Train delay evolution as a stochastic process

Pavle Kecman[1], Francesco Corman[2], Lingyun Meng[3]
[1]Department of Science & Technology, Linköping University
e-mail: pavle.kecman@liu.se
[2]Section Transport Engineering & Logistics, Maritime and Transport Technology, Delft University of Technology
e-mail: f.corman@tudelft.nl
[3]State Key Laboratory of Rail Traffic Control and Safety Beijing Jiaotong University
e-mail:lymeng@bjtu.edu.cn

Abstract
In this paper we present a method for modelling uncertainty of train delays based on a Markov stochastic process. The dynamics of a train delay over time and space is presented as a stochastic process that describes the evolution of the time-dependent random variable. Probability distribution of an arrival delay in a station changes over time in discrete steps as more information becomes available. We consider and compare the results and computational requirements of two discrete state space formulations. Moreover, we test the applicability of modelling train delays as a non-stationary Markov chain, meaning that the probability of a state change depends on the moment of transition. The model is applied on a set of historical traffic realisation data from the part of the high-speed corridor between Beijing and Shanghai. We analyse the accuracy of predictions as well as the evolution of probability distributions of all events over time. The presented method is important for making better predictions for train traffic, that are not only based on static, offline collected data, but are able to positively include the dynamic characteristics of the ever-changing delays, thus increasing the reliability of prediction by 71%.

Keywords
Prediction, Railway traffic, Stochastic processes, Train delay evolution

1 Introduction

Accurate prediction of train delays is an important requirement for proactive and anticipative real-time control of railway traffic and transport. Traffic controllers need to predict the arrival times of the trains within (or heading towards) their area in order to control the feasibility of timetable realisation. Similarly, the transport controllers on behalf of train operating companies may use the predictions to estimate the feasibility of planned passenger transfers, as well as rolling-stock and crew circulation plans. Valid estimates of arrival and departure times are therefore important for preventing or reducing delay propagation, managing connections, and providing reliable passenger information. The difficulty for predicting train event times comes from the uncertainty and unpredictability of process times in railway traffic. The models for real-time traffic and transport control have so far mostly focused on overcoming the great combinatorial complexity of train rescheduling [6, 23], delay management [11] and rolling-stock and crew rescheduling [25, 27]. The developed

approaches are able to solve complex instances in real-time, however they typically assume perfect deterministic knowledge of the input traffic state and subsequent traffic evolution.

In recent years, the uncertainty of train event times has been recognised as one of the major obstacles for computing feasible and implementable solutions for rescheduling problems in railway traffic [7]. The uncertainty of an event is usually represented by the probability distribution of its realisation. However, most of the existing approaches assume fixed probability distributions for the estimation of process times and do not consider the effect that real-time information on train positions and delays may have on (the parameters of) the corresponding distributions. In order to create realistic online tools for real-time traffic management, the dynamics of uncertainty of delays needs to be considered. When new information about train positions and delays becomes available, the uncertainty for predicting subsequent events is typically reduced. The main objective of this paper is to examine the effect that the prediction horizon and incoming information about a running train may have on the predictability of subsequent arrival and departure times of that train. In other words, we try to give an answer to the question: how does the probability distribution of delay of an event change over time? The concept of the problem is illustrated in Figure 1. With every update of train delay (arrivals to station A and B) probability distributions of arrival times to subsequent stations (C and D) are computed.

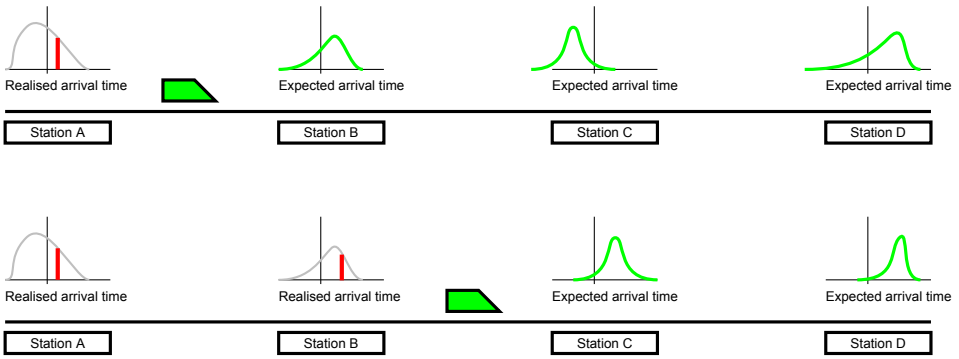

Figure 1: Dynamic evolution of probability density in time

Earlier approaches in delay prediction strongly depend on the availability of an underlying traffic model. Real-time train delay prediction typically relies on the deterministic estimates either (i) computed offline using theoretical values [10] or (ii) dynamically with respect to the actual traffic conditions using train motion equation [9] or historical traffic data [18]. The model-based real-time prediction tools require frequent (almost continuous) updates of train positions and delays that are further propagated through the model. This may be a problem for many railway networks since train delays are in current practice in Europe typically measured and registered only roughly at the main signals in a station, corrected with a fixed term and rounded to full minutes [17]. Moreover, the applicability of these models is conditioned on the availability of data that is needed to model the train dynamics or historical traffic data. The stochastic delay propagation models [3, 21] were mostly used for offline analyses of timetables. An approach that considers the dynamics of uncertainty of train delays was presented by Bauer & Schöbel [1]. The authors developed

a ‘delay generator’ for the purpose of integrating uncertainty in online traffic management. A uniformly distributed delay value is assigned in real-time to a set of randomly chosen events. However, their approach represents a rather theoretical concept that mimics the evolution of train delays in time in order to create realistic instances for validation of an online delay management tool.

