# Modeling and solving line planning with integrated mode choice

Johann Hartleb[*1,2], Marie Schmidt[†1], Dennis Huisman[‡3,4], and Markus Friedrich[§2]

[1]Rotterdam School of Management and Erasmus Center for Optimization in Public Transport, Erasmus University Rotterdam, The Netherlands
[2]Institute for Road and Transport Science, University of Stuttgart, Germany
[3]Econometric Institute and Erasmus Center for Optimization in Public Transport, Erasmus University Rotterdam, The Netherlands
[4]Process quality and Innovation, Netherlands Railways, Utrecht, The Netherlands

**Abstract**

We present a mixed-integer linear program (MILP) for line planning with integrated mode and route choice. In contrast to existing approaches, the mode and route decisions are modeled according to the passengers' preferences while commercial solvers can be applied to solve the corresponding MILP. The model aims at finding line plans that maximize the profit for the public transport operator while estimating the corresponding passenger demand with choice models. Both components of profit, revenue and cost, are influenced by the line plan. Hence, the resulting line plans are not only profitable for operators but also attractive to passengers. By suitable preprocessing of the passengers' utilities, we are able to apply any choice model for mode choices using linear constraints. We provide and test means to improve the computational performance. In experiments on the Intercity network of the Randstad, a metropolitan area in the Netherlands, we show the benefits of our model compared to a standard line planning model with fixed passenger demand. Furthermore, we demonstrate with the help of our model the possibilities and limitations for operators when reacting to changes in demand in an optimal way. The results suggest that operators should regularly update their line plan in response to changes in travel demand and estimate the passenger demand during optimization.

[*]hartleb@rsm.nl
[†]schmidt2@rsm.nl
[‡]huisman@ese.nl
[§]markus.friedrich@isv.uni-stuttgart.de

# 1    Introduction

In previous years, public transport operators around the globe recorded a continuous increase in passenger numbers. There is a variety of reasons that might explain changes in passenger numbers, such as a shift in passenger interests and behavior, the development of regions, or policy measures. Since the beginning of the Covid-19 pandemic in early 2020, however, the number of passengers has fallen dramatically to a fraction of its original size. The pandemic is likely to have long-term effects on travel behavior, as, for example, working from home is more accepted by many companies. In the case of the Netherlands Railways (NS), the expectation is that passenger numbers will not reach their pre-pandemic level before 2024[1], but increase slowly over time.

Before and after the pandemic, railway operators like NS were and will be constantly faced with changing travel demand. To adapt their service to small fluctuations, operators can make adjustments on the level of tactical planning. For example, small adaptions in the timetables and rolling stock schedules can be implemented relatively spontaneously and with comparatively little effort. However, such adjustments are not suitable to cope with greater and longer-term changes in demand. Instead, this issue needs to be approached from a strategic planning perspective and the line plan needs to be adjusted from time to time.

When designing a public transport line plan, it is important to distinguish between passenger and traveler demand. With *traveler demand* we refer to the total number of people who want to travel. Travelers may choose to use any available mode of transport, such as train or car. *Passenger demand* includes only those travelers who choose to use the public transport system. The decision of travelers for their mode of transport, and thus the number of passengers, depends to a certain extent on the quality of the service offered. This poses an interesting yet complex situation for operators: public transport services have to be designed to provide sufficient capacity for passenger demand, which in turn depends on the service.

In this paper, we consider the problem of finding a line plan and simultaneously estimating the corresponding passenger demand based on a prognosis for traveler demand. The aim is to estimate both the share of travelers deciding to use public transport (mode choice) and the passenger distribution in the network (route choice). The travelers' choices depend on the quality of the service offered: they value a service with fast, direct, and frequent connections. A line plan offering such connections between two stations will attract more passengers between those stations, while passengers between stations with slow and infrequent connections will be inclined to turn to other modes of transportation. Considering travelers' decisions allows an accurate estimation of passenger demand during optimization, and the resulting line plans are aligned with the demand they generate.

Although many approaches state that passenger demand and line plans are interdependent, we identify two reasons why demand estimation is mostly not modeled accurately. First, travelers' mode choice is in most cases neglected. Second, if a passenger distribution on routes is considered, usually one of the following two simplifications is applied: either all passengers traveling between two stations are required to use the same route, or the model can assign passengers to routes in favor of a system optimum, rather than considering passenger preferences. An imprecise demand estimate is obstructive to the search for

---

[1] https://nos.nl/artikel/2344006-ns-lijdt-185-miljoen-verlies-door-corona.html, accessed February 19

efficient line plans and carries the risk of insufficient seating capacity. Only a few publications deal with the integration of passenger choice models in line planning. However, these approaches are usually not computationally tractable and the quality of solutions found with heuristic approaches is hard to assess. In Section 2 we discuss the related line planning literature in detail.

In our paper, we present a mixed-integer linear programming (MILP) formulation for finding a line plan and corresponding passenger loads from given traveler demand. The passenger loads depend on the line plan and are estimated with passenger choice models. By suitable preprocessing of the utilities for the passengers' mode and route decisions, the choice models can be linearized and commercial solvers can be used to find solutions. Approaching the problem from the operator's perspective, our objective is profit maximization, where profit is equal to the revenue from serving passenger demand minus the operating costs. This yields line plans that attract many passengers, while being efficient concerning operational costs.

We test and analyze the model in experiments on the Intercity network of the Randstad region that is operated by NS, the largest Dutch railway operator. Additional constraints and branching strategies are tested to improve the computational performance. The model is compared to a basic line planning model with predetermined passenger loads which highlights the advantages of demand estimation during optimization. Furthermore, the integration of passenger decision models enables to conduct a sensitivity analysis of the service level and operator profit on fluctuations in traveler demand. This gives valuable insights for operators into their business models and the optimal solutions provide concepts for how the line plans should be adjusted in response to demand changes.

Summarizing, our main contributions are fourfold. First, we present a novel line planning model that considers both route and mode choice from a customer's perspective. In contrast to existing optimization approaches, passengers are not assigned to routes according to a system optimum but distribute on the best routes in our model. Second, we develop a linear formulation for this model which allows the usage of commercial solvers and we provide means to improve the computational performance. Third, we show in experiments that operators should include an estimation of mode and route choice *during* optimization to achieve the best possible profit and passenger shares. Fourth, we show the impact of drastic changes in travel demand on the modal split and the financial performance of the public transport operator.

The remainder of this paper is structured as follows. In Section 2, related line planning approaches are summarized. The modeling of line planning and passenger demand estimation is described in Section 3. This section discusses the used choice models in detail and the assumptions made in order to linearize them. Section 4 gives information about the experimental setup, used data sets, and parameters. The experiments are described and discussed in Section 5, followed by practical insights for operators in Section 6. The paper concludes in Section 7 with a summary of the findings.

## 2 Related literature

The goal of the line planning problem is to find a set of lines with corresponding frequencies such that conditions on operating costs and passenger service level are satisfied. In this context, lines are defined as a sequence of stations that are served by a vehicle. Schöbel (2012) summarizes different modeling

3

approaches and solution methods for the line planning problem in public transport and identifies several variations: In some formulations, the task is to select lines from a given pool of lines (Gattermann et al., 2017), while in others the line routes are constructed during optimization (Borndörfer et al., 2007). There also exist different objectives for line planning. On the one hand, cost-oriented objectives aim at minimizing operational costs while ensuring a certain passenger service level (Claessens et al., 1998; Friedrich et al., 2017; Goossens et al., 2006). On the other hand, passenger-oriented objectives mostly consider a budget constraint on operational costs to maximize passenger service level, represented by the share of passengers with direct connections (Bussieck, 1998; Bussieck et al., 1997; Schöbel and Scholl, 2006) or by passenger journey times (Goerigk and Schmidt, 2017; Schöbel and Scholl, 2006).

The impact of the service on passengers is often neglected. Most of the existing approaches have the assumption in common that (an estimate of) the passenger demand is known before the line plan is found. This means the number of passengers between each station pair is assumed to be fixed. In addition, passengers are in many cases assigned a priori to paths in the network to estimate the required capacity between stations. However, both the number of passengers and the passenger paths depend on the line plan and the corresponding passenger service level (de Dios Ortuzar and Willumsen, 2011).

