An iterative two-stage heuristic algorithm for a bilevel service network design and pricing model

Christine Tawfik[*1,2], Bernard Gendron[†3] and Sabine Limbourg[‡1]

[1]University of Liège (Uliege), HEC Management School, Liège, Belgium

[2]Zuse Institute Berlin, Applied Algorithmic Intelligence Methods Department, Takustraße 7, 14195 Berlin, Germany

[3]Département d'informatique et de recherche opérationnelle, CIRRELT (Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation), Université de Montréal, Montréal, Canada

August 20, 2021

Abstract

Building upon earlier research, we revisit a bilevel formulation of service design and pricing for freight networks, with the aim of investigating its algorithmic aspects. The model adds substantial computational challenges to the existing literature, as it deals with general integer network design variables. An iterative heuristic algorithm is introduced, based on the concepts of inverse optimization and neighbourhood search. The procedure alternates between two versions of restricted formulations of the model while inducing promising changes into the service assignments. The approach has proven a high performance for all of the considered real-world instances. Its efficiency rests on its ability to deliver results within a close proximity to those obtained by the exact solver in terms of quality, yet within a significantly smaller amount of time, and to land feasible solutions for the large-sized instances that could not be previously solved. In line with the sustainable transport goals, a deeper observation of the transport management side highlights the strategy of the algorithm favouring freight consolidation and achieving high load factors.

Keywords — heuristics; bilevel programming; network design; network pricing; inverse optimization [1]

[*]Corresponding author; christine.tawfik@uliege.be; tawfik@zib.de

[†]bernard.gendron@cirrelt.net

[‡]sabine.limbourg@uliege.be

[1]This is a preprint manuscript. The online published version is available through DOI: 10.1016/j.ejor.2021.07.052

1 Introduction

Planning freight transport networks is a non-trivial process requiring accurate cost assessment and clear insights into the target market. In addition to the efficiency and profitability goals of traditional systems, consolidation-based transport introduces a challenging management environment. As transport modes having large capacities and high operating expenses are involved, integrated planning is required, as well as a high resource utilization. In particular, tactical decisions are regarded as a vital link in the planning process of a freight transport system (Crainic and Laporte, 1997). At the operational level, those decisions deliver a transport plan that is used to determine the day-to-day policies, whereas, at the strategic level, they help evaluating plausible scenarios. The design of the service network is the main concern of the tactical planning, involving service type and routing decisions (Crainic and Laporte, 1997). With the current presence of deregulation and market competition in Europe, pricing decisions have become crucial for several transport industries. This is especially relevant for rail-based freight transport, where operators are typically required to make months-long slots' requests from the infrastructure managers (see Infrabel, 2016 for related cases). In the context of medium-term agreements and goals' tradeoffs, pricing and design issues are closely entwined. Nevertheless, they have mostly been treated separately in the literature (Brotcorne et al., 2008).

Bilevel programs, introduced by Bracken and McGill, 1973, represent a suitable mathematical programming framework, where hierarchical decision-making is relevant. More formally, a subset of the variables in a mathematical program, labelled the upper level, is constrained to assume the optimal value of a second optimization problem, labelled the lower level, while the latter problem is parametrized by the former's remaining variables. The diversity of contributions in bilevel optimization is outlined in Dempe, 2018. A bilevel service network design and pricing model is discussed in Tawfik and Limbourg, 2019a, presenting a path-based multicommodity formulation. The upper level portrays a transport operator - named a leader - that jointly selects the frequency levels of its offered services as well as their associated prices in the quest of profit maximization, whereas, at the lower level, shipping companies - named followers - decide on the flows of their demands to send over the leader's itineraries, or an always available competition's alternative. The presence of this alternative at the lower level is necessary to render the bilevel framework feasible. Mathematically, the model extends the original framework in Brotcorne et al., 2008 in the sense of including general integer design variables, in contrast to open/close decisions, thus complicating the network design part and introducing new computational challenges.

The modelling framework is valid for long-distance consolidation-based transport systems; in particular, it considers the relevant application context of *intermodal transport*. The intermodal freight transport scheme could be defined as a multimodal chain of transport services that links an initial shipper with the final consignee of the shipment, where the transfer between modes takes place at designated terminals/hubs without handling the goods themselves, according to the European conference of ministers of transport, 1997. Generally, environment-friendly transport modes, such as rail or IWW, are being used for most of the travelled route, known as the *main haulage*, and road for the shortest possible parts, to and from the origin and destination terminals respectively, known as the *pre- and post-haulage (PPH)* or *drayage operations*. The competition in this bilevel modelling context is thus represented by an all-road trucking option that is available for each shipping demand at a fixed price. Owing to its environmental advantages and the opportunities it provides to generate economies of scale (Kreutzberger, 2003; Kreutzberger et al., 2003; Mostert and Limbourg, 2016), intermodal transport is considered as a transport scheme with significantly high potentials to endorse sustainability and energy efficiency. Nowadays, the multiple user feature of intermodal hinterland networks is considered as a critical aspect that needs to be accounted for at the network design stage as pointed out by Bouchery et al., 2020.

The bilevel model of Tawfik and Limbourg, 2019a is tested on a subset of real-world data instances, based on the European rail freight corridors. Even though the exact numerical tests exhibit a reasonable optimality gap (no more than 5%) using an off-the-shelf solver, the considered data instances are still significantly reduced in size with respect to the full real-world cases and the computation times suggest a room for algorithmic developments. In this paper, we look more closely on the structure of the problem, with the aim of designing an appropriate solution algorithm. Acknowledging the underlying complexity of bilevel programs - strongly NP-hard (Hansen et al., 1992) - we seek to present an efficient approach that delivers near-optimal solutions for the considered real-world instances.

The main areas of complexity of the model are addressed within a proposed heuristic algorithm: namely, the network design and the lower-level optimality. A heuristic algorithm is thus developed, by building a potentially costly initial service assignment to accommodate the largest possible amount of shipping demands, guided by the competition's prices. An assignment in this sense denotes the weekly frequencies of offering the long-haul transport services, each defined by their origin, destination and transport mode. The algorithm then launches an iterative procedure to explore the neighbourhood of the services' assignment and their corresponding pricing decisions, respecting the optimality conditions of the lower-level problem in the bilevel framework, i.e., the demands are optimally routed. At each iteration, to reduce the computational effort, the problem is further divided so as the approach alternates between two versions of restricted mixed integer programming (MIP) formulations of the bilevel model; in each version, the model is solved with respect to the upper- and lower-level variables, respectively. The latter part is inspired by a concept often referred to as *inverse optimization*. The underlying idea in inverse optimization is to map a feasible solution of an optimization problem to a cost vector for which the given solution is indeed optimal (Ahmadian et al., 2018). This goal could be achieved by introducing minimal perturbations to the cost vector of the optimization problem, so that a given feasible solution becomes optimal (Ahuja and Orlin, 2001). The solution algorithm in Brotcorne et al., 2000, 2008 is indeed centered around that similar idea of optimizing an auxiliary objective while forcing some solution to be optimal.

The designed algorithm is invoked on the full set of real-world data representing the rail freight corridors in Europe, in the interest of assessing its performance. Further solution analysis is performed from a logistics management perspective, in order to elaborate on the algorithm's underlying strategies showing a clear endorsement of the freight consolidation levels, which help the algorithm land efficient solutions within a reasonable computation time. To the best of our knowledge, the proposed algorithm is the first in the literature that handles a bilevel network design and pricing model with general integer design variables.

The remainder of the paper is organized as follows. In Section 1, we revisit the formulation of the joint design and pricing model. In Section 2, we review the literature in terms of the algorithms presented for previous related frameworks and discuss their applicability to the current context. The details of the iterative two-stage heuristic and the computational experiments are presented in Section 3 and 4, respectively. The managerial insights of the approach are covered in Section 5. Closing remarks are given in the last section along with potential future perspectives.

2 Joint service network design and pricing formulation

In Tawfik and Limbourg, 2019a, an underlying physical network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ is considered, with node set $\mathcal{N}$ and arc set $\mathcal{A}$. A node can be regarded as a supply, demand or terminal node where the transshipment between the different modes takes place. $\mathcal{S}$ denotes a set of freight transport services, where each service $s \in \mathcal{S}$ is assigned: a physical arc $a_s \in \mathcal{A}$ in the network; a transport mode m_s (i.e., road, rail or IWWs); maximum allowed units of capacity u_s and a fixed cost f_s of operating service s once in the planning period, typically one week. A set of commodities $\mathcal{K}$ are shipped over the network, where each commodity $k \in \mathcal{K}$ is defined by: an origin and destination pair $(o_k, d_k) \in \mathcal{N} \times \mathcal{N}$; total demand volumes w^k in tonnes; a fixed competition's trucking service tariff R^k to deliver one tonne of commodity k relative to the direct all-road distance separating o_k and d_k and a variable cost v_s^k to transport one tonne of commodity k using service s. A set of feasible itineraries $\mathcal{L}^k$ is generated for each commodity k by a special procedure, where each itinerary $l \in \mathcal{L}^k$ is tantamount to a sequence of services ($l \subseteq \mathcal{S}$). The idea is, for each commodity, to scan all possible services emanating from its origin node for candidate paths. Then, starting from each of those services, the algorithm seeks to append a successor one, whose origin node corresponds to the destination node of the currently considered service. This procedure is iteratively repeated until either the maximum allowed number of services along an intermodal path is reached, or the destination of the current commodity coincides with the destination node of the last service along the path in construction. If the latter case is attained, an intermodal path of the commodity is assumed to be found and added to its set of feasible itineraries. When constructing a typical itinerary, some conditions are to be taken into consideration with respect to avoiding cycles, the total intermodal path length in comparison to its equivalent all-road distance and the maximum number of services on an itinerary. The full details of such a generation procedure are outlined in Tawfik and Limbourg, 2019b.

Additional indicator parameters δ_s^l are defined for each service $s \in \mathcal{S}$ and itinerary $l \in \mathcal{L}^k$ of commodity $k \in \mathcal{K}$, in order to link the services to their corresponding itineraries: $\delta_s^l = 1$, if service s is used within itinerary l (0, otherwise). We distinguish between the variables of the leader and those of the followers, represented by the service provider and the target shippers, respectively. At the upper level, for each service $s \in \mathcal{S}$, the leader's decisions are two-fold: a discrete frequency y_s of running s during the planning period (Z^* denotes the set of non-negative integers, including zero) and a real-valued associated price T_s of transporting one unit (of any) commodity over it. At the lower level, the real-valued variables h_l^k (respectively, z^k) denote the volumes of commodity $k \in \mathcal{K}$ shipped on the leader's itinerary $l \in \mathcal{L}^k$ (respectively, the competition). In this context, the commodities' demands can be split and the flows' bundling is, thus, made possible. Based on the above notation, the joint design and pricing problem can be expressed as a bilevel MIP formulation with bilinear objectives and linear constraints as follows:

$$\max_{T,y,h,z} \quad \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}^k}\sum_{s\in\mathcal{S}}\delta_s^l T_s h_l^k - \sum_{s\in\mathcal{S}} f_s y_s - \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}^k}\sum_{s\in\mathcal{S}}\delta_s^l v_s^k h_l^k \tag{1a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}^k}\delta_s^l h_l^k \leq u_s y_s \quad \forall s \in \mathcal{S}, \tag{1b}$$

$$y_s \in \mathbb{Z}^* \qquad \forall s \in \mathcal{S}, \tag{1c}$$

$$T_s \geq 0 \qquad \forall s \in \mathcal{S}, \tag{1d}$$

where (h, z) solves:

$$\min_{h,z} \quad \sum_{k\in\mathcal{K}}\left(\sum_{l\in\mathcal{L}^k}\sum_{s\in\mathcal{S}}\delta_s^l T_s h_l^k + R^k z^k\right) \tag{2a}$$

$$\text{s.t.} \quad \sum_{l\in\mathcal{L}^k} h_l^k + z^k = w^k \quad \forall k \in \mathcal{K}, \tag{2b}$$

$$h_l^k \geq 0 \qquad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}^k, \tag{2c}$$

$$z^k \geq 0 \qquad \forall k \in \mathcal{K} \tag{2d}$$

As in bilevel programming frameworks, the model is considered from the perspective of the leader. In this sense, the upper-level objective (1) serves to maximize the leader's net profit, expressed as the difference between the collected revenues from the services' prices T_s and the sum of fixed costs to run the services and the variable costs to transport the assigned commodities. The latter component is assumed to comprise the relevant operational costs for the intermodal operators, including the asset balance costs and the unit freight handling costs at the terminal. Note that the cost discounts associated with the flow bundling opportunities are absent in the considered objective. The reason is that it would change the functional form of the upper-level problem, potentially requiring further linearisation steps and adding considerable computational complexity.

