Pricing commercial train path requests based on societal costs

Abderrahman Ait Ali, Jennifer Warg and Jonas Eliasson

Tweet

LiU LINKÖPING UNIVERSITY

Pricing Commercial Train Path Requests Based on Societal Costs

Abderrahman Ait Ali[1,3,*], Jennifer Warg[2] and Jonas Eliasson[3]

[1]*Swedish National Road and Transport Research Institute (VTI), Malvinas väg 6, SE-114 28 Stockholm, Sweden*
[2]*KTH Royal Institute of Technology, Brinellvägen 23, SE-114 28 Stockholm, Sweden*
[3]*Linköping University, Luntgatan 2, SE-602 47 Norrköping, Sweden*

Abstract

On deregulated railway markets, efficient capacity allocation is important. We study the case where commercial trains and publicly controlled traffic ("commuter trains") use the same railway infrastructure and hence compete for capacity. We develop a method that can be used by an infrastructure manager trying to allocate capacity in a socially efficient way. The method calculates the loss of societal benefits incurred by changing the commuter train timetable to accommodate a commercial train path request, and based on this calculates a reservation price for the train path request. If the commercial operator's willingness-to-pay for the train path exceeds the loss of societal benefits, its request is approved. The calculation of these benefits takes into account changes in commuter train passengers' travel times, waiting times, transfers and crowding, and changes in operating costs for the commuter train operator(s). The method is implemented in a microscopic simulation program, which makes it possible to test the robustness and feasibility of timetable alternatives.

We show that the method is possible to apply in practice by demonstrating it in a case study from Stockholm, illustrating the magnitudes of the resulting commercial train path prices. We conclude that marginal societal costs of railway capacity in Stockholm are considerably higher than the current track access charges.

Keywords: Train timetables; societal costs; train conflicts; commuter trains; commercial trains; train path pricing; railway capacity allocation

(*) Corresponding author. Tel.: +46 8 555 367 81; E-mail address: abderrahman.ait.ali@vti.se

1. Introduction

The deregulation of railway markets in many countries have meant that it has become increasingly common that several operators run trains on the same track and hence compete for the same capacity. Inevitably, conflicts arise between different operators´ capacity requests, and in such cases the infrastructure manager needs to prioritize between conflicting path requests. Depending on the ownership of the railway infrastructure, the infrastructure manager can have different objectives. In this paper, we will consider the case where infrastructure is publicly owned, and the infrastructure manager wants to allocate capacity in a way that maximizes total societal benefits.

There is a clear need for methodological development to overcome these challenges. Very large social and commercial values are at stake, so making the capacity allocation process more efficient can potentially generate substantial benefits. Moreover, the process and decision criteria need to be transparent and consistent to ensure fair and efficient competition between operators. In current practice, however, capacity allocation processes are often opaque, and it is common that simplified decision criteria or rules of thumb are used, such as faster trains having priority over slower ones, or passenger services having priority over freight.

Highly simplified, there are two principal ways for an infrastructure manager to resolve conflicting path requests, either based on benefit judgments or on willingness-to-pay (WTP). In processes based on benefit judgments, the infrastructure manager compares some criteria or calculations intended to reflect the societal benefits generated by alternative capacity allocations. The allocation which scores highest, according to these criteria, is chosen. In processes based on WTP, the operator that is willing to pay the highest price gets priority. WTP-based processes include slot auctions and demand-differentiated track charges. Both methods have their respective strengths and weaknesses. The problem of processed based on benefit judgment is that the data necessary for judging the benefit of a service is not always available. There exist well-developed methods to calculate societal benefits and costs of transport services, but such calculations need detailed data about demand, ticket prices and operating costs. For commercial traffic, such data is almost never available to the infrastructure manager, either because it is sensitive business information known only to the operator(s), or because this data is unknown at the time of the decision. The problem in WTP-based processes is that the societal benefits of publicly controlled traffic – for example, the subsidized commuter trains serving many large urban regions – do not necessarily correspond to the responsible public agency's willingness (or ability) to pay. Since the societal benefits of urban public transport cannot easily be "observed" in the same way as a commercial operator can observe its profits, it is much more difficult for public transport agencies to correctly assess its "willingness to pay" for capacity; both over- and underestimations may occur. Moreover, the societal benefits generated by urban public transport accrue to society at large, not to the agency, so there is no obvious link between the societal benefits generated by commuter train

services and the responsible agency's financial resources or hence its ability to pay for capacity. This is illustrated by the observation that public transport agencies are often under great financial strain, despite generating substantial societal benefits, not least because urban commuter trains are usually heavily subsidized (often for good reasons).

In this paper, we propose a hybrid method to resolve capacity conflicts between publicly controlled traffic (i.e., train services where supply and fares are determined by a public transport agency striving to maximize social welfare) and commercial traffic (i.e., passenger or freight train services run by companies striving to maximize profits). For brevity, we will call the publicly controlled services "commuter trains" in the following. Such services are often run by one or more operators contracted by the agency, but this is inessential in this context.

The main idea of the proposed method is to calculate a reservation price for a commercial operator's path request by estimating the societal costs (i.e., loss of benefits) of the changes needed in a baseline commuter train timetable to accommodate this path request. If the commercial operator is willing to pay this reservation price, it is awarded the path and the commuter train timetable is adjusted; if not, the request is declined. The process can be extended to handle multiple commercial path requests which also allows an infrastructure manager to prioritize between several commercial operators competing for the same capacity, i.e., on-rail competition.

This circumvents the problems explained above – on the one hand that an infrastructure manager does not have access to information necessary to assess the societal benefits of a commercial train service, and on the other hand that there is no clear correspondence between public agencies' WTP:s and the societal benefits generated by the services they run. Instead of only comparing benefit calculations or only comparing WTP:s, the process proposed in this paper compares benefit calculations for commuter train services to the WTP of commercial train services. The advantages of such a process is that it utilizes the respective characteristics of commuter trains and commercial traffic: for commuter trains, the information needed for benefit calculations (fares, passenger volumes, operating costs etc.) is not secret and can be relatively obtained easily; for commercial traffic, commercial operators have a relatively good perception of the profit they would make on a given train path, and hence of the price they are willing to pay for a train path request.

The economic logic of the idea is most easily explained as a way to price the negative externality the commercial train service incurs on the commuter trains. When the commuter train timetable is adjusted to make room for the commercial train request, this causes a loss for the commuter train passengers and the operator(s), which is an external effect from the point of view of the commercial operator. It is well known that pricing an externality will increase total societal benefits, if there are no other market imperfections. But there are at least two types of market imperfection that are worth considering: unpriced externalities of road traffic, and monopoly power of commercial train services. If there are unpriced road traffic externalities, then it may be motivated to subsidize either commuter train services or commercial train services, or both (so-called second-best pricing). If commercial train operators have some degree of monopoly power, then they will supply less services than socially optimal, which may warrant various kinds of regulations. The assumption in this paper is that such market imperfections are handled outside of the capacity allocation process. There are a variety of ways to do this, and many are already in use, for example subsidizing commuter train fares, setting track charges lower than marginal costs, regulating commercial train monopolies, and pricing road traffic externalities through fuel taxes, parking charges and congestion pricing. All of these methods are both simpler and more efficient ways to handle market imperfections than trying to handle them in the capacity allocation process. Hence, we assume that we can ignore these market imperfections in this context, and then it follows that pricing the externality caused by commercial train services interfering with commuter trains will increase the overall social welfare.

The purpose of the paper is to explain and develop the proposed method, and to apply it in a real-world application. In a case study from Stockholm, we calculate reservations prices for a commercial train service by adjusting a baseline commuter train timetable in three different ways: removing a commuter train service, shifting its departure time, and increasing its running time by letting it wait along the line while the commercial train passes by. The corresponding reservation prices are calculated for three different time periods, morning and afternoon peak hours and mid-day off-peak, and compared to the current track charges (Trafikverket, 2017). The adjustments to the commuter train timetable are done in the microscopic simulation tool *RailSys* (Radtke and Bendfeldt, 2001) to guaranty conflict-free timetable solutions.

