Report
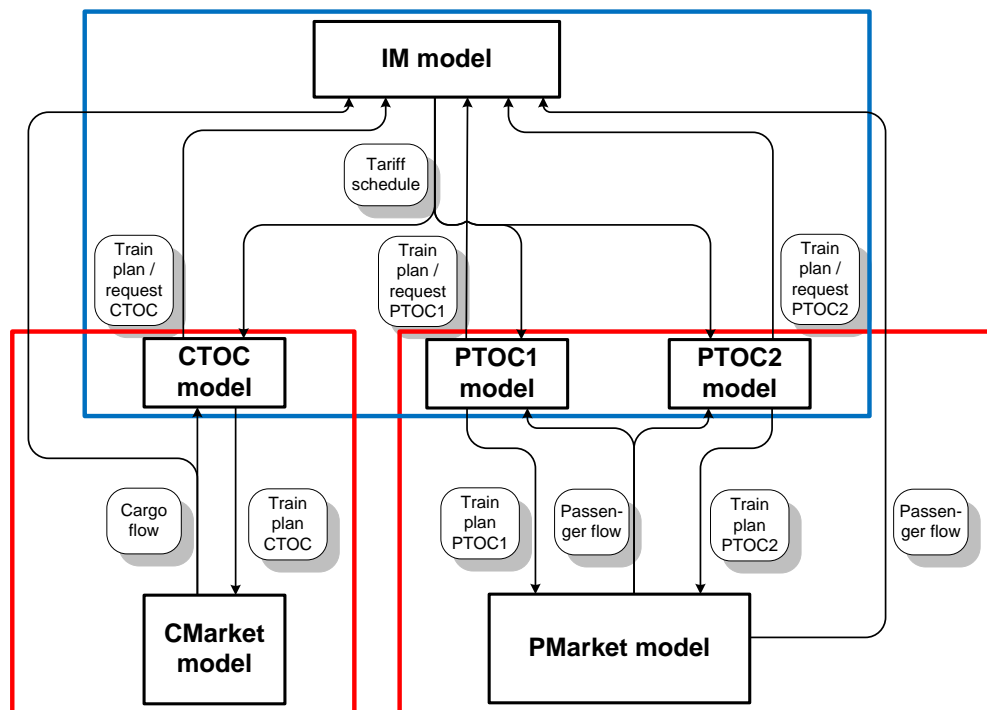
Socio-Economical Optimal Pricing of Railroad Infrastructure (SOPJI)

Mathematical Model Formulation and Implementation

Author(s)
Michal Kaut, Adrian Werner, Arnt-Gunnar Lium, Truls Flatberg, Hajnalka Vaagen, Teodor G. Crainic

SINTEF

Report

Socio-Economical Optimal Pricing of Railroad Infrastructure (SOPJI)

Mathematical Model Formulation and Implementation

VERSION	DATE
1.0	23rd January 2014

AUTHOR(S)
Michal Kaut, Adrian Werner, Arnt-Gunnar Lium, Truls Flatberg, Hajnalka Vaagen, Teodor G. Crainic

CLIENT(S)	CLIENT'S REFERENCE
The Research Council of Norway	Prosj.nr. 200575

PROJECT	NUMBER OF PAGES AND ATTACHMENTS
102000998	56

ABSTRACT
The project "Socio-Economic Optimal Pricing of Railroad Infrastructure" (SOPJI) developed an approach to support the process of granting access to railway infrastructure by way of scarcity tariffs. The idea can contribute to a more efficient utilization of existing and a clear identification of needs for future railway infrastructure.

This report presents the mathematical description of this model and a first prototype implementation. The implementation adjusts, through an iterative solution process, the scarcity tariffs such that over-utilization of the infrastructure is avoided. A socio-economic evaluation of the resulting train schedules is performed in a post-processing step.

We also describe some implementation aspects, the structure of the input parameter database, and potential data sources.

PREPARED BY	SIGNATURE
Adrian Werner	*A. Werner.*

CHECKED BY	SIGNATURE
Marte Fodstad	*Marte Fodstad*

APPROVED BY	SIGNATURE
Frode Rømo	*Frode Rømo*

REPORT NUMBER	ISBN
SINTEF A25843	978-82-14-05667-9

CLASSIFICATION	CLASSIFICATION THIS PAGE
Unrestricted	Unrestricted

Contents

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

3 of 56

1 Introduction

The aim of the project "Socio-Economic Optimal Pricing of Railroad Infrastructure" (SOPJI) was to develop a model and a prototype software implementation supporting the process of granting access to railway infrastructure. The project was financed by the Research Council of Norway under grant no. 200575 and ran over four years, from 2010 until 2013. The model has been developed with support from the infrastructure manager Jernbaneverket, a cargo train operator, Cargolink AS, and a passenger train operator, Flytoget AS. The passenger train operator NSB, operating the majority of the passenger trains on the Norwegian network, did not participate in the SOPJI project. However, due to this TOC's predominance, it appeared necessary to take their operations into account in some way when developing the model.

The approach shall contribute to a more efficient utilization of existing and a clear identification of needs for future railway infrastructure. The general idea is that the infrastructure manager (IM), in addition to approving route requests, can also affect demand for access to the infrastructure through a pricing mechanism called scarcity fees. Such a mechanism is intended to be separate from existing infrastructure usage charges. The tariff can be positive (charge) or negative (subsidy) and depends on the time period and the particular network element. It is assumed that the railroad operators or train operating companies (TOCs) using the infrastructure adapt their route requests to the tariffs according to their respective objectives: If it becomes too expensive to operate a train during times with high scarcity fees, the TOC may consider moving this train's operations to a lesser utilized slot. This forms a bilevel structure with the IM as leader and the TOCs as followers, each solving their own optimization problem. In theory, each TOC takes into account only the IM's price schedule but not the decisions of the other TOCs, i.e., they lay their plans as if they were the only TOC on the network. While the TOCs pursue goals such as profit maximization, a typical objective for the IM is the maximization of social welfare.

We describe the process from the IM's point of view, and the TOCs' models reflect the (limited) information the IM has about their decision process and their markets. In other words, they cannot be very detailed. For example, the IM will not have information about the TOCs' rostering, which trains shall be operated using the same (specific) rolling stock, or about commodity flows. Furthermore, all parameter values must be based on information available publicly or to the IM rather than information "internal" to the TOCs.

A crucial driver for the TOCs' decisions is, obviously, their respective customer demand, i.e., the market side (M). Indeed, the complete problem to be studied would consist of three decision levels: the IM, the TOCs, and their market(s). Clearly, including all in one model would make it too complex to be solved in a meaningful way. Hence, we employ a decomposition into two distinct—but connected—sub-problems, the IM–TOC relation, which is the main topic of this report, and the TOC–M relation outlined in appendix A.

The IM–TOC model implementation uses a crude description of the market through maximal achievable revenue and customer willingness to pay for each train (specified either per whole train, per car, or per capacity unit), modulated by multiplier profiles. These profiles reflect the change of the characteristics depending on departure and arrival times. For example, the transport capacity of a passenger train running outside rush hours will not be fully utilized.

If comprehensive information about the TOCs' customers' behaviour is available to the IM, a more detailed TOC–M model may be employed. Appendix A outlines such an approach and how this model may interact with the main IM–TOC model. It takes into account customer segments with different preferences and alternative trains or transportation modes. Hence, it enables to evaluate single trains in the light of the whole schedule and to see substitution or complementarity effects of other transportation options.

Note that these different approaches have different needs for input parameter values. The first approach requires statistics over transported passengers and cargo on the single routes and their variation over time. In the second approach, the transport amount is calculated, but this requires good information about the market heterogeneity, i.e., the single customer segments and their preferences, their willingness (probabilities) to substitute first and alternative choices, and their valuation of these choices. In other words, the challenge of finding good parameter values describing the market side has moved. Depending on the availability and quality of these data, the IM–TOC model may therefore be run with or without the TOC–M model.

Figure 1 illustrates the complete three-level structure and the scopes of both models, where the blue box indicates the IM–TOC model and the red box the TOC–M model. Note that this figure contains two passenger TOCs, "PTOC1" and "PTOC2", providing services to a common market "PMarket". The cargo TOC ("CTOC") caters to a cargo market "CMarket". Obviously, a full study of the three-level structure entailing an optimization of the train schedules with respect to the markets' response is beyond the scope of the SOPJI project.
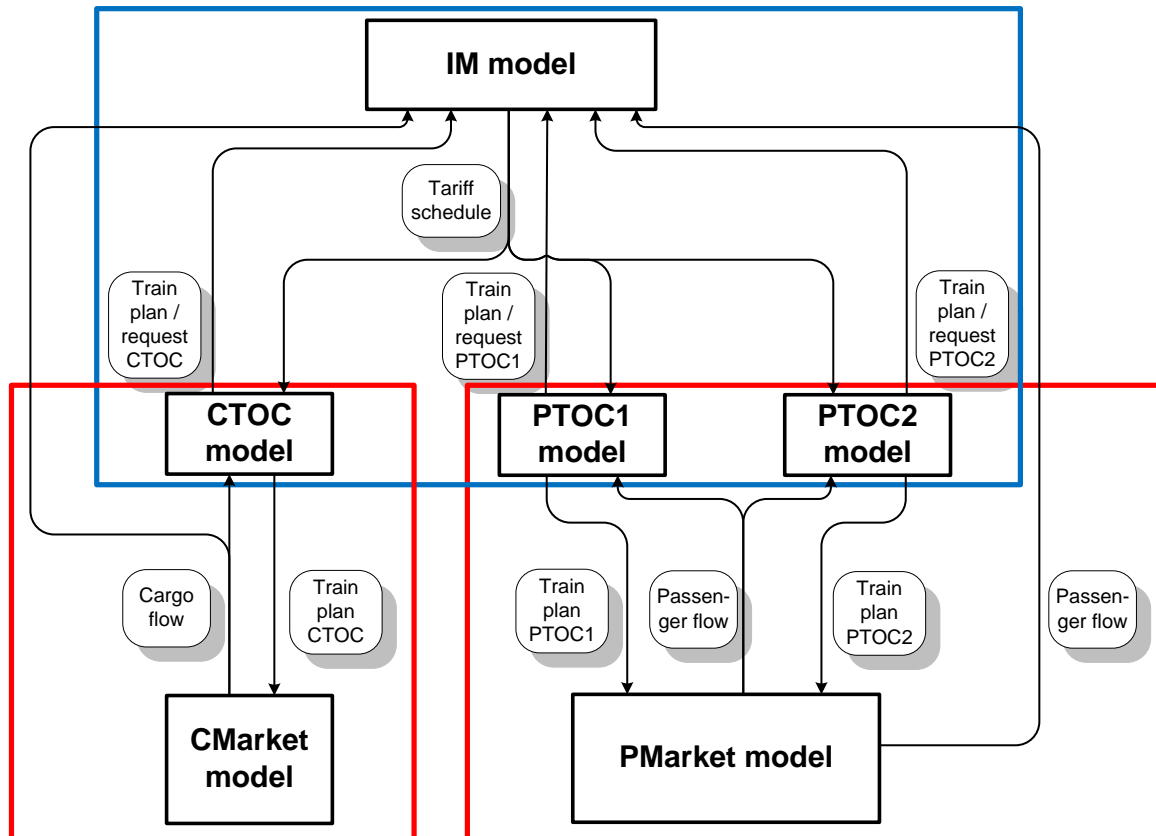
Figure 1: Three-level structure with two passenger TOCs and one cargo TOC; decomposition into IM–TOC and TOC–M sub-problems. The model presented in this report focusses on the IM–TOC relation.

This report focusses on providing a complete mathematical model formulation of the IM–TOC model and its implementation as a software prototype. Some background information including a preliminary qualitative model description can be found in Lium and Werner [2012][1]. Lium et al. [2014] discuss socio-economic aspects of the suggested scarcity pricing approach in the light of Pigovian taxes.

Observe that the IM pursues two possibly opposing goals, minimizing over-utilization (or removing capacity conflicts) and maximizing socio-economic utility. While we present a complete mathematical problem description taking into account both goals, the prototype implementation focusses only on the former goal which has been considered most important. It is, however, possible to evaluate the socio-economic utility of the resulting train schedules in a post-processing step.

An alternative implementation of the model and solution method which can accommodate both goals is discussed in Kaut et al. [2014]. This approach requires a different modelling of the train movements. Moreover, it uses a different solution algorithm, taking the points of view of the IM and the TOCs, respectively, and finding scarcity fees which lead to a compromise between these two extremes.

The remainder of this introductory section discusses some basic concepts and definitions used in our model formulation. Section 2 describes some train flow modelling issues necessary for describing the TOCs' operations

---

[1]The mathematical model in the present report has been updated with new information and knowledge. It is, therefore, not a one-to-one translation from this description.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

5 of 56

before section 3 presents the infrastructure manager's decision problem to derive an optimal schedule of charges. Then, sections 4, 5, and 6 formulate the IM's perception of the response problems of a cargo TOC, a small passenger TOC, and a large passenger TOC, respectively. Sections 7 and 8 are concerned with thoughts on solution approaches and a prototype implementation of the model and solution procedure, respectively. Section 9 describes a test case and presents results illustrating a run of the prototype. Appendix A outlines an idea for a more comprehensive market modelling and appendix B presents a few more aspects of interest—which, however, are not implemented in the prototype presented here. Appendix C lists the notation used in the mathematical model and appendix D describes relevant data sources and the structure of the input database for the prototype implementation.

1.1 Uncertainty

Generally, a significant degree of uncertainty and imperfect information can be observed, for example:

- The TOCs' schedules are built based on previous schedules and customer demand statistics and projections, i.e. on estimates about the development of the transportation demand. It is obvious that such projections cannot be very precise, especially for the cargo TOC (due to, e.g., industry dynamics and, hence, uncertainty about future contracts and prices or the behaviour of competitors).

- Price, cost, and (dis)utility parameters used in the model are typically average values determined using statistics and surveys. First, using average values instead of differentiated ones is an—often necessary—approximation of the situation observed in practice. Second, the quality of such average values depends very much on the quality of the underlying statistics and survey models, their level of aggregation etc. and may not give a sufficiently correct picture of reality.

- The TOCs' decision models used here are approximations of their "true" decision processes as the latter are typically much too complex to be modelled as mathematically tractable problems. They may include many business-modelling and other aspects of more operational or strategic character such as rolling-stock or personnel rostering, investment decisions, contract negotiations, modelling of strategic behaviour toward important (potential) customers, or fare modelling. To a certain degree, this holds also for the IM's decision model, especially with respect to socio-economic aspects.
  Moreover, the model discussed here takes the IM's point of view who has only a limited insight in the TOCs' business processes. That means that—even if there would be sufficient data and insight available to model the TOCs' problems per se—the model can only rely on what is (publicly) available to the IM. Hence, there is a certain degree of uncertainty about the TOCs' decision processes and their responses to given price plans.

- The IM–TOC model assumes that the TOCs find their schedules independently of each other and do not act strategically. This is not necessarily true in reality.

Some of this uncertainty can be dealt with using stochastic programming concepts such as scenarios while other aspects are more difficult to address and may need insights from agency theory (moral hazard). However, the focus of our model is on the most important relationships in the IM–TOC constellation. Although case results and effects will quite likely change, the basic functionality of the pricing mechanism will not be affected by a more sophisticated modelling of the uncertainty. For an initial formulation, we refrain therefore from a comprehensive treatment of this aspect. Appendix A discusses briefly how a dedicated TOC–M model may address uncertainty about customer demand behaviour.

1.2 Infrastructure network

Generally, a railroad infrastructure network can be understood as a set of nodes and a set of connecting edges. One may distinguish different kinds of nodes, for example, passenger stations, cargo terminals, or parking areas.

In the following, we employ a more detailed concept, considering *sections* with a length λ_{ij}^B as the edges $\mathcal{A}$, see **??**. If required (and practicable) we may consider the smallest unit of network infrastructure, *block sections*, with a capacity of at most one train at any time. Typically, however, a coarser graduation will be applied which, by and large, reflects the network structure as it is used in practice. Exceptions may be larger stations like Oslo S or Trondheim S which are considered as one (or a few) section(s) in reality but the model may require several sections (see also the comment in section 8). *Links* denote the arcs of the network, that is, they are directed sections. In this context, *network nodes* $\mathcal{N}$ denote points where at least two sections meet, e.g., signal points. All nodes are indexed such that each section $(i, j) \in \mathcal{A}$ is uniquely defined through a pair of nodes $i, j \in \mathcal{N}$. Note that the adjacency matrix of these nodes is typically very sparse, especially for a structure such as the Norwegian railroad network.

We consider terminals and stations as special sections, $\mathcal{A}^{Stop} \subseteq \mathcal{A}$ with a somewhat different capacity than "simple" sections along a line. For example, a section with a capacity of four trains at a time may actually denote a station with four tracks. We do not take into account that stations or terminals may be more or less efficient for different trains to move around. This is a rather operations oriented issue which may be included later if required. Also, the presented model does not explicitly include parking areas, but they may be modelled as special types of stations or terminals.

The costs of some external effects of train traffic on a local level are given per train-kilometre but differentiated only with respect to the type of area where the section is located. The set $\mathcal{F}$ denotes all such types, e.g. "city", "other urban regions", and "rural regions", and f(ij) is the type of section (i, j).

1.3 Routes, services, slots, and trains

Routes $r \in \mathcal{R}$ are unique paths through the network (i.e., a complete sequence of (block) sections $\mathcal{B}_r \in \mathcal{N} \times \mathcal{N}$). In other words, for each route, all nodes to be passed and their sequence are given. A *service ('tilbud')* $s \in \mathcal{S}$ denotes a group of trains a TOC operates on a given route such as suburban trains on a periodic schedule (e.g., a train each 15 minutes). We assume that these trains share some common properties such as the rolling stock used (e.g., type and number of engines and cars) or their stopping pattern: all trains in a service s are assumed to stop at the same stations $\mathcal{A}_s^{Stop} \subseteq \mathcal{A}^{Stop}$ for the same amount of time. For example, passenger train services may be "Airport express train Drammen–Gardermoen", "local train Moss–Skøyen morning rush", or "daytime train Trondheim–Oslo" while cargo services may be "car-transport train Drammen–Trondheim", "short container train Alnabru–Bergen, mid-week". A *train* $\xi \in \Xi$ denotes one occurrence of a service, together with actual time information about its travel along the route.

Consequently, several services can be defined on a given route, but a service cannot travel several routes. (In this case, a new route should be defined specifying this services' geographical path.) Likewise, several trains may operate on the same service, but one train cannot belong to several services. In other words, for each train $\xi \in \Xi$, its service is given by s(ξ) and for each service s, its route is specified by r(s) and the rolling stock configuration v(s).

The rolling stock configuration v(s) used for a service s, in turn, determines properties such as a train's standard speed[2] $\gamma_{ij}^{v(s)}$ on a section (i, j), the number λ_s^T of cars or car sets in a train, or the length $\lambda_{v(s)}^{RS}$ of these cars or car sets. Consequently, a train used in service s has a length of $\lambda_s^T \cdot \lambda_{v(s)}^{RS}$ which, in turn, determines its length class l(s).

We denote the first and the last sections on a route r by (i_r^{RO}, j_r^{RO}) and (i_r^{RD}, j_r^{RD}), respectively. These sections will, typically, be terminals or stations. Some services may also have given time windows when trains may depart, $[\sigma_s^{DE}, \sigma_s^{DL}]$, at the origin and arrive, $[\sigma_s^{AE}, \sigma_s^{AL}]$, at the destination terminal or station. Each train in a service s may have a maximum duration D_s on its travel time which includes driving, scheduled stopping or (off)loading, and waiting times.

A *slot* represents a combination of a section and a time interval. Hence, a train describes a sequence of slots. While the slots are pre-defined by the IM, the trains are defined by the TOCs. In other words, the IM's decision

---

[2]Given as the minimum of the rolling stock's standard speed and maximum allowed speed on this section.

problem is primarily focused on the pricing of the slots. The TOCs' decision problems, on the other hand, are concerned with the scheduling of trains, that is, whether and at which times to run the trains of their services. We introduce disjoint subsets Ξ^C, Ξ^{SP}, and Ξ^{LP} for the trains operated by the cargo TOC, the small passenger TOC, and the large passenger TOC, respectively, such that $\Xi^C \cup \Xi^{SP} \cup \Xi^{LP} = \Xi$.

Although the TOCs' slot requests do not specify which actual rolling stock (e.g., "engine no. 26 with cars no. 576, 577, and 578") will operate this train and the model is not concerned with operational aspects such as rostering, it may be useful to define additional constraints relating some trains to each other which the TOC intends to serve by the same rolling stock. A simple example is direction balance requiring that the number of trains travelling on a route over the optimization horizon must be the same for each direction of this route.

Another example is that adjusting a train such that it arrives later at the destination should not upset the rostering for subsequent trains from that destination. Approaches to accommodate this without becoming too detailed are described in the appendix, section B.2.

1.4   Time modelling

The model combines continuous and discrete time representations. For modelling train movements on the network, we employ a continuous time formulation, see section 2. It enables a precise and realistic description of departure, waiting and arrival times without the need for artificial slack variables.

Sometimes, however, it is necessary to operate with a discrete time representation, e.g., to define slots. For this purpose we use consecutively numbered time periods with a given length of Δ time units and a time index $p \in \mathcal{P} = \{0, 1, ..., P\}$. Hence, the length of the optimization horizon is $P \cdot \Delta$ time units. For example, with a time period length Δ of half an hour and an optimization horizon of a week, we would have $P = 336$.