Most existing approaches that have included passenger route choice either applied a single (shortest) route search (Guan et al., 2006; Liu et al., 2019; Nachtigall and Jerosch, 2008) or a distribution according to a system optimum (Borndörfer et al., 2007; Borndörfer and Karbstein, 2012). Both strategies are unlikely to accurately estimate a passenger distribution, bearing the risk for operators of crowded or underutilized vehicles. In a cross-entropy heuristic for integrated line planning and timetabling presented by Kaspi and Raviv (2013), passengers are distributed on shortest paths for evaluation, which serves as the basis to refine the search for an updated solution in the next iteration. Schmidt and Schöbel (2015) and Friedrich et al. (2017) present generic line planning models with integrated passenger route choice and discuss complexity and bounds. A passenger-optimal route search was introduced in Schmidt (2014) and Goerigk and Schmidt (2017) where sufficient seating capacity is ensured assuming that passengers distribute over the available shortest routes. This approach overcomes the problem that passengers are assigned to sub-optimal routes by the model and prevents capacity conflicts during operation. Schiewe et al. (2019) propose a game-theoretical approach where passengers are individual players choosing their routes with the highest travel quality.

All of the approaches discussed above consider a flexible passenger to route assignment or search but assume the total passenger demand to be fixed. Only a few publications consider the mode choice of travelers during line planning to estimate the number of passengers attracted by the solution. The integrated stop location and line planning approaches discussed in Laporte et al. (2005) and Laporte et al. (2007) aim at a maximum trip coverage. Similarly, Klier and Haase (2015) maximize the number of expected passengers and estimate the mode choice with the logit model as a traveler's decision between the best available route and an alternative mode of transport. Bertsimas et al. (2021) have the same objective to maximize ridership of the public transport system. In their model, each additional line increases the modal split by a predefined percentage, independent of which other lines are selected and their frequencies. De-Los-Santos et al. (2017) use the logit model for the mode choice as well and approximate it with a piecewise linear function. For the specific case of Intercity buses, Steiner and Irnich (2018) consider

different passenger demand levels depending on departure and travel time in a combined optimization approach for stop selection on a line and timetabling. A comprehensive model integrating network design, line planning, traveler mode choice, passenger route choice, and fleet investment is discussed in Canca et al. (2016). Later, the authors provide in Canca et al. (2017) an adaptive large neighborhood search metaheuristic for this model, limiting the passenger route search to a shortest route. For a revised model including an integrated passenger distribution on routes, Canca et al. (2019) present a two-level local search matheuristic which was successfully used to find solutions for real-world sized instances.

These models include a modeling of travelers' mode choice during line planning, but they contain at least one of the two limitations. Either, the passenger distribution on routes, although relevant for seat capacity estimation, does not reflect passenger preferences. Or, the quality of the solutions found by the applied metaheuristics is in many cases hard to assess.

# 3 Modeling

In this paper, we develop a mixed-integer linear programming model for finding a line plan that is tailored to the corresponding passenger demand. Lines are selected from a pool of potential lines. The passenger demand is estimated with a discrete choice model and the distribution on routes is modeled according to passengers' choices. This modeling of passengers provides a basis for an accurate estimation of the required seat capacity and the expected revenues from ticket sales. The objective is to find a profit-optimal line plan, which means, the difference between revenue and cost is maximized. In the following sections, we discuss all components of the line planning problem, the underlying assumptions made, and how we model it. Throughout this paper, we use the terminology of public rail transport, however, the proposed model applies to public transport services of any kind.

## 3.1 Network structure and line selection

We consider a public transport network with a set of *stations* $\mathcal{S}$ as nodes, a set of *tracks* $\mathcal{T}$ as direct connections between the stations, and symmetric traveler demand between the stations. Let $k$ denote an *origin-destination (OD) pair*, that is, an unordered pair of stations $s_1$ and $s_2 \in \mathcal{S}$, and let $\mathcal{OD}$ be the *set of OD pairs*. The *traveler demand*, that is, the number of persons wanting to travel between OD pair $k \in \mathcal{OD}$, is denoted by $\delta_k$. We assume a *line pool* $\mathcal{P}$ to be given, where each *line* $l \in \mathcal{P}$ is an undirected sequence of stations $(s_1, s_2, \ldots, s_{n_l})$.

The aim is to find a *line plan* $\mathcal{L} \subseteq \mathcal{P}$, a subset of lines from the line pool, for a regular, symmetric service. When selecting a line from the pool $\mathcal{P}$, it is assumed to operate in both directions.

The selection of lines is modeled using binary decision variables

$$z_l \in \{0,1\} \text{ with } z_l = \begin{cases} 0, & \text{Line } l \text{ is not selected,} \\ 1, & \text{Line } l \text{ is selected.} \end{cases}$$

This constitutes the basis of the line planning problem: Lines have to be selected from a line pool to

5

meet passenger demand. Higher line frequencies can be achieved by selecting multiple lines with the same itinerary. This indirect modeling of frequencies is accepted to unambiguously link lines and passenger routes in Section 3.3.1. In the following, the objective of the line planning model is defined, the procedure to estimate the passenger demand is explained, and, based on that, the required capacity is calculated.

## 3.2   Profit maximization

Our objective is to find a line plan that is optimal with respect to profit, which means, the difference of revenue generated by passenger fares and costs for installing and operating lines should be as high as possible. The *costs* $c_l$ for installation and operation of line $l$ comprise acquisition and maintenance of rolling stock as well as personnel and energy expenses. We assume an OD pair-dependent passenger pricing as it is applied in the Netherlands. There, passengers are charged based on the locations of their origin and destination only, and ticket prices are independent of the chosen route. Let $p_k$ denote the *ticket price* for a passenger of OD pair $k$ and let $w_k$ denote the *share of travelers* of OD pair $k$ that choose to use public transport. Then, the following function describes the profit made by a line plan $\mathcal{L}$.

$$\sum_{k \in \mathcal{OD}} p_k \cdot \delta_k \cdot w_k - \sum_{l \in \mathcal{P}} c_l \cdot z_l \tag{1}$$

The total cost calculation is defined as the sum of costs for all selected lines. The product of the total number of travelers $\delta_k$, the share of travelers using public transport $w_k$ and the ticket price $p_k$ gives the total revenue from passengers of OD pair $k$. Note that next to the cost, also the revenue is affected by the selected lines. With a better supply, more passengers decide to use public transport and generate more revenue. Therefore, the share of travelers deciding to use public transport $w_k$ is a variable in this context. How the value of $w_k \in [0, 1]$ is estimated based on the selected lines is described in the next section.

## 3.3   Demand estimation

Only travelers that decide to use public transport because of the quality of the line plan generate revenue for the operator. Travelers that choose to travel with an alternative mode, such as private car, do not contribute to the objective function (1). Mode choices of travelers can be estimated with discrete choice models (de Dios Ortuzar and Willumsen, 2011). We assume a traveler demand instead of a passenger demand to be given and include a mode choice based on the quality of the line plan.

Furthermore, passengers that choose to travel with public transport distribute over available routes. A *route* is a sequence of consecutive line segments, where each line segment corresponds to a track in the network operated by a certain line. This concept of a route corresponds to a path in the *change&go network* introduced by Schöbel and Scholl (2006). Estimating which routes passengers use is important to determine the required seating capacity correctly.

### 3.3.1 Route choice

Passengers distribute over multiple available routes if they are of reasonably good quality. A route is *available* if all lines that are used on the route are selected. Since the availability of routes is only known after the line plan is known, we consider a *choice set* $\mathcal{C}_k$ of routes for each OD pair $k$ to be given. The availability of a route $r \in \mathcal{C}_k$ is modeled with the binary auxiliary decision variable

$$y_r \in \{0,1\} \text{ with } y_r = \begin{cases} 0, & \text{Route } r \text{ is not available,} \\ 1, & \text{Route } r \text{ is available.} \end{cases}$$

The selection of lines and availability of routes can be linked with the following set of constraints

$$y_r \leq z_l \qquad\qquad \forall k \in \mathcal{OD}, \ r \in \mathcal{C}_k, \ l \in \mathcal{P}_r \qquad (2a)$$

$$y_r \geq \sum_{l \in \mathcal{P}_r} z_l - |\mathcal{P}_r| + 1 \qquad \forall k \in \mathcal{OD}, \ r \in \mathcal{C}_k \qquad (2b)$$

where $\mathcal{P}_r$ is the set of lines in the pool used on route $r$.

In practice, we observe that passengers predominantly distribute over the best available routes. Hence, we restrict the choice sets $\mathcal{C}_k$ to contain only the best routes. The quality of a route is determined by the driving time as well as the number of transfers. Let $j_r$ be the *journey time* of route $r$, that is, the approximate driving time including dwell times plus a transfer penalty for each transfer. Then, we consider only those passenger routes for OD pair $k$ that are among the journey-time shortest routes. By restricting the choice sets to the shortest routes, all alternative routes for one OD pair are of very similar quality to the passengers. As a consequence, we assume that passengers of one OD pair distribute uniformly on available routes.

