Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

SIMULATION AND OPTIMIZATION OF TRACTION UNIT CIRCULATIONS

Matthias Rößler
Matthias Wastian

dwh GmbH
Simulation Services & Technical Solutions
Neustiftgasse 57-59
Vienna, 1070, AUSTRIA

Anna Jellen
Sarah Frisch

Institute of Mathematics
Alpen-Adria-Universität, Klagenfurt
Universitätsstraße 65-67
Klagenfurt, 9020, AUSTRIA

Dominic Weinberger
Philipp Hungerländer

hex GmbH
Lakeside B10a
Klagenfurt, 9020, AUSTRIA

Martin Bicher
Niki Popper

Institute of Information Systems Engineering
TU Wien
Favoritenstrasse 9-11
Vienna, 1040, AUSTRIA

ABSTRACT

The planning of traction unit circulations in a railway network is a very time-consuming task. In order to support the planning personnel, the paper proposes a combination of optimization, simulation and machine learning. This ensemble creates mathematically nearly optimal circulations that are also feasible in real operating procedures. An agent-based simulation model is developed that tests the circulation for its robustness against delays. The delays introduced into the system are based on predictions from a machine learning model built upon historical operational data. The paper first presents the used data and the delay prediction. Afterwards, the modeling and simulation part and the optimization are presented. At last, the interaction of simulation and optimization are described and promising results of a test case are shown.

1 INTRODUCTION

The railway traffic in Austria is primarily executed by the Austrian Federal Railways (ÖBB), which provides passenger transport as well as freight services. The planning of the schedules and the traction unit circulations is done in a combined manner. This means that a traction unit can be used for passenger as well as freight traffic. The joint planning leads to a lot of trains with different requirements and conditions that have to be scheduled. For example, passenger services have almost fixed operation times affected by the traffic of previous years and of neighboring countries while the operation of freight services is more flexible but may only be scheduled at off-peak hours and is set to a lower priority than passenger traffic.

The circulation planning of traction units is one of the last steps in the preliminary planning phase. This step is based on a clustering of traction units and a mapping of each ordered service to a traction unit cluster. After the schedules for the ordered trains per week are set, each service is associated with a certain cluster of traction units. These clusters are identified by the type of the traction unit and a shorthand for the location. Table 1 shows some clusters and their number of weekly services. The table includes the clusters with the most, median and least services. It shows that even an average cluster has to deal with about 190 different services per week while big clusters have up to about 4500 assigned services. This leads to a lot of potential conflicts which require careful planning.

Table 1: Clustering of traction units, based on type and location, and number of associated services within a week for the timetable period 2017/18. Highest numbers of services on the left, the region of median number of services in the middle, and smallest numbers of services on the right.

type	location	# services	type	location	# services	type	location	# services
4020	F	4614	1047	MAV	192	2043	TS	6
1144	TR	3360	4024	F	187	247	DB2	4
1116	TR	3086	1216	RLS	182	6189	LTE	4

For the circulation, each cluster is scheduled into so-called circulation days adding empty runs for the locomotives where needed. The main condition for the circulation is that at the end of the week, equal numbers of traction units within a cluster have to finish their services at the stations where they have started at the beginning of the week. This ensures the reusability of the planned week for subsequent weeks. An optimal circulation has as few circulation days as possible while using as few empty run kilometers as possible. But in order to be feasible, the planned circulation has to be robust against delays as well. In the case of manual planning, typical delay reasons may be taken into account which could be called ”expert knowledge”. Acknowledging delays in an automated planning process might lead to not optimal results given their random nature.

To study the effects of different delays, a Monte Carlo simulation using a model that validly simulates a given circulation seems a natural choice. Therefore, this paper proposes a combination of optimization and simulation runs to perform an automated intelligent planning task whose results can be used as part of a decision support tool for the planner (see Figure 1). While the optimization provides an (nearly) optimal circulation with respect to the previously mentioned deterministic objective function, consisting of a number of traction units and an amount of empty run kilometers, the simulation estimates the impact of the circulation on the overall delay status. This feedback is, again, used by the optimization adapting certain constraints to increase robustness.

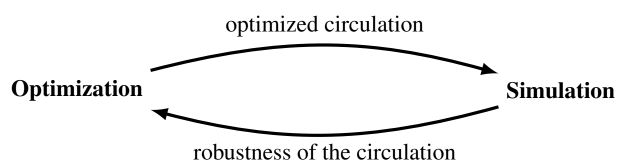

Figure 1: Interaction between simulation and optimization. While an optimization method improves the circulation based on a deterministic target function, the simulation gives feedback w.r. to its robustness.

1.1 Originality and Related Work

While neither the concept of using agent-based models for evaluation of train schedules nor using mixed linear integer optimization for optimization is a novel idea, its combination in form of a feedback-loop is. It combines the exploratory features of classic agent-based railway simulations with features of high-performance optimization algorithms leading to highly optimized **and** robust train schedules.