Looping.   Obviously, trains are operated continuously and may extend beyond the end of the optimization horizon. Hence, the time counting must be performed modulo P for all such trains, such that there are trains which, seemingly, arrive earlier than they depart. This must be taken into account for the train flow modelling in section 2.

1.5   Model structure

The IM controls the TOCs' usage of the infrastructure by way of access charges and subsidies (*scarcity fees*): A charge is levied on over-utilized slots (where demand exceeds track capacity) while subsidies may be given for using little-utilized slots. Ultimately, this results in a tariff schedule over all slots in the considered network and time horizon. The TOCs respond to this schedule by adapting (also cancelling) their requests, thus incorporating the tariffs into their utility functions.

This set-up describes a bilevel decision problem (BLP) structure with the IM as the upper-level decision maker or leader and the TOCs as the lower-level decision makers or followers. The scarcity fee is a decision variable for the IM, and a parameter for the TOCs. Bilevel optimization problems are inherently difficult to solve even in their simplest form. Properties such as integrality conditions on the TOCs' decisions make our problem class even more complex. Hence, we do not attempt to devise a direct solution approach but follow an iterative procedure as described in section 7.

We assume that each TOC plans the schedule of their trains in a selfish manner, independently of the other TOCs. That is, each TOC assumes there are no other companies operating on the infrastructure which may impede this TOC's operations. As explained in subsection 1.3, a TOC has information about all potential services (routes and rolling stock types), together with time windows for their operation. This information may come from previous schedules or the TOC's assessments. The TOC's decisions concern the timing of these services along the whole route—which will then form a train. In other words, the TOCs are concerned with the scheduling of each of their trains over its complete route, i.e. over several sections. In contrast, the IM deals with scheduling (possibly) several trains on each section or at each time slot. Hence, the TOCs' decision problems can be considered to be train oriented while the IM's problem is section or network oriented.

As the tariffs also imply subsidies, their effect should be a redistribution of traffic over time rather than just discouraging usage of the most popular slots, possibly even a migration from other transport modes. For a more in-depth discussion of such effects we refer to Lium et al. [2014].

The IM is not concerned with the actual operation of the trains (i.e., determining waiting times etc. for a train along a route such that stopping patterns or time windows are observed). Therefore, it appears sufficient to use a coarser granularity. For example, while the TOC models use a granularity of one minute, the IM model may work with a granularity of 30 or 60 minutes. The scarcity charges in a slot express, therefore, how much it costs the TOC to use this slot for one minute but they will vary only over each half or whole hour.

This set-up accommodates the primary goal of the model—efficient (and socio-economically optimal) utilization of the infrastructure—but does not prevent conflicts occurring due to two trains using a block section at the same time or driving in each other's headway. To identify such conflicts, the model may be run a second time (in the sense of a scheduling tool) with fixed price schedules and the same fine time granularity as the TOCs' models.

Note that the model is memoryless, i.e. all route requests are treated equally without observing "grandfather's rights" or the like. This implies also that revenue from charges cannot be set aside for infrastructure investments (which is permitted according to the European Parliament [2001] directive) but must be balanced by pay-outs in the form of subsidies. This balancing requirement may, though, be easily adapted to the case of user-financed lines as outlined in appendix B.1. Another consequence of using a memory-less model is that each run of the model may create completely different plans. This effect can be mitigated by adding a term to the objective function penalizing deviations from some reference train schedules.

1.6 Sets, parameters and variables

The scarcity fee z_{ijp}^v is levied by the IM on the TOCs as a charge for using a section (i, j) at a time period p with a train of rolling stock type v if this section is expected to be heavily utilized at this time. Vice versa, it can be granted as a subsidy ($z_{ijp}^v < 0$) if this section is under-utilized at a time period p.

In the *IM's* decision problem, several parameters express external effects of train traffic in monetary terms, differentiated with respect to sections and train types v: The parameter $c_{f(ij)v}^{LocalEm}$ denotes the costs of local emission effects per train-kilometre, $c_v^{GlobalEm}$ the costs of global emission effects, $c_{f(ij)v}^{Noise}$ the costs of local noise, and c_v^{Acc} the aggregated costs of accidents, respectively. For the costs of local emissions and noise, Jernbaneverket [2011] classifies the sections with respect to their location $f(ij) \in \mathcal{F}$. Infrastructure maintenance costs c_v^{Maint} are given per train-kilometre and differentiated with respect to rolling stock types. We ignore their dependency on the infrastructure elements' age. A dependency on the type of the infrastructure element (bridges, tunnels, etc.)—if such values are available—may be taken into account by including the section indices i, j.

The *TOCs'* main decisions concern the operation of their trains, in particular the timing: the binary variable y_{ijp}^ξ indicates whether, at time period p, (part of) train ξ is in section (i, j). In order to find these decisions, auxiliary variables are introduced in section 2. The standard speed γ_{ij}^v of rolling stock type v on section (i, j) is fixed, and the decision variables $w_{j\xi}$ denote train ξ's waiting time on that section. This way, travel times can be adjusted during the decision process. Both passenger and cargo trains may have required stops along their route and we assume that all trains in a service have the same stopping pattern, i.e., stop at the same stations or terminals (i, j) for the same time λ_{js}^{Stop}.

The following costs are related to operating a train of service s consisting of rolling stock type $v(s)$: Fixed operational costs c_s^{TrFix} may arise whenever the train is operated, independent of route length or travel time. Time-dependent operational costs $c_{v(s)}^{RST}$ are calculated per hour operating time of one unit of rolling stock type $v(s)$, and c_s^{TrHour} gives the time-based costs of operating a single train of service s (per hour), resulting in the time-dependent operating cost c_s^{TrMin} of a train of service s (per minute operation time). The former costs refer to the rolling stock while the latter comprise, e.g., train driver wages. Distance-dependent operational costs of one unit of rolling stock type $v(s)$ (per km) are given by $c_{v(s)}^{RST}$. Infrastructure usage charges $c_{v(s)\,r(s)}^{Use}$ may be differentiated according to the type of rolling stock used, see, for example, Jernbaneverket [2012, Ch. 6]. A further differentiation with respect to the train's route $r(s)$ allows to include the additional charges for using the

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

9 of 56

Gardermobanen (GMB) line. Costs c_s^{Wait} for waiting represent a penalty for (or the disutility of) a travel time of a train of service s longer than necessary.

Passenger TOCs may face a cost c_j^{Stop} for using a station $(i, j) \in \mathcal{A}_s^{Stop}$. At present, this concerns only prioritized services at the stations on GMB: Oslo S Airport Express Train Terminal, Lillestrøm, and Gardermoen [Jernbaneverket, 2012, pp. 105/106]. Note that Oslo S Airport Express Train Terminal is reserved for Flytoget and no other TOCs are allowed to use it [Jernbaneverket, 2012, footnote 44 on p.109]. Likewise, for cargo trains, the parameter c_j^{Stop} denotes the average cost of freight handling (loading or offloading) at a terminal $(i, j) \in \mathcal{A}_s^{Stop}$. As stations and terminals are mutually exclusive, a differentiation of c_j^{Stop} with respect to the train service is not necessary. For cargo trains, one may, however, differentiate c_j^{Stop} with respect to train type if required, see also comment in Appendix D.

For the *cargo TOC*, the parameter c_{jv}^{Park} denotes a penalty for a train of type v(s) occupying terminal (i, j) for too long. L_{js}^{Park} denotes the time limit from which on time spent at the terminal is considered excess time and $w_{j\xi}^{Park}$ this excess time spent at the terminal by train ξ.

For *passenger TOCs*, the set $\mathcal{U}$ contains all commissions of publicly purchased train services under the traffic contract, "Trafikkavtalen", with the Ministry of Transport (MoT). Each commission is specified through the route ρ_u and a (minimum) number ω_u of trains to serve a station i_u during a time interval $[\tau_u^E, \tau_u^L]$. The parameter $Comp_u$ gives the compensation paid by the MoT for commission u.

Appendix D outlines data sources for these parameters with focus on the Norwegian railway network.

Multiplier profile curves

The *value* (e.g., realized revenue or customer utility) of a train operating at a certain time appears difficult to estimate. In addition to the customers' willingness to pay for the transport, also their valuation of the train's duration plays an important role here.

For cargo trains, the report Halse and Killi [2012] discusses costs of delays for railway transport and may give indications for the latter aspect. However, both factors may vary quite significantly over the day and among different customers. Similar challenges arise when considering passenger trains. To our knowledge, research about variations in customers' willingness to pay, utility, or price sensitivity *throughout the day* is still lacking. However, exactly these variations are the motivation to change a train's schedule, balancing between varying consumer and producer surpluses and varying costs.

We tackle this challenge by using profile curves to describe varying user preferences for certain departure or arrival times of a train or service. This facilitates an approximation of user demand variation over the day without having to resort to a detailed representation of the TOCs' customers' behaviour. They are a series of multipliers between 0 and 1 expressing how much of the transportation potential of a train in a service s can be achieved if it departs and / or arrives at a certain time period p. For example, an Airport Express Train service towards Oslo Gardermoen may have a arrival profile with two tops, one in the morning and one in the afternoon as illustrated in Figure 2. A local passenger train in the Oslo area may have a profile for arrival times with a top around 8 o'clock in the morning for inbound services, see Table 1, while a departure profile for outbound services would have a top around 16 o'clock. A cargo train may have both a departure profile (topping around, say, 21 o'clock) and an arrival profile, topping at 4 o'clock. If there is no or a flat profile, the demand for transportation by a train of that service is the same no matter when the train is operated.

Time period p	6	7	8	9	10	11	...	19	20	21	22
Multiplier $\mu_{s(\xi)}^A$	0.15	0.7	1	0.9	0.8	0.6	...	0.75	0.6	0.4	0.2

Table 1: Mock example of multiplier profile for arrival, Oslo-bound local train

PROJECT
102000998

REPORT NUMBER
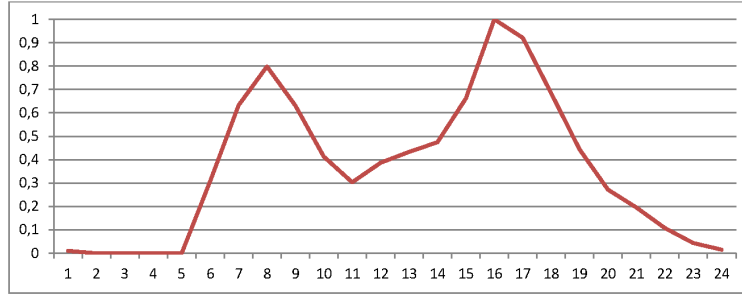SINTEF A25843

VERSION
1.0

10 of 56

Figure 2: Example of multiplier profile for arrival, Airport Express Train service toward Oslo Gardermoen

From the variables describing the train flow in section 2, one can derive binary variables

$$
\delta_{\xi p}^{D} = \begin{cases} 1, & \text{train } \xi \text{ left its departure node in time period } p \\ 0, & \text{else} \end{cases}
$$

$$
\delta_{\xi p}^{A} = \begin{cases} 1, & \text{train } \xi \text{ arrived at its final node in time period } p \\ 0, & \text{else} \end{cases}.
$$

They help to estimate the percentage of the maximal transportation demand potential of a train of this service which can be realised:

$$
\nu_{\xi}^{D} = \sum_{p \in \mathcal{P}} \mu_{s(\xi)p}^{D} \delta_{\xi p}^{D} \tag{1a}
$$

$$
\nu_{\xi}^{A} = \sum_{p \in \mathcal{P}} \mu_{s(\xi)p}^{A} \delta_{\xi p}^{A} \tag{1b}
$$

Then, in order to express the socio-economic utility (see subsection 3.2) or the revenue (see sections 4.1, 5.1, and 6.2) from operations of a train ξ at the given time, a multiplier ν_{ξ} is derived,

$$
\nu_{\xi} = \begin{cases} \nu_{\xi}^{D} & \text{if only departure profile } \mu_{s(\xi)p}^{D} \text{ exists} \\ \nu_{\xi}^{A} & \text{if only arrival profile } \mu_{s(\xi)p}^{A} \text{ exists} \\ (\nu_{\xi}^{D} + \nu_{\xi}^{A})/2 & \text{if both profiles } \mu_{s(\xi)p}^{D} \text{ and } \mu_{s(\xi)p}^{A} \text{ exist} \\ 1 & \text{if no profile exists.} \end{cases} \tag{2}
$$

2   Modeling the flow of trains

The following constraints apply to each TOC separately and are part of their respective decision problem, i.e. they are formulated for the trains $\xi \in \Xi^{C}$, $\xi \in \Xi^{SP}$, and $\xi \in \Xi^{LP}$.

It is paramount to ensure that the train movement is consecutive: trains do not appear or disappear along the way, and sections must be traversed in the correct sequence. Note that, as a train moves along, it may occupy more than one section at a given time. Hence, the constraints also model that a section may not be open to other trains during this time.

For defining slots with prices varying over time, is is convenient to use a discrete time scale. For modelling the flow of trains, however, this approach would mean to indicate whether the head of a train passed a node at a given time period. Obviously, this is impractical since it implies that all travel times between consecutive nodes must be rounded up to multiples of Δ, e.g. by using slack variables. Doing this for every section on the

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

11 of 56

train's route, the error may accumulate significantly. Hence, we employ a continuous time-keeping formulation for modelling the movement of trains: For each node i of the network, $t_{\xi i}^H$ denotes the time the head of train ξ passes node i. This gives, for each "normal" section $(i, j) \in \mathcal{B}_{\mathrm{r(s(\xi))}}$,

$$t_{\xi j}^H = t_{\xi i}^H + \gamma_{ij}^{\mathrm{v}(s)} + w_{j\xi} \,. \tag{3}$$

For some specific sections, this expression is amended, see (9).

In order to find out which sections are reserved for a train, we keep track of where its head and tail are. The variables $\tilde{h}_{\xi ip}^H$ indicate whether the head of train ξ *passes* node i at time period p while the variables $h_{\xi ip}^H$ state whether it *has passed* node i by time period p (that is, at any time up to p). The former is computed from the latter as

$$\tilde{h}_{\xi ip}^H = \begin{cases} h_{\xi ip}^H & \text{if } p = 0 \\ h_{\xi ip}^H - h_{\xi i,p-1}^H & \text{otherwise.} \end{cases}$$

2.1 Synchronizing continuous and discrete times

With the introduction of the above variables, we now have two sets of time-accounting variables in the model: the continuous time $t_{\xi j}^H$ and discrete time $\tilde{h}_{\xi ip}^H$. They are kept synchronized by introducing additional constraints. These constraints depend on whether a train crosses the horizon P or not, since the discrete time 'loops around' and starts from zero, while the continuous time continues past $P\Delta$.

We start with the simplest situation where the time windows are such that we know upfront that the train will not cross the horizon. In such a case, we add the following two constraints:

$$P\Delta \, h_{\xi ip}^H \geq p\Delta r_\xi - t_{\xi i}^H \,, \tag{4a}$$
$$P\Delta \, h_{\xi ip}^H \leq (p + P)\Delta - t_{\xi i}^H \,, \tag{4b}$$

where r_ξ is a binary variable controlling whether train ξ runs or not. For running trains, i.e. ξ with $r_\xi = 1$, constraint (4a) guarantees that, if the head of the train passes signal i before time period p, i.e. if $t_{\xi i}^H < p\Delta$, then $h_{\xi ip}^H = 1$. Constraint (4b) then secures the reverse relation, ensuring that $h_{\xi ip}^H = 1$ implies $t_{\xi i}^H \leq p\Delta$.

If the train is not run, the r_ξ in (4a) help to disconnect the continuous and discrete time-keeping by allowing all the $h_{\xi ip}^H$ variables to be zero regardless of the values of $t_{\xi i}^H$. Hence, a train not running is modelled as a train that does not use any resources; the continuous time-keeping variables are still being set, but they are disconnected from the rest of the model.

The situation is analogous for trains that are planned to cross the $p = P$ boundary. In fact, constraints (4) are still valid for the periods before the boundary. For the periods after we cross the boundary and start again from $p = 0$, we have to subtract $P\Delta$ from the continuous time, given that the train is run at all:

$$P\Delta \, h_{\xi ip}^H \geq p\Delta r_\xi - (t_{\xi i}^H - P\Delta r_\xi) \qquad = (p + P)\Delta r_\xi - t_{\xi i}^H \,, \tag{5a}$$
$$P\Delta \, h_{\xi ip}^H \leq (p + P)\Delta - (t_{\xi i}^H - P\Delta r_\xi) = (p + P + Pr_\xi)\Delta - t_{\xi i}^H \,, \tag{5b}$$

Finally, we consider the situation where we do not know in advance whether a train will 'loop in time', that is, cross the $p = P$ boundary. We illustrate this on a 'free' train that does not have any time window at all. We introduce the binary variable q_ξ that is equal to one if ξ loops over in time. Obviously, we have $q_\xi \leq r_\xi$. Moreover, the *duration* of a train of service s must be limited by some $D_s \leq P/2$. This gives, for all nodes $i \in \mathcal{N}$,

$$t_{\xi i}^H \leq (P - 1 + D_{\mathrm{s}(\xi)} q_\xi)\Delta \,. \tag{6}$$

The limit on the duration is necessary for the following reason: for a given period p, we need the 'looping-adjusted' constraints (5) only if $q_\xi = 1$ and $p < D_s$; otherwise, we know there is no looping at that period and

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

12 of 56

we can use (4). It follows that for $p \geq D_s$, we can always use (5), irrespective of the value of q_ξ, while for $p < D_s$ we need to adjust the constraints in the following way:

$$P\Delta \, h_{\xi ip}^H \geq p\Delta r_\xi - (t_{\xi i}^H - P\Delta r_\xi q_\xi) \qquad = (pr_\xi + Pq_\xi)\Delta - t_{\xi i}^H \,, \tag{7a}$$

$$P\Delta \, h_{\xi ip}^H \leq (p + P)\Delta - (t_{\xi i}^H - P\Delta r_\xi q_\xi) = (p + P + Pq_\xi)\Delta - t_{\xi i}^H \,, \tag{7b}$$

where the equalities use the fact that $r_\xi q_\xi = q_\xi$, since $q_\xi \leq r_\xi$.

2.2 Connecting the tail of the train

The next step is to compute the time when the *tail* of the train passes a given node i of the network. This can be done exactly, assuming that trains traverse each section at their standard speed and wait only at the end of the sections, i.e. with the head just before j. Now, since we know the lengths λ_{ij}^B of the sections and $\lambda_s^T \cdot \lambda_{v(s)}^{RS}$ of a train ξ in service s, we can easily find a section (k, l) where the head is in the moment the tail of the train passes node i (i.e. enters section (i, j)). Furthermore, we know how far from k the head of the train is—let us denote this by $\lambda_{v(s)ik}^{TB}$. With this notation, the (continuous) time when the tail passes i is given by

$$t_{\xi i}^T = t_{\xi k}^H + \frac{\lambda_{v(s)ik}^{TB}}{\lambda_{kl}^B} \gamma_{kl}^{v(s)} \,.$$

We can then define new binary variables $h_{\xi ip}^T$, denoting whether the *tail* of the train has passed node i by time period p. These have to be connected to $t_{\xi i}^T$, in the same way as for the train head in constraints (4)–(7), replacing the 'H' superscript by 'T'.

This makes it then very easy to determine the variables y. y_{ijp}^ξ is equal to one if the head of the train has passed i at period p, but the tail has not passed j yet:

$$y_{ijp}^\xi = h_{\xi ip}^H - h_{\xi jp}^T \,. \tag{8}$$

Note that for each section (i, j) and time p, only six out of the sixteen possible combinations of $h_{\xi ip}^H$, $h_{\xi jp}^H$, $h_{\xi ip}^T$, and $h_{\xi jp}^T$ are valid, because of the 'logical' constraints that the train must pass node i before node j and the head of the train must pass before its tail. The feasible combinations are:

	a	b	c	d	e	f
$h_{\xi ip}^H$	0	1	1	1	1	1
$h_{\xi jp}^H$	0	0	0	1	1	1
$h_{\xi ip}^T$	0	0	1	0	1	1
$h_{\xi jp}^T$	0	0	0	0	0	1
y_{ijp}^ξ	0	1	1	1	1	0

The columns represent, respectively, where the train is at period p: a) before node i; b) passing node i, with the head inside (i, j); c) the whole train inside (i, j); d) train "outside of (i, j)", i.e. the head has left (i, j), while the tail still has not passed i; e) leaving (i, j), i.e. the head has left the section, but the tail is still inside; and, finally, f) the whole train has left the section.

This formulation makes it also easy to consider train specific headways; they can be added either before or after a train, e.g. by adding to the train lengths.

2.3 Special sections

For sections $(i, j) \in \mathcal{A}_s^{Stop}$ representing stopping places (terminals or stations) of a train ξ in a service s, the stopping time λ_{js}^{Stop} must be included in equation (3),

$$t_{\xi j}^H = t_{\xi i}^H + \gamma_{ij}^{v(s)} + w_{j\xi} + \lambda_{js}^{Stop} \tag{9}$$

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

13 of 56

For cargo trains, λ_{js}^{Stop} represents the average time necessary to (off)load cargo while, for passenger trains, it denotes scheduled stops for passengers to board or alight.

