# What Do We Know When? Modeling Predictability of Transit Operations

Beda Büchel and Francesco Corman

***Abstract*— Predictions of transit delays are crucial to passengers and operators. Passengers utilize the predictions to decide on departure time, route choice, and mode choice, whereas operators decide on schedules, timetables, rolling stock allocation, and control actions. We introduce the concept of predictability of transit travel times as the study of the reduction of the predicted variability with the temporal approaching of a predicted event. We evaluate predictability on a real-life test case in Zurich, Switzerland, spanning multiple transit lines over one year of operations. The concept is shown based on predictions obtained by a state-of-the-art Bayesian network approach, where we show how predictability (in general) can be modeled as an exponential decay phenomenon. The study of predictability of transit operations leads to additional insights for control actions and system analysis compared to other complementary concepts such as punctuality or regularity, for instance, concerning bunching, identification of bottlenecks, and passenger routing.**

***Index Terms*— Transit operations, predictability, Bayesian network, stochastic predictions, travel time variability.**

## I. Introduction

TRANSIT operations are crucial to welfare in urban networks, providing mobility for a wide range of users. Smart cities are expected to leverage data about past operations to quantify, predict, and especially improve mobility. Transit systems are known to be subject to delays and non-punctuality, which constitutes a gap between passengers' desires, operators' goals, and realized performance. Observed data (running and dwell time, arrival and departure times) have often been used to evaluate punctuality or regularity of operations. When running times deviate from the timetable or show large variability, operations can be managed by inserting appropriate buffer time or control points. In this paper, we focus without loss of generality on urban bus systems.

The importance of accurate delay predictions in transit systems is undisputed. For control purposes (i.e., actions such as holding or short turning), a model of future operations is required [1]. Moreover, accurate predictions of delays, disseminated through real-time information systems, assist passengers in decision making considering route, mode and departure time [2].

Prediction models are getting progressively sophisticated based on analytic paradigms and the increased availability of data. Technologies such as automatic vehicle location (AVL) or identification (AVI) systems have been implemented in transit vehicles all over the world [3].

Passengers and operators need predictions for different purposes and are affected differently from their prediction horizons. We define a prediction horizon as the time period between the time the prediction of an event (e.g., the arrival of a bus) is made and the time the event takes place. For instance, passengers planning their trip in a transit system might be interested in a long prediction horizon (e.g., one day in advance). While en route, they are interested in their arrival time prediction horizon comparable to the length of the trip, say 15-30 minutes in urban areas. At transfer stops, they are interested in the arrival time of the next bus (prediction horizon comparable to headway; $\sim$ 5min in urban areas). On the other side, operators are interested in predictions horizons compatible with possible control actions. For instance, the prediction horizon for inserting additional vehicles from the depot is comparable to the length of a line run ($\sim$ 30 min - 1 hour). Generally, predictions can be made at any desired prediction horizon. Their accuracy is related to the error between the predicted and the realized outcome, aggregated over all the outcomes; this hints to a distribution of error. The smaller the variance of the distribution of the error (sharp), and the smaller the mean of the distribution of the error (unbiased), the more accurate the prediction is. The accuracy of a prediction is assumed to improve with a decreasing prediction horizon.

Here we address the problem of formally describing this phenomenon, propose mathematical models that represent the real-life dynamics that can be estimated from the study of networked systems over time, and explain its relevance.

We, therefore, introduce the concept of predictability, which determines the time horizon at which a prediction becomes sufficiently accurate. The predictability informs about the remaining variability given an event, a prediction model, and observed input data (variability describes the characteristic spread of a phenomenon; variation is its quantification; variance is a specific quantification function). Predictability improves with the temporal approaching of an event.

To quantify it, our framework processes data, considers a stochastic prediction, and repeatedly analyses the prediction over time. The evolution of the predictions is described by a specific model. A stochastic prediction (compared to a point-based estimate as current in the literature) is required, which can dynamically include online, real-time information (compared to not using it or using it only once as a snapshot) to predict bus operations interconnected in a network. We model predictability by the variance of the predicted outcome at

Manuscript received 30 March 2021; revised 17 August 2021, 1 November 2021, and 13 December 2021; accepted 19 January 2022. Date of publication 1 February 2022; date of current version 12 September 2022. The Associate Editor for this article was A. H. Sodhro. *(Corresponding author: Francesco Corman.)*

The authors are with the Institute for Transport Planning and Systems, Swiss Federal Institute of Technology, ETH Zürich, 8092 Zürich, Switzerland (e-mail: corman@ethz.ch).

Digital Object Identifier 10.1109/TITS.2022.3145243

different prediction horizons. We experimentally find a convergence from a maximum to a minimum value, modeled well as an exponential decay.

Predictability allows understanding which temporal and spatial factors have the strongest impact on predictive performance, highlighting the dynamics of the transit system. Predictability is relevant in prescriptive analytics frameworks or for customer information.

The contributions of the paper are as follows:

- The concept of predictability is introduced, which complements the descriptive (punctuality, regularity), predictive (prediction accuracy), and prescriptive analytics (observability, controllability) of networks.
- The predictability fits well an exponential decay based on a state-of-the-art Bayesian network approach.
- Comparing predictability of different events over time and space, transit operations can be analyzed, optimized, and communicated to passengers, for improved performance.

This paper is structured as follows: Section II proposes a literature review; Section III defines our interpretation of predictability. Section IV introduces the prediction model and the framework to analyze predictability. Section V and VI report on the test case and an empirical evaluation. Section VII and VIII discuss applications, and conclusions respectively.

## II. Literature Review

### A. Transit Operations

Transit systems are considered an essential backbone of suitable urban development [1]. Bus operation consists out of alternating running and dwelling processes while a bus travels along a route. Both running and dwell times are subject to stochastic variability (i.e., statistical dispersion), which can be attributed to different factors [4]. Among others, running times are affected by traffic conditions, infrastructure, and driver behavior [5]; dwell times are affected, for instance, by boarding and alighting passengers and vehicle characteristics [6].

Low-frequency transit systems are typically operated according to a timetable; hence punctuality (i.e., the adherence to timetable) is an important aspect of reliability. On the other hand, in high-frequency transit systems, regularity (i.e., adherence to headway) is crucial for reliable operations. To increase transit reliability, it is therefore vital to detect irregular operations [7], [8] and to implement control strategies [1]. These control strategies aim to reduce the variability of departure times or headways and improve the regularity and punctuality of timetable-based operations. Timetable-based bus operations are typically scheduled so that the timetable includes buffer times in addition to the minimal dwell and running times, such that operations are feasible also if a bus is impeded to operate at optimal conditions. At specific stops, called time points, early buses wait to adhere to schedule again. Introducing buffer times forces dealing with the trade-off between minimal travel times and maximal punctuality. The service quality is optimized when, among other indicators, buses show high punctuality (i.e., low deviations from the published timetable) and high reliability (i.e., low variability) [9].

Accurate predictions of delays, given as deviations from a planned timetable, are beneficial to passengers. Real-time information reduces the waiting time uncertainty of passengers [10]. Additionally, passengers can adapt their travel choices in terms of route, departure time, and mode according to the provided information leading to better decision-making [11].

Furthermore, in recent years research focus has been put on reliability issues, "arriving when planned" is an essential desire of transit users, and reliability and variability are critical for analyzing a bus system [12].

### B. Transit Operations

The prediction of bus arrival times or the linked delays got, given its practical usefulness, plenty of research attention in the past decades [13], [14]. Research contributions in bus arrival time predictions can be divided into three areas: route construction, impact factors (related to modeling the processes and their interconnection), and models (related to the algorithmic performance and accuracy) [15].

Routes are most commonly either constructed basing on links (e.g., [16]) or stops (e.g., [17]). Link-based models divide a bus route basing on infrastructure configuration (e.g., intersections), whereas stop-based models divide the bus route basing on stop locations [18]. It has been shown that stop-based models show a higher prediction accuracy, especially if dwell times are explicitly modeled [18], [19].

Most prediction models make use of traffic information either based on real-time data (online, e.g., [15]) or archived data (offline, e.g., [20]). Moreover, weather (e.g., [21]), passenger counts (e.g., [22]), crowdsourced data were used (e.g., [23]).

A variety of statistical and machine learning models have been proposed to predict bus travel times. Early models often applied historical averages (e.g., [19], [24]) or multivariate linear regression (e.g., [25]). Kalman filter models (e.g., [16]), support vector machines (e.g., [16]) or hybrid approaches (e.g., [26]) were applied. In recent years, prediction frameworks basing on artificial neural networks became popular. Various types of neural networks have been proposed for bus arrival time predictions, such as feedforward neural networks (e.g.,.( [27], [28]), recurrent neural networks (e.g., [26], [29]–[31]), or convolutional neural networks (e.g., [17]).

