departure and arrival times. The Line Planning Problem (LPP) consists of defining for each line the origin, the destination, the route and the frequency in order to satisfy the estimated demand, namely to enable the transport of all the planned passengers from their origin to their destination. There are two competing objectives: on the one hand, to minimize the operating costs of lines, and, on the other hand, to minimize user discomfort. User discomfort is usually measured by the total passenger travel time or the number of transfers during the ride, or both. The planner has to balance these two objectives. Indeed, if we only minimize user discomfort, we would install a line for every passenger, which would be exceedingly expensive, but if we only minimize the operating costs of lines, it would inflict extremely long travel times for the passengers. One of the first model in literature can be found in Bussieck et al. (1996). The model maximizes the direct travelers with a MILP formulation. Barber et al. (2008) present a model which maximizes passenger coverage. In their paper, lines are constructed from scratch, instead of using a set of preprocessed lines. Schöbel and Scholl (2006) present a model which integrates line planning with minimal travel time and traffic assignment. The mathematical model is based on the change & go graph. Bussieck (1998) developed two different linearizations for cost-oriented model and a Branch&Cut algorithm for its solution. Goossens et al. (2006) investigates the multi-line planning problem in which not all trains need to stop at all stations. This model is cost-oriented. Borndörfer et al. (2008) present two multi-commodity path formulations where the passenger paths can be freely routed and the lines are generated dynamically. Their objective is to minimize the travel time, to which they add fixed and variable costs for the line system. The first formulation concerns the case in which all paths are allowed as lines. Distinct flow variables for the passengers and for the lines ensure that lines are constructed from scratch and the passengers are routed freely through the network. The second formulation uses a variable for each potential line and each potential passenger path. The models are solved by heuristics and compared theoretically. Schöbel (2012) proposes a paper surveying the literature published on the line planning problem. The Transit Route Network Design Problem (TRNDP) is a related area of research defined as a “process which implies clear and consistent techniques for designing a public transportation network”: the focus is on defining an approach rather than a model. The TRNDP deals also with the LPP problematics, but may comprise other planning stages like timetabling or rolling stock scheduling. Kepaptsoglou and Karlaftis (2009) propose a paper surveying this topic.

In the literature, the conception or adaptation only takes travel time into account, whereas passenger comfort and service reliability are neglected. Some papers introduce indicators to characterize the performance of railway services, focusing on the passengers’ point of view. As such we want to emphasize the way the travel is perceived by the user through indicators of travel time, comfort and reliability. Reliability is of great importance since studies (Bates et al. (1971)) have proven passengers tend to prefer having a slightly longer trip if it is more reliable. It can be measured through the adherence to schedule – or punctuality –, day-to-day variability, the latter corresponding to the time one would budget on its travel time to be late at most once a month. Van Oort et al. introduce a design approach taking service reliability (van Oort et al. (2015)) and crowding (van Oort et al. (2016)) into account, but do not facilitate optimization. Indeed, they propose a modelling framework to measure the quality of the transport offer, but its objective is not to optimize it.

The main goal of this paper is to offer a multiobjective approach for Line Planning,

where both operating costs and passenger comfort are taken into account. To this aim, we present an algorithm based on the ϵ -constraint method.

The paper is organized as follows. We formally state the problem in Sect. 2. We present the adopted multiobjective approach in Sect. 3, based on a Integer Linear Program (ILP) which minimizes both the the cost and the passenger's comfort. Experimental results are presented and analyzed in Sect. 4. Finally, some conclusions and perspectives are given in Sect. 5.

2 Problem description

We aim at developing a decision support tool that proposes passenger-oriented lines. This section describes the Transilien characteristics and some assumptions considered in our problem formulation.

Our goal is to define a set of lines, and their frequencies. As several types of rolling stock are used on the same train line in Transilien, we want to define which one is used for each line, depending on its cost and its capacity. We assume that the travel time is the same for all rolling stock.

The aim of this decision support tool is to quickly propose a panorama of different line concepts, depending on the cost. That is why we introduce a multi-objective method in our work. The two objectives will be on the one hand the cost, defined by the number of commercial kilometers, weighted by the cost of the rolling stock per kilometer, and on the other hand, the passenger's comfort, measured by the generalized travel time (travel time and transfer time).

This model has to cover current use of the network, namely the number of passengers who use the network today.

