IMPROVING RAIL NETWORK SIMULATIONS WITH DISCRETE DISTRIBUTIONS IN ONTIME

BURKHARD FRANKE, DAN BURKOLTER & BERNHARD SEYBOLD
TrafIT Solutions GmbH, Switzerland

ABSTRACT
To analyse a rail network's punctuality and the operational quality of a timetable on a network-wide scale an advanced simulation is needed. Whereas most simulations use a Monte Carlo approach, we calculate delay distributions analytically and thus need only a single calculation run. Previously we used exponential distribution functions as they map the status in railway operations well and are suited for efficient calculation of delays. The resulting delay distributions due to primary delays along a train's itinerary as well as delay propagation from other trains is handled by convolution of these distribution functions. However, as the resulting distributions become more complex, a simplification step is needed from time to time to keep calculation times reasonable. Increased requirements for the accuracy of the simulation model and improvements in the computational potential led us to remodel the delays with discrete distributions. This has two main advantages. First, restrictions on the possible form of primary delays are much smaller compared to the previous exponential distributions and second, the simplification step is no longer needed, which increases accuracy considerably. We discuss the different options of distribution modelling and their use in railway applications.
Keywords: railway, simulation, delay distribution modelling, operational quality, punctuality.

1 INTRODUCTION

The analysis of the operational quality of a timetable is usually based on the examination of operational data covering the trains and their respective delays in a time interval. Common quality parameters to describe the performance of railways are key figures like punctuality, mean delays, or the variation of arrival times. Looking at the whole set of data reveals the stochastic nature of delays: The deviations can be best described as distributions as Fig. 1 shows for the example of the graphic timetable of planned and realised train runs over several weeks. The actual train runs are distributed around the scheduled train path depending on their delay characteristics on any recorded day.

The characteristics of the delay distributions will be different for every train, station, and time of day or time of year, as the underlying reasons for delays vary. Those reasons can be differentiated in primary delays (caused by external sources) and secondary delays (passed on due to train interaction), which will be discussed in Section 2.

Based on our experience with delay analysis and timetable improvement, our principal insight is that a railway timetable will unavoidably experience a certain level of disturbance and this level of disturbance can best be described as distributions of (primary) delays. It is therefore reasonable to assess the operational quality of a future timetable on a given infrastructure by applying distributions of (primary) delays on every train and simulate delay propagation and delays reduction, as this replicates the situation a timetable will experience during its existence (the timetable period). This idea was developed amongst others in the thesis of Büker [2] and is the basis of what was implemented as the tool OnTime. A description of OnTime and its methodical advantages is part of Section 3. The main topic of this paper is the modelling of distributions in OnTime's calculation kernel. Section 4 outlines the development of the modelling and reviews the reasons for and the implementation of a discrete modelling of distributions in OnTime. The paper is concluded in Section 5.

WIT Transactions on The Built Environment, Vol 213, © 2022 WIT Press
www.witpress.com, ISSN 1743-3509 (on-line)
doi:10.2495/CR220191

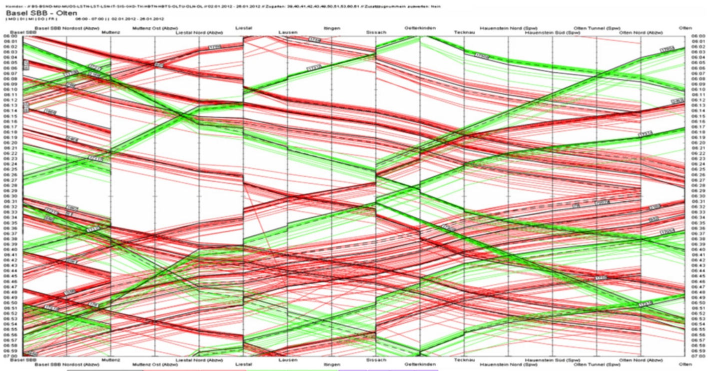

Figure 1: Visualisation of operational data [1].