The paper contributes to the literature on capacity conflict resolution in several ways. First, it describes a concrete method to resolve conflicts between commuter and commercial trains. This, in turn, is an important component in an overall process of capacity allocation, which also needs to include resolving other kinds of conflicts such as conflicts between different commercial operators, and between train services with different planning horizons (Broman et al., 2018). Second, the method can also be used as a decision tool for traffic planners to evaluate societal effects of alternative rescheduling scenarios. Third, the implementation and application of the method shows its practical feasibility, and shows how cost-benefit analysis can be combined with a large-scale commercial microscopic simulation tool such as *RailSys* for real-world train timetabling and planning applications.

Fourth, results from the case study indicate a realistic range of reservation prices that commercial trains should pay if capacity should be reallocated from commuter to commercial trains. Our results indicate that the reservation prices in our case study are considerably higher than current track charges in Stockholm.

The paper has the following structure: Section 2 gives a brief literature review of recent research on train timetable assessment. Section 3 describes the method for calculating the societal costs of changes in a baseline commuter train timetable. Section 4 presents the case study, with input data and results. Section 5 concludes.

2.   Previous related research

The idea proposed in this paper is similar to the one proposed by Johnson and Nash (2008). They examine the feasibility of using opportunity costs for each train slot to price scarce rail capacity on congested franchised lines in Britain. Opportunity costs are calculated based on consumer surplus, externalities and operation costs, and based on the additional traffic attracted by the slot, the additional quality for the users due to the slot, the savings of external costs for users that switch from road to rail, as well as the costs for not running that train. A demand model is used to evaluate the effects of changes in the transport system that are larger than the ones considered in this paper. The basic idea is similar to the current paper, but there are a number of differences between the suggested models. For example, we avoid having to develop and calibrate a demand model by working directly with origin-destination matrices. The advantage is that one can then accurately evaluate even minor changes in the timetables; on the other hand, keeping demand constant is a disadvantage when evaluating large timetable changes (we return to this issue below).

Using pricing to allocate scarce rail capacity is still rather uncommon in practice, even if several countries use it to some minor extent (Nash, 2005). Nash et al. (2004) point out that the lack of a price for scarce capacity is the major defect in the current British system for capacity allocation. Several authors such as Affuso (2003), Nilsson (2002) and Newbery (2003) discuss and propose various auction-based methods to allocate capacity, focusing on competing commercial services. The current paper, similar to the one by Johnson and Nash (2008), complements these approaches by proposing a method to resolve conflicts between publicly controlled services and commercial ones.

There are still not many papers that use cost-benefit analysis methods to evaluate train timetables, although there a few examples such as Adler et al. (2010) who study how competition between airlines and high-speed rail may affect service levels, and how this affects the benefits of high-speed rail investments. Eliasson and Börjesson (2014) also apply cost-benefit analysis on timetable construction, highlighting the impact of timetable assumptions on appraisal of railway investments. (Brännlund et al., 1998) is also an example of capacity allocation with an economic objective; they present an algorithm that schedules a set of trains to obtain a profit-maximizing timetable without violating track capacity constraints. More generally, however, there is a vast literature on valuing elements of timetables, such as waiting times, transfer and in-vehicle time, e.g., (Hensher, 1997), (Hensher and Ton, 2002) and (Balcombe et al., 2004).

There is also of course a vast literature evaluating timetables from various perspectives. A classic and widely studied question is how timetabling affects total capacity. This can be studied with a combination of analytical methods, simulation and optimization methods (Abril et al., 2008). For reproducing railway operations in the best way, dynamic, synchronous, microscopic, stochastic simulation tend to work best (Borndörfer et al., 2018). In addition to capacity effects, operational characteristics of timetables such as punctuality, stability and robustness have been extensively studied. For example, Delorme et al. (2009) focus on the stability aspect of the train timetables by developing a timetable stability module that is used to build an optimization model for railway operation planning.

There are also studies evaluating the performance of timetables from a passenger perspective, for example Kunimatsu et al. (2012), who use microsimulation of both passengers and trains in the railway network. Passenger behavioral aspects such as avoiding transfers and choosing best routes are also accounted for. The timetable is evaluated based on the disutility of passengers due to delays and crowding. This evaluation model is useful to capture detailed aspects of single line train timetables with a passenger perspective. The use of big data helps study even larger networks such as in the work by Jiang et al. (2016). Other studies have attempted to aggregate multiple passenger-related aspects of train timetables such as the sum of weighted waiting times, average of unit waiting time and maximum ratio of waiting time to travel time into a fuzzy analytic hierarchy process (Isaai et al., 2011).

3.   Evaluation Model

The general idea of the method is to calculate the loss of benefits in commuter train traffic incurred when adjusting their timetable to accommodate an additional commercial train path. This loss of benefits, measured in monetary terms, yields the price that a commercial train operator requesting the path has to pay. If the operator is

willing to pay the requested amount (the reservation price), the operator is awarded the requested path; otherwise the request is denied. As discussed in the introduction, this can be viewed as a way to price the externality caused by the commercial train service when it interferes with the commuter train services. It can also be seen as a way to correctly price the "input good" (the track capacity) need to produce the "good" that is the commercial train service. Both interpretations are well-known economic principles that lead to a socially efficient market equilibrium, provided that other market imperfections can be ignored. There are obviously several market imperfections in transport markets, such as road traffic externalities and some degree of monopoly power for commercial train operators, but there are both simpler and more efficient ways to handle these than in the capacity conflict resolution process, and many of them are also used in practice, such as subsidized fares and track charges, fuel taxes and congestion charges. Hence, it is reasonable to expect that introducing a price for track capacity that more accurately reflects its societal cost should lead to an overall increase in total societal benefits.

The method starts with an initial situation, where commuter trains are run according to a baseline timetable, and a train path request from a commercial operator which conflicts with the baseline timetable. Usually, commuter trains aim to be operated with regular headways (4-8 trains per hour) and frequent stopping pattern, while the commercial train operates longer distances less regularly with fewer stops and higher average speeds. Both the difference in average speed and the trains´ inflexibility to change their departure times due to marketing and capacity reasons makes it difficult to allocate capacity. The method calculates the value of the loss of societal benefits resulting from this modified commuter train timetable compared to the baseline timetable, and this loss of benefits (measured in monetary terms) constitutes the reservation price.

The loss of benefits ΔB consists of two parts: the change in consumer surplus ΔC and the change in producer surplus ΔP. The change in consumer surplus reflects all changes affecting passengers, for example changes in waiting times, crowding, transfers and travel time. The change in producer surplus reflects all changes affecting the operator, for example changes in operations costs and revenues. It is quite possible that a change may result in, for example, a loss of consumer surplus but a gain in producer surplus. Cancelling a train service, for example, would normally produce a gain in producer surplus by decreasing the costs of operations, but a loss of consumer surplus, since some passengers may experience increased waiting times and increased crowding in the remaining trains. Other types of societal benefits, e.g., changes in road traffic emissions and tax revenues, are ignored since they are either relatively small or are internalized in other ways, e.g., corrective taxes.