We consider a high-level representation of the problem, as the tactical level is the decision scope of interest in this work. This implies a limitation of not explicitly considering asset-balancing constraints in the model. In contrast, we assume the availability of the required quantities of vehicles and corresponding assets related to each transport mode. The exact projection of our itinerary and service frequency to detailed operational decisions are eliminated from this model. This problem could, however, be extended in the future to a more holistic framework involving several decision topologies, where the decisions projections and the resolution of the potential infeasibilities between the layers are explicitly addressed.

The lower-level objective (2), on the other hand, is to minimize the followers' out-of-pocket costs of the selected itineraries, which include the direct monetary costs to purchase the itineraries. Constraints $(1b)$ ensure that the sum of the commodities' volumes sent over a certain service does not exceed this service's capacity, according to its selected frequency. Constraints $(2b)$ express the followers' demand satisfaction requirement. Finally, $(1c)$ - $(1d)$ and

$(2c)$ - $(2d)$ define constraints on the decision variables of both levels.

Reformulation Following previous authors' methodologies (Labbé et al., 1998; Brotcorne et al., 2000, 2001), the traditional idea to reformulate the obtained bilevel model as a single-level MIP is to substitute the lower-level mathematical program by its Karush–Kuhn–Tucker (KKT) conditions as sufficient and necessary optimality conditions. Let $\lambda_k, \forall k \in \mathcal{K}$ denote the dual variables associated with constraints $(2b)$; note that they are not restricted in sign as they are associated with equality constraints. The following single-level bilinear optimization problem is obtained, that we call Joint Design and Pricing (JDP):

$$\max_{T,y,h,z,\lambda} \quad \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}^k}\sum_{s\in\mathcal{S}}\delta_s^l T_s h_l^k - \sum_{s\in\mathcal{S}} f_s y_s - \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}^k}\sum_{s\in\mathcal{S}}\delta_s^l v_s^k h_l^k \tag{3a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}}\sum_{l\in\mathcal{L}^k}\delta_s^l h_l^k \le u_s y_s \qquad \forall s\in\mathcal{S}, \tag{3b}$$

$$\sum_{l\in\mathcal{L}^k} h_l^k + z^k = w^k \qquad \forall k\in\mathcal{K}, \tag{3c}$$

$$\lambda_k \le \sum_{s\in\mathcal{S}}\delta_s^l T_s \qquad \forall k\in\mathcal{K}, \forall l\in\mathcal{L}^k, \tag{3d}$$

$$\lambda_k \le R^k \qquad \forall k\in\mathcal{K}, \tag{3e}$$

$$\left(\sum_{s\in\mathcal{S}}\delta_s^l T_s - \lambda_k\right) h_l^k = 0 \qquad \forall k\in\mathcal{K}, \forall l\in\mathcal{L}^k, \tag{3f}$$

$$\left(R^k - \lambda_k\right) z^k = 0 \qquad \forall k\in\mathcal{K}, \tag{3g}$$

$$y_s \in \mathbb{Z}^* \qquad \forall s\in\mathcal{S}, \tag{3h}$$

$$T_s \ge 0 \qquad \forall s\in\mathcal{S}, \tag{3i}$$

$$h_l^k \ge 0 \qquad \forall k\in\mathcal{K}, \forall l\in\mathcal{L}^k, \tag{3j}$$

$$z^k \ge 0 \qquad \forall k\in\mathcal{K} \tag{3k}$$

The lower-level optimality is guaranteed by the primal feasibility $(3c)$, dual feasibility $(3d)$ - $(3e)$ and the complementarity slackness constraints $(3f)$ - $(3g)$. Note that no complementarity is required for constraints $(2b)$ as equality is originally ensured. The resulting formulation is still, however, non-linear, as it comprises bilinear terms in the objective $(3a)$ as well as the complementarity constraints $(3f)$ - $(3g)$. Those terms have undergone a linearisation procedure through additional variables and constraints. The detailed steps of this procedure are discussed in Tawfik and Limbourg, 2019a.

3 Literature review

The principal complexity of the JDP model lies in its combination and inheritance of the properties of two-computationally expensive problem classes: service network design and bilevel programming. Whereas several solution algorithms have been previously proposed for each of these types of problems separately, few authors, as mentioned before, considered a joint service design and pricing optimization problem within a bilevel programming framework. The previous algorithmic advancements in this respect are, consequently, limited in methodologies and testing endeavours, which lack, in turn, to provide a reliable benchmark for results' comparison. The existing algorithmic attempts are covered in this section, in terms of their main points of contribution and drawn conclusions, with the aim of identifying a starting point for a new algorithmic framework.

A lot of work has been directed to solving service network design problems, which extend the traditional network design problem in more specific contexts that vary in their complexity according to the particularities of the applications and the size of the instances. Crainic, 2000 presents a generic framework for service network design in freight transport, with the aim of bridging the gap in the literature concerning modelling efforts in service network design tailored to specific transport modes and the mathematical programming developments in traditional network design formulations. A more recent review of service network design formulations incorporating different decisions, such as services' frequency, mode and routing, is considered in Wieberneit, 2008. Several assets (alternatively, resources) are involved in operating services, e.g., tractors, locomotives, trailers, loading/unloading units and crews. Pedersen et al., 2009 introduce generic service network design models with design-balance constraints that aim at balancing the number of asset units entering and leaving each terminal/node and optimizing the utilization of the coslty and limited resources. Different formulations of the problem have been investigated in the literature, considering arc- and cycle-based models (Andersen et al., 2009, 2011). Andersen et al., 2009 show, by a computational study, that the formulations based on cycle design variables may be solved faster than the formulations based on arc design variables.

As opposed to the deterministic cases, service network design involves in practice several aspects of uncertainties, e.g., in terms of unpredictable demands, travel delays, vehicle breakdowns and congestion. This issue is becoming more pronounced in the era of online shopping, where additional uncertainties related to the locations and purchasing behaviors of a large base of clients significantly contribute to the network planning complexity. In that respect, stochastic service network design has received some attention in the literature. To name a few, Lium et al., 2009 study demand stochasticity within the context of service network design and investigate the structural differences with respect to the solutions based on deterministic demands. Bai et al., 2014 incorporate vehicle rerouting in a stochastic freight service network design model. Ng and Lo, 2016 introduce robust formulations for a ferry transport application, where base assumptions about the passenger demand estimates are considered. An insightful study into the value of deterministic solutions in the stochastic environment of service network design problems is conducted in Wang et al., 2018, by quantitatively evaluating the potentially beneficial structural features to obtain better performing solutions. As the economies grow and the transport activities increase, an awareness is raising to limit the harmful effects of the transport externalities and switch to more environmentally friendly modes than road. A class of green network design problems has emerged as a consequence, with the aim of reducing the environmental impacts of logistics activities. However, at the tactical planning level, most considered decision problems consist of location-routing and allocation models (Dukkanci et al., 2019). Few examples can be noted in relation to service network design for green logistics. Tawfik and Limbourg, 2019b consider a case of path-based formulations of service network design models for intermodal transport, in combination with an assessment of the solutions' environmental impacts within a scenario-based parametric analysis. The two lines of research - green service network design and stochastic modelling - meet in Demir et al., 2016, where travel time and demand uncertainties are considered for a case of intermodal transport in relation to cost, time and emission objectives.

The algorithmic efforts to solve service network design problems spanned a large spectrum of exact and heuristic approaches, as well as their combination, as reviewed in Crainic, 2000 and Wieberneit, 2008. For instance, Andersen et al., 2011 propose a first branch-and-price framework for the mixed-integer formulation of the problem with asset-management constraints. For the capacitated version of the problem, the Lagrangian relaxation framework has been shown as a suitable approach, where one or several groups of constraints are relaxed to obtain a simpler problem, and the resulting bounds were often integrated in a branch-and-bound algorithm (Gendron, 2011). Several speed-up mechanisms are investigated for guided local search algorithm, along with feasibility restoration procedures (Bai et al., 2012). There is no general solution framework that is applicable to all variants of the service network design problem; the choice of appropriate approaches essentially depends on the particular complexity of the considered problem and the included modelling dimensions.

Bilevel pricing problems have been extensively studied as well in terms of solution approaches and formulation's strengthening procedures. However, the majority of such problems consider a bilevel program having, on the upper level, a single set of continuous variables analogous to the prices and, on the lower level, a shortest-path problem (Labbé and Violin, 2013). In the freight-tariff setting context, primal-dual heuristic algorithms have been designed to handle the problem's complexity, taking into account the special structure of the underlying network (Brotcorne et al., 2000, 2001). Graph processing techniques have been studied as well in the literature to reduce the practical size of the original network (Bouhtou et al., 2007; Van Hoesel, 2008). In their bilevel model combining revenue management with operations planning for railroad transport, Crevier et al., 2012 compared procedures to generate high-quality initial solutions to be fed to CPLEX, as well as the impact of adding valid cuts to the formulation. They acknowledged, however, the hindrance of the use of exact algorithms to solve large instances. Within an applicable framework to the telecommunication industry, Brotcorne et al., 2008 developed a generic heuristic solution procedure to handle the reformulated MIP of the bilevel joint pricing and design problem, having binary open/close design variables. The authors extended the previously introduced primal-dual heuristic algorithm in Brotcorne et al., 2000, where, within an iterative procedure, the problem is reformulated as a single-level problem through the use of an exact penalty function applied to the lower-level complementarity term and optimized sequentially with respect to the upper- and lower-level variables. This procedure was implemented in the context of a Lagrangian relaxation framework, treating the linking capacity constraints as the complicating ones. The framework showed an efficient and consistent performance, both on randomly generated and real data. Finally, Ypsilantis, 2016 developed a simple heuristic to address a further variant of this bilevel problem from the perspective of maritime container terminal operating companies, taking advantage of the special case in their considered network where every port to hinterland path can go through one tariff arc. The heuristic was based on the idea of aggregating individual solutions relative to the case where only one corridor controlled by the extended gate operator is allowed to open. By means of solving a set of restricted MIP formulations, feasibility was ensured and higher revenues were sought where possible.