Usage of agent-based models in transport logistics, in particular railway transport, has become quite popular in the last decades as the observed system provides perfect characteristics to being abstracted by this modeling approach (see also Davidsson et al. (2005)). The success of this modeling technique is best displayed by a vast amount of specific simulation environments that are particularly designed for this purpose, e.g. Open Track (Nash and Huerlimann 2004), RailSys (Bendfeldt et al. 2000), or MATSIM (Horni et al. 2016). Unfortunately, neither of these simulation environments provides the opportunities to optimize the circulation of trains on a large railway network. In Waraich et al. (2009), authors perform performance comparisons of different queuing systems within MATSIM and reach computation times of about 30 minutes per simulation run using a benchmark network slightly larger than ours. Consequently, performing simulation-optimization attempts requiring millions of iterations is impossible. The latter, being not only a problem of the specific simulator, but a general problem of large-scale agent-based network models, makes agent-based models incapable of optimizing large train-circuits.

On the contrast, classical optimization algorithms such as Mixed-Integer Linear Programs are quite capable of finding optima for problems formulated in a large network-like structure due to highly optimized branch-and-bound methods (see Mitra (1973)) or problem-focused meta-heuristic approaches. Unfortunately, when applied on circulation optimization, these algorithms lack of a feasibility evaluation of the found optimum in the meaning of a robust train schedule. Consequently, the proposed simulation-optimization loop joins the forces of both approaches.

A first version of the agent-based simulation model along with a feasibility study was presented in Rößler et al. (2018). The basis for the linked optimization routine was presented in Frisch et al. (2019).

2 DATA

The data was provided in the form of database extracts. It included planning data as well as historical, operational data. In the following, an overview of the provided data and its connection to the entities of the simulation model and the optimization are given. As the provided data sources are used in live operations, they contain a lot of information that is not relevant for the purpose of the simulation model or the optimization approach; this information is omitted in the following. As the whole system can be imagined as a large network, the nodes of the network, the operational points, are explained first.

2.1 Operational Points

In European Union (2019), an operational point is defined as follows:

> *Operational point* (OP) means any location for train service operations, where train services may begin and end or change route and where passenger or freight services may be provided; it includes locations at boundaries between member states or infrastructure managers.

The most important form of an operational point is a train station. Trains typically start and end at train stations and a change of traction units is performed there as well. But also switches and signals are operational points. The relevant data belonging to an operational point includes an ID, the name, short names from different systems, and also geographic coordinates.

For the optimization as well as the simulation model, the track network, on which the trains operate, is needed. The operational points are the nodes of the graph that corresponds to this network. The availability of geographical data allows the embedding of the network into maps which enables clear visualizations.

In the examples given in this paper, a naming code used internally by the ÖBB is shown. Table 2 gives an overview of the operational points used throughout the paper, most of them are train stations.

2.2 Trains

A train is a (passenger or freight) service from one train station to another with a specific departure and arrival time. A train is identified by a train number and can be operated on an arbitrary number of days (e.g. only on work days). Additional information on the train includes an association to a traction unit cluster (see Section 1), a time frame of validity, and handling times at departure and arrival (*h.dep., h.arr.*).

Within the data set, a train can be associated with more than one entry. A train is split into several entries if it crosses a border or if the train is (partially) operated by more than one traction unit. Table 3 shows the data entries associated to the train with train number 168. The train is run from Fws to ZUE.

Table 2: Examples for operational points.

id	code	name	id	code	name
363	Ams	Amstetten	2049	Sv	St.Valentin
541	Bc	Buchs (national border)	2454	Ws	Wien Westbf
697	Fws	Vienna International Airport	2777	ZUE	Zuerich HB (CH)
1035	I	Innsbruck Hbf	27505	Tfd	Tullnerfeld
1845	Sb	Salzburg Hbf	27536	Wbf	Wien Hbf
2043	Pb	St.Pölten Hbf			

On the way from Wbf to Sb, the train is operated with two traction units. In this case, these traction units are associated with the same cluster, but that is not a necessary condition.

Table 3: Data entries for train 168 in the timetable period 2017/18 (date format DD.MM.YYYY).

type	loc.	valid from	valid to	day	#	from	to	dep.	arr.	h.dep.	h.arr.
8090	Ws	10.12.2017	08.12.2018	We	168	Fws	Wbf	15:03	15:18	0	0
8090	Ws	10.12.2017	08.12.2018	We	168	Wbf	Sb	15:30	17:52	0	0
8090	Ws	10.12.2017	08.12.2018	We	168	Wbf	Sb	15:30	17:52	0	0
8090	Ws	10.12.2017	08.12.2018	We	168	Sb	I	17:56	19:44	0	0
8090	Ws	10.12.2017	08.12.2018	We	168	I	Bc	19:48	22:03	0	0
8090	Ws	10.12.2017	08.12.2018	We	168	Bc	ZUE	22:12	23:20	0	0

In addition to the regular trains, there are other services that need to be planned. These services are typically stationary, i.e. they are operated at the same station, so they can be omitted for the simulation model. Most of the stationary services are used to preheat train cars (typically passenger cars) over night.

