

Article

Train Delay Predictions Using Markov Chains Based on Process Time Deviations and Elastic State Boundaries

Thomas Spanninger * , Beda Büchel  and Francesco Corman Institute for Transport Planning and Systems (IVT), ETH Zurich, Stefano-Franscini-Platz 5,
8093 Zurich, Switzerland

* Correspondence: thomas.spanninger@ivt.baug.ethz.ch

Abstract: Train delays are inconvenient for passengers and major problems in railway operations. When delays occur, it is vital to provide timely information to passengers regarding delays at their departing, interchanging, and final stations. Furthermore, real-time traffic control requires information on how delays propagate throughout the network. Among a multitude of applied models to predict train delays, Markov chains have proven to be stochastic benchmark approaches due to their simplicity, interpretability, and solid performances. In this study, we introduce an advanced Markov chain setting to predict train delays using historical train operation data. Therefore, we applied Markov chains based on process time deviations instead of absolute delays and we relaxed commonly used stationarity assumptions for transition probabilities in terms of direction, train line, and location. Additionally, we defined the state space elastically and analyzed the benefit of an increasing state space dimension. We show (via a test case in the Swiss railway network) that our proposed advanced Markov chain model achieves a prediction accuracy gain of 56% in terms of mean absolute error (MAE) compared to state-of-the-art Markov chain models based on absolute delays. We also illustrate the prediction performance advantages of our proposed model in the case of training data sparsity.

Keywords: train delays; prediction; stochastic; Markov chain

MSC: 60J20



Citation: Spanninger, T.; Büchel, B.; Corman, F. Train Delay Predictions Using Markov Chains Based on Process Time Deviations and Elastic State Boundaries. *Mathematics* **2023**, *11*, 839. <https://doi.org/10.3390/math11040839>

Academic Editor: Nadir Farhi

Received: 10 December 2022

Revised: 31 January 2023

Accepted: 2 February 2023

Published: 7 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Railway systems suffer from train delays, defined as deviations of the observed operations from the planned timetable [1]. These unplanned deviations are inconvenient for passengers due to the resulting loss of time and increased anxiety. In case a train is delayed, it is of great interest for railway operators to predict the consequences of this delay to minimize the spread of the delay (i.e., knock-on delays [2]) and inform passengers about the expected delays to their final destination [3]. Therefore, accurate delay predictions in railway operations are essential [4].

Various approaches have been addressed to predict train delays. Among them, Markov chain (MC) models have been specifically applied to analyze and predict the evolution of train delays along their journey. Within the Markov chain framework, delays at certain points along a train's journey (typically arrival and departure events) are modeled as stochastic processes. Delays are classified into state and transition probabilities and describe the probability of transitioning from one state to another (from one point of the journey to the consequent point). The MC model is based on the name-giving Markov property (i.e., the assumption that the current state solely depends on the most recent state—earlier states of the process are irrelevant). For this reason, the MC model is simple to implement, provides an interpretable framework to assess the quality of a railway timetable, and allows for predicting the effects of real-time delays [5]. In spite of its simple framework,

the MC framework contains a lot of flexibility in its application of modeling the evolution of train delays.

In this study, we aimed to predict the delay of a train at its scheduled arrival or departure, given its current delay in the moment of departure or arrival, using a set of historical delay observations (and schedules). For this purpose, MC models have been outperformed lately by far more complex machine learning models (e.g., [6–8]) in terms of prediction accuracy. To some extent, this is due to the rigid implementations of MC models that do not allow exploiting the specific differences between the locations in the distribution of historical train operation data. In this study, we propose an advanced and more flexible Markov chain model for train delay prediction that keeps most of its major strength in terms of simplicity and increases the prediction accuracy significantly to state-of-the-art Markov chain applications. Therefore, we constructed a discrete Markov chain model that

- Is based on process time deviations as underlying variables instead of absolute delays.
- Relaxes the commonly used stationarity assumption for Markov chains (i.e., state transition probabilities are equal for all points of the stochastic process).
- Is built on an elastic definition of the state space (i.e., state boundaries can vary in absolute terms for different points of the stochastic process).

This study also comprehensively analyzes the impact of the dimension of the state space on the prediction accuracy of train delays. Hence, we formalized a more flexible Markov chain setting to predict train delays with significant improvement in prediction accuracy compared to state-of-the-art Markov chain approaches. In this context, we introduce the innovative approach of an elastic state space definition, analyze the effects of the choice of the underlying variable (delays at events or deviations of process times), and investigate the prediction accuracy with respect to the dimension of the state space.

The reasons for using the process deviations rather than the absolute deviations are multiple. Absolute train delays are typical key performance indicators of railway operations and, therefore, are the immediate targets of many modeling and prediction approaches. Most of the literature studies discuss how absolute train delays suffer from skewed distributions, which complicates the definition of large delays. Those are relatively infrequent but are relevant in delay prediction tasks. In [3], the variations in process times were shown to be relatively small and well approximated by non-skewed distributions. We, thus, focus on the incremental variations of delays, at the process levels. The unavoidable drawback is that the evolution of the system over multiple steps cannot be elegantly computed by a matrix multiplication representing Chapman–Kolmogorov equation types. We instead use a simple algorithm that determines the probability of being at any state. Moreover, we show in the results that such an approach gives good insight; and leads to good computational performance.

Our main findings include the advantage of using process time deviations as underlying variables to gain prediction accuracy. The prediction accuracy can further be improved by our introduced elastic state space structure, mainly because of a much higher exploitation of the elastic state space compared to the static state space. This is especially relevant in the case of training data sparsity because Markov chain models with elastically defined state spaces need less training data to achieve their highest prediction accuracies compared to Markov chain models based on statically defined states.

This work is beneficial for researchers as well as for practitioners as it shows the capabilities and limits of extending traditional applications of Markov chains to predict and analyze the propagation of train delays. Railway operators can use our proposed model for punctuality analyses, timetable stability assessments, and to analyze the effects of infrastructural changes (e.g., blocked track segments or potential infrastructure investments).

This paper is further structured into a literature review on the application of Markov chains to predict train delays in Section 2. We then present the methodology of Markov chains in Section 3 followed by details of our application on a busy railway corridor in Switzerland (in Section 4). Section 5 analyzes the performance of our introduced advanced Markov chain model and is followed by a discussion about the strengths and weaknesses

of the proposed approach in Section 6. Finally, Section 7 summarizes our findings and provides potential future research topics in this field.

2. Literature Review

2.1. Train Delay Predictions

Approaches for train delay prediction can be classified according to their underlying modeling paradigms, i.e., data-driven or event-driven [9]. Event-driven approaches explicitly model the dynamics within the prediction horizon. Most event-driven approaches generate stochastic predictions (i.e., probability distributions of future delays), while data-driven approaches typically result in single-value predictions.

Early contributions of data-driven approaches to predict train delays are based on queuing models [10,11] and linear regression models (e.g., [12,13]). Phase-type distributions are used by [11] to estimate delay propagation. Recently, multiple contributions used machine learning approaches to predict train delays, including neural networks (e.g., [6,7,14,15]), decision tree models, random forest models (e.g., [8,16,17]), and support vector machine models (e.g., [18–20]).

Concerning event-driven approaches, early publications on train delay prediction were based on general graph models (e.g., [2,21,22]) and aimed to describe the uncertainties of future delays by assuming and fitting probability distributions [23–25]. An event-activity graph is used by [23] to stochastically model delay propagation on a network level based on an expansion of exponential polynomials as flexible (i.e., easy to convolute) distribution functions. The theory of the max-plus algebra was introduced by [26] to model the train delay propagation for periodical timetables, and extended by [27]. A few studies were conducted based on Markov chain models [5,28–33], which we will review in detail in Section 2.2. Recently, Bayesian networks have been used intensively as part of a combination of event-based modeling and machine learning techniques to increase the performances of train delay predictions [3,34,35]. Moreover, Petri net models have been applied to predict train delays [36,37].

Among the variety of applied models to predict train delays, complex machine learning techniques might provide the most accurate predictions [9]. Compared to event-driven models, however, machine learning techniques typically provide single-valued predictions and lack interpretability in terms of the possibility of back-tracing predictions.

2.2. Markov Chain Models

Markov chains are a subset of stochastic processes describing the evolution of a variable over time. Markov chain models assume that the probability of the value of the variable only depends on the most recent value. This assumption intuitively holds for the evolution of delay along a train's journey and has been studied by [30].

As delays in railways are typically measured at stations (departure and arrival delays), most approaches consider discrete Markov chains (i.e., random variables are described at specific moments in time) to model and predict the delay evolution. However, also continuous Markov models (i.e., the random variable is described continuously in time) was studied by [38] in the context of train delay prediction.

In [28], the authors considered the sequence of arrival delays of consecutive trains at a certain station as a stochastic process and modeled this sequence as the Markov chain. They used their MC model to forecast the evolution of delays of freight trains. Specifically, they showed where delays were expected to be amplified and reduced.

Typically, however, the delays of a specific train along its journey are considered part of the stochastic process to be modeled. Therefore, in [29], the authors distinguished between early, small-delayed, and largely delayed trains to predict the evolution of delays along train journeys. Also, in [5], a stationary Markov chain model (i.e., transition probabilities are assumed to be the same for all steps of the stochastic process; more details in what follows) was applied to analyze the timetable stability of the Turkish railway network and

to predict the effects of running time supplements on section runs as well as buffer times at stations on the conflict resolution of delays.

The benefits of higher-order MC models to predict train delays—where not only the very last but the previous k delay observations are assumed to affect the consequent delay—were studied by [30]. They concluded (within their application environment on the Indian railway network) that a first-order Markov chain mostly delivers the most accurate prediction results. In [38], the authors estimated Markov transition intensities depending on the influences of weather indicators to provide a measure for the risk of a train falling into the state of being delayed along its journey.