As shown, none of the reviewed approaches can be applied in a straightforward manner to the bilevel program (JDP) in question, as a more generic case is considered and the necessary network simplification assumptions are not carried over to the present case. A heuristic solution approach is therefore devised, inspired by the inverse optimization concept of the primal-dual algorithm in Brotcorne et al., 2000, 2008, in its alternation procedure between two versions of restriced MIPs that consider each problem level variables separately. This procedure is implemented within a larger solution neighbourhood search framework that seeks to construct promising patterns of service design decisions. The inverse optimization technique is first depicted in the case of a multicommodity toll optimization problem (TOP) by Labbé et al., 1998. As mentioned in Brotcorne et al., 2000, the arcs' flows can be assessed through some heuristic procedure, then, the values of these flows are improved to maximize the leader's revenue.

4 An iterative two-stage heuristic solution approach

In order to gain significant efficiency over the traditional methods, this approach addresses in an alternative way the two main areas of complexity in the bilevel design and pricing model, namely: the network design part and the lower-level optimality. This is represented in the services' capacity (constraints 3b) and complementarity slackness (constraints 3f - 3g), respectively. The idea is to start with an initial, potentially costly, services' assignment that is able to accommodate a maximum quantity of the shipping demands. Then, in an iterative manner, this assignment is modified in such a way as to eliminate the services that do not contribute (or have a negative contribution) to the leader's overall profit. This idea is motivated by the concept of improving freight consolidation and tweaking the design of the offered services so as to serve the maximum number of shippers. Furthermore, each iteration, as it contains in itself a bilevel program to be solved, is divided into two steps, each addressing a partial aspect in the problem: a technique that is inspired by the penalty method of the primal-dual heuristic in (Brotcorne et al., 2000, 2001).

In contrast to network pricing problems, where the optimal solution involves all followers selecting the leader's alternative over the competition's, network design and pricing problems requires a solution that best handles the interplay between serving customers and saving costs. Such a solution, in most cases, implies an educated choice of demands to be *deliberately* lost to the competition, in the interest of attaining a higher final profit. This concept is better illustrated by an example. In Figure 1, a network consisting of four nodes is considered, assuming the existence of an authority (leader) owning a single rail service (dashed line) and two trucking services (solid lines) for the pre- and post-haulage (PPH) operations that take place at the beginning and the end of the intermodal transport chain. The market is represented in two shipping demands (followers), Antwerp - Vienna and Antwerp - Curtici, with 840 and 1500 tonnes, respectively. The distances considered in the example are based on the real-life situation, as well as the capacities with 24 and 1500 tonnes for a truck and a train, respectively. Trucking services (thick lines) are assumed to be available as the competition to accommodate all commodities' demands at a fixed price of 0.08 €/tkm.

The leader's revenues are therefore bound by *73920* and *192000 €* for the Antwerp – Vienna and Antwerp – Curtici routes, respectively. Finally, the cost components are taken to be similar to the actual real-life figures (Schroten et al., 2011), as follows:

- *PPH:* fixed costs = 0.00385 €/tkm and variable costs = 0.0016 €/tkm.
- *Rail:* fixed costs = 0.0541 €/tkm and variable costs = 0.0157 €/tkm.

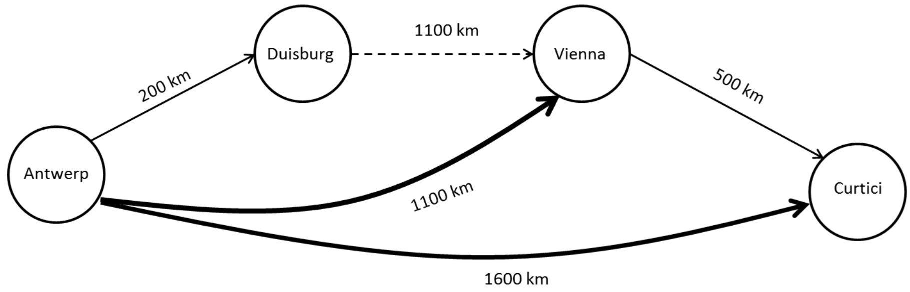


Figure 1: A numerical example

Three possible decisions will be analysed for the leader; commodities are assumed to be non-separable for the sake of simplicity:

1. *Serving both commodities:* This requires offering 2 rail services (Duisburg - Vienna), 98 PPH services (Antwerp - Duisburg) and 63 PPH services (Vienna - Curtici). The total costs approximately amount to 218941 € for rail and 6670 € for PPH, yielding a maximum profit of **40309 €**.
2. *Serving the Antwerp - Vienna commodity:* This requires offering a single rail service (Duisburg - Vienna) and 35 PPH services (Antwerp - Duisburg). The total costs approximately amount to 103771 € for rail and 915 € for PPH, yielding a negative profit in the best case. Therefore, this option is **discarded**.
3. *Serving the Antwerp - Curtici commodity:* This also requires offering a single rail service (Duisburg - Vienna) and 63 PPH services (Antwerp - Duisburg and Vienna - Curtici). The total costs approximately amount to 115170 € for rail and 5754 € for PPH, yielding a maximum profit of **71076 €**.

Therefore, decision 3, though it does not involve serving all demands, results indeed in a higher profit for the leader and qualifies as the optimal solution. The above calculations of the revenues were essentially equal to the upper bound on the competition price for each commodity: an idealistic case that is hardly reached in practice. Similarly, at the beginning of the proposed algorithm, a possibly suboptimal solution, relative to the maximum needed services' assignment is considered and iteratively modified in the interest of optimizing the final profit.

4.1 The algorithm

The procedure starts with an initial solution that corresponds to the original JDP model, with the integral design variables (services' frequency) being relaxed and the long-haul (rail) services' prices being set to an initial value that is bound by the corresponding competition's prices: this value is decided upon within a parameters' calibration procedure, as it is later shown in Appendix A. The problem with the previous restrictions eliminates the complexity coming from the network design and significantly reduces the pricing part. It can hence converge to optimality in a reasonable amount of time. The obtained *continuous* services' assignment for the long haul is then converted to integral values by fixing each service's frequency level to its rounded-up solution value to ensure that all the related shipping demands can be accommodated. The JDP model is (re)solved for optimized services' prices that correspond to the flows' repartition, yielding an initial, potentially costly situation for the leader (Step 0). An iterative procedure is consequently launched, where each iteration is composed of two complementary steps:

- In Step 1, given services' prices and a subset of fixed long-haul services' frequencies, a restricted version of the model penalizing the lower-level optimality conditions is solved for the flows on the itineraries, as well as the remaining services' frequencies.
- In Step 2, the services' prices that simultaneously maximize the leader's profit while maintaining the lower-level optimality are computed, with respect to the previously obtained itineraries' flows.

At the end of Step 2, the offered long-haul services are only fixed throughout the next iteration based on their contribution to the profit margin. The solution associated with the best objective value is memorized and returned when a termination criterion is met, i.e., when the objective has not improved nor undergone significant modification throughout a consecutive pre-defined number of iterations. More concisely, the underlying idea of the proposed approach is to address the bilevel model in an inverse manner by generating itineraries' flows that maximize the leader's profit, within a framework that induces promising changes to the services' assignment. The rationale behind this is, given the computational complexity of the JDP model and the practical inability to solve it for global optimality, we aim at obtaining lower-level solutions (itinerary flows and their corresponding required service levels), having the leader's interests in sight, by solving a slightly altered version of the tractable lower-level problem. Then, at each iteration, we recover the pricing decisions that are compatible with the obtained flows, i.e., making those flows indeed the optimal solution with respect to the original lower level. Through this scheme, we opt to arrive to a near-optimal combination of flows and pricing decisions within a reasonable time. Note that, the long-haul services are the point of interest in the service assignment's optimization, as they represent the significant part of the leader's costs. In what follows, each of the main components of the algorithm is elaborated upon.

Step 0 Following an initial run of the JDP model with relaxed design variables and fixed long-haul services' prices, the model is solved once again to optimize all the services' prices T and λ variables with respect to the resulting itineraries' flows - h and z. Let $\mathcal{S}_{rail} \subseteq \mathcal{S}$ denote the long-haul services that are assumed in our case to be associated with the rail transport mode. The corresponding continuous long-haul services' frequencies - $y_s, \forall s \in \mathcal{S}_{rail}$ - are rounded up and saved for Step 1. The restricted case of the model consequently reduces it to the following linear program (RES).

(RES(h, z))

$$\max_{T,\lambda} \quad \sum_{k\in\mathcal{K}} \sum_{l\in\mathcal{L}^k} \sum_{s\in\mathcal{S}} \delta_s^l T_s h_l^k \tag{4a}$$

$$\text{s.t.} \quad \lambda_k \leq \sum_{s\in\mathcal{S}} \delta_s^l T_s \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}^k, \tag{4b}$$

$$\lambda_k \leq R^k \quad \forall k \in \mathcal{K}, \tag{4c}$$

$$\left(\sum_{s\in\mathcal{S}} \delta_s^l T_s - \lambda_k\right) h_l^k = 0 \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}^k, \tag{4d}$$

$$\left(R^k - \lambda_k\right) z^k = 0 \quad \forall k \in \mathcal{K}, \tag{4e}$$

$$T_s \geq 0 \quad \forall s \in \mathcal{S} \tag{4f}$$

Step 1 Given T and λ that solve (RES) and fixed frequencies $\tilde{y}$ associated with a subset of positive long-haul services $\tilde{\mathcal{S}}_{rail} \subseteq \mathcal{S}_{rail}$, the following (PEN) model computes the itineraries' flows - h and z - and the remaining non-restricted services' frequencies - y - that maximize the leader's profit. This technique is in line with the concept of *inverse optimization*, where we aim at generating a lower-level solution that is potentially with desirable qualities for the leader. In this partial problem, the complementarity slackness constraints of the lower level are penalized by a value N. Note that constraints 3f - 3g are always nonnegative for feasible T and λ.

(PEN($T, \lambda, \tilde{y}$))

$$\max_{y,h,z} \quad \sum_{k\in\mathcal{K}} \sum_{l\in\mathcal{L}^k} \sum_{s\in\mathcal{S}} \delta_s^l T_s h_l^k - \sum_{s\in\mathcal{S}} f_s y_s - \sum_{k\in\mathcal{K}} \sum_{l\in\mathcal{L}^k} \sum_{s\in\mathcal{S}} \delta_s^l v_s^k h_l^k \tag{5a}$$

$$- N \sum_{k\in\mathcal{K}} \sum_{l\in\mathcal{L}^k} \left(\sum_{s\in\mathcal{S}} \delta_s^l T_s - \lambda_k\right) h_l^k - N \sum_{k\in\mathcal{K}} \left(R^k - \lambda_k\right) z^k \tag{5b}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} \sum_{l\in\mathcal{L}^k} \delta_s^l h_l^k \leq u_s y_s \quad \forall s \in \mathcal{S}, \tag{5c}$$

$$\sum_{l\in\mathcal{L}^k} h_l^k + z^k = w^k \quad \forall k \in \mathcal{K}, \tag{5d}$$

$$y_s = \tilde{y}_s \quad \forall s \in \tilde{\mathcal{S}}_{rail}, \tag{5e}$$

$$y_s \in \mathbb{Z}^* \quad \forall s \in \mathcal{S}, \tag{5f}$$

$$h_l^k \geq 0 \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}^k, \tag{5g}$$

$$z^k \geq 0 \quad \forall k \in \mathcal{K} \tag{5h}$$

Step 2 The linear program (RES) is solved for the final services' prices T of the current iteration that maintain the optimality of the previously resulting itineraries' flows - h and z. At the end of this step, the long-haul services' frequencies are being updated for the following iteration. Conversely, the algorithm returns if the stopping criterion is satisfied.

