

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Rail Transport Planning & Management

journal homepage: www.elsevier.com/locate/jrtpm

Full Length Article

Utilizing phase-type distributions for queueing-based railway junction performance determination

Tamme Emunds *, Nils Nießen *Institute of Transport Science, RWTH Aachen, Mies-van-der-Rohe Straße 1, Aachen, 52074, Germany*

ARTICLE INFO

Keywords:

Railway junction capacity
Queueing system
Timetable-independent
Continuous-Time Markov chain
Performance analysis

ABSTRACT

To ensure the effective and objective development of transportation networks, it is crucial to identify performance limitations across various subsystems. A timetable-independent assessment of infrastructure capacity at railway junctions is a fundamental aspect of long-term rail network planning. While recent research introduced queueing-based methods to quantify route-based railway junction performance, modelling arrival and service processes has been limited to exponential distributions. This work utilizes Phase-Type Distributions to propose an extension to a previously described Continuous-Time Markov Chain model. In a comparison between assumed distribution combinations, the effect of a more detailed stochastic process modelling is described. Furthermore, an analysis of the differences to a simulation method is conducted for an exemplary railway junction. The introduced method enables infrastructure managers to accurately model stochastic processes for performance determination in the early stages of the strategic planning phase.

1. Introduction

Due to the increasing demand for environmentally friendly modes of transportation, railway infrastructure managers worldwide need to develop their infrastructure further. Performance analysis of individual facilities serves as a valuable method to evaluate existing infrastructure and compare expansion and new construction scenarios. Some well established methods are particularly suited for analysing current infrastructure and timetables, for example those designed to determine capacity utilization (UIC, 2004). On the other hand, efficient resource management requires a focus on long-term planning. For this purpose, timetable-independent methods can be particularly useful.

While such methods for railway lines are already well established in their practical implementation, methods for railway junctions or nodes are still the subject of ongoing research. The currently utilized systems in practice employ single-channel approximations for multichannel service systems, such as railway route nodes and junctions. However, those systems with parallel utilizable routes are particularly critical points that need to accommodate the traffic from multiple railway lines.

In a preceding work (Emunds and Nießen, 2024a), a multichannel method was introduced to analyse the performance of railway junctions and estimate the necessity for an overpass structure to mitigate route dependencies. In that study, the arrival and service processes were modelled using only exponential distributions, necessitating the use of approximation formulas to accommodate other probability distributions.

In contrast, this work introduces a novel approach, modelling general independent arrival and service processes with phase-type distributions in a Continuous-Time Markov Chain for multi-channel railway systems. This model can be used to obtain more accurate

* Corresponding author.

E-mail address: emunds@via.rwth-aachen.de (T. Emunds).

<https://doi.org/10.1016/j.jrtpm.2025.100523>

Received 5 December 2024; Received in revised form 17 March 2025; Accepted 8 May 2025

Available online 26 May 2025

2210-9706/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

approximations of the queue-lengths of the different routes through the infrastructure, thereby enabling timetable-independent quantifications of the occupancy of each route, which allows for a detailed bottleneck analysis of the infrastructure.

The contribution of this work is manifold, including:

- A novel Continuous-Time Markov Chain model that utilizes phase-type distributions for modelling arrival and service processes in a railway junction.
- An algorithm designed to efficiently compute timetable capacity by comparing queue-length estimations to their respective thresholds.
- A comprehensive study analysing the influence and approximation quality of different queuing models for varying distributions of traffic types across routes.
- A case study demonstrating the usability of route-based analysis to evaluate the performance of a railway junction infrastructure that accommodates both freight and passenger traffic.

The remainder of this work is structured as follows. Section 2 presents the current state of research. In Section 3, a formal description of the junction capacity determination problem is provided. Subsequently, the novel model and the employed approximation formulas are introduced in Sections 4.3 and 4.5, along with an explanation of the algorithm for efficient capacity determination in Section 4.6. Section 5 offers a validation of the proposed method and a comparison with other models under different traffic distribution scenarios. Finally, an example railway junction combining a freight and a passenger traffic line is analysed in Section 6.

2. Related work

This section provides a summarized overview of the state-of-the-art for railway capacity analysis, focusing on key methodologies and their applications.

Railway capacity is defined in various ways depending on the planning stage and requirements (Jensen et al., 2020; Emunds and Nießen, 2024a). The *theoretical capacity* refers to the maximum number of trains that can be scheduled without conflicts, considering driving dynamics and control systems. The *timetable capacity* or *maximal capacity* includes additional factors such as train-mix and schedule quality. Finally, the *operational capacity* or *practical capacity* accounts for disturbances and delays, ensuring acceptable operational quality. Research also focusses on calculating the *capacity utilization*, which measures how much of the available capacity is used in a given timetable.

Various methodologies have been employed for railway performance analysis and vary in their dependency on timetables; some are timetable-dependent while others are not, making them useful for early infrastructure planning stages.

These methodologies differ based on the analysed infrastructure (lines, junctions, stations, networks), infrastructure decomposition, and solution methods like mixed integer programming (MIP) or matrix calculations.

The *UIC Code 406* (UIC, 2004, 2013) is widely used internationally for assessing railway line and station capacities by determining capacity utilization through compression methods requiring a timetable or randomly generated sequences of train types to overcome timetable dependencies (Goverde, 2007; Goverde et al., 2013; Abril et al., 2008; Bešinović and Goverde, 2018; Jensen et al., 2020).

Optimization methods estimate theoretical capacity through linear mixed-integer programming problems for lines, stations, and networks. They solve railway timetabling problems to build optimal timetables based on objective functions (Zwaneveld et al., 1996; Burdett and Kozan, 2006; Burdett, 2016; Harrod, 2009; Lusby et al., 2011; Cacchiani and Toth, 2012; Cacchiani et al., 2016; Yaghini et al., 2014). Some optimization approaches incorporate rolling stock information or model delay propagation effects using max-plus algebras (De Kort et al., 2003; Mussone and Wolfier Calvo, 2013; Liao et al., 2021).

Simulations provide detailed insights into operational parameters by simulating train operations based on given timetables or random variables representing arrival and service processes (Zieger et al., 2018; D’Acierno et al., 2019).

Analytical methods based on queueing theory efficiently analyse timetable or operational capacities during early planning stages by considering inter-arrival and service time distributions, thus typically remaining independent of a predefined schedule. Methods to determine the performance of railway lines (Schwanhäußer, 1974; Schwanhäußer and Schultze, 1982; Wendler, 2007; Weik and Nießen, 2017), railway junctions (Schwanhäußer, 1978; Wakob, 1981; Nießen, 2008, 2013; Schmitz et al., 2017; Weik, 2020; Emunds and Nießen, 2024a) and track groups (Potthoff, 1970; Fischer and Hertel, 1990) have been introduced.

These timetable-independent models inherently address uncertainties in arrival and service processes, such as train precedence, buffer times or minimum headway times. To incorporate these uncertainties into UIC Code 406 or simulation based methodology, multiple timetable samples must be considered (Jensen et al., 2020; Weik et al., 2020). Alternatively, optimization methods require additional complexity to ensure *robust* solutions (Liebchen et al., 2009).

The timetable-independent method proposed here is based on queueing theory. Table 1 compares other queueing-based methods for evaluating railway junction performance with the model introduced in this paper. For example, Schwanhäußer (1978) introduces a single-channel approximation to obtain performance indicators of railway junctions by scaling route-based waiting times with the occurrence probability of route conflicts. While Schwanhäußer considers exponentially distributed inter-arrival times, Wakob (1981) extends this model for Erlang distributed arrival processes.

Nießen (2008, 2013) analyses interlocking nodes for their timetable and operational capacity using loss probabilities in service systems. To this end, exponentially distributed inter-arrival and service times are employed and results are scaled to approximate non-exponential service times, utilizing the variation coefficient.

In Schmitz et al. (2017), a railway junction is divided into two service stations and is modelled as a Markov Chain. Notably, the exponential scaling of the state space due to the storage of the request type, as well as the assumption of an existing subdivision into

Table 1
Analytical junction capacity determination methodology.

Literature	Multi-channel model	Timetable capacity	Route-based decomposition	Route-based quality assessment	Explicit non exponential process modelling	Solution method
Schwanhäußer (1978)		✓	(✓)			closed-form formula
Wakob (1981)		✓	(✓)		(✓)	iterative formula
Nießen (2008, 2013)	✓	✓		(✓)		iterative formula
Schmitz et al. (2017)	✓	✓			✓	matrix-vector equations
Weik (2020)		✓	(✓)		✓	matrix-vector equations
Emunds and Nießen (2024a)	✓	✓	✓	✓		probabilistic model-checking
introduced here	✓	✓	✓	✓	✓	probabilistic model-checking

Remarks: Please note that a statement in bracelets means that this feature is partially supported.

independent service stations, are significant properties of this approach. This method allows a railway junction to be considered as a multi-channel service system and analysed using phase-type distributed arrival and service processes.

To analyse the significance of switch connections in station yards, Weik (2020) examines interlocking nodes by approximating them as a series of interconnected service stations in a single-channel system, following the methodology of Schwanhäußer (1978), Wakob (1981). In this model, the arrival, service, and an additional repair processes are represented using phase-type distributions.

In contrast, this paper is based on the model proposed by Emunds and Nießen (2024a). This model enables the performance evaluation of various junction infrastructures by discretizing the utilized infrastructure based on operational route paths and modelling the arrival and service of requests on this infrastructure using a Markov chain. Initially, only exponentially distributed inter-arrival and service processes were considered using the Markov chain model, which had to be converted into processes with variation coefficients not equal to 1 using approximation formulas for multi-channel service systems according to Fischer and Hertel (1990).

In this work, models are derived that enable the computation of phase-type distributions. In particular, several influencing factors on the accuracy of the approximation are presented and analysed.

3. Problem description

This work introduces a method to evaluating the performance of a double-track railway junction. The performance is assessed using the metric known as *timetable capacity*, which gauges the infrastructure's ability to accommodate train schedules, see also Wendler (2007), Emunds and Nießen (2024a). Specifically, it pertains to the timetabling process, wherein the infrastructure manager allocates service slots to various requests submitted by train operators. With this approach, random distributed requests can be modelled, encompassing all potential timetables that could result in such a procedure. Conflicts may arise due to the inherent randomness of the requests, which might necessitate imposing waiting times for some of these requests. Furthermore, it may be necessary to decline certain requests if no conflict-free allocation is possible. The performance of the infrastructure can be assessed by the estimated number of requests that are pending, referred to as the *estimated queue length* L . Although the method we propose is designed to model the allocation of scheduling requests, for clarity, we often refer to these requests as *trains*.

A *railway junction* describes the infrastructure that connects railway lines to more than two different destinations. It usually consists of entry- and exit-signals from and to every direction, as well as one or multiple switches to set the path for the different routes between origins and destinations. In comparison with a *railway station*, trains are not allowed to have scheduled stops in a railway junction, they only stop if the signalling system indicates occupied infrastructure in their planned route.

However, in the remainder of this work, a simple example of a double-track railway junction with four different routes will be considered. These routes can be distinguished regarding their direction, A-B (r_1) and A-C (r_2) start in A towards B or C, while B-A (r_3) and C-A (r_4) end in A, coming from B or C. This exemplary infrastructure is depicted in Fig. 1.

Since no use of turnout tracks is required for the two routes between A and B, we refer to the line between A and B as the *main line* and the line between A and C as the *branch line*. The proportion of the main line in the total traffic volume is then given by

$$p_{main} = \frac{n_{main}}{n_{total}}, \quad (1)$$

where it depends on the number of trains on the main line (n_{main}) and the total number of trains $n_{total} = n_{main} + n_{branch}$ on main and branch line (n_{branch}).

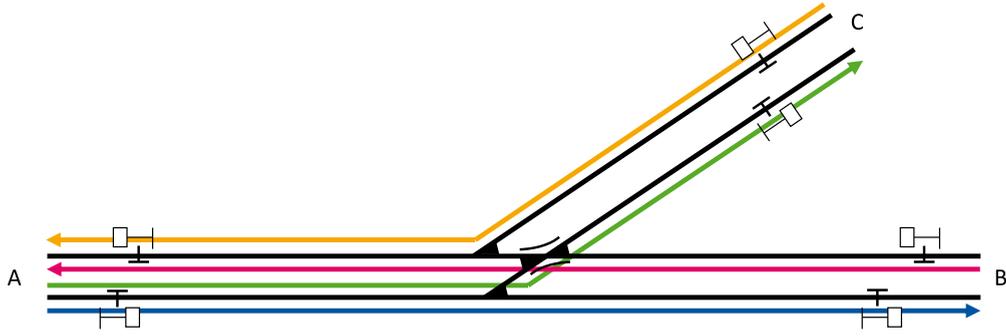


Fig. 1. Exemplary infrastructure of a railway junction.