3 A multiobjective approach for Line Planning

3.1 A graph theory model

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph, based on the *Change & Go network* (Schöbel and Scholl (2006)), defined as follows.

The set \mathcal{V} of nodes represents either station-line-pairs or the origins and destinations of the passengers. The set \mathcal{E} of edges represents:

- Directed edges between nodes of the same stations (representing transfers, boarding or alighting);
- Edges between nodes of the same lines (Representing the train paths).

We define weights on all edges $e \in \mathcal{E}$ representing the real travel time for passengers using this edge e :

- For an edge e between a node representing an origin or a destination and a station-line-pair, we have the weight $c_e = 0$;
- For an edge e between nodes of the same line, the weight is equal to the travel time required between the two stations;

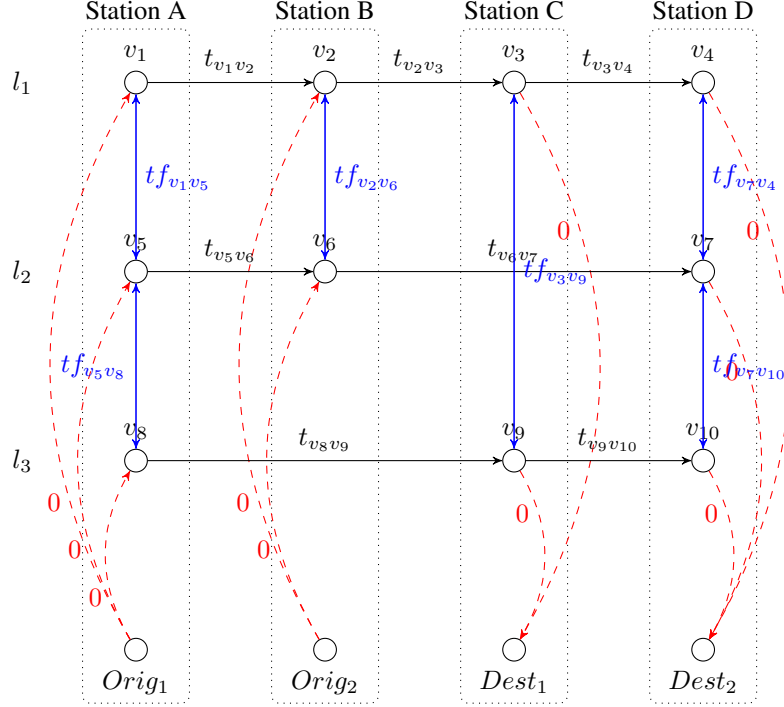


Figure 1: Change & Go Network: representation of 3 lines and 4 stations

- For an edge e between station-line-pairs of the same station, the weight is: $c_e = t_{f_e}$ where t_{f_e} is the required time to do the transfer between the two lines.

An example is illustrated in Figure 1. Three lines l_1 , l_2 and l_3 are running from station A to station D . Line l_1 is stopping at stations B and C , whereas line l_2 is stopping at station B and passing through station C , and line l_3 is passing through station B and stopping at station C . Black plain edges correspond to running pattern and run time between two consecutive stops of a train, e.g. $t_{v_1v_2}$ is the running time between the stop v_1 and the stop v_2 . Blue plain edges represents transfer time between two train stops of the same station, i.e. $t_{f_{v_1v_5}}$ is the minimal transfer time between v_1 and v_5 . Red dashed edges represent passengers boarding and alighting the train. $Orig_1$ and $Dest_1$ vertex correspond to the origin and the destination of the first group of passengers. $Orig_2$ and $Dest_2$ vertex correspond to the origin and the destination of the second group of passengers.

Given a set of lines L , determining the best lines for passengers amounts to calculate a minimum-cost multi-commodity flow where the capacity of edges are given by the maximum frequency of the line and the capacity of the rolling stock, and where the commodities are the origin-destinations of the passengers.

Notations

Let us consider the following sets:

- $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: the *Change & Go network*,

- \mathcal{S} : the set of stations of the network,
- \mathcal{L} : the set of considered lines,
- \mathcal{L}_i : the set of lines stopping at station $i \in \mathcal{S}$,
- \mathcal{L}_e : the set of lines crossing edge $e \in \mathcal{E}$,
- \mathcal{M} : the set of types of rolling stocks available,
- \mathcal{OD} : the set of Origin-Destination pairs covered by the demand.