3.1. Calculating the change in consumer surplus

Let T_{ijr} be the number of passengers arriving at time r to station i to travel to station j. The matrix $\{T_{ijr}\}$ is called the (dynamic or time-dependent) origin-destination (OD) matrix. Let c_{ijr} be the generalized cost of travelling from station i to station j starting at time r, and define it as the sum of the fare f_{ijr}, waiting time(s) w_{ijr} and the in-vehicle time t_{ijr}, converting the time components into money by multiplying with the value of waiting time β and the value of in-vehicle travel time α, see equation (1). As explained further below, α depends non-linearly on the crowding level in the train. The waiting time penalty β is in general a non-linear function of the headway, since when headways are long, passengers will spend part of the "waiting time" elsewhere by adjusting their schedule to the train's departure time. In this study, however, the service frequency is high, so the waiting time penalty can be assumed to be constant. Moreover, we do not distinguish between waiting at the first station and waiting at transfer stations in this study. Introducing this distinction, and allowing for non-linear headway valuations, is conceptually straightforward.

$$c_{ijr} = f_{ijr} + \alpha t_{ijr} + \beta w_{ijr} \tag{1}$$

A change in the baseline timetable will cause generalized travel costs to change, possibly inducing a change in passenger volumes as well. Let exponents 0 and 1 denote variables before and after the change, respectively. Using the rule-of-a-half approximation, the change in consumer surplus[1] is defined as in equation (2).

$$\Delta C = \sum_{ijr} T_{ijr}^0 (c_{ijr}^0 - c_{ijr}^1) + \frac{1}{2}\sum_{ijr}(T_{ijr}^1 - T_{ijr}^0)(c_{ijr}^0 - c_{ijr}^1) \tag{2}$$

---

[1] This definition rests on some conventional assumptions such as negligible income effects and a locally linear demand curve.

In the following, we ignore the second term in this expression, effectively assuming that $T_{ijr}^0 = T_{ijr}^1$. This is a reasonable assumption as we are considering small timetable changes. This simplifies the exposition and the following calculations, and the approximation error is small[2] for moderate changes in the generalized travel cost.

The timetable consists of a number of train services indexed by k. Given the timetable and the OD matrix, the number of passengers boarding and alighting train service k at station i, called B_{ik} and A_{ik} respectively, can be calculated (as explained in more detail further below). The number of passengers on train service k between station i and the subsequent station along the service line is $N_{ik} = \sum_{s=1}^{i} B_{sk} - A_{sk}$, where the summation is taken over the stations served by train service k. This allows us to rewrite the change in consumer surplus from equation (2) as the difference in the relevant part of the aggregate generalized cost, denoted by C, before and after the change, see equation (3).

$$\Delta C = C^0 - C^1, \quad where \quad C = \sum_k \sum_{i \in k} \alpha(N_{ik}) t_{ik} N_{ik} + \beta_{ik} w_{ik} N_{ik} \tag{3}$$

The notation $i \in k$ means that the summation should be taken over stations served by train service k; t_{ik} denotes the travel time from station i to the next station with train service k; w_{ik} denotes the average waiting time for train service k at station i.

A number of things should be noted. First, we assume that the timetable change does not affect fares, so the fare terms in the generalized cost cancel out. Second, the value of in-vehicle time α depends on the number of passengers in the train between each pair of stations – the more passengers in the train, the higher is the weight (i.e., the disutility) of in-vehicle travel time. A change in the timetable will typically change N_{ik} and hence α, since the crowding levels change. Third, the value of waiting time β_{ik} depends on i and k. This is because the marginal valuation of waiting time is falling since when waiting times (or rather, headways) are long, passengers can adapt their schedule to avoid waiting at the platform, see (Fosgerau, 2019) for a theoretical analysis of this. This matters mainly for long headways, however; in our case study, headways are so short that this distinction does not matter much. Finally, and most importantly, expressing ΔC in this way simplifies the implementation substantially, since there is no need to calculate and represent the entire generalized cost matrix $\{c_{ijr}\}$, or even have precise data on the full OD matrix $\{T_{ijr}\}$. It is enough to have data on the number of passengers in each train on each link (N_{ik}) and the arrival times of passengers to their departure stations (which determines their waiting times w_{ik} based on their destination station and the next train that is serving the latter). This data is much easier to measure than the full OD matrix or generalized cost matrix; station arrival data is often directly available from entry measurements (e.g., smartcard gantries), and passenger loads can be measured either by on-board counts, automatic counting or weighing systems. Of course, if the OD matrix is available, passenger loads, travel times and average waiting times can be calculated with the corresponding timetable. The latter method simplifies the calculation of ΔB. Further, using a simulation software also allows for introducing delays in the model.

The assumption that passengers take the next train that is serving their destination station is made possible since the commuter network in the case study has two lines forming an X without any branches and the studied train timetable does not include skip-stops or overtaking. In other possible case studies, networks and/or timetables can be more complex (with branches, skip-stops or overtaking). In this case, the passengers will take the quickest path to their destinations which may not necessarily correspond to the next train or to a direct trip (without transfer).

3.2. Calculating the change in producer surplus

The producer surplus is the difference between total fare revenues and operating costs. In the following, we assume that the timetable change does not affect fare revenues (although allowing for this is straightforward). This means that we only consider the change in operating costs. The total relevant operating cost can be separated into three parts: costs proportional to trains' running distances (e.g., maintenance), costs proportional to trains' running times (e.g., staff), and costs proportional to the number of wagons necessary to run the timetable (capital costs for vehicles). Hence, we can write the change in producer surplus ΔP as $\Delta P = P^0 - P^1$ where P is calculated as in equation (4).

---

[2] If the demand elasticity with respect to generalized cost is ε, the relative error of a relative change in generalized costs p is $\frac{\varepsilon p}{2}$. For example, a 10% change of the generalized cost with a demand elasticity of -0.5 gives a relative approximation error of $\frac{\varepsilon p}{2} = 2.5\%$. As we only consider small changes in the timetable, p will stay moderate.

$$P = (1 + K_{overhead})\left(K_{time} \sum_k \sum_{i \in k} t_{ik} + K_{distance} \sum_k \sum_{i \in k} d_{ik} + K_{wagon} N_{wagon} \right) \quad (4)$$

The number of wagons N_{wagon} necessary to run the timetable is estimated based on the total number of train services. Since the studied timetables include slight changes compared to the baseline timetable, it is possible to manually compute the changes in vehicle allocation compared to the reported number of wagons for the baseline timetable. For more substantial changes to the train timetable, an existing vehicle allocation model can be used to compute the changes in N_{wagon}. The parameters $K_{time}, K_{distance}, K_{wagon}$ and $K_{overhead}$ are calculated by analyzing the public transport agency's total operating costs by type (e.g., train staff, maintenance, capital costs etc.), allocating them to most relevant proxy for variable costs (distance, time or number of wagons) and dividing by the corresponding total (total distance, total time, total number of wagons).

3.3. Parameters

The parameters that are used in the benefit calculations in the case study are taken from various sources: partly from the research literature, partly from the Swedish cost-benefit guidelines (Trafikverket, 2016), and partly from the calculation manual of the Stockholm Public Transport Agency (SLL, 2017).

The notations, values and sources of the parameters that are used for the computation of the consumer surplus are provided in Table 1 whereas Table 2 provides the ones for the producer surplus (10 SEK is around 1€).

Table 1
Parameters for the computation of the consumer surplus.

