



HAL
open science

A microscopic public transportation simulation framework based on machine learning

Younes Delhoum, Olivier Cardin, Maroua Nouiri, Mounira Harzallah

► **To cite this version:**

Younes Delhoum, Olivier Cardin, Maroua Nouiri, Mounira Harzallah. A microscopic public transportation simulation framework based on machine learning. *Journal of Public Transportation*, 2024, 26, pp.100103. 10.1016/j.jpubtr.2024.100103 . hal-04719367

HAL Id: hal-04719367

<https://hal.science/hal-04719367v1>

Submitted on 3 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A microscopic public transportation simulation framework based on machine learning

Younes Delhoum¹, Olivier Cardin¹, Maroua Nouiri¹, Mounira Harzallah¹

¹ Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France (younes.delhoum@univ-nantes.fr; olivier.cardin@univ-nantes.fr; maroua.nouiri@univ-nantes.fr; mounira.harzallah@univ-nantes.fr)

Abstract

The evaluation of performance of public transportation, such as bus lines for example, is a major issue for operators. To be able to integrate specific and local behaviors, microscopic simulations of the lines, modelling each buses on a daily basis, brings an actual added value in terms of precision and quality. A scientific deadlock then appears regarding the parameterization of the simulation model. In order to be able to gather relevant performance indicators on a potential evolution of the configuration of the line, validated and modifiable simulation models need to be developed. This study aims at proposing a model development methodology based on a multi-agent simulation framework and data inputs extracted by a hybrid approach combining [machine learning \(ML\)](#) trained on actual bus data to predict travel times and probabilistic distributions to accurately estimate travel time variability. It also aims to propose a two-step validation framework that exhibits the performance of the obtained model on a case study based on actual data. The results of the proposed approach are validated by a real case study of three bus lines, including a number of simulation scenarios, to study the impacts of bus recovery time and bus control strategies on bus punctuality. The results obtained show that proposed hybrid approach combining ML with probabilistic distributions outperforms probabilistic distributions on average. Overall, the results show a good fit with the actual [Key Performance Indicator \(KPI\)](#) used by bus operators.

Keywords: Machine learning, Microscopic simulation, Public transport, Bus punctuality, Bus travel time, Bus holding control

1. Introduction

Nowadays, urban cities face multiple challenges, ranging from rapidly increasing population to increasing number of motorized vehicles and traffic congestion [1]. In order to reduce traffic congestion, and therefore its negative impacts, cities around the world are trying to shift personal traffic to public transport [2]. Public transportation (PT), public transit or mass transit is defined as transportation of passengers by group travel systems available for use by the general public. It is an important part of urban life, which plays a critical and essential role for mobility in the modern city. It is managed according to a predefined schedule and operated on established routes. The reliability of service can be understood as a combination of: punctuality of service provided within the scheduled times and the frequency of service of vehicles evenly spaced to accommodate passengers. In public transport, reliability is considered one of the most critical features to assess the quality of service from the perspectives of passengers and operators. Reliability is understood in PT as the certainty that passengers have about the level of service they will experience while travelling [3]. Travel time and waiting time, in addition to transfer time and comfort level, experienced during the transit trip are some of the important reliability attributes [4].

Travel time variability (TTV), also known as travel time uncertainty, is the key indicator for measuring the performance of the transport system, can be divided into three distinct components: day-to-day variability, variability over the course of a day and vehicle-to-vehicle variability [5]. Research emphasizes day-to-day (or inter-day) variability,

*Corresponding author. olivier.cardin@univ-nantes.fr

17 which describes the degree of variation in travel time for a trip taking the same route over a specific period [6]. Probabilistic distributions are capable of describing the nature and the pattern of travel time variability. Understanding travel time distributions and their components is a prerequisite for reliability analysis. Various studies have made considerable effort in fitting travel times with different types of distributions, such as normal [7], lognormal [8; 9; 10; 11; 12], gamma [13; 14], Weibull [15], Burr [16; 17], Generalized extreme value (GEV) distribution [18], Gaussian Mixture Model [19] were proposed. Such modelling aimed to find the best statistical distribution to describe and explain the shape and the pattern of TTV, because a distribution fitted to travel time values could illustrate a more comprehensive nature of the TTV [20].

25 Limiting transit delays is at the core of control metrics for the public transport operators, limiting these delays will help to maintain an appropriate quality of service with reasonable certainty. Delays can be caused by multiple factors: traffic congestion, passenger boarding, road delays and certain events like traffic accidents, etc... Historically, many researchers have adopted various methods for predicting the estimated time of arrival of buses, these methods can be divided into several groups: historical and real-time approaches, Kalman filtering approaches, statistical methods and ML techniques [21]. In practice, the performance indicators are often estimated in a purely statistical way by practitioners over several months or years of data. This method provides a relatively coherent image of the past behaviour of the bus lines with low modelling effort and computation cost, but does not allow to evaluate these indicators in different "what if?" scenarios (e.g. roadworks, new traffic priorities, etc.).

34 Another problem that the bus operators face is bus bunching. Bus bunching refers to the phenomenon where a group of two or more buses arrives at the same bus stop at the same time. It is a long-standing operational problem in urban public transport systems, and it is a major issue that concerns transit users and affects our perception of service reliability and efficiency [22]. This indicator is completely impossible to compute in a purely statistical way, as it relates to single bus travels, their history (delays) and the buses ahead and behind [23; 24; 25; 26]. Microscopic simulations are one of the most promising way to be able to tackle these kind of problems.

