The capacity analysis of a railway node: a saturation-based methodology and its application to Novara freight terminal

Nicola Coviello[1*], Bianca Pascariu[1]

[1]Politecnico di Torino, DIATI, Transport Systems

*Corso Duca degli Abruzzi, 24, 10129 Torino (Italy). nicola.coviello@polito.it

Abstract

Due to the growing interest in the last decade in rail freight transport, the infrastructure managers are now facing a considerable problem: networks congestion. An adequate level of service, in terms of capacity and operations quality, shall be ensured at the same time. The paper presents a microscopic simulation/saturation method for the railway capacity evaluation, implemented within an integrated analysis environment developed at Politecnico di Torino. The method is applied to validate the effects of a current RFI project which involves infrastructure improvements in the rail freight node of Novara, Italy. The capacity of the node is evaluated by saturating the current timetable adding as many additional paths as possible, while respecting a set of technical and operative constraints. In order to highlight the actual bottleneck of the network, proper KPIs are measured on the saturated timetables and an UIC-compression analysis is performed as well.

Keywords: Railway Capacity; Saturation; Timetable; Freight trains.

1 Introduction

1.1 Aims and purposes

Congestion of railway networks is becoming a central issue in several European corridors, due to the steady increase of rail freight transport demand in the last decades. A proper level of service shall anyway be ensured, both in terms of capacity and reliability. This can be accomplished by optimising the planned traffic, by means of infrastructure improvements of through a proper blend of the previous two measures.

A case-study for this phenomenon can be retrieved in the railway network of north-western Italy, pivoted in the Novara rail node. Two main European freight corridors converge in Novara, namely the Mediterranean and the Rhine-Alpine. The latter is taking advantage from the opening of the Gotthard and Ceneri (in 2020) base tunnels in Switzerland. In order to avoid or limit possible critical congestion deriving by an increase of the freight traffic through the Novara node, the Italian Infrastructure Manager RFI defined an infrastructure improvement programme. It aims at increasing the overall capacity of the node and improving the reliability of the related scheduled services.

The aim of this study is therefore to analyse the node capacity in two infrastructure configurations, i.e. with minimum and maximum improvements, thus considering two infrastructure scenarios. A timetable-based approach has been used, based on the microscopic simulation of rail traffic followed by the saturation of the current passenger timetable.

According to Delorme et al. (2001), saturation is the process in which additional courses (saturating courses) are inserted into a given timetable, while respecting some technical and operative constraints, until no more paths can be added without incurring in a constraint violation. In our approach to the saturation problem, saturating courses are modelled after those (called saturating course prototypes) belonging to a user-defined set, also considering a scheduling priority among them. The technical constraints regard the minimum/maximum run times of the courses and the minimum headways imposed by the signalling system in each block section. Their imposition ensures that the resulting timetable (called saturated timetable) is feasible, which means that courses can actually comply with the scheduled arrival and departure times during real operation. The operative constraints regard the minimum/ maximum dwell times of saturating courses in stations and connections times. Their imposition ensures that the saturated timetable actually match the commercial requirement for the saturating courses.

A saturated timetable is assumed to take advantage of all the available capacity. This can therefore be quantified by calculating proper Key Performance Indexes (KPIs) on the saturated timetable. A *compression analysis*, performed according to the UIC leaflet 406R (2013), is furthermore suitable to point out where the actual local bottlenecks lie within the studied network.

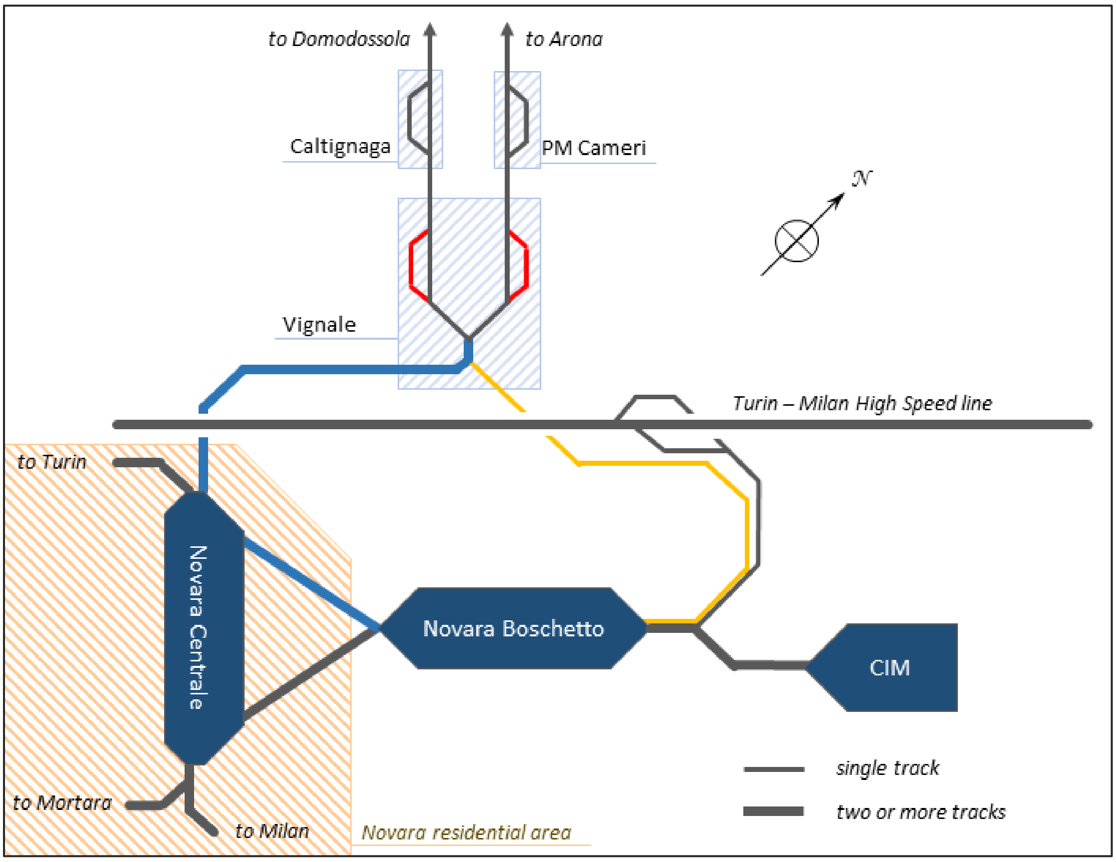


Figure 24 The Novara node with the planned infrastructure improvements.

The considered railway system (Figure 24) includes the two major railway stations of Novara (the passenger terminal) and Novara Boschetto (the freight yard, connected to the CIM intermodal terminal), and a part (about 20 km) of the two north-bound railways towards the Swiss border, with some minor stations. This system can be classified as a *complex railway node* according to Malavasi et al. (2014), since it includes different functional alternatives which shall be considered at the same time during the saturation process. In facts, available capacity is exploited by a synchronous selection of these alternatives, which can be listed as follows:

- the different routing alternatives within the two freight terminals of Novara Boschetto and CIM;
- the different routing alternatives between the two freight terminals, the Vignale station and hence the north-bound lines;
- the different routing/scheduling combinations within the Vignale, Caltignaga and PM Cameri stations, which make it possible to schedule crossings or overtakings.

This saturation-based approach is implemented in the SASTRE calculation package developed at Politecnico di Torino and devoted to the simulation and the analysis of railway systems.

1.2 Methodology framework

In this paper we quantify railway capacity by means of the maximum number of trains which can be operated on a given railway infrastructure, in a certain time window (one day) and with given operative constraints (Abril et al, 2008). In practice, this indicator shall be carefully considered, since it strongly depends on several operative parameters like the traffic mix and the utilised rolling stock and the minimum level of service required by commercial needs (KFH Group, 2013). As a result, an absolute definition of railway capacity does not exist, as stated by the very UIC: "*The capacity of the railway infrastructure is not static, it depends on the way it is utilised"* (UIC 2013). Because of this relativity, it is not possible to define a univocal method to evaluate railway capacity, instead a number of methods has been formulated in the last 50 years, each of them addressing a particular application range.