We express the railway junction as a tuple $J = (R, C)$ of a set R of k routes and a *conflict matrix* $C \in \{0, 1\}^{k \times k}$, describing whether two routes $r, r' \in R$ can be used at the same time ($C_{r,r'} = 0$), or are *conflicting* ($C_{r,r'} = 1$).

In addition to the infrastructure, performance can depend on the used train types T . We therefore formulate an occupation request type $o = (r, t) \in O \subset R \times T$ as a combination of a route $r \in R$ and a train $t \in T$.

For some examples, it might be suitable to give the distribution of a given total number n_{total} of requests to the number of requests of same type $o = (r, t)$. This can be formalized by a function

$$\theta : \mathbb{R} \rightarrow \mathbb{R}^{|O|} : n_{\text{total}} \mapsto \theta(n_{\text{total}}), \tag{2}$$

yielding the number of requests $n_{r,t} = \theta_{r,t}(n_{\text{total}})$ for every occupation request type $o = (r, t)$.

Determining the timetable capacity of a railway junction involves the definition and calculation of multiple parameters, some of which can already serve as an indicator about the quality of the planned junction.

Table 2 gives a list of the used notations.

Performance analysis of railway infrastructure is typically conducted with respect to a fixed *time horizon* t_U , which defines the duration of the investigation period in minutes. The *arrival rate* $\lambda_r = \frac{n_r}{t_U}$ represents the average number of trains per route that request service on route r per minute.

The performance of railway infrastructure is significantly influenced by the service times planned for train usage. For this purpose, the *minimum headway times* $h_{(r_i,t_i),(r_j,t_j)}$ describe the minimum time interval required for the route-train combination (r_j, t_j) to initiate service after the service of the preceding route-train combination (r_i, t_i) has commenced. These headway times are dependent not only on the infrastructure and the specific routes that the trains t_i and t_j are scheduled to traverse, but also on certain rolling stock parameters, such as acceleration and braking behaviour, as well as train length.

For comprehensive details on obtaining blocking and minimum headway times, readers are referred to Hansen and Pachtl (2014). In this paper, it is assumed that the minimum headway times are provided; typically, these would have been calculated using a microscopic tool prior to conducting a detailed infrastructure analysis. However, microscopic data might not be available for long planning horizons, and assumptions or other measures to estimate blocking times might be necessary.

Let $n_{r,t}$ denote the number of trains on route r of train type t . For a given pair r, r' of conflicting routes ($C_{r,r'} = 1$), the minimum headway time $h_{(r,t),(r',t')}$ of all possible sequences $(r, t), (r', t')$ of train-route combinations for this route pair r, r' , can be weighted with the total number of possible sequences for r and r' , $n_{r,r'} = \sum_t \sum_{t'} n_{r,t} \cdot n_{r',t'}$, to obtain the average minimum headway time

$$h_{r,r'} = \sum_t \sum_{t'} n_{r,t} \cdot n_{r',t'} \cdot h_{(r,t),(r',t')} \cdot \frac{1}{n_{r,r'}} \tag{3}$$

between all train pairs t, t' on route r and r' respectively. By further weighting the average minimum headway times $h_{r,r'}$ for each conflicting route pair with the total number of trains $n_{r'}$ per conflicting route r' , the average *service time* for a route r can be calculated as:

$$b_r = \sum_{\substack{r' \\ C_{r,r'}=1}} \frac{n_{r'}}{n_{r,\text{conflict}}} \cdot h_{r,r'}, \tag{4}$$

where $\frac{n_{r'}}{n_{r,\text{conflict}}}$ describes the probability of the pair (r, r') , i.e. a sequence describing any train on the route $r' \in R$ following any train on the conflicting route r . Here, the total number of conflicting trains for route r is given by

$$n_{r,\text{conflict}} = \sum_{\substack{r' \\ C_{r,r'}=1}} n_{r'}. \tag{5}$$

Using this service time b_r for each route, the *service rate* $\mu_r = \frac{1}{b_r}$ describes the average number of trains that can be serviced on route r per minute.

Table 2
List of notations.

R	Set of routes
r_i	Route i
$C \in \{0, 1\}^{k \times k}$	Conflict matrix
$J = (R, C)$	Junction infrastructure
t_U	Time horizon
T	Set of trains
O	Set of occupation request types
θ	Request distribution
λ	Arrival rate
μ	Service rate
ρ	Occupancy rate
v_A	Variation coefficient arrival process
v_S	Variation coefficient service process
S	Set of states
q_i	Number of requests in queue of route r_i
s_i	Status of Service of route r_i
$p_{A,i}$	Phase of arrival process on route r_i
$p_{S,i}$	Phase of service process on route r_i
M	Maximum rate
$MC = (S, Q_{\text{rates}}, t_{\text{init}}, AP, \mathcal{L})$	Continuous-Time Markov Chain
Q_{rates}	Transition rate function
(u, v)	Transition from state u to state v
L_r	Expected queue length of route r
m	number of waiting slots
L_{limit}	Threshold value for sufficient quality
p_{pt}	Share of passenger trains
p_{main}	Share of main line traffic
n_{total}	Total number of trains
n_{max}	Timetable capacity
p_{suburban}	Share of suburban trains
$p_{\text{regional freight}}$	Share of regional freight trains
$n_{(r,t)}$	Train numbers on route r and for traffic type t
$h_{i,j}$	Minimum headway time of the sequence train j after train i
b	Service time
n_r	Number of trains on route r

The occupation ratio ρ_r of a route can be determined by comparing the arrival- and service rates

$$\rho_r = \frac{\lambda_r}{\mu_r}. \quad (6)$$

Furthermore, to determine the timetable capacity of a railway junction, the expected queue-length L_r can be calculated for every route $r \in R$ utilizing the arrival rate λ_r and service rate μ_r .

To enforce a sufficient timetable quality, a limit to the expected queue length $L_{\text{limit},r}$ can be set by the infrastructure manager, see Section 4.2 for a detailed description. This limit depends on the share of passenger trains and can therefore vary between every route, depending on its operating programme. However, the expected queue-length of every route can be compared to the route-specific threshold value to determine the quality factor

$$qf_r = \frac{L_r}{L_{\text{limit},r}} \quad (7)$$

of this specific route. Therefore, the bottleneck in the analysed railway junction infrastructure layout can be identified by comparing the quality factors of the different routes.

The timetable capacity of a given railway junction infrastructure J can be formulated as the maximal number of train requests n_{max} that can be scheduled, under consideration of a fixed distribution θ to routes, while respecting given minimum headway times $h_{(r_i,t_i),(r_j,t_j)}$ and not exceeding a set limit $L_{\text{limit},r}$ of the expected queue-length per route.

A method to calculate the route-based queue length has been introduced in Emunds and Nießen (2024a). It corresponds to modelling the railway junction as a Continuous-Time Markov Chain (CTMC) and calculating the state probabilities utilizing probabilistic Model-Checking.

Since this model utilizes Markov Chains to model the arrival and service processes, it assumes exponentially distributed inter-arrival and service times. Therefore, the arrival and service processes have been supposed to be completely random, expressed by their coefficient of variation of $v_A = v_B = 1$. In correspondence with the Kendall notation (see Kendall, 1953), we call this fully exponential setting (M/M).

However, the coefficient of variation can vary between different settings and general planning rules specify the coefficient of variation for the arrival process to $v_A = 0.8$ and for the service process to $v_B = 0.3$. Hence, more general distributions need to

be analysed in a queueing system for railway infrastructure. We call this general independent setting (GI/GI), also following the Kendall notation.

The next section extends the Continuous-Time Markov Chain model using phase-type distributions and further explains quality thresholds.

4. Methods

In this section, the methods for calculating the timetable capacity of a railway junction are presented. First, the Continuous-Time Markov Chain model for exponential arrival and service processes is described. Afterwards, the threshold values that ensure sufficient infrastructure quality are defined in Section 4.2. Next, the approximation methods used to adapt exponential distributions to general independent settings are discussed in Section 4.3. The novel model, which incorporates phase-type distributions for the arrival and service processes, is introduced in Section 4.4. Finally, the algorithm designed to efficiently compute timetable capacity using queue-length estimations (Section 4.5) is described in Section 4.6.

4.1. A queueing based railway junction CTMC

Following the approach in Emunds and Nießen (2024a), a CTMC model can be obtained by describing the railway junction as a queueing system.

For this, we model the queueing system as a stochastic process $\{X(t)|t \in \tau\}$, where $t \in \tau$ describes a point in continuous Time τ and $X(t)$ corresponds to a random variable with values, called *states*, in a state set S .

This state set S can be used to formalize the stochastic process as a Continuous-Time Markov Chain $MC = (S, Q_{\text{rates}}, t_{\text{init}}, AP, \mathcal{L})$ with a transition rate function $Q_{\text{rates}} : S \times S \rightarrow \mathbb{R}_0^+$, an initial distribution $t_{\text{init}} : S \rightarrow [0, 1]$ with $\sum_{u \in S} t_{\text{init}}(u) = 1$, a set of atomic propositions AP and a labelling function $\mathcal{L} : S \rightarrow 2^{AP}, u \mapsto \mathcal{L}(u) \subset AP$. To obtain the estimated queue length of every route, this information needs to be accessible via the labelling function. Therefore, we define the set of atomic propositions as the set of $|R|$ -Tuples $AP = \{(q_r)_{r \in R} | q_r \in \mathbb{N}^0\}$, which represents all possible combinations of queue-lengths $(q_r)_{r \in R}$ for the different routes. The labelling function \mathcal{L} therefore maps a state $u \in S$ to the tuple $\mathcal{L}(u) = (q_r(s))_{r \in R}$, with $q_r(u)$ as the queue length of route r at state u . Finally, we utilize a simple initial distribution

$$t_{\text{init}}(u) = \begin{cases} 1 & \text{if } u = u_0 \\ 0 & \text{else,} \end{cases} \quad (8)$$

with one starting state $u_0 \in S$, describing an empty system.

In the former contribution (Emunds and Nießen, 2024a), the state set has been described as

$$\hat{S}_{M,M} = \{(q_1, s_1, \dots, q_k, s_k) | q_r \in \{0, \dots, m\}, s_r \in \{0, 1\}\}. \quad (9)$$

Hence, the number of waiting requests $q_r \in \{0, \dots, m\}$, and the state of use of the track $s_r \in \{0, 1\}$ are described for each route r . Regarding the transitions $(u, v) \in S \times S$, three different types have been distinguished: *Arrival* transitions $(\dots, q_i, s_i, \dots) \mapsto (\dots, q_i + 1, s_i, \dots)$ correspond to an arrival to route i ; *Service* transitions $(\dots, q_i, 1, q_{i+1} \dots) \mapsto (\dots, q_i, 0, q_{i+1} \dots)$ correspond to the service on route i ; and *Choice* transitions $u \mapsto v$ correspond to the choice which route to serve next. The transition rate function

$$Q_{M,M}(u, v) = \begin{cases} \lambda_r & \text{if arrival to } r \\ \mu_r & \text{if service at } r \\ M & \text{if choice transition,} \end{cases} \quad (10)$$

maps the transitions (u, v) to their rates. In this model, the choice rate M corresponds to a very high rate, in order to introduce only negligibly small delays to the system. Note that a state might have a single outgoing choice transition, corresponding to the start of the service on a route if the associated queue is not empty and no other conflicting route is being served. We still call this a choice transition, even though it technically does not correspond to a choice between multiple possibilities. It is therefore possible that multiple non conflicting routes might be served in parallel. The number of channels of the modelled queueing system depends on the currently served requests. However, we calculate the estimated queue length L_r for every route and therefore assume a single-channel queue in the approximation formulas introduced below.

To model non-exponential inter-arrival and service time distributions, we extend this formulation with phase-type distributions in Section 4.4. Before, thresholds for the performance determination are introduced in the next section.

4.2. Threshold values

The calculated queue-length L_r for a route can be compared with limit values for the queue-length $L_{\text{limit},r}$ to assess the quality of the provided transport service. For long-term infrastructure planning, the largest German infrastructure manager, DB InfraGO (2022), utilizes the threshold

$$L_{\text{limit},r} = 0.479 \cdot \exp(-1.3 \cdot p_{\text{pt},r}) \quad (11)$$

(Schwanhäußer and Schultze, 1982).

This formula is dependent on the share of passenger trains

$$p_{pt,r} = \frac{\text{number of passenger trains on route } r}{\text{total number of trains on route } r} \quad (12)$$

on route r . The more passenger trains are in the operating programme of this route, the higher the threshold for sufficient timetable quality. Some intuition into this can be obtained by looking at allowances for time deviations in the timetabling process in Germany (DB InfraGO, 2023). While the timetable constructor might only move passenger trains up to 3 min, freight trains may be shifted by up to 30 min - depending on the ordered type.