2.3 Schedules

The optimization of a traction unit circulation is based on the train data. For the simulation model, a finer resolution of the data is needed, in order to accurately predict the impact of delays.

A schedule defines a sequence of operational points for a specific train as well as arrival and departure times for each point. Additionally, the traveled distance up to this point is given in meters. The schedule data describes the maximum form of a train, but a train can also be operated on only a part of the schedule. Table 4 shows an example of a train that is operated differently on weekends. Moreover, it is operated by two traction units during the week (the given representation is shortened, in the real data the rows are duplicated for each corresponding day), but by only one traction unit on weekends.

Using the schedules for all trains, a network between all operational points can be created. Two consecutive entries in a schedule define an edge of the corresponding network graph, whose nodes are the operational points. The length of an edge can either be defined by the distance traveled between two entries or by the time needed for the distance defined by the difference between arrival at the following entry and departure at the previous entry.

The generated network is the basis of the simulation model. The assumption is that every edge of the graph (which represents real world tracks) is occupied if a train is currently using it and only a predefined number of trains can occupy a track at the same time. During optimization, the network is used to create empty runs in order to relocate traction units.

2.4 Delays

Punctuality and delays are widely used general concepts, but their exact definitions and computation methods vary among countries and railway companies. Additionally, punctuality targets are commonly different for

Table 4: Excerpt of schedule for train 1909 and its associated trains (date format DD.MM.YYYY).

arr.	dep.	nr	meters	at	arr.	dep.	nr	meters	at
05:26	05:26	1	0	Sv	07:06	07:06	57	103014	Pb
		⋮					⋮		
05:55	06:10	21	39542	Ams	07:19	07:20	65	133654	Tfd
		⋮					⋮		
06:55	06:55	56	103014	Pb	07:40	07:40	77	162850	Ws

type	loc.	valid from	valid to	day	#	from	to	dep.	arr.	h.dep.	h.arr.
4744	Ws2	10.12.2017	04.02.2018	Mo-Fr	1909	Sv	Ws	05:26	07:40	0	0
4744	Ws2	10.12.2017	04.02.2018	Mo-Fr	1909	Sv	Ws	05:26	07:40	0	0
4744	Ws2	10.12.2017	04.02.2018	Sa,Su	1909	Pb	Ws	07:06	07:40	0	0

freight trains, long distance passenger trains and regional/suburban trains per country and railway company. Both concepts can be attributed to individual operational points, train paths and trains as well as to the whole railway network. While punctuality refers to the number of trains that are not delayed compared to the total number of trains operated within the attribution context, delays are usually positive deviations between the realized and the scheduled times of activities (Cerreto et al. 2016); especially in the context of freight trains, it can be agreed on that this deviation has to be larger than a predefined threshold. If also considered, negative delays belong to the effect of a train gaining time, i.e. not using the full amount of timetable supplement included by the planner.

Delay-cause tracking is regulated and standardized under the UIC leaflet 450-2 (UIC 2009). A first classification of delay types can be made according to whether the delay occurs while a train is located at an operational point (usually train stations or signals) or at a section between two operational points. An additional delay classification approach is based on the specific parts of the train path:

- *Initial delay* refers to the time difference between the realized and the scheduled startup (see Section 3.3) or departure of a train.
- *Operational delay*: In literature, most of the delay classifications distinguish between
  - *Primary delays* that refer to unexpected extensions of the planned times of individual processes scheduled, e.g. due to equipment failures, temporary obstacles on the railway track, or passenger flows larger than expected, and
  - *Secondary delays* that are generated by operational conflicts arising due to primary delays; they are delays that emerge from queuing.
- *Final or total delay* refers to the time difference between the realized and the scheduled shutdown (see Section 3.3) or arrival of the train.

While operational secondary delays are created and propagated in a very natural way by the agent-based model introduced in Section 3, operational primary delays are introduced and calculated by using a separate machine learning model. The regression task of predicting the operational primary delay of a train was tackled by various machine learning algorithms such as neural networks, k-nearest neighbor regression and several random forest regression approaches. Some of the input features used are presented in Table 5. The data sets that were used to train and validate the machine learning models originated from database extracts from historical operational data from the period 10 December 2017 to 8 December 2018. 3-fold cross-validation was applied.

As the similarity of delays of freight and passenger trains is only marginal, it became clear at a very early stage of the model development that building separate delay prediction models for freight and passenger

Table 5: Selected input features of the primary train delay prediction model.

Feature Description	Feature Type	Feature Description	Feature Type
Planned departure and arrival	cyclic	System of drive of the traction unit	categorical
Lat. and Lon. of stations	continuous	Traction unit type series	categorical
Distance between two stations	continuous	Type of the train operation	categorical
Altitude difference of stations	continuous	Delay at departure	continuous

trains results in a significant reduction of the overall MSE. A further fineness of grain by building separate models for each circulation is currently under development.