In this paper we present a method for modelling the uncertainty of train delays based on a Markov stochastic process. The dynamics of a train delay over time and space is presented as a stochastic process that describes the evolution of delay as a time-dependent random variable [26]. Probability distribution of an arrival delay in a station changes over time in discrete steps as more information becomes available. A train run is represented as a Markov chain with state transitions in discrete moments that represent arrival and departure events from a scheduled stop. After every registered departure or arrival event, the conditional probability distributions of the downstream events are updated with respect to the essential assumption for Markov processes that given the present, future events do not depend on the past. We consider and compare the results and computational requirements of two discrete state space formulations. Moreover, we test the applicability of modelling train delays as a non-stationary Markov chain, meaning that the probability of a state change depends on the moment of transition. The model is applied on a set of historical traffic realisation data from the part of the high-speed corridor between Beijing and Shanghai in China. We analyse the accuracy of predictions and the evolution of probability distributions of all events over time.

The remainder of this paper is structured as follows. The next section presents the review of state-of-the-art in real time predictions of railway traffic. Sections 3 describes the methodological framework, followed by the description of the case study and data set in Section 4. Sections 5 and 6 present the computational experiments and results. Finally the main findings are summarised and the directions for future research given in Section 7.

2 Problem description and relevant approaches

The role of real-time prediction systems in a traffic control environment is twofold. The first requirement is the prediction of train trajectories until the next controllable point. In other words, train running times to the next station or point in the network where a dispatching action can be performed or where a passenger or logistic connection is planned. That way the accuracy of the input to the corresponding rescheduling models can be improved. The second role is connected to the procedure of finding a new feasible schedule as a result of the rescheduling process. During the optimisation stage, rescheduling systems evaluate different solutions in the search for the optimum. Traffic evolution according to each observed potential solution is predicted and summarised in the corresponding value of the objective function. By improving the predictions in this stage, the rescheduling models that assume the full knowledge of future traffic state can be significantly improved.

Real-time prediction models presented in this review can be classified, according to how they tackle uncertainty, to deterministic and stochastic. Deterministic models assume full knowledge of the future traffic evolution. On the other hand, stochastic models attribute each event with a probability distribution in order to model the uncertainty of its realisation. Moreover, we classify the presented approaches based on how they use the real-time information to update their predictions to static and dynamic. Whereas static prediction models are based on the offline computed estimates and parameters, dynamic models are updated in real-time as new information becomes available.

The deterministic prediction tools mostly model train traffic as a discrete event dynamic system represented by a directed acyclic graph. Considered events comprise only station events (arrivals, departures, through runs) or also signal passing events depending on the level of modelling. Process times are in the majority of approaches computed offline using theoretical estimates. Running time estimates are typically computed using train motion equation that considers the dynamic properties of rolling-stock and infrastructure. They are assumed to be fixed and do not change depending on the real-time information obtained from the monitoring system. Dwell times are predicted with fixed estimates. An example of such deterministic static prediction model has been implemented in the Swiss traffic control system RCS [10]. After each train position update, a critical path algorithm derives predictions of all event times in the graph.

Another example is presented by Fukami et al. [13] who described a prediction system for high-speed network in Japan. Given the information from the monitoring system, the train trajectory to the next station is simulated in real-time. The delay for the succeeding stations is then simply extrapolated. D'Ariano et al. [8] presented a graph-based model designed for real-time rescheduling. Accurate prediction of future train positions was an important requirement in order to estimate the effect of proposed rescheduling actions. Temporal decomposition was used to apply the model for predictions over several hours. In an improved approach, the running time estimates for trains hindered in route conflicts are updated online to model braking and re-acceleration [9].

The availability of historical traffic realisation data motivated another direction for the deterministic prediction models. Statistical analysis and data mining techniques can be applied to estimate the dependence of process times on factors such as delay and peak hours [17]. Process times are estimated online depending on the information received by the monitoring system. This method has been used by Hansen et al. [16] who presented a way to predict the realisation time of a station event by computing the critical path from the last realised event. Kecman and Goverde [18] extended this approach by modelling train traffic on the level of block sections. The online character of the approach is further improved by an adaptive component that monitors the accuracy of predictions in real-time and adapts the prediction of future events accordingly.

The deterministic approaches are focused on explaining the uncertainty of process times in railway traffic. Even though the more advanced data-driven models are able to explain a large percentage of process time variability using the values of explanatory variables, a certain degree of variability, especially for dwell times, still remains unresolved. Another important aspect for application of the presented models for real time prediction is that their accuracy depends to a great extent on the level of modelling and spatial and temporal resolution of updates obtained from the monitoring systems. Therefore, in order to obtain accurate predictions, a detailed model of railway traffic is required with frequent updates on train positions on the network.

Stochastic models in the relevant literature were mainly developed with a purpose to analyse timetable robustness, stability and resilience [15]. Büker and Seybold [3] modelled delays as random variables, described with suitable distribution functions, and applied analytical methods to compute delay propagation in a mesoscopic graph-based model. A stochastic delay propagation approach based on processed train event recorders data was presented by Medeossi et al. [21]. Delay propagation is computed by means of stochastic blocking times. The method relies on asynchronous simulation of individual train runs based on probability distributions of train motion parameters [20]. Meester and Muns [22]

formalised the model presented by Carrey [5] as a stochastic event graph. The distributions of free running times are given and the process dependencies are computed as a mixture of the corresponding distributions. A set of performance measures is derived as a linear combination of delay distributions. Yuan and Hansen [30] described a detailed analytical stochastic delay propagation model for complex station areas. First, analytical formulas are given for deriving secondary delays of departing and arriving trains separately. Conditional distributions of arrival and departure times are computed using convolutions for computing the distribution of the sum of random variables.

The described stochastic approaches approaches are inherently static. They do not consider the effect that the real-time information obtained from the monitoring system may have on reducing the uncertainty of the future events. The first stochastic approach that to some extent exploits the current traffic conditions to reduce uncertainty is presented by Berger et al. [2]. They created a stochastic graph-based model for delay prediction. The approach is suitable for online applications where updates about train positions are frequent. Running times are modelled as random variables. The authors proposed several theoretical probability distributions of running times, depending on departure time for each train type. Even though this way a real-time information on train departure time can be used to adapt the estimate of the subsequent running times, the proposed approach does not provide explicit conditional probability of running times depending on departure delays.

