Integrating yard, network and optimisation models towards real-time optimisation of rail freight yard operations

Riccardo Licciardello, Norbert Adamko, Samuel Deleplanque, Pierre Hosteins, Ronghui Liu, Paola Pellegrini, Anders Peterson, Magnus Wahlborg, Miloš Zaťko
[1] E-mail: riccardo.licciardello@uniroma1.it, Phone: +39 06 44585 138

Abstract

This paper describes the state of advancement achieved in the OptiYard research project in the use of optimisation algorithms in interaction with microsimulation of the rail-yard and surrounding network towards real-time yard management and communication with the network. Two case studies, a hump marshalling yard (mainly Single Wagon Load traffic) and a flat shunting yard (mainly intermodal traffic), were represented with state-of-the art microsimulation models, combined with innovative optimisation algorithms. Some specialistic information on the nature of the models is provided. However, the focus is oriented to railway engineers, with a description of the interactions between the models in producing outputs that are useful both to the yard dispatcher (decisions on staff, track, locomotive assignment, order of operations) and the infrastructure manager of the surrounding network (expected times of departure, availability of tracks in the yard).

1 Introduction

Rail freight yards are key elements of the rail system. Their operations affect the overall effectiveness of the door-to-door transport chain of any goods that use the rail mode. In order to contrast the decline in Europe of the use of the rail mode for freight transport, and in particular that of Single Wagon Load transport - see Guglielminetti et al. [1] - it is thus extremely important to streamline operations as much as possible, decreasing if possible their duration but above all improving their regularity, thus enabling a strong improvement in the performance of the rail mode and attracting traffic that otherwise would go mainly by road (Islam et al [2]).

The European Commission and the rail industry have taken this issue seriously - their Joint Undertaking Shift2Rail have included terminals, hubs, marshalling yards and sidings as a Technology Demonstrator in their Multi-Annual Action Plan (Shift2Rail [3]). A specific call on real-time yard and network management was issued by Shift2Rail in 2017. Real-time management of railways has been a hot topic in recent years - see e.g. Pellegrini et al. [4] - however the literature regarding management of yards specifically, and their interaction with the network is in fact scarce.

Real-time management capabilities are key in improving regularity in the flow of trains between yards and network in daily operations that are often highly perturbed.

The call resulted in the funding of a dedicated project. The OptiYard project received funding from the Shift2Rail Joint Undertaking under the European Union's Horizon 2020

research and innovation programme under grant agreement No 777594. The authors also acknowledge the support provided by all the consortium partners, in particular Adriafer and CD Cargo for the data and information regarding the yards.

OptiYard addressed the optimisation of rail yards targeting the yard dispatcher as end-user. The project worked on two case studies, a hump marshalling yard mainly for Single Wagon Load transport (Czech case) and flat shunting yard mainly managing intermodal traffic (Italian case).

The idea behind OptiYard was to tackle the real-time management issue through optimisation algorithms interacting with microscopic models of the yard and of a 'relevant' portion of the surrounding network. The project's Work Package n. 4 'Modelling' had the main objectives of development and validation of the yard and network models capable of real-time functionality and interaction with the optimisation model developed in Work Package 5.

It is the findings regarding the interactions between the abovementioned models and with the IT systems connected to the yard and the railway network that are the focus of this paper. These findings contributed to the design of the Decision Support System (DSS) and are now being followed up in the implementation, which is still in its first steps. The final aim is to develop a highly innovative yard management DSS, capable of generating optimal disposition decisions and recommendations for resource utilisation in real time, in parallel with the on-going innovation in real-time network management. This will allow in turn to improve decision-making in ad-hoc timetable planning to optimise operational processes that connect freight traffic in yards and terminals with timetable slots to and from the network.

In section 2 a brief overview of the state of the art is provided. Descriptions of the case study yards and the models follow in section 3 to 7. Section 8 describes the interactions between the models required to produce the desired outputs: between the optimisation and yard micro-simulation models, between these and the network model, and with outside systems. Conclusions and outlook are given in section 9.

List of acronyms

DSS	Decision Support System
ETA	Expected Time of Arrival
ETCS	European Train Control System
GPS	Global Positioning System
IM	Infrastructure Manager
ISR	International Service Reliability
IT	Information Technology
IVG	Intelligent Video Gate
KPI	Key Performance Indicator
JSON	JavaScript Object Notation
ORFEUS	Open Rail Freight EDI User System
PIC	Piattaforma Integrata di Circolazione (integrated traffic platform)
RailML	Railway Markup Language
RNE	Rail Net Europe
RU	Railway Undertaking
SWL	Single Wagon Load
TIS	Train Information System
TAF	Telematic Applications for Freight
TSI	Technical Specification for Interoperability
UIC	Union Internationale des Chemins de Fer
XML	Extensible Markup Language

YM Yard Manager

2 State of the art

The literature on rail yard modelling widely addresses the determination of yard capacity and other Key Performance Indicators (KPIs) for different types of yards (marshalling yards, flat rail yards, intermodal terminals etc., see e.g. in the Italian literature Dalla Chiara et al. [5], Bruno [6], Lo Russo et al.[7], and elsewhere a interesting descriptions by Shi and Zhou [8] and Antognoli et al. [9] for several types of rail terminals).

Both discrete and continuous (i.e. over time, second by second) simulation models are used. The purpose of this modelling has generally been to support decision-making regarding changes in infrastructure design or operations management. In OptiYard the purpose is relatively novel: using simulation as a support for real-time optimisation of rail yards (fixed infrastructure conditions), and thus combining continuous micro-simulation with optimisation algorithms to provide not only KPIs after a certain time of operation, but real-time results supporting the yard dispatcher's decision-making process, all of this in interaction with outside systems representing the rail network and rolling stock fleets. In this field the literature becomes scarcer.

In fact, when it comes to optimisation suitable for real-time functionality, the existing literature almost exclusively focuses on the marshalling process (Bohlin et al. [10], Boysen et al. [11], Daganzo [12]). The marshalling process consists in affecting wagons to tracks to compose outbound trains. Indeed, in the typical yard layout, inbound trains arrive on receiving tracks, their wagons are detached and pushed to the hump to roll to classification tracks (roll-in), and the resulting outbound trains are pulled to departure tracks before joining the network.

The marshalling process concretises differently in different yards, depending on the infrastructure layout and on the occurring traffic. The goal is in general the minimization of train delays at departure. In single-stage classification, wagons are moved only once from receiving to classification tracks. Once they reach the latter, they start composing the outbound train to which they are aimed (Haahr and Lusby [13], Jaehn et al. [14], Yagar et al. [15]). An alternative to single-stage classification is multi-stage classification with mixing tracks. This classification is relevant when there are not enough classification tracks to start building all outbound trains as soon as their first wagon is received. Hence, wagons are stored in one or more specific mixing tracks waiting for the appropriate moment to start building the corresponding outbound trains. These wagons need to be pulled back to the receiving tracks at least once. This is a costly operation, which needs to be performed attentively (Bohlin et al. [16], Kraft [17]). Finally, the third alternative is the multi-stage classification with wagon ordering. Here, wagons can be immediately used for composing the outbound train they are aimed to, but they must be placed in a specific order. Hence, it may be necessary to pull them back to the receiving tracks once or several times (Shi and Zhou [8]).

