Service Network Design for Freight Railway Transportation: the Italian Case*

Guglielmo Lulli† Ugo Pietropaoli‡ Nicoletta Ricciardi§

Abstract

In this paper we present a case-study on freight railway transportation in Italy. This case-study is a by-product of a research collaboration with a major Italian railway company. We highlight the main features of the Italian reality, and propose a customized mathematical model to design the service network, that is, the set of origin-destination connections. More specifically, the model suggests the services to provide, the number of trains traveling on each connection, the number of cars and their type. It considers both full and empty freight car movements and takes handling costs into account. All decisions are taken in order to minimize the total costs. The quality of service is guaranteed by satisfying all the transportation demand and by implicitly minimizing the waiting time of cars at intermediate stations.

Our approach yields to a multi-commodity network design problem with a concave arc cost function. To solve this problem, we implement a specialized tabu search procedure. Computational results on realistic instances show a significant improvement over current practice.

KEYWORDS: rail transportation, service network design, practice of OR, tabu search.

*Part of the contents of this paper appeared in preliminary version in the Proceedings of ATMOS 06.

†Università di Milano "Bicocca", Dipartimento di Informatica, Sistemistica e Comunicazione, Viale Sarca 336, 20126 Milano, Italy. E-mail: `lulli@disco.unimib.it`

‡Université Libre de Bruxelles, Département de Mathématique, Boulevard du Triomphe, B-1050, Bruxelles, Belgium. E-mail: `pietropaoli@disp.uniroma2.it`

§Sapienza Università di Roma, Dipartimento di Statistica, Probabilità e Statistiche Applicate, P.le Aldo Moro 5, 00185 Roma, Italy. E-mail: `nicoletta.ricciardi@uniroma1.it`

1 Introduction

Railway freight transportation is a relevant activity in many economies, as it supports and makes possible most other economic activities and exchanges. In the last twenty years, the privatization and reorganization of most of the national railway companies, combined with the removal of many entrance barriers and national protectionism to markets, and the consequent increase of competition, have boosted the development of Operations Research (OR) methods applied to railway transportation.

For comprehensive reviews on planning models for freight transportation, the reader may refer to the surveys by Crainic and Laporte [7] and by Cordeau, Toth and Vigo [4]. The former describes the main issues in freight transportation planning and operations and presents OR models, methods and tools relevant in this field. The latter is specifically focused on railway transportation. The authors provide a complete list of routing and fleet management models for freight transportation, scheduling models for train dispatching and assignment models to assign locomotives and cars. The two cited surveys also give a broad overview on several types of problems and issues arising within the freight transportation arena. As a consequence of the topological structure of the problem, most of the railway transportation models that are available in the literature belong to the class of network flow models, and they are usually customized to take specific aspects of the problem into account.

One of the most relevant tactical planning issues in railway transportation is the *Service Network Design problem*, i.e., the problem of designing the set of origin-destination train routes between yards of the network. This is a typical tactical planning problem, since the effects of decisions span a mid-term planning horizon (about 6-12 months). Service network design problems have been widely investigated. For a survey on this topic the reader may refer to [5]. Crainic et al. [6] provide a first example of a tactical planning model applied to rail freight transportation. In this paper the authors examine the train and traffic routing problems, and the allocation of work between yards. They also present an optimization model intended to compute efficient solutions that reduce costs and provide good quality of service in terms of transportation delays and reliability over a medium term planning horizon. A nonlinear, mixed-integer, multi-commodity formulation is given and a heuristic algorithm is tested on instances arising from the Canadian National Railroads. Kwon et al. [18] present a network flow model on a time-space network to model the dynamic problem on freight car routing and scheduling. The model is solved by a column generation technique. Holmberg and Hellstrand [14] propose a lagrangean heuristic within a Branch & Bound framework to solve the uncapacitated network design problem with single origins and destinations for each commodity, which can be used to model the rail freight transportation. Fukasawa et al. [9] present a network flow model to maximize the total profits, while satisfying the demands within a certain period of time, given the schedules and the train capacities. Jeong et al. [17] focus on (part of) the European freight railway system modeled as a hub-and-spoke network. The scope of their model, an integer linear program, is to find the transport routes, the frequency of services, the number of cars composing each train and the transportation volumes. To solve real instances of the problem, the authors propose a heuristic algorithm. In a very recent paper, Andersen and Christiansen [2] develop a strategical/tactical model for designing the transportation services in part of the European network (the so-called Polcorridor), including in it fleet-sizing considerations, and putting a particular emphasis on some service quality aspects. The resulting non-linear mixed integer program is linearized and solved via the XPRESS solver on different modelling scenarios arising in the Polcorridor network.

Several recent papers address the *Railroad Blocking Problem*, i.e., the problem of identifying a car blocking plan in order to minimize handling costs. A car blocking plan dictates the composition of blocks (groups of incoming cars for connection with outgoing trains) at each classification yard. On this problem, both Newton et al. [24] and Barnhart et al. [3] present a network design formulation. Real instances of the problem are solved in the former paper by a column-generation algorithm and in the latter by a Lagrangean heuristic approach. Ahuja et al. [1] develop a Very Large-Scale Neighborhood Search for the blocking problem and test it on data provided by three major US railroad companies. Liu et al. [20] study the interaction between the design of blocking plans and the modification of the network configuration, i.e., changing yards location and associated capacities. All these papers show that relevant cost savings can be obtained by "good" blocking plans.

In railway freight transportation the movement of empty cars represents an important source of cost with respect to the budget of railway companies. Dejax and Crainic [8] provide a complete review of models for empty car movement in freight transportation. Holmberg et al. [15] face the problem of identifying distribution plans for the movement of empty cars by solving an integer multi-commodity network flow model on a time-expanded network. Sherali and Suharco [25] propose a tactical model for the distribution and repositioning of empty rail cars for shipping automobiles. Jaborn et al. [16] analyze the cost structure for the repositioning of empty cars, and show that such costs often depend on the number of car groups handled at yards. In the cost structure they also compound costs due to car classification at intermediate yards. They observe the economy-of-scale behavior of total distribution costs, which can be reasonably decreased by building fewer but larger car groups. The authors model the empty freight car distribution problem as a capacitated network design on a time-dependent network and solve it with a tabu search meta-heuristic.

Many of the papers listed above refer and are customized for the North American freight railway transportation systems. Although the decision problems are similar in nature, there are several differences between European and North American freight railway transportation companies, both in the way of managing operations and in the way of handling tactical planning. Consequently, mathematical models developed for the European system may not be appropriate for the North American one and vice versa.

This paper is a by-product of a cooperation with a major Italian railway company, which is in charge of managing freight transportation in the whole Italian network. We have analyzed the freight operations of this company, singling out the differences and similarities with other freight railway transportation systems, and we have developed a mathematical model to design the service network. The mathematical model, customized for this company, takes into account specific features of the system and of the type of services. It represents both full and empty car movements, considers the specific cost structure of the Italian railway company, and guarantees important quality requirements. On this subject, all the transportation demand has to be satisfied and delivered on time. In fact, in case of late deliveries, transportation contracts and agreements often include fines and penalties for the railway transportation company.