A continuous state space approach based on the Markov property for stochastic processes was presented by [31]. To overcome the limitation of the discretization of the state space, they modeled delays as continuous variables and proposed simplifying assumptions for the evolution of the probability over time. In [32], the authors provided an MC model in the context of train delay predictions using process time deviations as variables and showed that this modification of the state space outperforms usual MC models with state spaces of absolute delays at events in terms of prediction accuracy. Moreover, [33] is based on a stationary Markov chain model used to analyze and predict the delay evolution for multiple steps along a train's journey. By relaxing the widely used stationary assumption for Markov chains, the authors of [39] achieved significant improvements in prediction accuracy.

All contributions of Markov chain models in the prediction of train delays are based on the following four model design elements: (1) definition of underlying variable, (2) dimension of the state space, (3) structure of the state space, (4) stationarity assumption/relaxation.

Concerning (1), the first step of the application of any Markov chain model is to define the variable of the stochastic process. In terms of predicting train delays, this can be the absolute amount of delays at the arrival, departure, and passing event of the train, or it can be the deviation of the planned and realized process time for consecutive running and dwelling processes.

For specifications (2) and (3): Delays or process time deviations are classified into states (i.e., in classes of time intervals). The amounts of those states are the dimensions of the state space. If the boundaries of these states are defined manually by expert domain knowledge, we call it a *classic* structure of the state space, as this procedure is most common in the literature. If the absolute values of the state boundaries are defined equally throughout all points of observation of the underlying variables by statistics and data analytics, we call it a *static* structure of the state space (so a classic state space structure is a special form of a static structure). In case the absolute values of the state boundaries differ throughout the points of the stochastic process, we call it an *elastic* structure of the state space.

A Markov chain is called *stationary* (4) if the transition probabilities are the same throughout the evolution of the stochastic process. Regarding train delays, the stationarity assumption is equivalent to the assumption that the probabilities of acquiring delay or making up delay are the same (homogeneous) for all processes of a train. As trains move in time and space, we can identify processes and steps linking them; those are ordered in both space (stations along a train's journey) and time (the time it takes to reach those stations). Therefore, the stationarity of the process is defined in an abstract way with regard to time, space, and ordered events in the train plan. Given a railway network with multiple train journeys, these transition probabilities might be different (inhomogeneous) for different processes. The non-stationarity discussed in the present paper enters via inhomogeneity of parameterizations for different events/processes (referring to different train lines, directions, and locations).

In this study, we want to distinguish the processes in terms of direction (going from station A to station B, there is a different transition probability than going from station B to station A), train line (a train line is defined by a unique stopping pattern), and location. Any of these three relaxations of homogeneity assumption for transition probabilities results in a non-stationary Markov chain and, naturally, any combination of these relaxations

is conceivable. Table 1 summarizes the definitions and assumptions of the existing MC applications to predict train delays.

Table 1. Specifications and assumptions of Markov chain model-based contributions in the context of train delay predictions.

Study	Test Case	Variable	Structure	Dimension	Stationary	Dir i.h. ¹	Line i.h. ²	Loc i.h. ³
[28]	Hupac network (Europe)	Events	classic	5	no	no	yes	yes
[29]	Beijing-Shanghai (China), 1318 km, 24 stations	Events	classic	3/10	no	no	yes	yes
[5]	Istanbul-Ankara (Turkey), 150 km single-track line	Events	classic	6	yes	no	no	no
[38]	Umeå-Stockholm (Sweden), 711 km	Events	static	2	yes	no	no	no
[31]	Zurich-Chur (Switzerland), 95 km, 40 stations	Events	static	1	yes	no	no	no
[32]	St. Margrethen-Buchs (Switzerland), 40 km, 9 stations	Processes	elastic	5	no	no	yes	yes
[33]	Arifiye-Çukurhisar (Turkey), 162 km, 18 stations	Events	classic	27	yes	no	no	no
[39]	Leiden-Breda/Roosendaal (Netherlands), 112 km, 26 stations	Events	classic	8	no	yes	yes	yes
[40]	Haarlem-Den Haag & Amstertdam Centraal-Almere (The Netherlands)	Events	static	7	no	yes	yes	yes
This work	Zurich-Chur (Switzerland), 95 km, 40 stations	Processes	elastic	2-30	no	yes	yes	yes

¹ Direction-inhomogeneous. ² Line-inhomogeneous. ³ Location-inhomogeneous.

The review of the existing literature in the context of Markov chain model applications to predict train delays shows differences between the model’s design specifications and assumptions. To the best of our knowledge, no researchers have analyzed the relaxation of all three homogeneity assumptions together with an elastic structure of the state space. Additionally, the importance of the dimension of the state space for the prediction accuracy has received little attention. Through this flexibilization, we want to gain deeper insight into the dynamics of delay propagation and we expect a higher prediction accuracy.

We aim to ‘flexibilize’ the application of Markov chains to predict train delays to the maximum in terms of the underlying variables, the relaxation of stationarity assumptions, and the structure and dimension of the state space. We provide analyses in which the setting of modeling options results in the best prediction accuracy and we give estimations for the importance and risks of possible assumptions/simplifications.

3. Methodology

3.1. Stochastic Process

A stochastic process $(X_t \mid t \in \mathbb{T})$ is a family of real-valued random variables, defined on a shared probability space (Ω, \mathcal{F}, P) , with the state space Ω , the σ -algebra \mathcal{F} (containing all potential concatenations of states of Ω) and the probability measure P , which assigns each element of \mathcal{F} its probability in $[0, 1]$.

In this study, we consider the evolution of absolute delays or process time deviations as random variables. The evolution is described along consecutive steps as the phenomenon we want to model. Those steps vary in time and space along the journey of a vehicle. We assume to observe these variables at specific moments, when processes, such as arrival and departures, are performed. Those moments are uniquely ordered along a time axis as the vehicles move in the network, according to a pre-specified plan, i.e., schedule. Train services typically repeat every day. We assume the runs of the same train services in the same direction, across multiple days, describe the same steps, and can be well-modeled as the same phenomena. For each train service, all of the steps happen at successive time instances. Hence, one could think of \mathbb{T} as the discretized set of moments in time $\mathbb{T} = \{t_1, t_2, \dots, t_z\}$ when we can observe delay/process time observations.

These moments in time are not necessarily distributed equidistantly. In general, those considered observation *steps* are part of unified terminology, modeling details of event-based (absolute delays) and process-based (process time deviations). $X_t \rightarrow X_{t+1}$, consequently, represents one step of the stochastic process. Together, the stochastic process can be seen as mapping from (Ω, \mathbb{T}) to \mathbb{R} :

$$\begin{aligned} X &: (\Omega, \mathbb{T}) \rightarrow \mathbb{R} \\ (\omega, t) &\rightarrow X_t(\omega) \end{aligned}$$

3.2. State Space

We classify the continuous variables of delay or process time deviation into n states. Those states are collected in the set $\Omega = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_n\}$, ranging from defined state boundaries $b_i \in \mathbb{R}$ to $b_{i+1} \in \mathbb{R}$, respectively, for $i = 1, \dots, n$. Note that b_1 can also be $-\infty$ and b_{n+1} can also be ∞ . The dimension n of the state space Ω (i.e., number of states) is assumed to be the same for all points of the stochastic process. Nevertheless, dimension n can vary as the parameter of our Markov chain model implementation for different experiments.

We introduce three model design options for the state space structure:

- *Classic Structure*
For the classic state space structure, the state boundaries are manually defined equally for all events (processes) by expert domain knowledge within the range of historical delay observations. Mostly, integers are used as state boundaries. In some cases, states are also defined according to analytical goals (e.g., to analyze how many trains are delayed by more than 5 min, it is reasonable to define one state boundary at the level of 5 min).
- *Static Structure*
For the static state space structure, we assigned quantiles q_x of the empirical distribution of all observed delays (process time deviations) of the training data to the state boundaries b_i , for $i = 1, \dots, n + 1$ to spread the quantity of observations per state evenly among the states of delay (process time deviations). To be more precise, we define the state boundaries with respect to the dimension n of the state space Ω as in Equation (1).

$$b_1 = q_0, \text{ and } b_i = q_{\frac{i+1}{n}}, \text{ for } i = 2, \dots, n + 1 \tag{1}$$

Note that it takes $n + 1$ boundaries to define n states of delay (process time deviation) in order to well-define the state space. For this reason, the lowest state boundary b_1 is defined as q_0 of the empirical distribution, which corresponds to the least observed delay (the deviation for the fastest realized process time) and b_{n+1} is defined as q_1 , corresponding to the highest delay (deviation for the slowest realized process time).

- *Elastic Structure*
The elastic state space structure is defined analogously to the static structure. In the elastic case, however, the delay (process time deviation) quantiles are done specifically for every event (process) with respect to the non-stationarity assumptions per direction, line, and location. Consequently, the absolute values of the state boundaries might vary from event (process) to event (process). The aim here is to incorporate direction-, line-, and location-specific peculiarities of transition probabilities (e.g., as a result of scheduled buffer times or running time supplements).

All three structures of the state space imply the same amount of states for all points of observation $\mathbb{T} = \{t_1, t_2, \dots, t_z\}$ of the stochastic process. Consequently, the proposed models are in line with the common literature on Markov chains based on a finite state space. As shown in Table 1, most approaches are based on the classic structure definition of the state space. Examples can be found in (e.g., [5,29,39]).

3.3. Markov Property

The Markov property assumption for stochastic processes implies that the probability of X at time t only depends on the value of X at time $t - 1$. Regarding the evolution of train delays (process time deviations), this corresponds to the delay (process time deviation) of a train at a certain event (process), which only depends on the delay (process time deviation) at the most recent event (process) of this train.

Due to assuming the Markov property for a stochastic process, the probability distribution of X_t conditioned on the history of the stochastic process (i.e., $X_1, \dots, X_{t-2}, X_{t-1}$) can be simplified as in Equation (2). For this reason, Markov chains are often called *memory-less*.

$$P(X_t = S_i | X_1 = S_j, X_2 = S_k, \dots, X_{t-1} = S_l) = P(X_t = S_i | X_{t-1} = S_l) \tag{2}$$

In other words, a Markov chain can be defined as a stochastic process $(X_t | t \in \mathbb{T})$, where X_1 follows an initial probability distribution λ and for all $t > 1$, the conditional probability of X_{t+1} only depends on the probability distribution of X_t and not on X_1, \dots, X_{t-1} . As λ needs to be defined only for the very first point of the stochastic process (corresponding to a train’s initial departure delay or first running time deviation), it only serves as input in our experimental setup. Therefore, λ will always have the form of a unit vector with zeros and one for the state of delay (process time deviation) that is observed for the first point of the stochastic process and used as input for the first one-step prediction.