Note that this assumption of a uniform distribution on routes of similar journey time does not perfectly represent passenger behavior. In fact, the distribution of passengers on routes also depends on the temporal spread of departure and arrival times of the available routes. Since the timetable is not known yet at the stage of line planning, it is only possible to estimate the passenger distribution based on the quality of the available routes. For the information available, the uniform distribution on the journey-time shortest routes is a good approximation. In any case, we believe that this modeling comes closer to an actual distribution than the common assumptions that all passengers of an OD pair use a single (shortest) route or can be assigned to routes according to a system optimum. Hence, this distribution is expected to give a better estimate of required seating capacity than existing approaches.

To assume that passengers distribute uniformly on routes is an essential component to achieve a tractable model. It is therefore important to consider comparable routes for each OD pair. In this paper, we use the journey time to compare routes, but in principle, any definition of quality that is independent of the timetable can be used.

Let the variable $w_r$ denote the *share of travelers using route $r$*. Then,

$$w_r \in [0, 1] \text{ with } w_r = \begin{cases} 0, & y_r = 0, \\ \frac{w_k}{\sum_{r' \in \mathcal{C}_k} y_{r'}}, & y_r = 1 \end{cases} \quad \forall k \in \mathcal{OD}, \ r \in \mathcal{C}_k \tag{3}$$

The values of $w_r$ are the same for all available routes for OD pair $k$ and they sum up to the share of travelers using public transport $w_k$. To linearize the computation of $w_r$ in Equation (3), the binary auxiliary variable

$$b_k^i \in \{0, 1\} \text{ with } b_k^i = \begin{cases} 0, & \text{OD pair } k \text{ has less than } i \text{ available routes}, \\ 1, & \text{OD pair } k \text{ has at least } i \text{ available routes} \end{cases}$$

to count the number of available routes of OD pairs is introduced. Using variable $b_k^i$, the route choice constraints (3) can be linearized with the following set of constraints

$$w_r \leq y_r \qquad \forall k \in \mathcal{OD}, \ r \in \mathcal{C}_k \tag{4a}$$

$$w_r \leq \frac{w_k}{i} - b_k^i + 1 \qquad \forall k \in \mathcal{OD}, \ r \in \mathcal{C}_k, \ i \in \{1, \ldots, |\mathcal{C}_k|\} \tag{4b}$$

$$\sum_{r \in \mathcal{C}_k} w_r = w_k \qquad \forall k \in \mathcal{OD} \tag{4c}$$

Constraints (4a) ensure that $w_r$ is positive only if route $r$ is available, that is, if $y_r = 1$. Constraints (4b) impose an upper bound of $\frac{w_k}{i}$ to $w_r$ if at least $i$ routes are available for OD pair $k$, that is, if $b_k^i = 1$. For increasing $i$, $\frac{w_k}{i}$ decreases and these constraints get tighter. If less than $i$ routes are available and $b_k^i = 0$, the right-hand side is greater than 1 and the constraints are redundant. Constraints (4c) ensure that the shares $w_r$ of passengers of OD pair $k$ using route $r$ sum up to the share of travelers using public transport $w_k$. Together, constraints (4) model that the passengers of OD pair $k$ distribute uniformly on all available routes from the set $\mathcal{C}_k$. Note that both the number of passengers and the number of available routes depend on the line plan.

The number of available routes can be counted within the model with the following constraints

$$\sum_{r \in \mathcal{C}_k} y_r \geq i \cdot b_k^i \qquad \forall k \in \mathcal{OD}, \ i \in \{1, \ldots, |\mathcal{C}_k|\} \tag{5a}$$

$$b_k^i \geq \frac{1}{|\mathcal{C}_k|} \left( \sum_{r \in \mathcal{C}_k} y_r - i + 1 \right) \qquad \forall k \in \mathcal{OD}, \ i \in \{1, \ldots, |\mathcal{C}_k|\} \tag{5b}$$

Constraints (5a) force $b_k^i$ to 0 if less than $i$ routes are available for OD pair $k$, and constraints (5b) ensure that $b_k^i$ equals to 1 if at least $i$ routes are available.

### 3.3.2 Mode choice

It remains to estimate the modal share $w_k$, that means, the share of travelers deciding to use public transport. The travelers' mode choice depends on both the utility of public transport and the utility of alternatives. We consider one alternative mode representing individual road transport, such as driving by private car. In general, changes in the public transport system affect the modal split, with this the congestion on roads, and eventually the utility of the alternative mode. However, as long as the utility of the public transport system is not changed substantially, the impact on the utility of the alternative mode is negligible. Hence, we assume that the utility of the alternative mode is independent of the utility of public transport in this research. As a consequence, the utility of the alternative mode is constant and the travelers' decision whether to take public transport or not solely depends on the utility of public transport.

Since a line plan is designed, not all information about the public transport system determining its utility is available. For example, departure and arrival times, as well as transfer durations are only known after a timetable is found. Nevertheless, the line plan determines the most important factors of influence for the travelers' decision about their mode of transport: the number of available routes, the approximate driving time, and the number of transfers.

While the number of available routes depends on the line plan, the approximate driving time and the number of transfers can be predetermined for each route $r \in \mathcal{C}_k$. Both are combined in the route journey time $j_r$ that is known for each route. Since all routes in the choice set $\mathcal{C}_k$ have a very similar journey time, the journey time $j_k$ for OD pair $k$ is very close to the route journey time $j_r$ for any route $r \in \mathcal{C}_k$ and does not depend on which routes are available. Hence, it can be predetermined as well and we define

$$ j_k = \frac{1}{|\mathcal{C}_k|} \sum_{r \in \mathcal{C}_k} j_r \qquad \forall k \in \mathcal{OD}. $$

The journey time $j_k$ captures the duration needed for OD pair $k$ when traveling with public transport. It is a travel time equivalent comprising drive and dwell times, and possibly transfer penalties. The number of available routes indicates the frequency of the service for the passengers. The more routes are available within a time period, the fewer passengers have to adapt to the schedule of the public transport. This is quantified in the *adaption time* $a_k$, the time passengers need to deviate from their preferred departure or arrival times. With a more frequent service, the expected adaption time decreases inversely proportional. We define the utility of the public transport for OD pair $k$ as the sum of the journey time and the adaption time.

$$ u_k = j_k + a_k \qquad \forall k \in \mathcal{OD} \tag{6} $$

The utility $u_k$ does not depend on *which* routes are available, but only on the *number* of available routes. Thus, the number of available routes is the main factor of influence on travelers' mode choice that cannot be predetermined. However, it is possible to express the utility of public transport, and consequently the mode choice of travelers for each OD pair as a function of the number of available routes.

Figure 1 shows the expected share of travelers using public transport for different numbers of routes, exemplified for one OD pair. In this paper, we use a logit model to estimate the travelers' mode choice
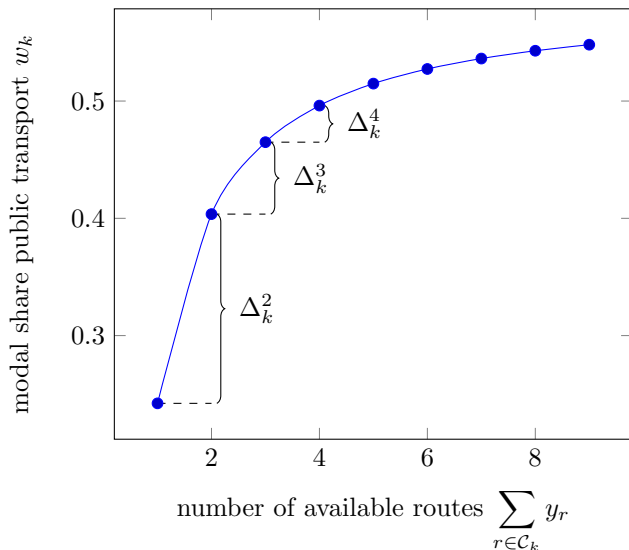
Figure 1: Estimated modal split for one OD pair $k$ as a function of the number of available routes

and to derive the modal split. The logit model is a discrete choice model that is commonly used to estimate travelers' choices. The probability that an alternative is chosen depends on the utilities of all available alternatives. For the mode choice of travelers, just two alternatives are considered: traveling by public transport and traveling by individual transport such as by private car. Let $u_k$ be the utility of public transport for OD pair $k$ from Equation (6) and $\hat{u}_k$ be the utility of the alternative mode. Then, the logit model estimates the modal share for public transport as

$$w_k = \frac{e^{\beta u_k}}{e^{\beta u_k} + e^{\beta \hat{u}_k}},$$

where $\beta$ is the logit coefficient to tune the model.

The *increments in modal share* $\Delta_k^i$ indicated in Figure 1 express the additional share of travelers of OD pair $k$ deciding to use public transport if $i$ instead of $(i-1)$ routes are available. Using these values, the mode choice of travelers can be modeled linearly with the constraints

$$w_k = \sum_{i=1}^{|\mathcal{C}_k|} \Delta_k^i \cdot b_k^i \qquad \forall k \in \mathcal{OD} \tag{7}$$

As discussed for the passenger route choice in Section 3.3.1, the binary variables $b_k^i$ equal 1 if OD pair $k$ has at least $i$ routes. Hence, constraints (7) set the modal share of public transport $w_k$ for OD pair $k$ dependent on the number of routes.