41 To deal with the challenges noted earlier, bus operators need efficient microscopic bus lines simulation models to evaluate the performance indicators of the current lines and forecast indicators on potential evolutions of these lines. To do so, the first requirement is to be able to model the current situation in a coherent way and validate the results it provides. In this study, we aim to validate public transport simulation framework, including simulated travel times and bus punctuality, with the available actual data. The focus is on retrieving simulation parameters that best match the actual data, particularly bus punctuality which is strongly related to bus travel time. A classical approach is to perform the simulation with scheduled travel times, for example from [General Transit Feed Specification \(GTFS\)](#) data. However these travel times are too theoretical and may differ greatly from observed travel times. We therefore propose to replace theoretical travel times with travel times extracted from actual data. An alternative is to use probabilistic distributions to estimate the actual travel time variability to feed PT simulation framework. We propose an effective ML technique along with probabilistic distributions, with aim of providing the simulation framework with more accurate travel times than those estimated with probabilistic distributions.

53 In this paper, we suggest a model development methodology based on a multi-agent simulation framework and data inputs extracted by a hybrid approach combining ML trained on actual bus data to predict travel times and probabilistic distributions to accurately estimate travel time variability. We also propose a two-step validation framework that exhibits the performance of the obtained model on a case study based on actual data provided by an operator. A first contribution of this paper is to propose effective ML techniques to predict the travel time in a real scenario, so as to bypass the computational efforts in the simulation platform to mimic congestion. A second contribution of this work is to couple ML and public transport simulation model in the same framework, which to our knowledge is rarely implemented.

62 The remainder of this article is structured as follows: Section 2 reviews related literature including ML models and their application for bus travel prediction, as well as simulation models with emphasis on traffic models. Section 3 provides an overview of the proposed ML-based microscopic simulation framework and then details each of the layers of the framework. Additionally, the two-step validation framework to assess the performance of the proposed ML-based simulation model will be presented in detail. A case study scenario will be given in Section 4. In Section 5, the simulation results will be analyzed and validated based on a set of performance indicators. A discussion of the analyses presented is in Section 6 and some concluding remarks in Section 7.

69 2. Literature Review

70 Over the past decade, the problem of bus travel time prediction has received wide attention. In this section, we first
71 review related works, which can be categorized into prediction of bus arrival time and bus travel time. Furthermore,
72 we present a number of traffic simulation frameworks and studies that attempt coupling ML with simulation models,
73 as well as the contributions of this work.

74 2.1. Bus Arrival Time Prediction

75 Estimated Time of Arrival (ETA), also known as expected time of arrival, is the time at which a transportation
76 system is expected to arrive at its destination. The term has long been used in aviation and maritime transport, but also
77 increasingly in road transport. Various methods have been adopted by researchers to predict the expected arrival time
78 of buses, these methods can be grouped as follows: 1) historical approaches predict the travel time at a given time
79 as the average travel time for the same period on different days [27; 28; 29; 30]; 2) real time approach predicts that
80 the next time interval travel time will be the same as the current travel time [21]; 3) regression models are classical
81 approaches for predicting travel time and predicting a dependent variable based on a function formed by a set of
82 independent variables [29; 30; 31; 32]; 4) Kalman filter is a popular tool for the recursive estimation of variables
83 that characterize a system, it is a model-based estimation scheme that takes into account the stochastic properties of
84 process disturbance and the measurement noise [31; 33; 34; 35; 36]; 5) ML techniques have been widely reported for
85 traffic and travel time prediction [33; 37; 38; 39; 40].

86 2.2. ML for Bus Arrival and Travel Time Prediction

87 In past decades, ML models have been widely applied in the transportation field, due to their ability to solve
88 complex problems and extract patterns. However, ML approaches are data-driven techniques, requiring a large data
89 set and high computation time. Many applications of traditional ML methods have been implemented to predict traffic
90 data [41; 42; 43; 44] and traffic congestion [45; 46], reinforcement learning approaches have been incorporated into
91 traffic control systems [47; 48], while the models of neural network and deep learning have been widely reported,
92 including traffic flow prediction [49; 50; 51; 52], travel demand forecasting [53; 54; 55; 56; 57], traffic signal control
93 [58; 59], driving behavior analysis [60; 61] and autonomous driving applications [62; 63].

94 *Bus Arrival Time Prediction.* ML models based on neural network (NN), tree-based, **Support Vector Machine (SVM)**
95 and regression have been proposed for bus arrival time prediction. Chien et al. [33] proposed two artificial neural
96 networks to address the dynamic bus arrival time prediction problem. Predicted travel times are then assessed with a
97 microscopic simulation model, calibrated and validated with real-world data. Pan et al. [64] introduced a self-learning
98 algorithm based on a back-propagation neural network for predicting bus arrival time based on historical data. Li et al.
99 [65] proposed a random forest-based approach for bus arrival time prediction, using travel data from the Beijing city
100 bus network. Yin et al. [66] proposed a model for prediction bus arrival time at stops with multiple routes using the
101 Support Vector Machine algorithm. Yu et al. [67] has applied several ML algorithms including: SVM, NN, k-nearest
102 neighbours algorithm (kNN) and linear regression (LR), to predict bus arrival time at bus stop with multiple routes.
103 In another work, Shalit et al. [68] introduced a supervised ML method for predicting passengers' boarding stops.