3.4. Transitions

The main advantage and core part of a Markov chain model is its simple yet very powerful ability to describe the probability of transitions in-between states per step of time ($t \rightarrow t + 1$) of the stochastic process $(X_t | t \in \mathbb{T})$. This can be done by setting up a transition probability matrix (TPM). Such a TPM matrix is of dimension n^2 and every element p_{ij} represents the probability of a state change from delay state S_i to delay state S_j . Hence, the TPM matrix is also very often referred to as a **one-step transition matrix** in the context of Markov chain models.

$$\text{TPM} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{bmatrix} \tag{3}$$

Every row of the TPM represents the discrete probability distribution of a state transition from a specific state (corresponding to the row index) to any other state (corresponding to the column index). Hence, the sum of every row of the TPM needs to equal 1, i.e., $\sum_{j=1}^n p_{ij} = 1$.

In a simple case of three states of delay (e.g., early, on time, and delayed),

$$\Omega = (S_1, S_2, S_3) = \{(-\infty, 0); [0, 3); [3, \infty)\}$$

the corresponding Markov chain’s transition probability matrix can be visualized as the directed graph (see Figure 1).

Markov chains that assume the same TPM for all time steps $X_t \rightarrow X_{t+1}$, $t \in \mathbb{T}$ are called *stationary*. In our study of delay development along a train’s journey, however, we assume differences (inhomogeneities) for different directions, lines, and locations. Hence, we propose a *non-stationary* Markov chain but also compare accuracy results to its stationary counterpart.

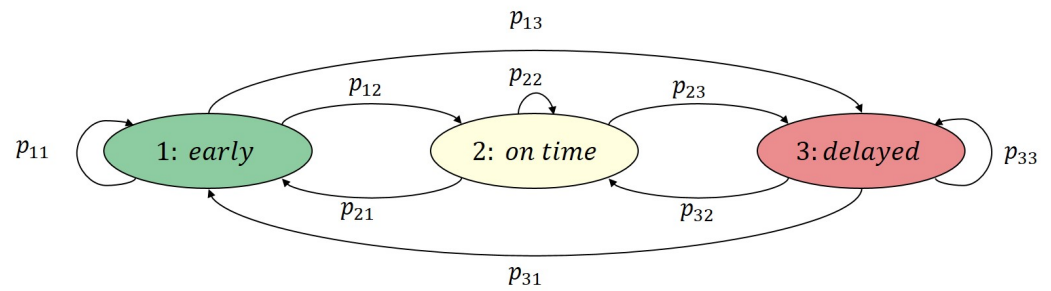


Figure 1. Visualization of a Markov chain for train delays with three states of delay.

Consequently, the one-step transition matrices (of transition probabilities p_{ij}) of the non-stationary Markov chain models are specified by the upper indices D for the direction within the corridor; L for the train line; and A and B for the initial station (section) and final station (section). Label A, B describes the locations and represents a unique step in the well-defined order of events or processes along a train service in a specific direction and, therefore, encodes all temporal aspects.

$$\text{TPM}(D, L, A, B) = \begin{bmatrix} p_{11}^{D,L,A,B} & p_{12}^{D,L,A,B} & \dots & p_{1n}^{D,L,A,B} \\ p_{21}^{D,L,A,B} & p_{22}^{D,L,A,B} & \dots & p_{2n}^{D,L,A,B} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1}^{D,L,A,B} & p_{n2}^{D,L,A,B} & \dots & p_{nn}^{D,L,A,B} \end{bmatrix} \tag{4}$$

3.5. Underlying Variable

According to Section 2, most approaches use the absolute departure delay (DD) and arrival delay (AD) as the underlying variable. Delay is defined as the deviation of the realized departure or arrival time from the scheduled departure or arrival time. Two consecutive scheduled departure (arrival) and arrival (departure) times determine the corresponding scheduled process time for the respective running (dwelling) process in-between. We call the realized deviation from this scheduled process time the process time deviation, or more precisely, running time deviation (rtd) and dwelling time deviation (dtd).

Analogous to the absolute delay variable, the process time deviation can be interpreted as a random variable. When process time deviations are considered to be the underlying variables of a Markov chain, the non-stationary levels, A and B , of the TPMs correspond to an initial process A and a final process B . Equations (5) and (6) point out the dualism of delays and process time deviations.

$$DD_t + rtd_{t+1} = AD_{t+1} \tag{5}$$

$$AD_t + dtd_{t+1} = DD_{t+1} \tag{6}$$

3.6. Dependency Assumptions

Following the idea in [32], we built our Markov chain model in the underlying variable of the process time deviation. We, therefore, refer to it as the MC-P model. As shown in Section 2, most applications of Markov chain models have used the absolute delay amounts (at arrivals and departures), which we will refer to as the MC-E model for comparison reasons.

To specify the dependencies in our MC-P model, we assume that the running (dwelling) time deviations are dependent on previous running (dwelling) time deviations. In other terms, the process time deviation is used to predict the next process time deviation, and not directly the delay. This means that we set up two parallel Markov chains to model the operations of a single train, as visualized in Figure 2. The indices of DD , AD , rtd , and dtd indicate the position of the consecutive departure and arrival events, as well as running

and dwelling processes. Once all the processes upstream of an event have been described, one can predict the delays of one train journey, by combining all terminal states weighted by their probabilities.

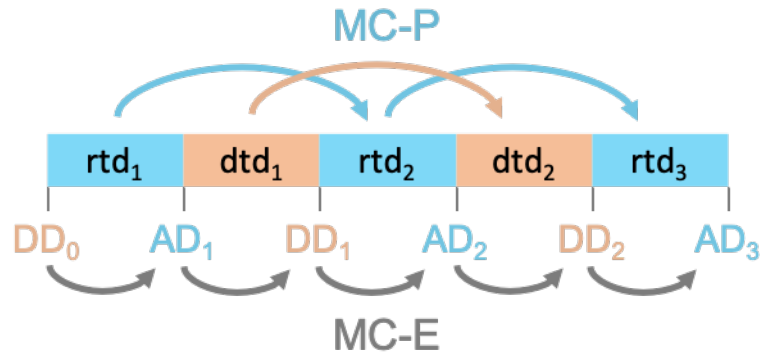


Figure 2. Visualization of dependency structures of the MC-E and MC-P model.

The interpretation of states and transitions in the context of the MC-P model is not straightforward as for the classic MC-E models. In MC-P models, states can be seen as categories of how much faster (or slower) a train completes a running or dwelling process with respect to its planned process time. Consequently, a transition can be interpreted as a faster (slower) than planned train remaining faster (slower) than planned or changing its process time deviation state from one point of the stochastic process to the next one.

The choice of the underlying variable essentially determines the assumed correlations on which the consequently constructed Markov chain model is built. While Markov chains based on absolute delays exploit the correlation of delay changes for consecutive train events, Markov chains based on process time deviations aim to exploit patterns of systematic, sequential deviations from planned operational durations.

3.7. Parameter Estimation

To estimate the parameters of the proposed Markov chain models, (i.e., all entries $p_{ij}^{D,L,A,B}$ of all TPMs) based on historical observations of delays/process time deviations, we use the maximum likelihood estimation (MLE), dating back to [41]. As a result of the Markov-properly assumption the MLE needs to be done separately for every TPMs (D,L,A,B), $d \in D, l \in L, a \in A, b \in B$. Therefore, we separate all historic train movement observations of the training data according to the direction (D), train line (L), and initial station/process (A) to the final station/process (B).

Without loss of generality, we omit the indices d, l, a, b for the following description of the MLE procedure. The concept of the MLE is to find the set of parameters $\theta = \{p_{ij} | i = 1, \dots, n; j = 1, \dots, n\}$ that makes the observed data samples of state transitions $X = \{X_1, X_2, \dots, X_N\}$ most likely. The MLE is based on the assumption that all observations are independent and identically distributed (IID). In our study, this assumption is equivalent to assuming that every train with delay/process time deviation state i faces the same probability distribution to transition to another (or the same) state during the next step of the stochastic process.

Following the IID assumption of the training data, the joint probability of all state transition observations X_1, X_2, \dots, X_N is the product of all corresponding probabilities. Let N_{ij} be the number of observations with initial state i and final state j within X . Then, the joint probability of the training data $F(X)$ can be written as

$$F(X) = \prod_{i=1}^n \prod_{j=1}^n p_{ij}^{N_{ij}} \tag{7}$$

Interpreting the joint probability distribution $F(X)$ as function of the parameter set $\theta = \{p_{ij} | i = 1, \dots, n\}$ gives the likelihood function

$$L(\theta) = \prod_{i=1}^n \prod_{j=1}^n p_{ij}^{N_{ij}} \tag{8}$$

To obtain the maximum likelihood estimates for all $p_{ij} | i = 1, \dots, n \& j = 1, \dots, n$; we have to maximize the likelihood function $L(\theta)$. As every row of a TPM corresponds to a well defined probability distribution by $\sum_{j=1}^n p_{ij} = 1, \forall i = 1, \dots, n$. To simplify the optimization, we maximize the log transformation of $L(\theta)$. As a result of the monotonic property of the logarithm, the maximizing values of θ for $L(\theta)$ and $\log L(\theta)$ are the same. Adding n Lagrangian multipliers $\lambda_i | i = 1, \dots, n$ and their respective constraints results in the log-likelihood function $\mathcal{L}(\theta)$, given in Equation (10).

$$\mathcal{L}(\theta) = \log(L(\theta)) - \sum_{i=1}^n \lambda_i (\sum_{j=1}^n p_{ij} - 1) \tag{9}$$

$$= \sum_{i=1}^n \sum_{j=1}^n N_{ij} \log(p_{ij}) - \sum_{i=1}^n \lambda_i (\sum_{j=1}^n p_{ij} - 1) \tag{10}$$

We determine the maximum of the log-likelihood function $\mathcal{L}(\theta)$ by setting the derivatives with respect to $p_{ij} \forall i = 1, \dots, n \& j = 1, \dots, n$ and $\lambda_i \forall i = 1, \dots, n$ to zero. Setting the derivatives with respect to p_{ij} to zero leads to n^2 conditions given in Equation (12).

$$\frac{\partial \mathcal{L}}{\partial p_{ij}} = \frac{N_{ij}}{p_{ij}} - \lambda_i = 0 \tag{11}$$

$$\Leftrightarrow p_{ij} = \frac{N_{ij}}{\lambda_i} \tag{12}$$

Setting to zero the derivatives with respect to $\lambda_i \forall i = 1, \dots, n$; and using Equation (12) leads to n conditions given in Equation (15).