Methods for railway capacity evaluation can be classified into three main categories (Kontaxi and Ricci, 2009): synthetic, analytical, and simulation-based ones. The latter can be considered optimisation methods when they produce a result which is optimised with regard to one or more variable, according to Abril et al. (2008) and Pouryousef et al. (2015). Figure 25 illustrates the main features of these three categories: pros are highlighted in green, cons in red. Synthetic and analytical methods rely on simplified models of the railway systems, require relatively few input data and provide a rather immediate applicability. By contrast, each of these methods is normally designed for its specific application scope, and it is therefore valid only within certain ranges of the parameters which do not explicitly figure in the related formulas. As a result, when applied to a case study laying outside it design scope, a synthetic or analytical method would produce potentially misleading or even wrong results. From another point of view, two or more different methods applied to the same case study would likely provide different results (Abril et al., 2008).

Simulation-based methods represent a general-purpose approach and are not limited to particular application cases. They are normally based on an explicit model of the railway system, whose granularity can range between microscopic and macroscopic. Even if the granularity of the model likely affects the overall accuracy, it preserves the versatility and scalability of these methods. On one hand, simulations reproduce the railway traffic behaviour, considering several technical and operative constraints as well as stochastic phenomena as perturbations. On the other hand, the setup of these models (mainly if with microscopic granularity) requires a big amount of input data which are then handled by complex simulation algorithms. The latter can moreover be affected by significant processing times and by the risk of not-converging because of the occurrence of deadlock conditions.

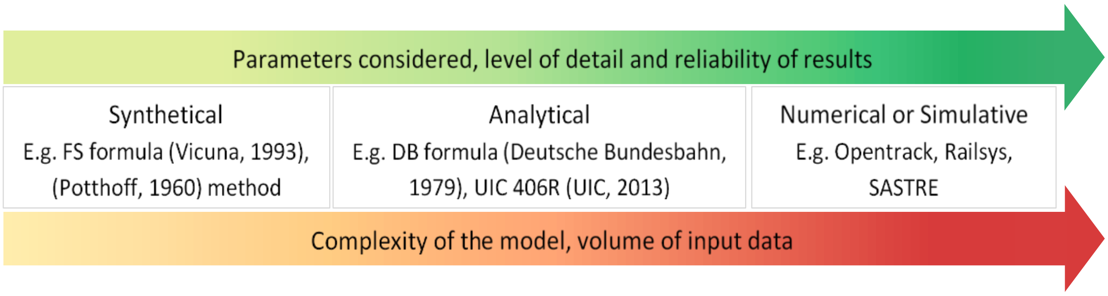


Figure 25 Methods for railway capacity calculation.

Analytical methods provide a trade-off between model complexity, amount of input data requested on one side and results accuracy and reliability on the other. The analytical method proposed by the UIC leaflet 406 R is adopted by several European Infrastructure Managers (RFI included) to calculate the capacity consumption of a given working timetable (Pouryousef et al., 2015). This method is of straightforward application in case of simple application case, namely those of linear networks where full-line operation is prevalent (a typical application case is provided by UIC, 2008).

Anyway both feedbacks from RFI analysts (gathered during preliminary technical meetings) and consolidated literarure sources (Landex and Jensen, 2013; Landex, 2011) agree upon that the UIC method is unsuitable to assess complex infrastructure topologies, like those of large stations or of highly interconnected railway nodes. In these cases, the reciprocal interdipendence of a huge number of possible alterative routes severely affects the symultaneous operation of trains in the same node area. Furthermore, additional technical and operative constraints could apply, which are not normally present in full-line operations. This behaviour cannot be properly grasped by analytical methods as the UIC one, thus resulting in a potential inaccuracy of the results.

Hansen (2000) points out how the application of analytical methods to large stations or nodes shall be supported by further empirical considerations drawn from real traffic data. In these case, simulations can be used to validate results.

In general, simulation methods represent the best and possibly the only universally valid approach to consistently analyse the capacity of complex railway nodes. Several simulation methods, embedded in calculation packages, are available, most of them under commercial licensing, as for instance: RTC (mostly utilised in North America, Sogin et al., 2013); MultiRail; RAILSIM; OpenTrack (Nash and Huerlimann, 2004); RailSys (Bendfeldt et al., 2000); CMS (Sameni at al., 2011, Abril et al., 2008). In general, these packages are conceived and designed with synchronous simulation in mind: they provide several and powerful functionalities for simulating and analysing a given timetable (assumed as feasible), possibly considering traffic perturbations. On the other side, they do not include tools for the automatic generation of feasible timetables: this task is up to the user, and can only be supported by graphical editors provided by some packages. Barber et al. (2007) and Pouryousef et al. (2015) provide comprehensive reviews about these tools..

Railway traffic simulation models can be used to arrange input datasets for the resolution of the Train Timetabling Problem (TTP, Hansen and Pachl, 2014). This classical problem in the field of operation research consists in defining the arrival and departure times of train in stations and in selecting their routing across the network. In the bargain, a solution of the TTP is represented by a timetable which is *optimal* according to one or more objective functions (e.g., minimisation of the total travel times, minimisation of the total connection time, …) and which respects a set of constraints. In some models, a controlled violation of certain constraints is accepted. Constraints can be roughly classified into technical and operative ones. The firsts make the resulting timetable feasible, i.e. ensure that in real operation trains can respect their schedule, at least without traffic perturbations (Goverde and Hansen, 2013). Operative constraints foster the compliance with commercial or organisational needs (passenger connections and transfer times, crew switches and rolling stock turnover, …).

Several approaches have been pursued to solve the TTP, like Mixed Integer Linear Programming (MILP) methods or meta-heuristic techniques. Cacchiani et al. (2016) provide an extensive and up-to-date review on this topic. A particular declination of the TTP is the railway timetable saturation problem, regarding the generation of feasible timetables which utilise all the available capacity (Delorme et al., 2001). This is generally accomplished by maximising the number of scheduled paths, by inserting additional courses into a given timetable while respecting given technical and operative constraints. By analysing a timetable assumed as saturated through a proper set of KPIs it is possible to get a numerical quantification of the capacity of the system. In this way a saturated timetable depends both on technical constraints (which are considered invariant) and operative ones, which figure out just one of the several different ways in which the system can be operated. As a result, capacity shall not be represented by a set of punctual values of the aforementioned KPIs, but rather by their variation ranges.

The open railway market is encouraging European Infrastructure Managers to carefully define and quantify the capacity of the relevant railway systems, together with the possible optimal ways to fully exploit it. Railway Operators, who are confronting themselves with growing competition, apply for more capacity both in terms of number of train paths and of their commercial requirements. The application of advanced methods for evaluating capacity represents a key action for a better exploitation of railway networks. Nowadays, significant research efforts are being carried on in order to improve the methods and the algorithms devoted to the solution of the TTP. Main current issues concern their actual applicability to real case study (characterised by big-sized problem instances), both in terms of computational time and resources utilisation. A review of the saturation method proposed so far is presented in Coviello et al. (2017). One of the most recent contributions is that of Pellegrini et al. (2017), which presents a saturation method based on a MILP formulation.

Significant research efforts have been devoted in developing and implementing methods for calculating full-line capacity. By contrast, node capacity appears to be a rather neglected topic, or tackled admitting a certain methodological weakness The very UIC (2013) reports a general deficiency of consolidated studies and data about node capacity, to be used as a reference. The present study presents an integrated simulation-saturation method to assess the capacity of a complex railway node. A timetable-based approach is used, in which saturated timetables are obtained through a MILP formulation of the TTP based on that proposed by Pellegrini et al. (2017).

Section 2 presents extensively the used method, addressing the microscopic model of the railway system, the simulation of the courses and the saturation of the timetable. The KPIs utilised to evaluate the saturated timetables are presented at the end of this section. Section 3 reports the application of the method to the freight node of Novara, with the description of the case study and the presentation and discussion of the relevant results. Finally, conclusions are drawn in Section 4, highlighting the strength and weakness point of the presented method as well possible future developments.