Note that the coefficients $\Delta_k^i$ can be predetermined for each OD pair $k$ and each possible number of available routes $1 \leq i \leq |\mathcal{C}_k|$. Hence, the $\Delta_k^i$ are constant parameters in the model formulation. This framework allows the integration of travelers' mode choice according to any choice model using the linear

constraints (7). Our model is thus not limited to the logit model which is used in this paper.

## 3.4 Operational requirements

### 3.4.1 Capacity constraints

The strength of the demand modeling in Section 3.3 is that the number of passengers on each route can be estimated accurately. It should be ensured by the operator that there is sufficient seating capacity on each available route $r$ for the number of passengers that is expected to choose route $r$. We model this as one capacity constraint per *line segment*, the part of a line $l$ traversing a track $t$. Let $\kappa_l$ be the *seating capacity* of line $l$, $\mathcal{P}(t)$ be the set of lines in the line pool that operate on track $t$ and $\mathcal{C}_k(l, t)$ be the choice set of routes for OD pair $k$ using line $l$ on track $t$. The following constraints ensure that on each line segment sufficient capacity is provided by the line plan for all passengers on their chosen routes.

$$\sum_{k \in \mathcal{OD}} \sum_{r \in \mathcal{C}_k(l,t)} \delta_k \cdot w_r \leq \kappa_l \cdot z_l \qquad \forall t \in \mathcal{T}, \ l \in \mathcal{P}(t) \tag{8}$$

Note that the presented model uses individual capacity constraints for each line segment and considers a distribution on routes according to passenger preferences. The combination of individual capacity constraints and passenger-optimal routes is important for accurate capacity estimation. It achieves that passengers use the same routes in the model as they would choose in real life, thus avoiding potential conflicts with capacity constraints. Existing line planning models often use only one capacity constraint per track $t$, aggregated over all lines operating on that track, or they assign passengers on a single route or according to a system optimum. Both can cause capacity conflicts unless passengers accept additional transfers to make space for other passengers, which is unrealistic.

### 3.4.2 Minimal service requirement

In addition to their aim to meet the passenger demand, most operators are required to offer a minimal service in certain parts of the networks. This ensures that all passengers have access to public transport, also in sparsely-populated areas. We model this as an additional set of constraints ensuring that at least $f_t$ vehicles service track $t$.

$$\sum_{l \in \mathcal{P}(t)} z_l \geq f_t \quad \forall t \in \mathcal{T} \tag{9}$$

## 3.5 Line planning model with integrated mode choice

In this section, we give the mixed-integer linear program for line planning with integrated mode choice $(LPwMC)$. As described in Section 3.2, the objective is to maximize profit, defined as revenue minus cost in Equation (1). The first part deals with the demand estimation explained in Section 3.3 including auxiliary modeling constraints. The linking constraints between route and line variables are discussed in Equations (2) and the uniform passenger distribution on available routes is modeled with Equations (4). The number of available routes is counted within the model with Equations (5) and the traveler mode

(**LPwMC**)

max profit = revenue - cost

$$\max \quad \sum_{k \in \mathcal{OD}} p_k \cdot \delta_k \cdot w_k - \sum_{l \in \mathcal{P}} c_l \cdot z_l$$

link between line and route variables

$$
\begin{aligned}
\text{s.t.} \quad y_r &\leq z_l & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k,\ l \in \mathcal{P}_r \\
y_r &\geq \sum_{l \in \mathcal{P}_r} z_l - |\mathcal{P}_r| + 1 & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k
\end{aligned}
$$

passenger route choice

$$
\begin{aligned}
w_r &\leq y_r & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \\
w_r &\leq \frac{w_k}{i} - b_k^i + 1 & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k,\ i \in \{1, \ldots, |\mathcal{C}_k|\} \\
\sum_{r \in \mathcal{C}_k} w_r &= w_k & \forall k \in \mathcal{OD}
\end{aligned}
$$

counting of available routes

$$
\begin{aligned}
\sum_{r \in \mathcal{C}_k} y_r &\geq i \cdot b_k^i & \forall k \in \mathcal{OD},\ i \in \{1, \ldots, |\mathcal{C}_k|\} \\
b_k^i &\geq \frac{1}{|\mathcal{C}_k|} \left( \sum_{r \in \mathcal{C}_k} y_r - i + 1 \right) & \forall k \in \mathcal{OD},\ i \in \{1, \ldots, |\mathcal{C}_k|\}
\end{aligned}
$$

traveler mode choice

$$
w_k = \sum_{i=1}^{|\mathcal{C}_k|} \Delta_k^i \cdot b_k^i \qquad \forall k \in \mathcal{OD}
$$

capacity constraints per line segment

$$
\sum_{k \in \mathcal{OD}} \sum_{r \in \mathcal{C}_k(l,t)} \delta_k \cdot w_r \leq z_l \cdot \kappa_l \qquad \forall t \in \mathcal{T},\ l \in \mathcal{P}(t)
$$

minimal service requirement

$$
\sum_{l \in \mathcal{P}(t)} z_l \geq f_t \qquad \forall t \in \mathcal{T}
$$

domains of variables

$$
\begin{aligned}
z_l &\in \{0,1\} & \forall l \in \mathcal{P} \\
y_r &\in \{0,1\} & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \\
b_k^i &\in \{0,1\} & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k,\ i \in \{1, \ldots, |\mathcal{C}_k|\} \\
w_r &\in [0,1] & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \\
w_k &\in [0,1] & \forall k \in \mathcal{OD}
\end{aligned}
$$

choice according to any choice model is estimated with Equations (7). The second part deals with the operational requirements from Section 3.4. It is ensured by Equations (8) that sufficient seating capacity is available for the expected number of passengers on each line segment. This links the line selection and estimation of passenger demand in the model. Furthermore, the minimal service requirement in Equations (9) ensures a minimal frequency on each part of the network. The last part defines the domains of the variables.

# 4  Experimental setup

We solve the model ($LPwMC$) with the branch and bound method implemented in the Fico® Xpress Optimizer version 35.01. All experiments are conducted on the Lisa cluster[2] operated by SURFsara with a time limit of one hour per model run. In the following sections, the instances are introduced, the derivation of passenger routes is explained and the choices for parameters described.

## 4.1  Instances

We consider the intercity network of the Randstad, a metropolitan area in the Netherlands. This is a partial network of the network operated by the largest Dutch railway operator, Netherlands Railways (NS). The network contains 21 stations connected by 31 direct tracks between them. The network is depicted in Figure 2 and denoted by $IC21$, indicating the number of stations in the network. The line pool $\mathcal{P}$ contains 107 lines, 43 of which are duplicates in order to be able to model higher line frequencies. The pool contains all relevant lines that currently operate in the considered area and is given in Appendix B. A reference line plan is available that is used as a feasible start solution. The reference line plan is a solution of a line planning model with fixed passenger assignment. Based on this reference line plan, a competing mode such as driving by car, and passenger count data from NS, traveler demand was estimated with the logit model. This resulted in 174 undirected OD pairs with positive traveler demand.
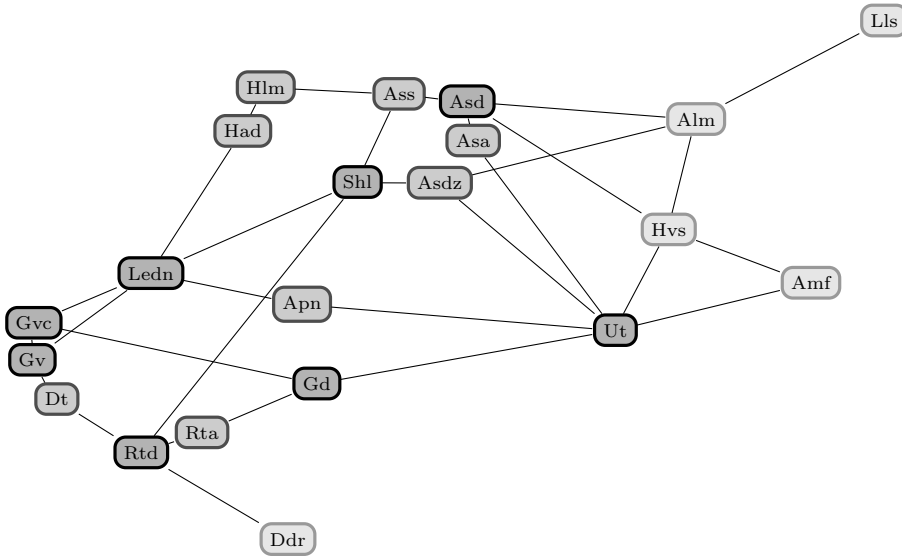


Figure 2: IC Network of Randstad region in the Netherlands

To obtain a variety of instances in some experiments, we consider two additional instances $IC08$ and $IC16$ and randomized demand situations for all networks. These instances have 8 and 16 stations, respectively,

---

and are sub-networks of network $IC21$. In Figure 2 the stations in network $IC08$ are marked by a dark node color scheme. The stations that are additionally in network $IC16$ have a lighter grey color scheme. The remaining stations with the lightest color scheme are only contained in network $IC21$. The *randomized demand situations* are generated by multiplying the number of travelers $\delta_k$ with a random number between 0.5 and 1.5 for each OD pair $k$.