Besides a detailed analysis of the italian freight rail transportation system, the contribution of this paper is twofold.

First, we present a new mathematical model to design the set of origin-destination service connections specifically customized for the Italian case. Indeed, we formulate the problem as a network design model with a non-linear objective function. The non-linear objective function is used to take into account the requirements on the quality of service. The mathematical model we propose can be cast as an Integer Concave-cost Multi-commodity Network Design Problem, which is $\mathcal{NP}$-hard [21].

Second, we describe a specialized tabu search algorithm to solve real instances of the problem. With the purpose of forcing the algorithm to visit a larger portion of the feasible region, we add a mechanism to perturb the current solution. The tabu search procedure takes advantage of the particular structure of the network, thus resulting in an efficient algorithm compared to other tabu search procedures applied to similar problems. We demonstrate that the algorithm is able to compute promising solutions on real instances of the problem.

The remainder of the paper is organized as follows. Section 2 describes the Italian freight railway transportation system, in order to provide the main hypothesis and constraints for our case-study; the mathematical model is discussed in Section 3; Section 4 presents the tabu search algorithm; Section 5 contains the computational results on numerical instances; Section 6 provides the conclusions.

2 Description of the Italian freight railway transportation

Railway transportation systems differ from country to country in several respects. We here present the main characteristics of the Italian railway transportation system. Many of these features are also valid for other European systems. The information provided in this section, resulting from our cooperation with an Italian railway company, will allow us to design an optimization model (described in Section 3) which is consistent with the Italian reality, and possibly extendable to other countries.

The Italian railway transportation takes place either with a complete train or a single car. *Complete trains* are dedicated point-to-point trains, rented by a customer, used to move goods when they are quantitatively and/or economically considerable. Complete trains are directly scheduled by the customer, who pays in full for the service received. Moreover, complete trains are often composed by private (customer owned) cars, and they are usually balanced in there-and-back shipments. In any case, there are no restrictive constraints on car availability.

In the *single car* transport modality, the railway company is in charge to move goods from their origins to their destinations either with a direct train, if one exists, or with a sequence of trains with one or more intermediate stops. Cars move on a *physical network* composed by a set of *yards* and a set of *tracks* connecting the yards. Using a hierarchical criterion of importance, yards are classified in first, second and third level and are connected in a hub-and-spoke mode. That is, a second level yard (resp., a third level yard) is connected to a first level yard (resp., a second level yard). First level yards, also called *classification yards*, are the only one equipped with special tracks, called *classification tracks*, each of them generally used for one specific destination. Once a train arrives at a classification yard, all cars are separated, classified and those, which are not yet arrived to their final destination, are recombined to compose new outgoing trains. In the single car transportation modality, trains are usually multi-client and unbalanced in the there-and-back shipments, and difficult decision problems are involved. Railway companies have to face several trade-offs between the cost structure and the service quality, e.g., all the demand of transportation has to be satisfied and possibly delivered on time.

Although cars traveling the Italian network may belong to foreign or to other private companies, freight *trains* are always arranged by the Italian railway company. In Italy, freight trains run according to a fixed schedule and are mostly operated at night. During daytime, they have to give priority to passengers trains. This is not the case in North America, where trains usually leave a yard only when a sufficient cargo tonnage is reached, and they have priority over passenger trains. In addition, some railroads, sometimes single track too, are of exclusive use of freight trains. In Italy, and in Europe as well, single track railroads are extremely rare, and no railroad is exclusively dedicated to freight trains.

The size of the Italian trains is quite small: a freight train generally does not have more than 20 cars, does not exceed 600 meters length, and is usually pulled by only one locomotive. Moreover, cars are generally managed individually and no "blocks" (groups of cars) are composed. At any classification yard, cars are uncoupled from the incoming train, reclassified and, if they have not reached their final destination yet, coupled again to a new outgoing train. Therefore, at any intermediate classification yard, car handling implies a coupling and an uncoupling manoeuvre, with significant impacts on the total transportation costs and delivery times. Hence any mathematical model, developed with the scope of managing efficiently freight operations, cannot avoid to take handling costs into account. Although the experience of North American systems suggests that large economies of scale can be achieved by managing blocks of cars, in Italy blocks of cars are not used for the following reasons. First of all, the Italian cars are not equipped with automatic hooks which facilitate the manoeuvre of groups of cars; second, the organization should be modified; and finally the size of trains is rather small with respect to the American counterpart, thus limiting the possibility to achieve large economies of scale.

Table 1 summarizes the main differences between railway transportation systems in North America and Italy. The comparison has highlighted three major features that are relevant for the model that we shall propose: first, freight trains are planned according to a fixed schedule; second, length and weight limits are rather tight, as a train can be assigned at most 20 cars; third, there is currently no possibility of grouping cars in blocks at intermediate yards.

2.1 Data analysis

We here report an analysis on traffic demand and on car movements. This analysis highlights some interesting features of the current Italian freight rail transportation, thus allowing us to design an optimization model (described in Section 3) which is consistent with the Italian reality. Data on car movements have been retrieved from historical databases of our industrial partner.

These data refer to cars owned by the company and moved on the Italian railroad network. This railway company owns about 35% of the total cars. The remaining part is owned by foreign companies (51%), both public and private, and other private Italian companies (14%).

There are several types of cars. As a first approximation, we may classify cars in seven major classes. Two of these classes, namely cars of type R/Rh and of type S/Sh (prevalently dedicated to containers, wood, iron and steel industry products), are those mostly used and currently cover about half of the total demand.

As for the cost structure, a first source of cost is the movement of empty cars. Even though they do not generate any economic return, these movements are necessary to overcome imbalances in the demand for cars at each node of the network. In Italy the movement of empty cars accounts for about 40% of the total car trips. Whenever a demand for empty cars arises, it is first satisfied with the internal offer (intra-plant), i.e., with cars available in the same geographical region. When intra-plant empties are not available, then inter-plant cars, i.e., cars coming from other geographical regions, are used. Inter-plant movements are 42% of the total. This large number of empty car movements reflects the Italian economy, characterized by production districts concentrated mainly in the northern part of the country.

The second relevant source of cost is the car handling process at intermediate yards. On average, car handling costs add up to *one third* of the total transportation costs that the railway company sustains to move a car from its origin to its final destination. In Figure 1 we report the number of manoeuvres for each car from its origin to its destination. The picture shows that more than one third of the empty cars change train at least twice before arriving at final destination (i.e., each of these cars is coupled to at least three trains). This large number of manoeuvres affects also the duration of the total travel time of a car from its origin to its destination. Reducing the number of manoeuvres, it is possible to guarantee a better quality of service to customers.