The only study considering a wider model of yard operations, not limited to marshalling, is the one proposed by Gestrelius et al. [18]. The authors present a mathematical model for track allocation and roll-in timing. The authors consider a specific yard layout with mixing track, and no departure tracks. Arrivals, roll-ins, pullbacks, mixing operations and departures are all considered.

Although the last model greatly broadens the scope of the existing literature, no existing contribution considers resource allocation (e.g. locomotives and staff to a given activity) and movements that must be executed to manage a yard. Moreover, most importantly for real-time yard optimisation, the interactions with IT systems external to the yard are not generally addressed. Also, the interaction between innovative optimisation and the state-of-the-art yard and network models described in the subsequent sections of the paper is not addressed.

3 Case studies

Two case studies - Česká Třebová in the Czech Republic and Trieste in Italy (Figure 1) - have been addressed in this study and models have been developed for the respective yards and surrounding networks. The former is a hump marshalling yard in which the vast majority of incoming trains are for SWL transport and require single wagons or groups of wagons to be sorted. The latter is a flat shunting yard serving a port area.

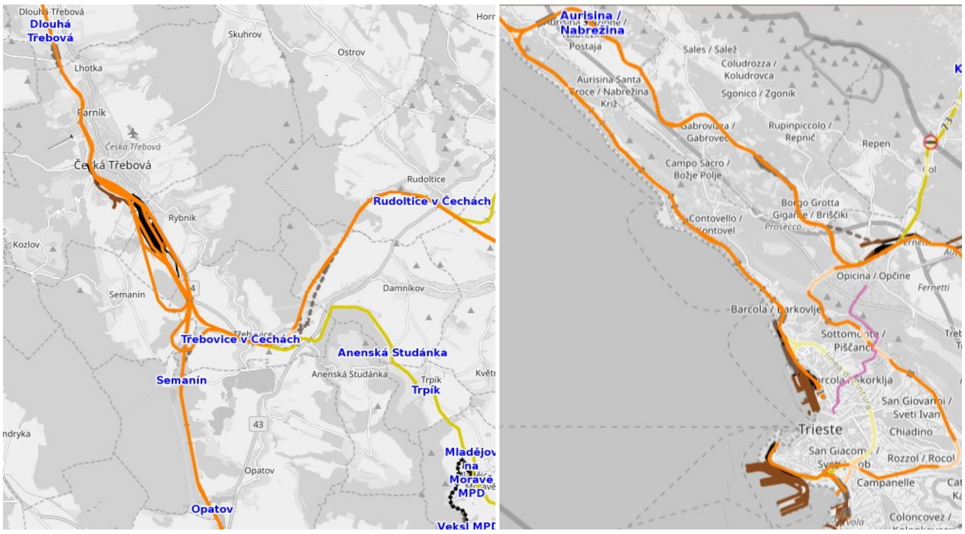


Figure 1: Rail networks surrounding the case-study yards

Situated in the centre of the Czech Republic rail network and with a marshalling capacity of up to 1,200 wagons in 24 hours, the marshalling yard Česká Třebová ranks among the seven key railway facilities utilized by ČD Cargo, the largest Czech operator. The current average marshalling throughput is around 700 wagons in 24 hours (2016), some 18% of which are intermodal wagons. The station is located on the busiest main Czech railway corridor East-West and on the main line between Prague and Slovakia with a section to Brno and Vienna. Česká Třebová has the most convenient railway marshalling topology, i.e. the entry yard, hump, classification yard and departure yard. The station currently operates about 30 domestic and international destinations, including Engelsdorf (Germany), Vienna (Austria), Bratislava/Žilina (Slovakia) and Wrocław (Poland).
Česká Třebová is designed with 38 sorting tracks (with a length of up to 1,000 meters), divided into 5 clusters, each equipped with 3 series of pneumatic retarders (up to final fine-braking) so that no shunters are necessary in the classification yard except for the couplers. The arrival yard is made up of 13 tracks with lengths of up to 900 meters and the departure yard consists of 15 tracks with a maximum length of 850 meters.
Typical optimisation decisions that are made by the dispatcher regard, in normal operation, the order of humping so as to minimise deviations on the schedule of departing trains or consideration of priority freight, and in disrupted situations the tactics needed to obtain maximum throughput.

The port of Trieste is served by the Campo Marzio rail yard. It is a flat shunting yard that handles around 8,000 trains per year (2018) of which around 25% are conventional SWL (essentially steel products coming from the Servola works, where a dedicated yard was reopened in 2016). The majority of trains are intermodal (mostly semi-trailers and containers exchanged with the vessels using the port facilities), travelling along the Baltic-Adriatic corridor with frequent connections to Austria through Tarvisio with 80 trains per

week, and a few connections with Slovenia through Villa Opicina. The operator Adriafer, subsidiary of Port of Trieste authority, performs the last mile operations in the region of Trieste (Campo Marzio – Villa Opicina, Monfalcone, Aquilinia).
The three port terminals see the arrival of vessels delivering the intermodal units to be loaded onto the wagon consists. Due to the limited length of the tracks in the terminals, the wagon consists corresponding to a whole train (550 m maximum length) often have to be split in half in order to be accommodated, and the two halves have to be joined together again with the shunting locos to form a departing train on one of the 6 tracks of 750 m length, after an eventual wait on dedicated sidings. 2 locomotives are available for the shunting operations and 2 Zephyr rail-road engines have been recently purchased to increase the efficiency of the operations and to enable a better use of existing tracks. For a flat yard, this is a relatively complex case in term of layout, variety of traffic, tight space conditions and rapidly growing activity and thus offers interesting optimisation possibilities. Typical optimisation decisions are the assignment of staff to locos, the assignment of locos and rail-road engines to consist and trains, the order in which consists and trains should be shunted considering the variations of train and vessel arrival times, as well as management of resource shortages (loco/track/staff unavailability) and disruptions on the surrounding network. The latter decisions must consider that different competing companies operate the port terminals. This is the problem that the operator needs to solve in daily operation, nowadays entirely based on experience and intuition.

4 Overview of the models

The simulation and optimisation models developed in OptiYard are designed to interact in a view to produce optimisation proposals for the dispatcher, as the basis for a Decision Support System (DSS). The DSS will essentially use an optimisation algorithm designed within the project. Periodically, the algorithm looks several hours into the future and works on a simplified yard model to identify the best way for managing yard operations given the observed situation. This algorithm makes decisions and communicates them to the simulator that plays the role of a real yard. If a perturbation occurs, the simulator communicates to the algorithm the updated operational conditions, which become the for the forthcoming optimisations. A further micro-simulation model of the surrounding network and the interaction with the IT systems of infrastructure manager and operators are foreseen in order to consider the state of the surrounding network in perturbation detection and path availability. The proposed DSS can in principle be deployed as a completely automated operations management system. However, yard dispatchers can be directly involved in the decision-making process. For example, the dispatcher may receive a set of options that are pre-assessed in terms of technical feasibility and respect of pre-defined priorities such as those linked to the treatment of dangerous or priority goods. The dispatcher can be expected to make a final assessment particularly regarding non-technical issues such as competition and other priorities. Indeed, by exploiting the proposed DSS, the dispatcher's decisions are expected to be optimised and less dependent on the person's experience and psychological conditions than they may be today.