The following parameters are defined:

- c_e : the cost of an edge $e \in \mathcal{E}$,
- n_m : the number of rolling stock units available for each type $m \in \mathcal{M}$,
- cap_m : the capacity of rolling stocks of each type $m \in \mathcal{M}$,
- k_l^m : the cost of a rolling stock unit $m \in \mathcal{M}$ used in line $l \in \mathcal{L}$,
- d_i^v : the demand for each Origin-Destination $i \in \mathcal{OD}$ at each vertex $v \in \mathcal{V}$,
- $freq_{min}^s$: the minimal frequency at a station $s \in \mathcal{S}$,
- $freq_{max}^s$: the maximal frequency at a station $s \in \mathcal{S}$.

The following integer decision variables are considered:

- T_l^m : the number of rolling stock units of type $m \in \mathcal{M}$ used during the planning horizon on line $l \in \mathcal{L}$,
- F_e^i : the number of passengers of the Origin-Destination $i \in \mathcal{OD}$ assigned to edge $e \in \mathcal{E}$.

Demand satisfaction and flow conservation constraints

F_e^i is the flow of passengers of the Origin-Destination $i \in \mathcal{OD}$ assigned to edge $e \in \mathcal{E}$. The number of passengers who want to ride this Origin-Destination is δ_i . The demand satisfaction constraints for this Origin-Destination are:

- for the Origin node $v \in \mathcal{V}$:

$$\sum_{e=(v,s) \in \mathcal{E}} F_e^i = \delta_i \quad (1)$$

- for the Destination node $v \in \mathcal{V}$:

$$\sum_{e=(s,v) \in \mathcal{E}} F_e^i = \delta_i \quad (2)$$

The flow conservation constraints for station-line-pairs nodes $v \in \mathcal{V}$ are:

$$\sum_{e=(v,s) \in \mathcal{E}} F_e^i - \sum_{e=(s,v) \in \mathcal{E}} F_e^i = 0 \quad (3)$$

We set the value of the demand for each Origin-Destination $i \in \mathcal{OD}$ at each vertex $v \in \mathcal{V}$ to:

$$d_i^v = \begin{cases} \delta_i & \text{if } v \text{ is the origin node} \\ -\delta_i & \text{if } v \text{ is the destination node,} \\ 0 & \text{if } v \text{ is a station-line-pair node.} \end{cases} \quad (4)$$

The demand satisfaction and flow conservation constraints can now be written as:

$$\sum_{e=(v,s) \in \mathcal{E}} F_e^i - \sum_{e=(s,v) \in \mathcal{E}} F_e^i = d_i^v, \forall i \in \mathcal{OD}, \forall v \in \mathcal{V} \quad (5)$$

An Integer Linear Programming Model

The mathematical model to solve the Multiobjective Line Planning Problem (MLPP) where both operating costs and user discomfort are taken into account is given below:

$$MLPP = (f_1, f_2) \quad (6)$$

$$f_1 = \min \sum_{i \in \mathcal{OD}} \sum_{e \in \mathcal{E}} c_e F_e^i \quad (7)$$

$$f_2 = \min \sum_{l \in \mathcal{L}} \sum_{m \in \mathcal{M}} k_l^m T_l^m \quad (8)$$

s.t.

$$\sum_{e=(v,s) \in \mathcal{E}} F_e^i - \sum_{e=(s,v) \in \mathcal{E}} F_e^i = d_i^v, \forall i \in \mathcal{OD}, \forall v \in \mathcal{V} \quad (9)$$

$$freq_{min}^s \leq \sum_{m \in \mathcal{M}} \sum_{l \in \mathcal{L}^s} T_l^m \leq freq_{max}^s, \forall s \in \mathcal{S} \quad (10)$$

$$\sum_{i \in \mathcal{OD}} F_e^i \leq \sum_{m \in \mathcal{M}} \sum_{l \in \mathcal{L}_e} cap_m T_l^m, \forall e \in \mathcal{E} \quad (11)$$

$$\sum_{l \in \mathcal{L}} T_l^m \leq n_m, \forall m \in \mathcal{M} \quad (12)$$

$$T_l^m \in \mathbb{N}, \forall l \in \mathcal{L}, \forall m \in \mathcal{M} \quad (13)$$

$$F_e^i \in \mathbb{N}, \forall e \in \mathcal{E}, \forall i \in \mathcal{OD} \quad (14)$$

The first objective function (7) minimizes the generalized travel time of Line Planning. The second objective function (8) aims to minimize the operating costs of the Line Planning.