Regarding the model performance, classical random forests, extremely randomized trees (Geurts et al. 2006) and gradient boosting algorithms (Friedman 2001; Friedman 2002) stood out (Leser et al. 2019). The gradient boosting models proved to be most feasible for being integrated into the agent-based model due to the significant shorter evaluation time, which is an important fact due to the very high number of evaluations needed.

3 MODEL

The used model is a standard agent-based model with a time-continuous, i.e. event-based, time update. Agents depict trains and traction units, but also the elements of the rail-network like stations and sections, are represented as agent-like objects with a certain behavior. It was developed to have a maximum of expansion capability while maintaining the greatest possible clarity. The basic idea is to have only one type of entity that registers events to the scheduler, while the other agents only react to the fired events.

3.1 Section

The edges of the network created from the provided schedules are called sections in the model. Sections are unique, i.e. there is a maximum of one section between two operational points in each direction. The main parameter of a section is its capacity. Roberts et al. (2010) show that the definition of capacity is very complex as it relies heavily on the objectives of the service. Given the fine resolution of the sections in the data it is assumed that the capacity of a section in the model reflects the number of tracks that are available on the corresponding segment. This assumption holds if the length of the sections is not too long.

An analysis of the provided data yields that the average distance of a section in Austria is 2.33 km while it is 4.68 km over all sections. The overall data includes international traffic, but the density of operational points is lower than in Austria. It is probable to assume that the above assumption does not hold outside of Austria. Therefore, the model is currently limited to Austria, where the provided data quality is assumed to be high enough to deliver reliable results. In order to incorporate international delays and delays at borders, trains that start or end at the Austrian border are also included in the model, but the corresponding sections get assigned an unlimited capacity.

In order to estimate the capacity of the sections in Austria an empirical ansatz is used. Given the provided schedules it is assumed that the planned time table does not lead to delays. As a result, the capacity of a section is estimated as the maximum number of planned trains that are scheduled at the same time at the specified section.

3.2 Traction Unit

A circulation associates each entry in the train data (see Table 3) related to the considered cluster with a traction unit. Additionally, empty runs for each traction unit are defined. This leads to a schedule for the traction unit that can be worked through similarly to the schedule of a singular train. Every time an entry of its schedule is finished, the traction unit moves on to the next element.

3.3 Schedule Entry

In contrast to the schedule described in Section 2.3, the simulation considers the schedule entries for sections instead of operational points, as secondary delays are created mainly on sections. A schedule entry is therefore a section linked to departure and arrival times. Additionally, handling times, a list of traction units, and the train the schedule entry is assigned to are defined. It should be mentioned that it is possible that no traction unit is assigned to a schedule entry. In this case, the schedule entry can be executed if the linked section is free, but does not have to wait for a traction unit to be ready. Figure 2 shows the flow chart of the schedule entry agent. Several types of events can be scheduled by the agent:

- *Startup:* The agent goes into startup phase, the scheduled time for the event is the departure time minus the time required for the start up (handling departure). For this event all assigned traction units have to be ready (i.e. they have to be at the right entry in their respective schedules). If this is the case, the next event (*departure*) is scheduled after the time for the start up has elapsed. If not, the *start up* event is rescheduled. The updated event time depends on the traction units that are not ready yet.
- *Departure:* If the assigned section is free, the next event (*arrival*) is scheduled after the duration of the schedule entry. If the section is blocked, the *departure* event is rescheduled. The scheduled time is the point in time, when the section is free again.
- *Arrival:* When the arrival event is executed, the corresponding section is freed and the next event (*shutdown*) is scheduled after the time the shutdown needs (handling arrival).
- *Shutdown:* After the shutdown, the next schedule element of the corresponding train can be initialized and its *startup* event is scheduled. The schedules of the assigned traction units are advanced.

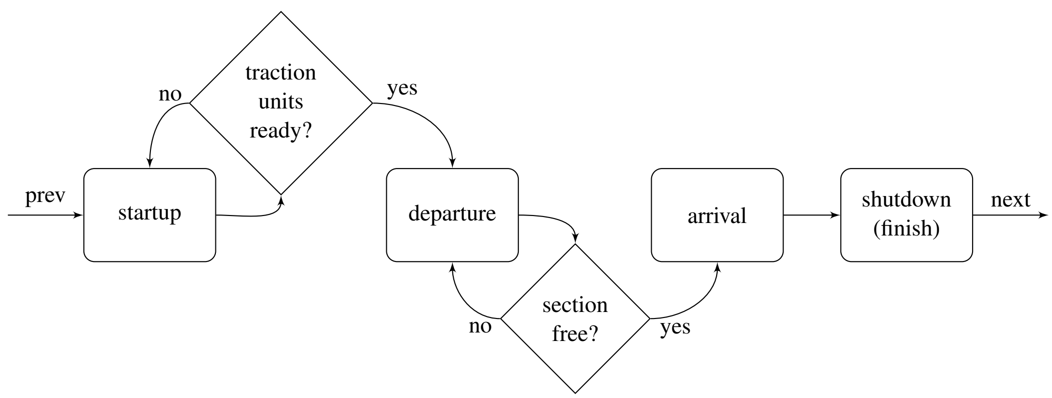

Figure 2: Flow chart of schedule entry agent.