The structure of the algorithm is schematically illustrated in Figure 2 and its main steps are summarized in Algorithm 1. Let $(\grave{y}, h^0, z^0)$ denote the obtained solution from the initial run of the JDP model, i.e., with relaxed frequency variables and fixed long-haul prices. Additionally, let y^0 be assigned the rounded-up values of the obtained frequencies: $y^0 \leftarrow \lceil \grave{y} \rceil$, and $\tilde{y}^0$ the initial frequencies to be fixed that are associated with a subset of long-haul services $\tilde{\mathcal{S}}^0_{rail} \subseteq \mathcal{S}_{rail} : \tilde{y}^0 \leftarrow (y^0_s), \quad \forall s \in \tilde{\mathcal{S}}^0_{rail}$. At Step 0, solving RES$(h^0, z^0)$ yields the solution (T^0, λ^0). The objective value is evaluated as follows: $Z(T, y, h) \leftarrow Th - fy - vh$. The remainder of the iterative procedure is highlighted in Algorithm 1.

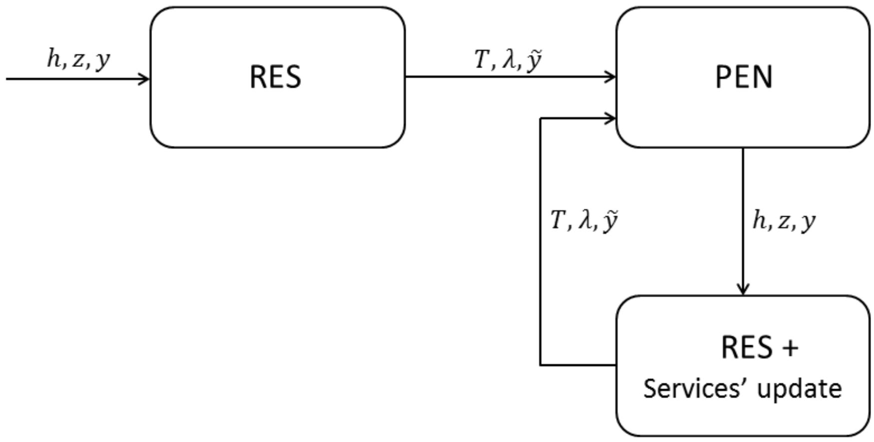


Figure 2: A schematic of the iterative two-stage heuristic

Updating the services' assignment At the end of each iteration, the long-haul services' assignment is updated based on the contribution of each service to the overall leader's profit. This is assessed through the computation of the related profit margin to each of these services. More precisely, for each offered long-haul service, the commodities' itineraries to which this service belongs are identified. The collective revenues are calculated in terms of the prices of the itineraries' services, as well as the total costs in terms of the fixed and variable components. The profit margin is calculated as the ratio of the profit to the costs. If this value falls below a pre-defined minimum threshold (m), the associated long-haul service is no longer fixed with a certain frequency. The remaining long-haul services that pass this test are assigned their current frequency levels throughout the next iteration of the algorithm.

4.2 Perturbation techniques

It is noteworthy that Algorithm 1 seeks solution's improvement in a purely greedy manner, always moving in the direction of solutions with better quality. A worse solution is not allowed to be accepted at any point of the

Algorithm 1 The iterative two-stage heuristic

$Z^* \leftarrow Z(T^0, y^0, h^0)$; $(T^*, y^*, h^*, z^*, \lambda^*) \leftarrow (T^0, y^0, h^0, z^0, \lambda^0)$
$i \leftarrow 1$
while Stopping criterion is not met **do**
Step 1:
  $(h^i, z^i, y^i) \leftarrow$ Solution of PEN$(T^{i-1}, \lambda^{i-1}, \tilde{y}^{i-1})$
Step 2:
  $(T^i, \lambda^i) \leftarrow$ Solution of RES(h^i, z^i)
  if $Z(T^i, y^i, h^i) \geq Z^*$ **then**
    $Z^* \leftarrow Z(T^i, y^i, h^i)$; $(T^*, y^*, h^*, z^*, \lambda^*) \leftarrow (T^i, y^i, h^i, z^i, \lambda^i)$
  end if
  Initialization $\tilde{y}^i \leftarrow \emptyset$; $\tilde{\mathcal{S}}^i_{rail} \leftarrow \emptyset$
  for each $s \in \mathcal{S}_{rail}$ **do** ▷ Updating the fixed service assignment
    if Profit margin of $y^i_s \geq$ Minimum required profit margin **then**
      Add y^i_s to $\tilde{y}^i$
      Add s to $\tilde{\mathcal{S}}^i_{rail}$
    end if
  end for
$i \leftarrow i + 1$
end while
return $(T^*, y^*, h^*, z^*, \lambda^*)$

procedure. Moreover, there is no guarantee that, after the services' assignment updating step, the same services that have been eliminated, will be revisited in such a way as to end up with a previously considered assignment. Therefore, the following techniques are designed in a *metaheuristic* fashion to force the algorithm to explore different services' assignments, possibly leading to solutions with a higher quality, and prevent it from getting stuck at a local minimum.

Avoidance of recurrent assignments The aim of this step is not to return to long-haul services' assignments that have been considered within previous solutions. This is implemented within the (PEN) model solved at Step 1 by keeping track of the sum of fixed costs $\tilde{F}^i$ of each of the previous solutions ($i \in \mathcal{I}$) and adding constraints that ensure that the absolute difference between each of these costs and the fixed costs of the currently considered solution is not null. In theory, duplications of the fixed costs can not be totally avoided, unless by the means of more sophisticated methods (e.g., hashing). Nevertheless, the fixed costs are selected as the basis of the comparison because of the practical improbability of their exact repetition, as they are composed of the distance between the origin and destination of the service in decimal point precision: a criterion that can safely identify a specific service in the majority of practical cases. Let D_i denote the absolute difference between the fixed costs of solution i and those of the currently considered solution. In order to represent this value, two sets of binary variables a_i and b_i are introduced to help design two sets of disjunctive constraints as follows:

$$
\begin{aligned}
0 \leq D_i - \sum_{s \in \mathcal{S}} f_s y_s + \tilde{F}^i \leq C a_i \quad \forall i \in \mathcal{I}, \\
0 \leq D_i - \tilde{F}^i + \sum_{s \in \mathcal{S}} f_s y_s \leq C b_i \quad \forall i \in \mathcal{I}, \\
a_i + b_i = 1 \quad \forall i \in \mathcal{I}, \\
D_i \geq f' \quad \forall i \in \mathcal{I}
\end{aligned}
\tag{6}
$$

where C denotes a *big M* parameter to keep the formulation feasible. In this case, C is represented by the sum of the fixed costs of all long-haul services, multiplied by a maximum frequency value. Since it is not possible to implement strictly greater or less than inequalities in CPLEX solvers, the last set of constraints ensures that D_i have indeed a positive value, where f' depicts a reasonable minimum costs' difference.

Tabooed services Each time a service is being removed from the list of long-haul services' frequencies to be fixed during the iterative procedure due to its inefficiency, its index is kept in a special *taboo list* with a fixed maximum size. The services' indices are added to this list in a queue fashion: first in, first out. The goal of this step is to block using services that were recently deemed inefficient in order to avoid re-visiting earlier services' assignments and explore new solutions. This is implemented also within the (PEN) model, where the design variables y_s associated with the tabooed services are constrained to assume the value zero.

5 Computational tests

In this section, we provide a set of computational tests of the proposed algorithm on real-world instances. Additionally, we provide a comparison with the exact results obtained from the commercial solver, CPLEX, when invoked on the linear reformulation of the JDP model as discussed in Tawfik and Limbourg, 2019a. The demand flows data regarded for these experiments were obtained from Carreira et al., 2012 at the NUTS 2 level, based on the accessible Worldnet database for Europe (Newton, 2009). The instances are defined based on the geographical information provided by RailNetEurope about the rail freight corridors across the continent. In particular, eight data instances are considered based on the nine defined corridors; corridor 5 and 9, relative to the Baltic-Adriatic and Czech-Slovak corridors, are combined in one instance due to their proximity and limited geographical coverage in separate. Only rail and road transport modes are considered to facilitate the results' analysis and comparison with real practices, however, the model can be easily generalized to wider contexts. The costs are compatible to those considered and validated within the BRAIN-TRAINS project (BRAIN-TRAINS, 2014). Experiments have been run on an Intel Xeon CPU ES-2620, 2.10GHz workstation with 32.0 GB RAM and 64-bit Windows 10 Pro. The code is implemented in Java using the IBM ILOG CPLEX 12.6 library as a Branch-and-Bound solver with default parametrisation.

There is a number of key parameters that need fine tuning in order for the algorithm to perform in a robust manner: namely, the services' price (α) set at the initial solution procedure, the profit margin threshold (m) at the assignment updating step and the optimality penalty (N) in the objective of the (PEN) model of the iterative algorithm. α is represented in terms of the ratio of the leader's service's price to that of the competition. In recent approaches, software packages, e.g. **irace** López-Ibáñez et al., 2016, have been used for the purpose of automatically finding the most appropriate settings of an algorithm, given a training set of the problem's instances. The algorithm can later be invoked with the reached configuration on new instances of the same problem. However, the decision was taken to resort to manual parameters' tuning for the present case. On the one hand, due to the scarce availability of real-world data, a sufficient number of instances could only be obtained by means of a complicated generation procedure in order to satisfy the size and the interleaving setup requirements of a realistic intermodal network. This suggests a potential overlap between the training and testing data sets, which leads to a case of an *over-fit* parametrization. On the other hand, the small number of key parameters in the algorithm, as well as the narrowness of their provisional range values makes a manual tuning a feasible endeavour.

The details of the parameters' calibration procedure are documented in Appendix A and the finally reached configuration ($\alpha = 1.0, m = 0.15, N = 10$) is used throughout the following experiments.

5.1 Results and discussion

Table 1 shows the results of the tests invoked on the eight considered real-world instances, corresponding to the rail freight corridors in Europe. $\frac{H}{C}$ denotes the ratio between the objective value obtained by the heuristic algorithm and the best - not necessarily optimal - objective obtained by CPLEX, where the time limit for the latter was set to be six hours. As the algorithm executes in a deterministic manner, no average is taken and the displayed heuristic result is the outcome of a single run.