Most of the proposed models predict single values. However, for smart decision-making, it is not only important to have predictions (as point estimates) as accurate as possible but also to quantify their associated variability (statistical dispersion). The variability of a prediction describes how the spread of possible outcomes is. Only a few studies have analyzed it in detail [32]. Formally, the variability of a prediction describes how the spread of possible outcomes is. Measures of variation, e.g., variance or interquartile range, can be used to evaluate it, as far as stochastic prediction techniques are used. For instance, quantile regression [33], relevance vector machine regression [34], or prediction intervals for neural networks [35] have been used to model the variability of bus arrival time predictions. Those studies provide, however, a prediction range rather than a fully-specified distribution.

TABLE I
OVERVIEW OF LITERATURE

| Problem or Approach | Descriptive / Predictive | Deterministic / Stochastic | Processes / Interconnection | Dynamics | References (examples) |
|---|---|---|---|---|---|
| Descriptive reporting (e.g., punctuality, regularity) | D | D | P | - | [5], [7] |
| Online anomaly detection, deterministic delay evolution | D | D | P / (I) | ✓ | [38] |
| Descriptive reporting (e.g., variability, reliability) | D | S | P | - | [39], [40] |
| (interconnected) Stochastic process, evolution of reliability | D | S | P / (I) | ✓ | [36] |
| Domain-unaware deterministic prediction | P | D | - | - | [18]–[20], [41] |
| Domain-aware deterministic prediction | P | D | P / (I) | - | [15]–[17], [22], [23], [26]–[30], [42]–[45] |
| Analysis of prediction horizon | P | D | P | ✓ | [37], [28] |
| Domain-aware stochastic prediction | P | S | P / (I) | - | [35, 46] |
| Bounding prediction error | P | S | P | ✓ | [33] |
| Predictability analysis (model of predicted and reduced variance) | P | S | P/I | ✓ | This paper |

### C. Influence of Prediction Horizon

It is generally reported that the accuracy of arrival predictions decreases with an increasing prediction horizon in terms of space and time (e.g. [10], [24], [36]). The reason for this behavior is that there is greater uncertainty regarding future conditions. However, the error of travel time prediction does not grow linearly. Reference [24] showed that the relative prediction error of a travel time prediction decreases as the remaining travel time increases. Finally, [37] investigated the horizon at which the prediction of the arrival time outperforms averaged past observations, which resulted in being 6km in their specific case.

### D. Research Gap and Contribution

To summarize, various models exist in the literature,

- *describing* past performance, or *predicting* future performance, given some knowledge of the system offline, and online, and some assumed mathematical model;
- using *deterministic* or *stochastic* approaches;
- modeling the performance of a *process* with its constraints and specific aspect (for instance, identify variables that physically influence performance); or relating multiple processes of a networked system into an *interconnected* chain of dependencies;
- studying the *dynamics* of how the prediction changes as more online information is revealed, and the event to be predicted becomes closer in time.

In Table I, we comparatively categorize the literature with respect to those four factors. For the most relevant combinations found, we describe some typical problem settings and provide some exemplary references. The table does not claim to be a complete summary of the vast literature but compares the different problems tackled so far.

Table I shows that many works cover specific aspects of the descriptive and predictive modeling of bus operations. Various approaches that predict bus arrival times identify more variation in a prediction (i.e., lower predictability) at a long prediction horizon than at a short horizon. However, in the current literature, no framework models quantitatively such a relation.

Hence, we propose a model to describe the information regarding an event over time, modeled as a process happening in a networked system, subject to online information; that is, what we know when. We define a characteristic, the predictability, modeling the variation of the predicted value of an event, computed a certain time before the event occurs. In other terms, we characterize how the information revealed over time results in changes for the prediction error of an event.

## III. PREDICTABILITY

Events, as running or dwell times of buses, do not take the same duration on different days but are subject to natural variability. This natural variability can be learned from past observations of an event and can be considered as the Day-to-Day variability [12]. An event observed with similar values on multiple days has a low natural variability, whereas an event that takes vastly different values has high natural variability.

For a specific event, its variability can be estimated by its natural variability (i.e., the variability between similar trips on different days [12]). Conditioning the event to specific other events (covariates) might decrease the variability. For instance, the travel time on rainy days is always a bit longer than on sunny days. The travel time across all days is more variable than on either sunny or rainy days. Prediction models exploit this to associate a reduced variability once the predictors are determined. A completely predictable event is an event where the predicted variability is negligibly small. A small predicted variability can either be achieved by a low natural variability (e.g., a bus can always run at the speed limit on a dedicated bus lane) or an informative (i.e., with high sharpness and low bias) prediction model (e.g., a model finding that a bus running time takes a fixed value on sunny days and another fixed value on rainy days). Prediction models can only reduce part of the variability. We call remaining variability the still outstanding variability associated with the output of the prediction model.

We refer to the predictability $\mathfrak{P}$ of an event at a given time, as the remaining variability of this event, given a prediction model using selected data, observed at the given time. High predictability means that an event is well predicted, resulting in a low remaining variability. Low predictability is associated with high remaining variability. $\mathfrak{P}$ is highly time-dependent, as the availability and precision of explanatory data depend on the time. With a shorter prediction horizon, the knowledge of the situation gets more complete and accurate, possibly enabling a more informative prediction. Nevertheless, even at the shortest horizon, there is a remaining variability. $\mathfrak{P}$ depends on the natural variability, the availability and precision of explanatory data, and the prediction algorithm.

TABLE II
PREDICTABILITY DEPENDING ON VARIABILITY AND PREDICTION MODELS

| | | Low natural variability | High natural variability |
|---|---|---|---|
| Non-informative prediction model | $\mathfrak{P}_{long}$ | High | Low |
| | $\mathfrak{P}_{short}$ | High | Low |
| Informative prediction model | $\mathfrak{P}_{long}$ | High | Low |
| | $\mathfrak{P}_{short}$ | High | High |

Note: $\mathfrak{P}_{short}$ is the predictability at a short prediction horizon and $\mathfrak{P}_{long}$ is the predictability at a long time horizon.

For a given prediction model, $\mathfrak{P}$ of an event is bounded between a maximal and a minimal value. The lowest predictability is observed at a long horizon. Assuming a perfect prediction model, the predicted variability at a long prediction horizon $\mathfrak{P}_{long}$ would correspond to natural variability. For any prediction model, the predicted variability is higher than the natural variability due to epistemic uncertainties (due to modeling simplifications). The maximum predictability is achieved when the predicted variability is minimal, i.e., just before the event takes place. Assuming a perfect prediction model, the predicted variability at a short prediction horizon $\mathfrak{P}_{short}$ corresponds to the aleatoric (process) uncertainty. In practice, any prediction model can only approximate the true dynamic of operations and interdependencies of events.

Table II displays the characteristics of predictability by comparing four different cases of operation. It shows which combinations of high/low natural variability and non-/informative prediction model result in what predictability at short and long time horizons. An informative prediction model can reduce the variability with a decreasing prediction horizon, whereas a non-informative prediction model cannot reduce the variability. A non-informative prediction model can be attributed to a combination of aleatoric and epistemic uncertainties.

Given a low natural variability, an investment in a more informative prediction model might not be worthwhile, as the variability is already low, i.e., there is nothing to improve. In the presence of high natural variability, an informative prediction model can reduce the variability significantly. However, we need to keep in mind that the prediction is typically only informative if the required covariates are observed or can be well predicted.

The predictability directly answers the question of how reliable a prediction is at a given time. Thus, it delivers relevant information for passengers and operators, which is more beneficial than a measure that solely evaluates the natural variability or how much the prediction model is informative. The predictability combines both the evaluation of the natural variability and the prediction model and hence gives information on the remaining uncertainty – without requiring assumptions on the reason for it.

## IV. METHODOLOGY

Analyzing the predictability over time requires probabilistic predictions. Hence, we first introduce a prediction framework in section A, which delivers predictions that can be analyzed later. In section B, we present a model for analyzing predictability. Fig. 1 shows the flowchart of the methodology for modeling predictability, with an explicit comparison against descriptive and predictive analytics.

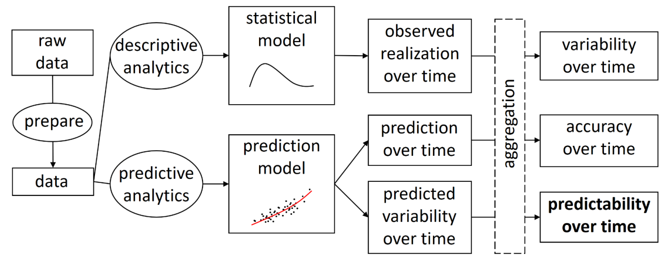


Fig. 1. Flowchart of the methodology.

### A. Prediction Framework

A prediction framework is needed to evaluate the predictability. We require the prediction approach to deliver probability distributions (as opposed to point estimates) to quantify the variability of the prediction. Furthermore, it needs to include online information such that the predicted variability changes with the temporal approaching of the predicted event. Any prediction approach fulfilling these requirements is applicable. In this paper, we use a Bayesian prediction framework, as it fulfills those requirements and is state-of-the-art [43], [45], [47].