Figure 2 summarises the interactions of the models developed in OptiYard. The interacting models are three for each case study: two microscopic models of yard and surrounding network and one model implementing the optimisation algorithm. They are developed on the basis of data obtained from the yard managers and from public sources of network information. In OptiYard, the optimisation and microsimulation models were implemented as software. The interactions between the optimisation and yard models were pioneered as described in this paper. The interactions with the network model and with external systems were conceptualised, as also described briefly

Although the outputs of the future DSS are directed mainly to the yard manager, they are also expected to provide information to the network infrastructure manager, in terms of expected times for trains to be ready and train path requests, and obtain information from infrastructure managers and operators. Through these interactions the DSS can thus also contribute to the optimisation of operations on the wider network.

The OptiYard DSS is intended to complement and receive input from the eventually existing (and nowadays increasingly developed) IT systems that the yard manager may use to monitor the state of the yard via e.g. a sensor network, portals, or manual data input. The OptiYard DSS is also conceived to draw input from the IT systems representing railway network status for timing data and train composition data. The target for this issue is full compatibility with RailNetEurope's Train Information System (TIS), along with the legacy IT systems of the infrastructure manager, as well as with the Telematics Applications for Freight Technical Specification for Interoperability (TAF TSI). Moreover data from wagon tracking & tracing systems, such as ISR (International Service Reliability) developed by the RailData group within UIC may be used as input.

In the current design and implementation, the outputs are produced by the optimisation model, with the virtual yard model validating results through micro-simulation and contributing to the representation of results and calculation of Key Performance Indicators (KPI). For the purposes of optimisation, the network micro-simulation model delivers (at the moment conceptually) Expected Times of Arrival (ETAs) of trains at the yard's home signals with higher accuracy than available from the network's IT systems. For departing trains, it determines the foreseeably available train paths. The DSS is capable also of providing expected times at which a given train will be ready for departure and suggesting when path requests are necessary.

Figure 2: Overview of the interactions between models and with external IT systems

5 The virtual yard micro-model

For the project OptiYard the state-of-the-art simulation tool Villon (Simcon, [19]), provided by Simcon, was selected as the core of the simulation framework. This tool was utilised to create flexible simulation models of the selected yards in Česká Třebová and Trieste. Villon has further been adjusted and its functionality has been extended in order to support real-time data exchange with the developed optimisation module.

Villon is a generic simulation tool, which allows microscopic modelling of various types of transportation logistics terminals containing railway and road infrastructures (e.g. marshalling yards see Figure 3, railway passenger stations, factories, train care centres, depots, airports, etc.). The simulation tool Villon is based on the flexible agent-oriented simulation architecture ABAsim (detailed description of its properties can be found in Kavička et al. [20]), which enables trouble-free extension of its functionality as well as cooperation with other software modules (e.g. railway network module or optimisation module in OptiYard).

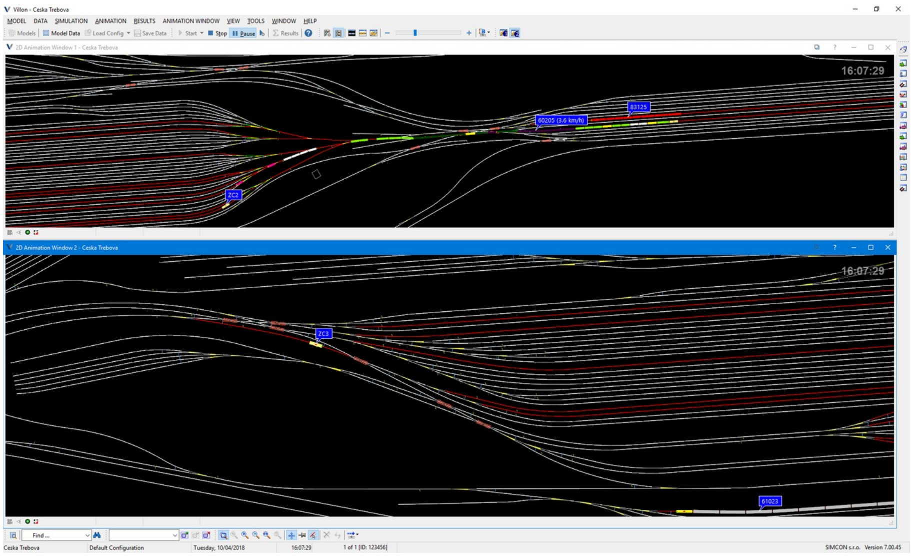


Figure 3: Animation of modelled processes during simulation of the Česká Třebová marshalling yard

In Villon, the modelled system (e.g. railway yard or container terminal) is considered as a service (queueing) system, composed of three subsystems, with specific roles in the system: resource subsystem, client subsystem and control subsystem. For example, in a rail yard the "client" may be a train, served by resources (locomotives, staff, tracks) following established procedures (controls), e.g. couple/un-couple loco, brake test etc., the latter being translated into flowcharts representing all the possible activities addressing the specific "client" (see e.g. Figure 4 illustrating the regrouping of wagons to form a group train on a sorting track of hump marshalling yard).

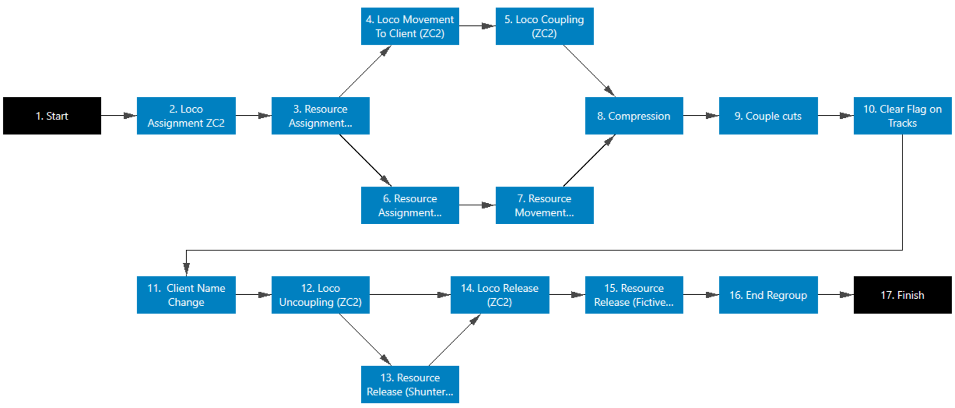


Figure 4: Example of a Villon flowchart (regrouping of wagons to form a group train in a hump marshalling yard)

Depending on the type of yard model, specific data for model building are required. However, both case-study simulation models share important data categories that need to be defined, reflecting the respective parts, processes and entities of the yard/terminal, namely:

- track layout/infrastructure model based on data in scale from CAD-drawings;
- working personnel (shunters, examiners, data collectors);
- mobile technical resources (shunting locomotives);
- yard processes (rules for sorting, un-/loading, parking);
- train handling/sorting/forming procedures in the form of flowcharts;
- input train flow (transiting and terminating trains).

The execution of a simulation run is presented using run-time animation and visualisation in order to support the validation and evaluation process of simulation runs. After the simulation run, an extensive set of post simulation evaluations is at hand – these are based on detailed simulation protocols recorded during the simulation run. Besides graphical presentation of simulation results using time dependent reports on utilisation of resources, waiting times, etc.; statistical evaluations are also provided – in the form of tables, graphs and charts (including resource utilisation statistics and many others).