Instance	No. of nodes	No. of commodities	Two-stage heuristic			CPLEX gap (%)
			No. of iterations	CPU (s)	$\frac{H}{C}$	
Corridor 1	17	270	29	262	0.95	3.29
Corridor 2	19	284	34	76	0.87	6.60
Corridor 3	20	193	45	386	0.93	10.83
Corridor 4	17	166	37	147	0.97	0.98
Corridor 5 and 9	15	204	21	135	0.93	16.72
Corridor 6	15	147	37	346	0.98	0.15
Corridor 7	14	137	33	589	0.81	8.13
Corridor 8	20	271	18	113	0.90	8.10

Table 1: Results of the two-stage heuristic on the rail corridor-based instances

The proposed heuristic exhibits a robust performance on all of the considered instances, with a maximum runtime that does not exceed ten minutes in the worst case. The CPU time is not necessarily directly proportional to the number of iterations nor the size of the problem instance, it rather potentially depends on the algorithm's capability to converge or, conversely, explore new solutions in each case. The heuristic approach also terminates with a solution within a close proximity to that obtained by CPLEX, with an obvious time advantage. Furthermore, Table 2 shows the instances for which CPLEX could previously find an optimal value, i.e., instance 4, 5 and 6 - a narrowed version of corridor 1, 2 and 8 - demonstrating the consistent performance of the proposed approach. Lastly, Table 2 comprises two instances - 7 and 8 - corresponding to the two halves of the network's demands over long distances (> 300km) in the European freight market, as depicted by the Worldnet database (Newton, 2009): a problem size for which CPLEX terminates in six hours with a negative objective value, whereas the heuristic approach lands a solution within both a reasonable CPU time and integral gap with respect to the best found bound.

Instance	No. of nodes	No. of commodities	Two-stage heuristic			CPLEX	
			No. of iterations	CPU (s)	Gap (%)	CPU (s)	Gap (%)
Instance 4	17	40	21	5	10.02	27	0 (opt)
Instance 5	19	44	21	3	12.49	2	0 (opt)
Instance 6	20	58	24	2	11.75	2	0 (opt)
Instance 7	45	849	12	1019	21.08	21600	352.33
Instance 8	43	936	16	950	20.99	21600	917.09

Table 2: Results of the two-stage heuristic on different-sized instances

5.2 Alternative stopping criterion

It is worth noting that, in the majority of the test cases, the number of iterations in the heuristic algorithm is not large. This could potentially be attributed to the selected stopping criterion, as it checks for stagnation in the objective throughout a sufficiently large consecutive number of iterations. In what follows, an alternative stopping criterion is put to the test.

The idea is to observe the change of flows on the itineraries at a certain iteration with respect to the previous one and to terminate the algorithm when this change becomes negligible. More precisely, let $h_l^k(i)$ be the flows of commodity $k \in \mathcal{K}$ on intermodal itinerary $l \in \mathcal{L}^k$ at iteration i, the procedure stops if the maximum relative difference of flows on the itineraries is less than a certain predefined level ε as follows:

$$\frac{|h_l^k(i) - h_l^k(i-1)|}{h_l^k(i)} < \varepsilon, \quad \forall k \in \mathcal{K}, \forall l \in \mathcal{L}^k \tag{7}$$

A range of values of ε is tested in Table 3, where $\frac{H}{C}$ denotes the ratio between the objective value obtained by the heuristic algorithm and the best objective obtained by CPLEX in a maximum runtime of six hours. It is obvious that the algorithm undergoes the most significant changes in the solution quality during its first iterations. A continuation of the procedure due to the chosen stopping criterion does not bring a notable advantage to the reached solution. This observation underlines a room for improving the developed algorithm in terms of the *diversification* and the exploration of the solution's search space. For instance, in addition to the implemented perturbation techniques, more influential operators could be applied on the solution - while keeping/restoring its feasibility - to force examine its neighbourhood, e.g., opening/closing services or re-routing demands on alternative itineraries.

6 Insights into transport management

In this part, we look more closely on the differences between the results obtained by the proposed heuristic approach and those obtained by CPLEX, from a transport management perspective. We start by displaying in Table 4 a

Instance	ε	No. of iterations	CPU (s)	$\frac{H}{C}$
Corridor 1	0.5	45	371	0.95
	1.0	14	141	0.95
	1.5	7	76	0.93
Corridor 2	0.5	135	301	0.87
	1.0	110	210	0.87
	1.5	3	11	0.75
Corridor 3	0.5	115	1000	0.93
	1.0	115	1000	0.93
	1.5	10	119	0.92
Corridor 4	0.5	7	82	0.97
	1.0	7	82	0.97
	1.5	3	74	0.96
Corridor 5 and 9	0.5	68	1000	0.93
	1.0	68	1000	0.93
	1.5	6	77	0.93
Corridor 6	0.5	22	219	0.98
	1.0	22	219	0.98
	1.5	3	42	0.90
Corridor 7	0.5	7	173	0.80
	1.0	7	171	0.80
	1.5	3	83	0.76
Corridor 8	0.5	47	241	0.90
	1.0	47	234	0.90
	1.5	14	94	0.90

Table 3: Relation between the relative change in itineraries' flows and the evolution of the two-stage algorithm

comparison in terms of the resulting modal split for the rail corridor-based instances. The rail modal share is shown, whereas, for more clarity, the road modal share is further differentiated in terms of the all-road and the PPH share.

It is noticeable that the proposed algorithm's strategy to attain high profits is by economizing in offering services, as the intermodal market share (PPH + Rail) of the results obtained by CPLEX mostly dominates those obtained by the new approach. However, this decrease in shares is significantly less for the rail services. The rail modal shares have not declined by more than 2% in most cases. In fact, the rail shares have increased for Corridor 6 and Corridor 8. This observation highlights the weight the algorithm grants to increase the freight consolidation over the rail services. Indeed, this is in line with the filtering condition invoked, at the end of each algorithm's iteration, to check for the long-haul services whose profit margin does not fall below a minimum requirement, in order to be deemed profitable and, thus, fixed throughout the following iteration.

In order to further visualize the differences, the resulting rail services are geographically displayed for three instances: Corridor 3 (Figure 3), Corridor 7 (Figure 4) and Corridor 8 (Figure 5). The NUTS 2 points, equivalently the network's nodes, are shaded on the maps according to each considered corridor. The segments joining the nodes denote the resulting rail services by the invoked solution approaches, and their thickness reflects their corresponding frequencies, as shown by the maps' legend. For simplicity and visibility purposes, the information about the direction of the services is omitted and both outward and return services - when existing - have their frequencies combined in a single illustrated service segment on the maps.

The aspect of a transport corridor is more pronounced for Corridor 3 and Corridor 8, in comparison to Corridor 7, where the network is separated in two clusters. This, potentially, could be due to the distribution of the shipping demands relative to each instance. The illustrated services' density is expected and in line with the observed results from Table 4, as the rail service network is noticeably less *dense* in general for the results obtained by the heuristic algorithm, with respect to those obtained by CPLEX. Nevertheless, some changes at the layout level can also be noted. For instance, we can observe that several low-frequency connections on Corridor 3 and Corridor 8 have disappeared when the heuristic method is applied, and replaced by higher frequencies on longer or more central connections to form longer chains. These observations suggest the ability of the proposed heuristic approach to adopt a searching strategy that favours freight consolidation opportunities, in contrast to offering tailored low-cost services for the long haul. This strategy understandably leads the algorithm to high quality - not necessarily the best - solutions in a shorter time, as is the case from the computational results.

Instance	Two-stage heuristic			CPLEX		
	All-road (%)	PPH (%)	Rail (%)	All-road (%)	PPH (%)	Rail (%)
Corridor 1	50.23	17.86	31.91	46.46	19.14	34.41
Corridor 2	58.02	17.72	24.27	53.00	20.20	26.81
Corridor 3	24.82	26.62	48.56	20.35	28.69	50.96
Corridor 4	18.66	65.97	15.36	16.29	67.69	16.01
Corridor 5 and 9	40.57	21.35	38.08	35.02	24.23	40.75
Corridor 6	32.25	27.27	40.47	33.05	28.64	38.31
Corridor 7	29.60	41.24	29.16	15.93	51.46	32.61
Corridor 8	27.11	27.22	45.67	30.84	27.58	41.58

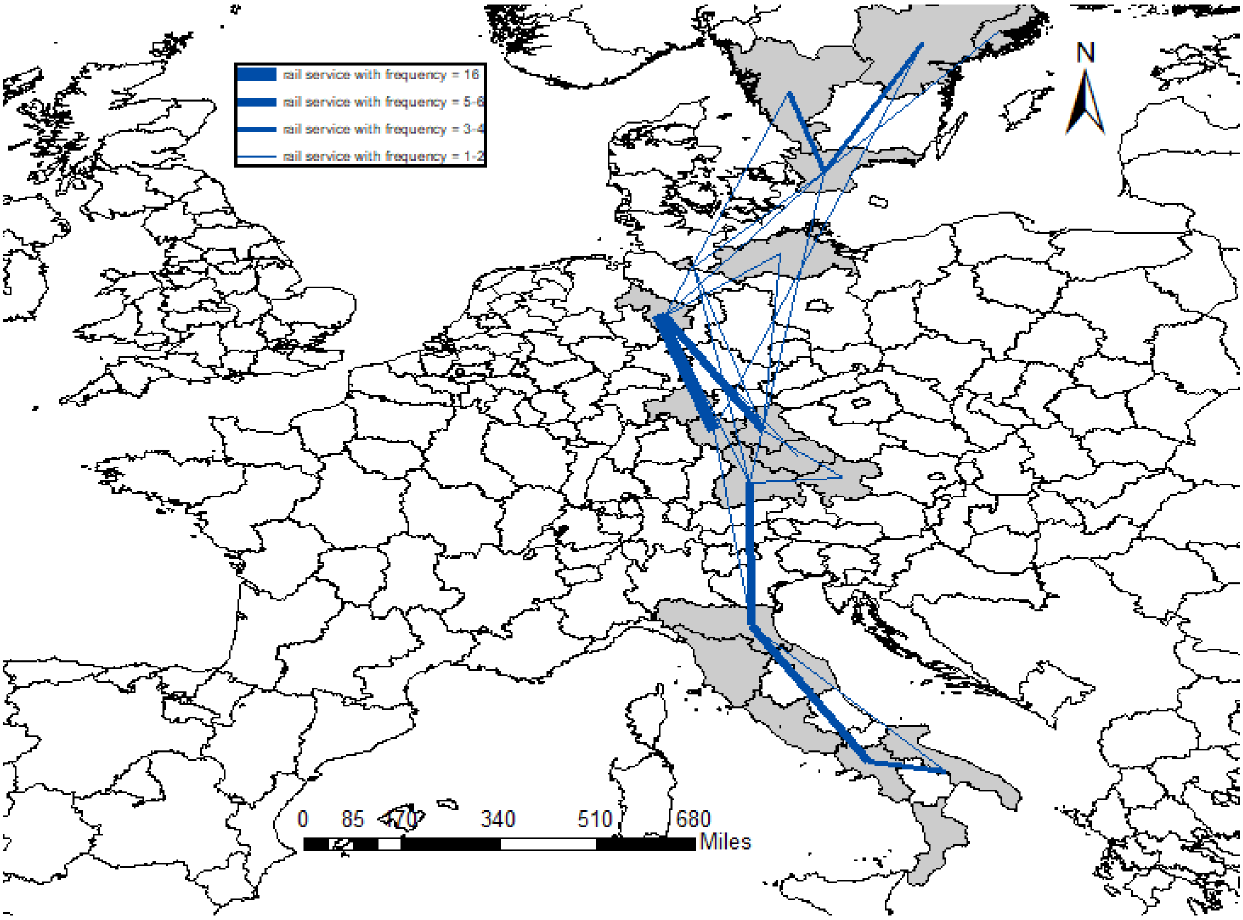
Table 4: Comparison of the modal split results between the heuristic approach and CPLEX