Parameter	Formula	Values	Source
In-vehicle travel time	$\alpha = K_{crowding} \alpha_0$	Leisure: $\alpha_0 = 57$ SEK/h Commuting: $\alpha_0 = 74$ SEK/h	(Eliasson and Börjesson, 2014)
Crowding factor	Piecewise linear function: $K_{crowding} = \begin{cases} K_{0-0.5}, 0 \leq \frac{n}{C} < 0{,}5 \\ \vdots \\ K_{1.75-2}, 1{,}75 \leq \frac{n}{C} < 2 \end{cases}$ $\frac{n}{c} = $ load factor $n = $ number of passengers onboard $C = $ number of seats in the trains	- Leisure trips, passengers sitting: 0-50 % is 1.04 50-100 % is 1.14 100-125 % is 1.26 125-150 % is 1.39 150-175 % is 1.53 175-200 % is 1.69 - Leisure trips, passengers standing: 100-125 % is 1.94 125-150 % is 2.15 150-175 % is 2.39 175-200 % is 2.64 - Commuting trips, passengers sitting: 0-50 % is 0.86 50-100 % is 0.95 100-125 % is 1.05 125-150 % is 1.16 150-175 % is 1.27 175-200 % is 1.4 - Commuting trips, passengers standing: 100-125 % is 1.62 125-150 % is 1.79 150-175 % is 1.99 175-200 % is 2.2	(Wardman and Whelan, 2011)
Waiting time	Piecewise linear function: $\beta = \begin{cases} \beta_{0-10}, 0 \leq w_{ijr} < 10 \\ \vdots \\ \beta_{30-60}, 30 \leq w_{ijr} < 60 \end{cases}$ $w_{ijr} = $ waiting time	- Leisure trips: 0-10 min is 86 SEK/h, 11-30 min is 70 SEK/h and 31-60 min is 34 SEK/h - Commuting trips: 0-10 min is 74 SEK/h, 11-30 min is 53 SEK/h and 31-60 min is 26 SEK/h	(Algers et al., 2010)

Table 2
Parameters for the computation of the producer surplus (SLL, 2017).

Parameter	Notation	Value
Time-related production costs	K_{time}	2 000 SEK/h
Distance-related production costs	$K_{distance}$	30 SEK/(wagon km)
Fixed vehicle costs	K_{wagon}	5 000 000 SEK/wagon/year
Production overhead	$K_{overhead}$	9 % on top of total production costs

To compute the consumer surplus, we use the daily average for the distribution of trip types, i.e., 50% leisure, 50% commuting (SLL, 2017). Given that the valuations (e.g., for in-vehicle travel and waiting times) are quite similar for leisure and commuter trips (see Table 1) that is considered to be a reasonable assumption in the absence of detailed data on the temporal distribution of trip types over the studied periods of the day, However, the model allows to easily include the temporal distribution for trip types given the corresponding data.

In the producer surplus, vehicle operation costs are estimated based on the total yearly travel distance and operation time for the trains and cost values according to the cost-benefit analysis recommendation from the local operator. Operating costs include capital costs and maintenance of the rolling stock, fuel and staff. Further, indirect costs including capital costs, overhead and administration are included based on the total number of passenger and rolling stock kilometers. The total costs are converted to daily costs by dividing yearly costs by the factor 320.

In Sweden, most of these valuations are compiled and explained in the guide for cost-benefit analysis, i.e., ASEK guidelines (Trafikverket, 2016). Several other countries have similar documents and handbooks. The British practical guide by Balcombe et al. (2004) provides extensive evidence on travel time, crowding, interchanges or transfers. These values can be used in various applications, e.g., public transport demand model, rail franchise specification. The guidelines that are used in these different countries are mostly based on a common body of research results.

3.4. Estimating passenger flow data

An important input to the calculation of the consumer surplus is the time-OD-matrix $\{T_{ijr}\}$. We will briefly describe how it can be estimated from station entry data (i.e., smartcards), and how it then can be used together with a given timetable to estimate passenger loads in the trains for calculating in-vehicle travel and waiting time costs. Handling trips with transfers is explained last, since this requires additional calculations.

Most public transport operators have access to data on the number of passengers entering each station during a certain interval of time, since this can usually be registered by most kind of entry gates. If some kind of smartcards are used, this becomes even easier. Let O_{ir} be the number of passengers entering station i at time r. In our case study, we only have data on the total number of passengers during each 15-minutes time period, so we assume that the arrival rate is constant over each such time period.

A common limitation (which is also the case in our case study) is that there is no station exit data, i.e., $\{D_{jr}\}$, the number of passengers exiting station j at time r, is unknown. To overcome this, we assume that trips are symmetric, meaning that the total number of passengers entering a station j during a day is equal to the number of passengers alighting at that station, i.e., $\sum_r O_{jr} = \sum_{ir} T_{ijr} = \sum_r D_{jr}$. Together, this allows us to estimate the time-OD-matrix with a standard entropy maximization problem. See (Ait-Ali and Eliasson, 2019) for more details.

In our formulas so far, we have tacitly assumed that the discretization of time, indexed by r, is sufficiently fine-grained compared to the discretization of the timetable. In a practical implementation, however, it is obviously inefficient to actually calculate and store the number of departures per station for each minute, or even second; instead, a constant rate of passenger departures per unit time is assumed to be known for each time interval. The number of passengers boarding a train is then easily calculated by multiplying the corresponding arrival rate(s) with the time elapsed since the previous train departure. The number of passengers alighting a train can be calculated in a similar way.

Let $T_{ij}(r)$ be the continuous rate of passengers entering station i to travel to station j at time r, and let r_{ik} be the departure time of train service k from station i. The number of passengers B_{ik} boarding train k at station i is then given by equation (5) where $j \in k(i)$ denotes that the summation is taken over all stations j served by train service k from station i.

$$B_{ik} = \int_{r_{i,k-1}}^{r_{ik}} \sum_{j \in k(i)} T_{ij}(r) \, dr \tag{5}$$

The time resolution of the rate $T_{ij}(r)$ depends on available data; in our case study, we have data on arrival rates per 15 minutes time period. Fig. 1 provides an illustration of the calculation of the number of boarding passengers B_{ik}. A similar method is used to calculate the numbers of alighting passengers A_{jk}.
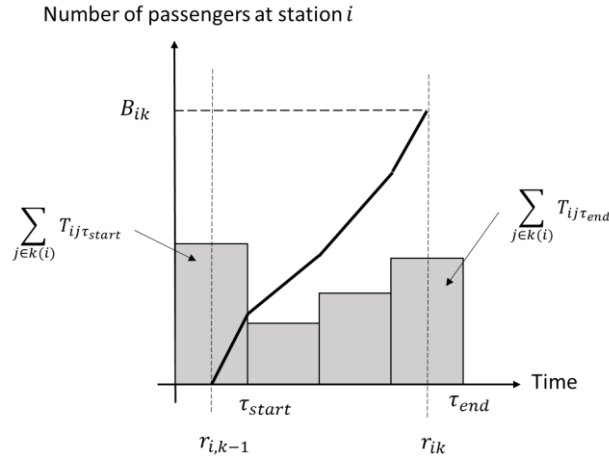

Fig. 1. Calculation of the number of passengers B_{ik} boarding train k at station i using OD matrix $\{T_{ijr}\}$.

The model is intended to be used for commuter services with high frequency leading to short headways between departures and therefore short waiting times for the passengers at the origin stations. This makes it possible to assume that passengers enter the origin stations at a uniform rate. However, for longer waiting times between two consecutive train departures, passengers may adapt their time of entrance to the stations. As previously explained in section 3.1, this adaptation is accounted for using a piecewise linear function which reflects the decreasing value of longer waiting times, see the last row in Table 1.