2 Method

The proposed method relies on 4 main processes, graphically represented by Figure 26. All these processes are carried out within the SASTRE environment (Coviello, 2018). SASTRE is a not-commercial simulation and analysis package for railways systems developed at Politecnico di Torino, DIATI – Transport Systems. Implemented in Python 3, SASTRE is composed by a set of modules which organise the data of the railway microscopic model and provide several main functionalities: to simulate single course runs as well as whole timetables, both with and without stochastic perturbations; to automatically generate feasible timetables; to analyse and compare set of timetables. Furthermore, SASTRE's open architecture allows for on-demand integrations and personalisation by the user. In the next sub-sections, the four main processes are addressed.

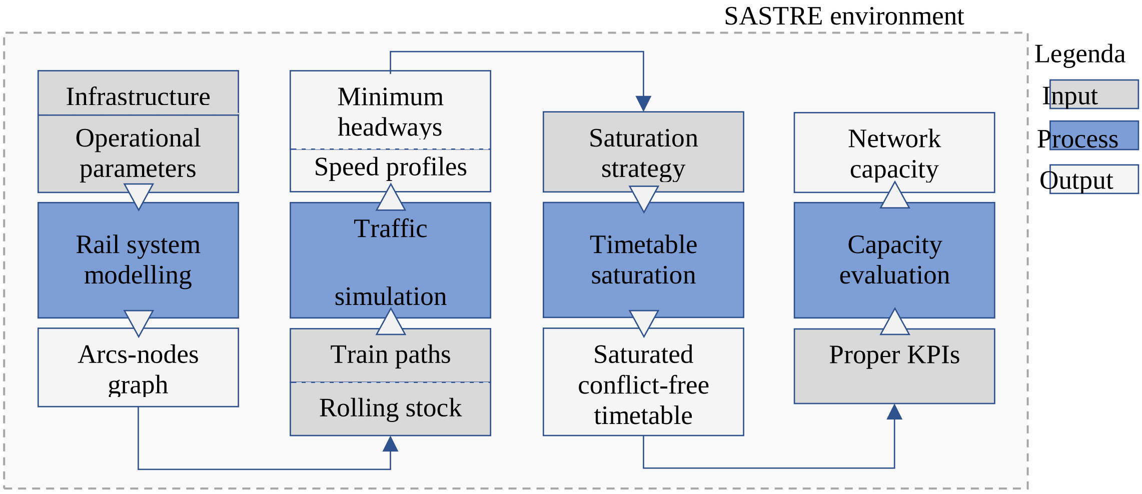


Figure 26 Flow chart of the main processes of the proposed method.

2.1 The microscopic model of the railway system

The microscopic model stores all the relevant data regarding: the infrastructure topology, i.e. the functional and spatial definition of the items composing the railway infrastructure[2]; the characteristics of the active signalling system; the characteristics of the active train control/protection systems, if present; the station routes interlocking dependencies and their possible speed limitations. SASTRE provides an infrastructure editor in which all these variables are uploaded by simply drawing the infrastructure scheme on a canvas, according to the real km-point of each item. After the editing phase, the tool automatically generates an infrastructure graph, where relevant characteristics are assigned to vertexes and edges (Hansen and Pachl, 2014). Each edge represents a *topological element (TE)*, characterised by constant values of the continuous infrastructure attributes as speed limits, grades, curve radii, etc. Nodes are defined between edges which differ in at least one continuous characteristic, or where punctual infrastructure items (switch points, signals, balises, timing points, isolated joints, etc.) are present.

This microscopic graph model is utilised to simulate courses and timetables. When dealing with operations on timetables (consisting in manual arrangements, simple feasibility checks or automatic resolution of the TTP), the SASTRE environment makes use of a more aggregated infrastructure model, whose fundamental element is the *track detection section* or TDS. The TDS is represented by an unordered set of topological element. Extensive descriptions of the TDS-based infrastructure model are provided in Coviello

[2] Tracks, switches and the related speed limitations, grades, curve radii, etc.; signals with their type (block, home, departure, distant) and aspect code; timing point; stations and service locations; extent and characteristic of power supply areas; position and possible additional route release constraints of level crossings.

(2018) and Pellegrini et al. (2017). As explained by the next section, this model is automatically arranged by SASTRE through the simulation of the courses runs.

2.2 Railway traffic simulation

This section addresses the simulation of the courses runs on the microscopic TE-based infrastructure model graph. Courses are defined by a rolling stock type and by a scheduled timetable, describing both the routing (or the different possible routing alternatives) of the course within the system and the programmed arrival, departure or passing times in each timing point along its journey (*Figure 27*).

With these information, the program simulates the courses and the concurrent behaviour of the signalling system. Speed profiles are calculated by numerically integrating the fundamental motion equation. Simulation is normally carried out considering the mutual interactions between courses produced by traffic perturbations or by a not-feasible timetable. In this case, the signalling system would intervene to regulate the course utilisation of shared resources (as TDS or station routes), thus affecting the relevant speed profiles.

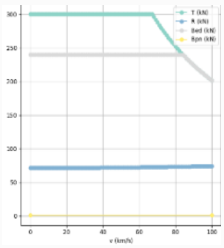
In SASTRE it is possible to perform simulations "turning off" the signalling systems, i.e. by ignoring the reciprocal interactions between courses. The resulting speed profiles are therefore those of trains that meet signals always displaying free-way aspects. By contrast, possible speed limitation imposed by the signalling system but not due to the interaction with other courses[3] are considered. These simulations are used to arrange the TDS-based model used to generate or edit feasible timetables, whose traffic is assumed to travel with free-way signals, at least in unperturbed conditions.

[3] Example of this kind of speed limitations are those imposed by control systems like the ETCS or the Italian SCMT while approaching a station stop or a diverging route.

TRAIN COURSES

Rolling stock

- Total mass
- Total length
- Max. acceleration
- Max. speed
- Braking mass %
- Type of train
- Speed class
- Motion resistance
- Tractive characteristics

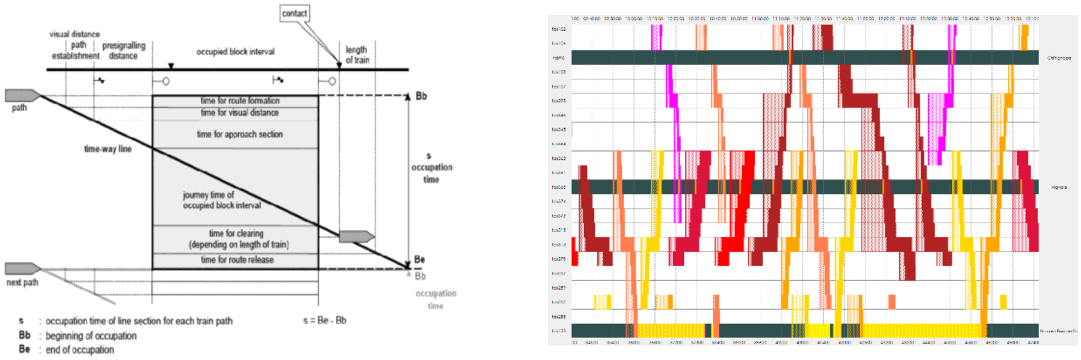


Train path

- Itinerary (topological segments)
- Time constraints
  e.g. dwell time at locations
- Constraints related to the signalling system
  e.g. speed during approach to a stop, speed limits, release speeds, …

Figure 27 Attributes of a course.

By applying the blocking time theory (UIC, 2013) to the simulated run profiles of courses, it is possible to compute the utilisation times of each TDS by the courses which travel through it. As described in Coviello (2018), the utilisation of a TDS by a course starts when the TDS is reserved by the signalling system and ends when it is cleared by the train and released by the signalling system (Figure 28.a). The utilisation intervals of more TDSs belonging to the same system can be represented in a Gantt chart (Figure 28.b). According to Szpigel (1973) and Liu and Kozan (2009), the Gantt chart help to figuring out how the TTP is actually a *job-shop scheduling problem*. Each TDS utilisation represents an activity, while all the consecutive TDS utilisations by the same course represent a process. The latter is actually graphically represented by a *utilisation time stairway*, and constitutes the capacity consumption of the course in point.



a) b)

Figure 28. a) Utilisation time of a block section, and its components (UIC, 2013). b) Gantt chart with the utilisation time stairways of a set of TDS (produced by SASTRE). Time is on the x-axis, each TDS is represented by a slot on the y-axis.