In a Bayesian prediction framework, the crucial assumption is that the time value of a node $X_i$ only depends on its parents $\boldsymbol{Pa}_{X_i}$ and it is independent of any other preceding node in the graph given its parents:

$$P\left(X_i|X_1,\ldots,X_{i-1}\right) = P(X_i|\boldsymbol{Pa}_{X_i}) \tag{1}$$

This property allows simplifying the joint distribution for the Bayesian network to:

$$P\left(X_1,\ldots,X_n\right) = \prod_{i=1}^{n} P(X_i|\boldsymbol{Pa}_{X_i}) \tag{2}$$

This allows reducing the amount of required computation since most nodes have few parents, even for large networks. The proposed Bayesian network uses the following (standard) assumptions. The dwell time $dt$ of a given bus run $r$ at a given stop $s$ depends on the dwell time of the same bus at the previous stop $dt_{r,s-1}$, dwell time of the previous bus at the same stop $dt_{r-1,s}$, and the headway from the previous bus $h_{r,s}$. The running time depends on the running time of the same bus in the previous section $rt_{r,s-1}$ and the running time of the previous bus in the same section $rt_{r-1,s}$. Hence, the dependency structure is given by

$$\begin{aligned} P\left(dt_{r,s}|\ldots\right) &= P\left(dt_{r,s}|dt_{r-1,s}, dt_{r,s-1}, h_{r,s}\right) \\ P\left(rt_{r,s}|\ldots\right) &= P|(rt_{r-1,s}, rt_{r,s-1}) \end{aligned} \tag{3}$$

The described framework builds on running times at a section level and dwell times at a stop level. The arrival time $AT_{r,s}$ and the departure time $DT_{r,s}$ of a bus run $r$ at a specific stop s can be described by summing up all upstream running and dwell times and adding to the initial departure time of the bus run. $DT_{r,0}$ describes the initial departure time at the start of a bus run $r$.

$$AT_{r,s} = DT_{r,0} + \sum_{\sigma=1}^{s} rt_{r,\sigma} + \sum_{\sigma=1}^{s-1} dt_{r,\sigma} \tag{4}$$

$$DT_{r,s} = DT_{r,0} + \sum_{\sigma=1}^{s} rt_{r,\sigma} + \sum_{\sigma=1}^{s} dt_{r,\sigma} \tag{5}$$

The model is furthermore adapted for controlled bus operations and can account for holding strategies. In practice, this means that buses ahead of schedule wait at predefined time points. By deliberately forcing buses to wait, headway variance and average waiting time of passengers can be reduced. Our model extends the dependency structure of dwell times at time points. The dwell time of early buses is prolonged such that the early arriving buses depart exactly on time at such time points.

The proposed model assumes that all bus trips are the same (i.e., same stops are served), buses do not overtake each other, buses are independent of other bus lines, only holding strategies are applied; and at terminal points, there are no dependencies between buses. To apply the model in a case study where these assumptions do not hold, the model should be adapted.

### B. Modeling Predictability Over Time

To mathematically characterize predictability at different time horizons $\mathfrak{P}(T)$, we identify few parameters, which describe its behavior well. Here, we quantify the $\mathfrak{P}$ by the predicted variance $\upsilon$. Hence, low $\upsilon$ symbolizes a high $\mathfrak{P}$. First, we model the variance $\upsilon$ at an event level, that is isolated running and dwelling events. At a long prediction horizon $T$ we can define the variance $\upsilon_{long}$, which corresponds for a good model approximately to the empirical variance of all recorded events used for training the model. At short $T$ we expect the variance $\upsilon_{short}$, which corresponds for a good model approximately to the aleatoric uncertainty in the data. A short $T$ symbolizes the time just before the event takes place, hence $\upsilon_{short}$ is the minimal predicted variance.

$$\lim_{T\to\infty} \upsilon(T) = \upsilon_{long}$$
$$\lim_{T\to 0} \upsilon(T) = \upsilon_{short} \tag{6}$$

A descriptive analysis of statistical moments of operations (e.g. standard deviation of delay) is comparable to an estimate of the variance $\upsilon_{long}$. Prediction algorithms aim to reduce prediction error, i.e., reduce the estimated variance $\upsilon_{short}$ to a minimum value. Predictability fills the gap between those two values, studying the variance depending of $T$.

With the approaching of a predicted event, i.e., a decreasing $T$, the model can explain more and more of the variance due to its real-time update. At a long prediction horizon, the variance explained by the real-time update is 0. This latter value increases with a decreasing prediction horizon and just before the predicted event is observed, the variance explained by the real-time update $\upsilon_{expl}$ is maximized.

$$\lim_{T\to\infty} \upsilon_{expl}(T) = 0$$
$$\lim_{T\to 0} \upsilon_{expl}(T) = \upsilon_{long} - \upsilon_{short} \tag{7}$$

To approximate the variance at different $T$, we assume that the predictors are observed regularly over time and that they can reduce the variance similarly in every step. This assumption is a major simplification, but can be defended as, in reality, following sections or bus stops share similar properties (e.g., in terms of infrastructure, passenger demand, traffic). The rate parameter $\lambda$ characterizes the continuous reduction. Given these assumptions, which can be accepted as a first approximation, we state:

$$\frac{d\upsilon_{expl}}{dT} = -\lambda \cdot \upsilon_{expl} \tag{8}$$

Through rearranging and integration, we find an expression for $\upsilon_{expl}(T)$, which contains an integration constant

$$\frac{d\upsilon_{expl}}{\upsilon_{expl}} = -\lambda \cdot dT$$
$$\ln\left(\upsilon_{expl}\right) = -\lambda \cdot T + C$$
$$\upsilon_{expl}(T) = e^{C} e^{-\lambda \cdot T} \tag{9}$$

From the expression for $\upsilon_{expl}$ of equation (7), derives

$$\upsilon_{expl}(T) = (\upsilon_{long} - \upsilon_{short}) e^{-\lambda \cdot T} \tag{10}$$

Finally, we find

$$\begin{aligned} \upsilon(T) &= \upsilon_{long} - \upsilon_{expl} \\ &= \upsilon_{long} - \left(\upsilon_{long} - \upsilon_{short}\right) e^{-\lambda \cdot T} \\ &= \upsilon_{long} - \upsilon_{expl}(0) \cdot e^{-\lambda \cdot T} \end{aligned} \tag{11}$$

Fig. 2 shows the proposed modeling, where the variance explained through the real-time update decays exponentially with an increasing prediction horizon.

An interpretation of how fast a prediction increases from $\upsilon_{short}$ to $\upsilon_{long}$ with an increase in prediction horizon, is given by the half-life, solely dependent on $\lambda$:

$$t_{1/2} = \frac{\ln(2)}{\lambda} \tag{12}$$

A large $t_{1/2}$ means that the operations become predictable early (i.e., the variance is reduced early). A small $t_{1/2}$ instead means that gains in predictability are only achieved just moments before the event will occur.

Arrival times, as well as departure times or delays, depend on multiple upstream running and dwell times (4),(5). Hence, the variability of a specific arrival time consists of the variability of the running and dwell events leading to it. We assume the total variance to be a linear combination of the variances of the relevant running and dwell events. This assumption is justified, as the dwell and running distributions are predicted distributions. Any assumed dependency structure is already described by the prediction model, i.e., we assume no covariance between the distributions of the individual running and dwell times. Hence, the total variance of the arrival time is given as:

$$\begin{aligned} Var(AT_{r,s}) &= \upsilon\left(DT_{r,0} + \sum_{\sigma=1}^{s} rt_{r,\sigma} + \sum_{\sigma=1}^{S-1} dt_{r,\sigma}\right) \\ &= \sum_{\sigma=1}^{s} \upsilon\left(rt_{r,\sigma}\right) + \sum_{\sigma=1}^{s-1} \upsilon\left(dt_{r,\sigma}\right) \end{aligned} \tag{13}$$

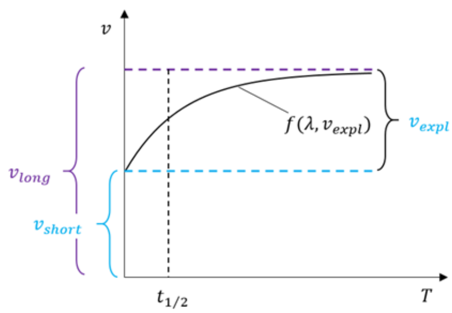


Fig. 2. Modeling the variance as a function of the prediction horizon.

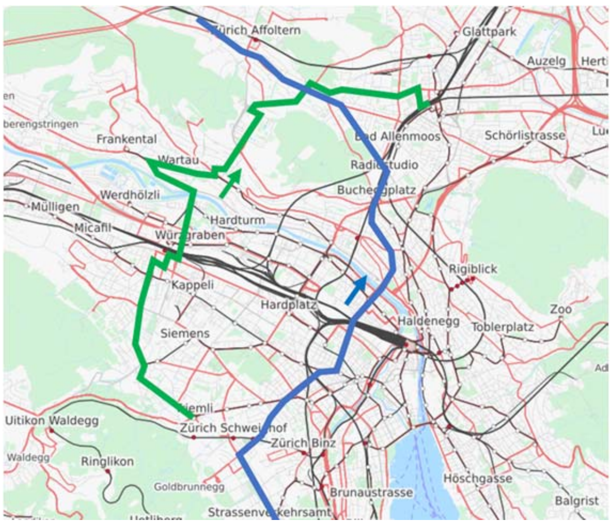


Fig. 3. Case study considering lines 32 (blue) and 80 (green) in Zurich.