In this paper we present a way to overcome the limitations of the presented approaches. A stochastic dynamic model is described where a train delay is represented as a time-dependent random variable. The probability distribution of a delay thus changes over time as more information about the train run becomes available. The main advantage of our approach is that it models the dynamics of uncertainty over time thus providing up to date probability distributions of train event times in real time. The approach does not depend on the detailed information about infrastructure, train orders routes and frequent updates of train movements that are often difficult to obtain. Moreover, this method for delay prediction is suitable even for the low granularity of historical traffic data that provides only arrival and departure times of each train. This approach enables a direct use of imprecise data to predict the evolution of train delays over time.

3 Methodological framework

3.1 General description

This section presents the methodology of modelling train delays as stochastic processes. Each train run is presented as an independent stochastic process. Such approach neglects the fact that multiple train runs are often interdependent because they use the same infrastructure or have scheduled passenger transfers or rolling-stock or crew connections. However, modelling those interdependencies requires a detailed knowledge of the train routes and departure orders which are often not available in the data. In fact, most often the data about train traffic include only station event times measured with a low level of precision, i.e., rounded to full minutes. Therefore, not enough information is available to build the detailed models that were dominant in previous approaches for delay prediction [18]. This justifies the stochastic approach that models the uncertainty of train event times instead of a deterministic approach.

A train run can be presented as a discrete event dynamic sequence of arrival and de-

parture events (Figure 7). The events are connected by the corresponding running and dwell processes. The number of events corresponds to the number of scheduled arrivals and departures. A through event can be separated into the arrival and departure events that occur simultaneously. The events occur in a fixed sequence $j \rightarrow k$, where $k = j + 1$, $j = 1, 2, \ldots, n - 1$, and n is the total number of events in the train run. This way a train run can be modelled by knowing only the events scheduled by the timetable. The variable of the system is train delay that may change at every event. We assume that no intermediate information about train delays is available, i.e., the system is event-driven and transitions in only at events that are predefined by the timetable. We therefore define the number of an event $i = 1, 2, \ldots, n$ in each sequence as the parameter of the system and in the remainder of this paper refer to it as time.

Figure 2: Representation of a train run as a discrete sequence of events

3.2 Train delay evolution as a Markov process

Train delay evolution over a sequence of events can be represented as a stochastic process by augmenting each event with a probability distribution of delay [3]. Such probability distributions are typically determined from historical traffic data. We extend this approach in order to make it compatible with the real-time environment where messages on train delays are received as each event gets realised. The idea is to model how the real-time information affects the uncertainty of the evolution of train delay by means of Markov property. We assume that the delay of a certain event in the future can be fully predicted based on the currently known delay. Therefore, a train delay in the future depends only on the current delay and not on the delay of events that preceded it [26]. This assumption limits the model to a "memoryless" approach where the past delays and delay jumps of a train cannot be used to predict the future. However, it is reasonable to assume that the propagation of any continuous deviation cause to the subsequent processes would be prevented by a traffic control action. Consequently, the assumption of Markov character of the process does not reduce the validity of the model.

Stochastic process of train delay evolution can be represented as a sequence of random variables $X_1, X_2, \ldots, X_n$. Each random variable represents a delay of event i, where $i = 1, 2, \ldots, n$. The Markov property is formally given in equation (1) where $x_i \in S$ the value of the corresponding random variable and S is the state space.

$$P\{X_{i+1} \mid X_i = x_i, X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2}, \ldots\} = P\{X_{i+1} \mid X_i = x_i\}. \quad (1)$$

Probability of transition form state x_i to state x_{i+1} in the time interval $i, i+1, \forall i \in 1, \ldots n-1$ is given by:

$$P\{X_{i+1} = x_{i+1} \mid X_i = x_i; i, i+1\} = P_{(i,i+1)}[x_i, x_{i+1}]. \quad (2)$$

In case that the distributions are independent on the time instant i of the transition, the resulting process is called a stationary or time-homogeneous Markov process for which a vast mathematical theory exists in terms of modelling the long term behaviour [12]. However, the application of stationary Markov process represents a limitation in modelling train delay evolution. A train delay increase or a decrease during a process execution (delay jump) typically depends on the observed process. Delay jumps may have a different probability distributions for running and dwell processes. Moreover, some processes are scheduled with more time reserves which may have a significant impact on the corresponding delay jump. For that reason, in this paper we examine the performance and computational requirements for modelling train delay evolution as a non-stationary Markov process.

Depending on the number of possible states that the observed random variable can take, a state space S can be discrete or continuous. In this paper we consider two discrete discrete state space definitions:

1. $S_1 = \{s_1 = [a_1, a_2], s_2 = (a_2, a_3], s_3 = (a_4, a_5]\}$
2. $S_2 = \{a_1, \ldots, -2, -1, 0, 1, 2, \ldots, a_2\}$

In the first case, the state space S_1 is separated into three subsets that represent: (i) early or punctual events, (ii) events with a small delay and (iii) events with a large delay. The limits of the subsets are determined by clustering the historical data. The second state space definition is a bounded set of integers. More details on state space definitions for a particular data set are given in Section 5.

The probabilities for the transition of a random variable in time, $X_i \rightarrow X_{i+1}$, are given by the stochastic matrix $\mathbf{P}$ of size $|S| \times |S|$. Each row of this matrix $\mathbf{P}[j, \cdot]$ sums to one and gives the conditional probability distribution of $P[X_{i+1} \mid X_i = j]$. Thus each field $\mathbf{P}(j, k)$ represents the conditional probability given in Equation (2) where $x_i = j$ and $x_{i+1} = k$. Note that due to a time-inhomogeneous property of the process each transition $X_i \rightarrow X_{i+1}, i = 1 \ldots n-1$ is modelled by a separate stochastic matrix $\mathbf{P}_{i,i+1}$. The sequence of stochastic matrices fully defines the time-inhomogeneous stochastic process for modelling train delay evolution.