Constraints (9) ensure that the demand is respected and that there is a flow conservation on each node of the network. Constraints (10) say that the number of rolling stock units which are crossing the station i have to respect the minimal and maximal frequencies of i . Constraints (11) ensure that the number of passengers on a line l cannot exceed the capacity of l . Constraints (12) limits the number of available trains for each type of rolling stock.

3.2 Methodology based on the ϵ -constraint method

As the two objective functions are two competing objectives, we choose to apply a multi-objective method to our model. There are three different types of multi-objective methods (Collette and Siarry (2003)):

- A priori methods, where preference information is first asked to the decision maker, and then a solution satisfying at best these preferences is found;
- A posteriori methods, where a set of Pareto optimal solutions is found, and then the decision maker chooses one of them;
- Interactive methods, where the decision maker can iteratively search for the solution satisfying at best his preferences.

As we want to provide a decision support tool where the decision maker does not know exactly his preferences, we choose to use an a posteriori method - the ϵ -constraint method - to solve our problem.

The ϵ -constraint method, introduced by Haimes et al. (1971), changes a multi-objective problem into a mono-objective problem. A single objective is chosen to be optimized and the others are transformed into constraints. The ϵ -constraint optimization problem is defined by

$$\min f_1 \quad (15)$$

$$s.t. f_i \leq \epsilon_i \forall i = 2, \dots, n. \quad (16)$$

where f_1, \dots, f_n are the n objective functions of the problem and ϵ_i is an upper bound for the objective i .

The Pareto Front is obtained by varying the value of ϵ_i for each optimization. This method is easy to use. It allows to find solutions on the Pareto Front, even in case of non-convex Front.

In our case, we choose to set the minimization of the generalized travel time as the objective to be optimized. The minimization of the operating costs is thus set as a constraint. We set the values of ϵ in the range between the optimal value of cost obtained by resolving the model only with the cost objective, and the value of cost obtained by only minimizing the travel time objective. Objective functions (7) and (8) become then:

$$\min \sum_{i \in \mathcal{OD}} \sum_{e \in \mathcal{E}} c_e F_e^i \quad (17)$$

s.t.

$$\sum_{l \in \mathcal{L}} \sum_{m \in \mathcal{M}} k_l^m T_l^m \leq \epsilon \quad (18)$$

Where (17) is the objective function of the problem, and constraint (18) ensures that the solution does not exceed the allocated budget.

4 Experimental results

4.1 Passenger traffic flows data

For this study, we need to predict – or estimate – Origin-Destination time-dependent passenger travel matrices. To this aim, we have data obtained by manual countings of travellers boarding and alighting at each train station, and Origin-Destination surveys. These data give us the number of passengers per Origin-Destination per hour (source Transilien). We assume that our data are reliable in order to focus on the modeling step.

4.2 Test instances

Several tests were performed on SNCF Transilien real instances. The chosen perimeter is composed of 60 stations, and there are 570,000 passengers a day. The instances concern 2 hours of the morning or evening peak hours, from the south to the north of the line. There are three types of rolling stock units, which can be used with one or two train-sets, and that have different passenger capacities. We created different pools of omnibus and semi-direct lines. Table 1 gives the characteristics of each instance.

Table 1: Instances description

Instance	$ L $	$ OD $
	(number of lines)	(number of Origin-Destination pairs)
A	5	441
B	5	399
C	9	441
D	7	399

4.3 Computational results

All computations were carried out on an Intel Core i7 processor running at 3.40 GHz, with a Windows 7 64 bits operating system. The algorithms were coded in Java and we use IBM ILOG CPLEX 12.4 as the MIP solver.

In this set of experiments, we decide that the minimum for ϵ is the solution of MLPP where we only consider f_2 , and the maximum for ϵ is set by the operating costs of the solution of MLPP where we only consider f_1 . So, we first solve the two MILP, one for each objective function. Let δ_ϵ be the minimal cost of a line. The parameter ϵ varies as follows: $\epsilon_{i+1} = \epsilon_i + \delta_\epsilon$.

Figure 2 shows the approximation of the Pareto front obtained by MLPP for the instance A. We can observe an inflection point, surrounded on the Figure 2, on the front. This observation is important because it shows that from a certain cost, the improvement of the offer in terms of travel time is not significant.