2 ON DELAYS AND OPERATIONAL QUALITY

Railway operations are vulnerable to delays as variability in operational process times and unexpected disruptive events lead to deviations from the planned timetable [3]. These deviations are called primary delays to discern them from secondary delays, which result from the interaction of trains and the propagation of delays. This differentiation helps to understand and counteract train delays. It is also the conceptual basis for the simulation tool described in the next section since a model cannot deduce the primary delays but must reproduce the interaction of trains and the consequences thereof.

However, these thoughts on primary and secondary delays do have the problem that operational delay data as presented in Fig. 1 do not differentiate between delay reasons as only the resulting delay can be directly measured. From our experience only few railways are able to identify primary delays, best known to us are studies in Switzerland by Labermeier [4] and Schranil [5]. The latter analysed the distribution of primary and total delays in the SBB network.

Although the shape of the distribution is probably SBB-specific, two findings are common to both papers:

- Small (primary) delays are frequent, large delays are rare.
- Very small delays are often recorded incorrectly or even neglected.

The first statement seems trivial, but even data presented in Fig. 2 does not support this completely. Whereas primary delays of 3 minutes are more frequent than primary delays of 4, 5 or more minutes, delays of 1 or 2 minutes hardly seem to occur at all. Our explanation for this phenomenon is given in the second statement: the recording of such small delays is often not correct. This is partly due to railway companies' regulations on how to record delays in the first place and partly due to difficulties in measuring small delays as records are often a spin-off product of the rail control system. Reports usually rely on the delay causes given by railway operators who are obliged to record a delay cause for delays exceeding a defined threshold. What happens below the threshold mostly stays in the dark.

Common sense says that the declining tendency of the frequency of larger delays has its continuation with smaller delays being more frequent: severe weather conditions causing a minute of running delay are less likely than weather conditions causing 30 seconds of delay. Large crowds of passengers delaying departure by 40 seconds are less frequent than crowds

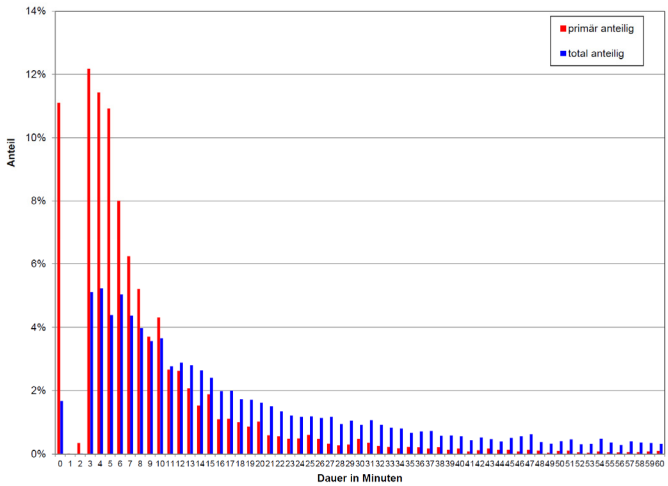

Figure 2: Density function of primary and total delays according to Schranil [5].

of passengers delaying a train by 20 seconds. This is backed by detailed analysis for instance by Yuan [6], who investigated train movement on the level of the rail control system. A similar approach was used by Labermeier [4], and both analyses end up focussing on delays in the range of seconds to few minutes, as these are the most frequent ones.
If we break down the general findings to particular locations or trains, the pattern is repeated as is shown in Fig. 3, small delays and – depending on the amount and allocation of margins – even small early arrivals are more frequent than larger deviations.