For the OptiYard case studies, the creation and validation of both models depended on the provided data from real operation. Undocumented data, processes and decisions were either not modelled or reasonable assumptions were made. Despite this fact, based on the comparison of simulation results for a real week of operation and discussion with the yard managers, the created models were considered as validated.

The main goal of modelling in both cases was to reach the same departure time of trains from the system as in real operation. Some processes inside the modelled yards are realised slightly differently or were simplified (handling of damaged wagons or hump restrictions for wagons in marshalling yard).

The model validation results however showed differences from reality in some situations, e.g. the order of train sorting over the hump in the yard model of Česká Třebová. This is mainly due to the fact that the reasons for the dispatcher's decisions about the sorting order of trains are unknown. A side-effect of this is a different number of wagons leaving the marshalling yard in outbound trains between model and in reality – some wagons were sorted sooner, some later. This further stresses the fact that dispatching decisions are probably not always based on purely technical considerations for maximum efficiency.

In the Trieste model the real start and finish times of shunting to/from port were respected. After the transloading process on the quays, the second most time-consuming process is the shunting of wagons within the port (due low maximal speed).

In order to investigate non-standard operation of the yard, additional simulation scenarios were prepared. The model validated by comparison with data from a real week's operation was used to implement scenarios with shunting locomotive out of order, track out of order, delayed trains on arrival and additional trains. In this way the possibility of delivering several solutions in a short time (real-time functionality) was tested.

The yard micro-simulation models have two important roles within the project. Firstly, a replacement of the real yard was needed during the research phase to feed the optimisation module with a current state of yard and the latest changes in operation. The second purpose of the simulation model, which remains in the concept of the DSS, is a validation platform for the optimisation module to able to evaluate and compare consequences of different decisions provided.

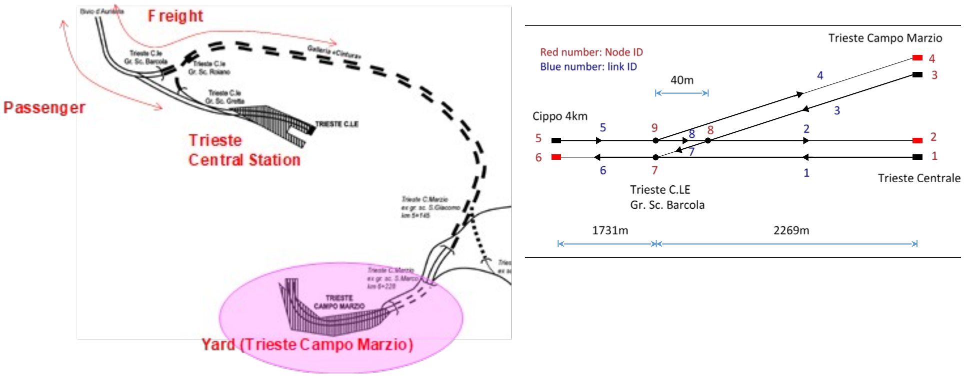
6 The virtual network micro-model

The network model developed for the OptiYard project is based on the state-of-the-art TrackULA (Track Unified simulation Algorithm) micro-simulation model developed by the University of Leeds (Liu et al.[21], [22]). The core functions of TrackULA include:

- simulation loop based on fixed time increments;
- railway network representation;
- railway timetable and train route representations;
- train and driver behaviour representations;
- train movement simulation;
- control command simulation;
- simulation outputs.

TrackULA is a microscopic simulation model which represents the movement of individual trains. It is based on discrete-time simulation where the train status is updated at a fixed time interval. It can model stochastic travel times (as opposed to deterministic, scheduled times) and disruption. It also allows heterogeneous train characteristics, train operating and train drivers' behaviour, as well as variations in drivers' experience and driving behaviour with a given probability distribution.

The network is described as a graph, with nodes representing the station and merging/diversion/crossing points, the links representing the tracks connecting nodes. This is a macroscopic representation of the infrastructure, typical of that for timetabling analysis. The track gradients are represented to model their impact on train running speeds. The graph representation of the surrounding network to the Trieste yard is shown in Figure 5.



(a) **(b)**

Figure 5: infrastructure layout of Trieste network model: (a) the surrounding network considered; and (b) its graph representation in the network model

TrackULA is a micro-simulation model because it simulates the movements of individual trains through a network, based on a 'train-following' model and the control-command of the signalling systems modelled. More specifically, it calculates the acceleration, speed and position of each train at every time interval, with given acceleration/deceleration profiles whose values are sourced from the literature and in consultation with the railway industry. Different signalling systems, ranging from fixed-block, to ETCS Level 2 and Level 3, are modelled in TrackULA.

The model outputs each individual train's second-by-second space-time trajectories as well as route/line-based and network-wide statistics. As a result of the stochastic modelling, the simulation outputs include not only the means but also the variances and probability distributions of performance measures.

For the virtual network model of Trieste, the network description includes the link locations on the network, link start and end locations (and link length), and gradient of the link, as well as four different speed limits on the link for four different train types: A represents freight trains, B regional trains, C Inter City trains, while P is used for tilting trains that are used for some high speed connections.

In OptiYard, simulation was conducted with the input timetables for a typical weekday passenger trains to/from Trieste Central (the passenger terminal station), and timetables for freight trains to/from Campo Marzio (the Yard). The possibility to achieve the real-time functionality for this type of software was assessed. Figure 6 shows the simulated trajectories of the outbound trains. The red lines are for freight trains, while blue ones the passenger trains.

It can be seen in Figure 6 that there are periods during the day when it was very congested on the network (with densely packed train trajectories), often involving both passenger trains and freight trains. Some of these congested periods are highlighted, and in one such example (shown in small box), there appears to be some delays to the passenger trains due to a slow freight train ahead.

Whilst during other times of the day, there are large time gaps in the train trajectories, such as between 10:00 – 12:00. The results illustrate that there is potential for improvement to ensure the best use of network capacity and to reduce conflicts between freight and passenger trains.

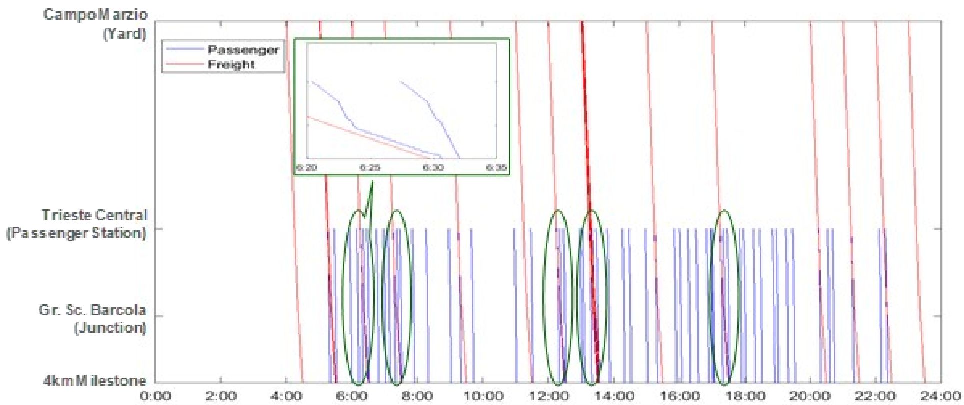


Figure 6: The simulated train trajectories with zoom of a typical conflict between passenger and freight train paths.