Finally, we verified the level of use of a car, in order to evaluate how the locked up capital is used. We refer to a *cycle* of a freight car as a trip between two consecutive loads. Figure 2 depicts the number of cars, ordered by number of cycles, in a period of three months. A large number of cars runs few cycles: half of the cars are used for less than one fourth of the total number of cycles. In addition, we define a *dispersion index* to measure the frequency of a car traveling the same service. More formally, the dispersion index of car i, denoted by $D(i)$, is defined as the ratio between the number of origin-destination pairs visited by a car i, and its total number of cycles:

$$D(i) = \frac{\text{number of O/D visited by car } i}{\text{number of cycles of car } i}, \qquad \text{for each car } i.$$

$D(i)$ assumes values in the interval $(0, 1]$. If it is equal to one, car i is used at each cycle for a different origin-destination pair (case of maximum dispersion). On the other hand, the more $D(i)$ is closed to 0, the more car i is travelling the same set of origin-destination pairs. To compute such an index, we considered only those cars running at least two cycles per month. In Figure 3 we report the dispersion index of cars with respect to the number of cycles. Half of the cars have a dispersion index which is close to one; on the other hand, less than 10% of the cars are always used for the same service. Private companies show quite different performances: their cars have a higher level of usage and a smaller index of dispersion (see Figure 3). These statistics reflect a natural situation, due to the public nature of the service provided by the Italian railway company. However, it is believed that, reducing the dispersion index of cars, it is possible to improve the efficiency of the operations. The concentration of freight flows on few links can be profitable in a hub-and-spoke transport modality.

In conclusion, the data analysis carried out on freight transportation has highlighted key features of the system, which can be summarized as follows. First of all, empty car movements represent a large portion (about 40%) of the total car movements, and they should be carefully managed. Second, we observe the necessity of taking into account the manoeuvres of cars at intermediate yards, since they have a big impact (more than 30%) on the total transportation cost and they can cause inefficiencies. Finally, we highlight a low level of use of cars, and the need for a reduction of their "dispersion" on the network. As we shall see in the following section, these issues will be considered in the proposed model.

3 A mathematical model for the freight service network design problem

In this section, we present a static mathematical model to design the set of services customized for an Italian freight railway transportation company. The scope of the model is to satisfy the demand of transportation at a minimum cost, over a mid-term planning horizon. The results of the mathematical model are: *i*) the definition of the service network and of the number of direct trains connecting the origin-destination pairs; *ii*) the routing of cars on the network. In the model, we consider both full and empty cars. In addition to the satisfaction of all the demand of transportation, the solution has also to satisfy the following requirement on service quality: goods have to be delivered on time. This requirement is enforced by minimizing the waiting time of cars at intermediate yards.

As an underlying structure of the model, we consider the network of *possible* services, represented by a directed graph $D = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ denotes the set of classification yards and $\mathcal{A}$ the set of all possible direct services. Since it is always possible to have a direct train between any pair of yards, we can reasonably assume that the network of possible services is *complete*, thus $\mathcal{A} \equiv \mathcal{N} \times \mathcal{N}$.

As mentioned above, the purpose of the model is to design the network of active services, namely $D' = (\mathcal{N}, \mathcal{S})$, which is a subgraph of the complete (directed) graph $D = (\mathcal{N}, \mathcal{N} \times \mathcal{N})$. More precisely, if $(i, j) \in \mathcal{S}$, then we say that yard $i \in \mathcal{N}$ is directly connected to yard $j \in \mathcal{N}$ by a *service*, that is, there exists at least one direct train connecting i to j. The *frequency* of a service is the number of trains planned on a service in the time unit, which is typically one week.

Cars move from their origins to their destinations using either a direct service, if one exists, or a sequence of trains with intermediate stops. Since blocks of cars are not allowed, we can assume without incurring in poor approximations that, whenever a train stops at an intermediate classification yard, *all its cars* are reclassified. In view of this hypothesis, we can directly associate handling costs to arcs $(i, j) \in (\mathcal{N} \times \mathcal{N})$ representing direct connections between pairs of classification yards $i, j \in \mathcal{N}$.

We also assume that the network of services is *uncapacitated*. According to the experience of the Italian railway company managers, at a tactical decision level each arc of the railroad network has enough capacity to route the cars needed to satisfy all the demand.

The mathematical model, in common with most of the transportation models available in the literature, minimizes the total costs. In view of the data analysis presented in Section 2.1 and of the experience of experts, the cost structure of freight transportation for the Italian company can be decomposed into two major components: *train* and *car* costs. These costs are accounted in the system of the railway company as costs per kilometre. The train cost per kilometre includes costs of depreciation, maintenance, power, tool and engine drivers. If δ denotes the train cost per kilometre, to each arc of the network we can assign a cost $f_{ij} = \delta \cdot d_{ij}$ where d_{ij} is the length of the link (i, j). The car cost per kilometre γ^p includes both car depreciation and maintenance of a car of commodity type p. In view of our assumptions, the total cost for a car of commodity type p traveling a link (i, j) is given by the formula $c^p_{ij} = \gamma^p \cdot d_{ij} + m^p_j$, where m^p_j is the handling cost of a car of commodity type p at yard j. The formulation of the model requires the definition of the following additional notation:

- $\mathcal{N} = 1, 2, \ldots, n$: the set of yards;
- $\mathcal{K} = 1, 2, \ldots, k$: the set of commodities, given by the cartesian product of the set of car types and the set of yards;
- b^p_i: the supply (if positive) or the demand (if negative) of cars of commodity p at yard i;

- α: the maximum number of cars that can be assigned to a train.

The decision variables of the model are:

- x_{ij}^p: the number of cars of commodity $p \in \mathcal{K}$ assigned to service (i, j);
- y_{ij}: the number of trains traveling on the direct service (i, j), i.e., the frequency of service (i, j).

The formulation here follows:

$$\text{Min} \sum_{(i,j)\in\mathcal{A}} h(y_{ij}) \cdot f_{ij} \cdot y_{ij} + \sum_{(i,j)\in\mathcal{A}, p\in\mathcal{K}} c_{ij}^p \cdot x_{ij}^p$$

s.t.

$$\sum_{j\in\mathcal{N}} x_{ji}^p - \sum_{j\in\mathcal{N}} x_{ij}^p = b_i^p \qquad \forall i \in \mathcal{N},\ \forall p \in \mathcal{K}. \tag{1}$$

$$\sum_{p\in\mathcal{K}} x_{ij}^p \leq \alpha \cdot y_{ij} \qquad \forall i, j \in \mathcal{N} : i \neq j. \tag{2}$$

$$x_{ij}^p \in \mathbb{Z}^+ \qquad \forall i, j \in \mathcal{N} : i \neq j,\ \forall p \in \mathcal{K}. \tag{3}$$

$$y_{ij} \in \mathbb{Z}^+ \qquad \forall i, j \in \mathcal{N} : i \neq j. \tag{4}$$

First of all, a few words on the objective function. As we have mentioned above, the model minimizes the total costs, which include both train and car costs. Train costs are multiplied by a parameter which is a function of the number of trains traveling link (i, j):

$$h(y_{ij}) = \left(a + \frac{b}{y_{ij} + 1}\right), \text{ with } 0 < a < 1 \text{ and } b > 0.$$

The introduction of this parameter $h(y_{ij})$ is motivated by the requirements on service quality. These requirements are soft constraints, i.e., their satisfaction is preferable but not mandatory. By means of the objective function, the model prefers service networks with freight flows concentrated on some links, thus guaranteeing a high frequency and possibly a lower dispersion of cars on the network. This implies a reduction of the waiting time of cars at intermediate yards. Note that $h(y_{ij})$ makes the objective function *concave*. Further details on the structure of the objective function are reported in Section 5.1.