3.3 Prediction using simulation of Markov process

Given the initial value of the first random variable in the sequence the remaining variables in the sequence can be computed recursively [26]. The probability distribution for each state is obtained using:

$$P[X_{i+1} \mid X_i = j] = \mathbf{x_i} \cdot \mathbf{P}_{i,i+1} \tag{3}$$

where $\mathbf{x_i}$ is a size $|S|$ vector representation of the one-point distribution where j^{th} element is equal to one and the remaining elements are zero. Given the conditional probability distribution, the value of the random variable X_{i+1} can simply be computed using the inverse transform sampling [19].

4 Case study

The methodology described in the previous section is used to model delay evolution of trains running over the high-speed line between Beijing and Shanghai in China. The Beijing–Shanghai High-Speed Railway is a 1318 km long double track line that connects two major

economic zones in the People's Republic of China. There are 24 stations in total along the route but only the data from the northern part of the line between the terminal in Beijing and Dezhou East were made available for this study (Figure 3). There are in total 58 G trains with maximum speed 300 km/h and 12 D trains with maximum speed 250 km/h daily per direction.

Figure 3: Railway line used in the case study

More than three months of historical traffic realisation data were used (between the 1st of December, 2013 and the 4th of March 2014). For each train run in the given period, the planned and realised time for each departure, arrival and through events are given in full minutes. No track occupation or signal passing information is given. The data set considered in this study contains 6803 train runs of south bound G trains. A test set containing 20% of randomly selected train runs is excluded from the analysis and is only used to test the accuracy of the models that are trained with the remaining data.

There are several peculiarities of the traffic on the high-speed that need to be considered:

- Not all trains that belong to the same train line have the same stopping pattern. In that sense, series types G and D indicate the rolling-stock used rather than the train line as it is commonly understood
- Trains are allowed to depart up to 5 minutes before their scheduled departure time.

Figures 4 and 5 depict the histograms of arrival and departure delays contained in data set. The vast majority of considered events occur before their scheduled time. This will to a great extent affect the results and analyses described in the following sections.

Data contains no information about the train routes, that is required in order to investigate the interdependence between trains that use the same infrastructure. This prevents building of the detailed traffic models that are required for deterministic prediction of train delays, such as those developed by Dolder et al. [10] or Kecman and Goverde [18]. The methodology presented in the previous section, which models the uncertainty of delays in real-time is therefore used to build the prediction models.

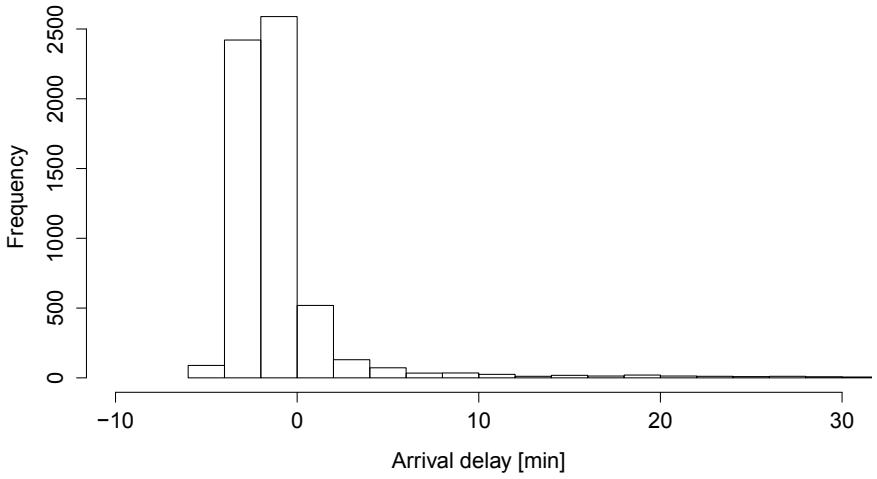

Figure 4: Histogram of arrival delays at all stations for southbound G trains

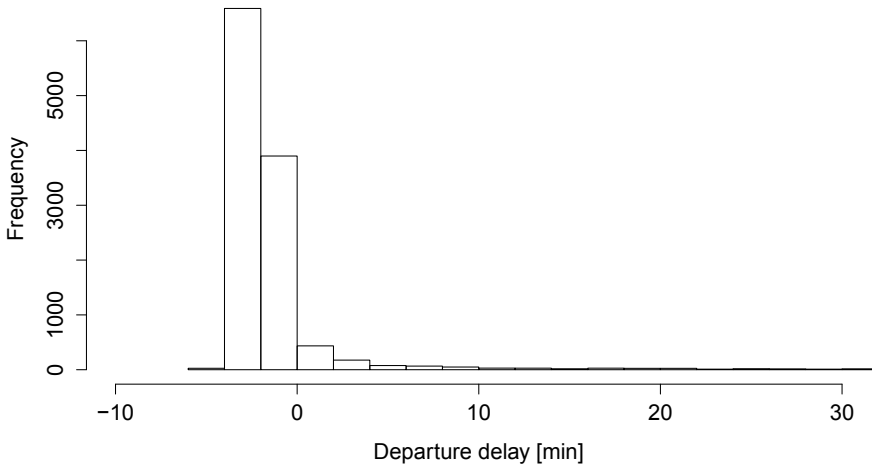

Figure 5: Histogram of departure delays at all stations for southbound G trains

5 Experimental setup

The methodology used for building dynamic stochastic prediction models is given in Section 3. The properties of the data set described in the previous section require some processing steps before building Markov chains from the training set. We examine the application of two state space definitions.

5.1 Clustering trains based on their stopping pattern

The first step in the analysis was to isolate the trains with the same stopping pattern. Each stopping pattern is treated as one train line. Seven different patterns were distinguished among the southbound G trains. This step is important in order to divide the data set into clusters of comparable train runs. Delay evolution of trains within a line is comparable because the updates of train positions in space and time are given at the same station events.