The duration of a train ξ in service $s = s(\xi)$ comprises driving, waiting, and planned stopping at sections $\mathcal{A}_s^{Stop}$ and is limited by D_s time units,

$$\sum_{(i,j) \in \mathcal{B}_{\mathrm{r}(s)}} \left(\gamma_{ij}^{\mathrm{v}(s)} + w_{j\xi} \right) + \sum_{(i,j) \in \mathcal{A}_s^{Stop}} \lambda_{js}^{Stop} \leq D_s \tag{10}$$

Reformulating, the limit D_s gives also a limit on the total waiting time of that train,

$$\sum_{(i,j) \in \mathcal{B}_{\mathrm{r}(s)}} w_{j\xi} \leq D_s - \sum_{(i,j) \in \mathcal{B}_{\mathrm{r}(s)}} \gamma_{ij}^{\mathrm{v}(s)} - \sum_{(i,j) \in \mathcal{A}_s^{Stop}} \lambda_{js}^{Stop} \tag{11}$$

On some sections (i,j), additional (explicit) waiting time limits L_{js}^{Wait} may apply for trains of a service s,

$$w_{j\xi} \leq L_{j\,\mathrm{s}(\xi)}^{Wait} \tag{12}$$

2.4 Further aspects

Start and end conditions

Some services may be limited to operate within some time windows $[\sigma_s^{DE}, \sigma_s^{DL}]$ or $[\sigma_s^{AE}, \sigma_s^{AL}]$ for departure or arrival, respectively. This means that all trains in this service s must pass the signal out of the departure terminal / station $(i_{\mathrm{r}(s)}^{RO}, j_{\mathrm{r}(s)}^{RO})$ or the signal into the destination terminal / station $(i_{\mathrm{r}(s)}^{RD}, j_{\mathrm{r}(s)}^{RD})$ not earlier or later than a given point of time:

$$\sigma_s^{DE} \leq t_{\xi j_{\mathrm{r}(s)}^{RO}}^{H} \leq \sigma_s^{DL} \tag{13a}$$

$$\sigma_s^{AE} \leq t_{\xi i_{\mathrm{r}(s)}^{RD}}^{T} \leq \sigma_s^{AL} \tag{13b}$$

Not all of the above time limits may apply to each service. It is, for example, also be conceivable that a train is only subject to a latest arrival time limit. Note that the limits (13), together with the standard driving times on the sections $(i,j) \in \mathcal{B}_{\mathrm{r}(s)}$ and the scheduled stopping times at stations or terminals help to define time windows for a train's earliest and latest passing of the nodes along its route. This, in turn, may help to decrease problem size and, hence, computation time.

Alternative services and cancellations

It may be required that, from a given set $\mathcal{S}_g^{Alt}$ of alternative services, at most one may operate: Similar to the variable r_ξ, one may define a binary variable $\bar{r}_s$ stating whether a service s is operated, i.e., if at least one train of the service runs,

$$\sum_{s \in \mathcal{S}_g^{Alt}} \bar{r}_s \leq 1$$

This means also that, if the service runs, at least one train in this service must operate. Vice versa, if no train is operated, the service is considered as not running.

$$r_\xi \leq \bar{r}_s \qquad \forall \xi \in \Xi : \mathrm{s}(\xi) = s$$

$$\bar{r}_s \leq \sum_{\xi \in \Xi : \mathrm{s}(\xi) = s} r_\xi$$

This helps, for example, to find the most appropriate choices between different trains, services, or even routes (such as Dovre- or Rørosbanen between Oslo and Trondheim) or to test different rolling stock configurations for a service.

Obviously, this formulation may easily be adapted to allow selecting several services from a group or to require a certain number of services to be selected.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

14 of 56

3 The infrastructure manager's problem

In the context of our model, the main responsibility of the IM (Jernbaneverket) is to ensure a best possible infrastructure utilization. This implies to accommodate as much traffic as possible in such a way that infrastructure capacity is not exceeded. Since each of the TOCs operates independently, the IM must coordinate the decisions of *all* TOCs. This means a trade-off between preventing conflicts between trains and finding a train schedule with maximum socio-economic utility. As the IM cannot plan the TOCs' train movements, these goals can be achieved only indirectly through incentives such as the scarcity tariffs. These guide the TOCs' requests—and, in consequence, infrastructure utilization—for access to the network and may vary over the day. The tariffs are used to encourage TOCs to shift their train traffic in the case of conflicts. (Encouraging second-best choices for "displaced" traffic through subsidies rather than just discouraging usage at peak times through charges contributes to keeping traffic on the railroad network instead of migration to other modes.)

Revenue-neutrality requirements mean that, over a "reasonable" time period, the IM shall balance revenues from scarcity charges and surpluses from other activities and state funding with infrastructure expenditure [European Parliament, 2001, art. 6, 1.]. However, the IM's other activities and state funding are beyond the scope of our model. Neither can it consider infrastructure investment planning which often spans several years. Moreover, these elements are already in place and are assumed to conform to revenue-neutrality conditions while scarcity charges are not implemented in practice yet. Consequently, also the scarcity charges/subsidies alone should balance out—or any increased income should be used for increased infrastructure expenditure to remove bottlenecks. With a memory-less model, we cannot assume that revenue can be set aside to improve network capacity but we discuss user financing briefly in appendix B.1.

Note that, as its main purpose is to guard the capacity utilization, the IM's model can be run with a coarser granularity such that slots consist of periods of a length Δ (say, 30 or 60 minutes) while the scarcity charges specify how much it costs for a train to use one minute of this slot.

3.1 Socio-economics

As mentioned earlier, we assume that the TOCs act in a selfish manner, not taking into account the decisions of other TOCs and their consequences. The IM, however, needs to find tariffs such that they are to the best of the whole society, including all TOCs as well as their customers and other individuals. Socio-economic aspects must be taken into account at the IM level. We briefly discuss some of the most important aspects included in the model in some form or other.

Utility

The utility of a (set of) train(s) to society is difficult to quantify as it is a way to express the involved actors' preferences. In the context of our model, we define the utility of the train operations as the sum of consumer surplus, producer (TOC) surplus, and the IM's surplus or deficit, diminished by a quantification of external effects of this train traffic.

Consumer surplus can be determined as the consumers' willingness to pay diminished by what they actually paid. In our model, the TOCs' customers' willingness to pay is considered on an aggregated level, per train. This is explained by the fact that, for cargo trains, a customer may actually book a whole train or several cars. For passenger trains, we don't track single passenger journeys between stations along a train's route but use an average transportation demand for a train of a service. Hence, it appears appropriate to describe customers' willingness to pay for a train rather than for individual passengers or cargo units. For a train of a service s, we denote the maximum achievable willingness to pay by W_s^{max}. Note that this also may include customer preferences for stopping pattern, (approximate) travel time, comfort/rolling stock, or other quality attributes. Using the profiles and resulting multipliers ν_ξ described on page 10, we express a variation in customer preferences for trains of a service with respect to departure and arrival times. Consequently, the consumers' willingness to

pay for transport on a specific train ξ can be written as

$$W_\xi = W_s^{max} \nu_\xi r_\xi \tag{14}$$

where the binary variable r_ξ indicates whether train ξ is operated. As a measure for what the customers paid for transport, one may use the TOC's actual revenue π_ξ from operating this train at its departure and arrival times. Section 4.2 describes how to determine this revenue. Then, the consumer surplus from all train operations is

$$\sum_{\xi \in \Xi} W_\xi - \pi_\xi$$

We do not include consumers' dis-utility from cancelled trains. Not at least, this is due to a scarcity of data.

The *TOCs' surplus* is defined as the revenue from transport of passengers or cargo diminished by all costs of running the trains—in other words, their objective function terms (21), (27), and (30), respectively.

We determine the *IM's surplus* as the sum of scarcity tariffs and other infrastructure usage charges paid by the TOCs diminished by expenditures on infrastructure maintenance. Basically, the fees c_{vr}^{Use} for infrastructure usage (other than the scarcity fees) shall cover all short-term marginal socio-economic costs of the train operations. Presently, however, they are set to zero for most trains [Jernbaneverket, 2012, ch. 6.1]. Marginal infrastructure maintenance costs c_v^{Maint} are given in Jernbaneverket [2011], differentiated for passenger and cargo trains.

Summing up over these three main components, some terms cancel each other out: customer payments π_ξ to the TOCs for transport on a train and payments between the TOCs and the IM in the form of scarcity tariffs and other infrastructure usage charges. The resulting expression for socio-economic utility from train transport is given in terms (15a)–(15g) of the IM's objective function.

External effects

In addition to the utility of train traffic to all involved parts, the IM's decision problem takes into account external effects of railroad traffic. First and foremost these are direct consequences of emissions (NO_x and PM_{10} (small particles)) on a local ($c_{fv}^{LocalEm}$) or global ($c_v^{GlobalEm}$) level, noise (c_{fv}^{Noise}), and accidents (c_v^{Acc}). Jernbaneverket [2011] gives values for these costs, differentiated with respect to rolling-stock type[3] v or the area f where the train is operated (city, other urban regions, rural regions).

Objective function terms (15h)–(15i) quantify the effects for the considered train schedules. Obviously, they can be extended or aggregated if necessary.

Other external effects such as health effects are too marginal to be taken into account in detail here. Moreover, they involve measurements of traffic migration from or to other modes of transportation which is neither addressed in our model.

Some consequences of including socio-economics in the model

Optimal solutions may still contain conflicts. As the IM's objective function takes into account also other aspects than conflicting slot requests (such as the socio-economic utility of the trains, external effects, or track maintenance costs), it may actually be optimal to keep some conflicts if this yields a better overall solution. This effect will not be as strongly pronounced when both level models are solved with a rough granularity: Most conflicting slot requests can be resolved by moving departure or arrival times by just a few minutes. With a rough granularity, the trains' running times are not controlled in such a great detail, and such conflicts are masked completely. Hence, the aspect of over-utilization due to lack of capacity will be more in focus.

The IM's decisions must be operator neutral. This implies that the scarcity tariffs can be differentiated only by train types rather than the trains' socio-economic utility. This, in turn, may lead to the "wrong" train (with lower socio-economic utility) being moved when resolving a conflict.

The IM's concern is about socio-economic effects and all decisions are based on this principle. Hence, the

---

[3]Typically, only diesel and electrical engines or goods and passenger trains are distinguished

scarcity pricing is directed toward improving *socio-economic* utility. The TOCs, however, respond with actions improving their *profits*. Also, a train with a low utility for society may have a high value to the TOC (e.g., since it is part of a larger portfolio of trains or a strategically important train). Consequently, the scarcity tariffs may work diametrically to the IM's intentions. These mechanisms are discussed closer in Lium et al. [2014].

3.2 Objective function

The IM's objective is to accommodate all traffic requests such that the available infrastructure is used efficiently and with highest socio-economic utility.

$$
\max_{z} \quad \sum_{s \in \mathcal{S}} \sum_{\xi \in \Xi : s(\xi) = s} \{ W_\xi \tag{15a}
$$

$$
- \left(c_s^{TrFix} + c_{\mathrm{v}(s)}^{RSL} \lambda_s^T \sum_{(i,j) \in \mathcal{B}_{\mathrm{r}(s)}} \lambda_{ij}^B \right) r_\xi - c_\xi^{TrTime} \tag{15b}
$$

$$
- c_s^{Wait} \sum_{(i,j) \in \mathcal{B}_{\mathrm{r}(s)}} w_{j\xi} \tag{15c}
$$

$$
- r_\xi \sum_{(i,j) \in \mathcal{A}_s^{Stop}} c_j^{Stop} \tag{15d}
$$

$$
- \sum_{(i,j) \in \mathcal{A}_s^{Stop}} c_{j\,\mathrm{v}(s)}^{Park} \cdot w_{j\xi}^{Park} \tag{15e}
$$

$$
+ \sum_{u \in \mathcal{U}} Comp_u \tag{15f}
$$

$$
- r_\xi c_{\mathrm{v}(s)}^{Maint} \sum_{(i,j) \in \mathcal{B}_{\mathrm{r}(s)}} \lambda_{ij}^B \tag{15g}
$$

$$
- r_\xi \sum_{(i,j) \in \mathcal{B}_{\mathrm{r}(s)}} \lambda_{ij}^B \left(c_{\mathrm{f}(ij)\,\mathrm{v}(s)}^{LocalEm} + c_{\mathrm{f}(ij)\,\mathrm{v}(s)}^{Noise} \right) \tag{15h}
$$

$$
- r_\xi \left(c_{\mathrm{v}(s)}^{GlobalEm} + c_{\mathrm{v}(s)}^{Acc} \right) \sum_{(i,j) \in \mathcal{B}_{\mathrm{r}(s)}} \lambda_{ij}^B \} \tag{15i}
$$

$$
- M \cdot |\pi^{Tariffs}| \tag{15j}
$$

Term (15a) expresses customers' willingness to pay for transporting passengers or freight on each operated train. It is based on an estimated maximum willingness to pay W_s^{max} for a train of a given service s and multiplier profiles, cf. (14). Expressions (15b)–(15e) specify the TOCs' costs of operating trains, see the explanations to (21b)–(21f) in subsection 4.1 and corresponding terms for the passenger TOCs for more detail, while (15f) states the MoT's compensation to a passenger TOC for carrying out publicly purchased train services. Observe that $c_{j\,\mathrm{v}(s)}^{Park}$ and $w_{j\xi}^{Park}$ are defined only for terminals such that term (15e) exists only for the cargo TOC's trains. Term (15g) describes the cost of maintaining the infrastructure due to wear and tear from the trains.

External effects of train traffic are quantified in (15h)–(15i): The former term describes costs of local effects such as local emissions and noise, while the latter estimates global emission and accident costs.

Term (15j) with M a sufficiently large number helps to balance the IM's aggregated income (or expenditure) $\pi^{Tariffs}$ from the scarcity tariffs $z_{ijp}^{\mathrm{v}(s)}$ which is defined in expression (19a).

Note that term (15j) allows for potentially high charges—as long as they are balanced by high subsidies.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

17 of 56

Alternatively, one may minimize the magnitude of all applied fees,

$$\min_z \sum_{\xi \in \Xi} \sum_{(i,j) \in \mathcal{B}_{\mathrm{r}(s)}} \sum_{p \in \mathcal{P}} |z_{ijp}^{\mathrm{v}(s)} t_{\xi ijp}^{S}| \tag{16}$$

Finally, an extension of (15j) and (16) may define a relation to the socio-economic utility of the conflicting trains following the discussion on page 16.

3.3 Constraints

As specified in (14), the customers' willingness to pay W_ξ for transport on a train ξ of service $s(\xi)$ can be determined from the maximum achievable willingness to pay for transport on a train of this service (if this train would travel at the best possible points of time), a multiplier ν_ξ derived from profile curves as described on page 10, and the indicator r_ξ whether the train is actually operated. As the multiplier ν_ξ depends on the decision variables for the train's actual departure and arrival times, this expression is quadratic. It can be linearized by

$$W_\xi \leq W_{\mathrm{s}(\xi)}^{max} r_\xi \tag{17a}$$

$$W_\xi \leq W_{\mathrm{s}(\xi)}^{max} \nu_\xi \tag{17b}$$

Values for W_s^{max} can be determined similarly to π_s^{max}: per whole train, per car, or per transported unit.

The IM needs to ensure that the capacity of the considered infrastructure elements is not exceeded. With a fine granularity, each block section (except stations and terminals) can accommodate only one train at a time. In a coarser granularity model, a slot may comprise several sections and/or time periods and several trains may use a section during the same time period. In this case, capacity calculations become much more involved.

The *capacity limit* of a section may be defined in several ways, e.g., as theoretical and practical capacity. More importantly, though, the number of trains which can be accommodated on a section during a given time period cannot be stated upfront, but depends on the train mix: the trains' speed (both individually and in relation to preceding and subsequent trains), length, direction, and other factors. Hence, the capacity of a section is typically determined by means of simulations and other empirical means [Skartsæterhagen, 1993], [Jernbaneverket, 2010, ch. 5], [Lium and Werner, 2012, ch. 7]. A simulation or even precise calculation of a given train mix' capacity usage is beyond the scope of the model considered here. An approximation takes into account several typical length classes $l \in \mathcal{L}$ of trains traversing a section. We assume that a section (i, j) has a capacity of κ_{ijl}^{S} trains of length class l during a time period p (if no other trains use this section at this time). Then, a constraint on the capacity utilization of this section during this time period can be written as

$$\sum_{l \in \mathcal{L}} \frac{1}{\kappa_{ijl}^{S}} \sum_{\substack{\xi \in \Xi: \\ \mathrm{l}(s(\xi))=l}} \left(y_{ijp}^{\xi} + y_{jip}^{\xi} \right) \leq 1 \tag{18}$$

Directed capacity bounds can be considered by defining the κ_{ijl}^{S} on links instead of sections and summing only over y_{ijp}^{ξ}.

As mentioned above, there should be a balance between revenue from charges and expenditures for subsidies. In other words, deviances should not exceed a certain percentage $\beta \in [0, 1]$, say, 20 percent. Note that the TOCs pay only for the time $t_{\xi ijp}^{S}$ (in minutes) they actually use a slot (i, j), not for the whole time period p. This time

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

18 of 56

variable can be derived from the train flow variables described in section 2.

$$\pi^{\text{Tariffs}} = \sum_{s \in \mathcal{S}} \sum_{\substack{\xi \in \Xi: \\ s(\xi)=s}} \sum_{(i,j) \in \mathcal{B}_{r(s)}} \sum_{p \in \mathcal{P}} \frac{z_{ijp}^{\text{v}(s)}}{\Delta} t_{\xi ijp}^{S} \tag{19a}$$

$$\sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{\text{v}(s)}>0}} z_{ijp}^{\text{v}(s)} t_{\xi ijp}^{S} \leq -(1+\beta) \sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{\text{v}(s)}<0}} z_{ijp}^{\text{v}(s)} t_{\xi ijp}^{S} \tag{19b}$$

$$- \sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{\text{v}(s)}<0}} z_{ijp}^{\text{v}(s)} t_{\xi ijp}^{S} \leq (1+\beta) \sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{\text{v}(s)}>0}} z_{ijp}^{\text{v}(s)} t_{\xi ijp}^{S} \tag{19c}$$

Constraints (19b) and (19c) contain a condition on the decision variables z and must, hence, be reformulated in order to be included in a model implementation. With $\bar{z}_{ijp}^{\text{v}(s)} = \max\left\{0, z_{ijp}^{\text{v}(s)}\right\}$, that is,

$$\bar{z}_{ijp}^{\text{v}(s)} \geq z_{ijp}^{\text{v}(s)}$$
$$\bar{z}_{ijp}^{\text{v}(s)} \geq 0$$

they become

$$\sum_{(i,j),p,s,\xi} \bar{z}_{ijp}^{\text{v}(s)} t_{\xi ijp}^{S} \leq (1+\beta) \sum_{(i,j),p,s,\xi} \left(\bar{z}_{ijp}^{\text{v}(s)} - z_{ijp}^{\text{v}(s)}\right) t_{\xi ijp}^{S} \tag{20a}$$

$$\sum_{(i,j),p,s,\xi} \left(\bar{z}_{ijp}^{\text{v}(s)} - z_{ijp}^{\text{v}(s)}\right) t_{\xi ijp}^{S} \leq (1+\beta) \sum_{(i,j),p,s,\xi} \bar{z}_{ijp}^{\text{v}(s)} t_{\xi ijp}^{S} \tag{20b}$$

Note that this constraint multiplicatively combines first and second level decision variables $z_{ijp}^{\text{v}(s)}$ and $t_{\xi ijp}^{S}$. However, this poses no problem in an iterative solution process such as the one described in section 8.

4 The cargo train operator's problem

The decision problem of the cargo TOC (Cargolink) is to find a profit maximizing schedule for the company's trains.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

19 of 56

4.1 Objective function

$$\max_{r,t^S,w} \sum_{s\in\mathcal{S}} \sum_{\substack{\xi\in\Xi^C:\\ \mathrm{s}(\xi)=s}} \left\{ \pi_\xi - \sum_{(i,j)\in\mathcal{B}_{\mathrm{r}(s)}} \sum_{p\in\mathcal{P}} \frac{z_{ijp}^{\mathrm{v}(s)}}{\Delta} \cdot t_{\xi ijp}^S \right. \tag{21a}$$

$$-c_s^{TrFix} r_\xi - c_\xi^{TrTime} \tag{21b}$$

$$-r_\xi \left(c_{\mathrm{v}(s)}^{RSL} \lambda_s^T + c_{\mathrm{v}(s)\,\mathrm{r}(s)}^{Use} \right) \sum_{(i,j)\in\mathcal{B}_{\mathrm{r}(s)}} \lambda_{ij}^B \tag{21c}$$

$$-c_s^{Wait} \sum_{(i,j)\in\mathcal{B}_{\mathrm{r}(s)}} w_{j\xi} \tag{21d}$$

$$- r_\xi \sum_{(i,j)\in\mathcal{A}_s^{Stop}} c_j^{Stop} \tag{21e}$$

$$\left. - \sum_{(i,j)\in\mathcal{A}_s^{Stop}} c_{j\,\mathrm{v}(s)}^{Park} \cdot w_{j\xi}^{Park} \right\} \tag{21f}$$

Term (21a) denotes the revenue π_ξ earned from operating train ξ at the given time, determined in constraint (22), and the cost or revenue from scarcity charges. The terms (21b) and (21c) represent all other costs of operating the trains. We divide this into four parts: Fixed costs and time-dependent operational costs are given in (21b). The latter term can be determined using constraints (23) and (24). Distance-dependent operational costs and infrastructure usage charges other than the scarcity fees are given in (21c). These costs arise for each running train $\xi \in \Xi^C$, independent of the time of day when it is operated, but not all cost elements need to be present for all trains.