104 Some of the above mentioned methods are used in a hybrid manner. Yang et al. [39] implemented an approach
105 in which SVM with Genetic Algorithm (GA) were used to predict bus arrival time, using different features including:
106 road length, weather conditions and travel speed. Chen et al. [69] presented a dynamic bus arrival time prediction
107 model based on real-world data. The model consists of: 1) An artificial neural network (ANN) model to predict the
108 bus travel time between two stops; 2) A Kalman filter-based dynamic algorithm to adjust the arrival time prediction
109 using the bus location information down to the minute. In another work, Zhang et al. [70] proposed a method called
110 CK-means (K-means Clustering) to predict the arrival time of buses using real time online bus locations. The K-means
111 clustering method is used to aggregate historical traffic data and calibrate the operating status of the road section.

112 *Bus Travel Time Prediction.* Many studies based on [deep learning \(DL\)](#), SVM and tree-based models have been
113 devoted for predicting bus travel time. Chen et al. [71] applied a deep learning method with a back-propagation
114 neural network to predict bus travel time. The proposed approach was then validated with real traffic data. In another
115 work, He et al. [72] used a DL model to predict bus travel time by taking into account passenger's riding time across
116 multiple bus trips and waiting time at transfer bus stops. To do this, the entire journey is partitioned into bus riding
117 components and waiting components, each of the components is predicted separately and the results are merged to
118 obtain the final bus travel time. Junyou et al. [73] applied the SVM algorithm for bus travel time prediction. In another
119 work, Yu et al. [74] predicted bus travel time using random forests based on near neighbors (RFNN). To predict the
120 bus travel time between adjacent bus stops, the proposed model takes as input the bus dwell time of the current stop
121 and the current traffic conditions on the predicted route segment and next segments. Recently, Ashwini et al. [75]
122 proposed a comparative study of number of linear and non-linear models including: LR, Support Vector Regression
123 (SVR), Regression Trees (RTs) and Random Forest Regression, to identify a suitable model for travel time prediction.
124 In another work, Serin et al. [76] applied and compared a number of ML methods, including tree-based regression
125 algorithms and SVR, to predict bus travel time.

126 In some works, ML methods combined with Kalman filtering-based algorithm were proposed. Bai et al. [77]
127 proposed a dynamic travel time prediction model for buses dealing with on-road cases with multiple bus routes, based
128 on SVMs and Kalman filtering-based algorithm. The SVM model predicts baseline travel times from historical bus
129 trip data. Travel times can then be adjusted using a Kalman filtering-based algorithm, based on the latest information
130 on bus operations and baseline travel times. In another work, Kumar et al. [78] proposed an approach based on kNN
131 classifier and model-based Kalman filtering for real-time bus travel time prediction.

132 *Transit Delay Prediction.* Many works focused on predicting transit travel delay. Shoman et al. [79] proposed a
133 deep learning-based framework to predict bus delays at the network level, in which the framework is fed by large,
134 heterogeneous bus transit data (GTFS) and vehicle probe data. Wu et al. [80] proposed a Random Forest based
135 approach to forecast multi-scenario train delays.

136 2.3. *Traffic Simulation Models*

137 Traffic simulation can be defined as the mathematical model of transportation systems, implemented through
138 the application of dedicated computer software. Traffic simulators can be divided according to the level of details
139 provided into: microscopic and macroscopic approaches. Microscopic traffic simulator focuses on the movement of
140 each individual entity in the system. It provides a detailed picture of each individual vehicle including: location, time
141 and speed [81]. Microscopic models are very effective in assessing heavily congested conditions. On the other hand,
142 these traffic models are time-consuming, costly, and can be difficult to calibrate. Macroscopic traffic simulators are
143 based on deterministic relationships of flow, speed and density of traffic flow [82]. They were originally developed to
144 model traffic in distinct transportation sub-networks, such as freeways and rural highways.

145 *Open-Source Simulation Models.* Over the past decades, a number of traffic simulation frameworks have been devel-
146 oped. TRANSIMS developed at Los Alamos National Laboratory (USA) is an integrated set of tools for performing
147 the analysis of a regional transportation system based on a cellular automata microscopic simulation model. It models
148 individual travelers and their multi-modal transportation based on synthetic populations and their activities. TRAN-
149 SIMS represents time in a consistent and continuous way, as well as detailed persons and households [83]. MIT
150 (Massachusetts Institute of Technology) developed MITSIMLab, an open-source microscopic traffic simulator that
151 evaluates the impacts of alternative traffic management systems, public transport operations and various [Intelligent](#)
152 [Transportation Systems \(ITS\)](#) strategies at the operational level and helps in their further refinement. In order to cap-
153 ture the sensitivity of traffic flows to control and routing strategies, the traffic and network components are detailed
154 in MITSIM [84]. MATSim (Multi-Agent Transport Simulation) developed by ETH Zurich, has in the recent years
155 become a major open-source framework, used to implement large-scale agent-based transport simulations, in which a
156 large number of synthetic persons (so-called "agents") are simulated. Designed as a framework for large-scale scenar-
157 ios, MATSim is highly modular, allowing for a very high level of customization [85]. SUMO (Simulation of Urban
158 MObility) was developed at the German Aerospace Centre. It is a microscopic multimodal traffic simulator capable
159 of simulating different types of traffic data, in which vehicles, public transport and persons are modeled explicitly. It
160 can provide a nice visual and understandable output for future tests and analysis SUMO [81]. Leblond et al. [86]

161 proposed Starling, an agent-based simulation software. It was developed as a generic framework to deal with spatial
162 issues of territories. The simulation framework is designed to evaluate a specific mobility service.