5.2 Construction of Markov chains

We consider two discrete state space definitions. In the first case, delay values are converted into sets of early (punctual) events, small delays and large delays in the following way:

- if delay $\leq 0 \rightarrow$ delay = 'early' (89% of data)
- if $0 <$ delay $\leq 5 \rightarrow$ delay = 'small' (6%)
- if delay $> 5 \rightarrow$ delay = 'large' (5%)

The bounds for state definition reflect the line specific factors such as the fact that the trains are allowed to depart five minutes before their scheduled departure time. Moreover, for performance analysis, trains are considered to be delayed if they arrive more than five minutes after their scheduled arrival times. Figure 6 shows the distribution of train delays in the first state space representation for one stopping pattern with scheduled stops in Tianjin South and Chanzhou West.

The second state space definition represents the bounded set of integers. Such Markov chain is simulated by sequentially generating the increments $\Delta_{i,i+1}$ for each event transition $i \rightarrow i+1$ and using the recursion $X_{i+1} = X_i + \Delta_{i,i+1}$ to compute the delays of succeeding events. We consider the increment $\Delta_{i,i+1}$ to be distributed dependent on the time i and state X_i. This formulation is therefore equivalent to the Markov chain formulation presented in Section 3.

The analysis of the data set revealed that 95% of the data are in the interval [-5,5] minutes. The distribution of the data points for the integer state space can be seen in histograms given in Figures 4 and 5. No delays lower than -5 were registered. The number of data points for each delay value exceeding 5 minutes observed separately is too small to make reliable transition probability estimates. For that reason, all delays larger than 5 minutes are clustered and treated as a single state that represents large delays.

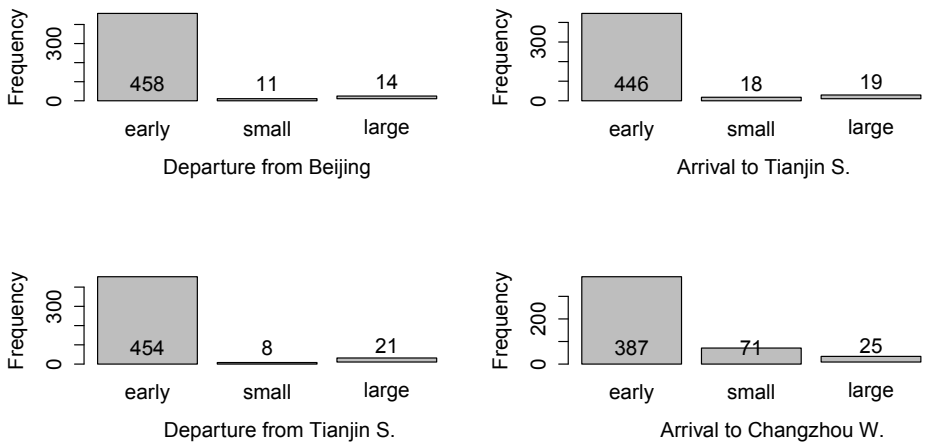

Figure 6: Overview of the number of data points used to derive transition probabilities

5.3 Computation of one-step transition matrices

The next step was to compute the one-step transition matrices of going from one state to another between two successive events. The matrices are computed as conditional probabilities from the data. Data are first converted to fit the above described state space formulations for each train line (stopping pattern) separately. Finally, the conditional probability transition matrix is computed for each process empirically from the data. This procedure is, for the sake of clarity, described for the smaller state space containing only three states: 'early', 'small' and 'large'. The run of G south bound trains that had a scheduled stop in Tianjin South and Cangzhou West (Figure 3) is represented in Figure 7. It is visible from the figure that there are four possible transitions.

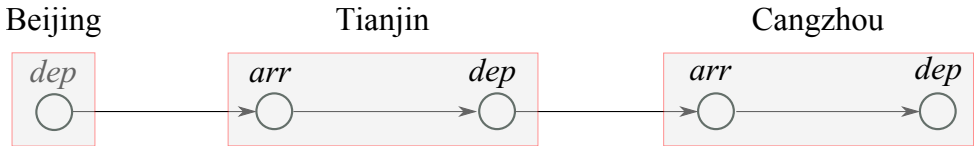

Figure 7: Representation of the example train run as a discrete sequence of events

The matrices computed from the data set are the following:

Departure Bejing - Arrival Tianjin =

$$\begin{array}{c} \\ early \\ small \\ large \end{array} \begin{pmatrix} early & small & large \\ 0.97 & 0.02 & 0.01 \\ 0.45 & 0.45 & 0.10 \\ 0.00 & 0.14 & 0.86 \end{pmatrix}$$

Arrival Tianjin - Departure Tianjin =

$$\begin{array}{c} \\ early \\ small \\ large \end{array} \begin{pmatrix} early & small & large \\ 0.99 & 0.00 & 0.01 \\ 0.67 & 0.28 & 0.05 \\ 0 & 0.11 & 0.89 \end{pmatrix}$$

Departure Tianjin - Arrival Cangzhou =

$$\begin{array}{c} \\ early \\ small \\ large \end{array} \begin{pmatrix} early & small & large \\ 0.85 & 0.15 & 0.00 \\ 0.00 & 0.50 & 0.50 \\ 0.00 & 0.00 & 1.00 \end{pmatrix}$$

Arrival Cangzhou - Departure Cangzhou =

$$\begin{array}{c} \\ early \\ small \\ large \end{array} \begin{pmatrix} early & small & large \\ 0.99 & 0.00 & 0.01 \\ 0.52 & 0.46 & 0.02 \\ 0.00 & 0.12 & 0.88 \end{pmatrix}$$

Closer analysis of the matrices reveals that the diagonal probabilities, i.e., the probabilities of keeping the current states are particularly high. For the early trains this implies that the extra time is not used for energy efficient driving. By running at lower speeds, such trains could consume less energy while still maintaining their schedule. On the other hand, for the delayed trains, this fact indicates that the time reserves implemented in the timetable may not be sufficiently large to make up or decrease their departure delays by running at the maximum speed.