Up to now the model is strictly deterministic, the schedule provides a strict sequence of events and the tie-breakers in case of simultaneous events are defined in a deterministic way as well. A stochastic element is provided by the inclusion of delays. The model considers the four types of delays presented in Section 2.4. While the *primary delays* are provided as input for the simulation model, the other types result through delay propagation during the model execution.

4 SIMULATION

The model was implemented in a self-developed, Java-based framework for agent-based models (dwh GmbH 2020). A special feature of this framework is the possibility to control the sequence of seeds for the random number generator. This, among other things, enables the creation of very specific Monte Carlo simulation runs.

For the optimization, the most relevant output is the effect of a specified traction unit circulation on the overall delay in the system. In order to differentiate between the delays that are caused by the circulation directly (because a train has to wait for a traction unit) and the delays inherent to the underlying schedule (because sections are blocked), the simulation is performed with and without the assignment of traction units. The simulation is run as a Monte-Carlo-Simulation. This means it is executed a certain number of times with different, random *primary delay allocations*. The mean of the resulting delays is calculated and used for further analysis. As mentioned before, the *primary delays* are the only stochastic elements in the model. If these are drawn at the beginning of the simulation, the rest of the simulation is strictly deterministic. This implies that the effects of a specific primary delay allocation over all schedule entries can be studied in a purely deterministic way. Therefore the simulations with and without assignment of traction units are simulated with the same set of random number seeds, allowing a better distinction between the two types of (secondary) delay causes.

The ratio of the sum of secondary delays is called *markup* and can be interpreted as an indicator for the effect of the circulation on the delays:

$$\text{markup} = \frac{\text{secondary delay with traction unit assignment}}{\text{secondary delay without traction unit assignment}}.$$

The markup can be calculated for different entities within the model. For single schedule entries, the markup does not make much sense, because the secondary delay will often be 0 and the markup not finite. For the markup of trains, the different delay types for a train have to be defined in a similar way as for schedule entries. The *total* and *initial delay* are defined in the same way as for schedule entries. The *primary delay* is the sum of all primary delays of schedule entries of the train. The *secondary delay* is then

$$\text{secondary delay} = \text{total delay} - \text{initial delay} - \text{primary delay},$$

similar to the delays of schedule entries. The markups for trains can be calculated for each entry in the train data, by restricting the associated train to the partial schedule defined in the train data entry.

For the overall effect of the provided circulation on the delays two indicators can be calculated. For both indicators the mean secondary delay with and without traction unit assignment over the train data entries is calculated and the markup of these two means is computed. For the *overall markup* the markup over all train data entries is calculated and for the *circulation markup* the markup over all train data entries where a traction unit was assigned is calculated.

5 OPTIMIZATION

In a mathematical context, the optimal assignment of traction units to scheduled, ordered trains is known as Locomotive Scheduling Problem (`LSP`). In more detail, the aim of the `LSP` is to assign individual locomotives to trains such that the overall operating costs are minimized. The `LSP` is a very well studied problem and was first introduced by Gleaves (1957). A variety of variants and solution approaches exist for the `LSP`, we refer to Piu and Speranza (2014) for a detailed overview. In Frisch et al. (2019), a sparse multi-graph model is introduced and a Mixed-Integer Linear Program for solving the `LSP` incorporating maintenance constraints is formulated. In this work, the modeling approach is adapted for the `LSP`.

5.1 Problem Formulation

The input is defined by a given railway network and a train schedule over a specific planning horizon as well as a pool of available locomotives (traction units). For each scheduled train, the corresponding train stations including departure and arrival times and the permitted locomotive types are stated. The information about the type and the current locations of all locomotives $L := \{k_1, \dots, k_m\}$, $m \in \mathbb{N}$, is available, too. Based on this information, a directed and weighted sparse problem graph with different types of nodes and arcs is constructed. Generally, nodes represent train stations at particular times and arcs represent possible trips of locomotives. Four different types of nodes are defined:

- *Locomotive starting nodes* V^s. For each locomotive $k \in L$ a node $i \in V^s$ including time information of availability of k at the current train station is modeled.
- *Departure nodes* V^d. For each departure train station of all scheduled trips, a node $i \in V^d$ with information about exact departure time and location is modeled.
- *Arrival nodes* V^a. For each arrival train station, nodes $i \in V^a$ are defined similar to departure nodes.
- The *artificial final node* v_f. This node is introduced for a technical reason, such that locomotives can end up in arbitrary train stations at any time.

Following, the set of nodes is defined by $V := V^d \cup V^a \cup V^s \cup \{v_f\}$. Arcs display two types of trips:

- *Scheduled trip arcs* A_1. For each scheduled trip from $i \in V^d$ to $j \in V^a$ an arc $(i, j) \in A_1$ between the corresponding departure and arrival node is modeled.
- *Deadhead trip arcs* A_2. For each possible journey between train stations that does not correspond to scheduled trips, an arc $(i, j) \in A_2$ is introduced passable by locomotives not pulling a train.