When dealing with passenger flows in complex train networks, it is necessary to handle trips that require transfers between train services. Given the OD matrix, transfer trips are split into two; one from the origin station to the transfer station and another one from the latter to the destination station. The number of passengers in each of the two trips is calculated using the same method as for any direct trip. These are added to the direct trips which gives the overall train loads including transfers. When there are several possible transfer stations, passengers are assumed to transfer at the first possible station. An illustration of the passenger flow per link segment including transfer trips is given in **Fig. *2***.
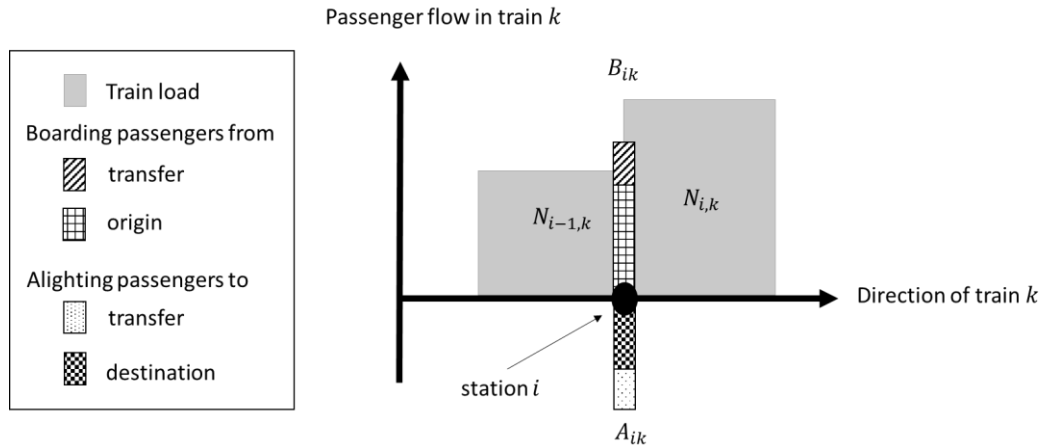

Fig. 2. Passenger flow in train service k at station i including transfers.

3.5. Implementation

The network and the train timetables are modelled using the microscopic simulation tool *Railsys* which allows for instance to easily create, handle and simulate train timetables. There are many reasons for this choice, but the main motives for the purpose of this paper are:

- The Swedish Transport Administration provides a network, train and timetable models in *RailSys* that can be easily adjusted.

- Manipulation and visualization of networks and train timetables is facilitated.
- Conflict detection (for timetable alternatives).
- Timetables can be simulated to allow evaluation of feasibility and robustness (e.g., delay propagation). This is out of the scope of the current study, but a major advantage for practical applications.
- Data can be exported easily (e.g., in XML/CSV format) for integration with other software environments.

The network is microscopically modelled in the simulation tool down to individual tracks, switches and signals. This gives more flexibility in investigating the effect of various aspects of the railway infrastructure on passengers and operators and the opportunity to check if a timetable solution is feasible. Due to the complexity of the network model, the network is always handled within the simulation tool and never exported outside.

The train timetables are however exported and used outside the simulation tool. This data is processed in order to extract information about arrival and departure times at every station, travel times and distances between stations. It is easy to create, modify and delete train paths for different train timetables scenarios. It is also possible to add potential delay distributions in the train timetable and simulate a certain number of days in order to estimate how the timetable will perform in operation.

4. Case study

This section applies the method on a case study from Stockholm, calculating reservation prices for a commercial train path request during three different time periods (morning peak, mid-day, afternoon peak). The case study illustrates the practical feasibility of the method, and also shows magnitudes of reservation prices, i.e., the loss of societal benefits in the commuter train system from allocating capacity to an additional commercial train.

4.1. Description of the Stockholm commuter trains

The Stockholm commuter trains (locally called *pendeltåg*) share parts of their tracks with commercial passenger[3] and freight trains (Froidh and Nelldal, 2015). Fig. 3 shows the network in 2016. For the central section, departure times and tracks are strictly regulated in order to coordinate the high frequency of services passing the double track line south of Stockholm Central. The signaling system in combination with the requirement to offer regular commuter traffic allows for 28 trains per hour (i.e., 24 commuter and 24 long distance), but four of these paths are not allocated in order to ensure some buffer (Trafikverket, 2013). While commuter and long-distance services are separated on most of the other lines, they share the tracks northwest of Stockholm between Karlberg and Bålsta (Fig. 3) which is the focus of the conflicts treated in this study. During peak hours, six commuter trains and two longer-distance trains per hour and direction have to be coordinated. However, in principle the whole commuter network is included in this study in order to cover all trips than can be affected by the adjustments. For the studied working day, the data includes around 346 commuter train departures allowing more than 230,000 individual trips.

---

[3] For regional passenger trains, frequent commuters can get their fares partly subsidized through a cap on their total fare expenses, where the excess is covered by a combination of regional public transport authorities. However, this complication is irrelevant for the present case study.

Fig. 3. Stockholm commuter network (Frohne, 2016).

The commuter network has an X-shape with around 50 stations along 240 km. All the trains pass a central section between *Karlberg* and *Älvsjö*. Commuter services are mainly operated in two channels, *Södertälje-Märsta* and *Nynäshamn-Bålsta*, complemented with additional services on shorter distances during peak hours. One of these complementary lines is prolonged to *Uppsala* via Sweden´s largest airport *Arlanda*. In addition, there is a branch connecting *Södertälje* to *Gnesta* (excluded in the case study for simplicity). Moreover, there are no skip-stop services or overtaking, the two daily departures that change branches are excluded in this study.

Between *Karlberg* and *Älvsjö*, the major lines share the infrastructure. *Stockholm central* is the hub of the network allowing connections to services such as long-distance passenger trains, local and regional buses, airport shuttles and metro.

In 2016, the Stockholm region had 2.3 million inhabitants (SCB, 2016). **Fig. *4*** presents the number of departing passengers from each station across the day. There are major travelling peaks in the morning and afternoon peak hours. Note the peak from Stockholm central station, especially in the afternoon peak hours.
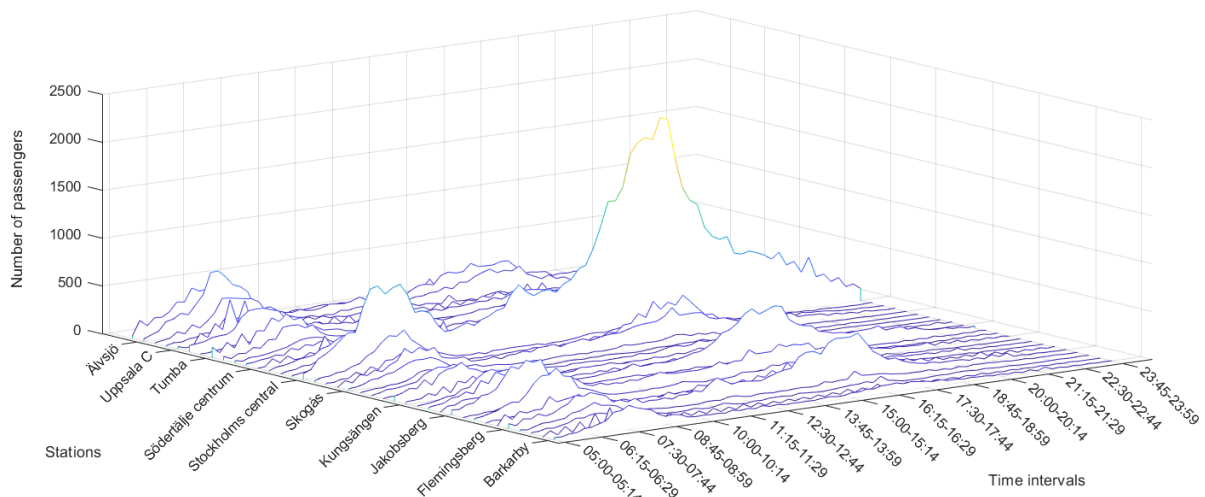

Fig. 4. Number of passengers (from smartcard data) travelling from different stations per 15 minutes time intervals.