Term (21d) sums up the costs of a train waiting along its route (or a penalty term) while (21e) gives the costs for (off)loading cargo at terminals. Finally, (21f) represents a penalty term if trains occupy terminals longer than a given time; the penalized time $w_{j\xi}^{Park}$ is calculated in (25). This part can easily be extended by several similar terms in order to mimic a stepwise increasing penalty. Also, observe that the penalty is different from the scarcity charges and vanishes out of the system considered in our model: it is not paid to the IM but, e.g., to some terminal owner.

4.2 Constraints

The main constraints for the cargo TOC's decision problem are those defining the flow of trains (see section 2).

The revenue π_ξ in term (21a) is calculated from the maximum achievable revenue π_s^{max} for a train ξ in service $\mathrm{s}(\xi)$ using the multiplier ν_ξ derived from profile curves and the indicator r_ξ whether the train is actually operated. This is quite similar to the determination of the willingness to pay in (14) and (17):

$$\pi_\xi = \pi_{\mathrm{s}(\xi)}^{max} \nu_\xi r_\xi \tag{22}$$

with a linearization

$$\pi_\xi \leq \pi_{\mathrm{s}(\xi)}^{max} r_\xi$$
$$\pi_\xi \leq \pi_{\mathrm{s}(\xi)}^{max} \nu_\xi$$

The maximum achievable revenue π_s^{max} for a train of service s, in turn, may be specified for a whole train, per car, or per unit of the transported commodity (using the capacity $\kappa_{\mathrm{v}(s)}^{RS}$ of the corresponding rolling stock type). For cargo, the former two options are more common.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

20 of 56

The time-dependent operational costs c_ξ^{TrTime} in (21b) can be expressed as

$$c_\xi^{TrTime} = c_{s(\xi)}^{TrMin}\, r_\xi \sum_{(i,j)\in\mathcal{B}_{r(s(\xi))}} \left(\gamma_{ij}^{v(s)} + \lambda_{j\,s(\xi)}^{Stop}\right) \tag{23}$$

if the train has a fixed service time, i.e. no waiting is allowed. The parameter c_s^{TrMin} denotes the time-dependent operational costs of a train in service s per minute,

$$c_s^{TrMin} = \lambda_s^T \frac{c_{v(s)}^{RST}}{\Delta} + \frac{c_s^{TrHour}}{\Delta}.$$

If the train's travel time is variable, the right hand side of (23) becomes quadratic,

$$c_\xi^{TrTime} = c_{s(\xi)}^{TrMin}\, r_\xi \sum_{(i,j)\in\mathcal{B}_{r(s(\xi))}} \left(\gamma_{ij}^{v(s)} + w_{j\xi} + \lambda_{j\,s(\xi)}^{Stop}\right).$$

Hence, equation (23) is reformulated to[4]

$$c_\xi^{TrTime} \geq c_{s(\xi)}^{TrMin} \left(\sum_{(i,j)\in\mathcal{B}_{r(s(\xi))}} \left(\gamma_{ij}^{v(s)} + w_{j\xi} + \lambda_{j\,s(\xi)}^{Stop}\right) - D_s(1 - r_\xi) \right). \tag{24}$$

Time spent at a terminal $(i,j)\in\mathcal{A}_s^{Stop}$ beyond a given limit L_{js}^{Park} is interpreted as parking. This excess time

$$w_{j\xi}^{Park} = \max\left\{0,\, w_{j\xi} + \lambda_{js}^{Stop} - L_{js}^{Park}\right\} \tag{25}$$

is calculated by

$$w_{j\xi}^{Park} \geq 0$$
$$w_{j\xi}^{Park} \geq w_{j\xi} + \lambda_{js}^{Stop} - L_{js}^{Park} \tag{26}$$

5 The small passenger train operator's problem

The main goal for the small passenger train operator (Flytoget) in the setting considered here is to maximize profit.

5.1 Objective function

$$\max_{r,t^S,w} \sum_{s\in\mathcal{S}} \sum_{\substack{\xi\in\Xi^{SP}:\\ s(\xi)=s}} \left\{ \pi_\xi - \sum_{(i,j)\in\mathcal{B}_{r(s)}} \sum_{p\in\mathcal{P}} \frac{z_{ijp}^{v(s)}}{\Delta} t_{\xi ijp}^S \right. \tag{27a}$$

$$-c_s^{TrFix} r_\xi - c_\xi^{TrTime} \tag{27b}$$

$$-r_\xi \left(c_{v(s)}^{RSL}\lambda_s^T + c_{v(s)\,r(s)}^{Use} \right) \sum_{(i,j)\in\mathcal{B}_{r(s)}} \lambda_{ij}^B \tag{27c}$$

$$-c_s^{Wait} \sum_{(i,j)\in\mathcal{B}_{r(s)}} w_{j\xi} \tag{27d}$$

$$\left. - r_\xi \sum_{(i,j)\in\mathcal{A}_s^{Stop}} c_j^{Stop} \right\} \tag{27e}$$

[4]An alternative formulation would calculate the actual travel time from the difference between departure and arrival time of the train, $t_{\xi i_{r(s)}^{RD}}^T - t_{\xi j_{r(s)}^{RO}}^H$, instead of summing up over its travel time on the individual links, $\sum_{(i,j)\in\mathcal{B}_{r(s)}} \left(\gamma_{ij}^{v(s)} + w_{j\xi} + \lambda_{js}^{Stop}\right)$. This results also in a quadratic expression. Then, in addition, possible looping must be taken into account, ensuring that $t_{\xi i_{r(s)}^{RD}}^T - t_{\xi j_{r(s)}^{RO}}^H$ is nonnegative.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

21 of 56

Term (27a) expresses the revenue π_ξ from each train and the cost or revenue from the scarcity charges (subsidies). Expressions (27b) and (27c) represent all other costs of operating the trains. Similar to the cargo train operator's objective function, this expression is divided into fixed and time-dependent costs of operating a train of service s in (27b) and distance-dependent operational costs and other infrastructure usage charges in (27c). The term (27d) expresses penalty costs for longer driving time (waiting). Finally, (27e) represents the cost of using a station along a train's route (at present, this concerns prioritized services at the Oslo S Airport Express Train Terminal and the Lillestrøm and Gardermoen stations on GMB).

5.2 Constraints

The main constraints for the small passenger TOC's decision problem are those defining the flow of trains $\xi \in \Xi^{SP}$ (see section 2). The revenue π_ξ can be calculated from the maximum achievable revenue π_s^{max} using multiplier profiles as in (22). For passenger trains, π_s^{max} will most likely be calculated from the maximum quantity of passengers determined, for example, from the number λ_s^T of cars (car sets) of the rolling stock type used for a train in service s and the number $\kappa_{v(s)}^{RS}$ of seats in each car (car set), and a fare which is averaged over the train's route $r(s)$. The time-dependent operational costs c_ξ^{TrTime} are defined similarly to (23)–(24).

6 The large passenger train operator's problem

This operator (NSB) did not participate in the project, but operates by far the majority of the trains on the network. In addition, the TOC is responsible for carrying out publicly purchased traffic. It appeared, therefore, paramount to model also the presence of this TOC.

6.1 Publicly purchased traffic

The **traffic contract ("Trafikkavtalen")** [Samferdselsdepartementet, 2012] between the Ministry of Transport and a TOC (at present, NSB) shall ensure the execution of unprofitable passenger transport by train where it is of general economic importance (public service obligation). For each commission, the contract specifies a frequency and some time information for when the trains shall be operated. In compensation for satisfying these requirements, the TOC receives a lump sum for each route [Samferdselsdepartementet, 2012, app. E]. This compensation shall, together with revenues from ticket sales and other revenues, cover all costs of producing the commissioned train traffic and give a "reasonable profit" (ch. 17.1). Such a requirement is difficult to translate into an exact mathematical expression. As publicly purchased train traffic is not the main topic of the model described here (but important enough to be included), we will not go into much detail to find such an expression. Instead, we only include the compensation, assuming that it is set sufficiently high in the contract.

For each commission ("bestilling") $u \in \mathcal{U}$, the contract specifies a minimum number of trains ω_u (or a frequency) on a route ρ_u over a given time period $[\tau_u^E, \tau_u^L]$, see Appendix A of the contract. The time period refers to either arrivals or departures at a given "reference" station (i_u, j_u), e.g., Oslo, Trondheim, Stavanger, or Bergen. The contract appears unclear on whether train arrivals or departures are considered in a commission. Therefore, we chose the following convention for the SOPJI model: If the "reference" station is the origin or an intermediary station of the route, then the time interval refers to departures from this station. If the station is the destination of the route, the interval refers to arrivals.

The auxiliary binary variables $a_{\xi u}^{From}$ and $a_{\xi u}^{Until}$ register whether a train ξ departs from the reference station

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

22 of 56

within the specified time interval[5]:

$$a_{\xi u}^{From} = \begin{cases} 1, & t_{\xi j_u}^H \geq \tau_u^E \\ 0, & \text{else} \end{cases}$$

$$a_{\xi u}^{Until} = \begin{cases} 1, & t_{\xi j_u}^H > \tau_u^L \\ 0, & \text{else} \end{cases}$$

which can be reformulated to

$$t_{\xi j_u}^H - \tau_u^E \leq a_{\xi u}^{From} P\Delta \tag{28a}$$

$$\tau_u^E - t_{\xi j_u}^H \leq (1 - a_{\xi u}^{From})P\Delta \tag{28b}$$

$$t_{\xi j_u}^H - \tau_u^L < a_{\xi u}^{Until} P\Delta \tag{28c}$$

$$\tau_u^L - t_{\xi j_u}^H < (1 - a_{\xi u}^{Until})P\Delta \tag{28d}$$

That is, $a_{\xi u}^{From}$ indicates whether the train departed station (i_u, j_u) at or after time τ_u^E while $a_{\xi u}^{Until}$ indicates whether it departed after time τ_u^L. Then, a minimum number ω_u of trains departing this station within the time interval $[\tau_u^E, \tau_u^L]$ is ensured through

$$\sum_{\substack{\xi \in \Xi^{LP}: \\ \mathrm{r}(s(\xi))=\rho_u}} \left(a_{\xi u}^{From} - a_{\xi u}^{Until} \right) \geq \omega_u \tag{29}$$

Replacing $t_{\xi j_u}^H$ with $t_{\xi i_u}^T$ in the above, we obtain formulations which hold when the commission u refers to the *arrival* of trains at a station (i_u, j_u).

There are, however, some issues which prevent a straight-forward inclusion in a real-life model implementation:

As the compensation shall cover all costs of operating the concerned trains, it shall also cover all *public fees* these trains are subject to. On the other hand, the compensation is explicitly stated as a lump sum for each route. Hence, a renegotiation would be required after the scarcity charges and subsidies are determined—possibly similar to the process described in paragraphs 5 and 6 in ch. 17.1.

Moreover, the traffic contract stipulates that the MoT is obliged to arrange for the TOC to be able to satisfy their commitments. The contract also builds on the assumption that all infrastructure necessary for delivering the purchased traffic production is available to the TOC but the TOC shall apply for infrastructure access according to the rules and guidelines laid down by the IM [Samferdselsdepartementet, 2012, ch.12]. This implies that—even if all explicit prioritization rules are set aside—publicly purchased traffic shall be prioritized in the case of capacity conflicts.

In principle, such a prioritization may be achieved either by reducing the available infrastructure capacity accordingly before granting access for all other trains or by not applying scarcity charging for prioritized trains. The charges may still be calculated and, thus, visualize the costs of operating the publicly purchased traffic at this slot. Both ideas, however, pose some practical challenges:

The contract states only a *minimum* number of trains to be operated and the TOC may schedule more trains on a concerned route in the given time interval. In such a case, it is challenging to decide which of these trains are to be considered covered by the contract with the MoT. Only these trains should be prioritized while all other trains should compete for infrastructure access on par with other TOCs' trains. This problem is aggravated by the fact that, typically, the time intervals given in the contract are longer than the slots used for determining the scarcity charges.

---

[5]These time intervals may be given at a finer resolution than the time slots for the scarcity fee and the reference station may also be an intermediary station. Hence, we cannot use the variables $h_{\xi i_u \tau_u^E}^H$, $h_{\xi j_u \tau_u^E}^T$, $\delta_{\xi P}^D$, or $\delta_{\xi P}^A$ from section 2 here.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

23 of 56

Also, as mentioned above, trains covered by the contract should be exempt from scarcity charges while the other trains of the TOC should be not. Again, there is the challenge of deciding which trains are to be included. Obviously, the MoT would consider trains outside the most popular slots (possibly even those collecting a subsidy) while the TOC would like the trains with the highest charges to be covered.

Within the context of and the expertise involved in the SOPJI project, clear and concise answers to these challenges cannot be found. Therefore, we leave these topics for further discussion.

6.2 Objective function

This TOC's objective function is quite similar to the small passenger TOC's function but includes the reimbursement from the MoT for publicly purchased services.

$$\max_{r,t^S,w} \sum_{s\in\mathcal{S}} \sum_{\substack{\xi\in\Xi^{LP}:\\ \mathrm{s}(\xi)=s}} \left\{ \pi_\xi - \sum_{(i,j)\in\mathcal{B}_{\mathrm{r}(s)}} \sum_{p\in\mathcal{P}} \frac{z_{ijp}^{\mathrm{v}(s)}}{\Delta} t_{\xi ijp}^S \right. \tag{30a}$$

$$-c_s^{TrFix} r_\xi - c_\xi^{TrTime} \tag{30b}$$

$$-r_\xi \left(c_{\mathrm{v}(s)}^{RSL} \lambda_s^T + c_{\mathrm{v}(s)\,\mathrm{r}(s)}^{Use} \right) \sum_{(i,j)\in\mathcal{B}_{\mathrm{r}(s)}} \lambda_{ij}^B \tag{30c}$$

$$-c_s^{Wait} \sum_{(i,j)\in\mathcal{B}_{\mathrm{r}(s)}} w_{j\xi} \tag{30d}$$

$$\left. -r_\xi \sum_{(i,j)\in\mathcal{A}_s^{Stop}} c_j^{Stop} \right\} \tag{30e}$$

$$+\sum_{u\in\mathcal{U}} Comp_u \tag{30f}$$

Term (30a) denotes the revenue from passenger transportation calculated similar to (22) and the loss or profit from scarcity charges and subsidies. Expressions (30b) and (30c) give the costs of operating the trains, divided into fixed and time-dependent costs of operating a train of service $\mathrm{s}(\xi)$ in (30b) and distance-dependent operational costs and other infrastructure usage charges in (30c). See also the explanations to the corresponding terms (21b) and (21c). Penalty costs for longer driving time are given by (30d) while (30e) represents the cost for stops at some specified stations (currently, Lillestrøm and Gardermoen stations on GMB). Term (30f) sums up over the (lump sum) compensations for each public purchase.

6.3 Constraints

The constraints are quite similar to those of the small passenger TOC: The main constraints define the flow of trains (section 2), including limits on the duration of a train and on its waiting time. Further constraints determine the revenue π_ξ from train operations and time-dependent operational costs c_ξ^{TrTime}. Constraints (28) and (29) express the requirements stipulated in the traffic contract between the MoT and the operator.

7 Solving the problem

The complete IM–TOC model describes a classical bilevel optimization problem with the IM as leader (upper level) and the TOCs as followers (lower level). Such problems consist of hierarchically nested sub-problems, each representing a fully-fledged optimization problem in itself: The IM's taxation scheme is a set of parameters for the TOCs' decision problems. Vice versa, the TOCs' responses must be included into the IM's decision

problem. Even if these responses could be expressed in a closed form, the resulting optimization problem would be non-convex and difficult to solve exactly, see, e.g., Dempe [2002].

In our case, the TOCs' decision problems are even mixed-integer problems, albeit linear. In addition, the constraints (20) describing the IM's revenue bounds contain the product of first- and second-level decision variables. However, the involved second-level decision variables are binary such that this constraint may be reformulated—at the cost of introducing (further) binary variables. As these would be at the upper level, this would be less of a complication, though. Several papers indicate that this class of bilevel programming problems is notoriously difficult to solve with exact methods—it appears even difficult to derive good optimality conditions, see Dempe [2001] and Dempe and Fanghänel [2009]. The former work discusses several solution approaches but stresses the lack of actually applicable methods. Also Gassner [2009] reported that good solution methods for this problem class had not been developed yet. Obviously, the situation does not become easier when considering large real-life instances such as (parts of) the Norwegian railway network.

7.1 Iterative solution processes

An iterative heuristic method appears most suitable for our problem structure. One may start, for example, with a tariff schedule with zero charges or a previously found schedule and determine the TOCs' responses to this schedule with their current demand parameters (or previous demand parameters with some updates or forward projections). The IM then heuristically adjusts the tariffs in order to achieve a better traffic distribution, for example, in terms of fewer conflicts, more even capacity utilization, or higher socio-economic utility. Then, the TOCs' best response to these updated tariffs is determined. This procedure continues until no over-utilized slots are found or a maximum number of iteration is reached. As the TOCs set their schedules without taking into account the other TOCs on the network, this step in the algorithm can be performed in parallel for all TOCs. Figure 3 illustrates this approach for a situation with one IM, one cargo TOC ("CTOC"), and two passenger TOCs ("PTOC1" and "PTOC2").
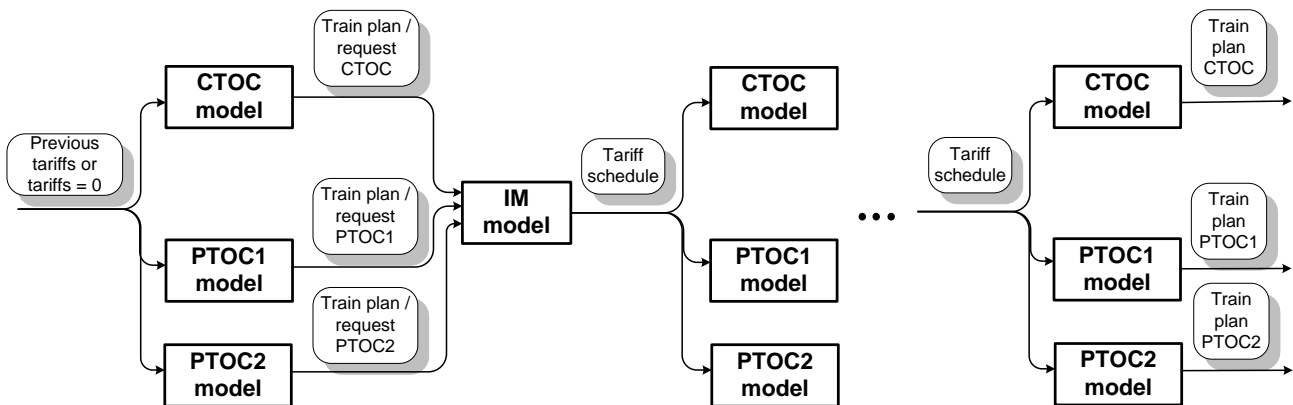
Figure 3: An iterative solution approach for the two-level IM–TOC problem

7.2 Setting the scarcity charges

The Norwegian railway network is rather sparse with very few alternative routes between origin / destination pairs. Hence, traffic is much more mobile with respect to time than geography. For example, operating a cargo train from Oslo to Trondheim some hours earlier than intended is (mostly) feasible, routing this train via Nordlandsbanen is obviously not feasible. Balancing the charges across the whole network would mean that, e.g., traffic passing the Oslo area will very likely subsidize traffic on little used sections somewhere else in the country. It is conceivable that, in areas with high infrastructure utilization, scarcity charges lead to a migration of traffic to other modes.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

25 of 56

In the model implementation, scarcity charges are balanced over the complete time horizon separately for each section rather than for the whole network. This helps to accommodate as much train traffic as possible on the network without affecting the TOCs' route choices too much. It is, however, possible to test alternative routes between the same origin/destination pairs (e.g., choosing between Dovre- and Rørosbanen for trains between Oslo and Trondheim), see page 14.

After calculating the capacity utilization in each slot (section × time period), the charges are determined in a two-step process. In the first step, fees are increased for all time periods where the section is over-utilized. At the same time, the fees in periods with a corresponding lowest utilization are decreased (i.e., subsidies are increased). This is accommodated by ranking all slots according to their capacity utilization. For example, increasing the fee for the slot with the second highest utilization leads to decreasing the fee in all slots with the same utilization as the second-last ranked slot.

In the second step, these charges are balanced based on the amounts which will actually be received or paid out, taking into account the trains' travel time over the considered section: If the sum of all collected fees is large compared to the sum of all paid-out subsidies, each fee and each subsidy on the section are divided by the square root of the ratio between fees and subsidies.

7.3   Computation times and decomposition

Ultimately, the model will be applied to real-life cases such as the complete Norwegian railway network. As such cases are potentially quite large, with over a thousand block sections, a week-long (even two-week-long) optimization horizon and several hundred trains each day, the resulting mathematical programming problem may become too large to be solved in reasonable time, even when solving the TOCs' sub-problems in parallel. Hence, it is important to find a reasonable level of detail, for example with respect to time granularity and aggregation of the network into sections.

Obviously, the problem size can be reduced by carefully analysing the case at hand before attempting to solve it. For example, narrow time windows for train departures and arrivals—also at intermediary stopping places, taking into account limits on wait and travel times—reduces the number of (binary) variables in the problem. Likewise, services with wide time windows such as cargo trains may be decomposed into several, mutually exclusive services with narrower time windows, see page 14. For example, the test runs presented in section 9 show that a passenger TOC's sub-problem with about 430 tightly scheduled trains took only about 2-3 seconds to solve. In contrast, the solution time of the cargo TOC's sub-problem involving 5 trains with wide time windows varied between 10 and 90 seconds.