Constraints (1) are the so-called mass balance constraints: they state that all the demand has to be satisfied and no goods are dispersed on the network. Constraints (2) link the decision variables: the total number of cars traveling on a specific origin-destination pair is closely related to the number of trains running on the same *o-d* pair.

The mathematical model belongs to the class of the multi-commodity network design problems with concave cost function. Minimum concave-cost network flow problems are known to be $\mathcal{NP}$-hard (refer to [13] for a survey on algorithms and applications). Several solution techniques have been developed, both exact and approximate. The first ones explicitly or implicitly enumerate the vertices of the polyhedron defined by the network constraints, and are based on Branch & Bound, extreme point ranking methods or dynamic programming [13]. The second ones can be subdivided into two classes: heuristic and approximation algorithms. Heuristics find local optima using standard convex programming techniques; approximation algorithms use piecewise linear approximation of the concave objective functions.

We here propose a specialized heuristic algorithm based on a tabu search, which computes good solutions in reasonable computational time.

4 A tabu-search heuristic algorithm

The application of railway transportation models to real operations generally requires the solution of large size instances, thus calling for a compromise between solution quality and computational time. The size of any realistic instance of the service network design model presented in Section 3 is very large. Even instances which consider only classification yards lead to formulations with a number of constraints and variables of order of several hundreds of thousands. Given the computational complexity of the problem, we propose a heuristic algorithm based on tabu search (see [10] and [11] for details on tabu search). This choice is motivated by the requirement of our industrial partner to have good solutions in reasonable computational time. We furnish the algorithm with a perturbing mechanism (diversification phase) which alters the current solution in order to explore a larger portion of the solution space.

Meta-heuristic algorithms have been widely implemented to solve railway transportation models. For instance, we can mention Marín and Salmerón [22], [23] and Gorman [12] among others. Marín and Salmerón [22], [23] compared the application of three different local search meta-heuristics to the tactical planning of rail freight networks. Gorman [12] discussed both genetic and tabu search meta-heuristics to solve the weekly routing and scheduling problem of a major US freight railroad.

In our model, decisions are classified into *service* decisions (whether to activate or not a direct connection between a pair of yards of the network) and *routing* decisions. In the proposed heuristic we treat the two types of decision variables in a hierarchical manner, i.e., we first fix the value of service decision variables and then we decide how to route cars using the activated links. Figure 4 reports a flow chart of the proposed heuristic algorithm. In particular, it is composed of three phases: the generation of a starting feasible solution; the tabu-search core procedure; and the so-called "perturbing phase".

Finding a starting feasible solution. Our heuristic procedure is initialized with starting feasible solutions. These solutions are computed with different methods, listed in the sequel.

Minimum spanning tree: The initial solution is given by a minimum spanning tree found on a complete graph, whose arc costs r_{ij} are computed as follows:

$$r_{ij} = \frac{d_{ij}}{g_{ij}}, \quad \forall (i,j) \in \mathcal{N} \times \mathcal{N}$$

where d_{ij} is the length of the link (i,j), and g_{ij} represents the total demand for full cars from origin i to destination j. Observe that the minimum spanning tree minimizes the number of services, i.e., direct links between specified origins and destinations.

Currently used service network: We use the solution that is currently implemented, which is computed on the base of practitioners' experience.

Direct service graph: We compute a solution which satisfies all the demands with a direct service. This solution clearly minimizes the car handling costs.

Minimum length graph: We compute a feasible solution selecting arcs from the network of possible services in a minimum length order until a connected graph is obtained.

The scope of all these methods is to compute a variety of feasible solutions quickly. Once the initial feasible service network is computed, we route cars minimizing the overall cost of the solution. Note that we consider full and empty cars separately, since they have different features. Full cars are characterized by a specific O/D pair, while empties have to satisfy only constraints on the type: any demand for empty cars can be satisfied with cars located at any origin.

As observed in Section 3, the service network is uncapacitated and we can schedule on any service as many trains as needed. Therefore, the flow of full cars can be computed using a shortest path algorithm for each commodity. In particular, we implement the Floyd-Warshall algorithm. To route the empty cars, we solve an uncapacitated minimum cost flow problem (one for each commodity).

Tabu search procedure. The core routine of our heuristic algorithm is a tabu search procedure. As a neighborhood of the current solution, we consider the set of all the networks that can be obtained from the current service network adding or dropping one arc at a time. Thus our tabu search procedure is based on two simple types of neighborhood *move*. Adding an arc corresponds to opening a new service, while dropping an arc corresponds to closing the corresponding service. If $D_c(\mathcal{N}, \mathcal{S})$ denotes the current service network, its neighborhood, namely $N(\mathcal{S})$, is defined as follows: $N(\mathcal{S}) = \{D_c(\mathcal{N}, \mathcal{S}') : \mathcal{S}' \in (\mathcal{S}^+ \cup \mathcal{S}^-)\}$ where $\mathcal{S}^+ := \mathcal{S} \cup \{(i,j)$ for some $(i,j) \notin \mathcal{S}\}$ and $\mathcal{S}^- := \mathcal{S} \setminus \{(i,j)$ for some $(i,j) \in \mathcal{S}\}$.

For each move a new solution is obtained. Among all these solutions we should select the one of minimum cost (to which corresponds the maximum improvement of the objective function value). An exact computation of the solution cost requires the solution of routing sub-problems, one for empties and one for full cars. This approach can be very time consuming thus leading to high computational time. To overcome this drawback, we use estimates of the solutions cost, computed as follows. Consider a new solution obtained by adding a new direct service (arc) between nodes i and j. In this case, all the flow going from i to j is re-routed on the directed arc (i,j). Given the updated flow vector, the new cost of the objective function is obtained by simply adding the new total transport cost and the new car handling costs. On the other hand, if an arc is removed, say (i,j), all the trains which previously travelled arc (i,j) are re-routed along the shortest path connecting i to j. Note that, in removing a service, not all moves are allowed since the arc deletion has to preserve graph connectivity. These procedures for the estimation of solution cost are fast and easy to implement; in fact, they do not compute any car routing but they only focus on the O/D pair whose connection status has been changed by the move. Yet, these procedures provide only an estimate of the cost. Figure 5 highlights the possible errors that could be made. However, once the most convenient move is determined on the basis of cost estimates, the new service network is considered, and full and empty cars are routed optimally, thus computing the *optimal* flow vector and the *exact* cost.