2.3 Timetable saturation

The definition of the utilisation stairways permits to calculate the minimum headways allowed by the signalling system between each pair of possible consecutive courses. Graphically, this can be performed by moving two stairways closer to each other until at least two TDS utilisation slots come into contact. A slots overlap represents a violation of a technical feasibility constraint, since it means that two trains are scheduled with a headway that cannot be respected in real operations. In that case, the simultaneous utilisation of the same TDS by the two courses will be prevented by the signalling system. It will stop one of the two concurrent courses which will inevitably suffer a delay. This delay would likely prevent the course from respecting its scheduled timetable.

The so computed minimum headways are used by the timetabling algorithms implemented in SASTRE to solve the TTP and produce feasible timetables. In order to improve their stability[4], it is possible to impose a certain amount of buffer time between the courses utilisation stairways, defined for each TDS as the minimum time span which shall separate the utilisation slots of two consecutive courses. Buffer times are useful to limit the spread of minor traffic perturbations, and actually involves a capacity consumption.

The proposed saturation approach is composed by two functional level: the base saturation algorithm (BSA) and the saturation strategy (SS).

[4] Goverde and Hansen (2013) define the stability of a timetable as it capability to cope with minor traffic perturbation making use of internal resources only (i.e. exploiting the scheduled run time margins and buffer times) without resorting to major dispatching actions, as course deletion, reordering or re-routing. In this paper we attain to this definition, being aware that different authors use the term “robustness” for the same concept.

The BSA implemented in SASTRE is based on a MILP formulation which extends that proposed by Pellegrini et al. (2017). Given a base timetable and a set of reference courses (called saturating courses prototypes, SCP), the algorithm produces a saturated timetable, obtained by adding copies of the SCPs into the base timetable within a given time window. Copies of the SCPs can differ from the prototypes since the algorithm can modify their station routing as well as their scheduling. In particular, their entry time in the system and their dwell times in stations can be altered, resulting in a shift or in a compression/stretching of the whole paths. The objective of the BSA is to generate a feasible timetable in which the number of saturating courses is maximised. Alternatively, an upper bound to the number of schedulable saturating courses can be set. This is useful when the BSA is launched within a saturation strategy, in order to leave capacity to the following iterations (see next paragraph). The based timetable can be set as fixed or as variable: in this latter case, the BSA can modified its courses (both in routing and scheduling) in order to accommodate more saturating paths.

The BSA do not apply any choice priority among the SCPs, their selection being guided just by the objective of maximising the overall number of saturating courses. To this purpose, the Saturation Strategy functional layer is designed, with higher hierarchical level than the BSA. The task of the SS is to iteratively launch several instances of the BSA with different SCPs. Each instance saturates the timetable obtained in the previous iteration. In this way, the priority of each SCP set is given by its position within the iterations sequence. The first iterations would have more available capacity, and so more courses copied from the relevant SCP sets will likely be scheduled. This applies if the courses of the different SCP sets share at least one TDS, since each iteration would consume capacity in just a portion of the whole considered system.

The characteristics of the courses of each SCP set implicitly implement the prioritisation criterion, which can be either functional or spatial. A functional criterion applies when each SCP set represents a particular kind of rail service: in this way, for instance, it is possible to saturate a timetable according more priority to passenger trains rather than to freight ones. In the Novara case study, functional prioritisation is used to favour the scheduling of additional “rolling highway” trains instead of conventional intermodal trains. A spatial criterion differentiates the courses of each SCP set according to their routing within the network, having fixed the start and final localities.

In our approach we explicitly differentiate between station and network routing. The first is managed by the BSA through the MILP formulation, and deals with the choice of tracks and platforms as well as of the entry and exit routes in each stations. Network routing is handled by the SS and, as the name suggests, controls the choice of alternative journeys which cross different parts of the system.

In the Novara case study, spatial prioritisation has been used to foster the utilisation of the new direct link between Novara Boschetto and Vignale instead of the old itinerary through the Novara passenger station.

As a result, the SS iterations produce a feasible saturated timetable which is assumed to exploit the 100% of the available capacity, with regard to those infrastructure areas utilised by the SCPs. We intentionally use the verb “assume” when mentioning the 100% capacity consumption: since the BSA resolution is normally subjected to a time limit (in order to contain the overall computation time), optimality cannot be guaranteed in all the SS iterations.

2.4 Capacity evaluation

A saturated timetable permits to estimate the system capacity in terms of number and type of scheduled courses. If the saturation is carried out considering buffer times (or run time margins, or in general *ex-ante* measures aiming at improving stability), the analysis returns a practical capacity. Differently, theoretical capacity is estimated. According to Abril et al. (2008), theoretical capacity is the absolute upper bound of the number of trains that can be operated on a railway system. A timetable exploiting all the theoretical capacity will probably be only nominally feasible. This means that even if it does not present scheduled conflicts, it would not be able to cope with the even minimal traffic perturbations that always affect real operations.

Theoretical capacity ignores traffic perturbations. Differently, practical capacity assumes that a part of the theoretical one shall be consumed by buffer times and run time margins. This improves timetable operations during real operations. The relevant maximum number of trains therefore depends on the desired level of service (mainly in terms of courses punctuality) and on the intrinsic reliability of the system (in terms of occurrence of primary delays). Practical capacity can be guaranteed in normal operative conditions[5].

In this study capacity is evaluated by means of two KPIs, computed on saturated timetables: the number of daily pairs of saturating freight courses (a pair being composed by an arriving train and by the corresponding departing one) and the infrastructure Occupancy Time Rate (OTR). The first KPI is directly obtained from the timetable, and represents an absolute estimation of the capacity (theoretical or practical, depending on the presence of buffer times). The second KPI is computed by means of the compression method presented by the UIC 406R leaflet (2013). Compression is applied on subsets of TDSs (called

[5] It is worthwhile to point out that the definition of the *normal* operative conditions varies according to the particular case study. For instance, traffic perturbations which are considered as acceptable in certain contexts can represent an exceptional anomaly in others. Or, from another point of view, the required level of service significantly depends on the application case.

compression sections, CSs), and consists in shifting the courses as closer as possible while avoiding overlaps of the utilisation stairways within the CS. With a proper definition of the CSs, this method can be utilised to evaluate how and how much capacity is consumed in different parts of the system. Figure 29 graphically illustrates the compression process as well as the relevant values, used to compute the OTR.

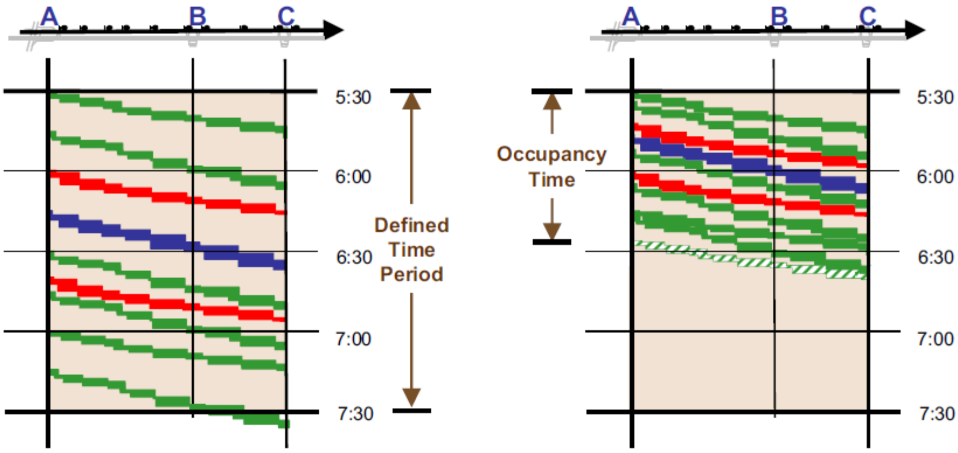


Figure 29. Graphical representation of the compression process (excerpt from UIC 406R, 2013)

$$\text{OTR} = \frac{Occupancy\ time}{Defined\ time\ period} \cdot 100 \qquad (\%)$$

CSs are defined by the user and have a major influence on the computation of the OTR. The UIC 406R leaflet provides indications about a correct definition of CSs in full-line stretches. By contrast, the leaflet admits the difficulties of applying the compression method to node areas, in which CSs shall be carefully defined according to each application case. Landex (2011) describes how the OTR value can radically change (thus figuring out misleading capacity levels) according just to the CS definition even in a very simple system constituted by a single-track line. The OTR index is therefore not suitable to evaluate the absolute capacity consumption of a complex railway node as a whole. On the other side, it can be used to identify bottlenecks within complex networks, characterised by a high *local* capacity consumption (Rotoli et al., 2015, Goverde and Hansen, 2013). These bottlenecks could represent a stability issue (because of the occurrence of knock-on delays) when the OTR is higher than given thresholds.