The reduction of the total variance has thus two reasons. First, when the running or dwell process is observed, it is a deterministic value and not probabilistic anymore. Hence, the variance of the arrival time is reduced due to the observation of events. Second, as discussed in the last paragraph, the variance is reduced due to the real-time update.

## V. Test Case

For this study, we consider two urban bus lines in Zurich, Switzerland. The lines 32 (blue) and 80 (green), shown in Fig. 3, are major lines that cross the city with similar characteristics. Both are about 11 km long; a journey from the origin to the terminal station takes ∼40 min, during peak hours. The service frequency is 10 bus/h at peak hours. Line 32 is a trolleybus line, whereas line 80 operates with conventional buses. Stop 14 of bus line 32 is a time point, whereas line 80 has no time points. At stations that are no time points, buses may arrive and depart earlier than planned. Both lines are amongst the most delayed lines in the public transport network of Zurich.

The planned and realized arrival and departure times of buses at stations of the network are publicly available on the open data portal of the city of Zurich (https://data.stadt-zuerich.ch). The dwell and running times are computed as the time stopped at a bus stop, and the time running between stations, respectively. For this study, the data of 2018 is used, available at an accuracy of seconds. The raw data are filtered for regular trips on lines 32 and 80, meaning special routes to and from depots/garages are neglected. After data collection

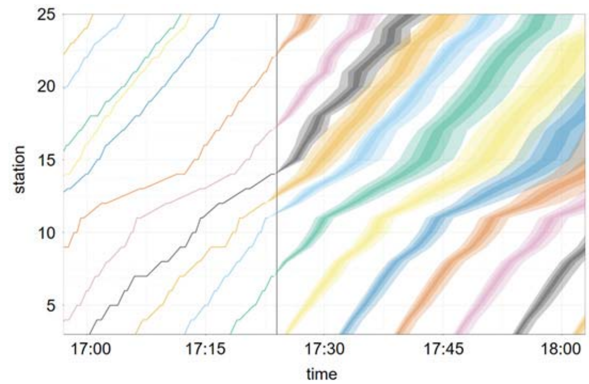


Fig. 4. Sample of a probabilistic predictions of bus line 32 in Zurich.

and cleaning, the next steps are the preparation in suitable data structure; fit of a stochastic prediction model, systematic analysis of the prediction model over time; and fit of the data to the decay model.

Automatically collected data can be subject to anomalies and inconsistencies, such as missing data points and unseen bus overtaking (see, e.g., [48], [49]). Overtaking is no issue in the selected case study, but we excluded the small number of runs with no recorded data, which might result in an overestimation of headways [48]. We do not impute bus runs (as done in [49]) as we have no way to assess if they took place in reality, or not.

Running times are assumed to be lognormally distributed (as shown in [39]), dwell times are modeled as normally distributed, and the initial departure of the route follows a normal distribution. A linear dependence structure between a node and its parents is assumed. Furthermore, the dependency of the time of the day was incorporated by discretizing the day into periods of similar behavior in terms of dwell and running time variability and fitting separate models for each period of 1h. Training the models for every hour of the day took ∼ 2h for the proposed Bayesian network models (Intel Xeon E3-1595Lv5 CPUs; 32GB of DDR4; 4 cores used). The prediction is evaluated for ∼200’000 single-step predictions, requiring ∼ 60min. Hence, the prediction of a single event takes less than 0.02s. Consequently, the prediction for all events of all buses of a line for the next 15min (∼ 150 events) would require ∼ 3s, making the model applicable in real-time.

## VI. Evaluation

### A. Fit of the Predictive Model

The application of the prediction model is shown in Fig. 4. It reports at 17:25 (indicated by the vertical line) the recorded and predicted variability of bus operations of line 32 in Zurich, Switzerland, for a peak hour (17:00-18:00) of a selected day. Past operation of buses (i.e., before 17:25) are known. They are shown by lines with different colors in the time-space diagram. The time-space position of buses is predicted for the future. The variability of the probabilistic prediction is shown with the 10th, 25th, 40th, 60th, 75th, and 90th percentile values in shades of color. The spread (i.e., the distance between the percentiles) is small for an event in the near future. However, even just before the event is observed, there is still some

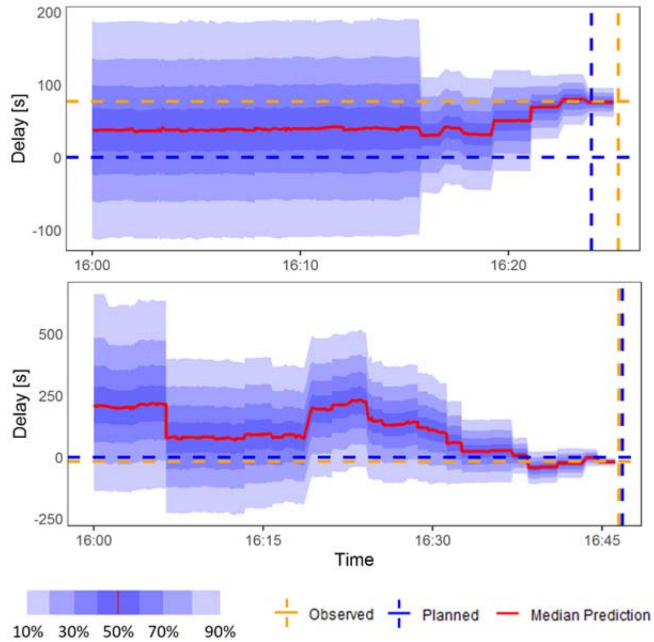


Fig. 5. Sample of evolutionary dynamics of predictions over time.

residual variability caused by exogenous factors, which can not be described by the model (aleatoric uncertainty) and modeling assumptions (epistemic uncertainty). The spread increases from near-future events to events in the far future. Predictions of events far in the future have a variability of operations that coincides approximately with the empirical distribution of all past observations of the event.

We further represent the evolutionary dynamics of increasing sharpness over time at the example of predicted arrival times in Fig. 5. We predict the arrival delay of the same bus run of line 32 at stop 8 (top) and stop 20 (bottom). The percentiles of the predicted delay are given as shades of blue. The dotted lines indicated observed (orange) and planned (blue) arrival times (vertical) and delays (horizontal). The top plot remains almost constant until $\sim$ 10min before the event, and then the accuracy of the prediction drastically increases. The bottom plot, on the other side, shows significant evolution already 45min before the event.

We first assess how well the prediction model can describe the realized operations. Table III shows the accuracy of the mean of the Bayesian network prediction for running and dwell times averaged over all events of lines 32 and 80. The accuracy values in terms of RMSE and MAE are only in the order of few seconds. The relative error quantified by MAPE for running times is $\sim$20% for running times and $\sim$30% for dwell times.

This accuracy could be improved by changing the prediction approach, variables, or modeling; however, we accept this quality as sufficient for the following predictability analyses.

### *B. Modeling Predictability of Single Running/Dwelling Events*

The resulting parameters of the fitted exponential functions are shown in Fig. 6 (top) for dwell times. For every event of the two bus lines, the diagrams depict the variances with two points connected by an arrow. To increase interpretability,

TABLE III

ACCURACY OF THE PREDICTION MODEL

| Time | line | RMSE [s] | MAE [s] | MAPE [-] |
|---|---|---|---|---|
| **Running time** | 32 | 18.0 | 13.6 | 0.21 |
| | 80 | 13.6 | 10.3 | 0.16 |
| **Dwell time** | 32 | 11.8 | 7.7 | 0.33 |
| | 80 | 8.4 | 5.6 | 0.29 |

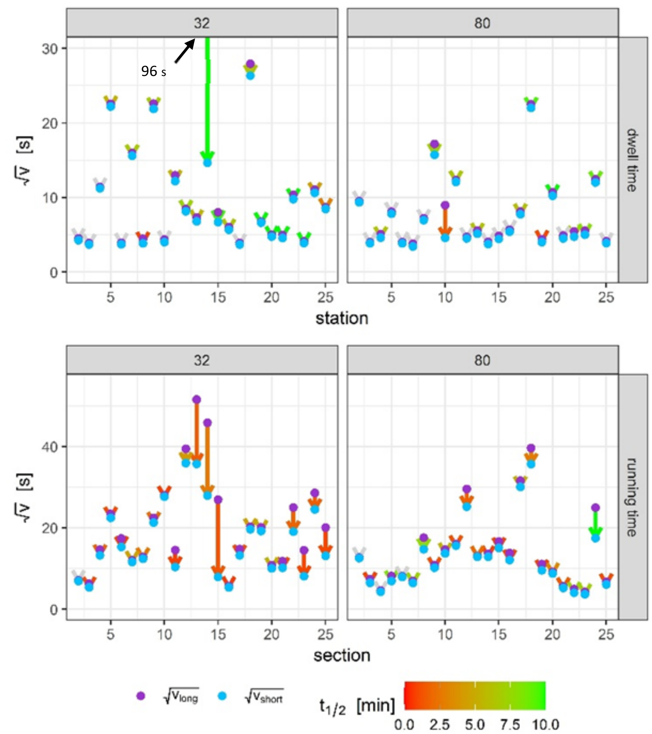


Fig. 6. Visual representation of parameters governing the reduction of variance for dwell time (top) and running time (bottom) over time.

we picture the square root of the variances, i.e., the standard deviation. The upper point (violet, see Fig. 2) gives $\sqrt{v_{long}}$, the lower point (light blue, see Fig. 2) gives $\sqrt{v_{short}}$ and the color of the arrow connecting the lines gives the half-life, where red is a short half-live and green a long one. If the difference between $\sqrt{v_{long}}$ and $\sqrt{v_{short}}$ is less than 1s, the half live is shown in grey.