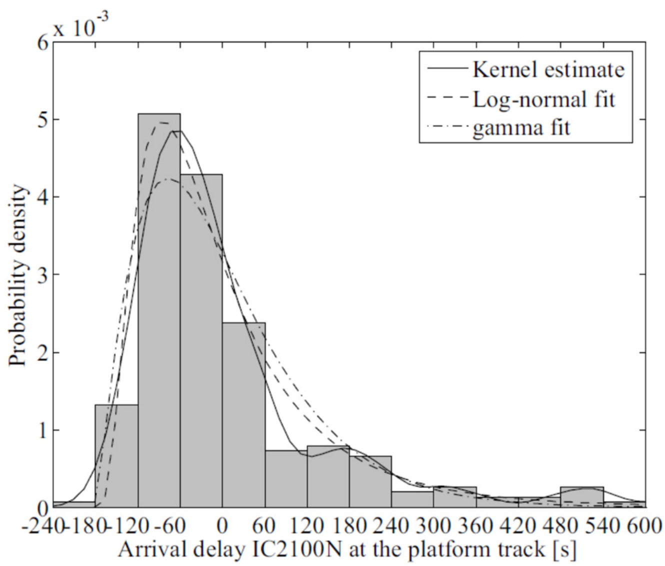

Figure 3: Histogram, fitting curves, and kernel estimate for arrival delays of a train [6].

Fig. 3 also shows some fitting curves for modelling the delays mathematically, a topic we cover in Section 4.

So, although individual disturbances cause discrete primary delays, the timetable over a timetable period is subject to a kind of random noise of disturbance. The distribution of this random noise is characteristic for each railway – a product of weather conditions, maintenance level, and other historic and current circumstances. For a realistic reproduction of operational quality, we need to use proper distributions of primary delays. As we do not know where disturbances will occur and which trains will be affected in a future timetable, we must apply a delay distribution on every train everywhere on its itinerary. This is the idea behind the stochastic simulation tool OnTime.

3 THE TOOL ONTIME

In 2011 SBB commissioned the development of a tool to complement their existing microscopic, deterministic simulation tool. The so-called “Stabilitätsmodell” (stability model) was meant to simulate the complete network to evaluate timetables and infrastructural measures for short- to long-term planning. This tool should not construct or change a timetable but evaluate a given timetable in its entirety regarding chosen punctuality measures. To win the tender the two companies TrafIT Solutions GmbH, Zurich, and VIA Consulting & Development GmbH, Aachen, joined forces to implement the tool OnTime based on research at RWTH Aachen by Weidner and Büker [7] and later Büker [2] and in use at railways such as SBB or DB

3.1 The application

OnTime (see Fig. 4) is designed to work as a complement to an existing timetabling tool as it imports timetables via standardized exchange formats and has no integrated planning capabilities or running time calculator. This is mostly because the first railways using the tool did not want to have “another running time calculator”, as keeping all necessary data up to date in several tools is quite costly and different calculators will always differ in running times at least to some degree. OnTime therefore depends on a timetable system to export a timetable with all the necessary data, most notably the differentiation between minimal running times and running time reserves, which can be used to reduce delays. Information on connections and turn-arounds are either imported or generated in a preprocessing step. OnTime uses a rule-based modelling [8] to represent the infrastructure properties, which allows working with the different microscopic or macroscopic levels of detail in different planning horizons for infrastructure. Primary delays can be provided either as default distributions for the whole network or specified for certain stations and sections, times, and trains or types of trains. The primary delay distributions are simply defined by giving a probability of delay and the average delay in case of delay and are differentiated between stopping, departing, and running delays. A set of input data – a scenario – forms an activity graph [1] for a selected day of operation. Trains accumulate delays incurred from primary delays as well as delays propagated from other trains. Since delays are modelled as distributions, the accumulated delay is calculated as a convolution of delay distributions. This delay is reduced by the amount of running or stopping time reserves present, which corresponds to a left shift of the delay distribution.

A single calculation of a scenario produces the results, which can be used to evaluate a timetable or to compare variants of timetables or infrastructure changes based on network-wide operational key figures. All resulting delays are modelled as distributions which enables