## 4.2 Passenger routes

In a preprocessing step, a choice set of passenger routes $\mathcal{C}_k$ for each OD pair $k$ is determined and used as input to the model ($LPwMC$). As described in Section 3.3.1, we consider the journey-time shortest passenger routes for the route choice sets. In our experiments, we model this with a tolerance coefficient $\alpha$ and a tolerance addend $\varepsilon$ to limit the maximally acceptable journey time. First, we derive the shortest possible journey time $\hat{j}_k$ for each OD pair based on an extensive line pool for each network. Then, a route $r$ is in the set $\mathcal{C}_k$, if and only if its journey time $j_r$ is at most $\alpha\hat{j}_k + \varepsilon$. Only routes that are at most 5% ($\alpha = 1.05$) and $\varepsilon = 10\,\mathrm{min}$ longer than the shortest possible journey time were accepted. The journey time of a route comprises drive and dwell times and a transfer penalty, if applicable. Average driving times per track are used and at all stations a dwell time of 4 min is assumed. A penalty of 20 min is added to the journey time of routes including a transfer. In the experiments, we restrict the route choice sets to routes with at most one transfer.

This yields a set $\mathcal{C}_k$ of passenger routes of comparable journey time for each OD pair $k$. For the intercity network of the Randstad $IC21$ from Figure 2, a total of 6391 routes are considered, which is on average 36.7 routes per OD pair.

## 4.3 Parameters

The monetary values for ticket price $p_k$ for a passenger of OD pair $k$, and cost $c_l$ related to line operation were chosen to represent a simplified situation for the Dutch railway operator NS. All lines are operated by trains with a capacity of $\kappa_l = 1000$ seats.

For estimating the mode choice, the logit model is used with a logit parameter of $\beta = -0.1$. We estimate the adaption time based on the assumption that demand is uniformly distributed over the period, and that route departures are spread evenly over the hour. We therefore arrive at an adaption time of half of the considered period of 60 min, divided by the number of available routes. The alternative mode resembling individual transport does not have an adaption time. Hence, the utility $\hat{u}_k$ of the alternative mode is quantified by the journey time only. We use the SAQ method (FGSV, 2008) to estimate the journey time based on the Euclidean distance between stations.

For the minimal service requirement (9), a minimal track frequency of $f_t = 2$ for all tracks $t \in \mathcal{T}$ is used. This is in line with the requirements for the Dutch railway operator NS.

# 5 Comparison and analysis

To test the line planning model with integrated mode choice, we conduct experiments on the Randstad network. We test means to improve the computational performance and compare our model ($LPwMC$) with a standard line planning model with fixed passenger demand to investigate differences in solution quality.

## 5.1 Improvement of computational performance

We test the impact of adding order constraints (*order*), adding symmetry breaking constraints (*sym*), and setting priorities for branching (*prio*) on the solution-finding process. We observe the CPU time until an optimal solution is found, and the gap to the best bound found in case the time limit of 1 hour is exceeded. The settings *order*, *sym* and *prio* are explained in the following.

***order*** The auxiliary variables $b_k^i$ are used for counting the number of available routes for each OD pair. By definition, $b_k^i$ equals 1 if for OD pair $k$ at least $i$ routes are available, and 0 otherwise. This implies that $b_k^i$ can only be 1 if $b_k^{i-1}$ equals 1. The other way around, $b_k^i$ can only be 0 if $b_k^{i+1}$ equals 0. This relation can be modeled with the order constraints

$$b_k^i \geq b_k^{i+1} \quad \forall k \in \mathcal{OD}, \ \forall i \in \{1, \ldots, |\mathcal{C}_k| - 1\}$$

***sym*** The binary decision variables $z_l$ model whether line $l$ is selected in the line plan or not. To model higher frequencies of a line, duplicates are considered in the line pool that can be selected independently. To break the symmetry implicated by this setting, we consider constraints enforcing an order of selection of a line and its duplicates.

***prio*** The solution of the line planning problem is uniquely determined by the line plan $\mathcal{L}$, that means, the solution values for the line selection variables $z_l$. The corresponding solution values of all other variables can be reconstructed from the solution values of $z_l$. However, model ($LPwMC$) uses three different sets of binary variables, $z_l$, $y_r$, and $b_k^i$, where $y_r$ and $b_k^i$ are auxiliary variables to model the availability and the number of routes. By default, any of these variables can be used for branching. We test whether branching first on the variables $z_l$ is preferable to the standard branching strategy of Fico Xpress. The Xpress Optimizer offers the option to set the branching priority of a variable between 0 and 1000, where always a variable with a lower priority number will be selected for branching. We set a high branching priority (1) for variables $z_l$, medium branching priority (500) for variables $y_r$, and low branching priority (999) for variables $b_k^i$.

The tests are conducted on the three networks $IC21$, $IC16$, and $IC08$ with 10 randomized demand situations per network. Table 1 gives the CPU times in seconds and the gaps to the best bound for solving the model ($LPwMC$). The CPU times in Table 1 are averaged over ten randomized demand situations on each network. Only those runs that exceed the time limit are included for computing the average gap. The number in brackets behind the average gap shows how often the time limit of 1 hour

is exceeded. The first and last row show results for the settings where *none* (reference case) or *all* three options are used, respectively.

| | average CPU time[s] | | | average gap[%] | | |
|---|---|---|---|---|---|---|
| | $IC08$ | $IC16$ | $IC21$ | $IC08$ | $IC16$ | $IC21$ |
| none | 166 | 3600 | 3600 | - | 30.5(10) | 8.9(10) |
| order | 144 | 3600 | 3600 | - | 11.4(10) | 17.2(10) |
| sym | 31 | 303 | 3194 | - | - | 1.6(5) |
| prio | 1034 | 3600 | 3600 | - | 2.5(10) | 9.3(10) |
| all | 32 | 248 | 1305 | - | - | - |

Table 1: CPU times in seconds and gaps in percent to the best bound for model ($LPwMC$) in different settings. The number in brackets gives the number of cases where the time limit of 1 hour is exceeded

Adding order constraints (setting *order*) reduces the average CPU time for the smallest network and the average gap for the midsize network. However, it increases the average gap for the largest network compared to the reference case (setting *none*). The reason might be that the additional constraints increase the problem size. This could initially make the search for a feasible solution more difficult, but accelerate the solution process at a later point in time. The symmetry-breaking constraints (setting *sym*) significantly reduce the CPU times and gaps for all instance sizes. For the largest instances with 21 stations, the time limit is exceeded in only five out of ten cases and the resulting average gap is with 1.6% very small. Setting branching priorities (setting *prio*) drastically increases the CPU time for $IC08$, showing that the solver was able to find better branching strategies for the small network. In contrast to that, the average gap for the medium network size is significantly reduced by setting the branching priorities, and the found solutions were close to the optimum. For the large instances, no improvement is found with the setting *prio*. The gaps are slightly higher than in the reference case without branching priorities. Tests with all strategies combined (setting *all*) yield by far the best CPU times and all instance sizes can be solved to optimality within the time limit of one hour. The largest instance considered with 21 stations is solved within an average CPU time of less than 22 minutes. Therefore, we keep this setting with all options (*order*, *sym* and *prio*) for further experiments.

## 5.2   Comparison of ($LPwMC$) with line planning without mode choice

To investigate the added value of estimating passenger loads during optimization, we compare our model ($LPwMC$) with a line planning model ($LP$) without integrated mode or route choice. Such a model requires a passenger assignment to tracks as input and assumes that passenger demand and distribution are independent of the solution found. The objective is to find a line plan meeting all demand with minimal cost. Since the passenger demand is assumed to be fixed, the revenue is constant and this objective corresponds with finding a profit-optimal line plan in model ($LPwMC$). Similar to model ($LPwMC$), model ($LP$) considers a minimal service requirement and seating capacity constraints. In contrast to model ($LPwMC$), the capacity constraints are aggregated per track. Individual constraints per line segment are not feasible since the passengers are assigned to the network before the line plan is

found. The MIP formulation for the line planning model $(LP)$ used in the experiments can be found in Appendix C.