7 Conclusion

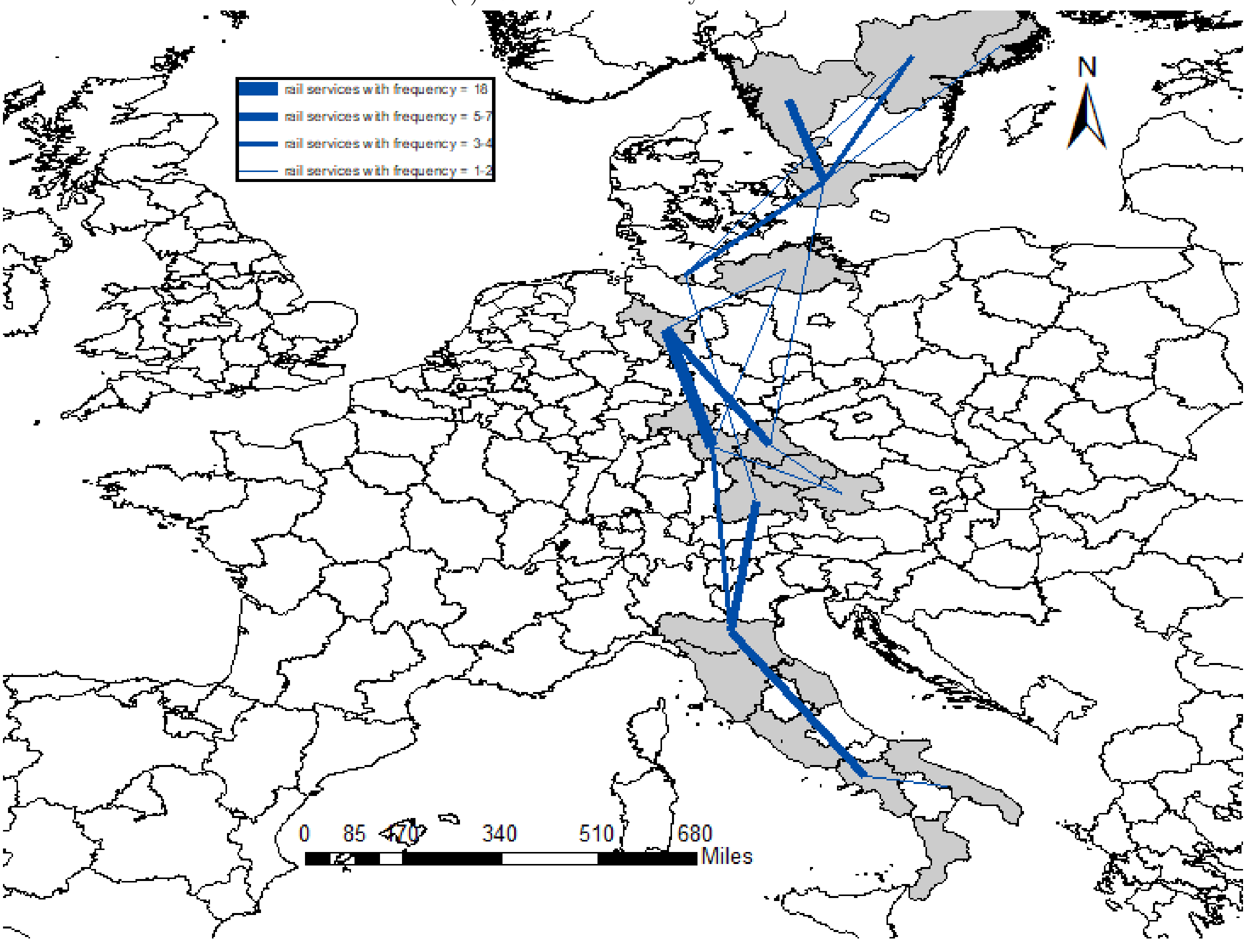
In this paper, a heuristic approach has been discussed with the aim of efficiently solving a bilevel network design and pricing model for consolidation-based transport, presenting a new complexity level to the current literature. An iterative two-stage heuristic has been introduced with the perspective of addressing the two areas of difficulty in the model - the network design and the lower-level optimality - in an alternative manner. More accurately, the algorithm starts with a services' assignment that is able to accommodate a maximum quantity of the shipping demands, then iteratively decrease the frequencies of those services that do not considerably contribute to the leader's profit. Each iteration is divided in two steps: generating the flows that are compatible with the updated services' frequencies while maximizing the leader's profits, then solving for the services' tariffs that guarantee the flows' optimality for the lower level.

The proposed approach demonstrates a superior performance with respect to the mathematical solver CPLEX for the corridor-based instances, in terms of the computation time. The obtained solutions are within a close proximity to the quality of the solutions that are reached by the solver. Furthermore, the developed algorithm is able to land solution within a reasonable gap with respect to the best bound, when invoked on the large-sized instances for which CPLEX terminates with a negative profit: a quality to be appreciated for the real-world problem's context. From a transport management perspective, the obtained results highlight the strategy adopted by the proposed algorithm of increasing freight consolidation in order to maximize the profit, as it was initially intended for it to perform.

Some perspectives could be envisaged to enhance the proposed approach. For instance, the solution stagnation during the early iterations of the algorithm suggests a potential room for improving the diversification force in exploring the search space, e.g., by occasionally manipulating the solution to investigate its neighbourhood, while maintaining its feasibility. With the generalization of the application scope, more relevant data instances could be obtained, and automatic parameters' tuning could consequently be applied. Finally, a computational challenge would have to be considered in the case of the exponential growth of the network's itineraries, which is currently restricted by the application scope of intermodal transport. A promising research direction would be then to decompose the problem by considering gradually growing subsets of the itineraries' space in a column generation fashion.