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \sum_{j=1}^n p_{ij} - 1 = 0 \tag{13}$$

$$\sum_{j=1}^n \frac{N_{ij}}{\lambda_i} - 1 = 0 \tag{14}$$

$$\Leftrightarrow \sum_{j=1}^n N_{ij} = \lambda_i \tag{15}$$

Using the result of Equation (15) $\forall i = 1, \dots, n$ again in Equation (12) leads to the maximum likelihood estimates, given by Equation (16).

$$p_{ij} = \frac{N_{ij}}{\sum_{j=1}^n N_{ij}} \tag{16}$$

Consequently, the estimation of the one-step transition probabilities with respect to direction (D), train line (L), initial station/process (A), to the final station/process (B) is given in Equation (17), where we use $N_i = \sum_{j=1}^n N_{ij}$.

$$p_{ij}^{D,L,A,B} = \frac{N_{ij}^{D,L,A,B}}{N_i^{D,L,A,B}} \tag{17}$$

3.8. Prediction

The aim of this study is to generate one-step predictions of train delays in real-time for a train's consequent arrival delay AD_{i+1} given its current departure delay DD_i or for a train's consequent departure delay DD_{i+1} given its current arrival delay AD_i . According to the underlying variable (absolute delays or process time deviations), the generation of predictions varies slightly.

Regarding the MC-E model, the current departure (arrival) delay DD_i (AD_i) is classified into delay state x with respect to the classically, statically, or elastically determined state boundaries. Consequently, the probability distribution of the respective row of the apposite TPM provides the one-step prediction for the consequent departure (arrival) delay state, $F(x)$, given the current delay state x . In terms of the elastic state space structure, the final state boundaries can be different in absolute values compared to initial state boundaries.

Regarding the MC-P model, the most recent running (dwelling) time deviation rtd_i (dtd_i) is classified into the state of running time deviation \hat{x} according to the classically, statically, or elastically defined process time deviation boundaries. The row corresponding to \hat{x} of the apposite TPM for the MC-P step $rtd_i \rightarrow rtd_{i+1}$ ($dtd_i \rightarrow dtd_{i+1}$) provides the probability distribution function $\tilde{F}(\hat{x})$ for the state of the consequent running (dwelling) process time deviation. To further obtain predictions for the consequent arrival (departure) delay out of the probability distribution for the consequent running (dwelling) process time deviation, we use the duality of delays and process time deviations given in Equations (5) and (6). Mathematically, this corresponds to a transformation $T(\cdot)$ of the state boundaries of process time deviation to state boundaries of delay by adding the current departure (arrival) delay DD_i (AD_i) to the state boundaries of the consequent process time deviation. In this way, we obtain the probability distribution for the consequent arrival (departure) delay $F(\hat{x}) = T(\tilde{F}(\hat{x}))$, given the most recent running (dwelling) time deviation. The probabilities for state transitions are unaffected by $T(\cdot)$.

The stochastic prediction output of both models (MC-E and MC-P) is a discrete probability distribution on the state of delay (dimension n) for the consequent event. Note that for the MC-P model, the first arrival and second departure event of a train journey cannot be predicted as those processes do not have a previous process of their kind. Therefore, we exclude those events from the test set for comparison reasons between the MC-E and MC-P models.

As we will discuss in Section 5, we need to provide single-valued predictions for future delays y for the reason of comparability. Therefore, we calculate the expected value of y , $\mathbb{E}_F(y)$, given our generated probabilistic prediction function $F(\cdot)$. By definition, the expected value $\mathbb{E}_F(y)$ is the sum of the products of the probabilities of states times their representing value. As states are defined as intervals (e.g., $[b_i, b_{i+1}]$), we need to define values as state representatives (i.e., values between b_i and b_{i+1}). We decided to use the median of delay observations/process time deviations that fall within the respective states as representative values. Other considered options include the mean value of the respective observations or the mean of the interval. As the mean of the interval is not defined according to historic observations, its risk of inheriting a bias is the largest. While the difference between the mean and median of the respective state observations might be minimal for internal states, the mean value of observations also inherits a risk of bias to represent a state for the lowest and highest state. For these reasons, we decided to use the medians of respective state observations as state representatives to calculate $\mathbb{E}_F(y)$ as a single-valued breakdown of $F(\cdot)$.

4. Application on Test Case

We test the above-discussed Markov chain models on the very busy railway corridor of the Swiss railway network from *Zurich HB* to *Chur*. This 95 km corridor is mostly double-track and has 34 stations. A wide range of 14 distinct train lines (defined by unique stopping patterns) are scheduled in this corridor (at least to some extent). Therefore, it is

well-suited for testing the differences between the model’s design elements. The line plan for all train lines of the corridor from Zurich to Chur is visualized in Figure 3.

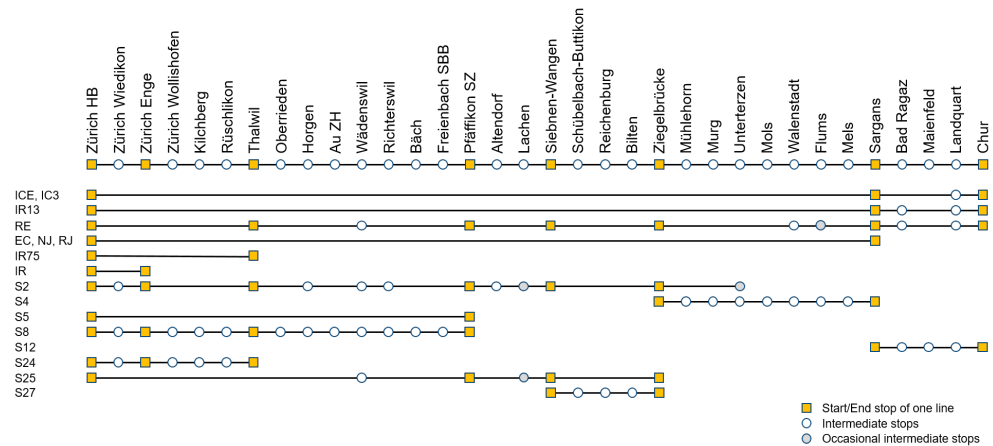


Figure 3. Line plan of services within the corridor Zurich–Chur.

4.1. Data

We used nine months of historical train operation data from January 2021 to September 2021, which are publicly available (<https://opentransportdata.swiss/en/>, accessed on 1 June 2022). The experimental setup consists of three phases: training, prediction, and evaluation phase. The training phase basically consists of estimating the TPMs(D, L, A, B) for all values of D, L, A, B . We used eight months of historical train observation data from January to August 2021, consisting of more than 1.3 million delay observations at departures and arrivals within the corridor of investigation. The estimated TPMs are consequently used in the predicting phase to generate one-step predictions of state transitions. To be more precise, we generate a prediction of delay for every departure (except the initial) and arrival of every train run within the corridor in September 2021 using the train’s previous delay as input.

Figure 4 provides visualizations of the empirical distributions of the observed delays and process time deviations per event type (process type) of the data used for training. We can observe that the majority of delay observations fall within the range of -2 to 4 min of delay. Figure 4a shows that departure events tend to happen with more delay compared to arrival events. The majority of process time deviations fall within the range of -3 to 2 min. The visualization of Figure 4b clearly shows that trains typically acquire delays during dwelling processes and can make up for delays during running processes.

4.2. State Space Structure

In line with common approaches for the classical state space structure of the MC-E and descriptive analyses of the delays in the training data, we (arbitrarily) define the state boundaries of the classical structure with 5 states of delay as $-\infty, -2, 0, 2, 4, \infty$. The intervals are skewed to the right to represent the well-known fact of a right skew of event delays, due to (among other things) minimum departure times imposed at stations. Similarly, for the MC-P model, we assign the state boundaries to $-\infty, -3, -1, 1, 3, \infty$.

Regarding the static state space structure, we need to calibrate the state boundaries that are the same for the delays (process time deviations) at all events (processes) according to the quantiles of the empirical distributions of all observed delays (process time deviations). For the MC-E (MC-P) model with five states of delay (process time deviation), the state boundaries are defined by the quantiles $q_0, q_{1/6}, q_{2/6}, \dots, q_{5/6}$, and $q_{6/6}$. As there is no reason to assume the test data cannot exceed the range of observed delays (process time deviations) of the training data, we set the lowest and highest state boundary to $+\infty$ and $-\infty$ respectively. Table 2 shows the calibrated boundaries of the classic and static state space structure for the MC-E and MC-P models with five states of delay or process time deviation.