4.3. Approximation formulas

Different techniques have been developed to approximate the general independent queueing system (GI/GI). In this section, we focus on two well-known approximation formulas in Hertel (1984) (Section 4.3.1) and Kingman (1961) (Section 4.3.2), while the utilizing of phase-type distributions in the Markov Chain to model general independent stochastic processes is presented in Section 4.4.

4.3.1. Hertel approximation

To achieve the coefficients of variation of $v_A = 0.8$ and $v_B = 0.3$, the calculated queue-lengths can be scaled with the approximation formula

$$L_r(M/M) \cdot \frac{1}{\gamma} \approx L_r(GI/GI) \quad (13)$$

from Hertel (1984) (see also Fischer and Hertel, 1990).

For this, the parameters

$$\gamma = \frac{2}{c \cdot v_B^2 + v_A^2} \quad (14)$$

and

$$c = \left(\frac{\rho_r}{s}\right)^{1-v_A^2} \cdot (1 + v_A^2) - v_A^2. \quad (15)$$

are determined in dependence on the occupation ration $\rho_r = \lambda_r/\mu_r$, the amount of channels $s = 1$ and the coefficients of variation v_A, v_B . This has been used in Emunds and Nießen (2024a) to model non-exponential arrival and service time distributions.

4.3.2. Kingman approximation

Another Approximation, that has been widely used for single-channel systems (see f.e. Gudehus, 1976), is the Kingman approximation formula (Kingman, 1961)

$$L_r(M/M) \cdot \left(\frac{v_A^2 + v_B^2}{2}\right) \approx L_r(GI/GI). \quad (16)$$

It gives a good approximation for single-channel scenarios with a high occupation ratio $\rho \rightarrow 1$.

4.4. Phase-type distributions

To use the estimated queue lengths L_r with the Markovian model, they are scaled with one of the approximation formulas in Section 4.3 to facilitate arbitrary coefficients of variation for the interarrival and service distributions.

In this section, we introduce a novel Continuous-Time Markov Chain formulation, enabling a direct modelling of general independent distributions.

For this, phase-type distributions (Cox, 1955), a type of distribution function, whose parameters can be fitted to approximate any other arbitrary probability distribution (Asmussen et al., 2003), can be used. Additionally, they can be represented by a Continuous-Time Markov chain, which facilitates their integration into the described model.

Specifically, the service and arrival processes on railways can be modelled with hypoexponential distributions because their assumed coefficients of variation $v_X < 1$, $X \in \{A, B\}$ are less than 1 (cf. Weik, 2020).

Fig. 2 illustrates two different types of service processes: one with exponentially distributed service times (Fig. 2(a)) and one with hypoexponentially distributed service times (Fig. 2(b)). In the Markov chain for the latter distribution, three intermediate states were introduced, which can be reached through transitions with different rates from the preceding states.

The number k of the minimum required phases to approximate a distribution X with coefficient of variation v_X can be determined using

$$k \geq \left\lceil \frac{1}{v_X^2} \right\rceil, \quad (17)$$

(see David and Larry, 1987).

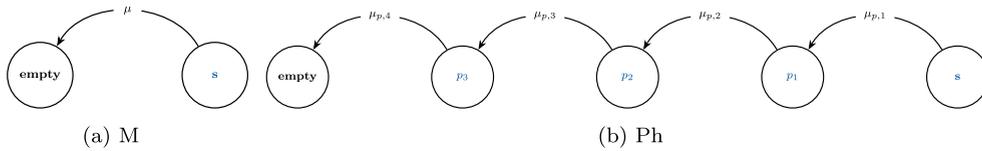


Fig. 2. Examples of the service process with exponential (M) or phase-distributed (Ph) service times.

4.4.1. Hypoexponential parameter fitting

To adapt the hypoexponential distribution \hat{X} to a distribution X , both the number of phases k and the transition rates μ_{p_i} of the corresponding phases can be used. Weik (2020) adapts an approach by Sommereder (2011), which divides the phases into two successive Erlang distributions X_1, X_2 . The k phases are determined using the coefficient of variation according to (17) and divided into $k_{X_1} = \lceil \frac{k}{2} \rceil$ and $k_{X_2} = k - k_{X_1}$ phases.

In a first step, a combined distribution (X_1^*, X_2^*) of two artificial Erlang distributions with k_{X_1}/k_{X_2} phases is fitted to the coefficient of variation v_X . For this, the expected value of X_1^* is fixed to $E(X_1^*) = 1$.

By determining the expected value

$$E(X_2^*) = \frac{k_{X_1} k_{X_2} v_X^2 + \sqrt{k_{X_1} k_{X_2} (v_X^2 (k_{X_1} + k_{X_2}) - 1)}}{k_{X_1} (1 - v_X^2 k_{X_2})} \tag{18}$$

for the artificial Erlang distribution X_2^* , the target coefficient of variation

$$v_X = \frac{\sqrt{Var(X_1^*) + Var(X_2^*)}}{E(X_1^*) + E(X_2^*)} = \frac{\sqrt{1/k_{X_1} + E(X_2^*)^2/k_{X_2}}}{1 + E(X_2^*)} \tag{19}$$

can be approximated together with the artificial Erlang distribution X_1^* .

To correctly estimate the expected value $E(X)$ of the distribution X , the expected values of the artificial Erlang distributions are then adjusted via

$$E(X_1) = \frac{E(X)}{1 + E(X_2^*)} \tag{20}$$

and

$$E(X_2) = \frac{E(X)E(X_2^*)}{1 + E(X_2^*)}. \tag{21}$$

Finally, the transition rate μ_{p_i} between phases $i - 1$ and i can then be determined by

$$\mu_{p_i} = \begin{cases} \frac{k_{X_1}}{E(X_1)} & i \leq k_{X_1} \\ \frac{k_{X_2}}{E(X_2)} & i > k_{X_1} \end{cases} \tag{22}$$

4.4.2. Example

In the example from Fig. 2(b), the following parameters would be used to define a corresponding hypoexponential distribution for a service process with coefficient of variation $v_B = 0.5$ and a mean service time of $\bar{b} = 3$: In total,

$$k = 4 = \frac{1}{0.5^2} \tag{23}$$

phases need to be considered, with $k_{X_1} = k_{X_2} = 2$ states for the two artificial Erlang distributions. Using (18)–(22), it follows that

$$\mu_{p_i} = \begin{cases} 1.33 & 1 \leq i \leq 2 \\ 1.33 & 2 < i \leq 4 \end{cases} \tag{24}$$

for the transition rates between the states of the 4 phases, where the first transition represents the transition from state **s** to state **p**₁, and the last transition represents the transition from state **p**₃ to state **empty**. Note that the states may have additional specifications, such as the number of waiting requests, which are omitted here.

4.4.3. States

To model hypoexponential distributions for the service and/or arrival processes, the current phase of the service and/or arrival processes must be added to the state description for each route.

The state set $\hat{S}_{M,M}$ with exponentially distributed inter-arrival and service times from Eq. (9) can be extended to the state set

$$\hat{S}_{Ph,Ph} = \{(q_1, s_1, p_{A,1}, p_{S,1} \dots, q_k, s_k, p_{A,k}, p_{S,k}) \mid (q_r, s_r, p_{A,r}, p_{S,r}) \in \hat{S}_r\} \tag{25}$$

where with

$$\tilde{S}_r = \{0, \dots, m\} \times \{0, 1\} \times \{0, \dots, k_{A,r} - 1\} \times \{0, \dots, k_{S,r} - 1\} \quad (26)$$

the set of tuples can be specified that describe the state of this track for a route r .

Thus, the number of waiting requests $q_r \in \{0, \dots, m\}$, the state of use of the track $s_r \in \{0, 1\}$, as well as the current phase of arrival $p_{A,r} \in \{0, \dots, k_{A,r} - 1\}$ and service process $p_{S,r} \in \{0, \dots, k_{S,r} - 1\}$ are described for each route r . Here, $k_{A,r}$ and $k_{S,r}$ denote the number of phases in the arrival or service process.

For readability, the notations $q_r(u)$ for the queue-length q_r , $s_r(u)$ for the service status s_r , and $p_{A,r}(u), p_{S,r}(u)$ for the phase of the arrival/service process $p_{A,r}, p_{S,r}$ of route r at state u are additionally used in the following.

Furthermore, the states-space \hat{S}_{p_h, p_h} can be restricted to

$$S_{p_h, p_h} = \left\{ u \in \hat{S} \mid \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k (C_{i,j} \cdot s_i(u) \cdot s_j(j)) = 0 \right\} \quad (27)$$

by using the conflict matrix C to exclude states $u^* \in \hat{S}_{p_h, p_h} \setminus S_{p_h, p_h}$, that cannot be visited because of conflicting ($C_{i,j} = 1$) routes, being serviced at the same time $s_i(u^*) = s_i(u^*) = 1$.

By formulating the transition rate function Q_{rates} , the railway junction can be modelled as a Continuous-Time Markov chain $MC = (S_{p_h, p_h}, Q_{\text{rates}}, t_{\text{init}}, AP, \mathcal{L})$. To map the hypoexponential distribution in the arrival or service process, the arrival or service transitions in Emunds and Nießen (2024a) are replaced by transitions between individual phases in the respective process.

4.4.4. Transitions

The three different transition types from the model in Section 4.1 can also be formulated on the set of states $S = S_{p_h, p_h}$.

The **arrival** process to a route r has been described by a transition $t = (u, v)$ with $q_r(v) = q_r(u) + 1$ in the M/M model. However, for the use of a phasetype distribution with a total of $k_{A,r}$ phases, a change in rates after $k_{A,r}^*$ phases and the two rates $\lambda_{r,a}$ and $\lambda_{r,b}$, multiple transitions (u, v) need to be formulated for the arrival of a train. They can be distinguished according to the arrival phase $p_{A,r}(u)$ from the start state of a transition. For $p_{A,r}(u) \leq k_{A,r}^*$, transitions from u to v with

$$p_{A,r}(v) = p_{A,r}(u) + 1, \quad (28)$$

are applied with a rate of $Q_{p_h, p_h}(u, v) = \lambda_{r,a}$. Furthermore, for $k_{A,r}^* \leq p_{A,r}(u) \leq k_{A,r}$, transitions of

$$p_{A,r}(v) = p_{A,r}(u) + 1, \quad (29)$$

with rate $Q_{p_h, p_h}(u, v) = \lambda_{r,b}$ are added. Lastly, transitions starting the service or adding the train to the queue are formulated for $p_{A,r}(u) = k_{A,r}$. When no conflicting routes are serviced in the start state u , the service of route r can start in state v

$$s_r(v) = 1, \quad (30)$$

otherwise a train is added to the queue

$$q_r(v) = q_r(u) + 1 \quad (31)$$

of the route. Both possible transitions reset the phase of the arrival process to

$$p_{A,r}(v) = 1 \quad (32)$$

in the end state v and use a rate of $Q_{p_h, p_h}(u, v) = \lambda_{r,b}$.

Service transitions between states u, v , with $s_r = 1$, are similarly modelled. Let $k_{S,r}$ be the number of phases in the service process with a change from rate $\mu_{r,a}$ to $\mu_{r,b}$ after $k_{S,r}^*$ phases. Then, for $p_{S,r}(u) \leq k_{S,r}^*$ transitions

$$p_{S,r}(v) = p_{S,r}(u) + 1, \quad (33)$$

with a rate of $Q_{p_h, p_h}(u, v) = \mu_{r,a}$ are used. For $k_{S,r}^* \leq p_{S,r}(u) < k_{S,r}$, the rate changes to $Q_{p_h, p_h}(u, v) = \mu_{r,b}$ and the same property (33) for u and v . Finally, if $p_{S,r}(u) = k_{S,r}$, the service process is terminated by resetting

$$p_{S,r}(v) = 1. \quad (34)$$

The last type, **choice** transitions, are used to model the initiation of service for a route when multiple routes $r \in R$ can be serviced due to the absence of conflicting routes being active in the initial state u . For each such possible route r^* , a transition (u, v) with an artificial rate $Q_{p_h, p_h}(u, v) = M$ is introduced, ensuring that

$$s_{r^*}(v) = 1 \quad (35)$$

and

$$q_{r^*}(v) = q_{r^*}(u) - 1 \quad (36)$$

are satisfied.

Ideally, these choice transitions would not introduce any additional time into the system. Therefore, the rate M should be selected to be sufficiently high, such that the expected induced time $1/M$ remains negligibly small. For all models presented in this paper, a rate of $M = 600$ has been selected, corresponding to a delay of $1/600$ minutes, which is considered to provide an adequately precise approximation (see also [Emunds and Nießen, 2024a](#)). Because all choice transitions occur at the same rate from any given state, there are no priorities assigned to different routes or requests.