For dwell times, $\sqrt{v_{long}}$ is for $\sim$ 75% of the stations low (below 10 s). Also, the difference between $\sqrt{v_{long}}$ and $\sqrt{v_{short}}$ is in many cases small; the real-time update often does not reduce the variance by more than 10%. The half-life, on the other side, is in many cases between 5 and 10 min. Consequently, the (small) reduction in variance can be predicted well ahead.

The dwell time at station 14 of bus 32 exhibits an exceptionally large $\sqrt{v_{long}}$ of 96 s, which is reduced due to real-time updates to $\sqrt{v_{short}} = 15$ s, just a fraction of the initial variation. This station serves as a time point. Thus, the arrival time at this stop is an important predictor. As this arrival time depends on the sum of all running and dwell times until this point, it has a rather high variance at a long prediction horizon.

TABLE IV
ACCURACY OF EXPONENTIAL DECAY MODEL FOR PREDICTED VARIANCE

| Time | Line | RMSE [$s^2$] | MAE [$s^2$] | MAPE [-] |
|---|---|---|---|---|
| **Running time** | 32 | 15.5 | 9.2 | 0.014 |
| | 80 | 9.3 | 6.7 | 0.026 |
| **Dwell time** | 32 | 134 | 109 | 0.052 |
| | 80 | 1.5 | 1.0 | 0.016 |

Fig. 6 (bottom) shows the parameters that govern the predicted variance at different time horizons for running times. The variance of running times is, in many cases, significantly higher than the variance for dwell times at all prediction horizons. Additionally, the real-time update improves the prediction significantly, what can be seen by the oftentimes big difference between $\sqrt{\upsilon_{long}}$ and $\sqrt{\upsilon_{short}}$. Further, we see that the half-life is typically lower than 5min; hence, the variance is only reduced shortly before the event is observed. Line 32, as compared to line 80, shows in many cases higher $\sqrt{\upsilon_{long}}$ and $\sqrt{\upsilon_{short}}$ values. Especially in the middle (sections 12-15) and at the end of the bus line (sections 22-25) the variability can be reduced with the real-time update. In both cases, the characteristics of the following sections are similar. Sections 12-15 is a corridor in the city center with many interactions with other transport modes and sections 22-25 is an arterial road exiting the city. In these sections, the prediction model can reduce the variance the most, but only at rather short notice ($t_{1/2} \sim 3$ min). Especially, high variances can be reduced through the prediction model. In general, the model can reduce the variance of running times compared to dwell times much better.

Now, we assess how well the exponential model can describe the predicted variance at different time horizons. Table IV shows the quality of the fit for the exponential model for running and dwell times averaged over all events of lines 32 and 80. The model is evaluated at prediction horizons of 1, 2, 5, 15, 30, and 45 min. The errors of the fit, quantified by RMSE and MAE, are in most cases small. The average errors for dwell time of line 32 are higher than for the other cases. This can mostly be attributed to the fit of the time point, where variances of up to $\sim$ 80'000 $s^2$ were observed. However, the percentage errors (MAPE) are on average between 1% and 5%. Hence, we can notice that the simple exponential model can fit the variance at different prediction horizons well.

### C. Modeling Predictability of Arrival Times

We here aggregate the running and dwell processes over multiple sections of a bus line, to discuss the predicted variance of the arrival times. Specifically, we look at the variance of the arrival time at a stop in the middle of the line and at the terminal station of the bus line.

Fig. 7 shows how the total variance is composed of the variances of dwell (blue) and running times (red) for lines 32 (top) and 80 (bottom). Each plot reports the stacked representations of multiple variance curves, analogous to Fig 2, where

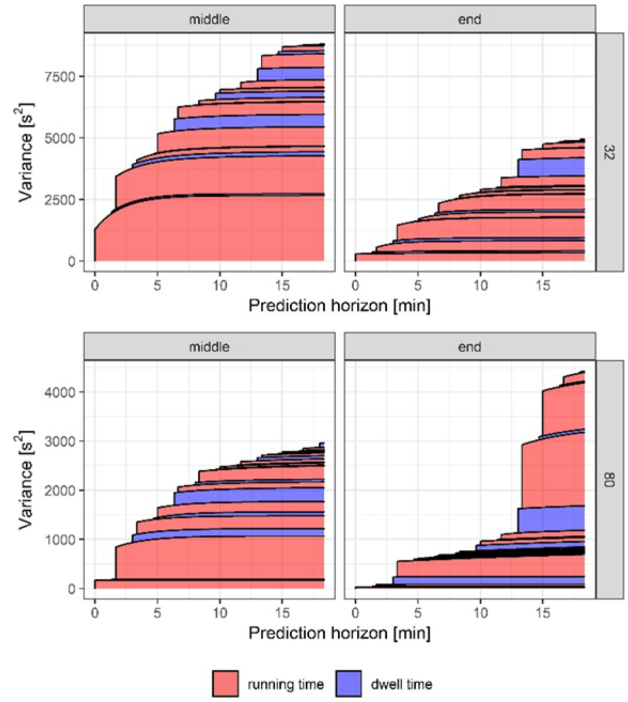


Fig. 7. Predicted variance, composed of running (red) and dwell (blue) events, for the arrival time at a stop in the middle (left) and the terminal stop (right) of lines of 32 (top) and 80 (bottom) at different prediction horizons.

the height of each curve describes the remaining variance over time for each specific stop. Note that in contrast to Fig. 6, we plot the variances and not the standard deviation. This, as we make use of the cumulative property of the variance, which would not hold for standard deviation. The colors identify how the variance is explained by the respective dwell/running times. For each event, we make the prediction at multiple moments in time. Furthermore, the running times last in average 80s and dwell times 20s (i.e., the average rounded to the closest ten of seconds). So we can build an approximation for the total variance by superimposing the variances of the running and dwelling processes. The four presented cases show different maximum values and temporal evolutions of the variance.

The first observed events are in the representation on top; events observed just before the arrival time at the terminal stations, i.e., running in late sections or dwell times at late stops, are on the bottom. A variance decreasing linearly is a stacked profile decreasing uniformly from top right to the bottom left. A large jump is a specific event, which explains most of the variance. Across all situations and prediction horizons, most of the variance is due to running processes (red). The influence of real-time prediction update is visible: the variance of individual events reduces with a decreasing prediction horizon. A big share of variance of the arrival time in the middle of route 32 (top left) is only reduced at a short horizon; the variance of the arrival time at the end of route 80 (bottom right) is already drastically reduced at a longer horizon.

The reduction of the variance of arrival times can have two reasons. First, the variance can be reduced due to updating

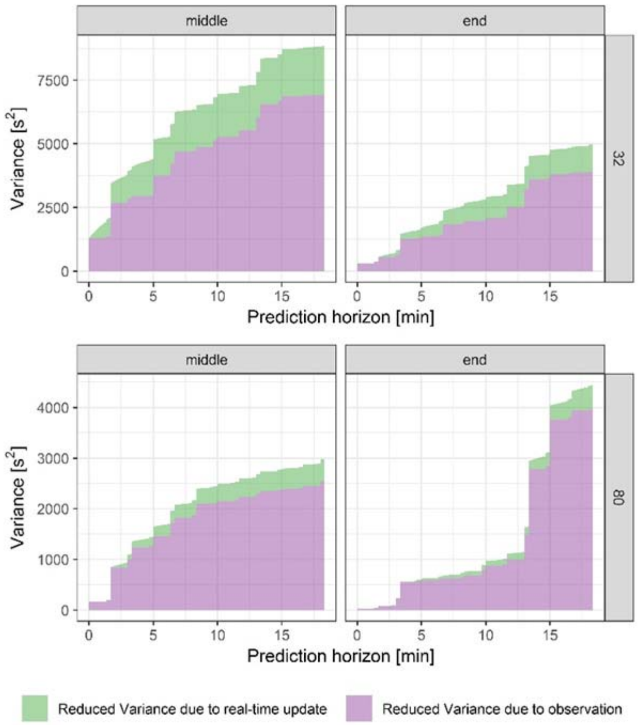


Fig. 8. Predicted variance for the arrival time at a stop in the middle (left) and at the terminal stop (right) of lines of 32 (top) and 80 (bottom) at different prediction horizons. The plot reveals how the total variance is composed of reduced variance due to real-time update (green) and observation (violet).

the prediction, by including more accurate data (e.g., the prediction of a dwell time upstream the predicted arrival time is improved as its covariates are observed). Second, the variance can be reduced due to observation of some events (e.g., the dwell time upstream the predicted arrival time is observed; thus it does not contribute any variability anymore). Fig. 8 shows what share of the total reduction can be attributed to what reason, with a 15-20% associated to real-time updates. For line 32 (top), the influence of the prediction is greater than for line 80 (bottom). The time point of line 32 lies in the middle of the line and is not shown in Fig. 8. At such a point the assumption of additive variance does not hold, as the reason for the high variance of a dwell time is to reduce the variance at later stops (i.e., the event has a duration negatively correlated with its predecessors).