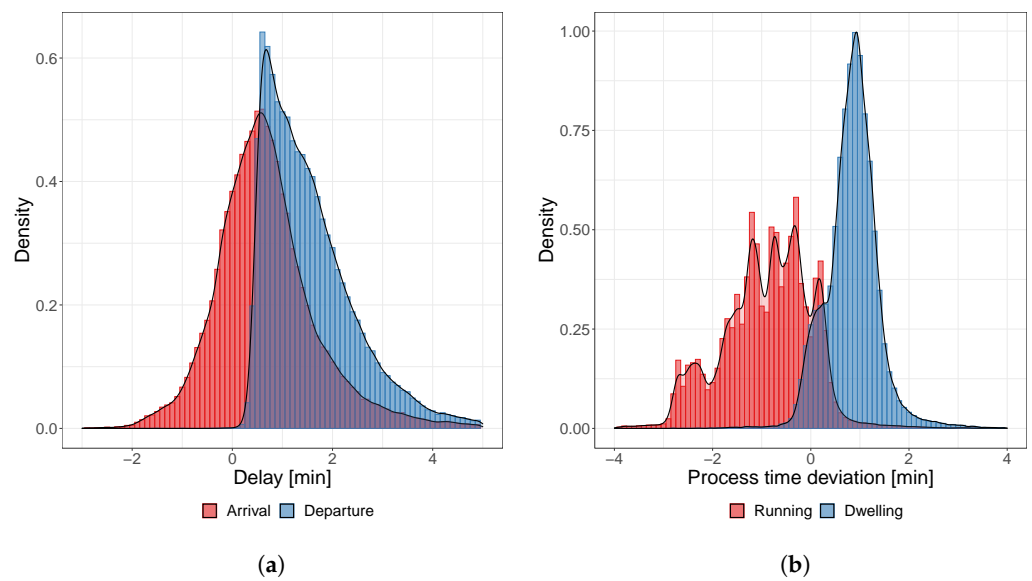


Figure 4. Histograms and empirical density estimations for delays and process time deviations. (a) Delay as the underlying variable of the MC-E model. (b) Process time deviation as the underlying variable of the MC-P model.

Table 2. Calibrated state boundaries [min] for the MC-E and MC-P models (i.e., based on delays at events or process time deviations of processes) with classic and static state space structures and five states of delay.

State Boundary	Events/Classic	Events/Static	Processes/Classic	Processes/Static
State Boundary 1	$-\infty$	$-\infty$	$-\infty$	$-\infty$
State Boundary 2	-2	0.45	-3	-1.1
State Boundary 3	0	0.85	-1	-0.31
State Boundary 4	2	1.32	1	0.32
State Boundary 5	4	2.05	3	0.92
State Boundary 6	$+\infty$	$+\infty$	$+\infty$	$+\infty$

For our introduced elastic state space structure, the state boundaries correspond to event-specific absolute values of delay in terms of the MC-E model, and process-specific process time deviations in terms of the MC-P model. Figure 5 shows the distribution of the calibrated elastically defined state boundaries of arrival delays (running process time deviations) for the MC-E (MC-P) model respectively—computed over all lines, directions and locations—with a state space dimension of five.

Figure 5 reveals a significantly higher variability within the absolute values of the MC-E state boundaries compared to the MC-P state boundaries. This is especially true for the highest (sixth) state boundary. Most likely, this is due to the fact that once a train has acquired a large delay, this delay is observed at multiple downstream arrival and departure events along its journey. However, in comparison, after a single large process time deviation corresponding to this primary delay, the following running times do not necessarily have to be longer or shorter than scheduled.

Moreover, the initial acquisition of a high delay can occur along a train’s journey before the train enters the corridor of our investigation. In this case, higher absolute delays would be observed for this train but process time deviations are not necessarily affected. For these reasons, the highest delay state boundary of the MC-P model is closer to the corresponding MC-P state boundaries.

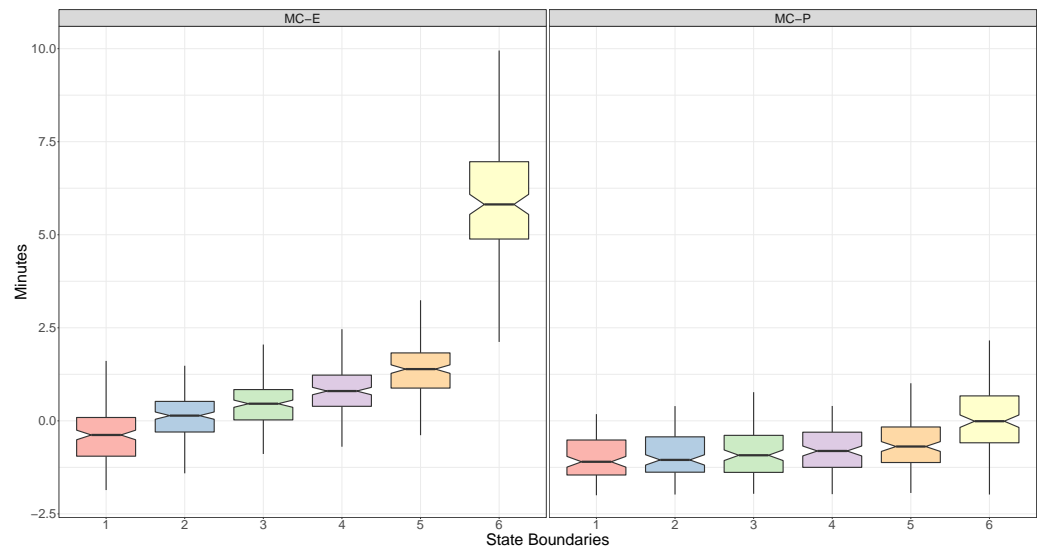


Figure 5. Distribution of elastically defined and calibrated state boundaries for the MC-E (left) and MC-P (right) with state space dimension five.

4.3. Transition Probability Matrices

We estimate the parameters for all one-step transition probabilities according to the maximum likelihood approach described in Section 3.7 separately for all levels of direction, train line, initial station/process, and final station/process. Figure 6 visualizes the average transition probability (per matrix entry) of the estimated TPMs for the MC-E (left) and MC-P (right) model based on an elastically (top) and statically (bottom) defined state space of dimension five. It points out the interpretability of MC models and their strength in terms of descriptive analytics of train delay propagation.

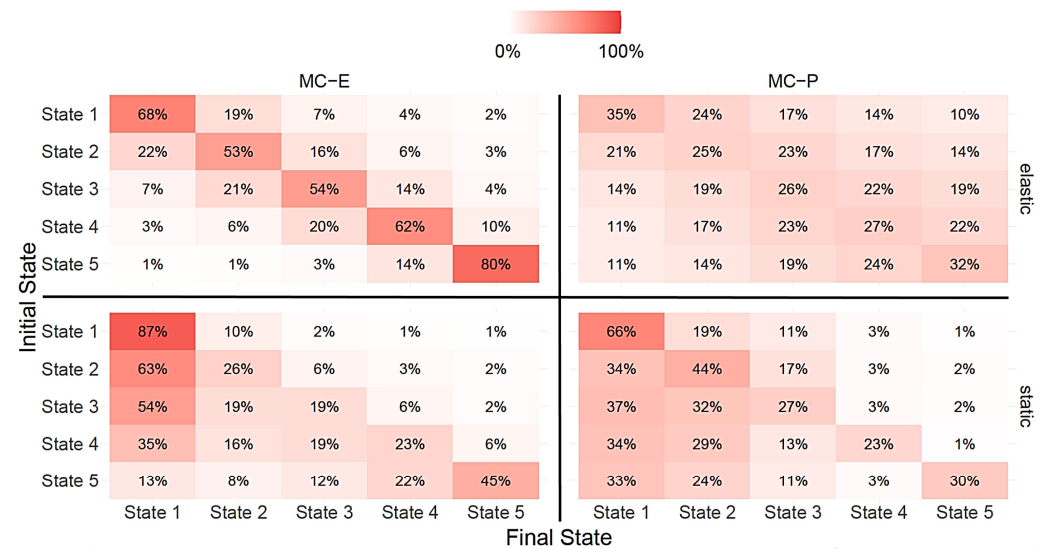


Figure 6. Heatmap of TPMs for the MC-E (left) and MC-P (right) model based on an elastically (top) and statically (bottom) defined state space of dimension five.

We can see in Figure 6 that the weight of the averaged TPMs of the MC-E model is diagonally centered. This is especially obvious in the elastic state space structure but also holds for the static state space structure. Consequently, the MC-E model tends to predict the same final state of delay as the input state. The averaged TPM of the MC-E model based on the static state space structure points out that it is more likely to transition to a lower final state of delay than to a higher state of delay. This clearly indicates that the schedule is

stable in the sense that trains can operate according to their planned times and make up for some acquired delays.

The estimated TPMs in the MC-P models are far less diagonally centered. For the static state space structure, we can observe, similar to the MC-E static averaged TPM, a tendency to transition to a lower state. For the MC-P model based on elastically defined state boundaries, the weight is quite balanced between all matrix positions. This indicates a more equally balanced usage of the state space by the MC-P model with elastic state boundaries in comparison to all other models.

Stationary Markov chains (same TPM for all steps) allow calculating steady-state probabilities $\pi = \{\pi_1, \pi_2, \dots, \pi_n\}$ by solving the linear equation system given in Equations (18) and (19), where $\mathbf{1}$ represents a column vector of dimension n with all entries being one. Steady-state probabilities can be interpreted as long-term behavior of the stochastic process, given that the transition probabilities for every step at the time remain the same.

$$\pi = \pi \times \text{TPM} \tag{18}$$

$$\mathbf{1} = \pi \times \mathbf{1} \tag{19}$$

We can use the averaged TPM of our non-stationary estimations for the transition probabilities in our study to provide a preliminary understanding of our model. This is especially useful for the MC-E model based on a static state space structure because the boundaries of the states remain the same as the values throughout the stochastic process and the probabilities are directly connected to the delays. Using the respective averaged TPM of Figure 6 (bottom left) to solve the equation system of Equations (18) and (19) results in the following steady-state probabilities

$$\pi = \{0.81, 0.12, 0.04, 0.02, 0.01\}. \tag{20}$$

The steady-state probabilities given in Equation (21) point out that the trains in this case study have a strong tendency (81%) to arrive or depart with less than 0.45 min of delay in the long run, have a small chance (12%) of having a steady-state delay between 0.45 and 0.85 min, and only a probability of 7% to arrive or depart with more than 0.85 min of delay.

Using the averaged TPM of the MC-P model with static states (bottom right of Figure 6) results in the steady-state probabilities

$$\pi = \{0.50, 0.29, 0.15, 0.04, 0.02\}. \tag{21}$$

As the MC-P model is built on dependencies of consecutive running time deviations and consecutive dwelling time deviations, these steady-state probabilities do not allow a similar interpretation as for the MC-E model. The significantly lower probability of state one, however, points out that process time deviations cannot only remain in the lowest states in the long run.