Additionally, transitions from other processes must also be incorporated into the introduced intermediate states. Consequently, the arrival of a request on a route i' can be modelled during phase j of the service process for another route i . A full description of a model in the PRISM modelling language can be found in the online repository ([Emunds and Nießen, 2024b](#)).

The modelling approach introduced here, which employs hypoexponential distributions for the arrival and/or service processes, is evaluated for accuracy in Section 5.

4.5. Queue-length analysis

The introduced Continuous-Time Markov Chain $MC = (S_{ph}, p_h, T) = (S, T)$ models the process of scheduling train requests on the considered railway junction. In order to assess the performance of this junction, the length L_r of the queue can be determined for all routes r by analysing the probability $p(u)$ of the states $u \in S$ with certain queue-lengths in the stationary distribution, i.e. the probability that the system is in state u in the long run. Hence, the expected length of the queue can be calculated with

$$L_r = \sum_{\substack{u \in S \\ q_r(u) > 0}} p(u) \cdot q_r(u). \quad (37)$$

The same basic concept has already been used by other analytical approaches in the performance determination of railway infrastructure ([Schmitz et al., 2017](#); [Weik and Nießen, 2017](#); [Weik, 2020](#); [Emunds and Nießen, 2024a](#)). Note that this estimation with stationary distributions assumes static rates in the CTMC, and therefore corresponds to a theoretical analysis of static traffic distributions for long time horizons. In such a capacity analysis, the influence of partial timespans with less or more traffic is therefore neglected. While various solution methods to determine the state probabilities $p(u)$ have been used, [Emunds and Nießen \(2024a\)](#) introduced an approach that builds the CTMC in the formal PRISM language ([Parker et al., 2000](#)) and applies probabilistic model-checking ([Hensel et al., 2022](#)) to obtain the expected queue-lengths L_r . The same mechanism is used for this work.

4.6. Capacity determination algorithm

To determine the timetable capacity of a railway junction, multiple n_{total} values must be tested to find the maximum train count n_{max} , where the expected queue lengths L_r do not exceed the limit $L_{\text{limit},r}$ (Section 4.2). With the introduced approach of modelling phase-type distributions directly within the CTMC, the number of states increases significantly (see also Section 5), leading to considerably longer computation times for the expected queue lengths L_r . It is therefore crucial to minimize the number of iterations in which the queue lengths must be recalculated, i.e., the number of tested n_{total} values.

To achieve this, a root finding problem can be formulated. Let the function

$$L_r : N_{\text{total}} \rightarrow \mathbb{R}, n_{\text{total}} \mapsto L_r(n_{\text{total}}) \quad (38)$$

describe the estimation of the expected queue-length $L_r(n_{\text{total}}) = L_r$ for a given train count $n_{\text{total}} \in N_{\text{total}}$ by utilizing probabilistic model-checking on a formulated CTMC.

By comparing it to the limit, the quality factor

$$qf_r(n_{\text{total}}) = \frac{L_r(n_{\text{total}})}{L_{\text{limit}}} \quad (39)$$

of a route r can be obtained. Since the limit has to be adhered to for every route, the optimal $n_{\text{total}}^* = n_{\text{max}}$ fulfils the property

$$qf_{\text{max}}(n_{\text{total}}^*) = \max_{r \in R} qf_r(n_{\text{total}}^*) = 1, \quad (40)$$

such that the maximum quality factor qf_{max} is exactly 1, and therefore the queue length at one route is exactly at the limit for this route.

The function

$$\phi : N_{\text{total}} \rightarrow \mathbb{R} : n_{\text{total}} \mapsto \phi(n_{\text{total}}) = qf_{\text{max}}(n_{\text{total}}) - 1 \quad (41)$$

can now be formulated, whose root $\phi(n_{\text{total}}^*) = 0$ describes the timetable capacity $n_{\text{total}}^* = n_{\text{max}}$ of the railway junction.

To determine the root of ϕ , the algorithm from [Brent \(1973\)](#), as implemented in `scipy` ([Virtanen et al., 2020](#)), was utilized. This algorithm efficiently computes the root of a function $f : \mathbb{R} \rightarrow \mathbb{R}$ within a specified interval $[a, b]$, where f changes its sign, achieving a solution up to a chosen level of accuracy. In the validation implementation (Section 5.3) and the case study (Section 6), a termination criterion was set to ensure both a total error of less than 10^{-3} and a relative error below 10^{-3} . An Algorithm, describing how to obtain the function evaluation $\phi(n_{\text{total}})$ for a given total train count n_{total} is given in [Appendix A](#).

This algorithm is then called consecutively from within Bent's method for varying n_{total} , until the set tolerances are met and n_{max} has been found.

Table 3
Parameters of the studied models.

Arrival process	Service process	Number of phases Arrival	Number of phases Service	v_A	v_S	Number of waiting slots m	Number of states	Number of transitions
M	M	1	1	1	1	5	10 368	63 688
Ph	M	2	1	0.8	1	5	141 108	829 825
M	Ph	1	12	1	0.3	5	623 376	3 664 703
Ph	Ph	2	12	0.8	0.3	5	8 192 448	46 447 056

5. Validation

To analyse the different distribution functions, the following distinguishes between four different applications.

- Exponentially distributed inter-arrival and service times (M/M)
- Phase-type distributed inter-arrival and exponentially distributed service times (PH/M)
- Exponentially distributed inter-arrival and phase-type distributed service times (M/PH)
- Phase-type distributed inter-arrival and service times (PH/PH)

For each distribution of arrival and service processes, a model of the junction from Fig. 1 was created.

In Table 3, the parameters of the different models are presented. Particularly noteworthy is the influence of the hypoexponential distribution on the size of the model. While the number of states for the model with exponentially distributed inter-arrival and service times is still in the order of 10^4 , more than 8 million states must be considered for the model with hypoexponentially distributed inter-arrival and service times.

The solving process has been automated using Python 3.10.9 (Van Rossum and Drake Jr., 1995; Python Software Foundation, 2022) and the probabilistic model-checker Storm (Hensel et al., 2022) with its Python-Interface stormpy (Junges and Volk, 2023). Furthermore, scipy (Virtanen et al., 2020) has been used for an implementation of the Brent's method (see Section 4.6). Used CTMC models (see Section 4.4) have been expressed in the PRISM modelling language (Parker et al., 2000; Kwiatkowska et al., 2011). The used model files can be found in the online repository (Emunds and Nießen, 2024b).

Results are compared to simulations, that have been implemented in Python 3.10.9 (Van Rossum and Drake Jr., 1995; Python Software Foundation, 2022), utilizing SimPy (Team SimPy Revision, 2023) and Ciw (The Ciw library developers, 2024; Palmer et al., 2019). An implementation can be found in the online repository (Emunds and Nießen, 2024b). In the simulation, inter-arrival times and service times are drawn randomly from phase-type distributions fitted according to Section 4.4.1 for every route. This process does not utilize any timetable information. Trains are managed using first-in-first-out (FIFO) queues for every route, and conflicting routes are prevented from starting service simultaneously by ensuring that shared resources are not utilized concurrently. The validity of this approach is verified through conflict analysis and the queue-length per route is monitored at each simulated minute and averaged for comparison purposes.

In the following sections, computation times and the approximation quality of the different experiments are analysed.

5.1. Computation time

All experiments have been conducted on an Intel Xeon Platinum 8160 Processor (2.1 GHz) and with a working memory of 32 GB. Additionally, 100 Simulations with 22 simulated hours each have been performed on the same processor with 4 GB of RAM for every considered train count n_{total} . To ensure the simulated system reaches a steady-state condition, characterized by minimal fluctuations after initial transient behaviours, the first and last hours of the 22-hour period were excluded from consideration.

When examining the computational times necessary to determine the queue lengths in the individual experiments, the size of the state-space becomes a significant factor. Fig. 3 shows the distribution of the respective calculations for each model. Furthermore, Fig. 3 includes the computational time required to serially simulate 2000 hours of traffic, which does not differ significantly between the assumed distributions of inter-arrival and service times. It is evident that the computation time increases substantially with the number of states.

Using the model for exponentially distributed inter-arrival and service times, approximately 0.22 s are required to compute the expected queue length on average. In contrast, computing the expected queue length using the model with arrival and service processes following a hypoexponential distribution requires over 140 seconds. However, all analytical methods can calculate the queue length faster than a simulation of 2000 hours of traffic, which takes a mean time of over 550 seconds. The computation time for the simulation could be significantly reduced by employing parallel simulations.

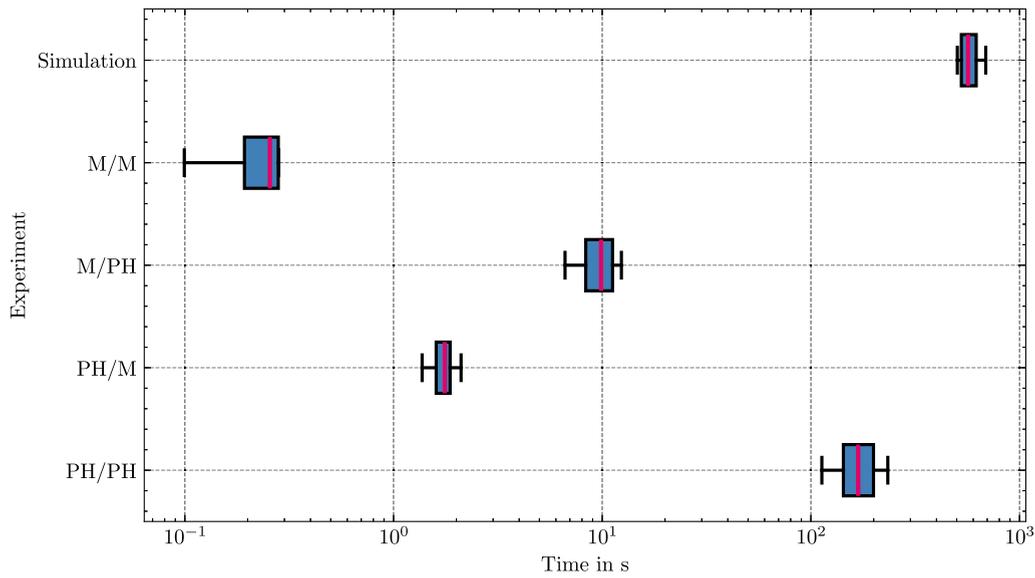


Fig. 3. Computation times of all considered models.

Table 4
Parameters of the approximation settings.

Arrival process	Service process	Approx. formula v_A	Approx. formula v_S	Service rate per route μ_r	Total train count n_{total}	Quality threshold L_{limit}	Share of main traffic p_{main}
M	M	0.8	0.3	0.3	4, ..., 40	0.13	0.5
Ph	M	1	0.3	0.3	4, ..., 40	0.13	0.5
M	Ph	0.8	1	0.3	4, ..., 40	0.13	0.5
Ph	Ph	1	1	0.3	4, ..., 40	0.13	0.5

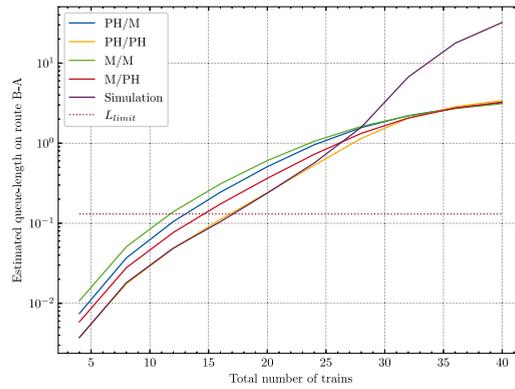
5.2. Queue-length approximation quality

In addition to the computation times of the different models, the approximated queue-lengths can be compared. For this, an exemplary railway junction with 4 routes (see Fig. 1) has been chosen. For this example, we assume passenger traffic only, hence resulting in a quality threshold (see Section 4.2) of $L_{limit} = 0.479 \cdot \exp(-1.3) = 0.13$. The service rates have been assumed to have artificial values of $\mu_r = 0.3$ for every route $r \in R$, corresponding to a mean minimum headway time of $\bar{b} = 1/0.3 \approx 3.33$ min. To highlight the dependencies of the results on the occupancy rate ρ , the total number of trains in the infrastructure per hour $n_{total} = \sum_{r \in R} n_r$ has been varied from $n_{total} \in \{4, 8, \dots, 40\}$ by setting the arrival rate $\lambda_r = n_r/60$ for every route $r \in R$. For a fixed traffic distribution of $p_{main} = 0.5$, the number of trains per route n_r is given by $n_r = n_{total}/4$ for every route. The arrival-rate per route λ_r has hence been set to values between 1/60 and 10/60.