Typical passenger loads for the different trains in a weekday in 2016 are presented in **Fig. 5**. All trains running between Bålsta and Nynäshamn in both directions are included. The x-axis shows the links between two consecutive stations on the line, the passenger load is given on the y-axis. The horizontal dashed line represents the train (i.e., two wagon units) seat capacity. In the case study, each train consists of two wagons of the model Cordia X60 with a seating capacity of 374 passengers for each wagon unit (ALSTOM, 2004). Although there are

10

extra train departures, crowding in the trains is common. The figure shows that passenger load on the line is well above the seating capacity for certain trains around the central station during peak hours.
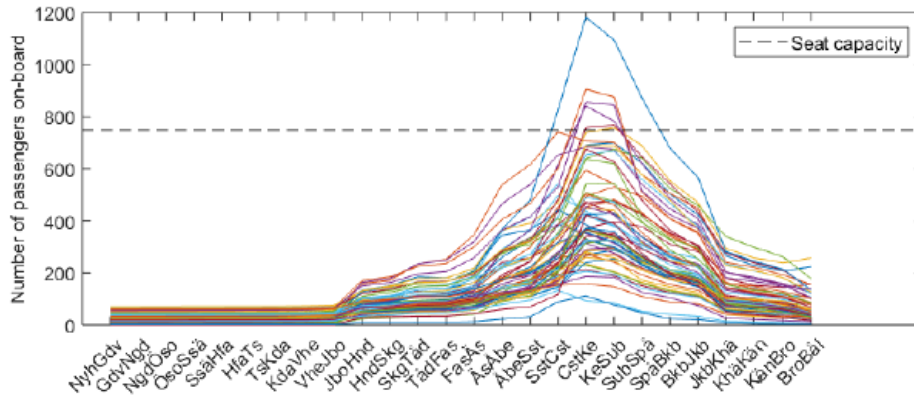

Fig. 5. Typical load of passengers of all the train paths during a weekday between Bålsta and Nynäshamn.

4.2. Allocating capacity for a commercial train path

In this case study, a working day in September 2016 on the line *Bålsta-Nynäshamn* is used as baseline timetable. **Fig. *6*** shows a graphical timetable, with a train path request in blue conflicting with a commuter train path in yellow. The conflict area is colored red, and the commuter trains not involved in the conflict are in purple. In a graphical timetable, each line represents the scheduled time (horizontal axis) and location (vertical axis) in a train run surrounded by squares showing how long each train blocks the belonging signal sequence (block section). Each block section can only be occupied by one train at a time.
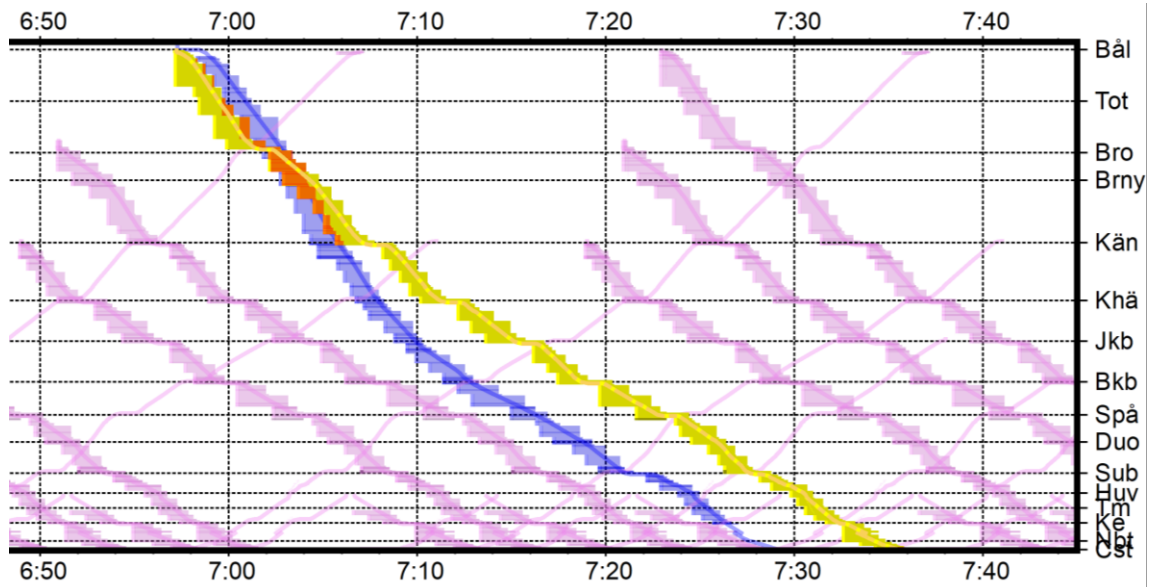

Fig. 6. Graphical timetable showing a conflict (red) between a commuter train (yellow) and a commercial train (blue) during the morning peak (around 7:00 AM) on the line between Bålsta and Stockholm central station. Commuter trains not involved in the conflict are purple.

In order to resolve the conflict, the commuter train timetable can be changed in either of three ways:
- Scenario S1 – Remove: the conflicting commuter train service is cancelled. Note that this means that the whole roundtrip train service needs to be cancelled.
- Scenario S2 – Delay: the commuter train´s dwell time at a certain stopping station is prolonged, in order to let the commercial train overtake.
- Scenario S3 – Shift: the commuter train´s departure time at the first station is shifted, so that the commercial train can pass first.
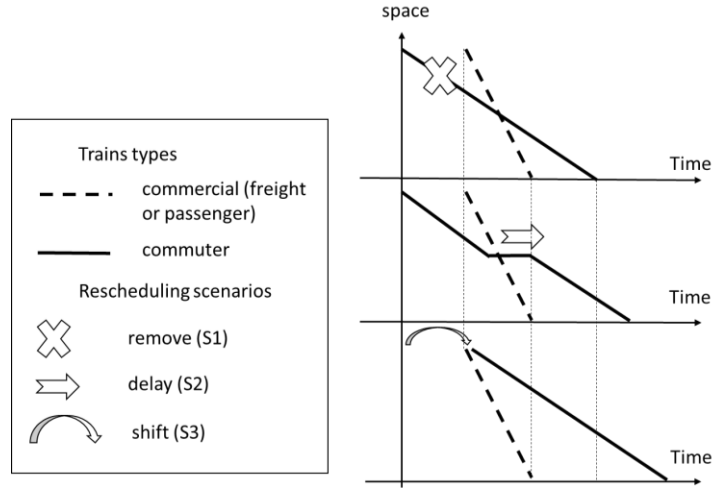
Fig. 7. Different experiment scenarios for conflicts between subsidized and commercial services

Changes were therefore made to the baseline timetable by rescheduling a certain train path using the different rescheduling scenarios, i.e., remove, shift and delay. These scenarios are illustrated in **Fig. *7*** with the graphical timetable showing the conflicts between the subsidized and the commercial trains and the different rescheduling scenarios. In order to avoid large costs, the adjusted timetables are designed not to largely affect vehicle and staff allocation. While the adjustments in S2 and S3 are within the available timetable margins, the whole roundtrip service is removed in S1.

The last two scenarios (i.e., delay (S2) and shift (S3)) differ from a passenger point of view mainly in the way delays are experienced. In both scenarios, average waiting and total crowding penalties will tend to increase, since passengers get less evenly spread across train services when departures are not at even intervals. In S2, passengers onboard also get a longer travel time.