The UIC 406R leaflet indicates some values for these stability thresholds, obtained from empirical evaluations on European case studies (Table 14). The leaflet explicitly differentiates between full-line stretches and station/node areas. For the latter, minor reliability of the data is admitted due to the lack of empirical studies as well as to the implicit higher variability of the phenomenon. Rather wide thresholds ranges are proposed for station areas, instead of the punctual values relevant to full-line stretches.

Table 14 OTR stability thresholds proposed by UIC leaflet 406R for lines and nodes.

		Peak hours	Daily period
Full lines	Suburban lines with passenger traffic	85%	70%
	High speed lines	75%	60%
	Mixed traffic lines	75%	60%
Nodes	Switch areas	60%...80%	
	Station tracks areas	40%...50%	

When the UIC compression method is applied to a saturated timetable, these bottlenecks indicate the spots which are preventing a further insertion of additional saturating courses. By relieving these bottlenecks (e.g. thanks to infrastructural improvements) more courses are likely to be accommodate, possibly moving the bottlenecks to another part of the system.

3 Case study: the Novara node

3.1 Input data

The evaluation approach has been applied to a case study provided by the rail freight node of Novara (Italy). The purpose is to analyse the effects, mainly in terms of capacity, of some infrastructure improvements planned by the national infrastructure manager RFI. In this study, with the term "Novara freight node" is meant:

1. The Novara Boschetto freight yard, which includes the storage yard (in this paper also referred to as NB), the rolling highway terminal and the intermodal terminal CIM (in Italian *Centro Intermodale Merci*);
2. The two northbound lines passing through Vignale stations (towards Borgomanero and Arona), up to the Caltignaga and PM Cameri stations respectively;
3. The Vignale - Novara Centrale - Novara Boschetto urban rail route, currently the sole access route to the Novara Boschetto yard.

A preliminary Origin/Destination survey of rail traffic supply and demand at the Novara Boschetto freight yard highlighted the predominance of northern routes traffic. On the remaining routes, freight traffic resulted negligible, and therefore has not been taken into account this study.

The infrastructure improvements planned in this area by RFI, with a time horizon of 2022, include:

1. The activation of a direct link route (in yellow in Figure 24) between the northbound lines to Domodossola and Arona and the northern-eastern tracks root of Novara Boschetto. This link route is intended to improve capacity by providing freight trains an alternative route to the current one which passes through the Novara Centrale passenger station (in blue in Figure 24). Meanwhile, the link would divert heavy freight traffic from densely inhabited residential areas which are nowadays affected by significant noise pollution caused by the railway;
2. The reconfiguration of the topology of Novara Boschetto yard, including the construction of a new rolling highway terminal (in blue in Figure 30) and the upgrade of the sidings to the European standard length of 750 m;
3. The activation of two passing loops in the Vignale station (in red in Figure 24, in red and orange in Figure 30), one for each northbound line. These loops are intended to provide a capacity buffer for freight trains to and from Novara Boschetto.

The analysis was carried out modelling two infrastructure scenarios that reproduce the gradual implementation of the described improvements, in order to point out the impact of each one. Figure 30 shows the microscopic model of the infrastructure as set up in SASTRE, highlighting the current infrastructure and the interventions introduced in each scenario.

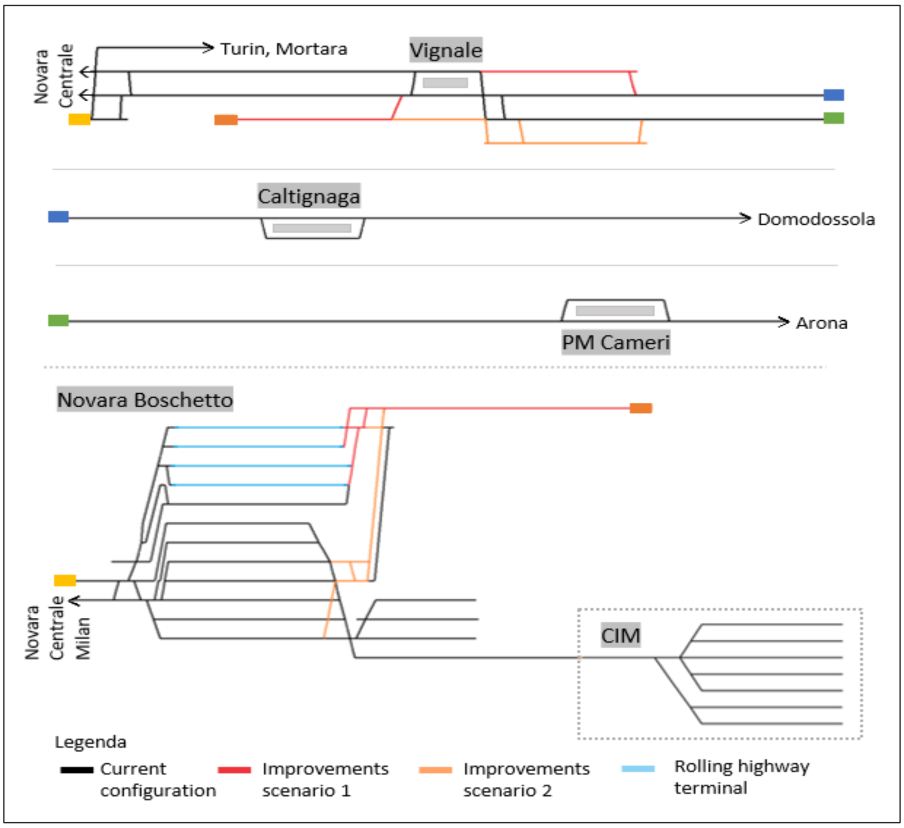


Figure 30 Microscopic infrastructure model.

The considered area has a total extension of about 30 km and is characterised by a maximum speed of 90 km/h due to the limitations imposed by the braking grade according to Tables B of the RFI PGOS (RFI, 2016) and to the active signalling and control systems. In fact, on the considered lines, the ETCS Level 1 is active, relying on fixed balises (automatic block signalling integrated with the Italian *Sistema Controllo Marcia Treni* SCMT control system). If ETCS Level 2 were used, the limit would have been 120 km/h, according to RFI prescriptions. After improvements, new switches will have a maximum speed of 60 km/h, while the old ones feature a 30 km/h limit. On some of them, for the rolling highway trains the limit is further reduced to 10 km/h because of their particular running gear. The SCMT system imposes a 30 km/h release speed[6]. Within the Novara Boschetto and CIM yards, the maximum speed limit is 30 km/h and 15 km/h respectively. Simultaneous entry is permitted in all the stations of the considered network.

A statistical sampling of the traffic from Domodossola/Arona has been carried out in order to define the rolling stock for the simulations. It emerged that the node is characterised

[6] The release speed is that imposed to a train which is approaching a closed signal, within a certain distance from it.

by a high volume of freight traffic, both rolling highway and intermodal, and by a relatively small volume of regional passenger traffic. The trainset type applied to each of these categories is that highlighted as most recurring by the sampling. Thanks to the lines upgrade to P/C 80 loading gauge, 750 m module and 22.5 t axle weight, we considered long and heavy freight trains composed as in Table 2. This table indicates the types of used locomotives and carriages and the corresponding number of units per trainset. For passenger trains, a trainset consisting of an E464 locomotive, a double-decker coach and three low-floor coach is considered. The freight trains feature a maximum speed of 140 km/h, while passenger trains have a maximum acceleration of 0.6 m/s^2 and a maximum speed of 160 km/h. However, the maximum operating speed is limited by the infrastructure to 90 km/h, as mentioned above.