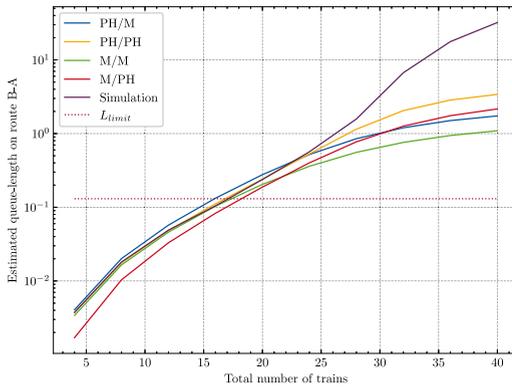
The expected queue length L_r for each route r has been calculated for every experimental distribution setting. Furthermore, for the models with exponential arrival and/or service processes, the approximation formulas by Hertel (Section 4.3.1) and Kingman (Section 4.3.2) have been applied to approximate the expected queue-length of the general independent system $L_r(GI/GI)$. In their formulas, the coefficients of variation v_A and v_B have been set according to the used distribution setting: If a coefficient of variation $v_X \neq 1$ has been used in the model, the corresponding coefficient of variation has been set to $v_X = 1$ in the approximation formula. Otherwise a coefficient of variation of $v_A = 0.8$ has been used for the arrival process and of $v_S = 0.3$ for the service process. Table 4 lists the values in the approximation formulas.

To analyse the approximation quality of a single route, a fixed traffic distribution of $p_{main} = 0.5$ has been set for varying n_{total} . Selecting route r_3 from B to A, Fig. 4 shows the estimated values for the queue-length at this route and compares them to the mean of the simulation. The values have been calculated with the specified model and either not scaled at all (Fig. 4(a)) or approximated with the Hertel (Fig. 4(b)) or Kingman (Fig. 4(c)) method. In this example, all four routes $r \in R$ are assumed to have the same number of trains $n_r = n_{total}/4$ in an hour.

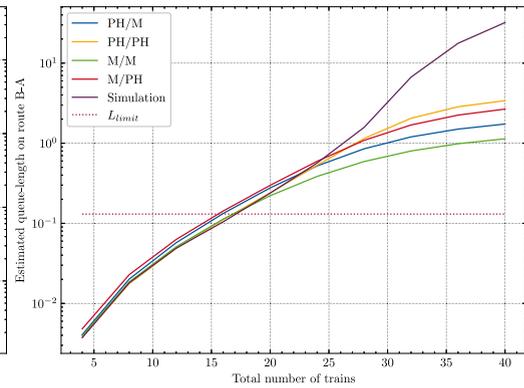
In all three Figs. 4(a)–4(c), the graphs of the PH/PH model and the simulation are almost congruent for a train count n_{total} below 23. At higher train numbers, the difference between the two calculates queue-lengths increases. This can be explained with the limitations in the maximum queue-length, or the number of waiting slots m in the queueing system, which has been set to $m = 5$ in the analytical experiments, but remained unlimited for the simulation setup. The probability of more than 5 trains having to wait for their service increases with an increasing number of trains. Consequently, the analytical models, which omit the state probabilities



(a) Without Scaling Factor



(b) Hertel Approximation



(c) Kingman Approximation

Fig. 4. Comparison of the queue-length estimations at route r_3 for different arrival- and service-process distributions with and without scaling factors.

with more than 5 simultaneous waiting trains, underestimate the total queue-length. However, since the quality threshold is defined for a limit of $L_{limit} = 0.13$, the PH/PH model gives an adequate estimation in the neighbourhood of the limit and can therefore be used without any scaling functions to evaluate the performance of the example railway junction.

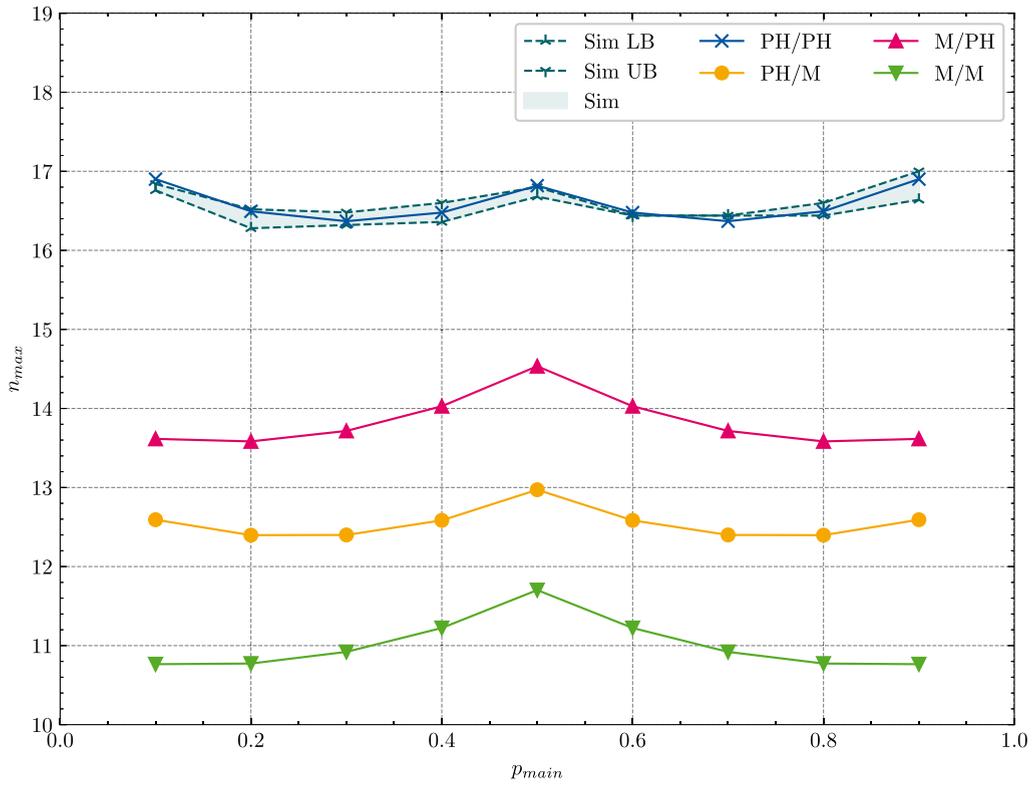
For the graph (Fig. 4(a)) without any applied scaling, it is clearly evident that there are significant differences between the simulation method and the models with exponential arrival and/or service processes. In these three experiments, only 12 – 14.5 trains per hour would be permitted to meet the given quality threshold specified quality threshold, whereas both the simulation method and the PH/PH model would set this limit at approximately 17 trains per hour. Additionally, the M/PH model demonstrates greater alignment with the simulation and PH/PH model compared to the PH/M model, which still exhibits more accuracy than the basic M/M model. These discrepancies can be attributed to variations in the coefficients of variation v_A and v_S from 1 — the M/PH model accurately represents the service process (characterized by a smaller coefficient of variation), while the PH/M model provides a more precise description of the arrival process (with v_A closer to 1).

However, when an approximation method is applied, the M/M model closely aligns with the simulation and the PH/PH method. In the case of the Hertel approximation (Fig. 4(b)), significant discrepancies become evident for a total train count of $n_{total} \geq 17$. Similarly, for the Kingman approximation (Fig. 4(c)), differences become apparent for $n_{total} \geq 19$.

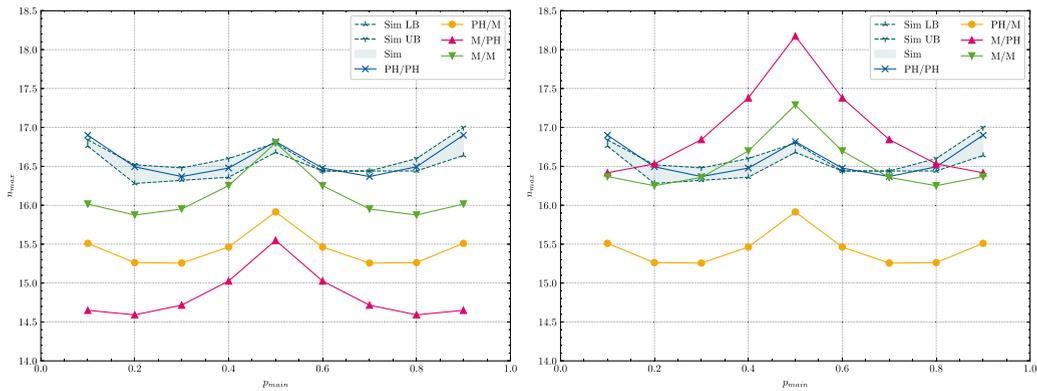
The models for the PH/M and M/PH settings consistently overestimate or underestimate the simulation results within the relevant region where the total train count is $n_{total} \leq 22$. In the case of the Hertel approximation, the PH/M model yields an overestimation while the M/PH model yields an underestimation. Alternatively, for the Kingman approximation, both models produce overestimations, albeit more closely aligned with the simulation results.

5.3. Performance determination quality

While the results for the raw queue-length estimations have been analysed in the previous section, this section focuses on the application of the introduced methods for performance assessment. In certain regions, such as those far from the quality threshold, a small deviation from the correct result in the analysed parameter may not significantly impact the overall approximation quality of performance indicators.



(a) Without Scaling



(b) Kingman Approximation

(c) Hertel Approximation

Fig. 5. Comparison of the performance determination by the share of main line traffic p_{main} for the different methods.

Therefore, we apply the capacity determination algorithm from Section 4.6 for all four analytical settings and both introduced approximation methodologies (see Section 4.3), the parameter settings can be found in Tables 3 and 4.

Table 5 summarizes the results for the different calculation scenarios.

The algorithm from Section 4.6 assumes monotonicity for the queue-length function $L_r : P_{main} \times N_{total} \rightarrow \mathbb{R}$, which cannot be guaranteed for the non-deterministic simulations. Therefore, 100 simulations, each lasting 20 h, are conducted for every combination of $p_{main} \in P_{main} = \{0.1, \dots, 0.9\}$ and $n_{total} \in N_{total} = \{12.00, 12.04, \dots, 19.92, 19.96\}$. From this, the bounds $n_{max, LB}^*$ and $n_{max, UB}^*$ can be identified to limit the optimal train-count n_{max} in the simulation that meets the quality thresholds (see Appendix B).

In Fig. 5, the results of the performance determinations are depicted for the different methods.

When comparing the simulation results to the PH/PH method (Fig. 5(a)), some minor deviations are apparent. For example, the analytical solution exhibits a clear symmetry around the axis $p_{main} = 0.5$, which is only qualitatively reflected in the simulation results. This demonstrates that even the averaged simulation results represent only a subset of all possible outcomes, whereas the analytical method performs the capacity determination deterministically.

Table 5
Selected indicators for the different methods.

Approximation formula	Arrival process	Service process	Mean number of iterations	Mean total computation time	Maximum capacity \hat{n}_{max}	Traffic share at max. capacity
None	M	M	9	2.22 s	11.70	0.5
None	PH	M	9	17.04 s	12.97	0.5
None	M	PH	8.44	78.86 s	14.53	0.5
None	PH	PH	9	1432.20 s	16.90	0.1
Kingman	M	M	8.89	2.35 s	16.80	0.5
Kingman	PH	M	8	15.76 s	15.91	0.5
Kingman	M	PH	7.78	74.65 s	15.55	0.5
Hertel	M	M	9.33	2.31 s	17.29	0.5
Hertel	PH	M	8	15.70 s	15.91	0.5
Hertel	M	PH	9.22	87.33 s	18.17	0.5

Overall, the simulation and PH/PH results are similar in terms of capacity n_{max} and exhibit the same general pattern: one local maximum for homogeneous traffic and two local maxima at the extremes, where nearly all traffic is concentrated on either the main or branch line.

For the raw analytical methods without any applied scaling factor (Fig. 5(a)), the M/PH, PH/M, and M/M methods significantly underestimate the available capacity and fail to exhibit the same general pattern of two maxima at the extremes.

Applying the Kingman approximation (Fig. 5(b)) significantly reduces the gap between the results of the M/M model and the PH/PH model, particularly for the homogeneous traffic scenario at $p_{main} = 0.5$. However, substantial differences in the calculated capacities persist in other traffic distributions. Notably, the order of approximation accuracy changes with the Kingman approximation: the M/M model offers the best approximation, followed by the PH/M model, and lastly the M/PH model.