The real-time update of the prediction model can only reduce the variance in a limited manner, whereas a major part of the variance reduction is attributed to the observation of events. Hence, it is crucial that the prediction system can incorporate observed information in real-time. The influence of the real-time update could be significantly higher in situations where the empirical variability of the running and dwell times are higher than in the presented case study. Furthermore, the predictability could improve by incorporating factors now assumed exogenous. Variations of running times could, for instance, be attributed to traffic variations and variations of dwell times could be attributed to passenger demand and behavior, which is outside the scope of the predictive model.

We can show how those predictability insights deliver added value compared to descriptive studies. Punctuality (measured as services with less than 3 minutes delay) is 92%, 85%, 99% and 95%, respectively for line 32, middle and end; and for line 80, middle and end. Regularity (measured as successive services with a deviation of less than 3 minutes from the planned headway) is respectively 89%, 82%, 95%, 90%. We thus have the case of highly punctual/regular operations (80 middle), which are having large variance still at 10 minutes ahead. The same line, analyzed at the last stop, has much worse punctuality/regularity performance, but 10 minutes ahead is actually twice more predictable than the middle stop, but with a late decrease. Similarly, the end stop of line 32 has much worse performance than the middle stop of that line, but its predicted variance is only half of this latter, at any time horizon.

## VII. Discussion and Applicability of Predictability

The predictability – as opposed to punctuality or regularity (realised variability computed after the event took place) – quantifies the temporal reduction of the predicted variability, as the event to be predicted gets closer in time. This concept is highly beneficial to use in non-punctual operations, which have low natural variability, or in the case of high natural variability, which is highly predictable. If operations show a low variability and high punctuality, such an approach does not provide novel insights as punctuality or variance-based interpretation of operations. Predictability focuses on the running/dwell times, and thus is independent on the timetable, unless time points are considered, when the buffer times play a role. We highlight the usefulness of predictability evaluation by three use cases:

### A. Bus Bunching Control

Bunching phenomena correspond to high differences in the headway of successive buses [50]. This phenomenon cannot fully be captured by means of punctuality as bunching requires differences in punctuality of following buses; regularity measures can identify this. Such phenomena can be prevented, e.g., by introducing a relief vehicle that can take over a service of a largely delayed bus at a specific stop. If only descriptive analytics are used, the relief vehicle must be reserved close to any specific stop in a static manner, i.e., assuming bunching either always occurs; or never. This results in very high cost due to the false positives (i.e., a relief vehicle is reserved for a specific stop, but bunching does not occur or occurs elsewhere). A predictability study allows, similar to probabilistic prediction frameworks [46], [51], determining the probability of such bunching phenomena (as defined as a critical headway between two successive buses) at any given time horizon ahead. When this situation is predicted, a relief vehicle can be proactively dispatched from a central depot. Achieving a very small prediction error shortly before the situation occurs is not useful, as the relief vehicle might arrive just too late. High predictability at a long time horizon is demanded to ensure efficient, proactive control. In specific, we need sufficient predictability of bunching at a time horizon where we can react (i.e., running time from the depot to the relief stop, say 30 minutes). Predictability can deliver additional insights, for instance, determining a priori the optimal parking locations of relief vehicles, to maximize their timely deployment to specific bunching situations. Considering the case of line 80 end station (Fig. 7 bottom right), an ideal location for a relief vehicle would be around 10min from the station, where the predicted variance is decreasing sharply.

We assume a determined required probability threshold for bunching to avoid false positives (i.e., a relief vehicle is sent, but the operation resolves on its own, resulting in costs as the sent vehicle is not used for service). A highly variable but lowly predictable operation would only at short notice provide a certain enough prediction of bunching. A highly variable but highly predictable operation would instead already at a long horizon provide a certain enough prediction. Given the specific certainty threshold, to accept a false positive, a predictability decay would determine the maximum intervention time that one can afford, given the recorded operations.

### B. Bottleneck Identification

The predictability-based analysis of a bus route allows the categorization of running and dwell events based on natural and predicted variability. The analysis below might apply to both cases of satisfactory or unsatisfactory punctuality and regularity, as we analyze only the ratio between $\upsilon_{long}$ and $\upsilon_{short}$, which is not studied by descriptive or predictive applications. If the predictability at a long horizon is much higher than at a short horizon ($\upsilon_{long} \gg \upsilon_{short}$), the variability can be largely reduced by the prediction. Thus the event is highly dependent on close-by events. For that reason, actions to reduce variability (e.g., dedicated bus lane, signaling) should there be planned on a multi-section level. If the predictability at a long horizon is only slightly higher than at a short horizon, the prediction model is not informative and the event is defined through aleatoric uncertainty. To reduce the variability of events with this characteristic of having a small $\upsilon_{expl}$, actions should focus on the section level. Especially in the case of budget limitations, actions need to be prioritized. Given two sections with high natural variability but different short-term predictabilities, the section with low short-term predictability should be treated with priority. The predictability has evidently an associated value, similar to reliability [52].

### C. Passenger Information

Real-time information results in many passenger benefits increasing passenger satisfaction [2]. Predictability can be crucial for improving real-time information systems. It allows informing passengers about associated variabilities, which has been related to an increased satisfaction [11], [53]. Furthermore, especially in complex interconnected networks, transfers are not always guaranteed. The probability of a successful transfer, from the perspective of passengers, depends on the arrival time of the first service and the departure time of the second service plus possible walking times at a transfer point (including close bus stops) (see, e.g., [54]). A descriptive approach would deliver an average success rate of transfer, and its actionable insight is an expected travel time to possibly choose a different route. A predictive approach might estimate, given some current conditions known shortly in advance, how much transfer time would be available. A predictability-based approach can, at any time horizon, estimate the probability of a successful transfer. Providing this information to passengers might reduce their anxiety, improve their transport experience, and lead to increased transit ridership. Furthermore, if the probability of a successful transfer drops below a given threshold, the passenger could be proactively rerouted to a different route, resulting in a lower total travel time. This contributes to the field of route choice under uncertainty, enriching the existing models and optimization strategies [55], [56].

## VIII. Conclusion

This paper introduces the concept of predictability to describe the knowledge we have of a network system over time. The concept is shown based on predictions obtained by a state-of-the-art Bayesian network approach, where the predictability is modeled as an exponential decay. It has been demonstrated that predictability has the potential to enhance the description, control, and communication of the current state of the system.

We evaluate the predictability based on a given prediction model, as opposed to a discussion of the entropy of the system to determine the maximum predictability limit. The limitations of the study are the need for data of sufficient quality to model normal operations. We considered a Bayesian network prediction model, but this is not a prerequisite. The presented analysis could be performed with any prediction framework capable of predicting variability and updating the prediction in real-time with additional information. Multiple predictions methods could also be compared with each other.

Future work should expand the proposed concepts in a variety of directions, which include other public transport systems, with associated different uncertainty factors and specific processes and constraints, for instance, modeling mobility in a smart city. Moreover, predictability could focus on specific non-performance, differentiating everyday peak hours; and planned/unexpected maintenance [57]. Other decay models, or considering lower/upper bounds, could be studied.

Reliability and its costs have been recently included in the assessment and appraisal of public transport projects [52]. Given the widespread availability of means to disseminate information to travelers, a low reliability but high predictability of services, could effectively be considered as a high reliability for transport economics purposes. We did not discuss the case of exceptional conditions, or erroneous data in the recorded data, or actual situation. Such cases might threaten the data-driven foundations of the concept and require a data collection and transmission method of high quality (as applied in, e.g., [8], [58]). The impact of higher predictability towards demand can also be identified, similar to [59]. A different complementary direction is to compute how much predictability is actually exploited by passengers in their choice process by fitting realized choices [60].