Effectively, we deal with relatively short Markov chains. Therefore, the asymptotic behavior can only provide a preliminary understanding of a hypothetical behavior of delay in the long run. Analyzing steady-state probabilities would make more sense if the train operations were coupled at terminal stations to the next run of the same vehicle (for instance if the same vehicle runs back and forth).

4.4. State Space Dimension

In most approaches, the dimension of the state space has been assumed reasonably with respect to observed delays or in the context of a specific target. Consequently, the classic state boundaries have been defined. In our implementation, however, the dimension of the state space n is a model parameter that we vary for different runs of experiments. As the static and elastic state boundaries are defined with respect to n , they adapt automatically within our implementation. A higher dimension of the state space, nevertheless, comes at the price of fewer data per state transition. Additionally, a higher dimension n increases

the computational effort of estimating the TPMs polynomial, because the dimension affects the computational time with order $O(n^2)$.

5. Results

5.1. Performance Evaluation

The evaluation of stochastic predictions is not straight-forward, as single-valued realizations (ground truth) need to be compared with predicted probability distributions. The most common way in literature to deal with that problem is to breakdown the predicted probability distribution to a single value. Typically, the mean value of the distribution is used. While this simplification/breakdown of the predicted probability distribution to a single value seems reasonable, the essence of stochastic models (describing, modeling, and using uncertainty) is entirely lost.

To compare the results of the predictions of our proposed model to other approaches within the field of train delay predictions we, nevertheless, also evaluate the mean absolute error (MAE) and root mean squared error (RMSE). Equations (22) and (23) provide the mathematical definitions of those key performance indicators, where p_i is a single-value prediction and o_i is the realized value of delay of the prediction instance i in a sample of N instances. The MC models compute a probability distribution over a set of states. To compute a single-value prediction p_i from our probability distribution, we use the mean value of the predicted probability distribution.

In total, our experimental setup consists of approximately 75 thousand one-step prediction instances for more than 13 thousand train journeys. These instances can be generated by both, the MC-E and MC-P model. Naturally, the very first departure delay observation is only used as input and no target to predict in real-time. Moreover, by construction, the MC-P model is unable to predict the first arrival and second departure delay of a train journey.

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_i - o_i| \tag{22}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (p_i - o_i)^2}{N}} \tag{23}$$

To address the stochasticity of our predictions in the evaluation, we additionally introduce and use the likeliness of the realization (abbreviated LoR), particularly for cross-comparison of the MC-E and MC-P models based on the elastic and static definition of the state space. The LoR is formally defined in Equation (24). It is calculated by integrating the probability density function f_i of the predicted probability distribution for sample i along a symmetric interval $\mathbb{I}(o_i)$ of length 2π around the realization o_i . In our case, f_i is a piecewise constant function described by the boundaries of the states (on the x-axis), and the probabilities of the states (on the y-axis). Hence, the LoR can be interpreted as the likeliness of the a-posterior realized delay with respect to the generated probabilistic prediction.

$$LoR = \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{I}(o_i)} f_i(x) dx \tag{24}$$

Moreover, we analyze the narrowness of the predicted probability density function f_i by calculating the inner support of f_i , defined as the difference of the highest non-infinite state boundary minus the lowest non-infinite state boundary. For five delay states, this is equivalent to $inner\ support = b_5 - b_2$.

Figure 7 visualizes our suggested evaluation performance measures for an exemplary prediction probability density function f_i of a one-step delay prediction with five states of delay $\Omega = \{S_1, S_2, \dots, S_5\}$. These states S_1 to S_5 are separated by the state boundaries b_2 to b_5 . Note that $b_1 = -\infty$ and $b_6 = \infty$. As defined above, the median of all delay observations within a state is assigned to be state representative for the calculation of the expected value

p_i of the stochastic prediction f_i and o_i is the realized delay observation of the exemplary prediction instance i .

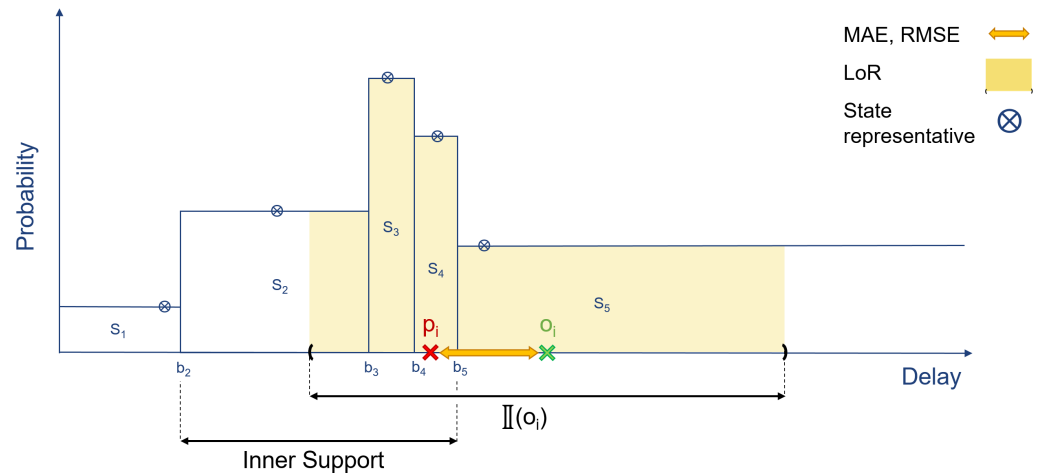


Figure 7. Visualization of evaluation methods for an exemplary probabilistic delay prediction (step-function) with five states.

The definition of the interval $\mathbb{I}(o_i)$ is independent of the dimension of the state space n . Hence, the LoR is useful to cross-compare the prediction performance of MC models with different dimensions of the state space n .

Without loss of generality, we set the interval length 2π of $\mathbb{I}(o_i)$ to 60 s in our test case prediction evaluation. Naturally, the interval length 2π influences the absolute value of the LoR significantly. Nevertheless, for cross-comparisons, we only focus on the relative differences in terms of LoR for different MC models.

5.2. Results

As explained in Section 4, we implemented the Markov chain model based on delays at events (MC-E) and process time deviations (MC-P) for the three introduced state space structure designs classic, static, and elastic. While the state space dimension for the classic structure is fixed at 5, it can vary for the static and elastic structures, where state boundaries are assigned according to quantiles of respective empirical distributions. For this reason, we can test and evaluate the prediction accuracy of the MC-E and MC-P models for statically and elastically defined state boundaries for dimensions 2 to 30. We evaluate all one-step arrival delay predictions that can be made of all models (see Section 3) to guarantee a fair comparison.

Figure 8 visualizes the MAE, RMSE, and LoR for an increasing dimension of the state space Ω for the MC-E (red) and MC-P (blue) model and elastically defined state boundaries (solid line) as well as statically defined state boundaries (dashed line).

Figure 8 contains multiple insights. First, the prediction accuracy increases with respect to an increase in the dimension of the state space Ω for both models and both structures of the state space (MAE and RMSE decrease and the LoR increases). This is interesting because it also points out that the commonly used MC-E model with statically defined state boundaries can increase prediction accuracy.

Second, the MC-P model based on the static definition of the state space (blue and dashed) increases the prediction accuracy significantly in terms of MAE and RMSE in comparison to the MC-E model (red) with statically or elastically defined state spaces.

Third, the prediction accuracy of the MC-P model with elastically defined state boundaries (blue, solid) is higher in comparison to the MC-P model based on static state boundaries—in terms of MAE, RMSE, and especially in terms of LoR. Eventually, the MC-P model with elastic state boundaries can achieve its high prediction performance already

with a small dimension of the state space. This is in contrast to all other MC model implementations of our study, which can all achieve a higher prediction performance with a higher dimension of the state space. However, the evaluation of the LoR points out a slightly increasing prediction performance of the MC-P model as well, with a saturation level of 12 states.

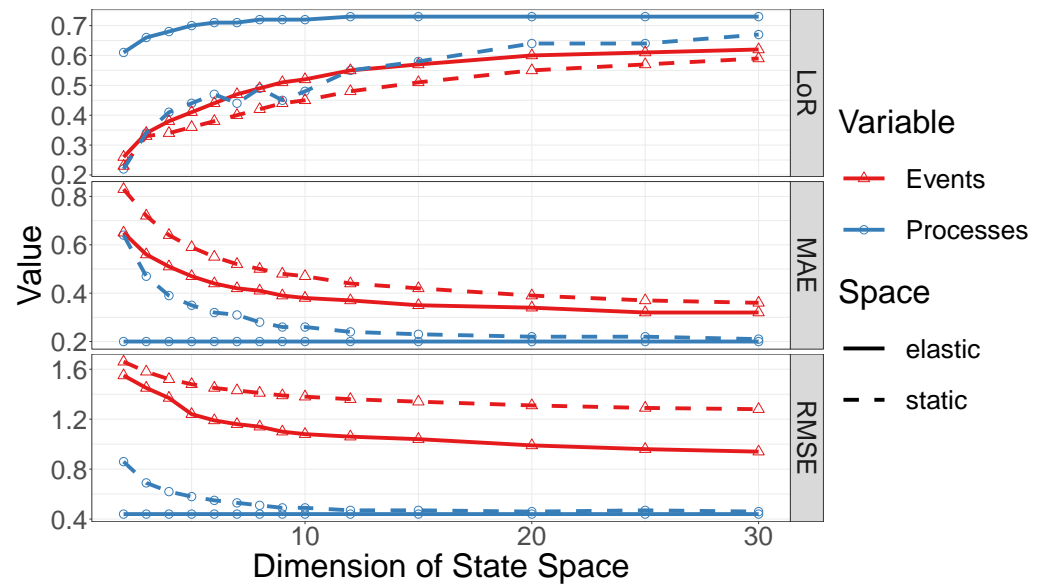


Figure 8. Evaluation of prediction accuracy of the MC-E and MC-P models in terms of MAE, RMSE and LoR (30 s interval) for increasing dimensions of the state space defined elastically (solid) and statically (dashed).