Two different data structures - a tabu list $TL1$ and a tabu list $TL2$ - are used in order to keep information on the last visited solutions. $TL1$ is a l-dimensional list which is filled with the last l solutions visited by the algorithm, while $TL2$ is an h-dimensional list which contains the last h moves performed by the procedure. The dimensions of the lists are fixed a-priori and their value are input parameters of the procedure. A trade-off exists on the dimension of the lists. Larger is the dimension of the lists and likely better solutions can be computed, at a cost of longer computational times. Once the list is full, i.e., the number of stored solutions in $TL1$ (resp., $TL2$) is equal to l (resp., h), and it is necessary to add a new solution to the list, then the solution of the list with maximum-cost is removed. Clearly, the tabu restriction of a solution may be ignored in case the produced solution is better than the current best solution. The algorithm ends when a maximum number of iterations is achieved.

Perturbing phase. In order to visit a larger portion of the feasible region, we furnish our tabu search algorithm with a diversification phase consisting of three perturbing mechanisms: the forced insertion, the forced removal and the serial elimination.

Whenever the current optimal solution is not updated within a certain (fixed) number of iterations, the algorithm is forced to perform a move. The move consists of adding or dropping an arc. In the first case, we say that we apply a *forced insertion*, while in the second case we say that we apply a *forced removal*. Among a set of candidate arcs, the arc that is either added or dropped is the one providing the highest improvement or the lowest worsening of the objective function. The set of candidate arcs is composed by a certain (fixed) number of less-recently-selected arcs. Again, if a removal is performed, network connectivity has to be verified.

The third perturbing mechanism, referred to as *serial elimination*, forces the algorithm to move from the current solution to a different one outside its neighborhood. Whenever the current optimal solution is not updated within a certain (fixed) number of iterations, the serial elimination forces the algorithm to re-start the local search from a service network which is composed by a "minimal" number of direct arcs. More precisely, the procedure forces the algorithm to keep on removing arcs until the solution becomes composed by a small fixed number of arcs. Again, removed arcs have to keep network connectivity and are those providing the highest improvement or the lowest worsening to the objective function.

5 Computational results

In this section, we present the computational results on instances of the problem. We first report the computation on random generated instances and then, in Section 5.1, we study a real instance provided by our industrial partner.

The set of random instances includes 25 "small" instances, each characterized by two commodities and a number of yards in the range between 5 and 10. Distances, demands and handling costs have been randomly generated using a uniform probability distribution. All the parameters used to generate the instances are summarized in Table 2.

We first solved these instances using MINLP-B&B [19] (with the default settings), a package for mixed-integer non-linear programs, which implements a branch & bound procedure. The problems at nodes of the b&b search tree are continuous nonlinear constrained optimization problems. These problems are solved using a sequential quadratic programming algorithm. All these instances have been solved to optimality within one second. However, the solver is not able to converge to an optimal solution within the time limit (set to 3600 seconds) on instances which are comparable in size to real instances, thus showing the low scalability of the b&b procedure.

We tested the computational performance of our heuristic algorithm comparing the optimal and heuristic solutions of the random generated instances described above. These results are listed in Table 3. In most cases the heuristic algorithm finds the optimal solution within few seconds. In one third of the instances the heuristic solution differs from the optimal one. However, the deviation from the optimum is rather low (1.78 %, on average).

5.1 Tests on a real instance

We here report the computational results for a real instance of the problem. Data on traffic demand and on car transportation and handling costs has been retrieved from historical databases and from the accounting system of the Italian railway company. The maximum number of cars which can be assigned to a train has been set to 20. We restrict our analysis to a network which includes all the *classification* yards and the border passes. Therefore the number of nodes of the network is 39. The traffic demand of lower-level yards has been aggregated to the demand of the corresponding classification yard. The instance also includes 375 commodities.

Computational experiments showed that best solutions are computed starting the algorithm procedure with either the minimum spanning tree solution or the currently used service network solution (see Section 4 for details).

A first phase of the computational experiments has been devoted to the fine tuning of the algorithm parameters. The best computational performance of the algorithm, in terms of quality of the solutions computed, are obtained with the following values of the parameters.

- The dimensions of the tabu-lists TL1 and TL2 are set to 10 and 7, respectively.
- The forced insertion and the forced removal modules may be activated, respectively, every 15 and 50 iterations of the standard procedure without improvement in the objective function value.
- The serial elimination module may be called every 150 iterations of the standard procedure without improvement in the objective function value, until the number of arcs of the solution becomes equal to 135.

A second phase of the tests has been devoted to investigated how much the heuristic solution is sensible to the choice of the multiplicative factor $h(y_{ij})$. The introduction of the multiplicative factor into the cost function has been motivated by the request of guaranteeing a certain level of quality of service. The railway company prefers direct services with a sufficiently *large* number of planned trains (see Section 3). We handled this preference introducing a multiplicative factor in the objective function. The scope of this factor is to reduce (to increase) train costs if enough (few) trains are scheduled on the service. The expression of the multiplicative factor, which has been decided with the experts of the railway company and has been corroborated by an internal and undisclosed data analysis, is the following (see Fig. 7):

$$h(y_{ij}) = \left(a + \frac{b}{y_{ij} + 1}\right),$$

with $0 < a < 1$, and $b > 0$.

The parameter a bounds above the maximum cost-reduction. For instance, if $a = 0.8$, then the maximum cost-reduction on a link is 20% of the total train costs. The parameter b is fixed in order to neutralize the effect of the multiplicative factor $h(y_{ij})$ (i.e., it is equal to 1) whenever the number of trains scheduled on the service is equal to the threshold-value set by the company preferences. In our case, this threshold-value has been set equal to ten trains per week for all the services. As an example, in Figure 6 we report the number of links in the optimal solution assuming that a varies from 1 to 0.1. It is clear that a smaller value of a corresponds to a higher sensibility to quality of service, while $a = 1$ implies that no "quality cost" is considered. As shown in the picture, the smaller the value of a is, the more the freight flow is concentrated on a small number of service links. In our case, all the parameters are fixed according to cost criteria which are currently used in the accounting systems of the railway company. In particular, the choice for the value of the parameters is $a = 0.9$ and $b = 1.1$. These values may be considered as those that better fit the cost structure of the company. Using these coefficients, the train cost is increased in case of a small number of trains planned on a service, while it is reduced (but never more than 10%) in case of a large number of trains (see Fig. 7).