## Nomenclature

### A. Prediction Framework

| | |
|---|---|
| $X_i$ | node of a Bayesian Network. |
| $\boldsymbol{Pa}_{X_i}$ | parents of $X_i$. |
| $P(X_i)$ | probability distribution of $X_i$. |
| $n$ | total number of nodes. |
| $rt_{r,s}$ | running time of bus run $r$ in the section before stop $s$. |
| $dt_{r,s}$ | dwell time of a bus run $r$ at a stop $s$. |
| $h_{r,s}$ | headway of a bus run $r$ at a stop $s$ compared to the previous bus. |
| $r$ | a bus run. |
| $\sigma$ | a generic stop of a bus. |
| $s$ | a stop of a bus. |
| $AT_{r,s}$ | arrival time of bus run $r$ at a stop $s$. |
| $DT_{r,s}$ | departure time of bus run $r$ at a stop $s$. |

### B. Predictability Analysis

| | |
|---|---|
| $\mathfrak{P}$ | predictability. |
| $\mathfrak{P}_{long}$ | predictability at a long time horizon. |
| $\mathfrak{P}_{short}$ | predictability at a short time horizon. |
| $\upsilon(\alpha, T)$ | variance of a measurement $\alpha$ at time horizon $T$. |
| $\upsilon(T)$ | variance at time horizon $T$, simplified version of the above for a generic measurement. |
| $\upsilon(\alpha)$ | variance of a measurement $\alpha$, simplified version of the above for a generic time horizon. |
| $\upsilon_{long}$ | variance at a long time horizon. |
| $\upsilon_{short}$ | variance at a short time horizon. |
| $\upsilon_{expl}$ | explained variance. |
| $\lambda$ | rate parameter of exponential distribution. |
| $t_{1/2}$ | half-life. |
| $T$ | a time horizon or time variable. |
| $C$ | an integration constant. |

## Acknowledgment

The authors would like to thank Sebastian Leisinger for his valuable help while building the prediction model.

## References

[1] O. J. Ibarra-Rojas, "Planning, operation, and control of bus transport systems: A literature review," *Transp. Res., Methodol.*, no. 77, pp. 38–75, 2015.

[2] C. Brakewood and K. Watkins, "A literature review of the passenger benefits of real-time transit information," *Transp. Rev.*, vol. 39, no. 3, pp. 327–356, May 2019, doi: 10.1080/01441647.2018.1472147.

[3] T. F. Welch and A. Widita, "Big data in public transportation: A review of sources and methods," *Transp. Rev.*, vol. 39, no. 6, pp. 795–818, Nov. 2019, doi: 10.1080/01441647.2019.1616849.

[4] E. Wong and A. Khani, "Transit delay estimation using stop-level automated passenger count data," *J. Transp. Eng., A, Syst.*, vol. 144, no. 3, Mar. 2018, Art. no. 04018005, doi: 10.1061/JTEPBS.0000118.

[5] M. D. Abkowitz and I. Engelstein, "Factors affecting running time on transit routes," *Transp. Res. A, Gen.*, vol. 17, no. 2, pp. 107–113, Mar. 1983.

[6] K. Dueker, T. Kimpel, J. Strathman, and S. Callas, "Determinants of bus dwell time," *J. Public Transp.*, vol. 7, no. 1, pp. 21–40, Mar. 2004, doi: 10.5038/2375-0901.7.1.2.

[7] B. Barabino, M. Di Francesco, and S. Mozzoni, "An offline framework for the diagnosis of time reliability by automatic vehicle location data," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 583–594, Mar. 2017, doi: 10.1109/TITS.2016.2581024.

[8] B. Barabino and M. D. Francesco, "Diagnosis of irregularity sources by automatic vehicle location data," *IEEE Intell. Transp. Syst. Mag.*, vol. 13, no. 2, pp. 152–165, Summer 2019, doi: 10.1109/MITS.2018.2889713.

[9] *Transportation—Logistics and Services—Public Passenger Transport—Service Quality Definition, Targeting and Measurement*, European Committee for Standardization, Brussels, Belgium, 2002.

[10] O. Cats and G. Loutos, "Real-time bus arrival information system: An empirical evaluation," *J. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 138–151, 2016, doi: 10.1080/15472450.2015.1011638.

[11] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson, "When (ISH) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems," in *Proc. Conf. human Factors Comput. Syst.*, 2016, pp. 5092–5103, doi: 10.1145/2858036.2858558.

[12] L.-M. Kieu, A. Bhaskar, and E. Chung, "Public transport travel-time variability definitions and monitoring," *J. Transp. Eng.*, vol. 141, no. 1, Jan. 2015, Art. no. 04014068.

[13] R. Choudhary, A. Khamparia, and A. K. Gahier, "Real time prediction of bus arrival time: A review," in *Proc. 2nd Int. Conf. Next Gener. Comput. Technol. (NGCT)*, Dehradun, India, Oct. 2016, pp. 25–29, doi: 10.1109/NGCT.2016.7877384.

[14] L. Lian, J. Shang, and X. Zong, "Review of research on prediction algorithms of bus headway," in *Proc. 7th Int. Conf. Inf., Cybern., Comput. Social Syst. (ICCSS)*, Guangzhou, China, Nov. 2020, pp. 789–792, doi: 10.1109/ICCSS52145.2020.9336859.

[15] J. Ma, J. Chan, G. Ristanoski, S. Rajasegarar, and C. Leckie, "Bus travel time prediction with real-time traffic information," *Transp. Res. C, Emerg. Technol.*, vol. 105, pp. 536–549, Aug. 2019, doi: 10.1016/j.trc.2019.06.008.

[16] C. Bai, Z.-R. Peng, Q.-C. Lu, and J. Sun, "Dynamic bus travel time prediction models on road with multiple bus routes," *Comput. Intell. Neurosci.*, vol. 2015, pp. 1–9, Jul. 2015, doi: 10.1155/2015/432389.

[17] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Syst. Appl.*, vol. 120, pp. 426–435, Apr. 2019, doi: 10.1016/j.eswa.2018.11.028.

[18] G. Chen, X. Yang, J. An, and D. Zhang, "Bus-arrival-time prediction models: Link-based and section-based," *J. Transp. Eng.*, vol. 138, no. 1, pp. 60–66, Jan. 2012, doi: 10.1061/(ASCE)TE.1943-5436.0000312.

[19] M. Æelan and M. Lep, "Bus-arrival time prediction using bus network data model and time periods," *Future Gener. Comput. Syst.*, vol. 110, pp. 364–371, Sep. 2020, doi: 10.1016/j.future.2018.04.077.

[20] F. Pili, A. Olivo, and B. Barabino, "Evaluating alternative methods to estimate bus running times by archived automatic vehicle location data," *IET Intell. Transp. Syst.*, vol. 13, no. 3, pp. 523–530, Mar. 2019, doi: 10.1049/iet-its.2018.5339.

[21] O. Alam, A. Kush, A. Emami, and P. Pouladzadeh, "Predicting irregularities in arrival times for transit buses with recurrent neural networks using GPS coordinates and weather data," *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 7, pp. 7813–7826, Jul. 2021, doi: 10.1007/s12652-020-02507-9.

[22] Y. Zhou, L. Yao, Y. Chen, Y. Gong, and J. Lai, "Bus arrival time calculation model based on smart card data," *Transp. Res. C, Emerg. Technol.*, vol. 74, pp. 81–96, Jan. 2017, doi: 10.1016/j.trc.2016.11.014.

[23] P. Wepulanon, A. Sumalee, and W. H. K. Lam, "A real-time bus arrival time information system using crowdsourced smartphone data: A novel framework and simulation experiments," *Transportmetrica B, Transp. Dyn.*, vol. 6, no. 1, pp. 34–53, Jan. 2018, doi: 10.1080/21680566.2017.1353449.

[24] D. Sun, H. Luo, L. Fu, W. Liu, X. Liao, and M. Zhao, "Predicting bus arrival time on the basis of global positioning system data," *Transp. Res. Rec.*, vol. 2034, no. 1, pp. 62–72, 2007, doi: 10.3141/2034-08.

[25] J. Patnaik, S. Chien, and A. Bladikas, "Estimation of bus arrival times using APC data," *J. Public Transp.*, vol. 7, no. 1, pp. 1–20, Mar. 2004, doi: 10.5038/2375-0901.7.1.1.

[26] H. Liu, H. Xu, Y. Yan, Z. Cai, T. Sun, and W. Li, "Bus arrival time prediction based on LSTM and spatial-temporal feature vector," *IEEE Access*, vol. 8, pp. 11917–11929, 2020, doi: 10.1109/ACCESS.2020.2965094.

[27] J. Zhang, J. Gu, L. Guan, and S. Zhang, "Method of predicting bus arrival time based on MapReduce combining clustering with neural network," in *Proc. IEEE 2nd Int. Conf. Big Data Anal. (ICBDA)*, Beijing, China, Mar. 2017, pp. 296–302, doi: 10.1109/ICBDA.2017.8078828.

[28] W. Treethidtaphat, W. Pattara-Atikom, and S. Khaimook, "Bus arrival time prediction at any distance of bus route using deep neural network model," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Yokohama, Japan, Oct. 2017, pp. 988–992, doi: 10.1109/ITSC.2017.8317891.