An important aspect that needs to be considered for verifying the correctness of such matrices is the number of data points used to derive each probability, in other words, how many data points were used to compute each row in the matrices. Figure 6 gives this information. The plots in the figure can be interpreted in the following way. For example the lower left plot in the figure gives the overview of departure delays from Tianjin. The graph reveals that the first row of the matrix Departure Tianjin - Arrival Cangzhou was computed based on 454 observations, while the second row was computed based on only 8 observations. From all graphs it seems that only the first row in each matrix (transition from early events) is computed based on the sufficient amount of observations. This implies that a larger data set is required to maintain the validity of the approach and avoid the aggregation of data which may be a major source of imprecision.

6 Results

A non-stationary Markov chain is fully defined with a transition matrix for each process in the sequence. The test data set was used to verify the accuracy of the described approach. Each train run from the test set was simulated separately. When information about an event is received, its delay is known with certainty and can thus be represented with a one-point distribution. The probability distributions of the remaining events in the sequence is updated recursively by sampling from the distributions in the corresponding rows of transition matrices. An example of dynamic updates of probability distributions in real time is given in Figure 8. Time is given on the x axis of the figure. The planned sequence of events is represented on the y axis. The realisation times of four events are captured. When an event occurrence is registered, the probability distribution of each future event in the sequence (given above the realised event) is updated. Therefore, each row in the graph gives the evolution of the probability distribution that is updated in discrete moments in time. As it was expected after analysing the data set, the occurrence of early events is very probable. However, after an arrival to Chengzhou with a small delay, probability distribution of the subsequent departure delay is significantly changed.

The accuracy of the dynamic stochastic prediction method is tested on a test set comprising 20% of train runs. Figure 9 shows the average accuracy of stochastic predictions for the first definition of the state space that classifies predicted delays in to three categories: early, small delay and large delay. The horizontal axis represents the prediction horizon which tells how far in advance was prediction performed. The vertical axis gives the prediction error. Since the state space definition implies classification of predicted delays into one of three classes, a coding system has been implemented in order to quantify the classification error. The three sets are coded with values 0,1 and 2 for early, small and large delay, respectively. A classification error is then computed as an absolute value of error. The figure shows the beneficial effect of implementing the dynamic algorithm that uses the real-time information to modify the probability distributions of future events. Static prediction is performed by using the fixed probability distributions computed offline based on the training data set. The advantage of dynamic predictions is visible for all considered prediction horizons.

Figure 10 shows the prediction accuracy of stochastic dynamic predictions for state space S_2 that considers integer delay values in full minutes. Dynamic predictions outperform the static case in terms of accuracy for every considered prediction horizon. Moreover, the accuracy of predictions decreases monotonically with the length of the prediction horizon. Prediction errors smaller than two minutes on average are achieved even for the longest prediction horizon of two hours. The prediction error is decreasing with a higher rate for shorter prediction horizons, thus indicating the increase in importance of the real-time information that is received close to the actual event realisation. This accuracy is still lower than the accuracy of predictions obtained using the deterministic prediction performed with frequent updates of train positions [18]. However, the low granularity of historical data used in this study prevented the application of a detailed model.