7 The optimisation module

The optimisation module consists of the implementation of an optimisation algorithm as well as the interface to communicate with the simulators (in OptiYard implemented only with the yard simulator).

The optimisation algorithm is the first appearing in the literature that is capable of dealing with all processes and movements taking place in a yard. The algorithm can deal with single wagon and block train operations without distinctions. Specifically, following the modelling in the yard simulation module, all yard activities are organised in flowcharts. Each node in these flowcharts corresponds to an operation, including, for example, the assignment of a resource to a specific operation, the movement of this resource up to the operation location, and the actual performance of the operation itself (refer to Figure 4 above). Specific flowcharts are associated to inbound and outbound trains. The optimisation algorithm proposed within the OptiYard project deals with each node in each of these flowcharts sequentially. Some of the operations, represented by nodes, require the allocation of a resource, being it a locomotive, an operator or a track. In this case, the algorithm chooses the specific resource to be allocated, if several alternative ones are available. Moreover, some operations require the execution of a movement, being it of a shunting locomotive, a wagon or a train. In this case, the algorithm chooses the specific route (sequence of tracks) along which the movement is performed.

Following the operations research and artificial intelligence literature, given the complexity of the problem at hand, a stochastic algorithm or randomized greedy algorithm is proposed. This type of algorithms explores the search space of the problem thanks to a random component. This is typically combined to a local search in order to assess the neighbourhood of visited solutions to improve their quality by shaping a greedy randomized adaptive search procedure (GRASP). Examples of this type of algorithms are the so-called meta-heuristics, which are often inspired to natural processes.

In practice, in the algorithm proposed, solutions are built incrementally. At each step, starting from the current state of the system, which corresponds to a node in the flowchart of operations of the present and expected trains, the algorithm randomly chooses the next operation to tackle considering all possible subsequent nodes. Then, if the operation requires a resource and then its route to execute a movement, this is randomly selected among the existing possibilities. For the schedule of these resources and movements, possible time windows are identified, setting it as soon as possible considering the constraints imposed by previously made assignments. As subsequent operations are included in the solution, additional constraints to previous assignments may apply, which imposes the tightening of the time windows. For example, consider the case of a previous assignment concerning the movement of a locomotive in the yard. Suppose the current operation requires the movement of a whole train, and this ends up needing to occupy the same track as the yard locomotive at the same time. In this case, the time window referred to the locomotive movement needs to be tightened to ensure that the train does not need to brake to give precedence. Indeed, stopping a train is much more expensive than stopping a locomotive, and hence the latter choice is always made. When tightening a time window, the algorithm checks for all the consequences on the already made assignments. For example, if the locomotive now needs to brake, it may arrive later than initially anticipated to push a given cut to the hump. The occupation of the hump by said cut will occur later as well and possibly delay further hump operations. In this sense, the algorithm propagates the constraints on operations scheduling and on resource availability.

The local search implemented consists in swapping the schedule of pairs of operations using the same resource. It is applied to each solution generated. To speed up the algorithm for real-time application, the exploration of a neighbourhood is stopped as soon as an improving solution is found. This is typically referred to as first improvement local search.

As mentioned at the beginning of this section, the optimisation module includes both the algorithm and the communication interface. Indeed, while a new state of the yard is communicated and a new optimised solution is being computed, the module ensures the continuity of operations by still communicating to the simulator the previously made decisions if their execution is imminent.

To maintain consistency between previously made decisions and the new decisions, the algorithm considers a short-term prediction starting from the state of the yard communicated by the simulator. When it receives the state of the yard, it computes what the state will be in a few minutes if the previously made decisions are applied. It then optimises starting at this future state, so as not to modify decisions which may have to be implemented while searching for the new solution.

8 Interactions to produce the required outputs

The interactions that were physically developed in OptiYard regard those between the optimisation module and the yard micro-model. This reflects the project's strong focus on providing a DSS for the yard manager. The interactions with the network micro-model and with the outside systems were developed conceptually, in a view to create a vision towards which future research and development activities could be devoted.

8.1 Interactions between the optimisation module and the yard micro-model

As explained in Section 7, the optimisation considers the state of the yard communicated by the simulator, and the simulation implements optimisation decisions as time elapses.

For performing this closed-loop interaction, a communication channel needs to be opened.

On the one hand, the optimisation gets information on the yard management through files. Specifically, a file containing static data (concerning, e.g., the infrastructure layout) is sent by the simulator to the optimisation module at the beginning of the process. Then, periodically, a file containing dynamic data is sent by the simulator to describe the state of the yard. Here, the operations currently being performed are indicated with their starting time, as well as novel information previously unavailable (e.g., a change in the expected arrival time of a train). These files are formatted in XML to be consistent with the RailML principles. In the future some dynamic data will come from outside systems via the network model as described in §8.2 and §8.3.

On the other hand, communication from the optimisation module to the simulator is continuous, based on client-server architecture, utilising socket communication. There are two client computers, one querying and one answering, both communicating only with the server computer (Figure 7) in a closed loop. The communication server provides connectivity, security and data validation (the server checks all pass-through communication for errors and completeness). All forwarded messages are in standard JSON format. Thanks to this system, as soon as a decision is needed on an imminent operation to be simulated, it is requested to the optimisation module that provides it immediately. For example, if it is the moment to perform a push operation on a cut, then the simulator will request which locomotive to use, and, supposing more than one is available, the optimisation module will indicate the most appropriate one given the current state in the yard and the overall plan.

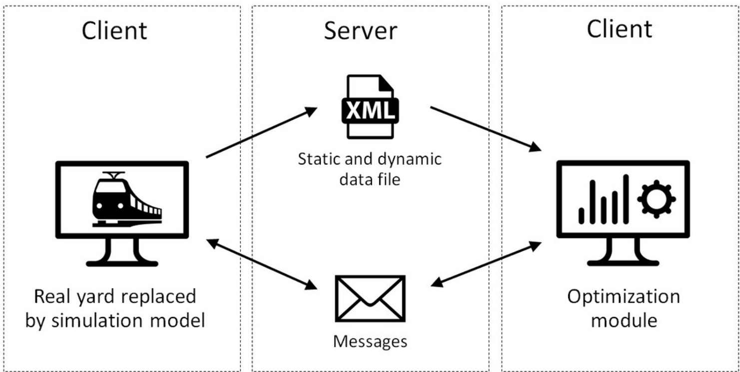


Figure 7: Closed-loop between optimisation and simulation

8.2 Interactions between the optimisation module / microsimulation model and the network micro-model

The diagram in Figure 8 shows the optimisation module / yard micro-model scopes on the one hand and the network model scope on the other, for the Trieste case-study.