163 *Commercial Models.* In addition to open-source models, a number of commercial software have been implemented.
164 PTV Vissim, a microscopic multi-modal traffic flow simulation software package developed by PTV (Planung Trans-
165 port Verkehr AG) in Karlsruhe, Germany. VISSIM allows users to define a full range of vehicle types including
166 passenger cars, buses, trucks, rail vehicles as well as pedestrians and cyclists. It is widely used to simulate, evaluate
167 and validate new transport policies and control systems [87]. TransModeler, a traffic simulation platform marketed
168 by Caliper Corporation (USA), it is used to model traffic planning, traffic management and emergency evacuation
169 scenarios over a wide-area [88].

170 2.4. Coupling ML and Simulation Models

171 In recent years, several studies attempt to perform ML with simulation models. Shafizadeh et al. [89] coupled ML
172 including SVM and ANNs, tree-based models and statistical models with cellular automata to simulate urban growth
173 in the city of Tehran. Tongal et al. [90] proposed a simulation framework to explore the effectiveness of different
174 ML approaches in streamflow simulation for four rivers in the United States. Yan et al. [91] proposed a method of
175 coupling NNs and numerical models to simulate and identify areas at high risk of urban flooding and to predict the
176 depth of water accumulation. The outputs of the simulation models are used to feed the neural network. Chabanet et
177 al. [92] proposed a method based on active learning concepts to combine a computationally costly simulator with ML
178 classifier (kNNs), which is less computationally costly to use online but whose predictions are only approximations
179 of the simulator. First, the framework tries to classify the unlabeled instances with kNNs classifier, otherwise, the
180 simulator will be performed then. Shahhosseini et al. [93] proposed an approach to couple crop modeling, a simulation
181 model with ML models, to improve corn yield prediction in the US Corn Belt. The integration of simulation results
182 improved the yield prediction accuracy of ML models. Abdelaty et al. [94] used the simulated energy consumption
183 scenarios to develop four different data-driven modelling techniques.

184
185 The use of framework combining ML models and public transportation simulators remains poorly implemented.
186 An interesting work that addressed a similar problem to our work has been proposed by Othman et al. [95], in which a
187 framework integrating NN models into a PT simulation model to improve real-time supply based on multiple demand
188 scenarios. First, a Multi-Layer Perceptron coupled with a LR model trained on traffic data and weather information
189 was used to predict the type of congestion, duration of congestion and hence delayed travel times. Then, the travel
190 times will be fed into a simulator, in which several scenarios will be simulated.

191 From a general point of view, the framework proposed in [95] and our framework follow a similar pattern, in which
192 the actual data is trained to provide public transport simulator with more accurate travel times. The main difference
193 lies in the type of data used as ML input data, with traffic data in [95] and actual bus stop times extracted from AVL
194 (Automated Vehicle Location) data in this proposed framework, respectively. In this proposed work, the impacts of
195 traffic on the bus routes (including traffic congestion) are indeed included in the actual bus travel times.

196 2.5. Study Contributions

197 ML is a powerful tool, but it faces certain limitations. One of the main drawbacks is that it can be biased.
198 Algorithms implemented in ML are designed to look for patterns in data. If there is a bias in the data, ML algorithms
199 will detect and reinforce that bias. Another disadvantage is that ML can be computationally expensive, slow and
200 resource-intensive to train successfully, especially on large datasets. On the other hand, simulation has long been
201 touted as a powerful tool for understanding complex systems. By building models of how systems work, we can
202 better understand how they might behave in the future. However, simulation has its drawbacks. First, the simulation
203 is only as good as the models built. If the built models are inaccurate, the simulation performed will be inaccurate
204 as well. Second, simulations are only approximations. In a number of situations, simulation models are not capable
205 of qualifying all the variables that affect the behavior of the system. In very large and complex problems, the large
206 number of variables and the relationships between them makes the problem very difficult to model.

207
208 In an effort to overcome the limitations of simulation and ML models, a framework for a ML-based microscopic
209 simulation model is proposed, to provide more accurate and realistic simulations for public transport operators, which

Reference	Classical ML Model	NN & DL	Simulation Model	ML and Simulation	Arrival Time Prediction	Travel Time Prediction	Delay Prediction	Historical Data	Real-time Data	Description
Ashwini et al. [75]	x					x		x		A study to identify a suitable ML model for bus travel time prediction
Bai et al. [77]	x					x		x	x	Dynamic bus travel time prediction models with multiple bus routes
Chen et al. [69]		x			x			x		Dynamic bus arrival time prediction model based on APC real-world
Chen et al. [71]		x				x			x	ML approach to predict bus travel time validated with real traffic data
Chien et al. [33]		x	x		x					Dynamic bus arrival time prediction with artificial neural networks
Gal et al. [96]	x				x	x		x	x	Traveling time prediction in scheduled transportation with trip segments
He et al. [72]		x				x		x	x	Travel-time prediction of bus journey with multiple bus trips
Jeong et al. [29]		x			x			x		Bus arrival time prediction using artificial neural network model
Junyou et al. [73]	x					x		x		Support Vector Machine algorithm for bus travel time prediction
Kumar et al. [97]		x			x					Kalman filtering and ANNs approaches for bus arrival time prediction
Kumar et al. [78]	x					x		x		kNN and Kalman filtering model for real-time bus travel time prediction
Lam et al. [98]		x			x				x	Prediction of bus arrival time using real time online bus locations
Li et al. [65]	x				x			x		Bus arrival time prediction based on random forest
Othman et al. [95]		x	x	x		x		x		Machine learning aided simulation of public transport utilization
Pan et al. [64]		x			x			x		Self-learning algorithm based on a BPNN for predicting bus arrival time
Panovski et al. [99]	x	x	x		x				x	Bus arrival time prediction at different bus stops on a given bus route
Shalit et al. [68]	x				x			x		A supervised ML model for imputing missing boarding stops
Serin et al. [76]	x					x		x		Predicting bus travel time using machine learning methods
Shoman et al. [79]		x					x			Deep learning framework for predicting bus delays on multiple routes
Wu et al. [80]	x						x		x	The bounds of improvements toward real-time forecast of train delays
Yang et al. [39]	x				x					Bus arrival time prediction SVM with GA
Yin et al. [66]	x	x			x			x		Prediction bus arrival time at stops with multiple routes
Yu et al. [67]	x	x			x				x	Bus arrival time prediction at bus stop with multiple routes
Yu et al. [74]	x					x				Predicted bus travel time based on near neighbors
Zhang et al. [70]	x				x			x		K-means approach to predict bus arrival times based on GPS data
Proposed work	x		x	x		x		x		A microscopic simulation of public transportation based on ML