Using the Hertel formula (Fig. 5(c)), the M/M model also provides the best fit to the PH/PH and simulation results. Although the M/M model still fails to accurately capture the capacity increase for highly heterogeneous traffic distributions, its overall estimates are quite close, with differences of less than 0.5 trains per hour at the local extrema of the PH/PH results. The M/PH model overestimates capacity by up to 1.5 trains per hour in the homogeneous scenario, intersects the PH/PH model near the boundaries, and underestimates performance at both the lowest and highest traffic shares. In contrast, the PH/M model shows the exact same poor approximation quality as with the Kingman approximation, with a clear underestimation of capacity by at least 1 to 1.5 trains per hour.

In summary, the PH/PH model delivers the most accurate results, although it requires the highest computational time. On the other hand, the Hertel approximation enables the M/M model to achieve a good fit for homogeneous traffic scenarios in a fraction of the time. The use of PH/M or M/PH models is generally not advisable, as these models are more inaccurate after applying an approximation formula, while also demanding more computational time. Depending on the application scenario, engineers may opt for the M/M model combined with the Hertel approximation for preliminary and rapid planning stages, while reserving the PH/PH model for more detailed analysis of junction infrastructure in selected long-term projects.

6. Case study

In this Section, we consider a double track railway junction, where a mixed-traffic railway line is divided into a freight line from and to a freight yard and a line for passenger traffic only (see Fig. 6).

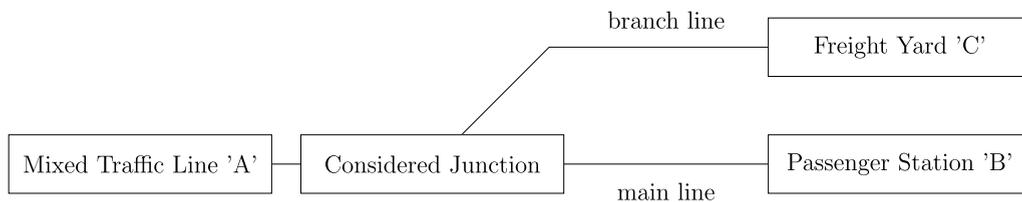


Fig. 6. Traffic scenario of the case study.

Consistent with the introduced notation, the line to the passenger station 'B' is called *main line* and the line to the freight yard 'C' *branch line*. For the railway junction infrastructure itself, we consider the same infrastructure as in Fig. 1, Fig. 7 redraws it for convenience.

6.1. Setup

The traffic on this infrastructure can be described with four different train types: *Suburban* (s) and *regional* (r) trains cover passenger traffic, while cargo trains can be distinguished between *long-distance freight* (lf) and *regional freight* (rf) trains. Depending on their origin and destination, only certain routes are feasible for these types of trains (see Table 6).

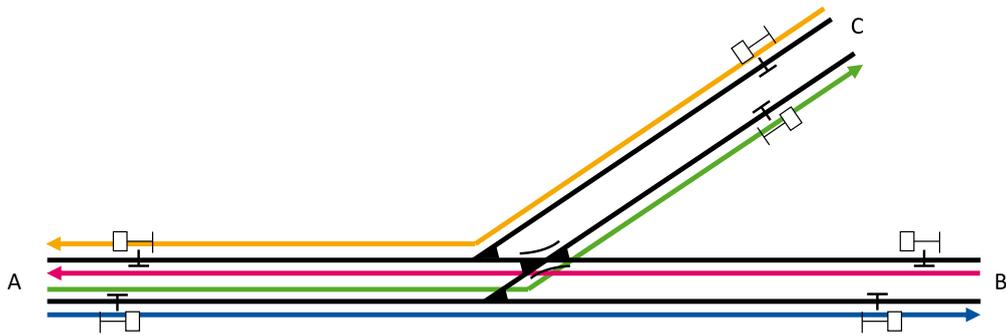


Fig. 7. Infrastructure of the considered railway junction.

Table 6
Traffic in the case study.

Train types	Route	Origin	Destination	Conflicts
suburban train, regional train	r_1	A	B	r_2
long-dist. freight train, regional freight train	r_2	A	C	r_1, r_3
suburban train, regional train	r_3	B	A	r_2, r_4
long-dist. freight train, regional freight train	r_4	C	A	r_3

To assess the junction’s performance, minimum headway times of the different routes have to be considered. Usually, an infrastructure manager would utilize microscopic infrastructure data to calculate minimum headway times with their corresponding toolset. However, for this made-up example, no microscopic data exists, therefore minimum headway times have been estimated in Table 7.

Furthermore, the number of trains per route-train combination has to be specified to calculate the arrival rates in the Markov Chain model from Section 4.4. Together with the service rates, calculated from the weighted minimum headway times, queue-lengths can be estimated for every route $r \in R$. Note that no priorities have been assigned to the different routes and train types, in correspondence with the definition of choice transition rates in Section 4.4.4.

In addition to the total number of trains n_{total} , the volume of traffic per combination of route-train $n_{(r,t)} \in \mathbb{R}$ depends on the share of main line traffic $p_{main} \in [0, 1]$, as well as on the share of suburban trains $p_{suburban} \in [0, 1]$ and regional freight trains $p_{regional\ freight} \in [0, 1]$. For the remainder of this case study, we assume a fixed rolling stock distribution of $p_{suburban} = 0.5$ for passenger and of $p_{regional\ freight} = 0.5$ for freight trains.

Therefore, we can formulate the functions $\theta_{(r,t)} : \mathbb{R} \rightarrow \mathbb{R}, n_{total} \mapsto n_{r,t}$ describing the number of trains per combination for every total number of trains.

To analyse the performance capabilities of this railway junction, the introduced CTMC model for phase-type inter-arrival and service time distributions can be formulated (Section 4.4). Since the model’s size depends on the number of phases in the service process, the variance coefficient for the service process per route $v_{S,r}$ must be calculated.

Table 7
Assumed minimum headway times $h_{i,j}$ in minutes for the considered junction.

$(r_i, t_i) \setminus (r_j, t_j)$	(r_1, s)	(r_1, r)	(r_3, s)	(r_3, r)	(r_2, lf)	(r_2, rf)	(r_4, lf)	(r_4, rf)
(r_1, s)	2.5	5.5			5	5		
(r_1, r)	3	2			3	3		
(r_3, s)			2.5	5.5	1.5	1.5	5	5
(r_3, r)			3	2	1.5	1.5	3	3
(r_2, lf)	3.5	5.5	3	2	3	6		
(r_2, rf)	8	8.5	2.5	2.5	7	4		
(r_4, lf)			3.5	5.5			3	6
(r_4, rf)			8	8.5			7	4

Table 8
Service Rates, Variance Coefficients and size of the CTMC Model (PH/PH) at different p_{main} Values.

p_{main}	μ_1	μ_2	μ_3	μ_4	$v_{S,1}$	$v_{S,2}$	$v_{S,3}$	$v_{S,4}$	Number of states	Number of transitions
0.1	0.25	0.20	0.36	0.19	0.27	0.36	0.52	0.33	5.36×10^6	3.05×10^7
0.2	0.26	0.21	0.36	0.19	0.29	0.40	0.51	0.33	4.32×10^6	2.46×10^7
0.3	0.26	0.21	0.35	0.18	0.31	0.43	0.51	0.34	3.91×10^6	2.23×10^7
0.4	0.27	0.21	0.35	0.18	0.33	0.45	0.50	0.34	3.70×10^6	2.11×10^7
0.5	0.28	0.22	0.34	0.18	0.34	0.47	0.49	0.34	3.44×10^6	1.96×10^7
0.6	0.28	0.22	0.34	0.17	0.36	0.49	0.48	0.34	3.17×10^6	1.81×10^7
0.7	0.29	0.22	0.33	0.17	0.37	0.51	0.47	0.33	3.01×10^6	1.72×10^7
0.8	0.29	0.22	0.32	0.16	0.39	0.52	0.45	0.33	2.95×10^6	1.68×10^7
0.9	0.30	0.22	0.32	0.16	0.40	0.53	0.44	0.32	3.09×10^6	1.76×10^7

Table 8 summarizes the calculated service rates μ_r , variance coefficients $v_{S,r}$ and model size for the analysed traffic distributions of $p_{\text{main}} \in \{0.1, \dots, 0.9\}$. Additionally, the variance coefficient for the arrival process has been set to $v_A = 0.8$. The results of those calculations are described and analysed in the following Section 6.2.

6.2. Timetable capacity results

Using the algorithm described in Section 4.6, the performance capabilities of the railway junction with fixed $p_{\text{main}}, p_{\text{suburban}}, p_{\text{regional}}$ freight traffic shares can be estimated. For this, a lower bound of $n_{\text{total, LB}} = 1$ and an upper bound of $n_{\text{total, UB}} = 40$ has been chosen.

Fig. 8 highlights the convergence of the method for the described example with a main line traffic share of $p_{\text{main}} = 0.5$.

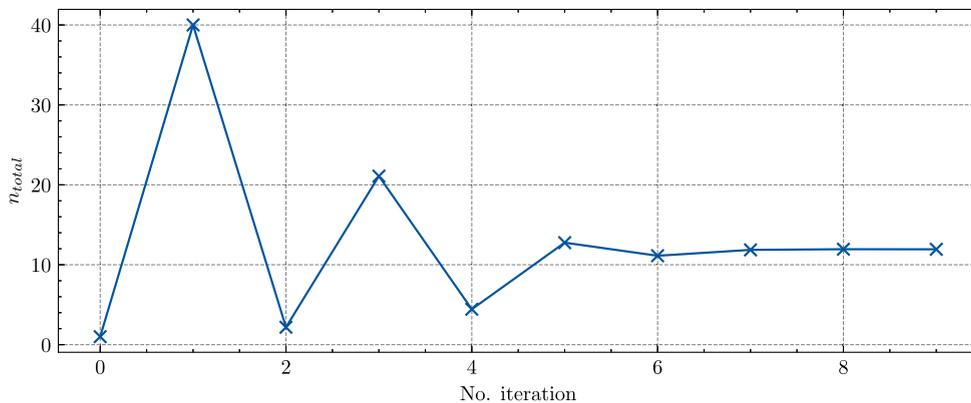


Fig. 8. The number of trains n_{total} in the junction with a traffic split of $p_{\text{main}} = 0.5$ for all iterations.

The figure shows the continuous convergence of the capacity determination algorithm for a growing number of iterations. While the given limits of 1 and 40 are explored in iteration 0 and 1, the algorithm terminates after 10 calls to the model-checking, respectively after 9 iterations at a value of $n_{\text{total}} = n_{\text{max}} = 11.93$, corresponding to the timetable capacity for the specified case.

The timetable capacity n_{max} for other main traffic shares $p_{\text{main}} \in \{0.1, \dots, 0.9\}$ has been calculated similarly. Fig. 9(a) presents the results for both the PH/PH model and the M/M model, with the latter further scaled using the Hertel approximation (see Section 4.3.1). Information about the computation times can be obtained from Appendix C.

In the PH/PH model, the highest timetable capacities are determined for scenarios involving asymmetric traffic, where movement predominantly occurs on either the branching or the main line. The maximum number of trains per hour is calculated as $n_{\text{max}} = 15.78$ for a main track probability of $p_{\text{main}} = 0.1$, and $n_{\text{max}} = 14.47$ trains per hour for $p_{\text{main}} = 0.9$.

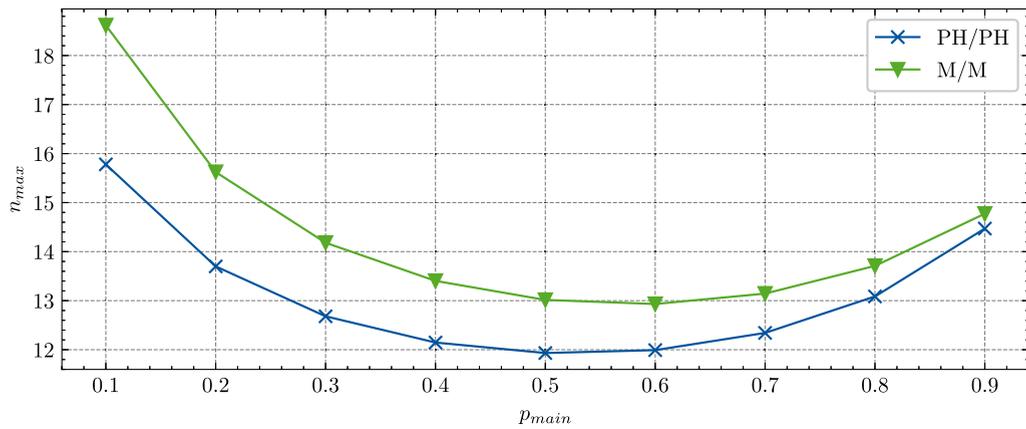
Compared to the PH/PH model, the M/M model significantly overestimates timetable capacity. While the difference is marginal for scenarios with a high main traffic (and therefore passenger traffic) share, it increases to a full train at $p_{\text{main}} = 0.5$ and exceeds 2.5 trains at $p_{\text{main}} = 0.1$.