While scheduled trips are predefined, rules for possible deadhead trips must be stated. Therefore, the notion of a *reachable node* is introduced. A node $j \in V$ is *reachable* by a locomotive $k \in L$ located in $i \in V$ if a path between i and j exists and k is able to arrive in j on time. Deadhead trip arcs can be inserted between $i \in V^s$ and $j \in V^d$, $i \in V^a$ and $j \in V^d$, $i \in V^s$ and v_f, and $i \in V^a$ and v_f. The set of arcs is defined by $A := A_1 \cup A_2$. Finally, the problem graph is given by $D := (V, A)$. The fact of modeling the time components in D implicitly leads to the advantage that time-dependent decision variables in the following solution approaches are not necessary.

5.2 Solution Approaches

For solving the `LSP`, all maintenance relevant components from the `MILP` stated in Frisch et al. (2019) are removed, which results in a well-performing solution approach due to the sharply decreased complexity. The resulting `LSP-MILP` reads as follows:

$$\min \quad c_{km} \sum_{(i,j)\in A_2} d_{ij} x_{ij} + c_{loc} \sum_{k\in L} s_k \qquad \text{(Objective)}$$

$$\text{s.t.} \quad \sum_{(i,j)\in A_2} x_{ij} = 1, \quad j \in V^d, \qquad (1)$$

$$\sum_{(i,j)\in A_2} x_{ij} = 1, \quad i \in V^a \cup V^s, \qquad (2)$$

$$\sum_{(i,j)\in A_2} x_{ij} = m, \quad j = v_f, \qquad (3)$$

$$s_k + x_{i_k v_f} = 1, \quad i \in V^s,\ k \in L,\ i = i_k, \qquad (4)$$

$$q_{ik} = 1, \quad i \in V^s,\ k \in L,\ i = i_k, \qquad (5)$$

$$\sum_{k\in F} q_{ik} = 1, \quad i \in V \setminus \{v_f\} \cup V^s, \qquad (6)$$

$$\left.\begin{aligned} &q_{ik} = q_{jk}, \quad (i,j) \in A_1, k \in F,\ \text{if}\ k \in P_{ij}, \\ &q_{ik} = q_{jk} = 0, \quad (i,j) \in A_1, k \in L,\ \text{if}\ k \notin P_{ij}, \end{aligned}\right\} \qquad (7)$$

$$q_{jk} \geq q_{ik} - (1 - x_{ij}), \quad (i,j) \in A_2, k \in L, j \neq v_f, \qquad (8)$$

$$x_{ij}, q_{ik}, s_k \in \{0,1\}. \qquad (9)$$

Using real costs as weights the (Objective) minimizes the overall cost for deadhead kilometers driven and locomotives used. Flow conserving constraints (1), (2) and (3) guarantee that all scheduled trips are conducted. By (4) the locomotive counting variable is updated only if the corresponding locomotive k uses an outgoing arc, connecting its starting node and a departure node. Constraints (5) link each locomotive with its starting node. Equations (6) ascertain that each node, except for the final node, is visited by exactly one locomotive. Constraints (7) ensure that only permitted locomotives conduct scheduled trips. Inequalities (8) guarantee that if an arc $(i, j) \in A_2$ is used, nodes i and j are visited by the same locomotive.

The `LSP-MILP` was tested on real-world based instances provided by the Austrian Federal Railways (ÖBB). These instances covered more than 2000 scheduled trains. For all instances, the `LSP-MILP` delivered optimal solutions within five minutes. However, in real-time applications respective instances should be solved in a few seconds. Thus, a fast and simple depth first search heuristic (`DFSH`) was developed additionally. The `DFSH` delivered high quality solutions for the `LSP` under very low computation time, which makes it usable for large-scale real-time applications.

All algorithms were implemented in Java. For solving the `LSP-MILP`, Gurobi 8.0.1 was used.

5.3 Integration of Simulation Results

The robustness of optimization solutions regarding delays in the whole railway system is of great interest. Thus, the goal of this work is to consolidate simulation and optimization. As described in Section 4, the effect of a locomotive assignment on (secondary) delays is measured by the *markup*. Markups are calculated as local effects for each scheduled trip as well as a statement about the overall effect of locomotive circulations, see Section 4.

For considering the effects on delays caused by optimal locomotive circulations, the single markups are integrated in the optimization model as follows. After analyzing the heaviness of the delays and their local concentrations, well-chosen time buffers are added to the respective arcs in the problem graph, i.e., arcs which show insufficient space of time are dropped out. Obviously, this procedure shrinks the solution space and results in a possible deterioration of the optimal solution. Thus, time buffers must be chosen wisely to not significantly worsen the solution.

In order to reach an effective integration process, a loop between optimization and simulation is most relevant. We iteratively generate optimal circulation plans, then calculate markups via simulation and again run the optimization under consideration of the beforehand calculated circulation markups. With increasing number of iterations robustness can be improved while locomotive circulations stay nearly optimal.