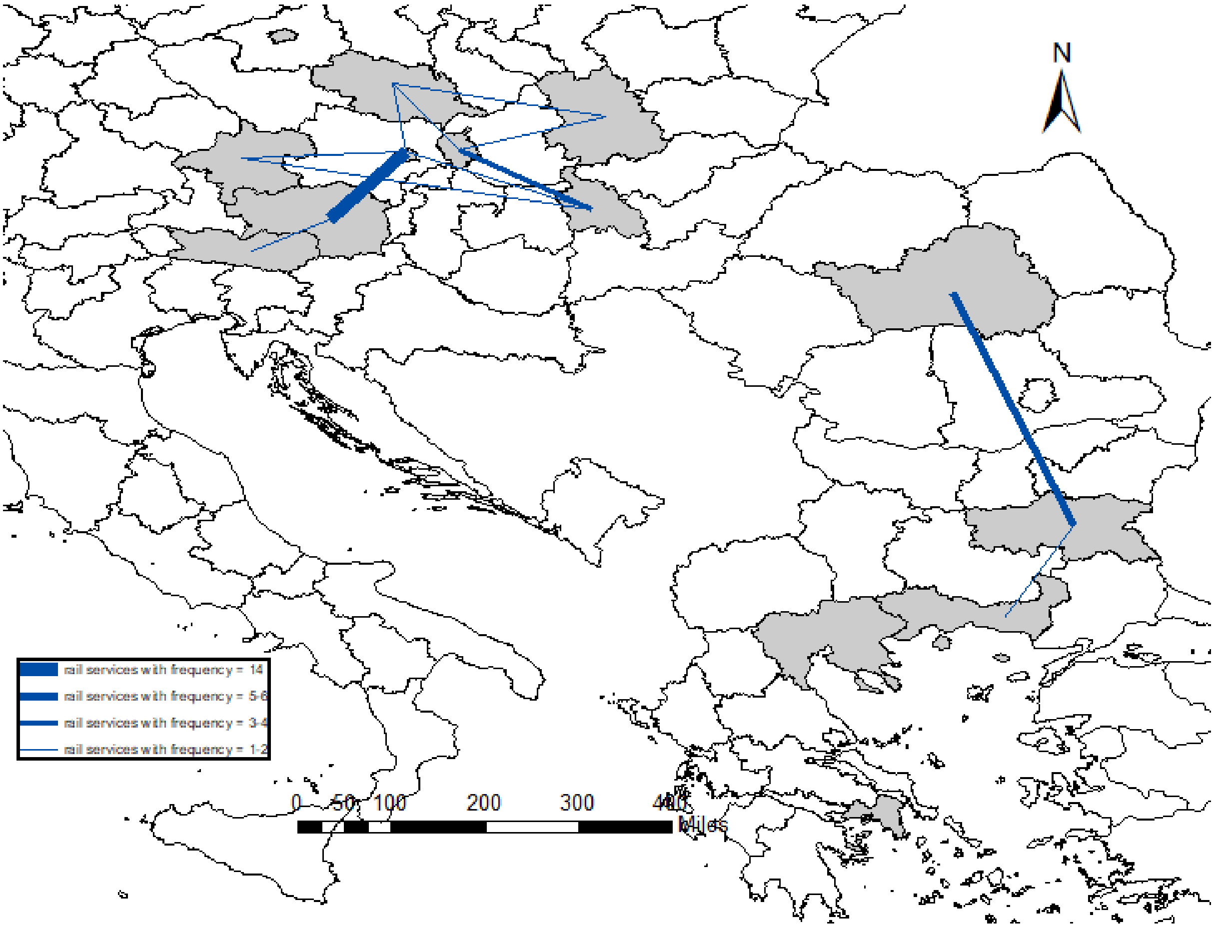


(a) Results obtained by CPLEX

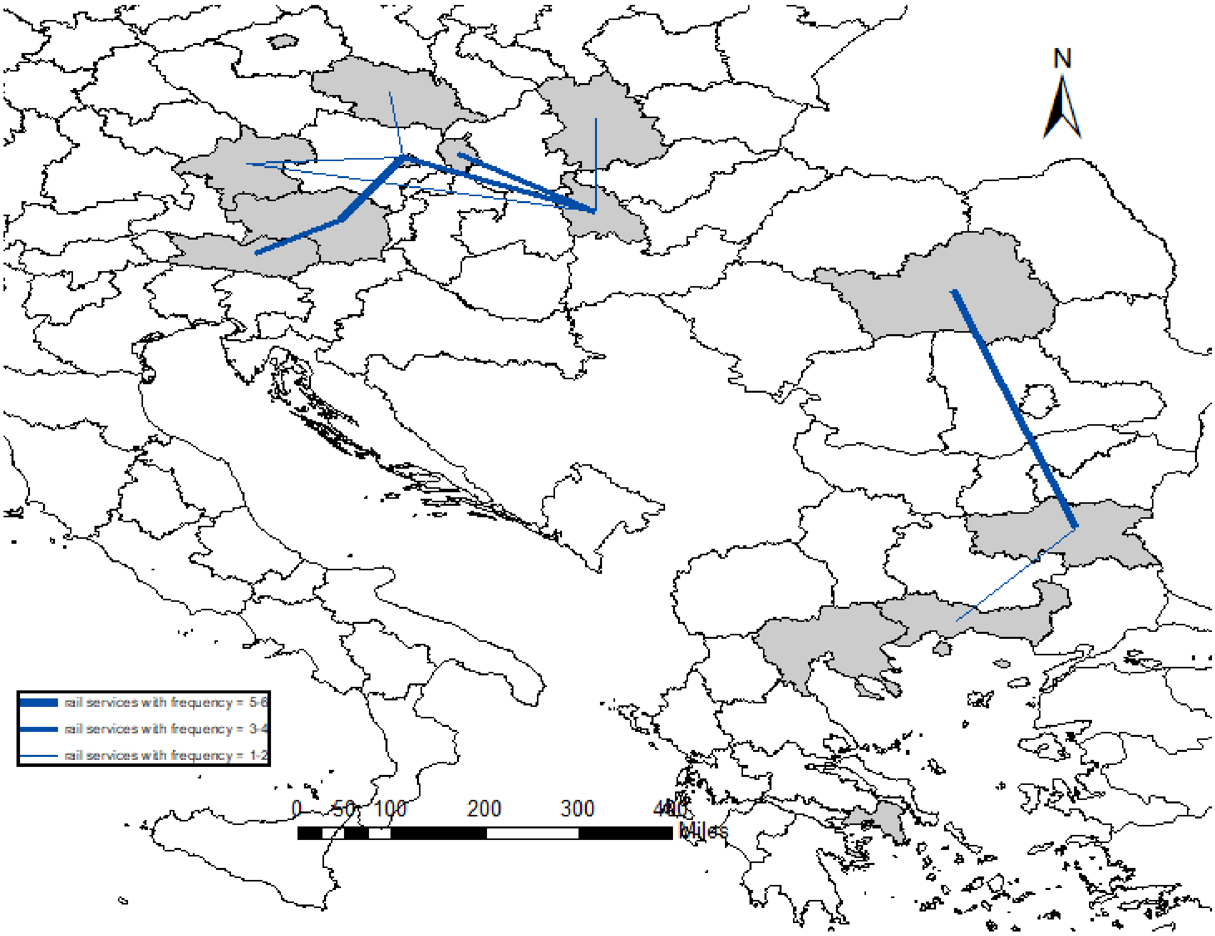


(b) Results obtained by the heuristic approach

Figure 3: The output rail services of Corridor 3

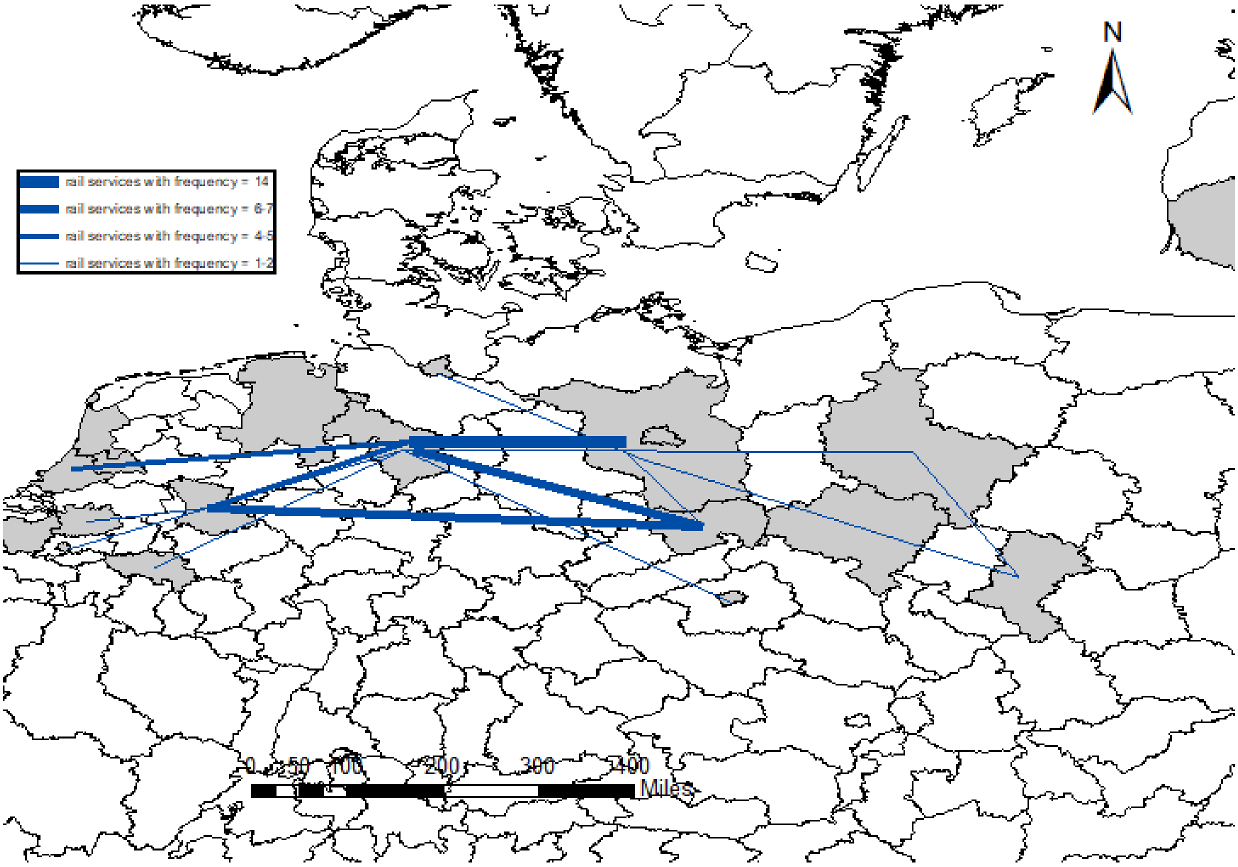


(a) Results obtained by CPLEX

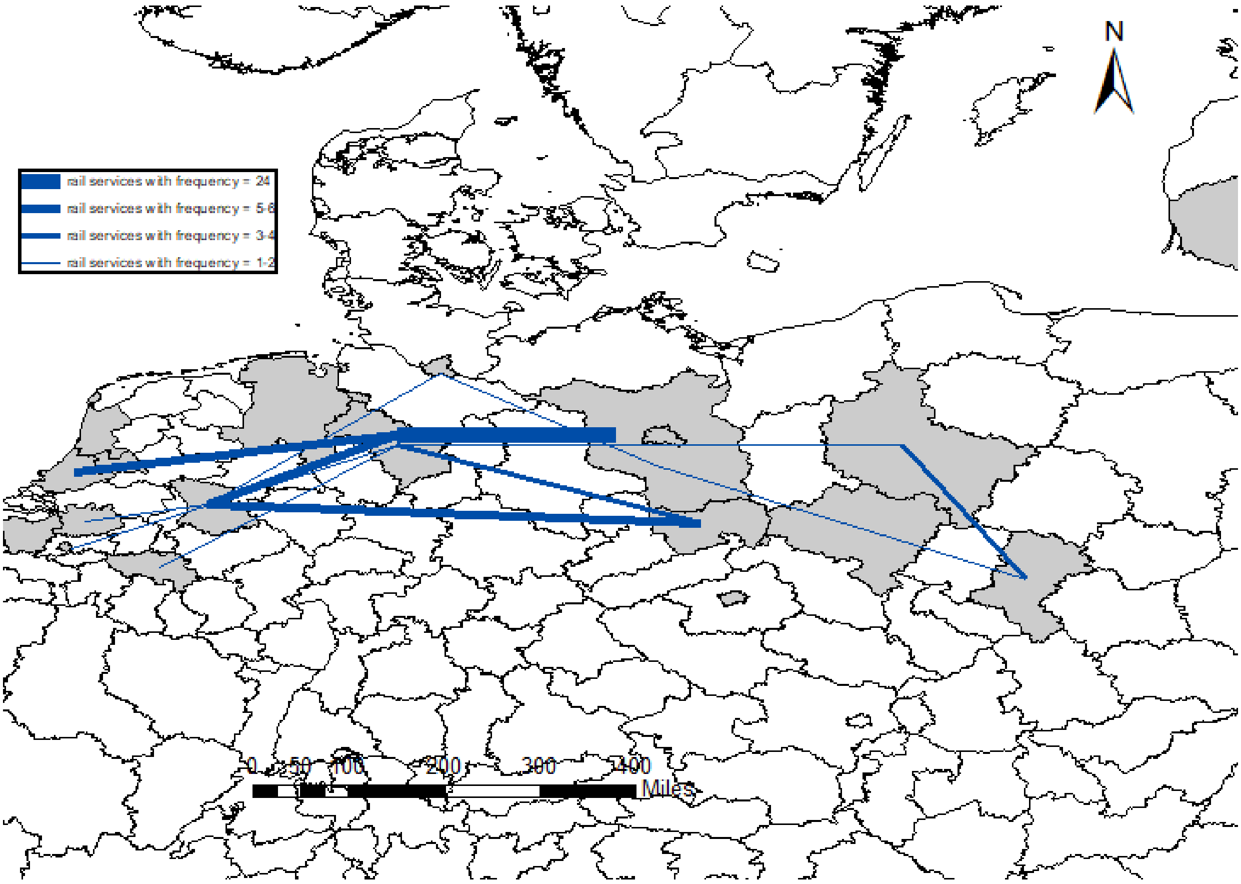


(b) Results obtained by the heuristic approach

Figure 4: The output rail services of Corridor 7



(a) Results obtained by CPLEX



(b) Results obtained by the heuristic approach

Figure 5: The output rail services of Corridor 8

A Parameters' tuning for the iterative two-stage heuristic

The following tests serve to fine-tune the key parameters involved in the iterative two-stage heuristic: namely, the services' price (α) set at the initial solution procedure, the profit margin threshold (m) at the schedule updating step and the optimality penalty (N) in the objective of the (PEN) model.

The calibration is performed in a sequential manner; when treating a parameter, the ones that have not been calibrated yet are set to random initial values ($m = 0.35, N = 100$) and the ones that have been calibrated are set to their updated values. The calibration follows an order of importance of the parameters in the course of the algorithm. The services' price α appears first in the procedure, playing a major role in determining the quality of the starting solution. The profit margin threshold (m) has a direct impact on the improvement step of the algorithm, as it is the filtering basis of the design decisions.

Tables 5 to 7 show the respective results of each of the considered parameters, based on the eight real-world instances that correspond to the rail freight corridors in Europe. In what follows, α and $\frac{H}{C}$ denote the ratio of the leader's service's price to that of the competition and the ratio between the objective value obtained by the heuristic algorithm and the best objective obtained by CPLEX in a maximum runtime of six hours, respectively.

In Table 5, the tested range of values for the α parameter was chosen not to exceed the competition's price, in which case, no flows will be sent over the leader's itineraries. Similarly, in Table 6, the tested values for the m parameter correspond to a range comprised between the maximum and minimum opted for profit margin in real-world cases. As shown in Tawfik and Limbourg, 2019a, the observed profit margin on a subset of real-world instances corresponding to the ones in this work has not fallen below the 20% threshold. Contrarily, in Table 7, as the N parameter does not signify a meaningful economic definition, its range of values has been decided based on multiple initial runs of the algorithm. The final value of each parameter is selected based on a tradeoff decision, balancing between the efficiency of the algorithm (its runtime) and the quality of the obtained solution. The selected parameters' configuration is consequently: ($\alpha = 1.0, m = 0.15, N = 10$).