Table 15 Composition of freight trainsets.

	Rolling Highway		**Intermodal Freight Trains**	
	Type	N° units	Type	N° units
Locomotives	E474	2	E484	2
Carriages	UIC-X	1	Sgns	20
	Saadkms	33	Sggmrs	10
Length (m)	700		700	
Mass (t)	1860		2720	

Figure 31 shows an example of the speed profile (obtained from SASTRE simulations) of a rolling highway train. The train is moderately affected by the high mass during acceleration, despite its high length and mass. However, its maximum speed on open lines does not deviates significantly from the operating limit of 90 km/h, leading to the conclusion that the mass does not constitute a stricter limitation to trains operation than those imposed by the infrastructure (specifically by the signalling and control system).

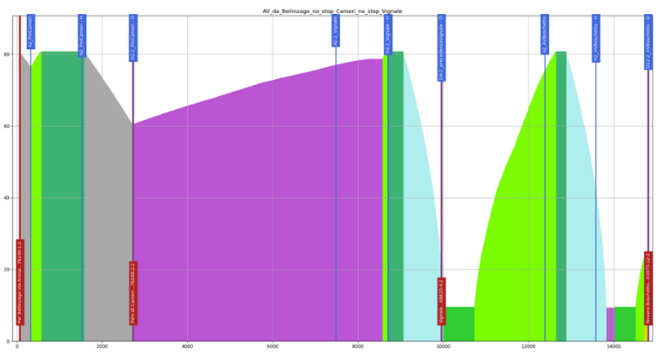


Figure 31 Speed profile of a rolling highway train.

The base timetable used as a reference in this analysis capacity is that of October 25th, 2018. It has been obtained from the RFI M53[7] modules. Figure 32 shows the current traffic in the node (train pairs/day) on the peak day of the week (Friday), for each location of the studied area. TCA stands for Accompanied Combined Transport (rolling highway), TCnA for Not Accompanied Combined Transport (intermodal freight trains) and REG for regional passenger trains. Future operative models foresee an increase of the demand for freight train paths, with a minimum of 8 more TCA pairs utilising the line passing through PM Cameri. Figure 33 shows the network routing of all the course types considered in this study.

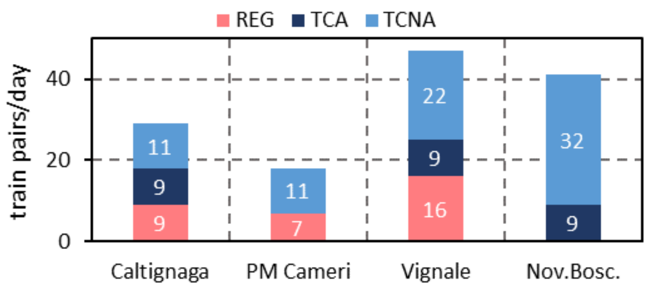


Figure 32 Current traffic in Novara node on peak day.

When generating saturated timetables, saturation is applied just with freight trains, while the passenger timetable is kept unchanged. 10 groups of saturating courses prototypes have been used to implement the saturation strategy, as displayed in Table 16 and Figure 33. These groups are actually composed by pairs of courses linked by continuation constraints, which impose that each arriving course shall be followed by a departing courses from the same platform/yard track. The time span between the linked arrival and departure represents a process time within the yard/terminal. For each group the *maxSat* parameter is defined, which sets the maximum number of saturating courses that can be scheduled during a single call of the BSA. The BSA can exploit the sidings in Vignale, Caltignaga and PM Cameri stations to schedule crossings and overtakings. In this case, the min and max dwell times are set to 3 and 60 minutes respectively.

It is worthwhile to point out that in real operations courses to/from CIM would actually end their travel in the NB yard, where the main line locomotive is replaced by a shunter. The latter shunts the wagons into the CIM terminal. Symmetrical operations take place

[7] M53 module is an official document of RFI, issued for each location of the network (whether it is a station or a simple service location) and for each operation day. It reports the scheduled arrival/departure trains of each trains, as well as the relevant station track and entry/departure station routes. By reading the M53 modules of the locations of the model, SASTRE can automatically upload the timetable (both in terms of routing and scheduling) on the microscopic model.

for the departing convoys. This is modelled by using a quadruple of courses linked by three continuations, instead of a pair.

Table 16. Saturation groups used in the saturation strategy

SCP group ID	Rolling stock	Or/Dest	Network routing	Process time min/max (h)	maxSat
1	TCA	AV – Domo.	link	1/1.5	11
2	TCA	AV – Arona	Link	1/1.5	11
3	TCnA	CIM – Domo.	Link	0.1/5 - 5/10 - 0.1/5*	14
4	TCnA	CIM – Arona	Link	0.1/5 - 5/10 - 0.1/5*	14
5	TCnA	CIM – Domo.	Urban	0.1/5 - 5/10 - 0.1/5*	14
6	TCnA	CIM – Arona	Urban	0.1/5 - 5/10 - 0.1/5*	14
7	TCnA	NB yard – Domo.	Link	1/10	6
8	TCnA	NB yard – Arona	Link	1/10	6
9	TCnA	NB yard – Domo.	Urban	1/10	6
10	TCnA	NB yard - Arona	Urban	1/10	6

* 1st and 3rd numbers pairs apply to the shunting continuation in NB yard. 2nd numbers pair applies to the process time within the CIM terminal.

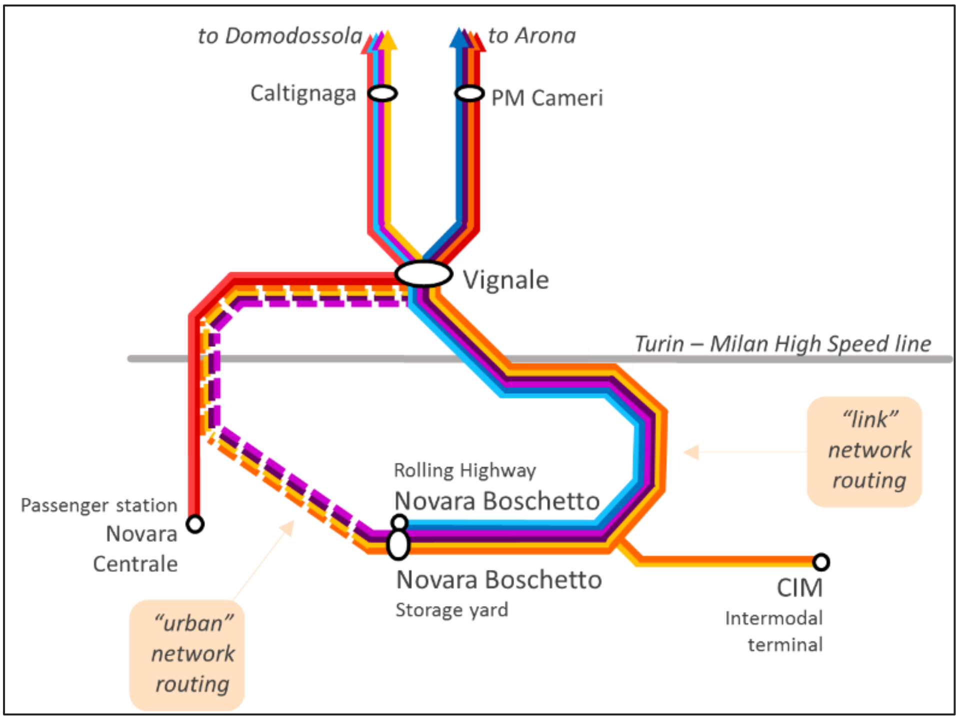


Figure 33 Network routing of the SCPs.

Table 17 displays a pseudo-code implementing the adopted saturation strategy. Firstly (lines 1,2) the rolling highway courses are scheduled. Since the network is still empty, in this phase all the maxSat possible courses are always scheduled. Lines 3 to 10 implement the saturation of the CIM courses. Priority is given to the usage of the new link. Only if it is not possible to schedule all the desired courses with this network routing, further BSA calls (lines 5 to 10) tries to schedule the remaining courses on the urban route, if possible. Finally, (lines 11 to 18) BSA is iteratively called until no more NB course can be added. In this stage too, priority is given to the usage of the link network routing (lines 12,13) over the urban one (lines 14 to 17).