Table 1: Comparative table of related works

210 will be useful to enhance the reliability of service. Although a number of studies have addressed the problem of bus
211 travel time and arrival time prediction, by implementing ML models. Most studies focus on prediction a single travel
212 time value. In this study, the focus is on more accurately estimating travel time variability rather than a single value.
213 This work contributes in the literature by: 1) using a ML model with a traffic simulator in the same framework for
214 public transportation simulation is not widely implemented; 2) A ML model for travel time prediction, trained on
215 actual data, combined with probabilistic distributions to estimate travel time variability, will be used to power the
216 simulation, helping to reduce simulation complexity and providing the simulator with appropriate parameters; 3) The
217 simulations will be more realistic, as they will be validated with performance indicators directly calculated from actual
218 data.

219
220 In Table 1, we present a comparison of the proposed work with other works on the following aspects: whether
221 ML models (e.g. classical model, neural networks and deep learning models) have been used; whether simulation
222 models have been carried out; if both ML and simulation models are a part of the implemented framework; whether
223 the focus of the work is travel metrics (e.g. arrival time, travel time and delay), as well as the nature of the data used
224 (e.g. historical and real-time data).

225 **3. Methodology**

226 **3.1. ML-Based Microscopic Simulation Framework**

227 Before outlining the proposed framework for a ML-based microscopic simulation model and its different layers,
 228 an overview of the proposed approach is presented as follows. In the first stage, actual data (GPS traces) and transit
 229 schedule data (e.g. GTFS data) will be matched and then aggregated at the route segment level. In the second stage, the
 230 process of building ML model is conducted. A ML algorithm is first selected, then feature importance techniques will
 231 be applied to select relevant features, then ML will be trained for bus travel time prediction. Finally, the generated
 232 ML models, combined with probabilistic distributions to estimate bus travel time variability, will be deployed and
 233 connected to the public transport simulation model. In the third stage, the transit scenario will be implemented, and
 234 then simulated. During the simulation process, bus travel times will be requested by the simulator. On the other hand,
 235 deployed models, will process each request and provide the simulator with the predicted travel times. At last, the
 236 simulation results will be validated with actual data using a set of metrics. An overview of the ML-based microscopic
 237 simulation framework is shown in Figure 1.

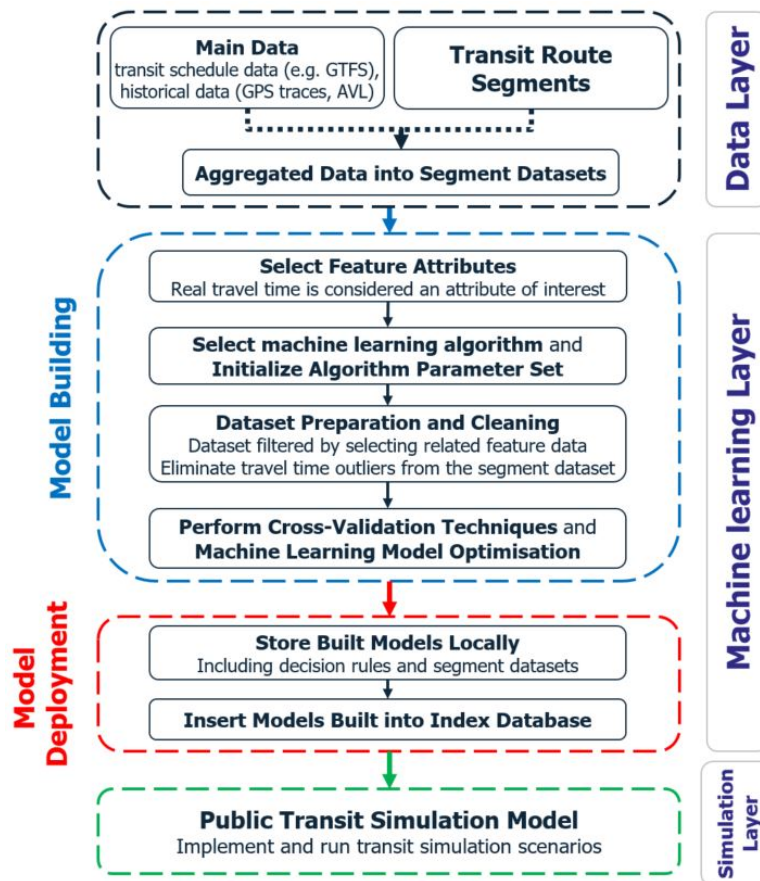


Figure 1: An overview of the ML-based microscopic simulation framework. The framework consists of three layers: 1) data layer, 2) ML layer, and 3) simulation layer. 1) Transit schedules and actual data are used to generate aggregated datasets at the segment route level. 2) The ML model for segment travel time prediction takes place in two stages: A) model building and B) model deployment. A) ML model is first initialized by selecting the model features, the supervised algorithm and configuring the model parameters. Next, the segment route dataset is prepared, cleaned, fed to train and adjust the model by performing cross-validation and optimisation techniques respectively. B) Model deployment consists of storing the model locally, and indexing it in a database. 3) A simulation scenario is implemented and then simulated, in which ML models will feed the simulation model with travel times