Overall, Figure 8 reveals that our proposed MC-P approach with elastic state boundaries can significantly outperform the benchmark state-of-the-art Markov chain MC-E approach with a classic state space. Table 3 summarizes the performance of the proposed MC-P model based on processes compared to the MC-E model based on events, for classically, statically and elastically defined state boundaries and stationarity assumption or relaxation. All models were evaluated with a state space dimension of five. Table 3 shows that the proposed non-stationary MC-P approach with elastic state boundaries results in an MAE of 0.2 min compared to an MAE of 0.55 min for the non-stationary state-of-the-art MC-E approach with classic state boundaries (a reduction of 56%). Moreover, in terms of the RMSE, the performance increases from 1.30 to 0.44 min (64%). Additionally, we evaluate the inner support (Supp., distance from state boundary 2 to state boundary n , measuring the narrowness of the predicted distribution).

Table 3. Comparison of the Markov chain model prediction performances in terms of MAE [min], RMSE [min], LoR [%], and inner support [min] for a state space dimension of five.

Variable	Structure	Dimension	Stationary	MAE	RMSE	LoR	Supp.
Events	classic	5	yes	0.88	1.51	0.26	6.00
Events	static	5	yes	0.95	1.67	0.24	1.60
Events	elastic	5	yes	0.94	0.67	0.23	1.65
Events	classic	5	no	0.55	1.30	0.30	6.00
Events	static	5	no	0.59	1.48	0.36	1.60
Events	elastic	5	no	0.47	1.24	0.41	1.24
Processes	classic	5	yes	0.90	1.17	0.26	6.00
Processes	static	5	yes	0.83	1.11	0.34	2.02
Processes	elastic	5	yes	0.83	1.10	0.34	2.13
Processes	classic	5	no	0.46	0.61	0.31	6.00
Processes	static	5	no	0.35	0.58	0.44	2.02
Processes	elastic	5	no	0.20	0.44	0.70	0.36

5.3. Data Sparsity

The results in terms of MAE, RMSE, and LoR visualized in Figure 8 indicate that the proposed MC-P approach with elastic state boundaries performs better in cases of few (2 to 10) states in comparison to the MC-P model with static state boundaries. This becomes especially relevant if data for training, calibrating, and estimating TPMs is sparse, because the dimension of the state space implies a data need of $O(n^2)$ to estimate the transition probability matrices.

To analyze the prediction performance with training data sparsity, we run experiments training the proposed non-stationary MC-P model with 12 states (level of saturation of the prediction accuracy in terms of LoR) and elastic state boundaries on a training data set of one up to 20 days of training data and compare it to the classic and static state space structure versions of the MC-P model. Figure 9 shows the prediction performance in terms of MAE and RMSE. It reveals that the prediction performance suffers regardless of the structure of the state space. However, the performance loss of the proposed MC-P model with elastically defined state boundaries is significantly lower compared to its static and classic counterparts, when days for training (i.e., data) become rare.

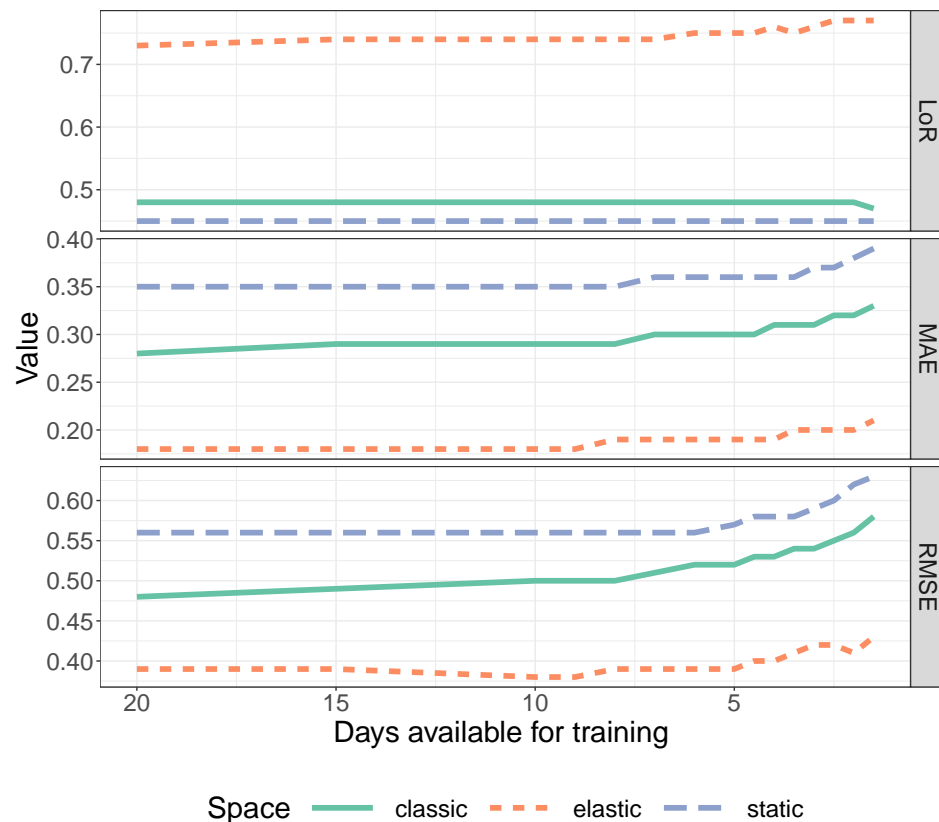


Figure 9. Performance evaluation of the MC-P model based on the classic, static, or elastic state space structure with respect to the availability of data in terms of days of historic train operation observations.

Our analyses on data sparsity also reveal that the prediction accuracy cannot be improved significantly by using more than 15 to 20 days of historical train operation data.

6. Discussion

The main advantage of the proposed MC-P model with an elastic state space to predict train delays in comparison to the classical approach of applying Markov chain models on absolute delays (MC-E) is obviously its much higher prediction accuracy. In contrast to far more complex machine learning approaches, the MC-P model also keeps its strength of

interpretability in terms of descriptive analytics of the dynamics of delay propagation and hence also provides easily understandable predictions.

As we presented in Section 5, the MC-P approach can overcome problems in the context of data sparsity in contrast to MC-E modeling. This is especially useful after changes in the timetable or even more after the introduction of new railway lines and practically no experience with the development of delays for their operation.

The reason for the high prediction accuracy of the MC-P model at a low-state space dimension can be back-traced to a far lower variability of process time deviations compared to absolute delays for different processes (in terms of the assumed inhomogeneities per direction, train line, and location). This is already indicated in Figures 4 and 5. The phenomenon is due to the fact that once a train has acquired a delay somewhere along its journey, only one process time deviation observation is exceptionally high whereas regarding absolute delay observations, the exceptionally high values remain for the rest of the journey until the train is able to make it up (using buffer time allocations). These multiple observations of extraordinarily high delay, however, play a key role in the calibration of the state boundaries and state representatives (median).

The significantly lower variability of the underlying variable in the MC-P model results in considerably narrower support of the probabilistic distribution, especially for a low number of the state space dimension. For this reason, the MC-P model outperforms any MC-E approach already with a low dimension of the state space.

A further advantage of the static and elastic structure of the state space compared to the classic structure is the ability to scale the approach easily throughout the desired number of dimensions of the state space. By definition of the state boundaries via respective quantiles, one does not have to care about defining state boundaries manually for every potential state space dimension.

Nevertheless, the change of the underlying variable from delays to process time deviations carries along to give up the simplicity and straight-forward interpretability of the Markov chain application. Additionally, multi-step predictions for absolute delays further downstream the chain of events cannot be derived that easily with the MC-P model. This is because the multiplication of TPMs for multiple steps results in the transition probabilities of changing from an initial process time deviation state to another process time deviation state later on during the process. However, the information on the development of the absolute delay is lost during that process.

A potential approach to overcome this problem and generate multi-step predictions with the MC-P model is to convolute the corresponding transition probability distributions along the stochastic process. Alternatively, one could think of a hybrid version of the MC-P model for the first step and the MC-E model for the remaining steps to combine the power of the MC-P model for the first step of the prediction and further use the simplicity of the MC-E model to generate predictions for the remaining steps.

7. Conclusions

In this work, we presented an advanced non-stationary Markov chain approach to predict train delays based on process time deviations. We relax commonly assumed stationarity assumptions of transition probabilities in terms of direction, train line, and location. Additionally, we introduced a static and elastic definition of the state space according to respective quantiles of delay and process time deviations. Further, we investigated the prediction quality for a great number of state space dimensions.

We applied and tested our proposed Markov chain setting based on process time deviations (MC-P) on the busy Swiss railway corridor from Zurich to Chur with fourteen different railway lines and more than 1.3 million points of delay observations within 2021.

We can show in our large test case that our proposed MC-P Markov chain model can increase the prediction accuracy by 56% in terms of mean absolute error (MAE) compared to the classic benchmark state-of-the-art Markov chain model based on delays at events

(MC-E) and by 64% in terms of the root-mean-square error (RMSE), all implemented with a state space dimension of five.

Additionally, we introduce the likeliness of the realization (LoR) as a performance indicator to account for the stochasticity in the evaluation of the predictions. The analysis of the LoR for an increasing state dimension reveals that the prediction performance of our proposed MC-P increases in the state space dimension and reaches its saturation level with 12 states in our test case.

Our analyses of the importance of the state space dimension for the MC-E and MC-P models revealed that our proposed MC-P model can already achieve its highest prediction accuracy for a low dimension of the state space (which is not the fact for MC-E models based on absolute delays at events). For this reason, our proposed MC-P model with an elastic state space structure is especially useful in terms of data sparsity (i.e., only a few days of training data are available).

Future research potentials based on the proposed MC-P model can be extensions to multi-step predictions (with or without probability distribution convolution) and the consideration of inter-train dependencies (e.g., headways, conflicts).