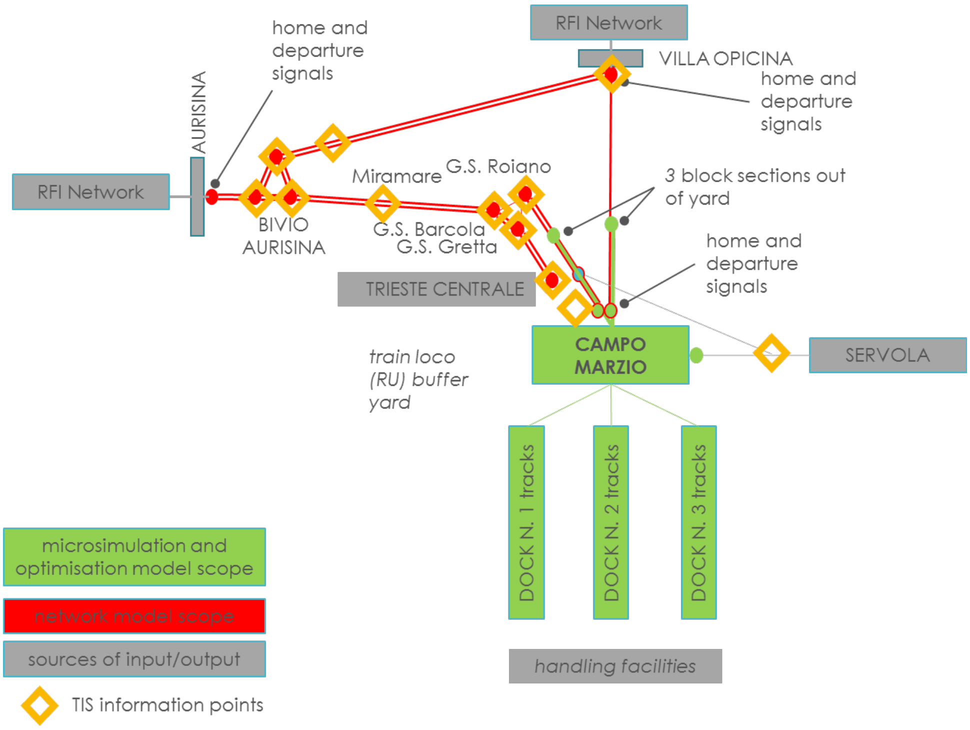


Figure 8: Scopes of the yard, network and optimisation models

The boundaries of the network model are limited to the closest "buffer sidings" to the optimised yard (specifically for Trieste Campo Marzio: Aurisina and Villa Opicina). These are the tracks (not included in the model) to which ideally any amount of trains per hour may be sent without any significant congestion occurring, and vice versa that can feed the yard with any amount of trains.

The network model scope overlaps with that of the optimisation/microsimulation part in that it arrives up to the home and departure signals of the yard. The optimisation/microsimulation scope extends to 3 block sections outward, in order to allow the trains to enter and exit smoothly in and out of the yard.

The model of the network surrounding Trieste and described in §6 simulates the train trajectories within the network scope defined above. The key issue in the network is the interaction between freight trains to and from Campo Marzio via Barcola (more often) and Villa Opicina (more rarely) with the passenger trains in and out of Trieste Centrale. The simulations show that the network is crowded but not close to saturation, and also that there should be no problem in performing the simulations in real-time.

The simulation tool as developed in OptiYard does not yet interact with the optimisation / microsimulation part but offers support in understanding what information needs to be passed to the optimisation module, as a basis for the development of the concept for the network software (i.e. the future product).

As inner boundary, it was established that the network model should comprise the tracks up to the entrance/exit of the yard and provide:

- ETAs,
- availability of train paths,

to the yard software based on real-time and static network information, and train-trajectory simulation. The availability of train paths is a key input for the Villon software used for the yard microsimulation.

The outer boundary of the network model is chosen, as mentioned, considering the closest yards with sufficient capacity to be considered as buffers, and also on the basis of the availability of reliable real-time train data. In Figure 8 the Train Information System (TIS) information points in the network are shown. These are sufficient for the network model to be able to draw up-to-date information on what is happening in the vicinity of Campo Marzio, and with that to make improved predictions, using the data available from outside the network within scope, of ETAs and available paths.

The grey blocks in Figure 8 represent the boundary elements for the optimisation-microsimulation yard+network model. The key information coming from the network blocks is the ETA of trains at the boundaries (Aurisina, Villa Opicina and Trieste Centrale). Servola provides few trains per day that are handled within the yard model. Key port-side information is expected times for arrivals of vessels and unavailability of resources to load the trains in the port terminals.

8.3 Interactions with the outside systems

Interactions with data analytics systems

The above three-way optimisation+yard-micro+network-micro interaction can benefit heavily, as already hinted, by interaction with systems outside the OptiYard scope.

The key interactions are with the Infrastructure Manager (IM) of the network surrounding the case study yard, and the Railway Undertakings (RU) operating in the yard.

The diagram in Figure 9 shows how the information from outside the OptiYard environment may conceptually be used. The inputs are both:

- static, timetable and network characteristics, that are generally available from the IM, in the example RFI for the Trieste case-study;
- and real-time, for which the legacy systems of the IM generally provide some information.

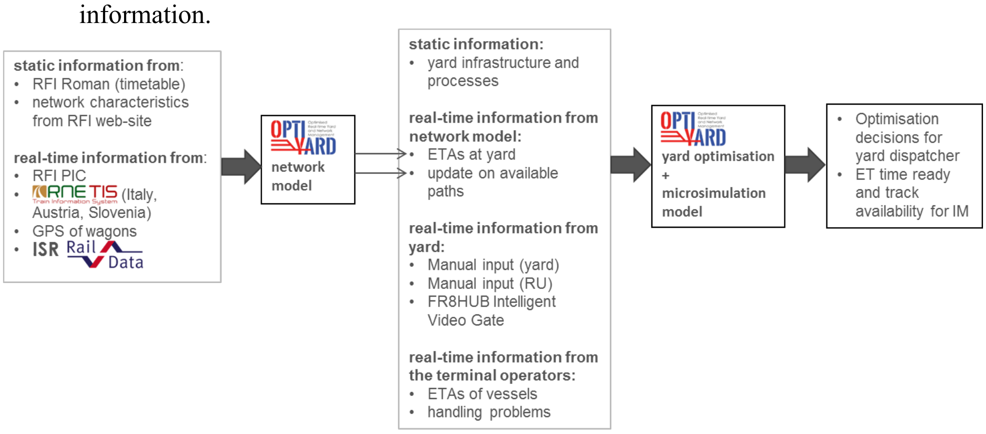


Figure 9: Use of outside inputs for the OptiYard environment

In particular for real-time information, possibilities are growing due to cooperation of different stakeholders. Both case study yards are surrounded by RNE TIS (tis@rne.eu) information points, which provide data e.g. regarding international train number, origin, planned and real arrival times, expected arrival tracks, customer RU. Already with this information, it is possible to conceive a network model identifying network problems in the vicinity of the yard (e.g. malfunctioning track circuit near Miramare in Figure 8) in order to make a more realistic ETA forecast than that made by TIS itself (or the IM's PIC). At the moment TIS has to be integrated with the legacy system (PIC for Trieste) since TIS's scope is only international trains.

Important for the Trieste case study, but much more so for Česká Třebová which is a hump marshalling yard, are the real-time information possibilities for single wagons. International collaboration within the International Union of Railways (UIC) has given rise to systems such as International Service Reliability (ISR), which uses train (TIS), consignment note (ORFEUS) and other information together in a platform bringing together the RUs for wagon information. Just as an example, the 'ReadyToPull' message that the RU inserts in the system on arrival may replace manual input directly into the OptiYard environment (see Figure 9). The message may be integrated with outside information coming from the increasing number of GPS equipped wagons. Moreover, the yard itself may be equipped with systems such as the IVG Intelligent Video Gate developed in the Shift2Rail FR8HUB project or other wagon identification portals, to integrate the information needs represented in Figure 9.