238 **3.1.1. Data Introduction**

239 In this section, the spatial and temporal components of the bus line will first be presented, then the data sources
 240 and data aggregation process used to train the ML models will be detailed.

241 **Bus Route Components.** A transit bus travels from an origin to a destination passing through a set of stops along
 242 the way. Spatial components can be grouped into several levels: stop, section, segment and route. The link between
 243 two consecutive stops is called *section*. A more aggregate section, called *segment*, the latter is made up of several
 244 consecutive sections. All sections from an origin terminal to a destination terminal form a *route* [100]. Additionally,
 245 and in order to maintain acceptable service reliability, most bus operators set up control points along the bus route.
 246 Control points are timing points, specific transit stops where the bus departure times are subject to regulation or
 247 to meet a specific buffer time [4]. An illustration of the spatial components of bus lines is given in Figure 2.

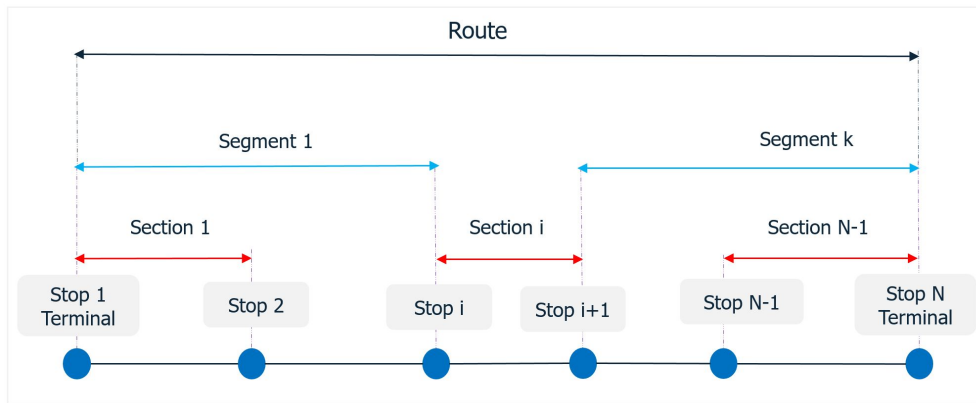


Figure 2: An illustration of the different spatial components of a transit line

248 The time components of bus routes can be distinguished into durations and time points. Travel time, dwell time
 249 and running time are durations, while arrival time and departure time at bus stops are points in time. The travel time
 250 is made up of dwell time and running time respectively, referring to the time a bus spends stationary at a scheduled
 251 stop and the time the bus spends while not stopping at a stop. Operating times can be expressed at the level of:
 252 section, segment or route. Section travel time corresponds to the sum of the dwell time and the running time, between
 253 two successive stops. Segment travel time corresponds to the sum of the travel times of the sections constituting the
 254 considered segment. While route travel time refers to the total travel time comprised between the first stop and the
 255 last stop. An illustration of the time components of bus lines at the section level is given in Figure 3.

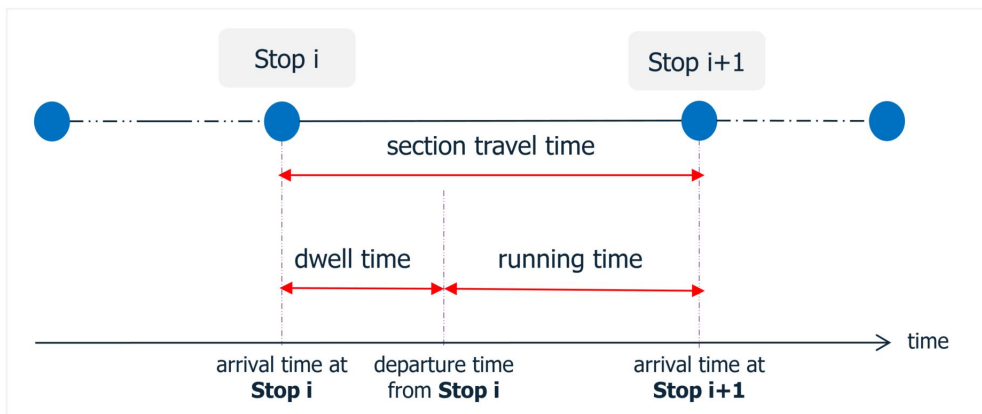


Figure 3: An illustration of the different components of travel time between two transit stops

256 *Data Used.* The main data for the application of the proposed approach can be distinguished into *GTFS* transit data
 257 (General Transit Feed Specification) and actual data. Transit schedules are grouped by transit line (e.g. bus line) and
 258 formatted as *GTFS* tables. A *GTFS* feed consists of several comma-separated files that mainly contain information
 259 about transit lines, stop locations, and timetables [101]. On the other hand, actual data provides information about
 260 transit trips during a period of the year, including: date, vehicle ID, corresponding trip, and recorded stop times. From
 261 a technical point of view, actual data is structured similarly to a *GTFS* stop times table. Practically, actual stop times
 262 are recorded from GPS trackers on transit vehicles (e.g. buses), at each transit stop. Therefore, scheduled and actual
 263 stop times can be matched in the same data structure based on mutual attributes, such as trip id and stop id.