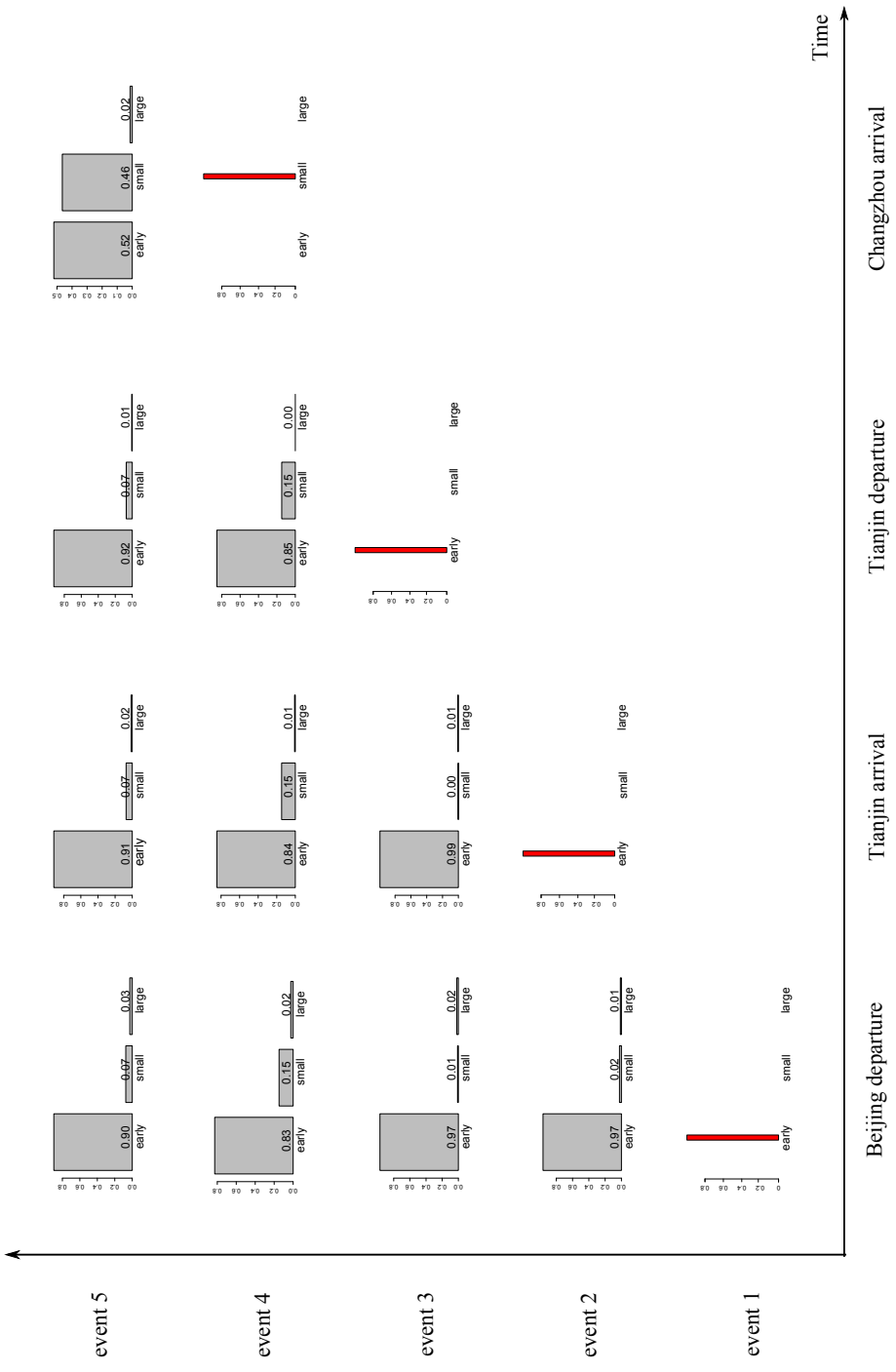

Figure 8: Evolution of probability distributions over time

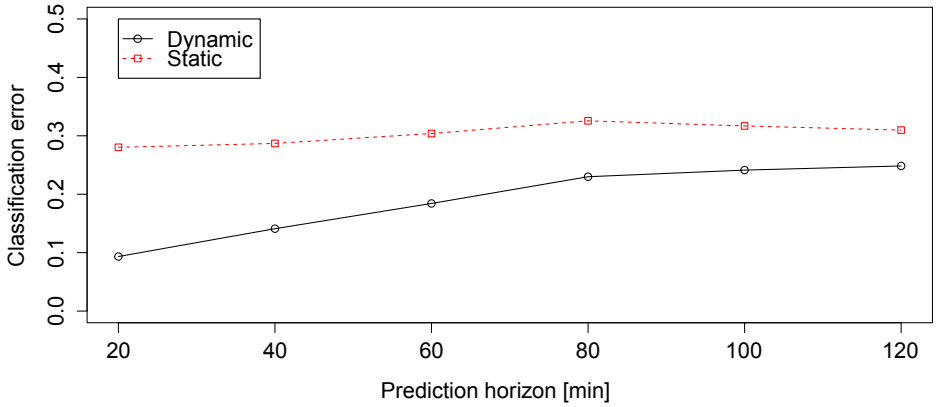

Figure 9: Prediction accuracy for state space $S_1 = \{\text{'ealy'}, \text{'small'}, \text{'large'}\}$

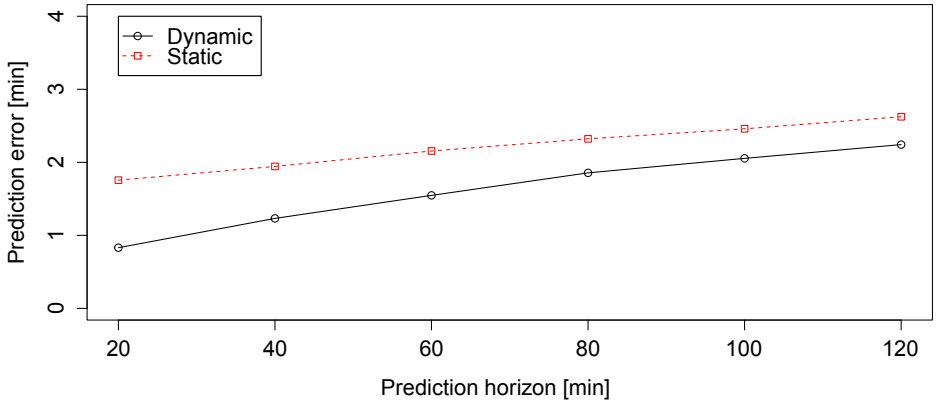

Figure 10: Prediction accuracy for state space $S_2 = \mathbb{Z}$

7 Summary of the main findings

This paper presented an analysis of stochastic dynamics phenomena performed in order to characterize the effect that the prediction horizon and incoming information about a running train may have on the future traffic evolution. The presented method is important for making better predictions for train traffic, that are not only based on static, offline collected data, but are able to positively include the dynamic characteristics of the ever-changing delays, thus increasing the reliability of prediction by 71%.

Having a better estimate of train delays could be greatly beneficial for validation and evaluation purposes of the state-of-the-art online traffic models. In particular, this approach enables the estimation of delay dynamics for the closed-loop [29, 4], online rescheduling [14, 1] and simulation [24, 28] tools. Moreover, the practical application of this method could increase the amount of information delivered to passengers, in the form of up-to-date probability for on-time arrival. It is shown by many policy studies that an informed passenger is way more likely to accept this delay, and giving probability margins will be an extra feature of projected travel time planners. Being able to characterize, analyse and predict the unavoidable dynamic uncertainty of process times can also result in better railway traffic planning and control.

As for the future works, modelling the dynamic interrelation of delays in a closed form would have the positive effect to allow for direct inclusion in planning and rules-of-thumb formulas, as well as allow generalization over different patterns of networks, train traffic, congestion. Furthermore, having available large amount of data would allow finer estimations of the parameters, and for instance being able to distinguish between off-peak dynamics, on-peak dynamics. This distinction can be further extended to within-day characteristics, for instance related to the passenger load, and within-year characteristics, for instance taking into account seasonal effect, bad weather, leaves on tracks.

Finally, the approach proposed is basically data driven and thus directly applicable to other networks, operated by trains, or sharing key operational characteristics. Thus a similar approach can be certainly of interest in transport networks, where delays are unavoidable, and spread easily over time and space, such as air traffic, maritime traffic, supply chains and freight flows.

Acknowledgments
This study has been carried out as a part of EU funded Capacity4Rail project with the support from the Knut and Alice Wallenberg Foundation, State Key Laboratory of Rail Traffic Control and Safety (RCS2013ZZ001), National Natural Science Foundation of China under Contract (No. 71201009) and Beijing Higher Education Young Elite Teacher Project (No. YETP0581).

References

[1] Bauer, R. and Schöbel, A. Rules of thumb: practical online-strategies for delay management. *Public Transport*, 6(1–2):85–105, 2014.