Figure 8 depicts the network of services currently used by our industrial partner (left side of the figure) and the network of service proposed by our algorithm (on the right side). Comparing the two solutions, it is evident that many direct services (links of the network) are the same for both solutions. However, the mathematical model suggests a larger number of services which are mainly concentrated in the northern part of Italy. Indeed, in the southern part, the two solutions show almost the same services, with minor differences. Without a large increment of the freight transportation demand, the current infrastructure and the geographical features of the region do not leave enough room for solution improvements. Therefore, in this part of Italy, the current freight transportation practice is the best practice. The northern part of Italy is the region where the two solutions are mostly different. In this area, the algorithm suggests a larger number of direct services. The current practice organizes freight operations in a hub-and-spoke modality with few major hubs connecting all the yards of the network. The algorithm suggests more yard-to-yard connections, thus allowing a reduction of handling costs, which represent a large portion of the total costs. At the same time, a larger number of direct services allows a higher efficiency in time deliveries. For a quantitative comparison of the two solutions, Table 4 summarizes both technical and economical statistics, hereafter listed:

1. C: the objective function value, representing a measure of the total cost.
2. S: the number of direct services between yards.
3. T: the number of trains necessary to deliver all the cars on the service network in order to satisfy all the transportation demands.
4. TK: the total amount of kilometers covered by all the trains traveling on the network.
5. CK: the total amount of kilometers covered by all the cars traveling on the network.
6. M: the number of manoeuvres.

It can be observed that total costs decrease of 3.82%, as well as the number of trains (−9.56%), car manoeuvres (−9.61%), and the total amount of kilometers covered by all the trains (cars, resp.) traveling on the service network (−3.17 and −3.16%, resp.). This cost decrease reflects onto an increase in the number of direct services which are activated in the solution proposed by the algorithm (+9.72%).

6 Conclusions

In this paper we have presented a case-study on the freight railway transportation, which has been promoted by an Italian railway company. As part of the study, a thorough analysis on the company operations has been carried out. The analysis has highlighted some specific features of the Italian transportation reality. Several differences with other freight railway transportation systems turned up, especially with the North American ones. Relevant differences arise with respect to the planning process, the structure of railroads, in the policies adopted to compose trains and in the train limits in terms of maximum weight and length. For the Italian case, we have detected the following key aspects:

1. A large portion of the total transportation costs is due to both empty car distribution and car handling at intermediate yards.
2. Blocks of cars are not formed.
3. The transport system has to guarantee some quality requirements, such as the punctuality of goods delivery.

All these aspects are considered in the mathematical model we have presented to design the network of services, i.e., the set of direct train connections between pairs of yards. The model, which is specifically tailored to the Italian reality, also determines the routing of freight cars on the network. It suggests the services to provide, the number of trains and both the number and type of cars traveling on each service. All the decisions are taken to minimize the total costs. The quality of the service offered to customers is enforced by a concave cost objective function.

The resulting model, which can be classified within the class of minimum concave-cost network design problems, is solved with a specialized tabu search meta-heuristic algorithm, which adopts some perturbing mechanisms in order to diversify the search process.

The computational results show a good behaviour of the algorithm on realistic instances of the problem, thus proving the viability for an integration in a decision support system.

Acknowledgment

The authors warmly thank Marco Campetella (previously at Trenitalia s.p.a.) for his valuable contributions, and Teodor G. Crainic for his suggestions on a preliminary version of the paper.

References

[1] R. K. Ahuja, K.C. Jha, and J. Liu. Solving real-life railroad blocking problems. *Interfaces*, 37(5):404–419, 2007.

[2] J. Andersen, and M. Christiansen. Designing new European rail freight services. *Journal of the Operational Research Society*, 60:348–360, 2009.

[3] C. Barnhart, H. Jin, and P.H. Vance. Railroad blocking: A network design application. *Operations Research*, 48(4):603–614, 2000.

[4] J.F. Cordeau, P. Toth, and D. Vigo. A survey of optimization models for train routing and scheduling. *Transportation Science*, 32(4):308–404, 1998.

[5] T.G. Crainic. Service network design in freight transportation. *European Journal of Operational Research*, 122:272–288, 2000.

[6] T.G. Crainic, J.A. Ferland, and J.M. Rousseau. A tactical planning model for rail freight transportation. *Transportation Science*, 18(2):165–184, 1984.

[7] T.G. Crainic and G. Laporte. Planning models for freight transportation. *European Journal of Operational Research*, 97(3):409–438, 1997.

[8] P.J. Dejax and T.G. Crainic. A review of empty flows and fleet management models in freight transportation. *Transportation Science*, 21(4):227–247, 1987.

[9] R. Fukasawa, M.V. Poggi de Aragão, O. Porto, and E. Uchoa. Solving the freight car flow problem to optimality. *Electronic Notes in Theoretical Computer Science*, 66(6), 2002.

[10] F. Glover. Tabu search - Part I. *Journal on Computing*, 1:190–206, 1989.

[11] F. Glover. Tabu search - Part II. *Journal on Computing*, 2:4–32, 1990.

[12] M.F. Gorman. An application of genetic and tabu searches to the freight railroad operating plan problem. *Annals of Operations Research*, 78:51–69, 1998.

[13] G.M. Guisewite and P.M. Pardalos. Minimum concave-cost network flow problems: Applications, complexity, and algorithms. *Annals of Operations Research*, 25:75–100, 1990.

[14] K. Holmberg and J. Hellstrand. Solving the uncapacitated network design problem by a lagrangean heuristic and branch-and-bound. *Operations Research*, 46(2):247–259, 1998.

[15] K. Holmberg, M. Joborn, and J.T. Lundgren. Improved empty freight car distribution. *Transportation Science*, 32(2):163–173, 1998.

[16] M. Jaborn, T.G. Crainic, M. Gendreau, M. Holmberg, and J.T. Lundgren. Economies of scale in empty freight car distribution in scheduled railways. *Transportation Science*, 38(2):121–134, 2004.

[17] S.J. Jeong, C.G. Lee, and J.H. Bookbinder. The European freight railway system as a hub-and-spoke network. *Transportation Research - A*, 41:523–536, 2007.

[18] O.K. Kwon, C.D. Martland, and J.M. Sussman. Routing and scheduling temporal and heterogeneous freight car traffic on rail networks. *Transportation Research - E*, 34(2):101–115, 1998.

[19] S. Leyffer. User manual for MINLP BB. Technical report, Department of Mathematics, University of Dundee, 1999.

[20] J. Liu, R.K. Ahuja, and G. Şahin. Optimal network configuration and capacity expansion of railroads. *Journal of the Operational Research Society*, 59:911–920, 2008.

[21] T.L. Magnanti and R.T. Wong. Network design and transportation planning: Models and algorithms. *Transportations Science*, 18(1):1–55, 1986.

[22] A. Marín and J. Salmerón. Tactical design of rail freight networks. part i: Exact and heuristic methods. *European Journal of Operational Research*, 90:26–44, 1996.

[23] A. Marín and J. Salmerón. Tactical design of rail freight networks. part ii: local search methods with statistical analysis. *European Journal of Operational Research*, 94:43–53, 1996.