These rescheduling scenarios, applied to the baseline timetable, are tested for conflicts in three different time periods, i.e., morning peak (6:00-9:00), mid-day off-peak (11:00-14:00) and afternoon peak (15:00-18:00). To illustrate this, **Fig. *8*** shows two graphical timetables presenting two rescheduling scenarios, i.e., delay (left) and shift (right). Lines show train paths; squares the infrastructure each path is blocking. Note that commercial trains that are not included in this study are not visualized here. The figure on the right shows the commuter timetable where the departure time of the conflicting commuter train path (yellow) has been shifted by +3 minutes whereas the one on the left shows that of delaying it in *Kungsängen* station for 4 minutes to let the commercial train pass. The third type of conflict solution is not presented in the figure but is about completely removing the conflicting commuter train path.
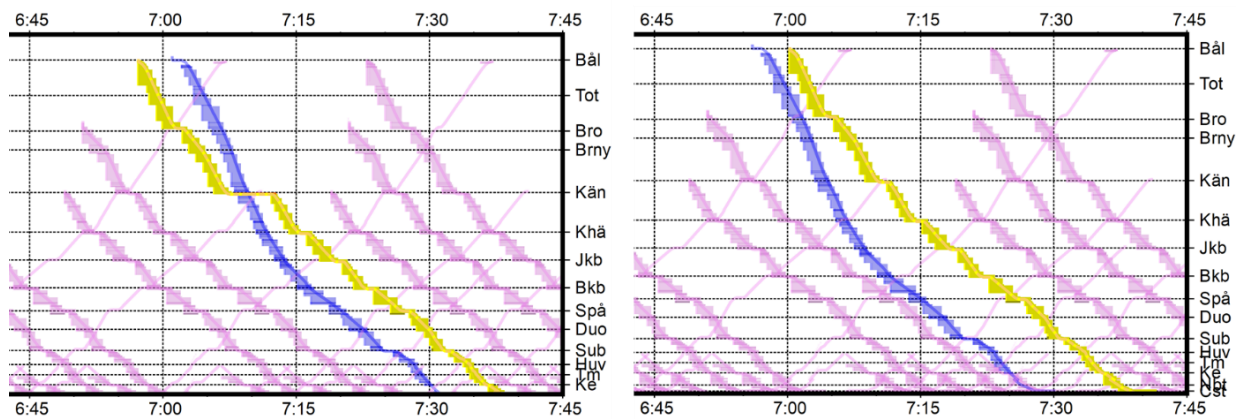

Fig. 8. Illustration of two different commuter timetable changes for conflict solution: S2-delay (left) and S3-shift (right). Adjusted commuter train path in yellow, other commuter trains in purple, commercial train in blue.

12

4.3. Results

The total societal costs of the different rescheduling scenarios are presented in **Fig. 9**. Changing the baseline timetable to accommodate the commercial train path can cause more than 100 kSEK (10 SEK is around 1€) in total societal losses in the commuter train system. Scheduling a commercial train leads to higher societal costs in the peak hours, especially in the morning. This can be explained by the variation in the trip distribution during the day, see Fig. 4. The minimal cost is incurred by shifting the departure time of the conflicting commuter train, resulting in a societal loss of around 9-16 kSEK. Removing the conflicting commuter train path (i.e., scenario S1) always lead to the highest societal costs, especially during peak hours. A less costly rescheduling strategy is to shift the departure time of the conflicting commuter train (i.e., scenario S3). Delaying the conflicting train at a certain station (S2) is also less costly but is more expensive than shifting the departures, especially in peak hours.
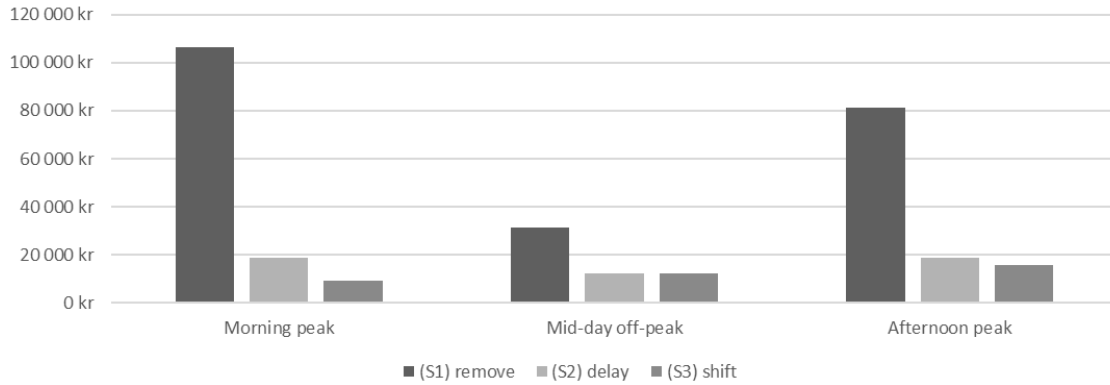

Fig. 9. Total societal costs of the changes to the baseline commuter train timetable (in SEK, 1€ is around 10 SEK).

In order to give more details about the results in **Fig. 9**, we present the different elements forming the total societal costs in Table 3 by showing the different consumer or passenger costs (in-vehicle travel time, waiting and transfer costs) and producer or operator costs (distance and time dependent costs and fixed costs). Negative costs in Table 3 indicate savings, such as decreased operating or shorter transfer times. A dash means no difference in the costs relative to the baseline timetable.

Table 3
Components of the total societal costs of changes relative to the baseline timetable (in SEK, 1€ is around 10 SEK).

Scenario	Time period	Consumer costs				Producer costs				Total costs
		Travel	Waiting	Transfer	Total	Distance	Time	Fixed	Total	
S1 – remove	Morning	97 499	18 835	-476	115 859	-6 370	-3 633	-900	-10 904	104 955
	Mid-day	32 433	9 482	-457	41 458	-6 370	-3 633	-900	-10 904	30 554
	Afternoon	70 817	21 298	-451	91 664	-6 370	-3 633	-900	-10 904	80 761
S2 – delay	Morning	14 059	2 038	-	16 096	-	133	12	145	16 242
	Mid-day	9 756	760	-	10 516	-	133	12	145	10 661
	Afternoon	12 138	6 920	-	19 058	-	133	12	145	19 203
S3 – shift	Morning	7 894	1 353	-	9 247	-	-	-	-	9 247
	Mid-day	9 571	1 059	-	10 630	-	-	-	-	10 630
	Afternoon	15 070	974	-	16 044	-	-	-	-	16 044

Removing the conflicting train always causes larger societal losses than delaying or shifting the commuter train. Indeed, cancelling the commuter train leads to higher consumer costs compared to the gains in producer costs, which in turns lead to a high total cost especially in peak hours. However, delaying the departure of the conflicting commuter train in a stopping station leads to higher costs in both consumer and producer costs. The latter is due to the increase in operation time due to the additional stopping time in the middle of the line. Shifting the departure times of the conflicting commuter train causes no additional producer costs and higher consumer costs. These results clearly show that shifting the trains is the societally optimal strategy to accommodate an additional commercial train path. Of course, this may not always be possible if the timetable is very dense.

Based on these societal costs, one can compute the minimum cost that the commercial operator(s) should pay to compensate the resulting losses in societal benefits. The last column of Table 3, i.e., total costs, shows this minimal price for the different rescheduling scenarios and time periods of the day. Compared to the current track charge for the train path, these total societal costs are substantially higher. Swedish track charges are made up of several components, one of which is set to partly reflect congestion (lack of capacity) on the tracks. For a train passing Stockholm, this "congestion charge" is 433 SEK (Trafikverket, 2017), i.e., less than 5% of the lowest

societal loss caused by accommodating an additional commercial train path. For instance, the total track charge for a commercial passenger train Västerås - Stockholm C sharing the track with the subsidized trains between Bålsta and Stockholm (45 km of the 105 km in total) is 1 454.46 SEK in total, including electricity and so on. Since the tracks are fully used for most parts of the day, the calculated costs can also be interpreted as the marginal value of increased capacity.

It is possible that certain adjustments of the commuter train timetable may result in positive societal benefits rather than losses, since there is no absolute guarantee that the baseline commuter train timetable is strictly optimal. In such cases, the reservation price of the accommodated train path is of course zero (except for the usual charges for wear-and-tear etc.). In some cases, however, such results may be a sign that the parameters of the benefit-cost calculation, or its input data, may need to be adjusted.

It is important to note that reliability is not included in this study. When disturbances are added and operation simulated, a timetable's robustness to delays can be studied and might change results (e.g., optimal strategy). Including reliability aspects, part of the future work, shows the importance and advantage of using a microscopic timetable simulation tool such as *Railsys*. Non-optimal timetables may also sometimes be due to political considerations; there is anecdotal evidence of over-supply from other commuting train systems. Exploring the optimality of timetables from a benefit-cost perspective is an interesting research area of its own but is left for future research.