[2] Berger, A., Gebhardt, A., Müller-Hannemann, M., and Ostrowski, M. Stochastic Delay Prediction in Large Train Networks. In Caprara, A. and Kontogiannis, S. (eds.),

11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems, pp. 100–111, Dagstuhl, 2011.

[3] Büker, T. and Seybold, B. Stochastic modelling of delay propagation in large networks. *Journal of Rail Transport Planning & Management*, 2(1-2):34–50, 2012.

[4] Caimi, G., Fuchsberger, M., Laumanns, M., and Lüthi, M. A model predictive control approach for discrete-time rescheduling in complex central railway station areas. *Computers & Operations Research*, 39(11):2578–2593, 2012.

[5] Carey, M. and Kwieciński, A. Properties of expected costs and performance measures in stochastic models of scheduled transport. *European Journal of Operational Research*, 83:182–199, 1995.

[6] Corman, F., D'Ariano, A., Pacciarelli, D., and Pranzo, M. Dispatching and coordination in multi-area railway traffic management. *Computers & Operations Research*, 44:146 – 160, 2014.

[7] Corman, F. and Meng, L. A review of online dynamic models and algorithms for railway traffic rescheduling. *Intelligent Transportation Systems, IEEE Transactions on*, 99:1–11, 2014. to appear.

[8] D'Ariano, A., Pacciarelli, D., and Pranzo, M. A branch and bound algorithm for scheduling trains in a railway network. *European Journal of Operational Research*, 183(2):643–657, 2007.

[9] D'Ariano, A., Pranzo, M., and Hansen, I. A. Conflict Resolution and Train Speed Coordination for Solving Real-Time Timetable Perturbations. *IEEE Transactions on Intelligent Transportation Systems*, 8(2):208–222, 2007.

[10] Dolder, U., Krista, M., and Voelcker, M. RCS – Rail Control System – Realtime train run simulation and conflict detection on a net wide scale based on updated train positions. In *Proceedings of the 3rd International Seminar on Railway Operations Modelling and Analysis (RailZurich2009)*, pp. 1–15, Zurich, 2009.

[11] Dollevoet, T., Huisman, D., Kroon, L. G., Schmidt, M., and Schöbel, A. Delay management including capacities of stations. *Transportation Science*, 2014. to appear.

[12] Ethier, S. N. and Kurtz, T. G. *Markov processes: characterization and convergence*. John Wiley & Sons, Hoboken, NJ, 2009.

[13] Fukami, K. and Yamamoto, H. A new diagram forecasting system for the Tokaido-Sanyo Shinkansen. In *Proceedings of the World Congress on Railway Research*, pp. 1–6, Köln, 2001.

[14] Gatto, M., Jacob, R., Peeters, L., and Widmayer, P. Online delay management on a single train line. In Geraets, F., Kroon, L., Schöbel, A., Wagner, D., and Zaroliagis, C. (eds.), *Algorithmic Methods for Railway Optimization*, vol. 4359 of *Lecture Notes in Computer Science*, pp. 306–320. Springer Berlin Heidelberg, 2007.

[15] Goverde, R. M. P. and Hansen, I. A. Performance indicators for railway timetables. In *Intelligent Rail Transportation (ICIRT), 2013 IEEE International Conference on*, pp. 301–306, Aug 2013.

[16] Hansen, I. A., Goverde, R. M. P., and Van der Meer, D. J. Online train delay recognition and running time prediction. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pp. 1783–1788, Madeira, 2010.

[17] Kecman, P. *Models for predictive railway traffic management*. Phd thesis, Delft University of Technology, Delft, 2014.

[18] Kecman, P. and Goverde, R. M. P. Online data-driven adaptive prediction of train event times. *Intelligent Transportation Systems, IEEE Transactions on*, (99), 2014. to appear.

[19] Liu, J. S. *Monte Carlo Strategies in Scientific Computing*. Springer series in statistics. Springer Science+Business Media, New York, 2008.

[20] Longo, G., Medeossi, G., and Nash, A. Estimating train motion using detailed sensor data. In *Transportation Research Board 91st Annual Meeting*, pp. 1–15, Washington, 2012.

[21] Medeossi, G., Longo, G., and de Fabris, S. A method for using stochastic blocking times to improve timetable planning. *Journal of Rail Transport Planning & Management*, 1(1):1–13, 2011.

[22] Meester, L. E. and Muns, S. Stochastic delay propagation in railway networks and phase-type distributions. *Transportation Research Part B: Methodological*, 41(2):218–230, 2007.

[23] Meng, L. and Zhou, X. Simultaneous train rerouting and rescheduling on an N-track network: A model reformulation with network-based cumulative flow variables. *Transportation Research Part B: Methodological*, 67(0):208 – 234, 2014.

[24] Nash, A. and Huerlimann, D. Railroad simulation using OpenTrack. In Allan, J., Brebbia, C. A., Hill, R. J., Sciutto, G., and Sone, S. (eds.), *Computers in Railways IX*, pp. 45–54, Southampton, 2004. WIT Press.

[25] Nielsen, L. K., Kroon, L., and Maróti, G. A rolling horizon approach for disruption management of railway rolling stock. *European Journal of Operational Research*, 220:496–509, 2012.

[26] Parzen, E. *Stochastic processes*. Society for Industrial and Applied Mathematics, Philadelphia, 1999.

[27] Potthoff, D. *Railway Crew Recheduling: Novel Approaches and Extensions*. Ph.D. thesis, Erasmus University Rotterdam, Rotterdam, 2010.

[28] Quaglietta, E. A simulation-based approach for the optimal design of signalling block layout in railway networks. *Simulation Modelling Practice and Theory*, 46:4–24, 2014. Simulation-Optimization of Complex Systems: Methods and Applications.

[29] Quaglietta, E., Corman, F., and Goverde, R. M. P. Stability analysis of railway dispatching plans in a stochastic and dynamic environment. *Journal of Rail Transport Planning & Management*, 3(4):137–149, 2013.

[30] Yuan, J. and Hansen, I. A. Optimizing capacity utilization of stations by estimating knock-on train delays. *Transportation Research Part B: Methodological*, 41(2):202–217, 2007.