264
 265 After matching scheduled and actual data in the same dataset. The next step is to aggregate the observations
 266 originally available at the stop-level, to a more aggregated level of details, for instance at the section and segment
 267 level. The data aggregation process aims to mitigate the impacts of imprecise stop times, since this latter is susceptible
 268 to measurement errors, which can therefore lead to a poor quality ML model. It takes place in two steps:

- 269 • As a first step, public transport observations at bus stop level will be aggregated at section level. From scheduled
 270 and actual stop times, the corresponding travel times can be simply derived at the section level.
- 271 • As a second step, certain criteria can be used to define control points, specific transit stops where the buses
 272 are subject to regulation or to meet a specific buffer time. Depending on these timing points, a set of route
 273 segments can be defined. Therefore, the observed transit stop times can be aggregated from the section level to
 274 the segment level. Arrival time, departure time and travel time will be derived.

275 For the actual data at hand, only bus arrival times are available. In the remainder of this paper, the bus stop time
 276 will refer to the bus arrival time. Bus route components, as well as the stop and the travel times of a transit trip,
 277 aggregated at the stop, section and segment levels, are shown in Figure 4.



Figure 4: An illustration of the components of a bus route with stop and travel times

278 The resulting stop-level and section level datasets are illustrated in more details in [Appendix A](#).

279 Transit schedules are mainly planned according to the type of the day. Since demand for public transport is strongly
 280 linked to the type of the day, the transit demand is greater on working days than on weekends or holidays. Therefore,
 281 the day type attribute will be used to implement particular scenarios, for instance the working days scenario. In
 282 addition, in order to understand the status of the transit vehicle relative to the scheduled times, delay measures can be
 283 estimated at the stop level, and therefore at the segment level. The stop-level delay formula is given in Equation 1.

$$D_{ijk} = AT_{ijk} - ST_{ijk} \quad (1)$$

284 Where:

- 285 • D_{ijk} : refers to the delay time for line i , on trip k , at bus stop j
- 286 • AT_{ijk} : refers to the actual stop time for line i , on trip k , at bus stop j
- 287 • ST_{ijk} : refers to the scheduled stop time for line i , on trip k , at bus stop j

An illustration of derived attribute data is given in Table 2.

from stop	to stop	segment index	...	from stop delay	to stop delay
STP-1883	STP-982	1	...	00:00:22	00:01:10
STP-982	STP-990	2	...	00:01:10	00:02:15
STP-990	STP-994	3	...	00:02:15	00:01:07
STP-994	STP-998	4	...	00:01:07	00:01:38
STP-998	STP-1005	5	...	00:01:38	00:00:10

Table 2: A sample of a segment-level aggregated dataset with focus on derived stop delays

288
289

In the remainder of this paper, the scheduled stop time will simply be referred as stop time.

290 3.1.2. ML: Building and Deploying Models

291 One of the main purposes of ML is to discover patterns, train a model on historical data, and then make pre-
 292 diction based on the constructed model. The ML model to built aims to predict the transit bus travel time at the
 293 segment-level. Therefore, *real travel time* attribute will be considered as the attribute of interest. Furthermore, feature
 294 importance techniques, in which a score is assigned to input features based on their usefulness in predicting a target
 295 variable, will be applied to build ML models with relevant features. Next, a ML algorithm will be chosen and ini-
 296 tialized with a set of predefined parameters. For instance, a decision tree supervised learning algorithm, which was
 297 originally designed to solve classification problems, has been extended to deal with regression problems [102]. On
 298 the other hand, the developed approach aims to predict transit travel time (numerical output), regression trees (RT) are
 299 one of the suitable models to perform. The proposed scheme for initializing the ML model is illustrated in Figure 5.

300 ML algorithms, for instance, regression trees designed to solve regression problems, are trained to return a single
 301 value as output. In order to capture travel time variability, we propose to combine ML algorithm with probabilistic
 302 distributions as follows. First, the ML algorithm will be trained to predict travel times, these will then be grouped into
 303 samples according to the resulting regression rules. Then, for each sample, travel time variability will be estimated
 304 using probabilistic distributions. An illustration of a regression tree model for bus travel time prediction and sample
 305 travel times resulting from the model output, fitted to a number of probabilistic distributions, to approximate travel
 306 time variability, are given in Figures 6 and 7, respectively.

307

308 After validating the ML model, the next step is to deploy it to feed the simulation model with bus travel times.
 309 At this stage, the model built will be registered in a specific database, in which each model is indexed by its key,
 310 as well as a set of information, including: model algorithm, transit line, route segment origin and destination stops,
 311 model input features and target attribute. In addition, the index database may contain the model parameters and its
 312 error value, such as the coefficient of determination R^2 . This last score value is extremely important for sorting the
 313 generated models. An illustration of the index file is presented in Table 3.