The section-wise determination of the scarcity tariffs described in subsection 7.2 helps to speed up computations as it decomposes the IM's problem into many much smaller sub-problems which can be solved independently.

One may employ further decomposition approaches with respect to either geography or time. For example, areas or corridors with high traffic density (e.g., Oslo and the Østlandet region) may be singled out in order to solve these sub-problems separately. Most likely, the majority of capacity conflicts requiring scarcity charges will occur in these regions anyway. Hence, measures affecting infrastructure utilization in these regions at certain times will not affect traffic in neighbour regions much except for some trains crossing over to or from the high-density region.

Likewise, if there are times with very little traffic, the optimization horizon may be decomposed along those times into separate smaller planning periods.

8   Model implementation

The mathematical model and the solution algorithm have been implemented in XPress MP, reading input parameters from a Microsoft Access database with a structure described in appendix D. Output from running the model is presented in the XPress MP user interface while selected results are written to text files, enabling the automatic creation of graphs using the open source program Gnuplot (`http://www.gnuplot.info/`).

The implementation can handle time periods with flexible duration—all time periods and their start and end time units are specified in the database table **Periods** (rather than the fixed length Δ used in the mathematical model formulation above).

Stations and terminals are handled as nodes in the implementation rather than sections. This does not seem to impose serious challenges with respect to our modelling choices. If necessary (for example, in order to control capacity utilization also in stations or terminals), one may introduce nodes like "Station boundary north" and "Station boundary south" and treat stations or terminals similar to other sections.

Pre-created trains. The mathematical model is based upon an existing set $\Xi = \Xi^C \cup \Xi^{SP} \cup \Xi^{LP}$ of the single TOCs' trains. These trains are grouped to services $s \in \mathcal{S}$ according to some common properties, see 1.3. Hence, in the implementation, it appears efficient to only specify the services in the database and to create the trains from this. The database table **Services** gives the earliest starting time σ_s^{DE} for a train of a service, `ServiceStartTime`. Then, using the number `ServiceRepeat` of trains in the service and the time `ServicePeriod` between each train, the earliest starting time for *each* train of this service can be determined while `ServiceMaxPostpone` gives an upper limit for how much each train's starting time can be postponed. This is treated similarly to the maximum waiting times at the trains' intermediary stopping places given in database table **ServiceWait**.

The maximum allowed travel time for trains of the service is calculated from either a given fixed travel time, a maximum travel time explicitly stated in database table **Services**, or from the maximum waiting times at all nodes and the standard travel time on each link of the service's route. The latter, together with the departure time at the first node, defines also time windows for arrival and departure of each train on intermediary and destination nodes.

Socio-economic optimality. Observe that the IM pursues two—possibly opposing—goals, minimizing over-utilization (or removing capacity conflicts) and maximizing socio-economic utility. The inherent complexity of the bilevel model structure discussed on page 25 poses challenges to developing an efficient solution method for our problem. As the first and foremost goal of the scarcity charges is to avoid over-utilization, the prototype implementation discussed in this report focusses on this goal. This means that, when adapting the schedule of scarcity tariffs, the iterative procedure does not regard the IM's objective function (15). However, it is possible to calculate the corresponding values (both socio-economic utility and profit contribution) of each single train and, hence, evaluate the schedules in this regard.

Kaut et al. [2014] investigate an alternative implementation of the model and solution procedure which can accommodate both goals, using a sequence of subproblems. In a first step, the two extremes are explored: TOCs independently planning their routes (without regarding the other operators or socio-economic aspects) and the IM planning all routes (without taking into account the TOCs' profit maximization goals). The former finds the maximum achievable TOC profits and the latter gives the maximum achievable socio-economic utility. Then, a scarcity tariff is determined which avoids over-utilization and ensures at least a certain percentage of the maximum achievable socio-economic utility while the TOCs realize as high profits as possible. As a last step, the TOCs' response on this tariff is tested. If this response does not satisfy the IM's requirements, another sequence of these steps is performed, forming an iterative solution procedure.

9 Illustrative case

A smaller test case illustrates the capabilities and functionality of the implemented prototype.

9.1 Case description

We consider selected services in the triangle between Drammen, Ski, and Lillehammer / Trondheim which includes the highly utilized area around Oslo. For our case, this triangle comprises 14 stations which are used by the five lines Hovedbanen, Gardermobanen, Dovrebanen, Drammenbanen, and Østfoldbanen, see figure 4.
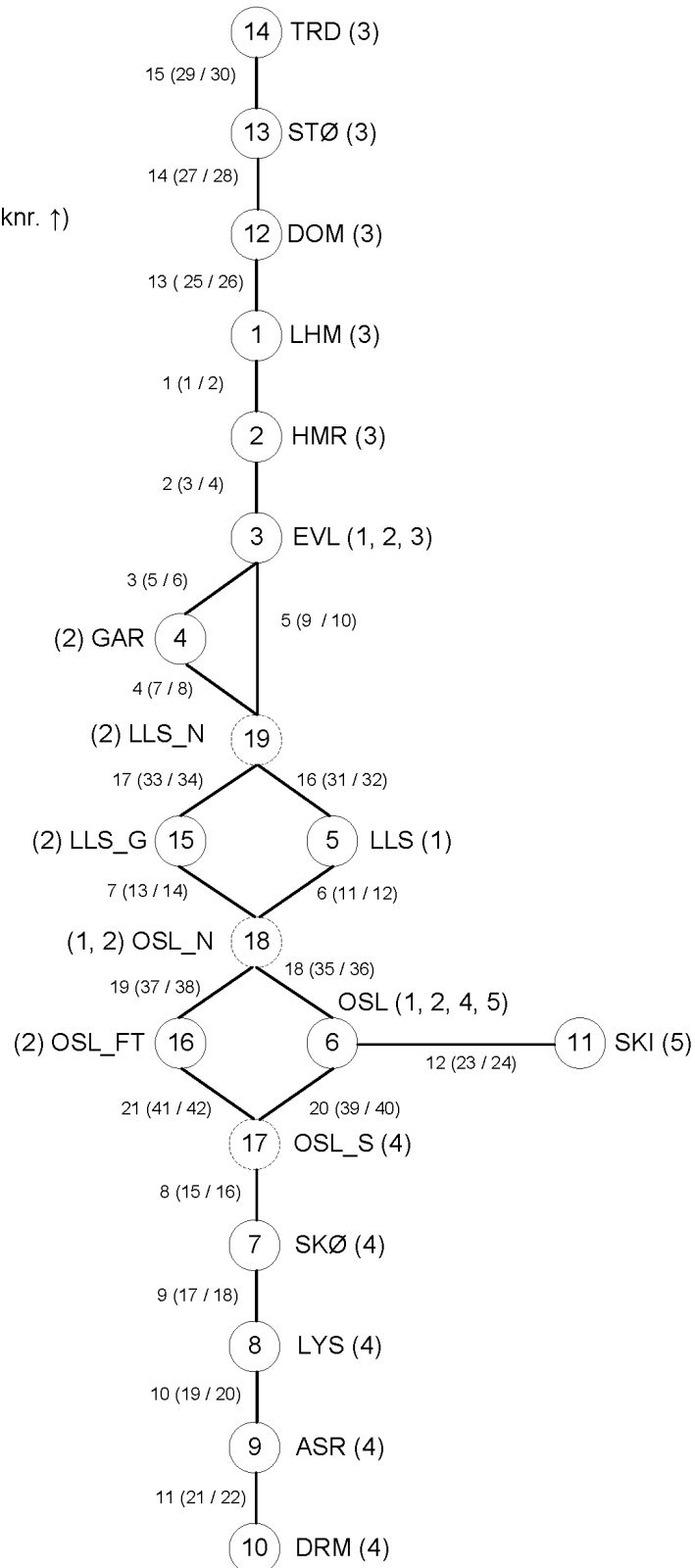
Figure 4: Layout of the infrastructure network used in the case, including node, section, link, and line numbers (in parentheses after node name codes)

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

28 of 56

Due to the specifics of Gardermobanen, it was necessary to introduce further nodes, denoting Oslo S Airport Express Train Terminal, Lillestrøm Airport Express Train Terminal, and three nodes for crossing over between Gardermobanen and Hovedbanen or Drammenbanen, respectively (north of Lillestrøm and both north and south of Oslo S). This results in 21 bi-directional sections and, hence, 42 links.

Three TOCs are studied, each with a specific portfolio of train services. In total, there are 28 different routes with 38 services. The large passenger TOC NSB operates about 415 trains, mostly commuter trains and a few long-distance trains. The smaller passenger TOC Flytoget runs about 215 Airport Express trains between Gardermoen and Oslo or Drammen, respectively. All these trains run on a fixed schedule where the only flexibility is the option to cancel single trains or a complete service. The cargo TOC Cargolink runs a handful of trains between Oslo and Trondheim and between Drammen and Trondheim. These trains are far more flexible with respect to departure and travel times. We also include options to choose between alternative trains: one choice is between a short, early departing and a long but later Oslo–Trondheim train while another choice is between an early-morning and an afternoon departure for a Drammen–Trondheim train.

Model input parameters defining the physical network, the train operations, and economic aspects are based on the data sources specified in appendix D.

9.2 Results and discussion

The heuristic algorithm described in 7.1 takes about seven iterations to arrive at a solution. Due to the large number of trains involved, the subproblems of the passenger TOCs comprise about 16.500 (Flytoget) and 48.400 (NSB) variables, respectively. As the trains of these TOCs do not have much flexibility (no waiting time or postponement for trains), the presolve step of XPress MP reduces the problem sizes drastically to only a few hundred variables. Therefore, in each iteration step a solution is found in only a few seconds. The cargo TOC's subproblem concerns only very few but quite flexible trains. The initial subproblem of this TOC contains about 11.200 variables which presolve can reduce only to about 8.200 variables. Hence, each iteration step takes considerably longer time, up to 90 seconds. However, the total solution time of the heuristics amounts to only a few minutes. In addition to some output in the XPress MP user interface during the solution process, we visualize selected results obtained during the iteration automatically using Gnuplot. Figures 5 and 6 show examples of these graphs.

Figure 5: Initial network utilization and scarcity tariffs after first iteration

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

29 of 56

Figure 5 illustrates the initial network utilization (without scarcity tariffs) in all time slots; time periods are shown on the horizontal axis and sections on the vertical axis. Obviously, a northbound train increases utilization on the Dovrebanen line during morning hours. In particular, the Lillehammer–Dombås section (LLH–DOM) has not sufficient capacity during three hours mid-morning. The input data show that this train is a goods train which is flexible. Therefore, one may expect that a scarcity tariff on these line segments will be able to resolve this problem.

Additionally, some line segments around Oslo are highly utilized: Lysåker–Asker (LYS–ASR), Skøyen–Lysåker (SKØ–LYS), and Lillestrøm–Oslo on the Hovedbanen line (LLS-OSL_N (HB)). Here, no significant variations throughout operation hours are discernible. Therefore, a scarcity tariff might not be of much help as it mainly seeks to balance traffic on a line over the day.

The graph on the right hand side in figure 5 shows a scarcity tariff schedule trying to address this over-utilization (the iteration is started with zero fees and subsidies). Indeed, fees are set at times with high utilization, balanced with subsidies at times with no or little traffic on the same line.
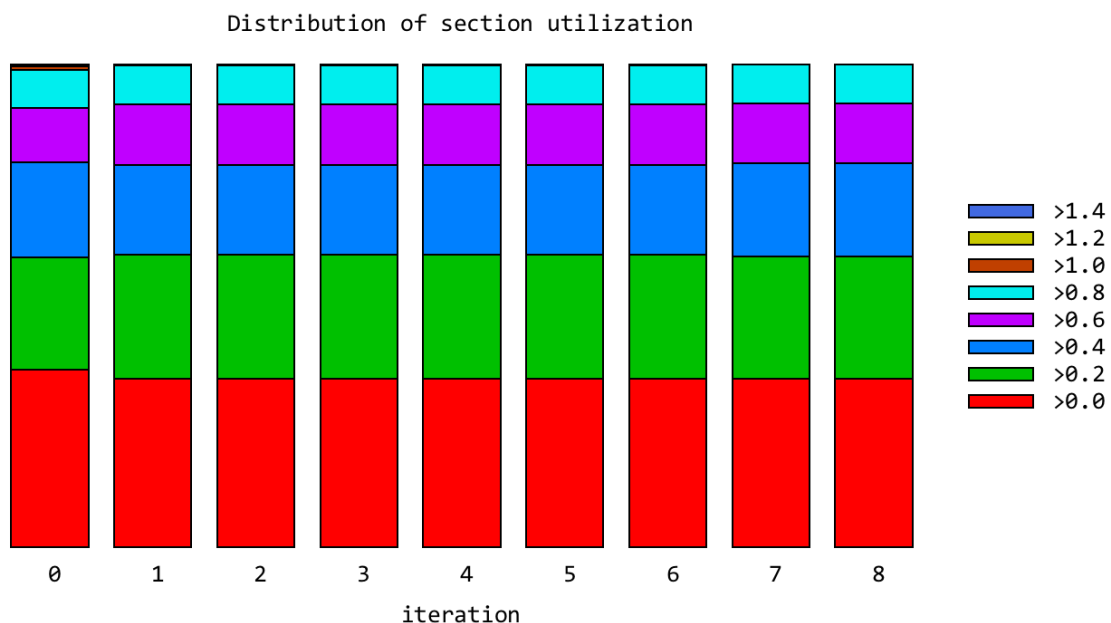

Figure 6: Development of distribution of network utilization during the solution process

The solution after seven iterations indicates that a tariff schedule could be found which removes the worst over-utilization as also evidenced by figure 6. It does not affect the passenger TOCs' train schedules: Flytoget operates all trains and NSB cancels four of the scheduled trains, both before and after the iteration. As expected, though, the cargo TOC adapts operations: In the initial schedule, the TOC operates the shorter Oslo–Trondheim train with departure at 8 o'clock and a waiting time of 66 minutes along the way. For the Drammen–Trondheim train, the later alternative is chosen, departing at 17 o'clock without waiting time. The Trondheim–Drammen train departs at 6 o'clock without waiting. After seven iterations, the cargo TOC still operates the shorter Oslo–Trondheim train, but it departs now an hour earlier while waiting time along the way has increased by one hour. The Drammen–Trondheim train runs now with the earlier option, starting at 7 o'clock while the Trondheim–Drammen train departs now much later, at 16 o'clock. There is no waiting scheduled for both trains.

The left-hand side graph in figure 7 shows that traffic on the Dovrebanen line segments between Eidsvoll and Dombås (HMR–EVL, LHM–HMR, and LHM–DOM) is spread out more evenly over the day. In particular on the latter, fees and subsidies of a relatively high magnitude are required. Note, however, that the right-hand side graph shows a suggested tariff while the algorithm balances the fees and subsidies based on what is actually paid out and received. Hence, there are only subsidies visible for the Skøyen–Lysåker line segment: The algorithm

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

30 of 56

Figure 7: Network utilization and scarcity fees/subsidies after iteration 7

tries to attract trains to these night time slots. However, if it were expected to actually pay out these subsidies, also fees would have to be raised at some time during the day.

Finally, figure 8 shows the total fees and subsidies actually paid and received by the operators and their balance in each of the iteration steps. The small passenger TOC Flytoget is not affected by this tariff schedule at all as they operate only on line segments where no tariffs exist—the night-time subsidies on SKØ–LYS are outside of Flytoget's operating hours.

The large passenger TOC NSB has a dense train schedule, which cannot be altered except for cancellations, and pays a substantial sum of scarcity fees. In theory, the operator may avoid the fees by cancelling trains but—as most fees occur during rush hour when many passengers travel—it is very likely that the fees are more than off-set by revenues from these trains. (It may, however, be incidental that there are no NSB trains at slots where subsidies can be collected.)

The cargo TOC Cargolink both pays fees and collects subsidies, resulting in a net gain from the tariff schedule. As the operator's trains cover long distances it appears that it is not possible to avoid slots with fees entirely. It is also conceivable that a schedule avoiding such slots leads to less subsidies collected on other slots, resulting in a lower net gain. In addition, the income profiles describe the TOC's variation in revenue actually realized depending on the train's departure and arrival times. Therefore, it may be worthwhile to pay some fees if this results in a more profitable operation.

The grey dashed line illustrating the net fees received by the IM shows how fees and subsidies gradually balance out during the solution process. These values are based on the suggested schedule—which again is based on the TOCs' operations in the previous iteration step—and lag therefore by one step. The implementation does not require that fees and subsidies balance perfectly and some small deviation is permitted. Hence, the IM experiences a small net gain also in the final iteration step.

PROJECT
102000998
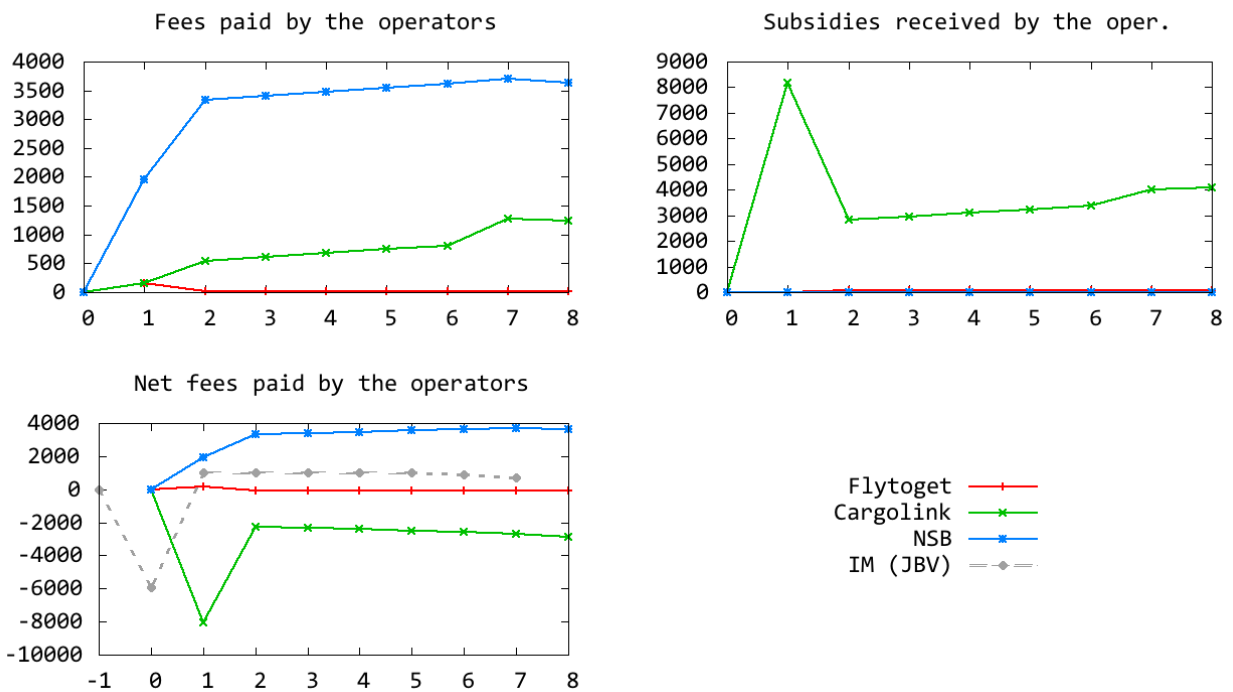
REPORT NUMBER
SINTEF A25843

VERSION
1.0

31 of 56

Figure 8: Development of fees and subsidies paid and received by the operators during the solution process

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

32 of 56

A   Customer behaviour (TOC–M model)

The introduction and section 1.6 describe how the IM–TOC model addresses the behaviour of the TOCs' customers in a simplified way: For each train, the maximum achievable TOC income and customer willingness to pay are modulated by multiplier profiles reflecting the variation of user preferences with respect to departure and arrival times over the day. If more detailed information about the customer markets is available, one may model customer response on the TOCs' train schedules and, hence, determine transportation demand more precisely.

The simplified approach considers the maximum achievable demand and the profiles as exogenous values and, hence, as parameters to the IM and TOCs' models. In contrast, determining transportation demand through a dedicated TOC–M model acknowledges that the TOCs' decisions about their train operations affect transportation demand.

As illustrated in Figure 1, the complete constellation forms a three-level optimization problem between the IM and the TOCs on the one hand and between the TOCs and their customers on the other hand. Section 7 mentions that solving the two-level IM–TOC optimization problem fast and exactly is very challenging. Obviously, this is even more true when including another, non-trivial level: In order to evaluate the quality of a given tariff plan, the IM would need to take into account the TOCs' response (train schedules) on this plan. In turn, to find these decisions, the customer demand must be determined as response on these train operations. Often, this demand is highly uncertain, see also the discussion in section 1.1. Segmenting the market or using discrete demand scenarios may alleviate this challenge but, still, a stochastic mixed integer problem with a bilevel structure must be solved.

An idea which emerged recently in the field of stochastic programming is to identify the "skeleton", a bearing structure, of a complex stochastic mixed-integer problem which may be considered as deterministic (preferably comprising the integer variables). Then, a simpler stochastic linear problem can be used to fill in the finer details. For commodity flow problems serving as illustrative examples for this approach, the skeleton consists of edges of the network while capacities on the edges are found in the stochastic model. This is quite similar to the structure of our TOC–M model. Actually, the train schedules emerging from the IM–TOC model represent already the skeleton of the TOC-M model to be solved. A first model formulation (DC1) evaluates the deterministic train schedules with fixed train capacities against uncertain customer preferences. A refinement of this model (DC2) seeks to find optimal train capacities.