Interactions with network management

The diagram of Figure 10 indicates the way the interaction between the real-time yard and network systems has been imagined.

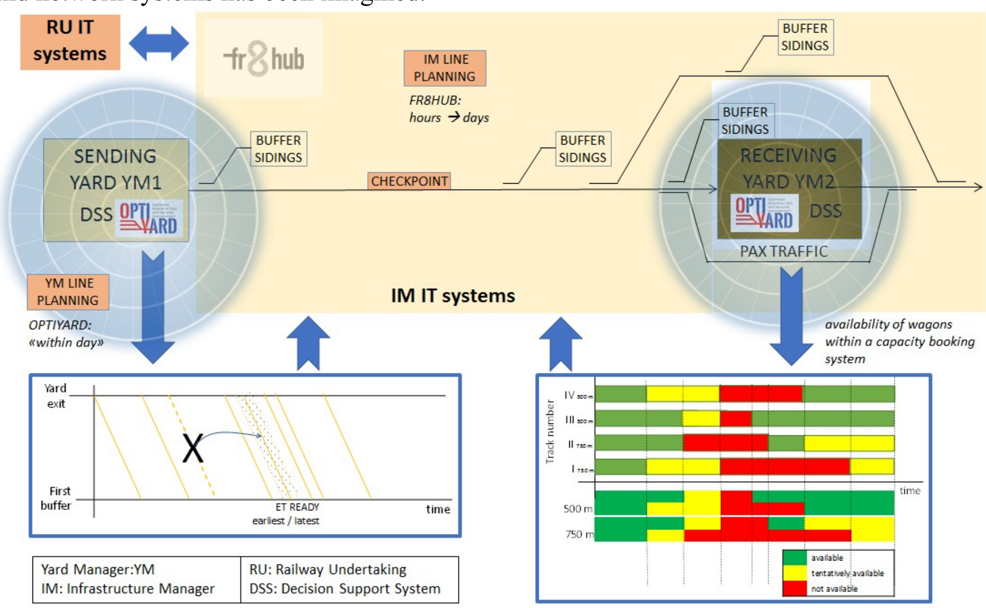


Figure 10: Inputs and outputs exchanged between the OptiYard Decision Support System and the outside information systems, with a focus on the IM

The diagram is focussed on the interaction with the IM for which a specific case study has been initiated. Two rail yards (of any type: marshalling yards, yards for intermodal traffic etc.) are connected by means of a railway line. The yard managed by Yard Manager YM1 is considered as the sending yard, and YM2 manages the receiving yard. Both yards have a real-time OptiYard Decision Support System. Of course, these two roles can be exchanged - this representation aims at showing the difference in terms of output of the yards to the Infrastructure Manager's IT systems. The latter are considered to comprise real-time network management capabilities based on the Single Train Insertion algorithm of Ljunggren [23] which was used as a reference for OptiYard. The real-world case study that was taken as an example is the Hallsberg – Malmö connection in Sweden. Of course, this does not preclude similar analyses to be applicable to lines starting in Česká Třebová or Trieste which are the OptiYard case studies.

Network and yards operate as closely as possible to predefined planning. However, with rail freight operations it is difficult to maintain punctuality to within minutes, even on days when there are no particular disruptions. One such day is considered. Regarding the time-horizon, it was noticed during the development of the case study that the focus of the OptiYard DSS is essentially "within day" (hours only) – which is the main interest of the yard manager - whereas the IM is more interested in the range from hours to days. The latter time is that required e.g. to complete an ad-hoc path request.

An example of how actual operations might develop in such a context could be as follows.

At a given point in time, the sending yard's DSS calculates, a few hours before scheduled departure, that a certain departing train is expected to suffer a significant delay, (Figure 10, bottom left), for example due to sudden unexpected lack of yard personnel of which the DSS was informed a few minutes earlier via manual input. The IM systems are automatically informed. This should be done via the SHIFT2RAIL integration layer whose requirements are set out in IN2RAIL, XRAIL2 and IMPACT 2 project deliverables, in RailML format (the need for integrations to such elements is to be assessed, e.g. in future SHIFT2RAIL projects).The delay is above a given threshold for the IM (to be determined, e.g. 15 minutes) which triggers the need for real-time rescheduling on the network. The IM systems thus immediately require Yard 1 dispatcher (human) confirmation of the train delay. Upon receipt of confirmation, the IM systems notify Yard 2 so that their DSS can plan accordingly and eventually use up the capacity that is now free. The receiving yard's DSS recalculates a prediction of track availability time windows considering the major train delay now confirmed. This sets a requirement for the time horizon of the DSS based on the travel time between sending and receiving yard. For the Hallsberg - Malmö example, the yard DSS should be looking at least about 12 hours into the future – 4 hours for advance notice of expected delay to the IM + 8 hours' travel time. The availability is provided in the form "track available, tentatively available, not available" (Figure 10, bottom right), based on given criteria (to be developed, e.g. considering number of available tracks for the required train characteristics, degree of yard saturation…). The Single Train Insertion algorithm receives this input, possibly in the form of a time-window, consistently with the algorithm's approach and uses the updated time-windows for yard 2, along with those coming from the intermediate service points along the track including buffer yards, and makes a first couple of path calculations (e.g. ETA,Robust, high-robustness, long-travel-time path, and ETA,Fast, low robustness, quick path) for the delayed train. The IM dispatcher selects a preference (e.g. Fast, in order to free up network capacity as soon as possible). A request is issued to the dispatcher of Yard 2, who examines actual availability as proposed by the OptiYard DSS (particularly if the result is "tentative") and accepts or refuses. In case of acceptance the train path is confirmed by the IM. In case of refusal by the receiving yard, the next-fastest path is proposed, and so on until acceptance by the yard.

For ad-hoc time-timetabling, where a path request could be made a few days before the actual train is to be run e.g. upon request of a customer, although the OptiYard time-horizon is within day, the annual timetable is present as an input and the yard manager's planned track occupations available in advance. Therefore, the algorithm for single-train insertion could search for an available path a few days in advance, thus supporting ad-hoc timetabling through its interaction with the yard's DSS. As the interaction evolves towards full real-time capabilities, this search could be updated if actual operations do not actually follow the timetable.

In the interaction with network management, key elements of rail operations complexity are considered in the representation of Figure 10, as listed below.

- Presence of mixed passenger/freight. A passenger line contained within the scope of the OptiYard DSS is represented for Yard 2, as well as a "through-line" for freight traffic. The OptiYard DSS is conceptually capable of re-estimating the ETAs to the yard by simulating the train trajectories in the vicinity of the yard considering the real-time information coming from IM systems (on corridors the Train Information System TIS of RailNetEurope, otherwise legacy systems, integrated with information on on-going disruptions via the asset management information through the SHIFT2RAIL integration layer). How to integrate this with the Single Train Insertion algorithm is an open issue to be solved once the network model of OptiYard is further defined.
- Use of network capacity to make up for the lack of yard capacity. This is represented through the "buffer sidings".
- Presence of "checkpoints" along the line where time-window constraints are possible (e.g. driver change, loco change, dropping wagons etc.).
- Interactions with the railway undertakings, of which several could be operating in the yards. For the moment the above is developed considering that the RU(s) agree with all decisions by IM and YM.