To obtain a passenger assignment for model $(LP)$, we assume a fixed percentage of travelers to use public transport and distribute them uniformly on the routes in the choice set $\mathcal{C}_k$. This yields a passenger load on each track in the network. We test model $(LP)$ with an assigned modal share for public transport ranging from 40% to 90%.

Table 2 shows the modal share $MS$ for public transport in decimals, and the revenue $R$, cost $C$, and profit $P$ of the found line plans. The values are given for model $(LP)$ with different assigned passenger shares and for model $(LPwMC)$. The column $P^{(LP)}$ gives the objective value of model $(LP)$, that is, the profit assuming the assigned passenger demand used as input. This value is only available for model $(LP)$ with assigned passenger shares. All monetary values are given in relation to the profit of the line plan found by model $(LPwMC)$, which is normalized to the value 100. That means a value of 110 implies that the corresponding monetary value is 10% higher than the profit of the solution of model $(LPwMC)$.

|            | $MS$ | $R$   | $C$   | $P$   | $P^{(LP)}$ |
|------------|------|-------|-------|-------|------------|
| $(LP)$ $(0.4)$ | 0.36 | 174.1 | 115.3 | 59.2  | 98.4       |
| $(LP)$ $(0.5)$ | 0.36 | 176.0 | 115.3 | 60.7  | 151.7      |
| $(LP)$ $(0.6)$ | 0.42 | 205.0 | 128.7 | 76.3  | 191.6      |
| $(LP)$ $(0.7)$ | 0.42 | 203.4 | 129.3 | 74.1  | 244.5      |
| $(LP)$ $(0.8)$ | 0.44 | 211.2 | 133.6 | 77.6  | 293.8      |
| $(LP)$ $(0.9)$ | 0.48 | 229.9 | 155.5 | 74.5  | 325.2      |
| $(LPwMC)$      | 0.49 | 240.5 | 140.5 | 100.0 | -          |

Table 2: Results of models $(LP)$ with different passenger shares and $(LPwMC)$. The modal share is given as decimal and the monetary values are normalized such that the profit of the solution of $(LPwMC)$ equals 100

The numbers in brackets in the first column give the share of travelers in decimals that are *assigned to use* public transport in the input for model $(LP)$. Based on the found solution, we estimate the expected share of travelers that *decide to use* public transport in a subsequent distribution of passengers with the logit model. This average modal share $MS$ coincides with the integrated modal share estimate in model $(LPwMC)$ and is given in the second column of Table 2. For all tests with model $(LP)$, the *assigned* share of travelers is higher than the *estimated* modal share $MS$. While the modal share is comparable for a low number of assigned travelers, the assigned and estimated modal share significantly differ for high numbers. This shows that in our experiments, the solutions of model $(LP)$ do not attract as many passengers as they were planned for.

In particular, line planning models without an integrated estimation of mode choice are not suitable for operators that strive for increasing their modal share. It is striking that the highest modal share is achieved by the solution of model $(LPwMC)$.

The monetary observation variables revenue $R$, cost $C$, and profit $P$ are based on the modal share $MS$ as estimated with the logit model. Both revenue and cost increase with the given passenger share for model $(LP)$ but the profit stagnates at around 75% of the profit generated by the solution of

model ($LPwMC$). The last column gives the anticipated profit $P^{(LP)}$ based on the *assigned* share of travelers, i.e., the objective value of model ($LP$). Especially for high values of the assigned passenger share, the estimated profit $P^{(LP)}$ is significantly higher than the corresponding profit based on estimated passenger numbers. By fixing the modal split before making the line plan, line planning models such as ($LP$) are prone to drastically overestimating the number of passengers and, with this, their revenue and profit. This causes these models to choose for solutions in which many lines are established, without attracting sufficient passengers to be profitable in the end.

The experiments clearly show the superiority of our model ($LPwMC$) over line planning models without an integrated mode choice. Due to an accurate estimation of passenger demand, profitable line plans of high quality can be designed that attract a high number of travelers.

# 6    Practical insights

Over years, travel demand is constantly changing and public transport operators need to react with an adjustment of their service. For operators, it is important to understand the impact of changes in travel demand on the quality of their service and the generated profit. Due to the integration of route and mode choices, model ($LPwMC$) is capable of computing profit-optimal line plans for different levels of travel demand. In this section, we outline which insights this can provide for operators.

The integrated traveler mode choice allows conducting a sensitivity analysis of the travel demand on passenger service level and operator performance. The different levels of traveler numbers are obtained by multiplying the original traveler demand with a factor ranging from 0.5 to 1.5. Figure 3 shows the results per traveler factor. The corresponding data is given in Table 4 in Appendix D. We analyze the number of lines $|\mathcal{L}|$ in the line plan, the number of available routes $|\mathcal{R}|$, the modal share $MS$, and the profit $P$ of the line plans found by model ($LPwMC$) for different levels of traveler demand. The profit $P$ is normalized such that the profit of the solution for the traveler factor 1.0 equals 100. The number of available routes $|\mathcal{R}|^{OD}$ averaged per OD pair and the number of available routes $|\mathcal{R}|^{pax}$ averaged per passenger are examined separately.

With higher traveler numbers, more lines $|\mathcal{L}|$ can be installed and the supply for passengers improves. Accordingly, the number of available routes for OD pairs $|\mathcal{R}|^{OD}$ and passengers $|\mathcal{R}|^{pax}$ increase. Large OD pairs mostly have direct routes in the route choice sets, while for small OD pairs we often observe transfer routes. Since the number of lines increases, many more line combinations form feasible transfer routes for passengers. Hence, the number of available routes increases proportionally more for small OD pairs than for large OD pairs, which explains the steeper increase of $|\mathcal{R}|^{OD}$ than of $|\mathcal{R}|^{pax}$.

The higher number of available routes implies an improved service level for passengers and with this a higher modal share for public transport. The increase in modal share from 44% to 52% is rather moderate, considering that the overall number of travelers as specified in the input data triples, and the average number of routes per OD pair almost doubles. The low effect on the modal split can be explained with the structure of the choice model that we use to estimate the travelers' mode choice. As indicated in Figure 1, the increase in modal share flattens for a higher number of available routes. This is in line with observations in the real world where the modal split is hardly affected by improvements in service
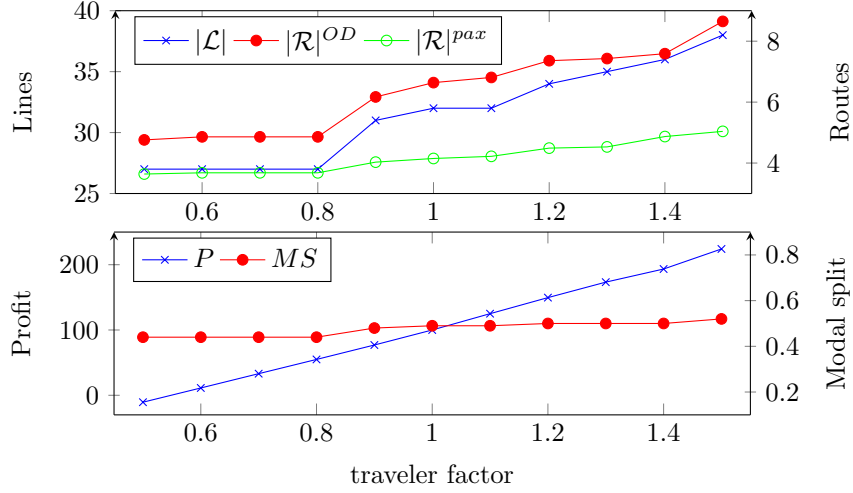
Figure 3: Results of model ($LPwMC$) for varying passenger load

for passengers, once a certain service level is reached.

Nevertheless, the profit increases approximately linearly from approximately negative 10% to 225% of the reference profit for an increasing number of travelers. The negative profit values for low passenger numbers are a result of the minimal service level required by governmental regulations, which enforces to operate two lines on each track, even if this cannot be done profitably. This increase in profit can be explained by the high costs for operating a basic line plan on the whole network. For up to 80% of the travelers, the capacity of such a basic line plan is sufficient on the considered instance. Until then, the number of lines, the number of routes, and the modal split stay almost constant. Only for more travelers, and thus for more passengers, it pays off to install more lines. It is interesting to see that the slightly higher modal share and thus the higher revenues offset the costs for additional lines and, overall, lead to an approximately linear increase in profit.

The evaluation shows that the operator's profit is very sensitive to changes in the traveler demand. In case of declining traveler demand, operators cannot prevent losses even if they react to demand changes in an optimal way. This is in line with recent observations where operators incur tangible losses due to considerably lower traveler demand caused by the Covid-19 pandemic despite efforts to reduce the service level. Conversely, operators can profit greatly from growing traveler demand when exploiting the full potential of passenger demand estimation during line planning.