Both TOC–M models use a characterization of the TOCs' services by attributes and a segmentation of the customer market. Such *discrete choice models* can accommodate uncertainty about customer behaviour and estimate satisfied and lost transportation demand flows and, consequently, revenue and socio-economic utility from given train schedules. Furthermore, they can show variations in the traffic flows along a train's route whereas the IM–TOC model considers only an averaged value for each train. The models take a decision-independent starting point which reflects the information available at the time of optimization [Vaagen et al., 2011]. Actual substitution is an outcome of the optimization process, constrained by available substitutes.

Customers do not select travel choices per se but look at particular service attributes such as departure time, punctuality / variability, travel duration, fares, or comfort. We focus on attributes that can cause traffic shifts from peak to non-peak periods, namely time information (departure and arrival times, travel duration), or fares. Combining these with a further attribute, origin–destination pairs, we define *choice possibilities*. Similarly, the TOCs' customers are distinguished by way of some characteristics, creating a set of *customer segments*. For example, passenger customer segments may be "Commuter", "Leisure traveller", or "Business traveller" while cargo customer segments may be defined according to price and time sensitivies. It is conceivable that passenger trains involve more choice possibilities but fewer customer segments and vice versa for cargo trains. If their first preference is not available, customers in a given segment may switch to another choice or not avail of any choice (e.g., they migrate to other modes of transport or do not travel at all) with a certain probability. These probabilities constitute a so-called *substitutability matrix* (section A.4). Obviously, the chance of finding acceptable substituting choices increases with the number of "similar" trains.

Then, taking into account train capacity limitations for the origin–destination pairs of the single choice possibilities, the model determines the actual traffic flow (each customer segment's first and second preferences

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

33 of 56

as well as lost traffic) for each choice. From these flows, the single trains' "value" (both socio-economic utility and the revenue for the considered TOC) can be estimated. It is conceivable that customers travelling on a second-preference choice perceive a lower utility than passengers on a first-preference choice. Lost traffic can be valuated using opportunity costs or losses (gains) of migration to other transport modes.

Note that the models rely on optimally allocating traffic flows. That is, we assume that the TOCs can control (or at least influence) the customers' transportation choices by, e.g., making passengers switch from their first to a second preference even when there are first-preference seats left. Clearly, this is not realistic: Travellers do not consult their TOC's profit maximization problem when purchasing a journey. However, one may assume that individual customer preferences are not ranked in decreasing order, but that each choice possibility has a set of choices which appear nearly equally good to a certain share of the customer market. In this case, an optimal allocation of first and second preferences can be considered as rather realistic from the TOCs' perspective although it still provides an optimistic upper bound on what is practically achievable.

Exact solutions to the customer-behaviour driven discrete choice problem, with first preferences allocated before second preferences, cannot be achieved by this model. However, discounting substitution in the objective function, e.g., through reduced revenues from second-preference flows, first-preference flows are given priority over second preferences [Vaagen et al., 2011, Section 3.4].

With the iterative solution approach described in subsection 7.1 and Figure 3, the TOC–M model can be used in at least two ways: a) It can help to improve train schedule evaluations which were based on averaged customer parameters, for example, in a post-processing step in the iteration (Figure 9). b) It can also be used to optimize (rather than evaluate) the TOCs' decisions with respect to market response, resulting in the full three-level structure. This may be achieved using two nested iterations (Figure 10): Starting with a given tariff schedule, an initial TOC response in terms of a train schedule is found (using, e.g., a previous market response). This schedule is then improved iteratively, taking into account the market response such as satisfied demand (Figure 11). Finally, the optimized train schedule is communicated to the IM that adjusts the tariff schedule, inducing the TOCs to find new responses.
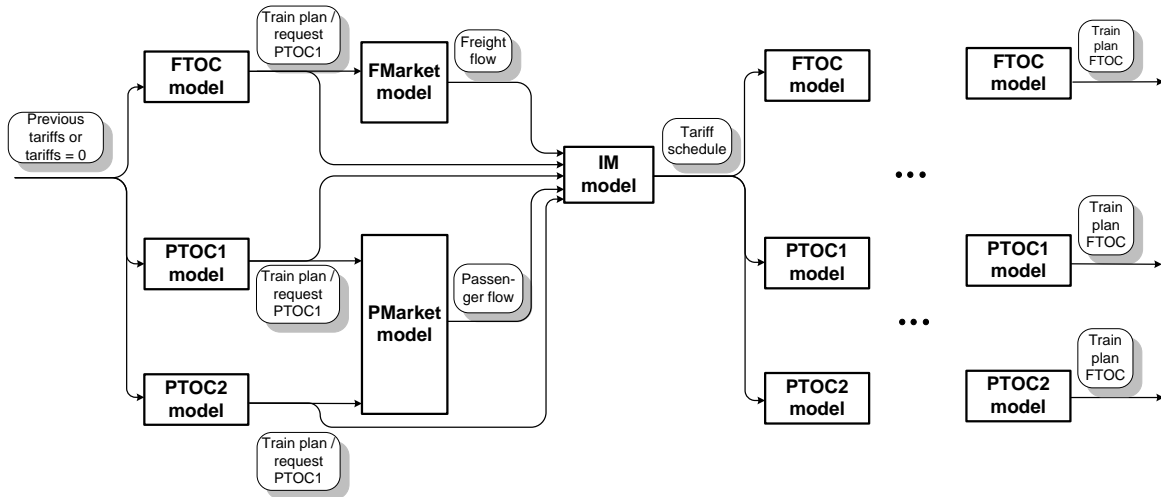
Figure 9: An iterative solution approach for the two-level IM–TOC problem. Market models help evaluating customer utility from the TOCs' train schedules.

A.1 Sets, parameters, and variables

Customer segments $k \in \mathcal{K}$ denote groups of the TOCs' customers with similar preferences with respect to a number of predefined attributes. The sets $\mathcal{K}^C$, $\mathcal{K}^{SP}$, and $\mathcal{K}^{LP}$ describe the customer market segmentation for the cargo TOC, the small passenger TOC, and the large passenger TOC, respectively (one may also distinguish only between sets for cargo and passenger transportation). Traffic choice possibilities $e \in \mathcal{E} = \mathcal{E}^C \cup \mathcal{E}^{SP} \cup \mathcal{E}^{LP}$
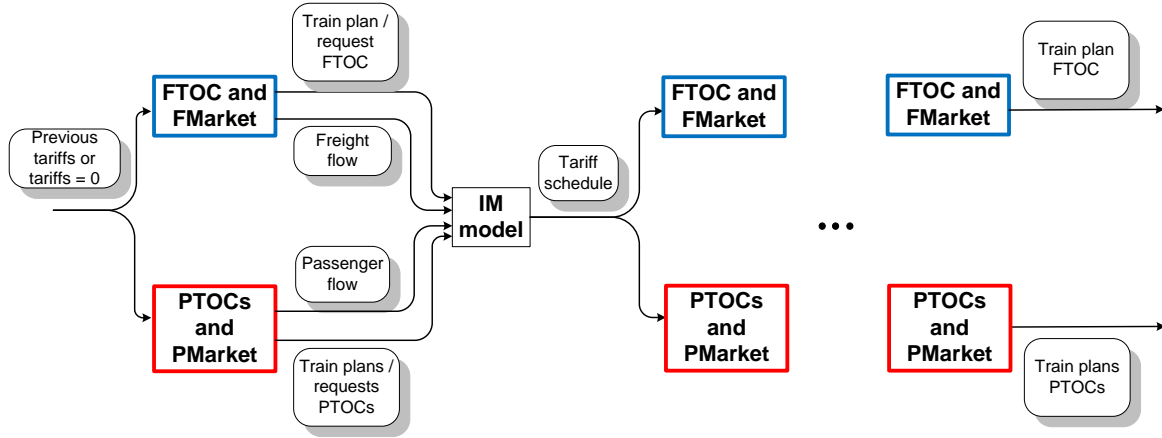
Figure 10: An iterative solution approach for the three-level IM–TOC–M problem with the TOC–M relation parts of the solution process illustrated by blue and red boxes for the cargo and passenger TOCs, respectively.
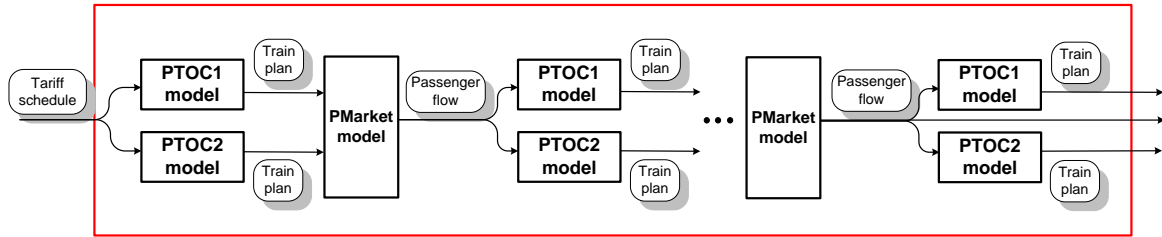

Figure 11: Iterative solution of passenger TOC–M problem with two passenger TOCs serving a common market. The cargo TOC–M problem in the three-level structure is solved in a similar way.

are defined by attributes such as origin or destination, time (departure and arrival times and / or travel duration), or fares. They indicate the various travelling alternatives available to the TOCs' customers, derived from the corresponding train schedules. The index $\mathrm{tr}(e) \in \Xi$ denotes which train provides choice possibility e.

We express uncertainty about the transportation demand of customer segment $k \in \mathcal{K}$ on the single choice possibilities $e \in \mathcal{E}$ by parameters D_{kes} where demand scenarios $s \in Scen$ have a probability $Prob_s$. The substitution probability that a customer of segment k selects choice e' if choice e is not available is denoted by $ProbS^k_{ee'}$.

Simplifying, we specify the socio-economic utility per passenger or TEU according to the defined customer segments k by a coefficient: u^{Dir}_k denotes utility from direct travel, u^{Sub}_k from substituted travel, and u^{Lost}_k (dis)utility from transportation demand lost to the system.[6] A fare π^{Choice}_{ke} applies per passenger or TEU of customer segment k transported on choice possibility e of a TOC's trains and may be distance (and, potentially, time) dependent. The cost c_e denotes the general costs of providing train services on choice e. It is condensed into a single parameter, based on the detailed cost description in the IM–TOC model, and then broken down to each choice e, see the description on page 38 and table 2. The cost c^{Lost}_k denotes (opportunity) costs for lost transportation demand while c^{Term}_{kj} describes terminal or handling costs for (off)loading cargo of customer segment k at terminal j.

In the first formulation DC1 of the discrete choice model, we assume that the train capacity $\widehat{Cap}_e$ available for the single choices is given, and a binary decision variable U_e indicates whether the choice e shall be used. Note that, rather than removing the whole train from the schedule, this only means that the concerned stops defining

---

[6]If no values can be found for u^{Sub}_k, one may derive an approximation through a functional dependency on u^{Dir}_k (e.g., based on the willingness to shift). If even more detailed values are available, a differentiation with respect to choice possibilities $e \in \mathcal{E}$ is conceivable, expressing varying utility with respect to time and distance. For example, a commuter travelling a short distance in the morning may experience a different utility than a commuter travelling a long distance in the evening.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

35 of 56

the choice may not be served. Moreover, this suggestion must be evaluated taking into account all choices using the concerned stops for the considered train. In the second formulation DC2, the train capacity Cap_e is considered a decision variable with an upper bound B_e. The variables f_{kes}^{Dir}, f_{kes}^{Sub}, $\hat{f}_{kee's}^{Sub}$, and f_{kes}^{Lost} describe direct transportation flow (customers using their preferred choice), substituted transportation flow (customers switching over from other choices), both aggregated and separately for each choice, and lost transportation (demand not satisfied in the system), respectively.

A.2 Model formulations

Fixed input capacity formulation – DC1

This model is a single stage stochastic program optimally allocating between direct and substitution travel. The train capacities are fixed input parameters and define an upper bound on the transportation flow. Hence, the model can only suggest to close down or to keep particular choice possibilities. Closed choices and, consequently, unmet demand indicate the extent of migration to competing transport modes. This can trigger discussions about, e.g., the socio-economic impact of traffic migration or the design of incentives (quality, price) to increase customer willingness to shift between different train departure times. The expressions given here complement the TOC models described in sections 4 – 6.

Objective function.   Maximize profit from direct and substitution traffic taking into account losses from lost demand (traffic migration to competing transport modes), for all customer segments and all demand scenarios. Decision variables are the transportation flows f_{kes}^{Dir}, f_{kes}^{Sub}, $\hat{f}_{kee's}^{Sub}$, and f_{kes}^{Lost} and the binary variable U_e.

$$\max \sum_{s \in Scen} Prob_s \sum_{k \in \mathcal{K}} \sum_{e \in \mathcal{E}} \left\{ \left(f_{kes}^{Dir} + f_{kes}^{Sub} \right) \pi_{ke}^{Choice} - f_{kes}^{Lost} c_k^{Lost} - c_e \widehat{Cap}_e U_e \right\} \tag{31}$$

Capacity constraints.   Ensure that total flow of passengers or goods from direct and substitution transportation does not exceed available train capacity on choice possibility e.

$$\sum_{k \in \mathcal{K}} f_{kes}^{Dir} + f_{kes}^{Sub} \leq \widehat{Cap}_e, \quad \forall e \in \mathcal{E}, s \in Scen \tag{32}$$

In this model, $\widehat{Cap}_e$ is considered to be a parameter, and the sum over all capacities $\widehat{Cap}_e$ using a given line segment and a given train $\xi = \text{tr}(e)$ running on a service $s = \text{s}(\xi)$ must not exceed this train's total capacity $\lambda_s^T \kappa_{\text{v}(s)}^{RS}$.

The binary variable U_e allows to suggest to close down a choice e if the expected traffic on this choice falls below a threshold value M_e.

$$\sum_{s \in Scen} Prob_s \sum_{k \in \mathcal{K}} f_{kes}^{Dir} + f_{kes}^{Sub} \geq M_e U_e, \quad \forall e \in \mathcal{E} \tag{33}$$

Total substitution travel   to choice $e \in \mathcal{E}$ from all other choices $e' \in \mathcal{E} \setminus \{e\}$

$$f_{kes}^{Sub} = \sum_{e' \in \mathcal{E} \setminus \{e\}} \hat{f}_{kee's}^{Sub}, \quad \forall e \in \mathcal{E}, k \in \mathcal{K}, s \in Scen \tag{34}$$

Demand constraint.   Total transportation flow of customer segment k on choice e and on other choices e' substituting unmet demand for choice e cannot exceed the demand of this customer segment for this choice e in demand scenario s.

$$f_{kes}^{Dir} + \sum_{e' \in \mathcal{E} \setminus \{e\}} \hat{f}_{kee's}^{Sub} \leq D_{kes}, \quad \forall e \in \mathcal{E}, s \in Scen \tag{35}$$

This equation can be rewritten to express that only unmet demand can be transferred to other choice possibilities. Without this constraint, the model might lead to increased demand, which is not feasible.

Substitution constraint.   Unmet demand for choice e is substituted by transportation on another choice e' only with some probability $ProbS_{ee'}^k$.

$$\hat{f}_{kee's}^{Sub} \leq ProbS_{ee'}^k (D_{kes} - f_{kes}^{Dir}), \quad \forall e', e \in \mathcal{E}, k \in \mathcal{K}, s \in Scen \tag{36}$$

Lost demand   for choice e is the difference between demand for this choice and the total transportation flow of customer segment k on this choice, comprising direct and substituted flows.

$$f_{kes}^{Lost} = D_{kes} - f_{kes}^{Dir} - f_{kes}^{Sub}, \quad \forall e \in \mathcal{E}, k \in \mathcal{K}, s \in Scen \tag{37}$$

Variable capacity formulation – DC2

In this model, the TOC receives feedback about the optimal capacity available for transportation on a choice e. Hence, the capacity Cap_e available on a choice e becomes a decision variable while the parameter $\widehat{Cap}_e$ and the decision variable U_e have been omitted. Note that, here, there is no option to close down a choice; this model optimizes train capacities on all 'active' choices. In addition, there are the decision variables f_{kes}^{Dir}, f_{kes}^{Sub}, $\hat{f}_{kee's}^{Sub}$, and f_{kes}^{Lost} defining the single transportation flows. We mention here only the expressions which are different to model DC1.

Objective function.

$$\max \sum_{s \in Scen} Prob_s \sum_{k \in \mathcal{K}} \sum_{e \in \mathcal{E}} \left\{ \left(f_{kes}^{Dir} + f_{kes}^{Sub} \right) \pi_{ke}^{Choice} - f_{kes}^{Lost} c_k^{Lost} - c_e Cap_e \right\} \tag{38}$$

Capacity constraints.

$$\sum_{k \in \mathcal{K}} f_{kes}^{Dir} + f_{kes}^{Sub} \leq Cap_e \tag{39a}$$

$$Cap_e \leq B_e \tag{39b}$$

$$\sum_{k \in \mathcal{K}} f_{kes}^{Dir} + f_{kes}^{Sub} \geq M_e. \tag{39c}$$

$$\forall e \in \mathcal{E}, s \in Scen$$

The upper bound B_e on the available transportation capacity can be determined by, for example, the total capacity of the train on which choice e is considered, see the explanations after (32).

In addition, the model comprises constraints (34)–(37) of DC1.

A.3   Relations between the IM–TOC and TOC–M models

The solution approach for a three-level IM–TOC–M model sketched in Figures 10 and 11 consists of two nested iteration loops, an outer loop dealing with the IM–TOC relations and an inner loop for the TOC–M relations. When running the outer loop, all information about the TOCs' markets is considered a parameter (i.e., not affected by the solution process). Likewise, the scarcity charges (the IM's decision variables) are considered as constant cost parameters when solving the TOC–M model.

Switching between the loops, some information needs to be submitted between the models: The TOCs' train schedules represent the most important information from the IM–TOC model to the TOC–M model, in addition to operational costs for each train (some of these costs are travel-time dependent and can, hence, not be treated as simple input parameters). In turn, the TOC–M model delivers back detailed information about transportation demands on first and second preferences and lost demand as responses on the single train schedules. In the following, we describe this information exchange and a conceivable adaptation of the models.

Choices	(A,B)	(A,C)	(B,C)	(A,D)	(B,D)	(C,D)
Sections	1	1, 2	2	1, 2, 3	2, 3	3
Aggregated costs	$\frac{c_1}{3}$	$\frac{c_1}{3} + \frac{c_2}{4}$	$\frac{c_2}{4}$	$\frac{c_1}{3} + \frac{c_2}{4} + \frac{c_3}{3}$	$\frac{c_2}{4} + \frac{c_3}{3}$	$\frac{c_3}{3}$

Table 2: Example: distributing aggregated costs to the single choices

From IM–TOC to TOC–M

For each TOC[7], all choice possibilities $e \in \mathcal{E}^{TOC}$ involving this TOC's trains $\xi \in \Xi^{TOC}$ are created. That is, for each train, all sub-routes (origin / destination combinations) are determined and the according departure / arrival times read from the train schedule. Additionally, the train $\mathrm{tr}(e) = \xi$ for choice possibility e is recorded.

For the implementation, it may be more efficient to proceed route-wise rather than train-wise: First, all sub-routes of a route are created and then time information is assigned by iterating over all trains using the considered route, thus creating the choice possibilities. It is also conceivable that the time information of the single choices does not refer to departure and arrival times but to, e.g., departure time and length of travel—or to time intervals such as "morning" or "early afternoon".

The discrete-choice model uses a single parameter for the *aggregated operational costs* on choice e. Often, a choice covers several sections while a section is used by several choice possibilities. Hence, when calculating this parameter from the cost parameters in the IM–TOC model, the costs for the single sections should be accounted for only proportionally. For example, consider a route comprising of four stations A, B, C, D and, consequently, three sections 1, 2, and 3 connecting them. For a train stopping at all these stations, we have the six choice possibilities indicated in the first line of Table 2. The second line shows which sections are covered by each of these choices. Consequently, section 1 is used by three choices, section 2 by four, and section 3 by three. We break down the total costs for a train traversing this route to the single sections, for example, proportional to their length or the travel time, resulting in the costs c_1, c_2, and c_3. The third row in the table shows how these costs are distributed among the six choices.

From TOC–M to IM–TOC

From the TOC–M model, the flows f_{kes}^{Dir}, f_{kes}^{Sub}, and f_{kes}^{Lost} of direct, substituted, and lost traffic for each choice e and customer segment k are read out. If the TOC–M model uses several scenarios $s \in Scen$ to reflect uncertainty about customer demand, aggregated flows are calculated as the expectation over the flows in the single scenarios.

IM's decision problem   Objective function term (15a) expresses the utility of transporting passengers and cargo on all trains using average utility values and estimated transportation demands on each train. With a discrete-choice TOC–M model, this term is replaced by

$$\sum_{s \in Scen} Prob_s \sum_{k \in \mathcal{K}} \sum_{\xi \in \Xi} \sum_{\substack{e \in \mathcal{E}: \\ \mathrm{tr}(e)=\xi}} u_k^{Dir} f_{kes}^{Dir} + u_k^{Sub} f_{kes}^{Sub} + u_k^{Lost} f_{kes}^{Lost}$$

summing up the expected utility of both direct (first preferences) and substitution travel (coming from other choice possibilities) as well as the dis-utility of transportation demand lost to other modes.