9 Conclusions and outlook

This paper describes the state of advancement reached in the OptiYard research project in using optimisation algorithms in interaction with microsimulation of the rail yard and surrounding network towards real-time yard management and communication with the network. Algorithms and microsimulation models have been developed that are capable of functioning in real-time, and physical interaction between optimisation and yard microsimulation has been established which forms the basis for a demonstration of the OptiYard results.

Yard optimisation is based on the interaction between optimisation algorithms, that are capable of making choices between alternatives in order to optimise specific KPIs (dwell time of wagons, deviations from scheduled departure times), and microsimulation, capable of validating the choice as a virtual yard within which to test the alternatives; in the first stages of implementation for a given yard, microsimulation is heavily used, and as the optimisation algorithms are improved on the basis of interaction during real operation, the microsimulation is gradually reduced, until the next significant change occurs to the infrastructure or operations.

The network model is concerned about the immediate vicinity of the yard in order to improve predictions (ETAs) to support yard management, and deals with network and RU communication.

The legacy information systems and the systems under development such as TIS and ISR are increasingly used in yard management to avoid manual inputs and duplications.

Contractual agreements between the stakeholders are developed to support all the above.

The OptiYard DSS for a case such as Česká Třebová (hump marshalling yard) could support decisions in dispatching trains towards a surrounding network that is quite congested, with the objective of minimising dwell times of wagons and delays on departure. In future developments of the optimisation algorithms, it could be used to exploit the current under-capacity situation to improve the "quality of trains" by supporting the reliable creation of trains with wagons in specific order thus enabling dynamic Railway Undertakings to develop more flexible production methods (group trains etc.).

For a case like Trieste, the OptiYard DSS can support difficult decisions on the order in which to shunt given trains/parts of trains so as to keep the throughput of the yard at the current levels close to saturation.

The optimisation+yard-simulation environment allows the inclusion in the picture of priority and dangerous goods on a wagon basis.

For both cases, decision-making in the case of shunting locomotive breakdowns, staff unavailability, tracks out of order, additional trains could be improved, along with the daily management of deviations with respect to schedule.

In the framework of corresponding agreements between yard managers, network managers and operators, the automatic support of yard management could become a key part of balancing the optimisation needs of all stakeholders towards an optimised rail freight system, thus bringing benefits to all, including society which would ultimately reap the social and environmental benefits of fewer goods on the roads.

References

[1] Guglielminetti, P., Piccioni, C., Fusco, G., Licciardello, R. and Musso, A. (2017). Rail Freight Network in Europe: Opportunities Provided by Re-launching the Single Wagonload System. Transportation Research Procedia, 25, pp. 5185-5204.

[2] Islam, D.M.Z., Ricci, S. and Nelldal, B.L. (2016). How to make modal shift from road to rail possible in the European transport market, as aspired to in the EU Transport White Paper 2011. European transport research review, 8(3), p.18.

[3] Shift2Rail Joint Undertaking, (2015). Multi-Annual Action Plan, Brussels

[4] Pellegrini, P., Marlière, G. and Rodriguez, J. (2016). A detailed analysis of the actual impact of real-time railway traffic management optimisation. Journal of Rail Transport Planning & Management, 6(1), p.13-31

[5] Dalla Chiara, B., Manti, E. and Marino, M. (2013). Intermodal terminals with gateway function: simulation of their engineering on a case study. Ingegneria Ferroviaria, 6, p. 587.

[6] Bruno F. (2013). Definition and application of a mathematical model for the simulation of railway intermodal terminals. Ingegneria Ferroviaria, 11, p. 965.

[7] Lorusso, R., Peresso, A and Malavasi, G. (2009). Automazione e applicazioni delle procedure di analisi della capacità di circolazione degli impianti ferroviari (Automation and applications of capacity analysis procedures of rail yards). Ingegneria Ferroviaria, 6, p. 547.
[8] Shi, T. and Zhou, X. (2015). A mixed integer programming model for optimizing multi-level operations process in railroad yards. Transportation Research Part B: Methodological, 80:19 39.
[9] Antognoli, M., Capodilupo, L., Marinacci, C., Ricci, S., Rizzetto, L., & Tombesi, E. (2018). Present and Future Operation of Rail Freight Terminals. In Transport Systems and Delivery of Cargo on East–West Routes (pp. 233-273). Springer, Cham.
[10] Bohlin, M., Hansmann, R., and Zimmermann, U. T. (2018). Optimisation of railway freight shunting. In Handbook of Optimisation in the Railway Industry, p. 181-212. Springer
[11] Boysen, N., Fliedner, M., Jaehn, F., and Pesch, E. (2012). Shunting yard operations: Theoretical aspects and applications. European Journal of Operational Research, 220(1):1 14.
[12] Daganzo, C. F. (1986). Static blocking at railyards: Sorting implications and track requirements. Transportation Science, 20(3):189 199.
[13] Haahr, J. T. and Lusby, R. M. (2016). A matheuristic approach to integrate humping and pullout sequencing operations at railroad hump yards. Networks, 67(2):126 138.
[14] Jaehn, F., Rieder, J., and Wiehl, A. (2015). Single-stage shunting minimizing weighted departure times. Omega, 52:133 141.
[15] Yagar, S., Saccomanno, F., and Shi, Q. (1983). An efficient sequencing model for humping in a rail yard. Transportation Research Part A: General, 17(4):251 262.
[16] Bohlin, M., Gestrelius, S., Dahms, F., Mihalæk, M., and Flier, H. (2015). Optimisation methods for multistage freight train formation. Transportation Science, 50(3):823 840.
[17] Kraft, E. (2000). A hump sequencing algorithm for real time management of train connection reliability. In Journal of the Transportation Research Forum, volume 39.
[18] Gestrelius, S., Aronsson, M., Joborn, M., and Bohlin, M. (2017). Toward a comprehensive model for track allocation and roll-time scheduling at marshalling yards. Journal of Rail Transport Planning and Management.
[19] Simcon. 2019. Villon simulation tool. Simcon. Accessible on internet: < http://www.simcon.sk/en/ tools/villon>
[20] Kavička, A., Klima, V., Adamko, N., (2007). Simulations of Transportation Logistic Systems utilizing Agent- Based Architecture. International Journal of Simulation Modelling 6 (1): 13-24
[21] Liu, R., Ye, H. and Whiteing, T. (2015) Dynamic Simulation for Real Time Operations of ERTMS Level 3. DITTO Project Deliverable 3.1, RSSB. (http://www.its.leeds.ac.uk/fileadmin/documents/research/F2-DITTODeliverable31Milestone6.pdf)
[22] Liu, R., Ye, H. and Whiteing, T. (2016) Simulation and Control of ERTMS Level 2. DITTO Project Deliverable 3.2, RSSB. (http://www.its.leeds.ac.uk/fileadmin/documents/research/F3-DITTOMilestone7Deliverable_32Final.pdf)

[23] Ljunggren, F., Persson, K., Peterson, A. and Schmidt, C. (2018). Maximum Robust Train Path for an Additional Train Inserted in an Existing Railway Timetable, Conference on Advanced Systems in Public Transport and TransitData (CASPT 2018), 23-25 July 2018 Brisbane, Australia.