[24] H.N. Newton, C. Barnhart, and P.H. Vance. Constructing railroad blocking plans to minimize handling costs. *Transportation Science*, 32(4):330–345, 1998.

[25] H.D. Sherali and A.B. Suharko. A tactical decision support system for empty railcar management. *Transportation Science*, 32(4):306–329, 1998.

Figures

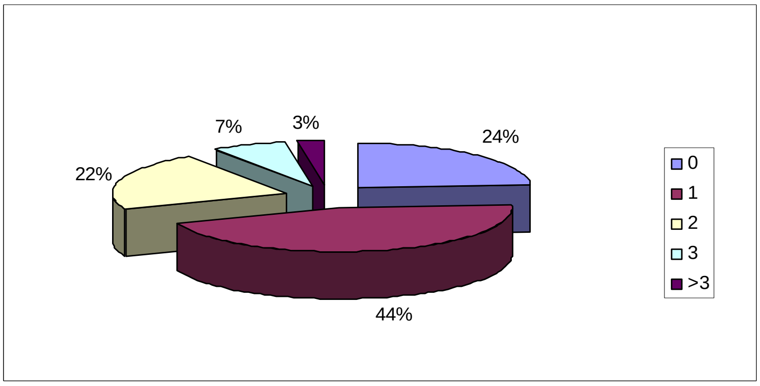


Figure 1: Number of manoeuvres per car per cycle.

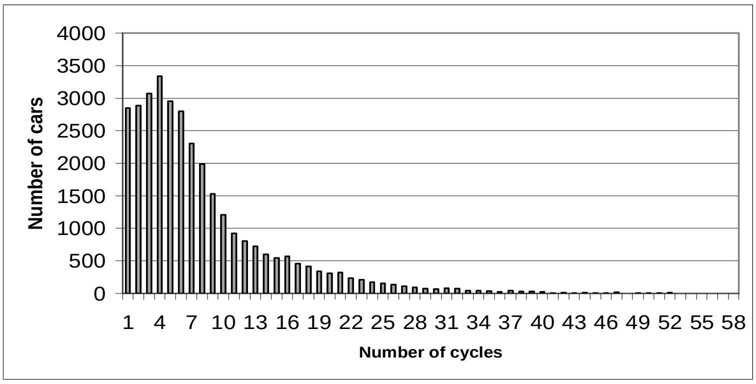


Figure 2: Number of cycles per car.

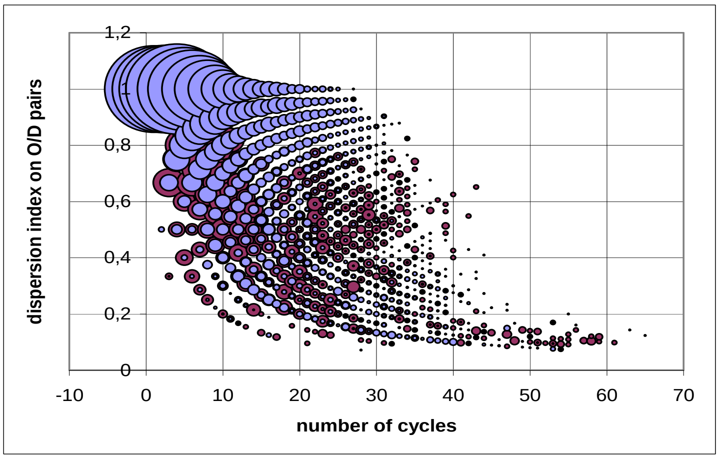


Figure 3: Dispersion of cars on O/D pairs. Blue circles refer to the Italian rail company, while purple circles refer to other private companies. The ray of the circles is proportional to the number of cars.

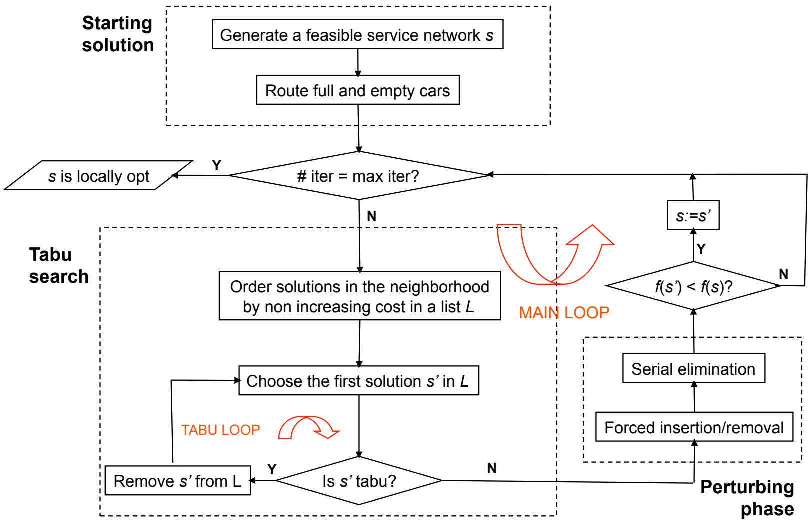


Figure 4: Flow chart of the heuristic algorithm.

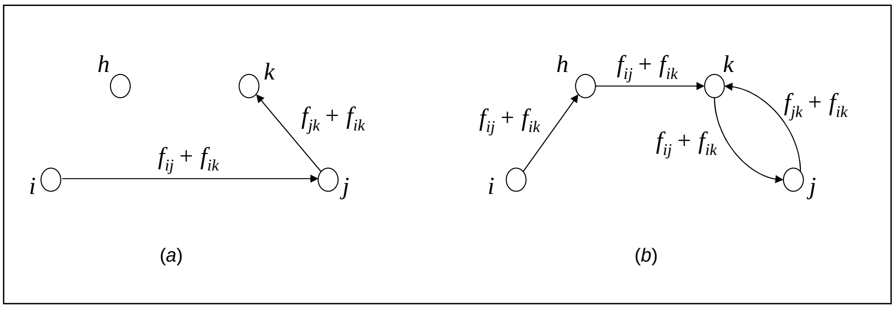


Figure 5: Cost-stimate for a solution obtained by removing a direct link. (a) The routing before the removal of arc (i, j). (b) The routing after the removal of arc (i, j). Let f_{ij}, f_{jk}, f_{ik} be the flow of cars with origin-destination pair $i - j$, $j - k$ and $i - k$, respectively, and suppose that arc (i, j) is removed. Then, flow $f_{ij} + f_{jk}$ is moved from arc (i, j) to the (new) shortest path connecting i to j, namely $\{(i, h), (h, k), (k, j)\}$. It follows that, when estimating the cost of the new solution, the algorithm will take into account the amount of flow f_{ik} on arcs (k, j) and (j, k); however, this flow should not be routed at all.