Furthermore, Fig. 9(b) depicts the quality factor for each route, which can be used to assess the infrastructure's bottleneck. In the given example, route r_3 consistently has a quality factor of 1 for every p_{main} , indicating that this route is the bottleneck for every described traffic scenarios.

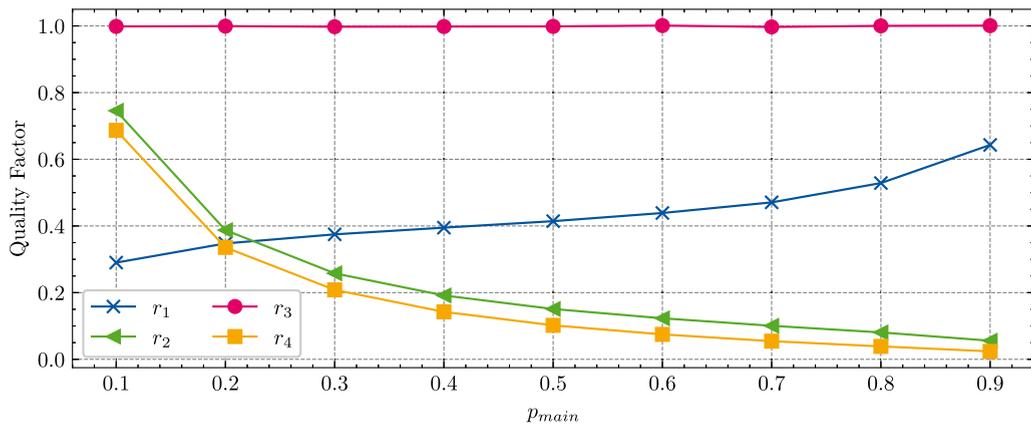
This can be explained by the lower capacity limit for passenger traffic compared to freight traffic, and by the fact that route r_3 conflicts with both branching routes r_2 and r_4 , which have high service times due to their exclusive utilization for freight traffic.

The influence of the traffic share on the other routes' quality is as expected: the freight traffic routes r_2 and r_4 improve their quality factors as traffic on these routes decreases, while route r_1 sees its quality factor increase inversely.

In summary, the analysis reveals that the introduced PH/PH model provides a more accurate estimation of timetable capacity. Its route-based application can be used to identify the bottleneck in the infrastructure for different traffic scenarios.



(a) Timetable Capacity



(b) Quality Factor

Fig. 9. Timetable capacity n_{max} of the two models PH/PH and M/M and quality factor per route for varying p_{main} .

7. Discussion and conclusion

In this work, a novel model for performance estimation of railway junctions has been introduced, utilizing Continuous-Time Markov chains with various distribution functions for inter-arrival and service times. The study examines the impact of using phase-type distributions compared to exponential distributions on the approximation accuracy of the models, with results compared against simulations.

Four different models, each describing distinct distributions for inter-arrival and service times, were analysed for their approximation quality and computation time. Special attention was given to the differences in approximation quality for various traffic distributions between the two intersecting railway lines in the analysed railway junction. Additionally, a case study was conducted to investigate the timetable capacity for a junction that splits a mixed traffic line into freight and passenger lines.

While the introduced model generally produces more accurate results, its computation times are significantly longer compared to models that utilize only exponential distributions. This is due to the use of phase-type distributions, which introduce additional states into the CTMC. As a result, the number of states becomes large even for relatively simple junctions. Consequently, the scalability of the introduced method is highly limited. To apply this approach to more complex railway junctions or even entire railway stations, further research is needed on decomposition techniques and alternative approximation methods. If such techniques allow for efficient sub-calculations of complex infrastructure without adding significant approximation errors, they would enable the use of models with phase-type distributions to benefit from their increased accuracy. However, given the current computational demands, it may initially be more practical to explore more complex systems using purely exponential arrival and service process models and enhancing their results with approximation formulas.

A major parameter influencing both the model size and the timetable capacity of railway junctions is the variation coefficient of the service process. In some chapters of this work (see Section 5), a default value was used, while in the case study, it was calculated

based on minimum headway times. For efficient use of the introduced method, it is essential to obtain an accurate estimation of this value to ensure the generation of reliable and usable results.

To assess the performance of the railway junction, queue lengths computed with CTMC models are compared to thresholds in the capacity determination algorithm (see Section 4.6). Although these thresholds are commonly used in practice, they were originally designed for line capacities. Therefore, new thresholds may need to be developed to appropriately assess the quality of multi-channel systems.

Future work may also explore techniques to incorporate priorities into the formulation. With the current route-based decomposition, priorities between routes could potentially be included by adjusting the rates of choice transitions. To incorporate priorities between different train types while maintaining the route-based decomposition, alternative techniques may be required, such as manipulating service times using priority coefficients, as demonstrated in the works of Schwanhäuser (1978) and Wakob (1981).

The proposed method, despite its longer computation times, allows for the substitution of approximation techniques (Section 4.3) in application scenarios that demand high accuracy. The introduced method demonstrates significantly shorter computation times compared to simulation approaches while producing deterministic results, making it more practical than the numerous simulation iterations needed to achieve sufficiently representative outcomes. However, a parallel implementation of simulation iterations might outperform the computation time of the introduced method.

The proposed model, which explicitly models phase-type distributions within the Continuous-Time Markov chain, can be utilized by infrastructure managers to obtain precise, timetable-independent estimations of the capacity of railway junctions. This enables a reduction in the reliance on computationally expensive simulations in long-term railway planning scenarios.

CRedit authorship contribution statement

Tamme Emunds: Conceptualization, Methodology, Software, Formal Analysis, Writing – original draft. **Nils Nießen:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 2236/2. Computational Experiments were performed with computing resources granted by RWTH Aachen University under projects rwth1413 and rwth1635.

Appendix A. Capacity determination algorithm

In Section 4.6, the root-finding problem has been introduced, which determines the timetable capacity of a railway junction by comparing the estimated queue-length per route L_r to their respective maximum thresholds $L_{r,limit}$, while varying the total train count n_{total} .

Since some parameters of the railway junction, such as the arrival rates λ_r , service rates μ_r and variation coefficients $v_{A,r}, v_{S,r}$ may depend on the total train count n_{total} , the CTMC model has to be formulated from inside the function call of the root-finding method.

Algorithm 1 therefore describes the complete inner function, that computes all relevant parameters before the calculation of the quality factor and the result $\phi(n_{total})$ for a chosen (n_{total}) in an interval $[n_{total, LB}, n_{total, UB}]$.

In the initial `foreach` loop, the number of trains per route, denoted by n_r , along with other computable parameters, is established. Subsequently, the frequencies of train sequences $p_{i,j}$ are determined, which allows for the setting of the remaining parameters in the following `foreach` loop. The next phase involves formulating the Continuous-Time Markov Chain (CTMC) and calculating the queue lengths L_r . Finally, the difference between the maximum quality factor and the optimal ratio of 1 is computed and returned.

With this information, Brent's method determines whether the set tolerances are met and the timetable capacity n_{max} has been found. Otherwise, the interval $[a, b]$ is shortened, selecting a new bound n_{total} replacing a or b , where ϕ needs to be evaluated. An example of the convergence of this method can be found in Section 6.2.

Algorithm 1: Inner Function for Brent's Method

Data: Railway Junction $J = (R, C)$, Minimum Headway Times $h_{(r_i,t_i),(r_j,t_j)}$, Total Number of Trains n_{total} , Request Distribution θ

Result: $\phi(n_{\text{total}})$

- 1 **foreach** route $r \in R$ **do**
- 2 Determine number of trains on the route n_r ;
- 3 Determine arrival rate $\lambda_r = 1/n_r$;
- 4 Determine share of passenger trains p_{pt} ;
- 5 Determine $L_{\text{limit}, r}$;
- 6 **end**
- 7 Determine frequencies of train sequences $p_{i,j}$;
- 8 **foreach** route $r \in R$ **do**
- 9 Determine service times b_r ;
- 10 Determine service rates $\mu_r = 1/b_r$;
- 11 Determine variation coefficients $v_{S,r}$;
- 12 Determine phase-type parameters $k_{A,r}, k_{S,r}, \lambda_{r,a}, \lambda_{r,b}, \mu_{r,a}, \mu_{r,b}$;
- 13 **end**
- 14 Build the CTMC $MC = (S, T)$;
- 15 Get $L_r(n_{\text{total}})$ with probabilistic model checking;
- 16 Determine $qf_{\text{max}}(n_{\text{total}})$;
- 17 Determine $\phi(n_{\text{total}})$;
- 18 **return** $\phi(n_{\text{total}})$

Appendix B. Performance determination for simulations

In Section 5.3, we conduct 100 simulations, each lasting 20 h, for $p_{\text{main}} \in P_{\text{main}} = \{0.1, \dots, 0.9\}$ and $n_{\text{total}} \in N_{\text{total}} = \{12.00, 12.04, \dots, 19.92, 19.96\}$.

This extensive computation was necessary because the algorithm from Section 4.6 assumes monotonicity for the function $L_r : N_{\text{total}} \rightarrow \mathbb{R}, p_{\text{main}} \mapsto L_r(n_{\text{total}})$.

However, since the simulation is not deterministic, even taking an average of 100 simulation runs does not guarantee a monotonic function approximation. It is therefore possible that there exist $n_{\text{total},1}, n_{\text{total},2} \in N_{\text{total}}$ for a fixed $p'_{\text{main}} \in P_{\text{main}}$, with:

$$L_r(n_{\text{total},1}) \geq L_r(n_{\text{total},2}) \quad (\text{B.1})$$

and $n_{\text{total},1} < n_{\text{total},2}$.

Therefore, it might not be possible to find a single value $n_{\text{max}}^* \in N_{\text{total}}$ that meets the quality threshold requirement:

$$L_r(n_{\text{max}}^*) \leq L_{\text{limit}}, \quad (\text{B.2})$$

where for all larger $n' \in N_{\text{total}}$, with $n_{\text{max}}^* < n'$, the limit

$$L_r(n') > L_{\text{limit}} \quad (\text{B.3})$$

is exceeded **and** for all smaller $n'' \in N_{\text{total}}$, with $n_{\text{max}}^* > n''$, the limit

$$L_r(n'') \leq L_{\text{limit}} \quad (\text{B.4})$$

is not violated.

Hence, we introduce the two bounds to the capacity $n_{\text{max, LB}}^*, n_{\text{max, UB}}^* \in N_{\text{total}}$, where the lower bound $n_{\text{max, LB}}^*$ meets the condition (B.4): For all smaller $n'' \in N_{\text{total}}$, with $n_{\text{max, LB}}^* > n''$, the limit

$$L_r(n'') \leq L_{\text{limit}} \quad (\text{B.5})$$

is not violated.

Similarly, $n_{\text{max, UB}}^*$ meets the condition (B.3): For all larger $n' \in N_{\text{total}}$, with $n_{\text{max, UB}}^* < n'$, the limit

$$L_r(n') > L_{\text{limit}} \quad (\text{B.6})$$

is exceeded.

These bounds are chosen to be as tight as possible, so that $n_{\text{max, LB}}^*$ represents the highest value and $n_{\text{max, UB}}^*$ represents the lowest value that satisfy the corresponding conditions.

Table C.9
Performance Metrics for Different Model Types at Various p_{main} Values.

Model type	p_{main}	# Iterations	# Calls to model-checking	Mean model-checking time (s)	Total computation time (s)
M/M	0.1	10	11	0.13	2.89
M/M	0.2	9	10	0.13	2.75
M/M	0.3	9	10	0.12	2.63
M/M	0.4	9	10	0.22	9.75
M/M	0.5	9	10	0.14	2.81
M/M	0.6	9	10	0.13	2.19
M/M	0.7	10	11	0.17	4.15
M/M	0.8	10	11	0.21	9.58
M/M	0.9	10	11	0.13	2.86
PH/PH	0.1	10	11	93.09	1025.04
PH/PH	0.2	9	10	65.20	656.58
PH/PH	0.3	8	9	67.71	610.58
PH/PH	0.4	9	10	60.55	606.76
PH/PH	0.5	9	10	48.74	488.77
PH/PH	0.6	9	10	61.68	617.89
PH/PH	0.7	9	10	43.84	439.48
PH/PH	0.8	9	10	40.65	407.56
PH/PH	0.9	8	9	50.39	454.54

Appendix C. Computation times case study

Table C.9 lists the computation times for both models and different p_{main} values from the case study in Section 6.