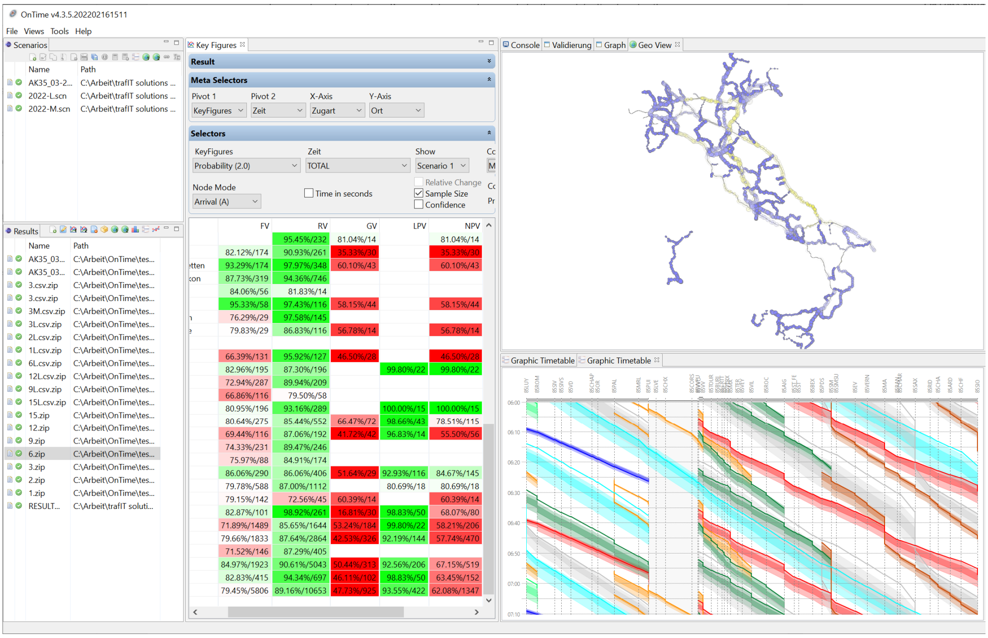

Figure 4: The application OnTime and some views on results.

calculating (almost) any desired key figure. As an example, since OnTime calculates the resulting probability distribution for every train and station, it is possible to determine the frequency of arrivals and departures within the chosen punctuality threshold, e.g. the 5 minute punctuality is given by the integral of the probability density function up to the 5 minute threshold.

3.2 Methodical advantages

Using distributions to model delays in a simulation has significant advantages over a Monte Carlo approach that uses individual disturbances in a (series of) deterministic simulations.

3.2.1 Real world key figures

Traditional tools of railway operation science do not use the same key figures that are common for describing railway operation quality. As Bär et al. [9] pointed out, analytical approaches and simulations do not match in their respective statements and even simulations often do not match empirical analysis of operation. This is mostly because simulations are restricted to parts of a network, so key figures are limited in scope or of a derivational nature. The ability of OnTime to simulate whole networks and to reproduce the operational quality of an average day or a timetable period makes it possible to forecast a future timetable's punctuality. Effects of timetable or infrastructure changes can be assessed consistently for the whole network to compare and prioritise measures.

3.2.2 Calibration

The verification and calibration of a simulation model is a long and laborious process. Using a Monte Carlo approach with repeated simulation runs to mimic all disturbance situations makes this task almost impossible. The fact that disturbances can (and do) occur everywhere in the network and cause a wide range of initial delays (see Section 2) should be reproduced by the tool, which is possible with OnTime. For example, SBB used primary delays extracted as described by Labermeier [4] to verify the OnTime model and as a starting point for a finer calibration.

3.2.3 Combination with demand models to evaluate passenger punctuality

Evaluating the operational quality of a timetable from a passenger's perspective is often quite demanding: a train delay of a few minutes can be just a nuisance, but if it results in missing a connection, the passenger's delay could become an hour or more. For example, Landex [10] described ways to assess passenger delays. Some of the previous problems are resolved when using the stochastic simulation in OnTime as all delays and the probabilities of catching or missing connections are available for the whole network as distributions. This suits the needs of demand modelling tools. In 2016 a project together with SBB showed the feasibility of the combination of OnTime and SBB's demand modelling tool SIMBA [11]. The combination allows for a holistic evaluation of timetable and infrastructural measures [12].