For each infrastructure scenario, three saturated timetables are obtained considering buffer times of 0, 60 and 120 seconds. The latter is the actual buffer time used by RFI to schedule the train paths on this network.

Table 17. Pseudo-code of the saturation strategy

```
1  BSA.saturate (SCP_1)  # BSA.saturate(SCP_x) launch the BSA with a given SCP group, as in Table 16
2  BSA.saturate (SCP _2)

3  nSat_Domo ⇐ BSA.saturate(SCP_3)  # nSat_x is the number of sat. courses scheduled by a BSA call
4  nSat_Arona ⇐ BSA.saturate(SCP_4)

5  IF nSat_Domo < SCP_3.maxSat:
6          SCP_5.maxSat ⇐ SCP_5.maxSat – nSat_Domo  # the maxSat parameter of SCP_x is updated
7          BSA.saturate(SCP_5)
8  IF nSat_Arona < SCP_4.maxSat:
9          SCP_6.maxSat ⇐ SCP_6.maxSat – nSat_Arona
10         BSA.saturate(SCP_6)

11 DO:
12         nSat_Domo_L ⇐ BSA.saturate(SCP_7)
13         nSat_Arona_L ⇐ BSA.saturate(SCP_8)
14         IF nSat_Domo_L > 0
15                 nSat_Domo_U  ⇐ BSA.saturate(SCP_9)
16         IF nSat_Arona_L > 0
17                 nSat_Arona_U ⇐ BSA.saturate(SCP_10)
18 WHILE (nSat_Domo_L + nSat_Domo_U > 0 AND nSat_Arona_L + nSat_Arona_U > 0)
```

3.2 Results

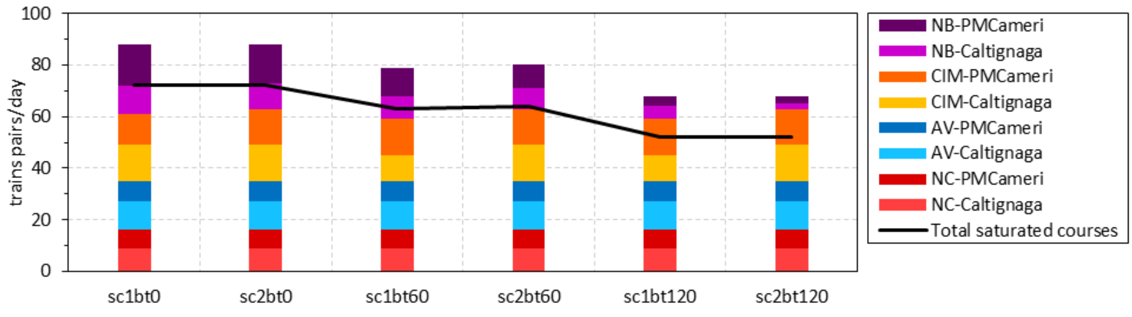


Figure 34 Absolute capacity of the Novara freight node.

Figure 34 displays the absolute capacity (in terms of pairs of trains/day) of the studied network calculated through the application of the presented saturation strategy. A colour code differentiates the different types of saturating courses. NC (Novara Centrale, in red) courses are the passenger ones, whose number is constant in all the saturated scenarios. It emerges that the rolling highway terminal can manage 8 extra pairs of trains with regard to the current situation (see Figure 32), thus matching the RFI objective for the future operating model. The CIM records an increase of saturating courses higher in scenarios 2 than in scenarios 1. We deduce that the infrastructural interventions introduced in scenario 2 facilitate the access to the CIM terminal, thus increasing its capacity. Finally, as far as Novara Boschetto yard is concerned, a relatively low number of saturating courses is scheduled, due to their lowest priority in the saturation strategy.

The black line in Figure 34 shows the trend of the total number of saturating courses highlighting, the dependence on the buffer time. When the buffer time increases, there is a significant decrease in practical capacity, with a ratio of about -14% per each 60 s step: in scenarios bt0 the algorithm inserts 72 pairs of freight trains per day, in bt60 63/64 pairs (in scenarios 1 and 2 respectively), and finally in bt120 52 pairs. Considering the same buffer time, it can be observed that the infrastructural improvements introduced between scenario 1 and scenario 2 do not produce a significant increase in the overall capacity. By contrast, at a local level the analysis of the saturated timetables reveals that the activation of the Vignale-Novara Boschetto link diverts significant traffic from the current urban route (Figure 35). The activation of the new sidings in Vignale station is also an effective measure, as crossings and overtakings are actually scheduled in the saturated timetables. In general, each siding is used by 15 to 20 daily courses trains.

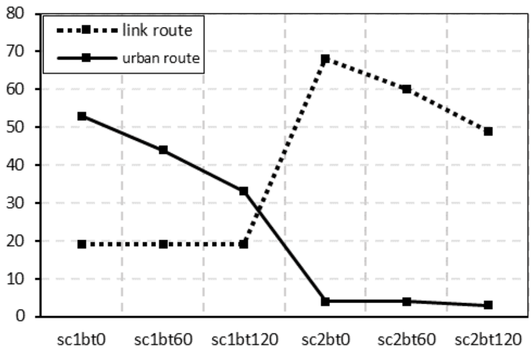


Figure 35 Utilisation of the link route (pairs of trains/day).

Figure 36 shows the Time Occupation Rate (OTR) calculated on main nodes and corridors. By observing Figure 36.a we can remark that the station tracks areas of CIM and Novara Boschetto represent the current bottleneck of the system, due to the high dwell times required by terminal/shunting operations combined with a relatively modest number of available tracks. These evidences would represent the starting point for further focused interventions, in the event that more capacity would be needed.

Comparing the calculated OTR values with the thresholds recommended by the UIC 406R leaflet (Table 14, 60% for mixed traffic over the 24-hour period), it emerges that the stability requirements are satisfied only by imposing a 120 s buffer time (Figure 36.b). In general, the OTR values are rather high. This is due to the fact that the considered lines are single-track, and in the saturated timetables intensive traffic is forced to use all available capacity. In scenario 2, the NB-Vignale link route exceeds the OTR threshold even with a 120 seconds buffer time (Figure 36.c). On the other hand, dealing with freight traffic only, lower punctuality can be accepted in comparison with mixed traffic lines, as stated by RFI technicians. Moreover, in case of delay, the two sidings in Vignale would represent a sort of "capacity buffer" to avoid disruptions of passenger services.