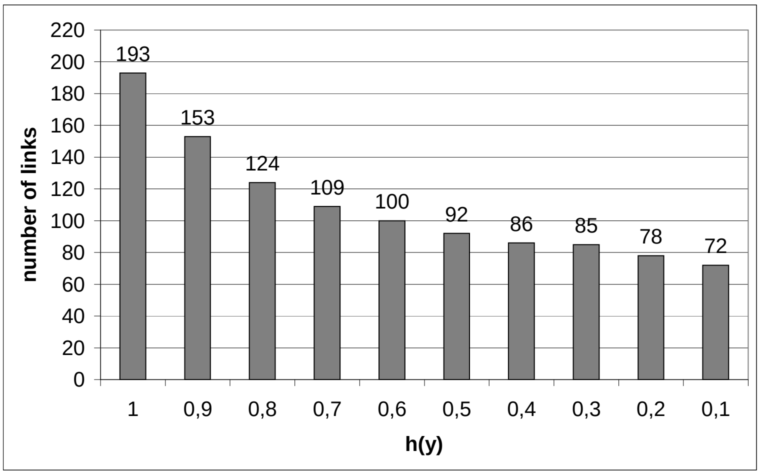


Figure 6: Tests on the multiplicative factor $h(y)$.

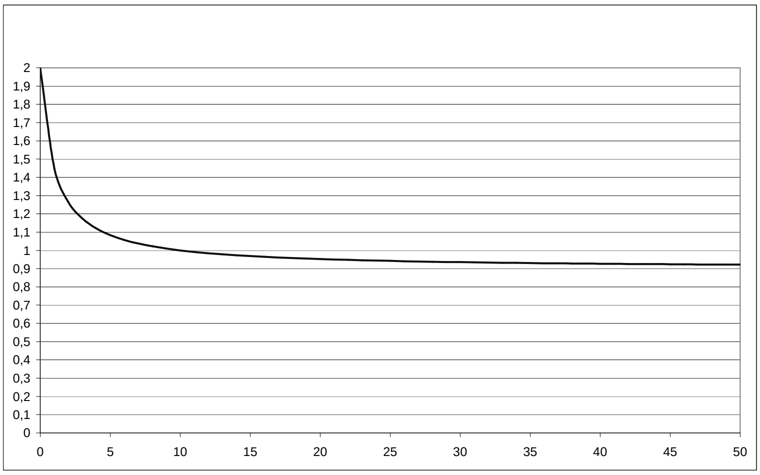
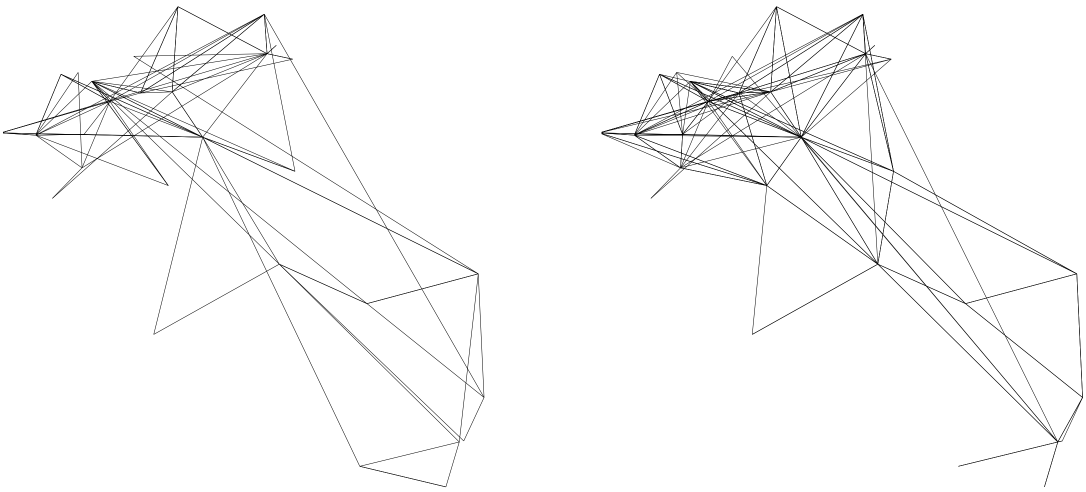


Figure 7: The multiplicative factor $h(y_{ij})$.



(a) (b)

Figure 8: (a) The currently used service network. (b) The service network computed by the proposed algorithm.

Tables

	North America	**Italy**
Planning process	Trains leave when a sufficient cargo tonnage is reached. Priority to freight trains.	Trains leave on schedule. Priority to passenger trains.
Railroads	Many single railroads dedicated to freight trains.	Few single railroads. No railroad is dedicated to freight trains.
Length/weight limits	Up to 5000 m. Trains with several pulling locomotives. A container can be carried on top of another one.	Up to 600 m. Trains with at most 2 or 3 pulling locomotives. No container can be carried on top of another one.
Blocks of cars	Widely used. Automatic hooks.	Rarely used. Manual manoeuvre.

Table 1: Freight rail transportation in North America and in Italy.

	Random instances	Real instance
Number of instances	25	1
Number of yards (n)	5 to 10	39
Number of commodities (p)	2	375
Distances (Km)	$U[0, 2000]$	0 to 2000
Demands (number of cars)	$U[-2000, 2000]$	-2000 to 2000
Car handling cost (euro)	$U[0, 50]$	0 to 50

Table 2: Characteristics of the random generated and of the real instances of the problem. Here $U[a, b]$ denotes a uniform probability distribution on the interval $[a, b]$.

ID	# of yards	OPT	HEUR	Deviation
1	5	2466557	2466557	0%
2	5	962219	962219	0%
3	5	1208717	1208717	0%
4	5	3003811	3003811	0%
5	5	1177365	1177365	0%
6	6	1067901	1167883	9.36%
7	6	2144221	2144221	0%
8	6	600882	650166	8.20%
9	7	3342363	3342363	0%
10	7	2670854	2670854	0%
11	7	1401017	1527085	9.00%
12	7	1362462	1362462	0%
13	7	1098599	1100798	0.20%
14	8	1563992	1746920	11.70%
15	8	2314114	2314114	0%
16	8	1422406	1422406	0%
17	8	1506203	1506203	0%
18	9	843020	875676	3.87%
19	9	453658	453658	0%
20	9	2004919	2004919	0%
21	9	1947572	1947572	0%
22	10	2058998	2080492	1.04%
23	10	1536016	1536016	0%
24	10	1586018	1586018	0%
25	10	1404944	1420350	1.10%
Average deviation				1.78 %

Table 3: Testing the algorithm on the set of random generated instances.

Parameter	Current network	Algorithm solution	% Comparison
C	11767851	11318144	-3.82%
S	144	158	+9.72%
T	4258	3851	-9.56%
TK	1060889	1027310	-3.17%
CK	21217665	20546329	-3.16%
M	170390	154014	-9.61%

Table 4: Comparison between the currently used service network and the service network proposed by the heuristic algorithm.