4 MODELLING DISTRIBUTIONS

4.1 Exponential

Exponential distributions are widely used for modelling delays since on the one hand they fit measured delays quite well [6], [13] and on the other hand their convolution is very easy to calculate, which makes calculating delay propagation fast. However, in order to better fit measured delays, Büker [2] introduced delay distributions with segment-wise exponential distributions, with cumulative distribution functions of the following form:

$$F_V\,(t) = \begin{cases} c^0 - \sum_{i=1}^{n^0} a_i^0 \cdot t^{d_i^0} \cdot e^{(-\lambda_i^0 \cdot t)} & t < g^1 \\ \vdots & \\ c^s - \sum_{i=1}^{n^s} a_i^s \cdot t^{d_i^s} \cdot e^{(-\lambda_i^s \cdot t)} & g^s \le t \end{cases}$$

Thus, the probability of a delay ≤ t is a constant c^n reduced by an exponentially decreasing part for every segment of the distribution.

In theory, delays can be fit arbitrarily accurately using this type of delay distributions by adding more segments. However, though the main mathematical advantage of exponential distributions – namely that the convolution is again an exponential distribution – is retained, the number of segments grows exponentially. This leads to severe limitations in practical applications due to running times becoming too large. Therefore, Büker [2] developed methods for simplifying these distributions by reducing the number of segments. This method is implemented in OnTime. Initially, delay distributions have up to three exponential segments. During calculation the number of segments will quickly grow, thus, when a distribution becomes too large a simplification will be triggered, which reduces the number of segments again.

The approach has worked fine in OnTime and is still widely used, yet it has two drawbacks:

1. Obviously, a simplification of a delay distribution does not fit the original exactly leading to small inaccuracies, which do not matter when a whole railway network is simulated. However, they can become an issue if two simulation runs of the same network with slightly changed parameters are compared. In this case, it is not always easy to distinguish whether the differences in the results are real, in the sense that they are caused due to different parameters, or just a result of simplifications taking place at different stages of the simulation. An example is shown in Fig. 8.
2. The simplification step is numerically very sensitive, which makes using a higher floating-point precision than double necessary. Unfortunately, this increases the running time for a simulation considerably.

With increased requirements on the accuracy of the simulation and increased computational potential changes were made to alleviate these issues. Yet, as raising the number of segments and increasing the floating-point precision did not solve the problems satisfactorily, a new approach became necessary.

4.2 Discrete modelling

In order to tackle the two drawbacks to the exponential distribution approach, we have implemented delay distributions modelled as discrete distributions (Fig. 5) with a fixed step size s, a minimum delay d_{min} (which can be negative, i.e., an earliness), and a maximum delay d_{max}. Delay distributions modelled this way fulfil the necessary requirement of their convolution again being a discrete distribution. Additionally, the step size is not changed by convolution. And though the values of the minimum and maximum delays would decrease or increase respectively, we avoid this by cutting of delays with a probability lower than a chosen minimum probability, e.g., one in a million. This leads to the main advantage of the discrete distribution modelling compared to segment-wise exponential distributions: no simplification steps are needed. Consequently, a floating-point accuracy of double proved to be sufficient for the implementation in OnTime.

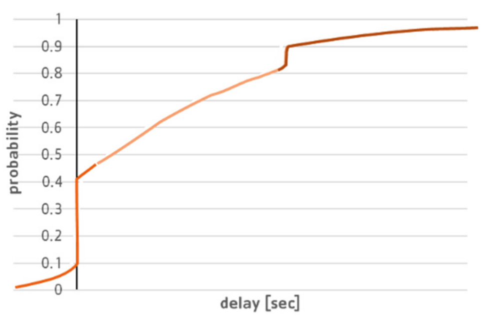

Figure 5: Delay distribution as a segment-wise exponential distribution compared to a discrete distribution. The exponential segments are indicated by the different shades of orange.

A further advantage of modelling delays with discrete distributions is that practically any form can be fitted very accurately further increasing the accuracy of the delay prognosis. As an example, delay distributions often contain discontinuities due to dispatching decisions, which can be better fitted with discrete distributions.

4.3 Experimental results

Since the step size s plays a crucial role both for running time and accuracy, we investigated the delay prognosis for test data of a large national network. All calculations were conducted on a standard laptop with an Intel i7 hexa-core CPU and 32 GB RAM.