[29] J. Pang, J. Huang, Y. Du, H. Yu, Q. Huang, and B. Yin, "Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3283–3293, Sep. 2019, doi: 10.1109/TITS.2018.2873747.

[30] C.-H. Chen, "An arrival time prediction method for bus system," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 4231–4232, Oct. 2018, doi: 10.1109/JIOT.2018.2863555.

[31] A. A. Agafonov and A. S. Yumaganov, "Bus arrival time prediction using recurrent neural network with LSTM architecture," *Opt. Memory Neural Netw.*, vol. 28, no. 3, pp. 222–230, Jul. 2019, doi: 10.3103/S1060992X19030081.

[32] M. Yang, Y. Liu, and Z. You, "The reliability of travel time forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 162–171, Mar. 2010, doi: 10.1109/TITS.2009.2037136.

[33] A. O'Sullivan, F. C. Pereira, J. Zhao, and H. N. Koutsopoulos, "Uncertainty in bus arrival time predictions: Treating heteroscedasticity with a metamodel approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 11, pp. 3286–3296, Nov. 2016, doi: 10.1109/TITS.2016.2547184.

[34] H. Yu, Z. Wu, D. Chen, and X. Ma, "Probabilistic prediction of bus headway using relevance vector machine regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1772–1781, Jul. 2017, doi: 10.1109/TITS.2016.2620483.

[35] E. Mazloumi, G. Rose, G. Currie, and S. Moridpour, "Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction," *Eng. Appl. Artif. Intell.*, vol. 24, no. 3, pp. 534–542, 2011, doi: 10.1016/j.engappai.2010.11.004.

[36] M. M. Rahman, S. C. Wirasinghe, and L. Kattan, "Analysis of bus travel time distributions for varying horizons and real-time applications," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 453–466, Jan. 2018, doi: 10.1016/j.trc.2017.11.023.

[37] C. Coffey, A. Pozdnoukhov, and F. Calabrese, "Time of arrival predictability horizons for public bus routes," in *Proc. 4th ACM SIGSPATIAL Int. Workshop Comput. Transp. Sci.*, Chicago, IL, USA, 2011, pp. 1–5, doi: 10.1145/2068984.2068985.

[38] Z.-Y. Wang, B.-H. Jin, T. Ge, and T.-F. Xue, "Detecting anomalous bus-driving behaviors from trajectories," *J. Comput. Sci. Technol.*, vol. 35, no. 5, pp. 1047–1063, Oct. 2020, doi: 10.1007/s11390-020-9933-3.

[39] B. Báchel and F. Corman, "Meaningful modeling of section bus running times by time varying mixture distributions of fixed components," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 8, pp. 626–637, May 2020, doi: 10.1177/0361198120918576.

[40] B. Báchel and F. Corman, "Review on statistical modeling of travel time variability for road- based public transport," *Frontiers Built Environ.*, vol. 6, p. 70, Jun. 2020, doi: 10.3389/fbuil.2020.00070.

[41] W.-C. Lee, W. Si, L.-J. Chen, and M. C. Chen, "HTTP: A new framework for bus travel time prediction based on historical trajectories," in *Proc. 20th Int. Conf. Adv. Geographic Inf. Syst.*, Redondo Beach, CA, USA, 2012, p. 279, doi: 10.1145/2424321.2424357.

[42] A. Achar, D. Bharathi, B. A. Kumar, and L. Vanajakshi, "Bus arrival time prediction: A spatial Kalman filter approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1298–1307, Mar. 2020, doi: 10.1109/TITS.2019.2909314.

[43] L. Deng, Z. He, and R. Zhong, "The bus travel time prediction based on Bayesian networks," in *Proc. Int. Conf. Inf. Technol. Appl.*, Chengdu, China, Nov. 2013, pp. 282–285, doi: 10.1109/ITA.2013.73.

[44] P. He, G. Jiang, S. K. Lam, and D. Tang, "Travel-time prediction of bus journey with multiple bus trips," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 11, pp. 4192–4205, Nov. 2019, doi: 10.1109/TITS.2018.2883342.

[45] B. Hu, K. Xie, H. Cui, and H. Lin, "A Bayesian spatiotemporal approach for bus speed modeling," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Auckland, New Zealand, Oct. 2019, pp. 497–502, doi: 10.1109/ITSC.2019.8917523.

[46] L. Moreira-Matias, J. Gama, J. Mendes-Moreira, and J. F. de Sousa, "An incremental probabilistic model to predict bus bunching in real-time," in *Advances in Intelligent Data Analysis XIII* (Lecture Notes in Computer Science), vol. 8819. Cham, Switzerland: Springer, 2014, pp. 227–238.

[47] F. Corman and P. Kecman, "Stochastic prediction of train delays in real-time using Bayesian networks," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 599–615, Oct. 2018, doi: 10.1016/j.trc.2018.08.003.

[48] B. Barabino, M. Di Francesco, and S. Mozzoni, "Time reliability measures in bus transport services from the accurate use of automatic vehicle location raw data: Accurate time reliability measures," *Qual. Rel. Eng. Int.*, vol. 33, no. 5, pp. 969–978, Jul. 2017, doi: 10.1002/qre.2073.

[49] F. McLeod, "Estimating bus passenger waiting times from incomplete bus arrivals data," *J. Oper. Res. Soc.*, vol. 58, no. 11, pp. 1518–1525, 2007.

[50] G. F. Newell and R. B. Potts, "Maintaining a bus schedule," in *Proc. Austral. Road Res. Board (ARRB) Conf.*, vol. 2, Melbourne, VIC, Australia, 1964, pp. 1–5.

[51] L. Moreira-Matias, O. Cats, J. Gama, J. Mendes-Moreira, and J. F. de Sousa, "An online learning approach to eliminate bus bunching in real-time," *Appl. Soft Comput.*, vol. 47, pp. 460–482, Oct. 2016, doi: 10.1016/j.asoc.2016.06.031.

[52] M. Fosgerau and A. Karlstrom, "The value of reliability," *Transp. Res. B, Methodol.*, vol. 44, no. 1, pp. 38–49, 2010.

[53] M. M. Rahman, S. C. Wirasinghe, and L. Kattan, "Users' views on current and future real-time bus information systems: Users' views on current and future real-time bus information systems," *J. Adv. Transp.*, vol. 47, no. 3, pp. 336–354, Apr. 2013, doi: 10.1002/atr.1206.

[54] B. Barabino, M. Di Francesco, G. Maternini, and S. Mozzoni, "Offline framework for the diagnosis of transfer reliability by automatic vehicle location data," *IEEE Intell. Transp. Syst. Mag.*, early access, Mar. 4, 2021, doi: 10.1109/MITS.2021.3051977.

[55] A. Carrel, A. Halvorsen, and J. L. Walker, "Passengers' perception of and behavioral adaptation to unreliability in public transportation," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2351, no. 1, pp. 153–162, Jan. 2013, doi: 10.3141/2351-17.

[56] M. Schmidt, L. Kroon, A. Schöbel, and P. Bouman, "The travelers route choice problem under uncertainty: Dominance relations between strategies," *Oper. Res.*, vol. 65, no. 1, pp. 184–199, Feb. 2017, doi: 10.1287/opre.2016.1564.

[57] A. D'Ariano, L. Meng, G. Centulio, and F. Corman, "Integrated stochastic optimization approaches for tactical scheduling of trains and railway infrastructure maintenance," *Comput. Ind. Eng.*, vol. 127, pp. 1315–1335, Jan. 2019, doi: 10.1016/j.cie.2017.12.010.

[58] A. H. Sodhro *et al.*, "Towards 5G-enabled self adaptive green and reliable communication in intelligent transportation system," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 8, pp. 5223–5231, Aug. 2020, doi: 10.1109/TITS.2020.3019227.

[59] L. Tang and P. Thakuriah, "Ridership effects of real-time bus information system: A case study in the city of Chicago," *Transp. Res. C, Emerg. Technol.*, vol. 22, pp. 146–161, Jun. 2012, doi: 10.1016/j.trc.2012.01.001.

[60] A. D. Marra and F. Corman, "Determining an efficient and precise choice set for public transport based on tracking data," *Transp. Res. A, Policy Pract.*, vol. 142, pp. 168–186, Dec. 2020, doi: 10.1016/j.tra.2020.10.013.

**Beda Büchel** received the B.Sc. and M.Sc. degrees in civil engineering from the Swiss Federal Institute of Technology, ETH Zürich, Switzerland, in 2013 and 2016, respectively, and the Ph.D. degree from ETH Zürich. His research interests include descriptive and predictive modeling of public transport systems.

**Francesco Corman** received the B.Sc. and M.Sc. degrees from University Roma Tre, Rome, and the Ph.D. degree in transport sciences from TU Delft, The Netherlands. Since 2017, he has been the Chair of transport systems at the Institute of Transport Planning and Systems, ETH Zürich. He is interested in the application of quantitative methods and optimization to transport sciences.