6 RESULTS & OUTLOOK

As a test case, the circulation for the traction unit cluster *1144-TR* was used. This is one of the biggest clusters and it includes national and international traffic as well as passenger and freight services. The manual circulation planning of the ÖBB resulted in a circulation using 83 traction units. For the best results provided by the optimization algorithm, 79 traction units were needed. For each schedule, 10 simulation runs were carried out (5 with and 5 without traction unit assignment). Figure 3 shows the relation between the optimization and simulation results.

All shown schedules are well below the manual planning result. Interestingly, in this scenario the two optimization results that need 79 traction units yield the best results regarding the circulation markup. A trade-off between empty run kilometers versus markup can be seen between the two results. Overall the results show that a better optimization result does not necessarily lead to less robustness. The concrete choice of circulation requires more detailed evaluation as greater delays and markups may be more tolerable for certain parts of the train schedule than for others. In order to provide a fully automated planning algorithm that incorporates the optimization as well as the simulation, such constraints have to be defined very clearly. At the moment the iterative process described in Section 5.3 is done manually (human in the loop).

For future work, a thorough validation of the agent-based model using the historical data is planned and a study on the number of Monte-Carlo-Simulation runs based on the findings in Bicher et al. (2019) will be conducted. Furthermore, maintenance constraints will be considered in the model. As mentioned, the

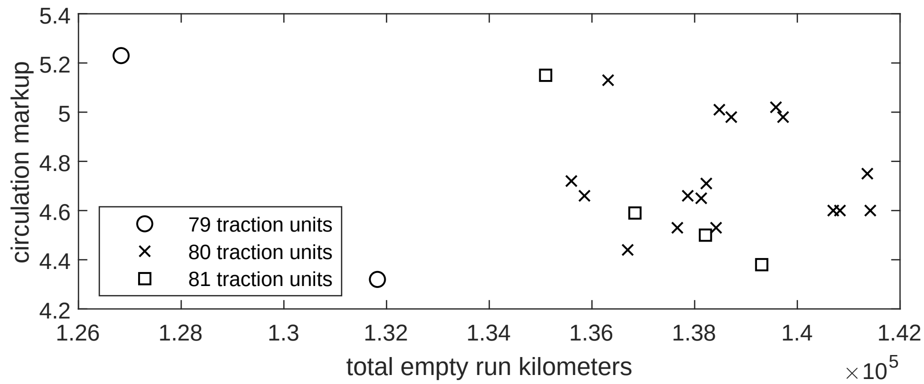

Figure 3: Scatter plot of circulations. Empty run kilometers are plotted against the resulting *circulation markup*. The number of traction units needed is indicated by the different types of markers.

optimization already incorporates maintenance constraints. In order to consider them in the simulation model as well, an interface to provide the relevant data has to be created between optimization and simulation. In addition, consequences for missed maintenance appointments have to be defined. Other planned extensions to the model include the energy consumption of the traction units, the consideration of train drivers and other train personnel, and widening the scope of the model to international traffic.

ACKNOWLEDGMENTS

This work was funded by the Austrian Research Promotion Agency (FFG) in the course of the project *AundO* (project number 867178). The data was provided by Austrian Federal Railways (ÖBB).

REFERENCES

Bendfeldt, J., U. Mohr, and L. Muller. 2000. "RailSys, a system to plan future railway needs". *WIT Transactions on The Built Environment* 50:249–255.

Bicher, M., M. Wastian, D. Brunmeir, M. Rößler, and N. Popper. 2019. "Review on Monte Carlo Simulation Stopping Rules: How Many Samples Are Really Enough?". In *Proceedings of the 10th EUROSIM Congress on Modelling and Simulation*. July 1st-5th, Logrono, Spain.

Cerreto, F., O. Nielsen, S. Harrod, and B. F. Nielsen. 2016. "Causal Analysis of Railway Running Delays". In *11th World Congress on Railway Research*. May 29th-June 2nd,Milan,Italy.

Davidsson, P., L. Henesey, L. Ramstedt, J. Törnquist, and F. Wernstedt. 2005. "An analysis of agent-based approaches to transport logistics". *Transportation Research part C: emerging technologies* 13(4):255–271.

dwh GmbH 2020. "dwh GmbH news entry for the ABT simulation framework". http://www.dwh.at/en/news/the-power-of-the-abt-simulation-framework/ accessed 23rd April 2020.

European Union 2019. "Commission Implementing Regulation (EU) 2019/777 of 16 May 2019 on the common specifications for the register of railway infrastructure and repealing Implementing Decision 2014/880/EU (Text with EEA relevance.)". *OJ* L 139I:312–355. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32019R0777 accessed 23rd April 2020.

Friedman, J. H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine". *The Annals of Statistics* 29(5):1189–1232.

Friedman, J. H. 2002. "Stochastic Gradient Boosting". *Computational Statistics & Data Analysis* 38(4):367–378.