Fig. 6 shows the calculation time and derived punctuality measures for several step sizes. As expected, the calculation time decreases quadratically with the step size s as the calculation time increases quadratically for the main driver of running time – calculating convolutions – in the number of supporting points of the discrete distributions. However, after a step size of 9 s no more improvement in running time can be realised since pre- and postprocessing steps independent of the step size s become dominant. On the other hand, the chosen punctuality measures are very robust with respect to changes in the step size. The results are nearly independent of s up to 6 s when they start to rise slightly, i.e., the prognosis starts to show higher punctuality values.

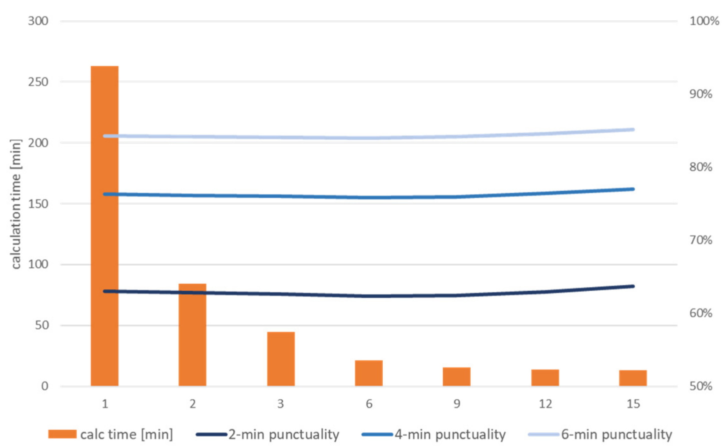

Figure 6: Increment versus punctuality.

Fig. 7 shows the resulting delay measures for several values of s. Again, the measured values do not change much up to a step size of 6 s, where measured values start to decrease a bit. As a result of these tests, OnTime now uses a step size of 3 s for the discrete distributions, which we feel results in a good balance between running time and accuracy.

Table 1 compares the calculation time for the same test scenario using either segment-wise exponential distributions or discrete distributions. The calculation time needed using discrete distributions is down to about a quarter.

Finally, Fig. 8 shows the unwanted changes in punctuality values due to the first drawback mentioned in Section 4.1. The figure shows the difference between two simulation runs, where the input is equal except for a small timetable improvement in the centre of the network. The result of the improvement is visualised by the green areas. However, on the left-hand side where the simulations were calculated using segment-wise exponential

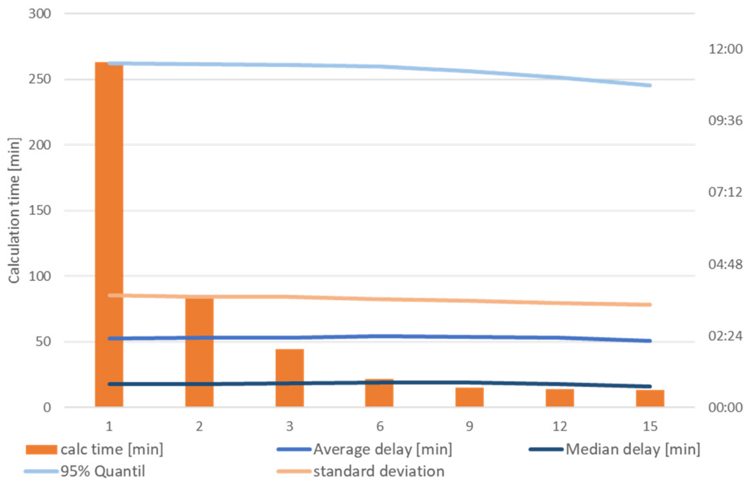

Figure 7: Increment versus delay.

Table 1: Calculation time.