Table 2 summarizes the numerical results. For each instance, we provide, for each objective function, the value of the minimum obtained and the value of the maximum obtained.

We also provide the number of solutions obtained by our resolution.

Table 2 shows that we can provide a large set of solutions on the Pareto front in a single run.

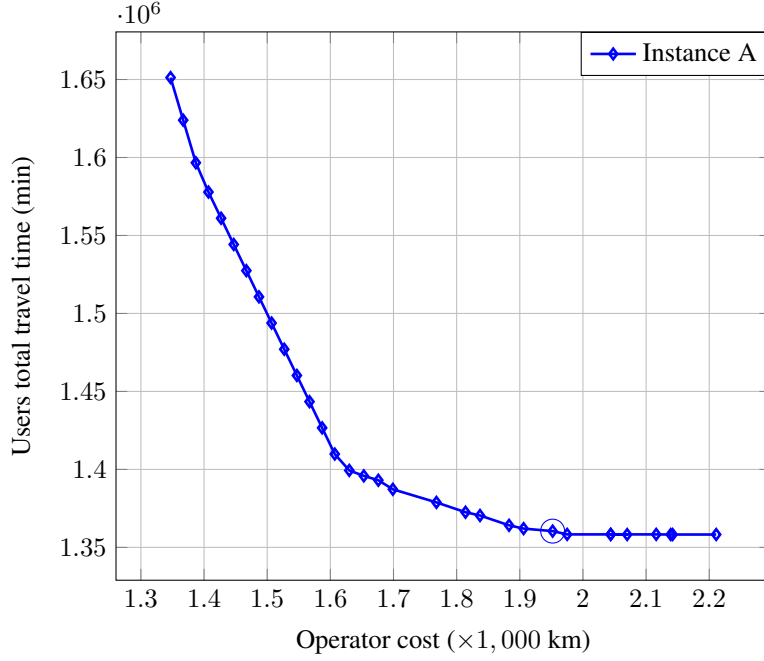


Figure 2: Approximation of the Pareto front

Table 2: Results of MLPP

Inst.	Operator costs (km)		User total travel time (min)		Nb sol.
	min	max	min	max	
A	1347	2211	1358238	1651227	32
B	1095	1265	1032845	1034258	7
C	1333	2179	1357398	1680282	25
D	980	1676	1032845	1064967	29

Then, we define the max-dominance point as the solution which dominates the greatest part of the objective space. Let max_p (resp. min_p) be the maximum (resp. minimum) obtained for the objective function f_1 minimizing the generalized travel time. Let max_c (resp. min_c) be the maximum (resp. minimum) obtained for the objective function f_2 minimizing the cost. Let Sol be the set of solutions of MLPP. Calculating the max-dominance point MDP means identifying:

$$MDP = \max_{x \in Sol} \{ \mathcal{A} = (max_p - f_1(x))(max_c - f_2(x)) \} \quad (19)$$

The max-dominance point characterizes the best compromise between the objectives for the decision-maker.

We introduce two indicators to qualify this point:

- The Normalized Cost Rate (NCR): let x be the value of the cost for the max-dominance point. The Normalized Cost Rate is defined as:

$$NCR = \frac{x - \min_c}{\max_c - \min_c} \quad (20)$$

- The Normalized Profit Rate (NPR): let y be the value of the profit for the max-dominance point. The Normalized Profit Rate is defined as:

$$NPR = \frac{\max_p - y}{\max_p - \min_p} \quad (21)$$

The calculated rates for all instances are reported in table 3. The main conclusion from these results is that at this max-dominance point, the minimum value of the generalized travel time is almost reached ($NPR > 95\%$), while the value of cost is quite far from the maximum ($NCR < 65\%$). These solutions are the most relevant for the decision-maker, because between the max-dominance point and the maximum obtained for the objective function f_2 minimizing the cost, a reduction in term of costs does not degrade the solution for travelers.

Table 3: Analysis of the max-dominance point

Inst.	Operator costs (km)		User total travel time (min)	
	max-dom. point	NCR	max-dom. point	NPR
A	1883	0.62	1364141	0.98
B	1121	0.15	1034270	0.99
C	1843	0.6	1358320	0.99
D	1059	0.11	1032850	0.99

5 Conclusion and Perspectives

In this paper, we have developed a multiobjective model for the LPP based on a graph. This model takes into account both operating costs and user discomfort, by minimizing the number of kilometers and the generalized travel time. Numerical experiments on test instances with real data show that the proposed multiobjective model for Line Planning Problem is capable of proposing a large set of non-dominated solutions in a single run. This model can be useful for the decision maker to assess the value of a solution, considering both user profit and operating costs.