TOCs' decision problems   Using the TOC-M model, the revenue π_ξ of operating a train ξ can be specified closer as composed of the expected revenues from first and second choice and lost transportation demand. For cargo, the latter can be interpreted as either a penalty for not being able to deliver on time (or not at all) or a cost associated with outsourcing the cargo transportation to a competitor. For passengers, this is transportation

---

[7]As the passenger TOCs operate in a common market, it may be appropriate to use a "super passenger TOC" comprising all passenger TOCs. The same holds for cargo TOCs.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

38 of 56

Customer segment k	Departure time $ProbS_{ee'}^{k,1}$	$w^{k,1}$	Travel duration $\alpha_{ee'}^{k,2}$	$w^{k,2}$	Travel fare $\alpha_{ee'}^{k,3}$	$w^{k,3}$	Average substitutability $ProbS_{ee'}^{k}$
Business	0.005	0.5	0.005	0.5	0.005	0	0.005
Commuter	0.005	0.6	0.005	0.3	0.4	0.1	0.085
Leisure	0.7	0.1	0.7	0.2	0.8	0.7	0.77

Table 3: Weighted average substitutability. Example for a pair e, e' of choices, three customer segments and three service attributes (departure time, travel duration, and fare)

demand lost to other modes of transport. Hence, expression $\sum_{\xi \in \Xi^C} \pi_\xi$ in the cargo TOC's objective function terms (21a) is replaced by

$$\sum_{s \in Scen} Prob_s \sum_{k \in \mathcal{K}^C} \sum_{\substack{e \in \mathcal{E}^C: \\ \mathrm{tr}(e) = \xi}} \pi_{ke}^{Choice} \left(f_{kes}^{Dir} + f_{kes}^{Sub} \right) - c_k^{Lost} f_{kes}^{Lost}.$$

Similar expressions hold for the passenger TOCs' objective function terms (27a) and (30a), respectively.

For the cargo TOC, some constraints need to be adapted. It is conceivable that freight handling costs at terminals also depend on the type of freight handled, i.e., on the customer segment $k \in \mathcal{K}^C$. This may be due to different amounts of time required for (off)loading operations, cf. Grønland [2011, ch. 4.4]. In this case, the sum in term (21e) is replaced or amended by

$$\sum_{s \in Scen} Prob_s \sum_{k \in \mathcal{K}^C} \sum_{\substack{e \in \mathcal{E}^C: \\ \mathrm{tr}(e) = \xi}} \left(c_{ki_e^{CO}}^{Term} + c_{ki_e^{CD}}^{Term} \right) \left(f_{kes}^{Dir} + f_{kes}^{Sub} \right),$$

distinguishing origin and destination terminals i_e^{CO} and i_e^{CD}, respectively, of the choices e associated with a train ξ. In equation (3), the time for (off)loading cargo of customer segment k is added, taking into account the transportation demand on both first and second customer preferences and terminal handling times. Also constraints (10), (25), and (26) are adapted accordingly.

A.4 Substitutability: parameter estimation

To estimate the substitutability parameters, one may analyse how the considered choice attributes (travel time and duration, fares, available origin–destination pairs) affect customer preferences [Samstad et al., 2010].

The substitutability measure $ProbS_{ee'}^{k} \in [0, 1]$ denotes the probability that, in case it cannot be satisfied, the demand of segment k for choice e may be replaced by demand for choice e'. This measure is the weighted average of the individual substitutability measures $ProbS_{ee'}^{k,sa} \in [0, 1]$ with respect to the single service attributes $sa \in \mathcal{SA}$

$$ProbS_{ee'}^{k} = \frac{\sum_{sa \in \mathcal{SA}} w^{k,sa} ProbS_{ee'}^{k,sa}}{\sum_{sa \in \mathcal{SA}} w^{k,sa}}, \quad \forall k \in \mathcal{K}$$

where $w^{k,sa} \in [0, 1]$ denotes the weight assigned to service attribute sa by customers of segment k. Table 3 gives an example to estimate the aggregated substitutability $ProbS_{ee'}^{k}$ with respect to three service attributes departure time (peak / off-peak), travel duration (speed or variability), and travel fare.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

39 of 56

B Further advanced features

This section contains suggestions to model features which were discussed during the development of the model but were not considered to be given priority when implementing the first prototype. They may, however, become interesting for subsequent versions.

B.1 User financing

The objective function term (15j) and constraints (19) of the IM's decision problem ensure that the total sum of charges collected by the IM is (approximately) matched by the total sum of subsidies dispensed. This is to comply with revenue-neutrality requirements as stated in the EU Directive [European Parliament, 2001] and in "Fordelingsforskriften" [Samferdselsdepartementet, 2003] within the non-dynamic context of our model. It is, however, easy to include user financing for given lines, for example, for necessary infrastructure capacity expansion projects. The IM's decision problem may address this aspect in several ways:

1. The IM may keep a certain pre-specified and fixed sum Z of the collected charges rather than pay out the whole amount as subsidies. In this case, conditions (19b) and (19c) become

$$\sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{v(s)}>0}} \frac{z_{ijp}^{v(s)} t_{\xi ijp}^{S}}{\Delta} \leq (1+\beta) \left(Z - \sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{v(s)}<0}} \frac{z_{ijp}^{v(s)} t_{\xi ijp}^{S}}{\Delta} \right) \tag{40a}$$

$$Z - \sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{v(s)}<0}} \frac{z_{ijp}^{v(s)} t_{\xi ijp}^{S}}{\Delta} \leq (1+\beta) \sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{v(s)}>0}} \frac{z_{ijp}^{v(s)} t_{\xi ijp}^{S}}{\Delta} \tag{40b}$$

   These constraints can then be reformulated to linear expressions similar to (20).

2. The single charges must be at a certain minimum level $\underline{z}$ which changes (19b) and (19c) to

$$\sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{v(s)}>0}} \left(z_{ijp}^{v(s)} - \underline{z} \right) t_{\xi ijp}^{S} \leq - (1+\beta) \sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{v(s)}<0}} z_{ijp}^{v(s)} t_{\xi ijp}^{S} \tag{41a}$$

$$- \sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{v(s)}<0}} z_{ijp}^{v(s)} t_{\xi ijp}^{S} \leq (1+\beta) \sum_{\substack{(i,j),p,s,\xi \\ s(\xi)=s \\ z_{ijp}^{v(s)}>0}} \left(z_{ijp}^{v(s)} - \underline{z} \right) t_{\xi ijp}^{S} \tag{41b}$$

   Also this requires a reformulation as in (20).

3. A pre-determined fee for user financing can be included in the "general" infrastructure usage charges (other than scarcity charges) c_{vr}^{Use} on this line, cf. Gardermobanen. In this case, *all* users of the line have to pay the fee, not only those who intend to travel sections of the line at their busiest time—which are the users benefiting most from a capacity expansion.

B.2 Combining train schedules

Generally, there may be relations between some of a TOC's trains, reflecting, e.g., rolling stock or personnel rostering. This means that changing or even cancelling a train may affect earlier or later trains as well if the same

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

40 of 56

rolling stock or personnel shall be used. However, this issue is located at the TOCs' operational level—while our model is concerned with tactical planning from the IM's point of view. Due to the IM's limited insight into and information about the TOCs' operational planning, train rostering aspects cannot be included in due detail in the model described here. Moreover, constraints defining precedence or other relations must be stated between *pairs* of (potential) trains ("Train A must arrive at X before train B leaves.", "If train A is cancelled, then also train B must be cancelled."). This increases the size of the problem to be solved significantly.

However, completely ignoring relations between trains may lead to solutions where, say, a TOC cancels more trains going into one direction than trains going back. As the model includes *all* train movements on the network—including positioning—such an uneven cancellation should be infeasible. Hence, a very simplified train rostering in the shape of a constraint ensuring directional balance may be necessary. Depending on the considered case, different formulations are conceivable.

As a route describes a directed path between two nodes (stations or terminals), one may require that the number of trains on two routes r_1 and r_2 in opposite directions between these two nodes should be equal over the horizon $\mathcal{P}$ or some shorter time interval,

$$\sum_{p \in \mathcal{P}} \sum_{\substack{\xi \in \Xi: \\ \mathrm{r}(\mathrm{s}(\xi))=r_1}} y_{ijp}^{\xi} = \sum_{p \in \mathcal{P}} \sum_{\substack{\xi \in \Xi: \\ \mathrm{r}(\mathrm{s}(\xi))=r_2}} y_{jip}^{\xi} \tag{42}$$

for some section $(i, j) \in \mathcal{B}_{r_1}$ (and, consequently, $(j, i) \in \mathcal{B}_{r_2}$). (A train does not travel only part of its route. Hence, if it traverses a section on its route it will operate on the whole route.)

Other formulations may combine several routes in the summation or—by way of "conservation constraints"—ensure that the number of trains (of an operator) entering a station equals the number of trains leaving this station. Actually, we need to be concerned only about the trains starting or ending there. The largest time period for which such constraints should hold, is the whole optimization horizon $\{1, ..., P\}$. After all, the schedules are intended to be repeatable. If the conservation constraints do not hold, then rolling stock would be "generated" or vanish at some place. Of course, also shorter time periods (e.g., half a day) are conceivable for these constraints or they may be formulated separately for the single train types, thus reflecting operational planning processes more precisely.

B.3 Periodic schedules ("stive ruter")

A periodic schedule defines a subset of a TOC's trains on the same route $r \in \mathcal{R}$ with a fixed time interval between any two consecutive trains.

The prototype implementation accommodates such train schedules by using pre-created train services. These services are characterized through a fixed time interval between consecutive trains in the service or a given number of trains to be operated (which, then, are spread out evenly). In other words, all specified train services represent periodic schedules per definition.

In this section, we describe a more formal way to ensure periodic schedules which is independent of how the model's input data are organized.

In the strictest form of a periodic schedule, no deviations from the fixed time interval between trains are allowed, neither for departure nor for arrival times at any station. This means that only the timing of the first train ξ_1^n in a given schedule Ξ_n (with n an identifier for the periodic schedule) is determined according to equations (3)–(8). All other trains $\xi_m^n \in \Xi_n \setminus \{\xi_1^n\}$ relate then directly to the y-variables of this train. Assuming the schedule Ξ_n is valid within time periods[8] $[\Phi_n^S, \Phi_n^E]$ and the fixed interval between the trains is d_n time periods long, the schedule comprises

$$M_n = \left\lceil \frac{\Phi_n^E - \Phi_n^S}{d_n} \right\rceil$$

---

[8]This time period refers to the departure station (i_r^{RO}, j_r^{RO}) of the considered route. Consequently, this time period will shift corresponding to the trains' travel (incl. stop and waiting) time.

trains. This number may need to be increased by one if the M_nth train would leave at $\Phi_n^E - d_n$ and the schedule is considered in isolation. The next train would then leave precisely at Φ_n^E which, strictly seen, falls just outside the defined period but commonly is still included. If there is another (periodic) schedule starting at Φ_n^E, this train would be included there and M_n does not have to be incremented. Note that d_n should be chosen such that train departures fall on the same minute in each hour during the schedule's duration, i.e. it should be a divisor of the number of time periods per hour. For example, $\Delta = 5$ gives 12 time periods per hour, such that $d_n \in \{1, 2, 3, 4, 6, 12\}$.

Now, keeping in mind that, for this first train ξ_1^n, expressions (3)–(8) define the $y_{ijp}^{\xi_1^n}$ for all sections $(i, j) \in \mathcal{B}_r$ and all time periods $p \in \mathcal{P}$, the corresponding variables for all subsequent trains in this schedule are given by

$$y_{ij,p+md_n}^{\xi} = y_{ijp}^{\xi_1^n}, \quad \forall\, m = 1, ..., M_n,\ \xi \in \Xi_n \setminus \{\xi_1^n\}$$

which replaces (3)–(8) for these trains ξ.

Observe that this ensures only a fixed time interval between the slots reserved for trains of the periodic schedule, i.e., for the discretized time periods. The intervals between the precise (continuous) times $t_{\xi i}^H$ and $t_{\xi i}^T$ when the trains pass a node i may still vary within the periods, in particular for long periods (large Δ). However, as the $t_{\xi i}^H, t_{\xi i}^T$ are not used for anything else in the model than defining the y—and we are concerned with which slots are occupied by which train rather than with a precise operational planning—the formulation is sufficient for the IM model.

A more precise formulation may start with defining the $t_{\xi i}^H$ and $t_{\xi i}^T$, for example, by

$$t_{\xi i}^H = t_{\xi_1^n i}^{Head} + d_n \cdot \Delta \tag{43}$$

(possibly also redefining d_n to be the actual time units rather than number of periods) and following the steps laid out in (3)–(8). In this case, the inclusion of periodic schedules would increase the number of constraints while the former formulation actually decreases it. However, it allows for a relaxation by defining (43) only for the exit signals of the stopping stations on a route. Hence, it permits a certain variation in the trains' arrival times.

C Notation

C.1 Sets and indexes

Name of set	Description
$\mathcal{A}^{Stop} \subset \mathcal{N} \times \mathcal{N}$	All stopping places (terminals or stations)
$\mathcal{A}^{Stop}_s \subset \mathcal{A}^{Stop}$	Terminals or stations where services of type s stop
$\mathcal{B}_r \subset \mathcal{N} \times \mathcal{N}$	Sections (line segments) on route r
$f \in \mathcal{F}$	All section types (urban, densely populated, ...)
$\mathrm{f}(ij)$	Type of section (i,j)
$g \in \mathcal{G}$	All alternative choice decisions
$l \in \mathcal{L}$	All length classes of trains
$\mathrm{l}(s)$	Length class of trains used in service s
$\mathcal{N}$	Nodes in the network
$\mathcal{N}^D \subset \mathcal{N}$	Dummy nodes
$(i,j) \in \mathcal{N} \times \mathcal{N}$	Section or line segment
$p \in \mathcal{P} = \{0,...,P\}$	Time periods (index) in the optimization horizon
$r \in \mathcal{R}$	All routes (indices)
$\mathrm{r}(s)$	Route traversed by service s
$s \in \mathcal{S}$	All train services
$\mathcal{S}^{Alt}_g \subseteq \mathcal{S}$	Services to choose between in alternative choice decision g
$\mathrm{s}(\xi)$	Service of train ξ
$u \in \mathcal{U}$	All instances ("bestillinger") of public purchases
$v \in \mathcal{V}$	All rolling stock types/configurations
$\mathrm{v}(s)$	Rolling stock type/configuration used for service s
$\xi \in \Xi$	All potential trains to be operated
$\Xi^C \subseteq \Xi$	Trains to be operated by cargo TOC
$\Xi^{LP} \subseteq \Xi$	Trains to be operated by large passenger TOC
$\Xi^{SP} \subseteq \Xi$	Trains to be operated by small passenger TOC

C.2 Parameters

Name	Description
β	Percentage for balancing income and subsidies from scarcity fees
$c^{LocalEff}_{ijpv}$	Costs of negative impacts of rolling stock configuration v using (i,j) at p
c^{Maint}_{ijv}	Maintenance cost of section (i,j), depending on rolling stock configuration v causing the wear & tear
c^{Park}_{jv}	Penalty for parking rolling stock type v at terminal (i,j)
c^{RSL}_v	Distance-dependent cost of operating a unit of rolling stock type v (per km)
c^{RST}_v	Time-dependent cost of operating a unit of rolling stock type v (per hour)
c^{Stop}_j	Cost of using a terminal or station (i,j) by a train of service s
c^{TrFix}_s	Fixed costs of running a train of service s
c^{TrHour}_s	Time-based cost of operating a train of service s per hour (independent of rolling stock)
c^{TrMin}_s	Time-dependent cost of operating a train of service s per minute (incl. rolling-stock dependent costs)
c^{Use}_{vr}	Infrastructure usage charges for running a train of type v on route r
c^{Wait}_s	Waiting cost for a train of service s (kr/minute)
$Comp_u$	Compensation by MoT for carrying out public purchase u
D_s	Maximum duration of a train in service s

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

43 of 56

Name	Description
$D_{s,i}$	Maximum duration of a train in service s on a section (i, j)
Δ	Length of a time period $p \in \mathcal{P}$ (in time units, e.g., minutes)
γ_{ij}^v	Time for rolling stock of type v to traverse section (i, j) at standard speed
$(i_r^{RD}, j_r^{RD}) \in \mathcal{N} \times \mathcal{N}$	Last ("destination") section on route r
$(i_r^{RO}, j_r^{RO}) \in \mathcal{N} \times \mathcal{N}$	First ("origin") section on route r
$(i_u, j_u) \in \mathcal{N} \times \mathcal{N}$	Station the public purchase u refers to for arrivals / departures
κ_v^{RS}	Capacity of rolling stock type v (passengers or TEU)
κ_{ijl}^S	Capacity of section between nodes i and j (number of trains of length class l)
L_{js}^{Park}	Time limit for stay at terminal $(i, j) \in \mathcal{A}^{Term}$ until penalty applies
L_{js}^{Wait}	Time limit for a train of service s for waiting on section (i, j)
λ_{ij}^B	Length of section (i, j)
λ_s^T	Number of cars (or car sets) of same type in a train used for service s
λ_v^{RS}	Length of rolling stock type v (car or set of cars)
ω_u	Minimum number of trains to be operated under public purchase u
$\mu_{s(\xi)}^A$	Multiplier profile for arrival of a train of service s at time p
$\mu_{s(\xi)}^D$	Multiplier profile for departure of a train of service s at time p
π_s^{max}	Maximum achievable revenue when operating a train of service s
$[\sigma_s^{DE}, \sigma_s^{DL}] \in \mathcal{P} \times \mathcal{P}$	Time window for departure of trains of service s at origin section $\left(i_{\mathrm{r}(s)}^{RO}, j_{\mathrm{r}(s)}^{RO}\right)$
$[\sigma_s^{AE}, \sigma_s^{AL}] \in \mathcal{P} \times \mathcal{P}$	Time window for arrival of trains of service s at destination section $\left(i_{\mathrm{r}(s)(s)}^{RD}, j_{\mathrm{r}(s)(s)}^{RD}\right)$
λ_{js}^{Stop}	Scheduled stopping time at station or terminal $(i, j) \in \mathcal{A}^s$ for trains of service s
$[\tau_u^E, \tau_u^L]$	Time window for arrival / departure of trains (public purchase) [min]
W_s^{max}	Maximum achievable willingness to pay when operating a train of service s

C.3 Variables

Name	Description	IM	TOC	Comment
$a_{\xi u}^{From}$	Did train ξ arrive at (i_u, j_u) at or after τ_u^E?		x	binary
$a_{\xi u}^{Until}$	Did train ξ arrive at (i_u, j_u) after τ_u^L?		x	binary
c_ξ^{TrTime}	Total time-dependent cost of operating train ξ		x	continuous
$\tilde{h}_{\xi ip}^H$	Head of train ξ passes node i at time p		x	binary
$h_{\xi ip}^H$	Head of train ξ passed node i by time p		x	binary
$h_{\xi ip}^T$	Tail of train ξ passed node i by time p		x	binary
λ_{vik}^{TB}	Distance of tail of a train of rolling stock type v from node k when head of train enters section (i, j)		x	continuous
ν_ξ	Profile multiplier for train ξ (derived from ν_ξ^A and ν_ξ^D) to adjust π_s^{max} or W_s^{max}		x	continuous $(\in [0, 1])$
ν_ξ^A	Multiplier for arrival of train ξ		x	continuous $(\in [0, 1])$
ν_ξ^D	Multiplier for departure of train ξ		x	continuous $(\in [0, 1])$
π_ξ	Actual revenue from operating train ξ		x	continuous
$\pi^{Tariffs}$	Aggregated income from access charges or subsidies	x		continuous
q_ξ	Train ξ crosses $p = P$ time boundary		x	binary
$\bar{r}_s$	Service s is running		x	binary
r_ξ	Train ξ is running		x	binary
$t_{\xi i}^H$	Time when head of train ξ passes node i		x	continuous
$t_{\xi i}^T$	Time when tail of train ξ passes node i		x	continuous

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

44 of 56

Name	Description	IM	TOC	Comment
W_ξ	Willingness to pay when operating train ξ as scheduled	x		continuous
$w_{j\xi}$	Waiting time of train ξ on section (i,j)		X	continuous
$w_{j\xi}^{Park}$	Time spent at terminal $(i,j) \in \mathcal{A}^{Term}$ in excess of L_{js}^{Park}		x	continuous
y_{ijp}^{ξ}	Train ξ uses section (i,j) in time period p		X	binary
z_{ijp}^{v}	Scarcity fee for access to section (i,j) at time p by rolling stock type v	X		continuous
$\hat{z}_{ijp}^{v}$	Auxiliary variable for calculating $\pi^{Tariffs}$		x	continuous

C.4 Additional notation used in the TOC–Market model

Sets and indexes

Name	Description
$e \in \mathcal{E}$	Choice possibilities
$\mathcal{E}^C \subseteq \mathcal{E}$	Choice possibilities involving cargo TOC's trains
$\mathcal{E}^{LP} \subseteq \mathcal{E}$	Choice possibilities involving large passenger TOC's trains
$\mathcal{E}^{SP} \subseteq \mathcal{E}$	Choice possibilities involving small passenger TOC's trains
$k \in \mathcal{K}$	Customer segments
$\mathcal{K}^C \subseteq \mathcal{K}$	Customer segments for cargo TOC
$\mathcal{K}^{LP} \subseteq \mathcal{K}$	Customer segments for large passenger TOC
$\mathcal{K}^{SP} \subseteq \mathcal{K}$	Customer segments for small passenger TOC
$i_e^{CD} \in \mathcal{N}$	Destination of passenger / cargo flow in choice possibility e
$i_e^{CO} \in \mathcal{N}$	Origin of passenger / cargo flow in choice possibility e
$s \in Scen$	Demand scenarios
$\mathrm{tr}(e) \in \Xi$	Train (index) associated with choice possibility e

Parameters

Name	Description
B_e	Upper bound for finding Cap_e in DC2 (e.g. total train capacity)
c_e	Generalized costs of providing traffic choice e
c_k^{Lost}	Opportunity cost of passengers / cargo of segment k not transported by the TOC (per pax or TEU)
c_{kj}^{Term}	Cost of handling cargo of segment k at terminal j (per TEU)
Cap_e	Subroute capacity available for choice e (DC1)
D_{kes}	Transportation demand of segment k on choice e in scenario s (in pax or TEU)
M_e	Threshold value for closing down a choice e (minimum transportation demand)
$Prob_s$	Probability of demand scenario s
$ProbS_{ee'}^k$	Probability that segment k substitutes choice e by e' if demand cannot be met
π_{ke}^{Choice}	Revenue from transporting passengers / cargo of segment k on choice e (per pax or TEU)
u_k^{Dir}	Socio-economic utility of transporting passengers / cargo of segment k on their first preference (per pax or TEU)
u_k^{Lost}	Socio-economic (dis)utility of lost demand from segment k (per pax or TEU)
u_k^{Sub}	Socio-economic utility of transporting passengers or cargo of segment k on their second preference (per pax or TEU)

Variables

Name	Description
Cap_e	Subroute capacity available on train for choice e (DC2)
f_{kes}^{Dir}	Direct passenger / cargo flow of segment k (choice e is first preference) in scenario s
f_{kes}^{Lost}	Not satisfied ("lost") transportation demand of segment k on choice e in scenario s
f_{kes}^{Sub}	Substituted passenger / cargo flow of segment k (choice e is second preference) in scenario s, aggregated
$\hat{f}_{kee's}^{Sub}$	Substituted passenger / cargo flow of segment k (choice e is second preference), substituting for choice e', in scenario s
U_e	Suggestion to close down choice e (DC1)

D Database structure and data sources

Comments:

- Infrastructure data: 'Jernbaneverkets digitale infrastrukturmodell' is available in OpenTrack and RailML formats from
  `www.jernbaneverket.no/no/Marked/Informasjon-for-togselskapa/Digital-infrastrukturmodell/`.