Instance	α	No. of iterations	CPU (s)	$\frac{H}{C}$
Corridor 1	0.0	76	902	0.85
	0.2	41	424	0.84
	0.4	38	810	0.89
	0.6	22	495	0.88
	0.8	58	645	0.92
	1.0	32	314	0.96
Corridor 2	0.0	45	96	0.79
	0.2	19	50	0.75
	0.4	21	144	0.77
	0.6	21	71	0.80
	0.8	19	46	0.80
	1.0	43	93	0.80
Corridor 3	0.0	25	186	0.69
	0.2	22	211	0.73
	0.4	21	235	0.75
	0.6	37	556	0.87
	0.8	37	477	0.83
	1.0	31	279	0.86
Corridor 4	0.0	33	139	0.85
	0.2	37	284	0.89
	0.4	30	289	0.92
	0.6	21	126	0.92
	0.8	22	126	0.95
	1.0	26	364	0.97
Corridor 5 and 9	0.0	25	113	0.67
	0.2	55	221	0.81
	0.4	26	114	0.83
	0.6	25	108	0.85
	0.8	21	122	0.81
	1.0	29	153	0.70
Corridor 6	0.0	29	292	0.91
	0.2	25	254	0.93
	0.4	57	540	0.93
	0.6	61	557	0.95
	0.8	26	250	0.94
	1.0	98	1000	0.97
Corridor 7	0.0	40	1000	0.81
	0.2	42	800	0.84
	0.4	29	584	0.74
	0.6	21	484	0.76
	0.8	25	570	0.73
	1.0	25	522	0.70
Corridor 8	0.0	23	104	0.81
	0.2	26	117	0.81
	0.4	29	137	0.82
	0.6	18	99	0.85
	0.8	19	169	0.81
	1.0	19	113	0.85

Table 5: Calibration of the initial service price (α)

Instance	m	No. of iterations	CPU (s)	$\frac{H}{C}$
Corridor 1	0.15	33	285	0.95
	0.30	21	185	0.95
	0.45	26	231	0.92
	0.60	18	172	0.92
Corridor 2	0.15	29	60	0.81
	0.30	61	129	0.82
	0.45	22	51	0.80
	0.60	19	43	0.75
Corridor 3	0.15	25	225	0.85
	0.30	34	270	0.84
	0.45	25	233	0.82
	0.60	18	183	0.79
Corridor 4	0.15	41	399	0.97
	0.30	23	361	0.97
	0.45	24	356	0.97
	0.60	26	371	0.97
Corridor 5 and 9	0.15	25	138	0.70
	0.30	25	139	0.74
	0.45	29	153	0.73
	0.60	63	293	0.67
Corridor 6	0.15	33	320	0.97
	0.30	60	567	0.97
	0.45	43	409	0.97
	0.60	59	551	0.97
Corridor 7	0.15	37	739	0.77
	0.30	25	536	0.70
	0.45	25	528	0.72
	0.60	25	508	0.65
Corridor 8	0.15	21	123	0.88
	0.30	19	122	0.85
	0.45	21	129	0.81
	0.60	19	125	0.78

Table 6: Calibration of the profit margin threshold (m)

Instance	N	No. of iterations	CPU (s)	$\frac{H}{C}$
Corridor 1	5	37	333	0.96
	10	29	262	0.95
	15	25	222	0.95
	20	25	230	0.95
	25	25	238	0.95
Corridor 2	5	26	55	0.86
	10	34	76	0.87
	15	22	49	0.83
	20	29	64	0.83
	25	33	61	0.83
Corridor 3	5	21	212	0.94
	10	45	386	0.93
	15	45	389	0.94
	20	58	464	0.95
	25	38	330	0.93
Corridor 4	5	40	142	0.97
	10	37	147	0.97
	15	49	161	0.99
	20	49	160	0.99
	25	21	55	0.96
Corridor 5 and 9	5	25	154	0.92
	10	21	135	0.93
	15	29	189	0.86
	20	37	196	0.89
	25	37	195	0.87
Corridor 6	5	37	347	0.98
	10	37	346	0.98
	15	41	395	0.98
	20	53	498	0.98
	25	33	331	0.98
Corridor 7	5	33	694	0.82
	10	33	589	0.81
	15	25	528	0.82
	20	49	931	0.80
	25	38	700	0.80
Corridor 8	5	18	116	0.88
	10	18	113	0.90
	15	19	119	0.90
	20	18	110	0.90
	25	18	119	0.89

Table 7: Calibration of the optimality penalty (N)

Acknowledgement

The underlying research in this work has been funded by the Federal Science Policy according to the contract n.°BR/132/A4/BRAIN-TRAINS.

References

Ahmadian, S., Bhaskar, U., Sanità, L., and Swamy, C. (2018). Algorithms for inverse optimization problems. In *26th Annual European Symposium on Algorithms (ESA 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Ahuja, R. K. and Orlin, J. B. (2001). Inverse optimization. *Operations Research*, 49(5):771–783.

Andersen, J., Christiansen, M., Crainic, T. G., and Grønhaug, R. (2011). Branch and price for service network design with asset management constraints. *Transportation Science*, 45(1):33–49.

Andersen, J., Crainic, T. G., and Christiansen, M. (2009). Service network design with asset management: Formulations and comparative analyses. *Transportation Research Part C: Emerging Technologies*, 17(2):197–207.

Bai, R., Kendall, G., Qu, R., and Atkin, J. A. (2012). Tabu assisted guided local search approaches for freight service network design. *Information Sciences*, 189:266 – 281.

Bai, R., Wallace, S. W., Li, J., and Chong, A. Y.-L. (2014). Stochastic service network design with rerouting. *Transportation Research Part B: Methodological*, 60:50–65.

Bouchery, Y., Slikker, M., and Fransoo, J. C. (2020). Intermodal hinterland network design games. *Transportation Science*, 54(5):1272–1287.

Bouhtou, M., Grigoriev, A., Hoesel, S. v., Van Der Kraaij, A. F., Spieksma, F. C., and Uetz, M. (2007). Pricing bridges to cross a river. *Naval Research Logistics (NRL)*, 54(4):411–420.

Bracken, J. and McGill, J. T. (1973). Mathematical programs with optimization problems in the constraints. *Operations Research*, 21(1):37–44.

BRAIN-TRAINS (2014). Brain-transversal assessment of intermodal new strategies. belspo project. Http://www.brain-trains.be/.

Brotcorne, L., Labbé, M., Marcotte, P., and Savard, G. (2000). A bilevel model and solution algorithm for a freight tariff-setting problem. *Transportation Science*, 34(3):289–302.

Brotcorne, L., Labbé, M., Marcotte, P., and Savard, G. (2001). A bilevel model for toll optimization on a multicommodity transportation network. *Transportation Science*, 35(4):345–358.

Brotcorne, L., Labbé, M., Marcotte, P., and Savard, G. (2008). Joint design and pricing on a network. *Operations Research*, 56(5):1104–1115.

Carreira, J., Santos, B., and Limbourg, S. (2012). Inland intermodal freight transport modelling. *ETC Proceedings*.

Crainic, T. G. (2000). Service network design in freight transportation. *European Journal of Operational Research*, 122(2):272–288.

Crainic, T. G. and Laporte, G. (1997). Planning models for freight transportation. *European journal of operational research*, 97(3):409–438.

Crevier, B., Cordeau, J.-F., and Savard, G. (2012). Integrated operations planning and revenue management for rail freight transportation. *Transportation Research Part B: Methodological*, 46(1):100–119.

Demir, E., Burgholzer, W., Hrušovskỳ, M., Arıkan, E., Jammernegg, W., and Van Woensel, T. (2016). A green intermodal service network design problem with travel time uncertainty. *Transportation Research Part B: Methodological*, 93:789–807.

Dempe, S. (2018). *Bilevel optimization: theory, algorithms and applications.* TU Bergakademie Freiberg, Fakultät für Mathematik und Informatik.

Dukkanci, O., Bektaş, T., and Kara, B. Y. (2019). Green network design problems. In *Sustainable Transportation and Smart Logistics*, pages 169–206. Elsevier.

European conference of ministers of transport (1997). Glossary for transport statistics.

Gendron, B. (2011). Decomposition methods for network design. *Procedia-Social and Behavioral Sciences*, 20:31–37.

Hansen, P., Jaumard, B., and Savard, G. (1992). New branch-and-bound rules for linear bilevel programming. *SIAM Journal on scientific and Statistical Computing*, 13(5):1194–1217.

Infrabel (2016). Document de référence du réseau.

Kreutzberger, E. (2003). Impact of innovative technical concepts for load unit exchange on the design of intermodal freight networks. *Transportation Research Record: Journal of the Transportation Research Board*, (1820):1–10.

Kreutzberger, E., Macharis, C., Vereecken, L., and Woxenius, J. (2003). Is intermodal freight transport more environmentally friendly than all-road freight transport? A review. In *Nectar conference*, number 7, pages 13–15.

Labbé, M., Marcotte, P., and Savard, G. (1998). A bilevel model of taxation and its application to optimal highway pricing. *Management science*, 44(12-part-1):1608–1622.

Labbé, M. and Violin, A. (2013). Bilevel programming and price setting problems. *4OR*, 11(1):1–30.

Lium, A.-G., Crainic, T. G., and Wallace, S. W. (2009). A study of demand stochasticity in service network design. *Transportation Science*, 43(2):144–157.

López-Ibáñez, M., Dubois-Lacoste, J., Cáceres, L. P., Birattari, M., and Stützle, T. (2016). The irace package: Iterated racing for automatic algorithm configuration. *Operations Research Perspectives*, 3:43–58.

Mostert, M. and Limbourg, S. (2016). External costs as competitiveness factors for freight transport-A state of the art. *Transport Reviews*, 36(6):692–712.

Newton, S. (2009). Deliverable 7. freight flows final. worldnet project (worldnet. worldwide cargo flows) deliverable 7. *Funded by the European Community under the Scientific Support to Policies (Framework 6), http://www. worldnetproject. eu/documents/Public D*, 7.

Ng, M. and Lo, H. K. (2016). Robust models for transportation service network design. *Transportation Research Part B: Methodological*, 94:378–386.

Pedersen, M. B., Crainic, T. G., and Madsen, O. B. G. (2009). Models and tabu search metaheuristics for service network design with asset-balance requirements. *Transportation Science*, 43(2):158–177.

Schroten, A., Van Essen, H., Otten, M., Rijkee, A., Schreyer, C., Gohel, N., Herry, M., and Sedlacek, N. (2011). External and infrastructure costs of freight transport paris-amsterdam corridor. part 1. overview of cost, taxes and levies. Technical report, CE Delft.

Tawfik, C. and Limbourg, S. (2019a). A bilevel model for network design and pricing based on a level-of-service assessment. *Transportation Science*, 53(6):1609–1626.

Tawfik, C. and Limbourg, S. (2019b). Scenario-based analysis for intermodal transport in the context of service network design models. *Transportation Research Interdisciplinary Perspectives*, 2:100036.

Van Hoesel, S. (2008). An overview of stackelberg pricing in networks. *European Journal of Operational Research*, 189(3):1393–1402.

Wang, X., Crainic, T. G., and Wallace, S. W. (2018). Stochastic network design for planning scheduled transportation services: The value of deterministic solutions. *INFORMS Journal on Computing*, 31(1):153–170.

Wieberneit, N. (2008). Service network design for freight transportation: a review. *OR spectrum*, 30(1):77–112.

Ypsilantis, P. (2016). *The Design, Planning and Execution of Sustainable Intermodal Port-hinterland Transport Networks*. PhD thesis, Erasmus University Rotterdam.