We are currently working on a model to solve the Timetabling Problem. Moreover, as the Timetabling Problem and the Line Planning Problem are always tackled separately in the literature as distinct problems, our goal is to deal with both problems simultaneously in order to improve the quality of the solution of our transportation plan design. To this aim, we will give a particular attention to the connections between the Line Planning Phase and the Timetabling phase. For that, we can imagine an iterative loop which creates new constraints for the Line Planning phase by analyzing the results of the Timetabling Phase. Experimental results on realistic use-cases will be carried out and analyzed. We will compare the results obtained by our tool combining line planning and timetabling with existing timetables, in terms of quality of service (generalized travel time, congestion...). Another point of our research will be the improvement of the objective functions. Indeed, we want to include

the crowdedness of the trains, the number of seats or the number of different lines in our criteria, in order to improve the level of service of our offer.

Then, a careful attention will be given to the robustness of the timetable: the Paris area is a very dense network, so if the timetable is not robust, a minor incident can affect a lot of trains, in particular in peak hour. So, we will have to evaluate it, and to find a way to improve it. To this aim, some indicators will allow the calculation of the reliability of the timetables, in term of quality of service. We think that we could include the calculation of these indicators in our design process, in order to improve passengers' experience. To this end, we think of using a simulating tool, to change our model into an iterative simulation-optimization loop.

6 Acknowledgement

This work is partially financed by ANRT (Association Nationale de la Recherche et de la Technologie) through PhD grant number 2015/0637 with CIFRE funds.

References

- Barber, F., Ingolotti, L., Lova, A., Marin, A., Mesa, J., Ortega, F., Perea, F., Tormos, P., 2008. "Integrating timetabling, network and line design" *Technical Report*, ARRIVAL project.
- Bates, J., Polak, J., Jones, P., and Cook, A., 2001. "The valuation of reliability for personal travel", *Transportation Research Part E: Logistics and Transportation Review*, 37(2), 191-229.
- Borndörfer, R., Grötschel, M., Pfetsch, ME., 2008. "Models for line planning in public transport", In: Lecture note in economics and mathematical systems, *Computer-aided scheduling of public transport (CASPT)*, vol 600, pp 363-378, Springer Berlin Heidelberg.
- Bussieck, M.R., 1998. *Optimal lines in public transport*, PhD thesis, Technische Universität Braunschweig.
- Bussieck, M.R., Kreuzer, P., Zimmermann, U.T., 1996. "Optimal Lines for Railway Systems", *European Journal of Operational Research*, pp. 54-63.
- Collette, Y., Siarry, P., 2003. *Multiobjective Optimization: Principles and Case Studies*, Springer.
- Haimes, Y.Y., Ladon, L.S., Wismer, D.A., 1971. "On a Bicriterion Formulation of the Problems of Integrated System Identification and System Optimization", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol 1, No 3, pp. 296-297.
- Goossens, J., van Hoesel, CPM., Kroon, L.G., 2006. "On solving multi-type railway line planning problems", *European Journal of Operational Research*, 168(2):403-424.
- Kepaptsoglou, K., Karlaftis, M., 2009. "Transit route network design problem: Review", *Journal of Transportation Engineering*, 135(8):491-505.
- Van Oort, N., Brands, T., de Romph, E., Flores, J.A., 2015. "Unreliability effects in public transport modelling", *International Journal of Transportation*, Vol.3, No.1, pp 113-130.
- Van Oort, N., Brands, T., de Romph, E., Yap, M., 2016. *Ridership Evaluation and Prediction in Public Transport by Processing Smart Card Data: A Dutch Approach and Example, Chapter 11, Public Transport Planning with Smart Card Data*, eds. Kurauchi F., Schmöcker, J.D., CRC Press.

- Schöbel, A., 2012. "Line planning in public transportation: models and methods", *OR Spectrum*, pp 491-510.
- Schöbel, A., Scholl, S., 2006. "Line planning with minimal travel time", *5th workshop on algorithmic methods and models for optimization of railways*, number 06901, Dagstuhl Seminar proceedings.