- All values from Jernbaneverket [2011] are given in 2009-NOK.

- (I) ... integer number or ID, (R) ... positive real number, (S) ... string / text

AlternativeGroups

Specifies which services ('tilbud') to choose from for each group of alternative choices; at most one service has to be chosen from each group, see page 14.
Model parameters: $\mathcal{S}_g^{Alt}$

- (I) **Alternative** Alternative group ID (given in table **Alternatives**)

- (I) **Service** Services to choose between (given in table **Services**)

Alternatives

Lists all instances (groups) of alternative choices.
Model parameters: $g \in \mathcal{G}$

- (I) **Alternative** Unique ID

- (S) **AlternativeName** Description

Commodities

Lists all commodities (or commodity classes) to be transported; a rough classification is sufficient. (Note: the implementation requires a commodity "Person" to identify RS types used for passenger transport.)

- (I) **Commodity** Unique ID

- (S) **CommodityName** Description

CostsLocalExtEffects

Specifies costs of local external effects of train traffic.
Model parameters: $c_{fv}^{LocalEm}, c_{fv}^{Noise}$
Data source: Tables 7.14 and 7.17 in Jernbaneverket [2011]

- (I) **SectionType** Type of section (from table **Sections**)

- (I) **RSType** Type of rolling stock (from table **RSTypes**)

- (R) **Emission** Costs of local emission effects [kr/train-kilometre]

- (R) **Noise** Costs of local noise effects [kr/train-kilometre]

CostsInfraUse

Costs for infrastructure usage.

Model parameter: c_{vr}^{Use}

Data source: Pages 101–103, Network statement [Jernbaneverket, 2012]

- (I) **RSType** Rolling stock type (from table **RSTypes**)

- (I) **Line** Line on which the costs apply (from table **Lines**, assigned to routes using this line in the implementation)

- (R) **UsageCost** Charge for using the infrastructure with a train of the given type [kr/trainkm]

Note: for cargo trains with axle weight below 25t, the standard charges for wagon loads are given per gross ton km.[9] This may, however, be transformed to train-kilometres, e.g., by assuming some "standard" load per train.

LengthClasses

Categories of lengths of trains using the network.

Model parameters: $l \in \mathcal{L}$

Data source: JBV's Capacity report [Jernbaneverket, 2010].

- (I) **ClassID** Unique ID (lower numbers for longer trains)

- (R) **MaxLen** Length of the train [m]

- (S) **ClassName** Description

Lines

Lists all lines on the network (e.g., Hovedbanen, Gardermobanen).

- (I) **Line** Unique ID

- (S) **LineName** Name or description

Links

Links are directed sections.

- (I) **Link** Unique ID

- (S) **LinkName** Description

- (I) **Section** Section the link is defined on (given in table **Sections**)

- (I) **NodeFrom** Start node (index, given in table **Nodes**)

- (I) **NodeTo** End node (index, given in table **Nodes**)

---

[9]Unit is missing for trains with axle load over 25t.

Nodes

Lists all nodes (points where sections meet) on the network.
Model parameters: $i, j \in \mathcal{N}$

- (I) **NodeID** Unique ID

- (S) **NodeCode** Abbreviated name

- (S) **NodeName** Description (full name)

- (R) **StopCost** Cost for using the station (applies only on nodes on GMB) [kr/train]
  (model parameter: c_j^{Stop})
  Data source: Network statement [Jernbaneverket, 2012, p. 103/104]

Operators

Lists all train operating companies.

- (I) **Operator** Unique ID

- (S) **OperatorName** Description

Periods

Defines all time periods for the IM's problem, e.g., 24 hours when the problem horizon is a day.
Model parameters: $p \in \mathcal{P}$

- (I) **Period** Unique ID

- (I) **PeriodStart** Time (e.g., minute) at which the period starts

- (I) **PeriodEnd** Time at which the period ends

Profiles

Lists all profiles of multipliers which may modulate the maximum possible income or consumer willingness to pay per train depending on its departure or arrival time (includes both departure and arrival profiles).

- (I) **ProfID** Unique ID

- (S) **Description**

ProfMult

Defines each income or willingness-to-pay profile through multipliers for each time period.
Model parameters: μ_{sp}^A and μ_{sp}^D (coupling against the services s happens in table **Services**).
Data source: We received some travel data from Flytoget AS to construct examples of such profiles for the Flytoget lines Gardermoen–Oslo S and retour. For other passenger trains, it is conceivable to utilize values in Samstad et al. [2010] on travelers' valuation of time, comfort and reliability or other socio-economic evaluations although this may require further theoretical work. No data sources or theory have been identified for cargo trains.

- (I) **ID** Profile ID

- (I) **Period** Time period when the multiplier is applied (given in table **Periods**)

- (R) **Mult** Multiplier to modulate maximum possible income or consumer willingness to pay when the train departs / arrives at this time period ($\in [0, 1]$)

PublicPurch

All public purchases of trains (commissions) as specified in the traffic contract between the MoT and NSB.
Model parameters: $u \in \mathcal{U}$
Data source: "Trafikkavtalen" [Samferdselsdepartementet, 2012].

- (I) **Index** Commission ID ("Bestilt togprodukt")

- (S) **Description**

- (I) **Operator** (from table **Operators**) As of 2013, all public purchases are operated by NSB. However, for the sake of generality, the operator may be specified explicitly.

- (I) **Route** Route the commission ("bestilling") refers to
  (model parameter: ρ_u)

- (I) **Repeat** Number of trains comprised by the commission, i.e., minimum number of trains to be operated on the route in the specified time interval
  (model parameter: ω_u)

- (I) **TimeFrom** Start of the time interval for the commission
  (model parameter: τ_u^E)

- (I) **TimeUntil** End of the time interval for the commission
  (model parameter: τ_u^L)

- (I) **Node** Station the time interval for departures or arrivals refers to
  (model parameter: i_u or j_u)

- (R) **Compensation** ("vederlag") from the MoT to the contracted operator for operating all trains in this commission [kr]
  (model parameter: $Comp_u$)

RouteLinks

Defines the (geographical) routes through the links they consist of and in which sequence the links are travelled on.

- (I) **Route** (given in table **Routes**)

- (I) **Index** Sequence in which the links are traversed

- (I) **Link** (given in table **Links**)

Routes

Lists all (geographical) routes on the network (ordered sequence of links).
Model parameters: $r \in \mathcal{R}$

- (I) **Route** Unique ID

- (S) **RouteName** Description

RSTypes

Defines the various rolling stock types used to form trains (we assume that a train is formed of only identical rolling stock), their costs as they arise to the TOC, and the costs of external effects of train traffic arising to the IM. The data source for the costs of external effects distinguishes only between passenger and cargo trains and between diesel and electrical (and not at all for accident costs—these are included here as they seem to be the only parameter not depending on section or train type etc.)

Data sources:

Length, speed, capacity: internet (Wikipedia etc.);

costs arising to the cargo TOC: estimates based on Grønland [2011] (the report uses values obtained from Jernbaneverket and CargoNet, including wages, costs of capital and maintenance of the rolling stock, and energy costs).

costs arising to the IM: Tables 7.15 and 7.12 and page 77 in Jernbaneverket [2011].

- (I) **RSType** Unique ID
  (model parameter: $v \in \mathcal{V}$)

- (S) **RSName** Description

- (R) **CarLength** Length of the rolling stock (car/set) [m]
  (model parameter: λ_v^{RS})

- (R) **Speed** Maximum speed [km/h]

- (I) **Commodity** Commodity type to be transported (given in table **Commodities**)

- (R) **TransportCap** Maximum capacity per car/set (number of passengers or TEU)
  (model parameter: κ_v^{RS})

- (R) **CostPerKm** Running cost per car/set per km travelled [kr/km]
  (model parameter: c_v^{RSL})

- (R) **CostPerHour** Running cost per car/set per hour travelled [kr/h]
  (model parameter: c_v^{RST})

- (R) **Emission** Costs of global emission effects [kr/train-kilometre]
  (model parameter: $c_v^{GlobalEm}$)

- (R) **Maintenance** Maintenance costs [kr/train-kilometre]
  (model parameter: c_v^{Maint})

- (R) **Accident** Accident costs [kr/train-kilometre]
  (model parameter: c^{Acc})

- (S) **Comment**

SectionCapacity

Capacity of each section if only trains of a given length class travel this section at a time period. For a mix of trains of different length classes, the total capacity of the section can be calculated as a linear combination of the single capacities.

Data source: based on JBV's Capacity report [Jernbaneverket, 2010]. The report gives trains pr. block section pr. day. This must be translated to trains pr. section pr. hour.

- (I) **Section** (given in table **Sections**)

- (I) **LengthClass** (given in table **LengthClasses**)
  (model parameter: $l \in \mathcal{L}$)

- (I) **Capacity** Number of trains of this LengthClass accommodated on the section per time period if only this type used the section
  (model parameter: κ_{ijl}^{S})

- (S) **Comment**

Sections

Defines all sections the network consists of and in which sequence they are arranged to form a line ('bane').
Model parameters: $(i, j) \in \mathcal{N} \times \mathcal{N}$
Data sources: Length: internet (Wikipedia), MaxSpeed: Annex I, 3.3.2.4 "Line Speed", Network Statement [Jernbaneverket, 2012].

- (I) **Section** Unique ID

- (S) **SectionName** Description

- (R) **Length** Length of the section [km]
  (model parameter: λ_{ij}^{B})

- (R) **MaxSpeed** Maximum speed the section can be travelled on, regardless of rolling stock type [km/h]

- (I) **Line** Line the section belongs to (given in table **Lines**)

- (I) **Index** Sequence number of the section on the line

- (I) **SectionType** Type of section (1: "City", 2: "Other urban region", 3: "Rural region")
  (model parameter: f(ij))

SectionTypes

Defines all types of sections.
Model parameters: $f \in \mathcal{F}$

- **SectionType** Unique ID

- **Description**

Services

Defines all services ('tilbud'). Services may consist of several similar trains traversing the service's route during a given time interval (evenly distributed). IncPerTrain, IncPerCar, and IncPerCap can be combined (are not mutually exclusive); likewise, CostPerTrain and CostPerHour can be combined.
Model parameters: $s \in \mathcal{S}$
Data sources:
Time tabling data for passenger trains: from time tables, for cargo trains: examples (no data sources identified).
IncPerCap for passenger trains: estimates based on ticket prices
Other costs, income and utility data: examples with a magnitude such as to correspond to other model data. Values may be derived from tables 7.6 (income passenger traffic), 7.7 and 7.8 (material costs) and p. 73 (personnel costs) in Jernbaneverket [2011]. Also Samstad et al. [2010] or [Jernbaneverket, 2011, ch. 7.2] may contain data to estimate customers' willingness to pay.

- (I) **Service** Unique ID

- (S) **ServiceName** Description

- (I) **Route** Geographical route the service is defined on (given in table **Routes**)
  (model parameter: $r(s)$)

- (I) **RSType** Rolling stock type used for trains on this service (given in table **RSTypes**)
  (model parameter: $v(s)$)

- (I) **Cars** Number of cars of the rolling stock type in a train on the service
  (model parameter: λ_s^T s)

- (R) **StartTime** Earliest starting time of a train on this service at the departure node [min]
  (model parameter: σ_s^{DE})

- (R) **MaxPostpone** Maximum time the start of (first) train of the service can be postponed after StartTime [min]
  (model parameter: used to calculate σ_s^{DL})

- (I) **Repeat** How many trains in the service

- (I) **Period** Time interval between consecutive trains in the service [min]

- (I) **Operator** Operator of the service (given in table **Operators**)

- (R) **MaxTime** Maximum allowed travel time for a train on the service [min]
  (model parameter: D_s)

- (I) **Commodity** Commodity to be transported (given in table **Commodities**)

- (R) **CostPerTrain** Fixed cost of operating a single train in the service [kr]
  (model parameter: c_s^{TrFix})

- (R) **CostPerHour** Time-based cost of operating a single train (driver salary etc.) [kr/h]
  (model parameter: c_s^{TrHour})

- (R) **IncPerTrain** Income per train [kr/train]
  (model parameter: to calculate π_s^{max})

- (R) **IncPerCar** Income per car [kr/car]
  (model parameter: to calculate π_s^{max})

- (R) **IncPerCap** Income per unit of capacity [kr/unit]
  (model parameter: to calculate π_s^{max})

- (I) **IncProfDep** Income multiplier profile to be applied for departure time (given in table **Profiles**)
  (model parameter: selects μ_{sp}^D)

- (I) **IncProfArr** Income multiplier profile to be applied for arrival time (given in table **Profiles**)
  (model parameter: selects μ_{sp}^A)

- (R) **WTPPerTrain** Maximum consumer willingness to pay per train [kr/train]
  (model parameter: to calculate W_s^{max})

- (R) **WTPPerCar** Maximum consumer willingness to pay per car [kr/car]
  (model parameter: to calculate W_s^{max})

- (R) **WTPPerCap** Maximum consumer willingness to pay per unit of capacity [kr/unit]
  (model parameter: to calculate W_s^{max})

- (I) **WTPProfDep** Willingness-to-pay multiplier profile to be applied for departure time (given in table **Profiles**)
  (model parameter: selects μ_{sp}^D)

- (I) **WTPProfArr** Willingness-to-pay multiplier profile to be applied for arrival time (given in table **Profiles**)
  (model parameter: selects μ_{sp}^A)

- (R) **PenWait** Penalty for waiting time, i.e. actual travel time exceeding optimal/shortest travel time (excluding postponement at departure node), per train [kr/h]
  (model parameter: c_s^{Wait})
  Values from table 7.4 in Jernbaneverket [2011] or from p. V in Halse and Killi [2012] may be utilized.

- (S) **Comment**

ServiceWait

Specifies limits on waiting times for each train in a service in nodes along the service's route. These can be used, e.g., to model stopping patterns for passenger trains.
Data source: examples (no data sources)

- (I) **Service** (given in table **Services**)

- (I) **Node** (given in table **Nodes**)

- (R) **MaxWait** Maximum allowed wait for a train of the service at the node [min]
  (model parameter: L_{js}^{Wait})

- (R) **StopTime** Planned stopping time for a train of the service at the node [min] (not included in waiting time)
  (model parameter: λ_{js}^{Stop})

Travel

Specifies travel time on the links of the network, depending on rolling stock types. As links are directed, different speeds can be modeled for the different directions on the same section (e.g., up- and downhill or other conditions). If no travel time is specified here, the model calculates the time based on the section's length (from table **Sections**) and the rolling stock type's standard speed (from table **RSTypes**).
Data source: from time tables; just a few selected routes, times for remaining routes are calculated in the model (using section length and maximum speed and rolling stock standard speed).

- (I) **Link** (given in table **Links**)

- (I) **RSType** (given in table **RSTypes**)

- (R) **TravelTime** Time needed to traverse this link by this rolling stock type [min]
  (model parameter: γ_{ij}^v)

References

S. Dempe. Discrete bilevel optimization problems. Technical report, Institut für Wirtschaftsinformatik, Universität Leipzig, 2001.

S. Dempe. *Foundations of Bilevel Programming*, volume 61 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 2002.

S. Dempe and D. Fanghänel. Bilevel programming with discrete lower level problems. *Optimization*, 58(8): 1029–1047, 2009.

European Parliament. On the allocation of railroad infrastructure capacity and the levying of charges for the use of railroad infrastructure and safety certification. Directive 2001/14/EC, 2001. URL http://europa.eu/legislation_summaries/transport/rail_transport/l24201_en.htm.

E. Gassner. The computational complexity of continuous-discrete bilevel network problems. Technical Report Report 2009–16, Institut für Optimierung und Diskrete Mathematik, Graz University of Technology, 2009.

S. E. Grønland. Kostnadsmodeller for transport og logistikk. Technical Report 1127/2011, Transportøkonomisk Institutt (TØI), Oslo, Norway, 2011.

A. H. Halse and M. Killi. Verdsetting av tid og pålitelighet for godstransport på jernbane. Technical Report 1189/2012, TØI, 2012.

Jernbaneverket. Kapasitetsrapporten, 2010.

Jernbaneverket. Metodehåndbok JD 205. Samfunnsøkonomiske analyser for jernbanen, versjon 3.0, 2011. URL http://www.jernbaneverket.no/no/Marked/Leverandorinfo/Jernbaneverkets-Metodehandbok-i-Samfunnsokonomiske-analyser---JD-205-.

Jernbaneverket. Network Statement 2013. 10th edition, 1 Dec. 2011. Validity period 9 Dec. 2012 – 14 Dec. 2013, 2012. URL http://www.jernbaneverket.no/en/Marked/Informasjon-for-togselskapa/Network-Statement-2013/.

M. Kaut, T. Flatberg, and A. Werner. Heuristic for a bilevel railroad-infrastructure access-pricing problem. Work in progress, 2014.

A.-G. Lium and A. S. Werner. Socio-economic social optimal pricing of railroad infrastructure (SOPJI). Qualitative problem description. Technical Report A23471, SINTEF, Trondheim (Norway), October 2012.

A.-G. Lium, A. S. Werner, and T. Bjørkvoll. Introducing positive and negative Pigovian taxes on railway infrastructure usage. Working paper, SINTEF, 2014.

Samferdselsdepartementet. St.prp. nr. 52 (1999–2000) "Om NSB Gardermobanen AS og oppfølging av NOU 1999: 28 Gardermoprosjektet. Evaluering av planlegging og gjennomføring", 1999. URL http://www.regjeringen.no/nb/dep/sd/dok/regpubl/stprp/19992000/stprp-nr-52-1999-2000-.html?id=203026.

Samferdselsdepartementet. Forskrift om fordeling av jernbaneinfrastrukturkapasitet og innkreving av avgifter for bruk av det nasjonale jernbanenettet (fordelingsforskriften), 2003. URL http://www.lovdata.no/for/sf/sd/sd-20030205-0135.html.

Samferdselsdepartementet. Avtale om utførelse av persontransport med tog som offentlig tjeneste i 2012–2017 "Trafikkavtalen" mellom Samferdselsdepartementet og NSB AS, 2012. URL http://www.regjeringen.no/Upload/SD/Vedlegg/Jernbane/nsvkjopsavtale2012c.pdf.

PROJECT
102000998

REPORT NUMBER
SINTEF A25843

VERSION
1.0

55 of 56

H. Samstad, F. Ramjerdi, K. Veisten, S. Navrud, K. Magnussen, S. Flügel, M. Killi, A. H. Halse, R. Elvik, and O. S. Martin. Den norske verdsettingsstudien. Sammendragsrapport. Technical Report 1053/2010, TØI / Sweco, 2010.

S. Skartsæterhagen. Kapasitet på jernbanestrekninger. Technical report, Institutt for energiteknikk, Norway, 1993. Utarbeidet for NSB Banedivisjonen.

H. Vaagen, S. W. Wallace, and M. Kaut. Modelling consumer-directed substitution. *International Journal of Production Economics*, 134:388–397, 2011. doi: 10.1016/j.ijpe.2009.11.012.