Extensions to continuous modeling of the train trajectory, by means of stochastic process modeling (similar to [42]), or the reconstruction of incomplete, noisy, or unknown structures of processes (similar to [43]) could be further relevant extensions.

Author Contributions: Study conception and design: T.S. and B.B. Data processing, analysis, and manuscript preparation: T.S. Interpretation of results and manuscript review: T.S., B.B. and F.C. All authors reviewed the results and approved the final version of the manuscript.

Funding: This work is supported by the Swiss National Science Foundation under Project 1481210/DADA.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in this study of the historical train movement operations are available for the public: (<https://opentransportdata.swiss/en/>, accessed on 1 June 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Zilko, A.A.; Kurowicka, D.; Goverde, R.M. Modeling railway disruption lengths with Copula Bayesian Networks. *Transp. Res. Part C Emerg. Technol.* **2016**, *68*, 350–368. [CrossRef]
- Carey, M.; Kwiecieński, A. Stochastic approximation to the effects of headways on knock-on delays of trains. *Transp. Res. Part B* **1994**, *28*, 251–267. [CrossRef]
- Corman, F.; Kecman, P. Stochastic prediction of train delays in real-time using Bayesian networks. *Transp. Res. Part C Emerg. Technol.* **2018**, *95*, 599–615. [CrossRef]
- Hansen, I.A.; Goverde, R.M.; Van Der Meer, D.J. Online train delay recognition and running time prediction. In Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, 19–22 September 2010; pp. 1783–1788.
- Şahin, İ. Markov chain model for delay distribution in train schedules: Assessing the effectiveness of time allowances. *J. Rail Transp. Plan. Manag.* **2017**, *7*, 101–113. [CrossRef]
- Oneto, L.; Fumeo, E.; Clerico, G.; Canepa, R.; Papa, F.; Dambra, C.; Mazzino, N.; Anguita, D. Train Delay Prediction Systems: A Big Data Analytics Perspective. *Big Data Res.* **2018**, *11*, 54–64. [CrossRef]
- Huang, P.; Wen, C.; Fu, L.; Lessan, J.; Jiang, C. Modeling train operation as sequences: A study of delay prediction with operation and weather data. *Transp. Res. Part E* **2020**, *141*, 102022. [CrossRef]
- Nair, R.; Hoang, T.L.; Laumanns, M.; Chen, B.; Cogill, R.; Szabó, J.; Walter, T. An ensemble prediction model for train delays. *Transp. Res. Part C Emerg. Technol.* **2019**, *104*, 196–209. [CrossRef]
- Spanning, T.; Trivella, A.; Büchel, B.; Corman, F. A review of train delay prediction approaches. *J. Rail Transp. Plan. Manag.* **2022**, *22*, 100312. [CrossRef]
- Greenberg, B.S.; Leachman, R.C.; Wolff, R.W. Predicting Dispatching Delays on a Low Speed, Single Track Railroad. *Transp. Sci.* **1988**, *22*, 31–38. [CrossRef]
- Meester, L.E.; Muns, S. Stochastic delay propagation in railway networks and phase-type distributions. *Transp. Res. Part B Methodol.* **2007**, *41*, 218–230. [CrossRef]

12. Gorman, M.F. Statistical estimation of railroad congestion delay. *Transp. Res. Part E Logist. Transp. Rev.* **2009**, *45*, 446–456. [[CrossRef](#)]
13. Murali, P.; Dessouky, M.; Ordóñez, F.; Palmer, K. A delay estimation technique for single and double-track railroads. *Transp. Res. Part E Logist. Transp. Rev.* **2010**, *46*, 483–495. [[CrossRef](#)]
14. Wen, C.; Mou, W.; Huang, P.; Li, Z. A predictive model of train delays on a railway line. *J. Forecast.* **2019**, *39*, 470–488.
15. Bao, X.; Li, Y.; Li, J.; Shi, R.; Ding, X. Prediction of Train Arrival Delay Using Hybrid ELM-PSO Approach. *J. Adv. Transp.* **2021**, *2021*, 7763126. [[CrossRef](#)]
16. Gao, B.; Ou, D.; Dong, D.; Wu, Y. A Data-Driven Two-Stage Prediction Model for Train Primary-Delay Recovery Time. *Int. J. Softw. Eng. Knowl. Eng.* **2020**, *30*, 921–940. [[CrossRef](#)]
17. Arshad, M.; Ahmed, M. Train Delay Estimation in Indian Railways by Including Weather Factors Through Machine Learning Techniques. *Recent Adv. Comput. Sci. Commun.* **2021**, *14*, 1300–1307. [[CrossRef](#)]
18. Marković, N.; Milinković, S.; Tikhonov, K.S.; Schonfeld, P. Analyzing passenger train arrival delays with support vector regression. *Transp. Res. Part C Emerg. Technol.* **2015**, *56*, 251–262. [[CrossRef](#)]
19. Huang, P.; Wen, C.; Fu, L.; Peng, Q.; Li, Z. A hybrid model to improve the train running time prediction ability during high-speed railway disruptions. *Saf. Sci.* **2020**, *122*, 104510.
20. Wang, Y.; Wen, C.; Huang, P. Predicting the effectiveness of supplement time on delay recoveries: A support vector regression approach. *Int. J. Rail Transp.* **2022**, *10*, 375–392. [[CrossRef](#)]
21. Hollowell, S.F.; Harker, P.T. Predicting on-time line-haul performance in scheduled railroad operations. *Transp. Sci.* **1996**, *30*, 364–378. [[CrossRef](#)]
22. Yuan, J.; Hansen, I.A. Optimizing capacity utilization of stations by estimating knock-on train delays. *Transp. Res. Part B Methodol.* **2007**, *41*, 202–217. [[CrossRef](#)]
23. Büker, T.; Seybold, B. Stochastic modeling of delay propagation in large networks. *J. Rail Transp. Plan. Manag.* **2012**, *2*, 34–50.
24. Keyhani, M.H.; Schnee, M.; Weihe, K.; Zorn, H.P. *Reliability and Delay Distributions of Train Connections*; Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik: Wadern, Germany, 2012.
25. Lemnian, M.; Rückert, R.; Rechner, S.; Blendinger, C.; Müller-Hannemann, M. *Timing of Train Disposition: Towards Early Passenger Rerouting in Case of Delays*; Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik: Wadern, Germany, 2014.
26. Goverde, R.M. A delay propagation algorithm for large-scale railway traffic networks. *Transp. Res. Part C Emerg. Technol.* **2010**, *18*, 269–287.
27. Ma, H.; Qin, Y.; Han, G.; Jia, L.; Zhu, T. *Forecast of Train Delay Propagation Based on Max-Plus Algebra Theory*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 661–672.
28. Barta, J.; Rizzoli, A.E.; Salani, M.; Gambardella, L.M. Statistical modeling of delays in a rail freight transportation network. In Proceedings of the 2012 Winter Simulation Conference (WSC), Berlin, Germany, 9–12 December 2012; pp. 1–12.
29. Kecman, P.; Corman, F.; Meng, L. Train delay evolution as a stochastic process. In Proceedings of the 6th International Conference on Railway Operations Modelling and Analysis—RailTokyo2015, Tokyo, Japan, 23–26 March 2015; pp. 1–27.
30. Gaurav, R.; Srivastava, B. Estimating train delays in a large rail network using a zero shot markov model. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 1221–1226.
31. Spanninger, T.; Büchel, B.; Corman, F. Probabilistic predictions of train delay evolution. In Proceedings of the 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2021, Heraklion, Greece, 16–17 June 2021.
32. Büchel, B.; Spanninger, T.; Corman, F. Modeling evolutionary dynamics of railway delays with Markov chains. In Proceedings of the 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2021, Heraklion, Greece, 16–17 June 2021.
33. Şahin, İ. Data-driven stochastic model for train delay analysis and prediction. *Int. J. Rail Transp.* **2022**, *1*–20. [[CrossRef](#)]
34. Lessan, J.; Fu, L.; Wen, C. A hybrid Bayesian network model for predicting delays in train operations. *Comput. Ind. Eng.* **2019**, *127*, 1214–1222. [[CrossRef](#)]
35. Huang, P.; Lessan, J.; Wen, C.; Peng, Q.; Fu, L.; Li, L.; Xu, X. A Bayesian network model to predict the effects of interruptions on train operations. *Transp. Res. Part C Emerg. Technol.* **2020**, *114*, 338–358. [[CrossRef](#)]
36. Milinković, S.; Marković, M.; Vesković, S.; Ivić, M.; Pavlović, N. A fuzzy Petri net model to estimate train delays. *Simul. Model. Pract. Theory* **2013**, *33*, 144–157. [[CrossRef](#)]
37. Zhuang, H.; Feng, L.; Wen, C.; Peng, Q.; Tang, Q. High-Speed Railway Train Timetable Conflict Prediction Based on Fuzzy Temporal Knowledge Reasoning. *Engineering* **2016**, *2*, 366–373.
38. Wang, J.; Granlöv, M.; Yu, J. Effects of winter climate on delays of high speed passenger trains in Botnia-Atlantica region. *J. Rail Transp. Plan. Manag.* **2021**, *18*, 100251. [[CrossRef](#)]
39. Artan, M.Ş.; Şahin, İ. Exploring Patterns of Train Delay Evolution and Timetable Robustness. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11205–11214. [[CrossRef](#)]
40. Artan, M.Ş.; Şahin, İ. A stochastic model for reliability analysis of periodic train timetables. *Transp. B Transp. Dyn.* **2022**, *1*–18. [[CrossRef](#)]

41. Fisher, R.A. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond. Ser. A* **1922**, *222*, 309–368.
42. Corman, F.; Trivella, A.; Keyvan-Ekbatani, M. Stochastic process in railway traffic flow: Models, methods and implications. *Transp. Res. Part C Emerg. Technol.* **2021**, *128*, 103167. [[CrossRef](#)]
43. Bertuccelli, L.F.; How, J.P. Estimation of non-stationary Markov Chain transition models. In Proceedings of the 2008 47th IEEE Conference on Decision and Control, Cancún, Mexico, 9–11 December 2008; pp. 55–60. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.