5.   Conclusions and future work

This paper describes an approach to resolve conflicting capacity requests between commercial trains and public controlled traffic ("commuter trains"), by calculating the loss of societal benefits caused by accommodating the commercial train path and using this to price the commercial train path. The calculation of benefits takes into account in-vehicle times, waiting times, transfers, crowding and operating costs. The economic logic of this approach can be seen as a way to internalize the externality the commercial train causes on the commuter trains by having them to adjust their schedule. This means that if the commercial operator finds the price worth paying, overall social welfare increases by granting capacity to the commercial train service, and vice versa.

The case study of the commuter train services in Stockholm shows that the evaluation model can be used in different situations to help planners evaluate the impact of their timetable choices. Results also provide insights into how the model can be used to price commercial train paths in conflict with the commuter train services. The results show that accommodating additional train paths in the busy commuter train timetable comes at a high societal cost – much higher than the current track charge intended to partly reflect scarce capacity in Stockholm. We also show that it is possible to substantially reduce the costs of changes in commuter train timetables by choosing the right rescheduling alternative. This best alternative might not be evident to the planners without the help of a cost assessment model as the one presented in this paper.

The evaluation model that is presented in this paper is not limited to the experiment in the case study. It can be used in a number of other real-world situations to help railway planners make efficient changes to the timetables with minimal societal costs. Ideas for possible applications include the assessment of the societal effects of using a train timetable with skip-stops instead of all-stops. It is also possible to link the model to an optimization model in order to find optimal train timetables given a certain trip distribution. Moreover, the results of the evaluation model can be used to assess project proposals for railway capacity expansion by comparing the costs of such projects with the societal benefits of having a new train timetable (with expanded capacity) in the long run. This point is further discussed by Eliasson and Börjesson (2014).

Additionally, one may also improve the model by including extra costs and accounting for negative and positive externalities. Many possible future works can help improve the quality of the assessment model and the case study that is presented in this paper. As mentioned in the limitations, the OD-data was incomplete, hence only an estimate was used. Applying the model in a case study with complete input data may provide more accurate results and insights. Moreover, socio-economic data on the passengers, if available, can be used with an improved version of the model where disaggregated results can be computed. In this way, one can get insights as to which (how much each) socio-economic group is winning or losing. This gives further insights on the equity of the train service operations.

Acknowledgements

References

ABRIL, M., BARBER, F., INGOLOTTI, L., SALIDO, M. A., TORMOS, P. & LOVA, A. 2008. An assessment of railway capacity. *Transportation Research Part E-Logistics and Transportation Review,* 44**,** 774-806.

ADLER, N., PELS, E. & NASH, C. 2010. High-speed rail and air transport competition: Game engineering as tool for cost-benefit analysis. *Transportation Research Part B: Methodological,* 44**,** 812-833.

AFFUSO, L. 2003. Auctions of rail capacity? *Utilities Policy,* 11**,** 43-46.

AIT-ALI, A. & ELIASSON, J. 2019. Dynamic Origin-Destination-Matrix Estimation Using Smart card Data: An entropy maximization approach. *RailNorrköping2019.* Norrköping, Sweden.

ALGERS, S., BÖRJESSON, M., SUNDBERGH, P., BYSTRÖM, C. & ALMSTRÖM, P. 2010. Valuation of Time in Transport – The National Studies 2007/08 in Sweden. *WSP report.*

ALSTOM 2004. CORDIA 60X Stockholm Transport Renews its Commuter Fleet.

BALCOMBE, R., MACKETT, R., PAULLEY, N., PRESTON, J., SHIRES, J., TITHERIDGE, H., WARDMAN, M. & WHITE, P. 2004. The demand for public transport: a practical guide.

BORNDÖRFER, R., KLUG, T., LAMORGESE, L., MANNINO, C., REUTHER, M. & SCHLECHTE, T. 2018. *Handbook of Optimization in the Railway Industry*, Springer International Publishing.

BROMAN, E., ELIASSON, J. & ARONSSON, M. 2018. A Mixed Method for Railway Capacity Allocation. *21st Meeting of the Euro Working Group on Transportation 2018.* Braunschweig.

BRÄNNLUND, U., LINDBERG, P. O., NÕU, A. & NILSSON, J.-E. 1998. Railway Timetabling using Lagrangian Relaxation. *Transportation Science,* 32**,** 358-369.

DELORME, X., GANDIBLEUX, X. & RODRIGUEZ, J. 2009. Stability evaluation of a railway timetable at station level. *European Journal of Operational Research,* 195**,** 780-790.

ELIASSON, J. & BÖRJESSON, M. 2014. On timetable assumptions in railway investment appraisal. *Transport Policy,* 36**,** 118-126.

FOSGERAU, M. 2019. The marginal social cost of headway for a scheduled service. *Transportation Research. Part B: Methodological,* 43**,** 813-820.

FROHNE, E. 2016. Stockholm commuter train system map. Wikimedia Commons.

FROIDH, O. & NELLDAL, B. L. 2015. The impact of market opening on the supply of interregional train services. *Journal of Transport Geography,* 46**,** 189-200.

HENSHER, D. A. 1997. A practical approach to identifying the market potential for high speed rail: A case study in the Sydney-Canberra corridor. *Transportation Research Part A: Policy and Practice,* 31**,** 431-446.

HENSHER, D. A. & TON, T. 2002. TRESIS: A transportation, land use and environmental strategy impact simulator for urban areas. *Transportation,* 29**,** 439-457.

ISAAI, M. T., KANANI, A., TOOTOONCHI, M. & AFZALI, H. R. 2011. Intelligent timetable evaluation using fuzzy AHP. *Expert Systems with Applications,* 38**,** 3718-3723.

JIANG, Z., HSU, C.-H., ZHANG, D. & ZOU, X. 2016. Evaluating rail transit timetable using big passengers' data. *Journal of Computer and System Sciences,* 82**,** 144-155.

JOHNSON, D. & NASH, C. 2008. Charging for scarce rail capacity in Britain: a case study. *Review of Network Economics,* 7.

KUNIMATSU, T., HIRAI, C. & TOMII, N. 2012. Train timetable evaluation from the viewpoint of passengers by microsimulation of train operation and passenger flow. *Electrical Engineering in Japan,* 181**,** 51-62.

NASH, C. 2005. Rail Infrastructure Charges in Europe. *Journal of Transport Economics and Policy (JTEP),* 39**,** 259-278.

NASH, C., COULTHARD, S. & MATTHEWS, B. 2004. Rail track charges in Great Britain—the issue of charging for capacity. *Transport Policy,* 11**,** 315-327.

NEWBERY, D. M. 2003. Network capacity auctions: promise and problems. *Utilities Policy,* 11**,** 27-32.

NILSSON, J.-E. 2002. Towards a welfare enhancing process to manage railway infrastructure access. *Transportation Research Part A,* 36**,** 419–436.

RADTKE, A. & BENDFELDT, J.-P. 2001. Handling of railway operation problems with RailSys. Proceedings of the 5th world congress on rail research, 2001 Cologne. World congress on rail research.

SCB 2016. Population in the country, counties and municipalities on 31 December 2016 and Population Change in 2016. Statistics Sweden.

SLL 2017. Dokumentation av SAMS 3.0. Stockholm.

TRAFIKVERKET 2013. Planning conditions for Stockholm sector T15.

TRAFIKVERKET 2016. English summary of ASEK recommendations.

TRAFIKVERKET 2017. Network Statement 2018. Swedish Transport Administration.

WARDMAN, M. & WHELAN, G. 2011. Twenty Years of Rail Crowding Valuation Studies: Evidence and Lessons from British Experience. *Transport Reviews,* 31**,** 379-398.