# 7   Conclusion and outlook

## 7.1   Conclusion

The line plan significantly determines the service level of public transport for passengers. It has an impact on how many travelers decide to use public transport, and which routes passengers use. In this paper,

we present a line planning model with integrated mode and route choice models. In contrast to most existing approaches, both choices are modeled from a passenger's perspective and are not driven by a system optimum. This allows an accurate estimation of passenger demand during optimization, resulting in line plans that are tailored for the demand they generate.

Considering passenger choice models during line plan optimization is a complex and hard-to-solve problem. In order to obtain a tractable model, we assume (1) that passengers distribute uniformly on the best available routes and (2) that the utilities of alternative modes are independent of the designed line plan. By considering only shortest routes for passengers and making these two assumptions, the presented model can be linearized and solved with existing branch and bound methods.

The linear model presented in this paper may be combined with any choice model to estimate the mode choice of travelers. In particular, the choice model does not need to be linear. Due to the two assumptions made, the traveler mode choice can be preprocessed using the preferred choice model without affecting the solving of the line planning model. In the paper, a logit model is used to estimate the mode choice. We discuss additional constraints and branching priorities for improving the computational performance and show their effectiveness in experiments. The advantages of integrated passenger choice models are outlined in a comparison with a standard line planning model that relies on predetermined passenger loads. Based on the results of this comparison, considering demand estimation during line planning is strongly recommended. Integrated demand estimation yields line plans that are well-suited for the demand they generate. They are more profitable for operators and feature a higher level of service for passengers compared to line plans found based on fixed passenger demand. Furthermore, we analyze the sensitivity of the public transport service level and operator profit on fluctuations in travel demand. This gives valuable insights into the business models of operators and suggests that operators should react to changes in travel demand regularly.

## 7.2   Outlook

In additional experiments, we have noticed that the profit of two line plans can be very similar while their respective costs and revenues are different from each other. Indeed, two line plans with the same profit do not need to have any line in common. This is an interesting observation from both an algorithmic and an application point of view.

On the one hand, this implies that the concept of 'neighborhood' of solutions, although a key element in many (heuristic) approaches to solve line planning models, is less useful than in other contexts. On the other hand, for operators it can be very valuable to see different solutions with similar profit. It would allow them to choose between solutions with different modal shares. This particularly motivates the search for multiple good solutions instead of one optimal solution only. One approach to obtain multiple solutions could be to modify the ticket price $p_k$ for passengers or the operational costs $c_l$ of lines to shift the weights between revenue and cost in the objective function.

Future research should address customized solution approaches for model ($LPwMC$) to find line plans for larger instances, and possibly with the feature to compute multiple good solutions.

**Acknowledgements**

The authors thank Robin Koot for his help with the experiments and Joël van't Wout from Netherlands Railways (NS) for the data and input. We thank SURFsara[3] for the support in using the Lisa Compute Cluster.

---

# References

Bertsimas, D., Y. S. Ng, and J. Yan (2021). "Data-Driven Transit Network Design at Scale". In: *Operations Research*.

Borndörfer, R., M. Grötschel, and M. Pfetsch (2007). "A column-generation approach to line planning in public transport". In: *Transportation Science* 41.1, pp. 123–132.

Borndörfer, R. and M. Karbstein (2012). "A direct connection approach to integrated line planning and passenger routing". In: *OASIcs-OpenAccess Series in Informatics*. Vol. 25. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Bussieck, M. (1998). "Optimal lines in public rail transport". PhD thesis. Citeseer.

Bussieck, M., P. Kreuzer, and U. Zimmermann (1997). "Optimal lines for railway systems". In: *European Journal of Operational Research* 96.1, pp. 54–63.

Canca, D., A. De-Los-Santos, G. Laporte, and J. Mesa (2016). "A general rapid network design, line planning and fleet investment integrated model". In: *Annals of Operations Research* 246.1, pp. 127–144.

Canca, D., A. De-Los-Santos, G. Laporte, and J. Mesa (2017). "An adaptive neighborhood search metaheuristic for the integrated railway rapid transit network design and line planning problem". In: *Computers & Operations Research* 78, pp. 1–14.

Canca, D., A. De-Los-Santos, G. Laporte, and J. Mesa (2019). "Integrated Railway Rapid Transit Network Design and Line Planning problem with maximum profit". In: *Transportation Research Part E: Logistics and Transportation Review* 127, pp. 1–30.

Claessens, M., N. van Dijk, and P. Zwaneveld (1998). "Cost optimal allocation of rail passenger lines". In: *European Journal of Operational Research* 110.3, pp. 474–489.

De Dios Ortuzar, J. and L. Willumsen (2011). *Modelling transport*. John Wiley & Sons.

FGSV (2008). *Richtlinien für integrierte Netzgestaltung (RIN)*. (in German). Forschungsgesellschaft für Straßen-und Verkehrswesen, Köln.

Friedrich, M., M. Hartl, A. Schiewe, and A. Schöbel (2017). "Integrating Passengers' Assignment in Cost-Optimal Line Planning". In: *17th workshop on algorithmic approaches for transportation modelling, optimization, and systems (atmos 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Gattermann, P., J. Harbering, and A. Schöbel (2017). "Line pool generation". In: *Public Transport* 9.1, pp. 7–32.

Goerigk, M. and M. Schmidt (2017). "Line planning with user-optimal route choice". In: *European Journal of Operational Research* 259.2, pp. 424–436.

Goossens, J.-W., S. van Hoesel, and L. Kroon (2006). "On solving multi-type railway line planning problems". In: *European Journal of Operational Research* 168.2, pp. 403–424.

Guan, J., H. Yang, and S. C. Wirasinghe (2006). "Simultaneous optimization of transit line configuration and passenger line assignment". In: *Transportation Research Part B: Methodological* 40.10, pp. 885–902.

Kaspi, M. and T. Raviv (2013). "Service-oriented line planning and timetabling for passenger trains". In: *Transportation Science* 47.3, pp. 295–311.

Klier, M. and K. Haase (2015). "Urban public transit network optimization with flexible demand". In: *OR Spectrum* 37.1, pp. 195–215.

Laporte, G., Á. Marín, J. Mesa, and F. Ortega (2007). "An integrated methodology for the rapid transit network design problem". In: *Algorithmic methods for railway optimization*. Springer, pp. 187–199.

Laporte, G., J. Mesa, F. Ortega, and I. Sevillano (2005). "Maximizing trip coverage in the location of a single rapid transit alignment". In: *Annals of Operations Research* 136.1, pp. 49–63.

Liu, D., P. Vansteenwegen, G. Lu, and Q. Peng (2019). "An iterative approach for profit-oriented railway line planning". In: *Linköping University Electronic Press*, pp. 806–825.

De-Los-Santos, A., G. Laporte, J. Mesa, and F. Perea (2017). "The railway line frequency and size setting problem". In: *Public Transport* 9.1, pp. 33–53.

Nachtigall, K. and K. Jerosch (2008). "Simultaneous network line planning and traffic assignment". In: *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Schiewe, A., P. Schiewe, and M. Schmidt (2019). "The line planning routing game". In: *European Journal of Operational Research* 274.2, pp. 560–573.

Schmidt, M. (2014). *Integrating routing decisions in public transportation problems*. Springer.

Schmidt, M. and A. Schöbel (2015). "The complexity of integrating passenger routing decisions in public transportation models". In: *Networks* 65.3, pp. 228–243.

Schöbel, A. (2012). "Line planning in public transportation: models and methods". In: *OR Spectrum* 34.3, pp. 491–510.

Schöbel, A. and S. Scholl (2006). "Line planning with minimal traveling time". In: *5th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS'05)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Steiner, K. and S. Irnich (2018). "Schedule-based integrated intercity bus line planning via branch-and-cut". In: *Transportation Science* 52.4, pp. 882–897.