Data availability

The processed data required to reproduce the above findings are available to download from [Emunds and Nießen \(2024b\)](#).

References

- Abril, M., Barber, F., Ingolotti, L., Salido, M.A., Tormos, P., Lova, A., 2008. An assessment of railway capacity. *Transp. Res. Part E: Logist. Transp. Rev.* 44 (5), 774–806. <http://dx.doi.org/10.1016/j.tre.2007.04.001>.
- Asmussen, S., Asmussen, S., Asmussen, S., 2003. *Applied Probability and Queues*, vol. 2, Springer.
- Bešinović, N., Goverde, R.M.P., 2018. Capacity assessment in railway networks. In: Borndörfer, R., Klug, T., Lamorgese, L., Mannino, C., Reuther, M., Schlechte, T. (Eds.), *Handbook of Optimization in the Railway Industry*. In: *International Series in Operations Research & Management Science*, vol. 268, Springer International Publishing, Cham, pp. 25–45. http://dx.doi.org/10.1007/978-3-319-72153-8_2.
- Brent, R., 1973. *Algorithms for Minimization Without Derivatives*. Prentice-Hall, Englewood Cliffs NJ.
- Burdett, R.L., 2016. Optimisation models for expanding a railway's theoretical capacity. *European J. Oper. Res.* 251 (3), 783–797. <http://dx.doi.org/10.1016/j.ejor.2015.12.033>.
- Burdett, R.L., Kozan, E., 2006. Techniques for absolute capacity determination in railways. *Transp. Res. Part B: Methodol.* 40 (8), 616–632. <http://dx.doi.org/10.1016/j.trb.2005.09.004>.
- Cacchiani, V., Furini, F., Kidd, M.P., 2016. Approaches to a real-world train timetabling problem in a railway node. *Omega* 58, 97–110. <http://dx.doi.org/10.1016/j.omega.2015.04.006>.
- Cacchiani, V., Toth, P., 2012. Nominal and robust train timetabling problems. *European J. Oper. Res.* 219 (3), 727–737. <http://dx.doi.org/10.1016/j.ejor.2011.11.003>.
- Cox, D.R., 1955. A use of complex probabilities in the theory of stochastic processes. In: *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 51, (2), Cambridge University Press, pp. 313–319.
- D'Acerno, L., Botte, M., Pignatiello, G., 2019. A simulation-based approach for estimating railway capacity. *Int. J. Transp. Dev. Integr.* 3 (3), 232–244. <http://dx.doi.org/10.2495/TDI-V3-N3-232-244>.
- David, A., Larry, S., 1987. The least variable phase type distribution is Erlang. *Stoch. Models* 3 (3), 467–473.
- DB InfraGO, 2022. *Richtlinie Fahrwegkapazität*.
- DB InfraGO, 2023. *Nutzungsbedingungen Netz der DB Netz AG*.
- De Kort, A.F., Heidergott, B., Ayhan, H., 2003. A probabilistic (max, +) approach for determining railway infrastructure capacity. *European J. Oper. Res.* 148 (3), 644–661. [http://dx.doi.org/10.1016/S0377-2217\(02\)00467-8](http://dx.doi.org/10.1016/S0377-2217(02)00467-8).
- Emunds, T., Nießen, N., 2024a. Evaluating railway junction infrastructure: A queueing-based, timetable-independent analysis. *Transp. Res. Part C: Emerg. Technol.* 165, 104704. <http://dx.doi.org/10.1016/j.trc.2024.104704>.
- Emunds, T., Nießen, N., 2024b. Research data and code: Utilizing phase-type distributions for queueing-based railway junction performance determination. <http://dx.doi.org/10.5281/ZENODO.14281662>.
- Fischer, K., Hertel, G., 1990. *Bedienungsprozesse im Transportwesen: Grundlagen und Anwendungen der Bedienungstheorie*. Transpress-Verlag.
- Goverde, R.M., 2007. Railway timetable stability analysis using max-plus system theory. *Transp. Res. Part B: Methodol.* 41 (2), 179–201. <http://dx.doi.org/10.1016/j.trb.2006.02.003>.
- Goverde, R.M., Corman, F., D'Ariano, A., 2013. Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions. *J. Rail Transp. Plan. Manag.* 3 (3), 78–94. <http://dx.doi.org/10.1016/j.jrtpm.2013.12.001>.
- Gudehus, T., 1976. Staueffekte vor Transportknoten. *Z. Für Oper. Res.* 20 (6), B207–B252. <http://dx.doi.org/10.1007/BF01918395>.
- Hansen, I., Pachel, J., 2014. *Railway Timetabling and Operations: Analysis, Modelling, Optimisation, Simulation, Performance, Evaluation*. Eurail Press.

- Harrod, S., 2009. Capacity factors of a mixed speed railway network. *Transp. Res. Part E: Logist. Transp. Rev.* 45 (5), 830–841. <http://dx.doi.org/10.1016/j.trc.2009.03.004>.
- Hensel, C., Junges, S., Katoen, J.-P., Quatmann, T., Volk, M., 2022. The probabilistic model checker storm. *Int. J. Softw. Tools Technol. Transf.* 24 (4), 589–610. <http://dx.doi.org/10.1007/s10009-021-00633-z>.
- Hertel, G., 1984. Exakte Lösung zur Berechnung der Wartegleiszahl vor im Einrichtungsbetrieb befahrenen Streckengleisen bei Nicht-Poisson-Ankünften (G/M/1-Wartesystem). *Wiss. Z. der Hochsch. Für Verkehrswesen* 31, 195–205.
- Jensen, L.W., Schmidt, M., Nielsen, O.A., 2020. Determination of infrastructure capacity in railway networks without the need for a fixed timetable. *Transp. Res. Part C: Emerg. Technol.* 119, 102751. <http://dx.doi.org/10.1016/j.trc.2020.102751>.
- Junges, S., Volk, M., 2023. Stormpy - Python bindings for storm: Version 1.7.0. URL: <https://moves-rwth.github.io/stormpy/index.html>.
- Kendall, D.G., 1953. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *Ann. Math. Stat.* 338–354.
- Kingman, J.F.C., 1961. The single server queue in heavy traffic. *Math. Proc. Cambridge Philos. Soc.* 57 (4), 902–904. <http://dx.doi.org/10.1017/S0305004100036094>.
- Kwiatkowska, M., Norman, G., Parker, D., 2011. PRISM 4.0: Verification of probabilistic real-time systems. In: Gopalakrishnan, G., Qadeer, S. (Eds.), *Proc. 23rd International Conference on Computer Aided Verification. CAV'11*, In: LNCS, vol. 6806, Springer, pp. 585–591.
- Liao, Z., Li, H., Miao, J., Corman, F., 2021. Railway capacity estimation considering vehicle circulation: Integrated timetable and vehicles scheduling on hybrid time-space networks. *Transp. Res. Part C: Emerg. Technol.* 124, 102961. <http://dx.doi.org/10.1016/j.trc.2020.102961>.
- Lieichen, C., Lübbecke, M., Möhring, R., Stiller, S., 2009. The concept of recoverable robustness, linear programming recovery, and railway applications. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Ahuja, R.K., Möhring, R.H., Zaroliagis, C.D. (Eds.), *In: Robust and Online Large-Scale Optimization*, vol. 5868, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–27. http://dx.doi.org/10.1007/978-3-642-05465-5_1.
- Lusby, R., Larsen, J., Ryan, D., Ehrgott, M., 2011. Routing trains through railway junctions: A new set-packing approach. *Transp. Sci.* 45 (2), 228–245. <http://dx.doi.org/10.1287/trsc.1100.0362>.
- Mussone, L., Wolfler Calvo, R., 2013. An analytical approach to calculate the capacity of a railway system. *European J. Oper. Res.* 228 (1), 11–23. <http://dx.doi.org/10.1016/j.ejor.2012.12.027>.
- Nießen, N., 2008. Leistungskenngrößen für Gesamtfahrstraßenknoten (Ph.D. thesis). Verkehrswissenschaftliches Institut der RWTH Aachen (German only).
- Nießen, N., 2013. Waiting and loss probabilities for route nodes. In: *International Conference on Railway Operations Modelling and Analysis. RailCopenhagen*.
- Palmer, G.I., Knight, V.A., Harper, P.R., Hawa, A.L., 2019. Ciw: An open-source discrete event simulation library. *J. Simul.* 13 (1), 68–82. <http://dx.doi.org/10.1080/17477778.2018.1473909>.
- Parker, D., Norman, G., Kwiatkowska, M., 2000. PRISM. URL: <https://www.prismmodelchecker.org/>.
- Potthoff, G., 1970. Verkehrsströmungslehre - Die Zugfolge auf Strecken und in Bahnhöfen. Transpress (German only).
- Python Software Foundation, 2022. Python 3.10.9. URL: <https://docs.python.org/release/3.10.9/index.html>.
- Schmitz, C., Weik, N., Zieger, S., Nießen, N., Schmeink, A., 2017. Markov models for the performance analysis of railway networks. In: *International Conference on Railway Operations Modelling and Analysis. RailLille, Lille, France*, p. 23.
- Schwanhäüßer, W., 1974. Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn (Ph.D. thesis). Verkehrswissenschaftliches Institut der Rheinisch-Westfälischen Technischen Hochschule Aachen (German only).
- Schwanhäüßer, W., 1978. Die ermittlung der leistungsfähigkeit von großen fahrstraßenknoten und von teilen des eisenbahnnetzes. *Arch. Für Eisenbahntechnik* 1978 (38), 7–18.
- Schwanhäüßer, W., Schultze, K., 1982. Ermittlung von Qualitätsmaßstäben für die Berechnung der Leistungsfähigkeit eines Streckenabschnittes und Entwicklung eines Rechenverfahrens zur Ermittlung von Endverspätungen: Forschungsarbeit für die Deutsche Bundesbahn. (Project report not formally published, German only).
- Sommereder, M., 2011. *Modelling of Queueing Systems with Markov Chains: An Introduction to Basic and Advanced Modelling Techniques*. BoD-Books on Demand.
- Team SimPy Revision, 2023. Simpy 4.0.1. URL: <https://simpy.readthedocs.io/en/4.0.1/>.
- The Ciw library developers, 2024. Ciw: v3.2.2. <http://dx.doi.org/10.5281/zenodo.12667778>.
- UIC, 2004. Code 406 - capacity.
- UIC, 2013. Code 406 - Capacity.
- Van Rossum, G., Drake Jr., F.L., 1995. *Python Tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods* 17, 261–272. <http://dx.doi.org/10.1038/s41592-019-0686-2>.
- Wakob, H., 1981. Wahrscheinlichkeitsmodell zur Ermittlung von Wartezeiten bei Zugfolgefällen. *Arch. Für Eisenbahntechnik* 36, 15–22, (German only).
- Weik, N., 2020. Long-Term Capacity Planning of Railway Infrastructure – A Stochastic Approach Capturing Infrastructure Unavailability (Ph.D. thesis). RWTH Aachen University, <http://dx.doi.org/10.18154/RWTH-2020-06771>.
- Weik, N., Nießen, N., 2017. A quasi-birth-and-death process approach for integrated capacity and reliability modeling of railway systems. *J. Rail Transp. Plan. Manag.* 7 (3), 114–126. <http://dx.doi.org/10.1016/j.jrtpm.2017.06.001>.
- Weik, N., Warg, J., Johansson, I., Bohlin, M., Nießen, N., 2020. Extending UIC 406-based capacity analysis – New approaches for railway nodes and network effects. *J. Rail Transp. Plan. Manag.* 15, <http://dx.doi.org/10.1016/j.jrtpm.2020.100199>.
- Wendler, E., 2007. The scheduled waiting time on railway lines. *Transp. Res. Part B: Methodol.* 41 (2), 148–158. <http://dx.doi.org/10.1016/j.trb.2006.02.009>.
- Yaghini, M., Nikoo, N., Ahadi, H.R., 2014. An integer programming model for analysing impacts of different train types on railway line capacity. *Transport* 29 (1), 28–35. <http://dx.doi.org/10.3846/16484142.2014.894938>.
- Zieger, S., Weik, N., Nießen, N., 2018. The influence of buffer time distributions in delay propagation modelling of railway networks. *J. Rail Transp. Plan. Manag.* 8 (3–4), 220–232. <http://dx.doi.org/10.1016/j.jrtpm.2018.09.001>.
- Zwaneveld, P.J., Kroon, L.G., Romeijn, H.E., Salomon, M., Dauzère-Pèrès, S., van Hoesel, S.P.M., Ambergen, H.W., 1996. Routing trains through railway stations: Model formulation and algorithms. *Transp. Sci.* 30 (3), 181–194. <http://dx.doi.org/10.1287/trsc.30.3.181>.