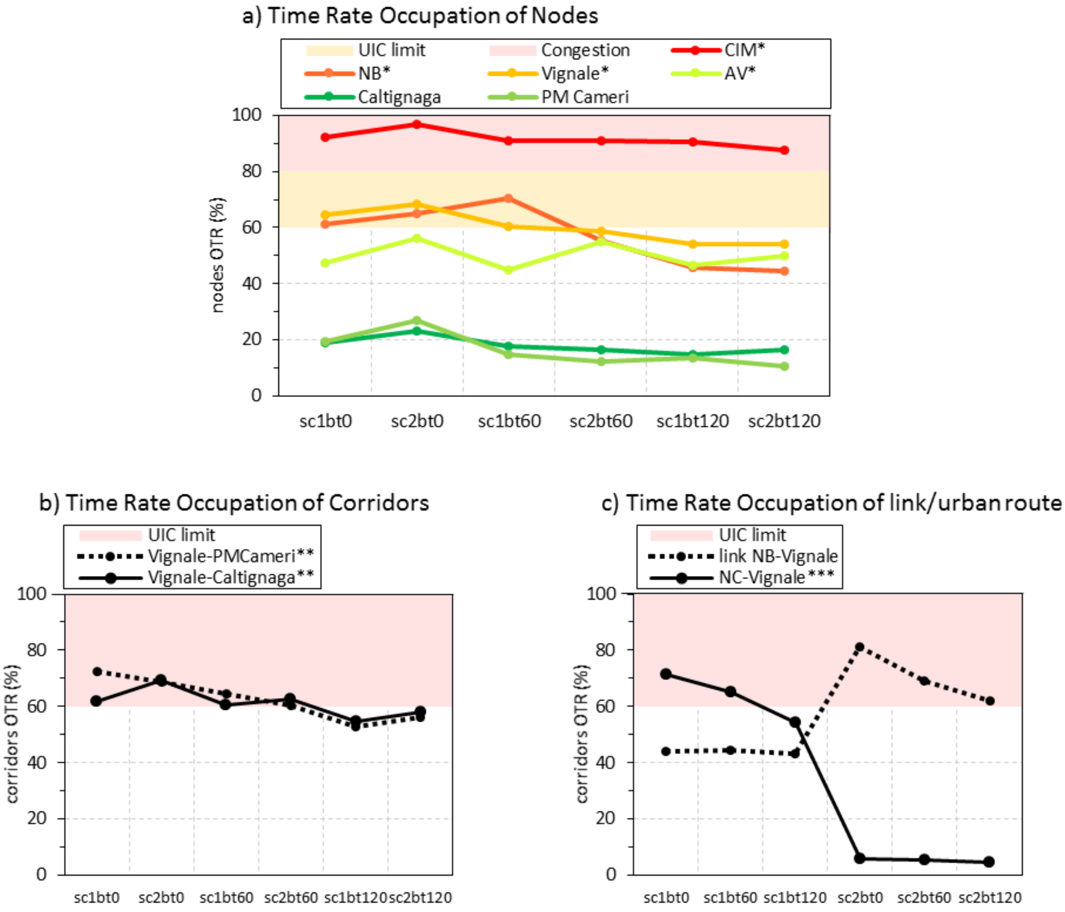


*Figure 36 Occupation Time Rate (OTR) of the infrastructure of: (a) nodes; (b) lines; (c) link route/urban route; (*station tracks + switch area; **including Vignale station; ***excluding Vignale station).*

4 Conclusions

The paper presented a capacity analysis performed on the Novara freight node, assessing the impact of infrastructural improvement measures planned by RFI. The analysis was carried out applying a timetable saturation-based method entirely implemented within the non-commercial software SASTRE, developed at the Politecnico di Torino.

A novel approach to the timetabling saturation problem has been presented, defining two functional layers: the base saturation algorithm and the saturation strategy. The latter implement priorities between the saturating courses prototypes, according to functional and spatial criteria.

The study led to the following considerations:

1. To model operations in a complex node raised the computational complexity of the simulation and saturation processes, therefore a future goal could be to keep

on optimizing the algorithms in order to reduce resolution times even in more complex cases;

2. The extension of the study area to secondary elements (arrival/departure yards, intermodal terminals, junctions, etc.) connected to the main network has allowed to state that the global bottleneck lies in a group of station tracks, and not in a main line;
3. The saturation strategy functional layer represents a novel methodological contribution of this study. It appears worthy of interest to keep on developing this approach, with the aim of figuring out a standard process to define the train priorities.

The use of an integrated method of simulation and optimization permitted to analyse and test in detail different study scenarios in an effective way in terms of time effort, as recognised by RFI analysts. Therefore, in the field of railway planning and management, the relationship between the academic world and the business one should certainly be enhanced, given the mutual relevant benefits.

Acknowledgements

This research activity was carried out in collaboration with the Italian railway infrastructure manager RFI (Rete Ferroviaria Italiana), within the infrastructure improvement project " Novara Node, 1st Phase".

References

Abril, M., Barber, F., Ingolotti, L., Salido, M. A., Tormos, P., & Lova, A. (2008). An assessment of railway capacity. *Transportation Research Part E: Logistics and Transportation Review*, *44*(5), 774-806

Barber, F., Abril, M., Salido, M. A., Ingolotti, L., Tormos, P., & Lova, A. (2007). Survey of automated systems for railway management. *Technical Report*

Bendfeldt, J. P., Mohr, U., & Muller, L. (2000). RailSys, a system to plan future railway needs. *WIT Transactions on The Built Environment*, *50*, 249-255

Cacchiani, V., Furini, F., & Kidd, M. P. (2016). Approaches to a real-world train timetabling problem in a railway node. *Omega*, *58*, 97-110

Coviello, N. (2018). Timetable-based capacity evaluations: Analysis of a sub-urban single-Track railway. *Ingegneria Ferroviaria*, *73*, 287-319

Coviello, P. Pellegrini, S. Sobieraj Richard, and J. Rodriguez (2017) Stability of saturated timetables:the influence of buffer times. *In7th International Conference on Railway Operations Modelling andAnalysis - RailLille2017, Lille, France*, 243–259

Delorme, X., Rodriguez, J., & Gandibleux, X. (2001). Heuristics for railway infrastructure saturation. *Electronic Notes in Theoretical Computer Science, 50*(1), 39-53

Bundesbahn, D. (1979). Richtlinien für die Ermittlung der Leistungsfähigkeit von Fahrstraßenknoten, 405

Goverde, R. M., & Hansen, I. A. (2013, August). Performance indicators for railway timetables. In *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings* (pp. 301-306). IEEE

Hansen, L. A. (2000). Station capacity and stability of train operations. *WIT Transactions on the Built Environment, 50*

Hansen, I. A., & Pachl, J. (2014). Railway timetabling & operations. *Eurailpress, Hamburg*

KFH Group. (2013). Transit capacity and quality of service manual.

Kontaxi, E., & Ricci, S. (2009). Techniques and methodologies for carrying capacity evaluation: comparative analysis and integration perspectives. *Ingegneria Ferroviaria*, *64*(12), 1051-1080

Landex, A. (2011). Capacity at railway stations. In *9th World Congress on Railway Research*

Landex, A., & Jensen, L. W. (2013). Measures for track complexity and robustness of operation at stations. *Journal of Rail Transport Planning & Management*, *3*(1-2), 22-35

Liu, S. Q., & Kozan, E. (2009). Scheduling trains as a blocking parallel-machine job shop scheduling problem. *Computers & Operations Research*, *36*(10), 2840-2852

Malavasi, G., Molková, T., Ricci, S., & Rotoli, F. (2014). A synthetic approach to the evaluation of the carrying capacity of complex railway nodes. *Journal of Rail Transport Planning & Management*, *4*(1-2), 28-42

Nash, A., & Huerlimann, D. (2004). Railroad simulation using OpenTrack. *WIT Transactions on The Built Environment, 74*

Pellegrini, P., Marlière, G., & Rodriguez, J. (2017). RECIFE-SAT: A MILP-based algorithm for the railway saturation problem. *Journal of Rail Transport Planning & Management*, *7*(1-2), 19-32

Potthoff, G. (1963). Verkerhrsstnomungslehre 1

Pouryousef, H., Lautala, P., & White, T. (2015). Railroad capacity tools and methodologies in the US and Europe. *Journal of Modern Transportation*, *23*(1), 30-42

RFI (2016). Prefazione Generale all'Orario di Servizio in uso sulla Infrastruttura ferroviaria nazionale

Rotoli, Francesco & Ricci, Stefano & Elena, Navajas & Malavasi, Gabriele. (2015). Capacity versus punctuality assessment procedures and accessibility measures for rail networks. *Ingegneria Ferroviaria, 70*

Sameni, M. K., Dingler, M., Preston, J. M., & Barkan, C. P. (2011). Profit-generating capacity for a freight railroad. In *Transportation Research Board 90th Annual Meeting*

Sogin, S., Dick, C. T., Lai, Y. C., & Barkan, C. P. (2013). Analyzing the incremental transition from single to double track railway lines. In *Proceedings of the International Association of Railway Operations Research (IAROR) 5th International Seminar on Railway Operations Modelling and Analysis, Copenhagen, Denmark,* 1-20

Szpigel, B. (1973). Optimal train scheduling on a single line railway

UIC (2008). Influence of ETCS on the line capacity, *International Union of Railways (UIC), ERTMS Plateform, Infrastructure Department*

UIC (2013). Leaflet 406 R: Capacity, *Paris*

Vicuna, G. (1993). Organizzazione e tecnica ferroviaria, *CIFI*