Exemplary network	Calculation time (min)	
	Exponential	Discrete
Germany	220’	60’
Switzerland	125’	30’

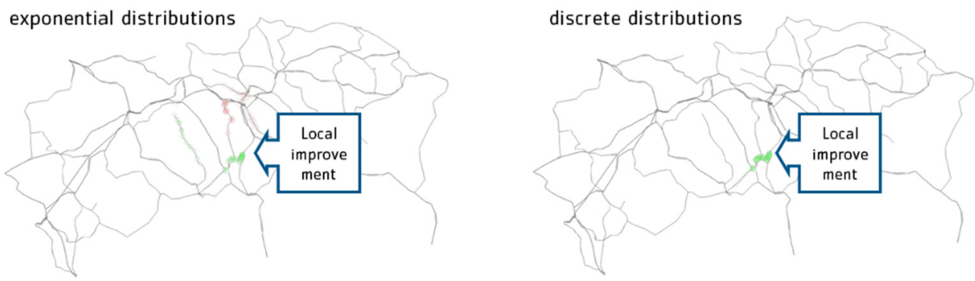

Figure 8: Comparison of difference in punctuality for two simulation runs that are equal up to a local timetable improvement. Green areas show improvements and red areas declines in operational quality.

distributions, some negative effects on operational quality visualised by red areas appear. This change is not caused by the difference of the two inputs of the simulations but are caused by simplification steps happening at different stages of the calculation. As shown on the right-hand side, no such effects occur if the simulation is calculated using discrete distributions.

5 CONCLUSIONS

OnTime is a simulation tool for calculating a delay prognosis for a timetable. While most tools apply a Monte Carlo simulation approach, OnTime calculates resulting delay distributions analytically, thus reducing calculation time considerably and enabling simulations of much larger railway networks. In addition to the previous delay modelling with exponential distributions, we have implemented calculations using discrete distributions. Comparisons between the two approaches have shown that on the one hand calculation time is reduced fourfold and on the other hand accuracy for small changes in the input data is much improved.

REFERENCES

[1] Büker, T. & Seybold, B., Stochastic modelling of delay propagation in large networks. *Journal of Rail Transport Planning and Management*, **2**(1), pp. 34–50, 2012.
[2] Büker, T., Ausgewählte Aspekte der Verspätungsfortpflanzung in Netzen. PhD thesis, RWTH Aachen, Germany, 2010.
[3] Spanninger, T., Trivella, A., Büchel, B. & Corman, F. A review of train delay prediction approaches. *Journal of Rail Transport Planning and Management*, **22**, 2022.
[4] Labermeier, H., On the dynamic of primary and secondary delay. *Proceedings of IAROR,* Copenhagen, 2013.
[5] Schranil, S., Prognose der Dauer von Störungen des Bahnbetriebs. PhD thesis, ETH Zurich, 2013.
[6] Yuan, J., Stochastic modelling of train delays and delay propagation in stations. Dissertation, Delft, Technical University Delft, 2006.
[7] Weidner, T. & Büker, T., Stochastische Simulationen zur Beurteilung der Stabilität von Personenverkehrsfahrplänen. *21. Verkehrswissenschaftliche Tage: Innovation und Investition*, Technische Universität Dresden, 2007.
[8] Seybold, B. & Kuckelberg, A., Adaptive rule-based infrastructure modelling. *Proceedings of IAROR,* Copenhagen, 2013.
[9] Bär, M., Lenze, W., Eisold, J. & Nießen, N., Benchmark von Qualitätsmaßstäben für Kapazitätsermittlungen. *Proceedings of the 2nd International Railway Symposium Aachen,* 2019.
[10] Landex, A., Methods to estimate railway capacity and passenger delays. Dissertation, TU of Denmark Lyngby, 2008.
[11] Olesen, A., Bützberger, P. & Lieberherr, J., Modellierung und Bewertung von Fahrplanangeboten der Zukunft. *Schweizer Eisenbahn-Revue,* **1**, pp. 16–19, 2016.
[12] Franke, B. & Lieberherr, J., Kombinierte betriebliche und verkehrliche Bewertung von Bahn-Angeboten. *Proceedings of the 1st International Railway Symposium Aachen,* 2017.
[13] Engelhardt-Funke, O. & Kolonko, M., Simulating delays for realistic time-table optimization. *International Conference on Operations Research 2001*, 2002.