Frisch, S., P. Hungerländer, A. Jellen, and D. Weinberger. 2019. "A Mixed Integer Linear Program for Optimizing the Utilization of Locomotives with Maintenance Constraints". In *Operations Research Proceedings 2018*, 103–109. Springer.

Geurts, P., D. Ernst, and L. Wehenkel. 2006. "Extremely Randomized Trees". *Machine Learning* 63(1):3–42.

Gleaves, V. 1957. "Cyclic scheduling and combinatorial topology: Assignment and routing of motive power to meet scheduling and maintenance requirements Part I-A statement of the operating problem of the Frisco railroad". *Naval Research Logistics Quarterly* 4(3):203–205.

Horni, A., K. Nagel, and K. W. Axhausen. 2016. *The multi-agent transport simulation MATSim*. Ubiquity Press London.

Leser, D., M. Wastian, M. Rößler, M. Landsiedl, and E. Hajrizi. 2019. "Comparison of Prediction Models for Delays of Freight Trains by Using Data Mining and Machine Learning Methods". *Simulation Notes Europe SNE* 1(29):45–47.

Mitra, G. 1973. "Investigation of some branch and bound strategies for the solution of mixed integer linear programs". *Mathematical Programming* 4(1):155–170.

Nash, A., and D. Huerlimann. 2004. "Railroad simulation using OpenTrack". *WIT Transactions on The Built Environment* 74:45–54.

Piu, F., and M. G. Speranza. 2014. "The locomotive assignment problem: a survey on optimization models". *International Transactions in Operational Research* 21(3):327–352.

Roberts, C., F. Schmid, P. Connor, V. Ramdas, and J. Sharpe. 2010. *A new railway system capacity model*. Published Project Report PPR541. ed. Chicago: Crowthorne: Transport Research Laboratory. https://trl.co.uk/sites/default/files/PPR541.pdf accessed 23rd April 2020.

Rößler, M., M. Wastian, M. Landsiedl, and N. Popper. 2018. "An agent-based model for robustness testing of freight train schedules". In *Proceedings of the 17th International Conference on Modeling and Applied Simulation (MAS 2018)*, edited by A. G. Bruzzone, F. De Felice, C. Frydman, F. Longo, M. Massei, and A. Solis, 101–105. Via Umberto Nobile 80, 87036 Rende (CS), Italy: CAL-TEK S.r.l.

UIC 2009. *Leaflet 450—2 - Assessment of the performance of the network related to rail traffic operation for the purpose of quality analyses - delay coding and delay cause attribution process*. 5th ed. 16 rue Jean Rey F-75015 Paris, France: Union Internationale des Chemins de fer.

Waraich, R. A., D. Charypar, M. Balmer, and K. W. Axhausen. 2009. "Performance improvements for large scale traffic simulation in MATSim". In *9th STRC Swiss Transport Research Conference: Proceedings*, Volume 565. Swiss Transport Research Conference.

AUTHOR BIOGRAPHIES

MATTHIAS RÖSSLER is researcher at dwh GmbH. His fields of research include modeling and simulation in the fields of health care, logistics, and energy simulations using various modeling techniques including ODE- and DAE-based models as well as agent-based systems. He is currently working on his PhD thesis at TU Wien. His email address is matthias.roessler@dwh.at.

MATTHIAS WASTIAN is senior data scientist at dwh GmbH. His scientific focus lies on artificial intelligence, machine learning as well as modeling and simulation in the fields of logistics, health care and energy simulations. He is currently working on his PhD thesis at TU Wien. His email address is matthias.wastian@dwh.at.

ANNA JELLEN is researcher at AAU Klagenfurt. Presently, she is writing on her PhD thesis, developing optimization approaches for NP-hard routing and scheduling problems. Her focus is on providing high-performance optimization algorithms that can be deployed in real life. Her email address is anna.jellen@aau.at.

SARAH FRISCH is researcher at AAU Klagenfurt. She is researching on real-world problems in the railway sector including locomotive scheduling, railway crew scheduling, freight car routing, and more. In the course of her PhD thesis she is working on developing exact and heuristic approaches to tackle these problems. Her email address is sarah.frisch@aau.at.

DOMINIC WEINBERGER is senior software developer at hex GmbH and Master student in computer science. He is specialized in programming optimization software for companies with routing and scheduling problems. His email address is dominic.weinberger@hex-solutions.com.

PHILIPP HUNGERLÄNDER is associate professor at AAU Klagenfurt and CEO of hex GmbH. His research focus lies on NP-hard optimization problems emerging from various projects with partners from industry. His email address is philipp.hungerlaender@aau.at.

MARTIN BICHER is research associate at TU Wien and scientific employee at dwh GmbH. He finished his PhD in Technical Mathematics at TU Wien in winter 2017. His doctoral thesis was about mean-field behavior of microscopic models. His email address is martin.bicher@tuwien.ac.at.

NIKI POPPER is research associate at TU Wien and CEO of dwh GmbH. He is responsible key-researcher of K-Project DEXHELPP and head of the corresponding association. His research focus lies on comparison of different modeling techniques. His email address is nikolas.popper@tuwien.ac.at.