# A    Notation

**Sets**

|  |  |
|---|---|
| $\mathcal{C}$ | Route choice set |
| $\mathcal{L}$ | Line plan |
| $\mathcal{OD}$ | Set of OD pairs |
| $\mathcal{P}$ | Line pool |
| $\mathcal{R}$ | Available routes |
| $\mathcal{S}$ | Stations |
| $\mathcal{T}$ | Tracks between stations |

**Indices**

|  |  |
|---|---|
| $i$ | Iterator used to count available routes |
| $l$ | Line |
| $k$ | OD pair |
| $r$ | Route |
| $s$ | Station |
| $t$ | Track between two stations |

**Parameters**

|  |  |
|---|---|
| $a_k$ | Adaption time for OD pair $k$ |
| $\beta$ | Logit parameter |
| $c_l$ | Cost for operator for installing line $l$ |
| $\delta_k$ | Traveler demand of OD pair $k$ |
| $\Delta_k^i$ | Increment in modal share if $i$ instead of $i-1$ routes are available |
| $f_t$ | Minimal frequency on track $t$ |
| $j_k$ | Journey time for OD pair $k$ |
| $\kappa_l$ | Capacity of line $l$ |
| $p_k$ | Ticket price for OD pair $k$ for using public transport |
| $u_k$ | Utility of public transport for OD pair $k$ |
| $\hat{u}_k$ | Utility of alternative mode for OD pair $k$ |

**Variables**

| | |
|---|---|
| $b_k^i$ | $\in \{0,1\}$: OD pair $k$ has at least $i$ available routes or not |
| $w_k$ | $\in [0,1]$: Share of travelers of OD pair $k$ using public transport |
| $w_r$ | $\in [0,1]$: Share of passengers using route $r$ |
| $y_r$ | $\in \{0,1\}$: Route $r$ is available or not |
| $z_l$ | $\in \{0,1\}$: Line $l$ is selected or not |

**Observation variables**

| | |
|---|---|
| $C$ | Cost for operator |
| $MS$ | Modal share according to logit model |
| $P$ | Profit (= revenue - cost) |
| $R$ | Revenue generated by transporting passengers |

# B  Line pool

| count | sequence of stations | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | Alm | Asd | | | | | | | |
| 1 | Alm | Asd | Ass | Hlm | | | | | |
| 1 | Alm | Asdz | Shl | Ledn | | | | | |
| 1 | Alm | Asdz | Shl | Ledn | Gv | | | | |
| 1 | Alm | Asdz | Shl | Ledn | Gvc | | | | |
| 1 | Alm | Asdz | Shl | Rtd | | | | | |
| 1 | Alm | Asdz | Shl | Rtd | Ddr | | | | |
| 1 | Alm | Hvs | Amf | | | | | | |
| 2 | Alm | Hvs | Ut | | | | | | |
| 1 | Alm | Lls | | | | | | | |
| 2 | Amf | Hvs | Asd | | | | | | |
| 1 | Amf | Hvs | Alm | Lls | | | | | |
| 1 | Amf | Hvs | Asd | Ass | | | | | |
| 2 | Amf | Hvs | Asd | Ass | Hlm | | | | |
| 1 | Amf | Ut | | | | | | | |
| 1 | Amf | Ut | Apn | Ledn | | | | | |
| 2 | Amf | Ut | Gd | Gvc | | | | | |
| 4 | Amf | Ut | Gd | Rta | Rtd | | | | |
| 1 | Amf | Ut | Gd | Rta | Rtd | Ddr | | | |
| 1 | Asd | Alm | Lls | | | | | | |
| 1 | Asd | Ass | Hlm | Had | Ledn | | | | |
| 3 | Asd | Ass | Hlm | Had | Ledn | Gvc | | | |
| 2 | Asd | Ass | Shl | Ledn | Gv | | | | |
| 1 | Asd | Ass | Shl | Ledn | Gvc | | | | |
| 1 | Asd | Ass | Shl | Ledn | | | | | |
| 1 | Asd | Ass | Shl | Rtd | | | | | |
| 1 | Asd | Ass | Shl | Rtd | Ddr | | | | |
| 1 | Asd | Shl | Gv | Rtd | | | | | |
| 3 | Asd | Shl | Rtd | | | | | | |
| 1 | Asd | Shl | Rtd | Ddr | | | | | |
| 2 | Asdz | Shl | Ledn | Gvc | | | | | |
| 1 | Ass | Asd | Alm | Lls | | | | | |
| 4 | Ass | Asd | Asa | Ut | | | | | |
| 1 | Ass | Shl | Ledn | Gv | | | | | |
| 2 | Ass | Shl | Ledn | Gvc | | | | | |
| 1 | Ass | Shl | Rtd | Ddr | | | | | |
| 1 | Ddr | Rtd | | | | | | | |
| 1 | Ddr | Rtd | Dt | Gv | | | | | |
| 1 | Ddr | Rtd | Dt | Gv | Gvc | | | | |
| 1 | Ddr | Rtd | Dt | Gv | Ledn | | | | |
| 1 | Ddr | Rtd | Dt | Gv | Ledn | Had | Hlm | | |
| 2 | Ddr | Rtd | Dt | Gv | Ledn | Shl | Ass | Asd | Alm | Lls |
| 2 | Ddr | Rtd | Rta | Gd | Ut | | | | |
| 2 | Ddr | Rtd | Shl | Asdz | Alm | Lls | | | |
| 1 | Gv | Ledn | | | | | | | |
| 1 | Gv | Ledn | Had | Hlm | | | | | |
| 1 | Gv | Ledn | Shl | Asdz | Alm | Lls | | | |
| 2 | Gv | Ledn | Shl | Ass | | | | | |
| 5 | Gvc | Gd | Ut | | | | | | |
| 3 | Gvc | Gv | Dt | Rtd | | | | | |
| 1 | Gvc | Ledn | | | | | | | |
| 1 | Gvc | Ledn | Had | Hlm | | | | | |
| 2 | Gvc | Ledn | Shl | Asdz | Alm | Lls | | | |
| 2 | Gvc | Ledn | Shl | Ass | Asd | Alm | Lls | | |
| 2 | Hlm | Ass | Asd | Alm | Lls | | | | |
| 1 | Hlm | Had | Ledn | | | | | | |
| 2 | Hlm | Had | Ledn | Gv | Dt | Rtd | | | |
| 3 | Ledn | Apn | Ut | | | | | | |
| 1 | Ledn | Gv | Dt | Rtd | | | | | |
| 3 | Ledn | Shl | Asdz | Alm | Lls | | | | |
| 1 | Lls | Alm | Asdz | Shl | Rtd | | | | |
| 2 | Lls | Alm | Hvs | Ut | | | | | |
| 4 | Rtd | Rta | Gd | Ut | | | | | |
| 4 | Shl | Asdz | Ut | | | | | | |

Table 3: Line pool for Intercity network of Randstad. The first column gives the number of occurrences of the line in the pool and the remaining columns give the sequence of stations on the line

# C  MIP formulation for a standard line planning model with fixed passenger demand

max profit = revenue - cost

$$\max \quad \sum_{k \in \mathcal{OD}} p_k \cdot \delta_k \cdot w_k - \sum_{l \in \mathcal{P}} c_l \cdot z_l$$

capacity constraints per track

$$\sum_{k \in \mathcal{OD}} \sum_{r \in \mathcal{C}_k(t)} \delta_k \cdot w_r \quad \leq \quad \sum_{l \in \mathcal{P}(t)} z_l \cdot \kappa_l \quad \forall t \in \mathcal{T}$$

minimal service requirement

$$\sum_{l \in \mathcal{P}(t)} z_l \quad \geq \quad f_t \qquad \qquad \forall t \in \mathcal{T}$$

domains of variables

$$z_l \quad \in \quad \{0, 1\} \qquad \forall l \in \mathcal{P}$$

The objective is to maximize profit, defined as revenue minus cost. Note that the mode choice of travelers $w_k$ is assumed to be known and fixed in this model. Hence, the revenue is constant and the objective is equivalent to minimizing cost. Similarly, the passenger assignment $w_r$ to routes is predetermined and constant in the model. Consequently, the constraints for determining passenger loads are omitted and only the constraints ensuring sufficient capacity and a minimal service on each track remain. Passengers are assigned to routes in the route choice set to obtain the approximate passenger load on each track between to stations. Since it is not known yet which routes will be available in the solution, only aggregated capacity constraints per track can be applied. The set $\mathcal{C}_k(t)$ denotes the set of all routes in the choice set for OD pair $k$ via track $t$.

# D Results sensitivity analysis on travel demand

|  | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $|\mathcal{L}|$ | 27 | 27 | 27 | 27 | 31 | 32 | 32 | 34 | 35 | 36 | 38 |
| $|\mathcal{R}|^{OD}$ | 4.76 | 4.86 | 4.86 | 4.86 | 6.17 | 6.64 | 6.81 | 7.36 | 7.43 | 7.59 | 8.65 |
| $|\mathcal{R}|^{pax}$ | 3.64 | 3.68 | 3.68 | 3.68 | 4.03 | 4.15 | 4.22 | 4.49 | 4.53 | 4.87 | 5.04 |
| $MS$ | 0.44 | 0.44 | 0.44 | 0.44 | 0.48 | 0.49 | 0.49 | 0.5 | 0.5 | 0.5 | 0.52 |
| $P$ | -10.6 | 11.2 | 33.1 | 54.9 | 77.0 | 100.0 | 125.0 | 149.6 | 173.3 | 193.4 | 224.1 |

Table 4: Results of model ($LPwMC$) for varying passenger load. The number of lines and routes are counts. The modal share is given as decimal and the monetary values are normalized such that the profit of the solution with traveler factor 1.0 equals 100