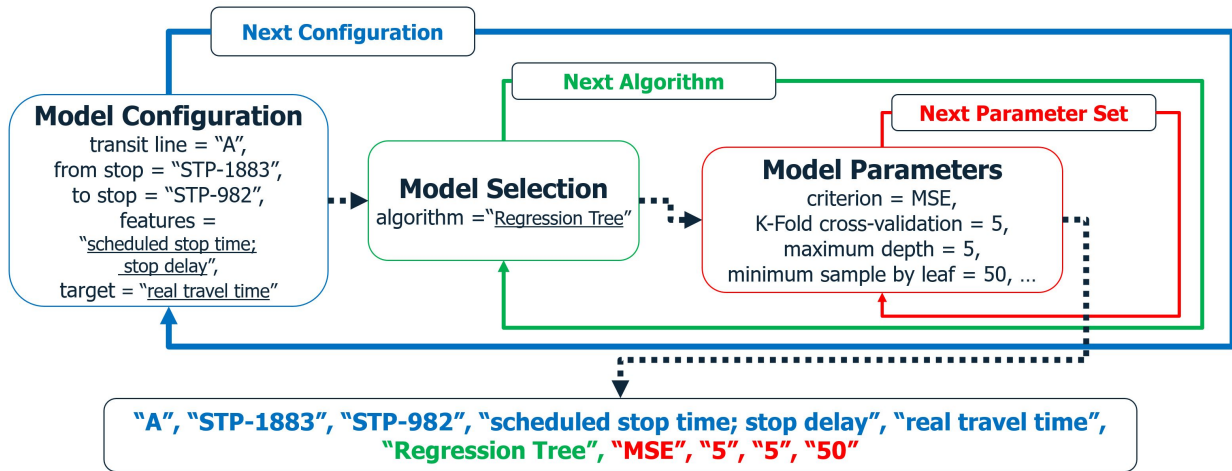


Figure 5: An illustration of the proposed scheme for setting up ML models. The initialization of a ML model is done at three levels. 1) Model configuration by initializing ML model features and target attribute, as well as route segment information such as transit line, origin and destination stops. 2) Selection of the supervised algorithm, for instance the regression tree. 3) Configuration of the algorithm parameters

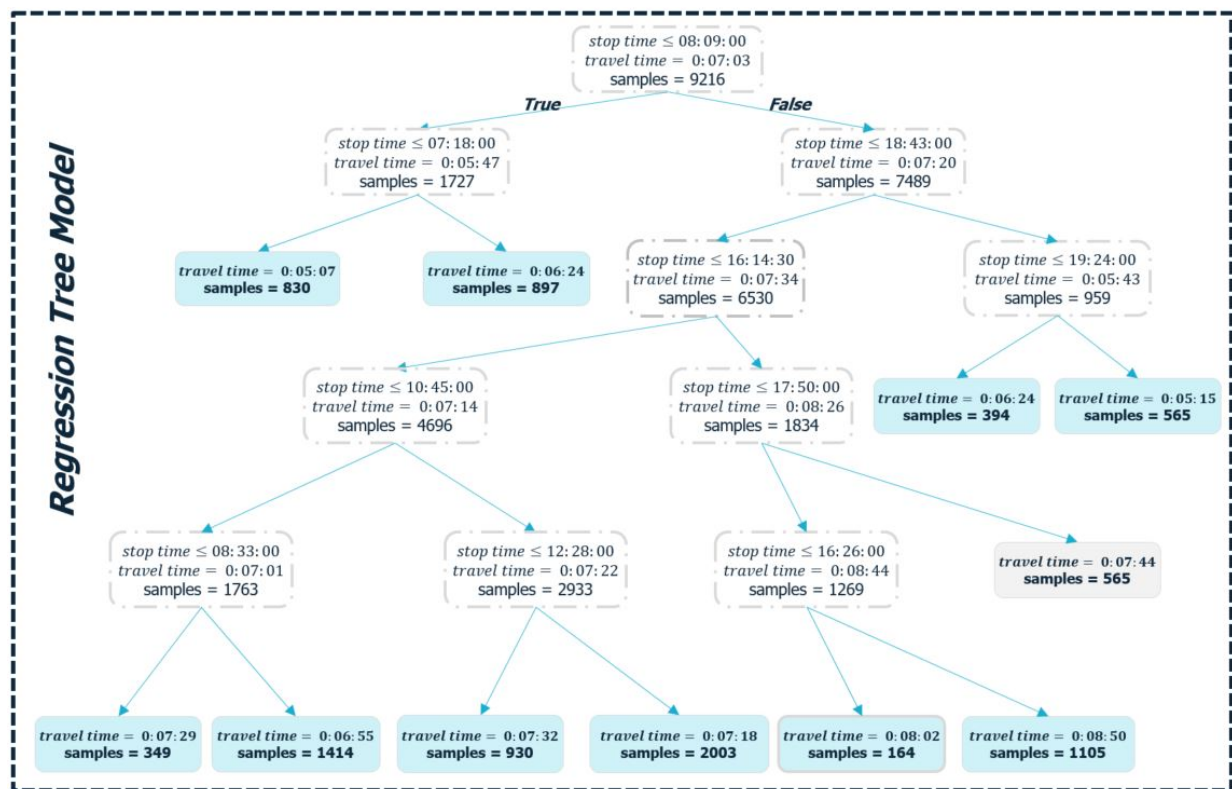


Figure 6: An illustration of a regression tree for travel time prediction. The white nodes refer to the decision conditions, while the cyan nodes are the terminal leaves of the tree

314 In sum, a ML algorithm is initialized with predefined parameters, takes as input: a route segment dataset, features
 315 to build the model, and an attribute of interest for prediction. It will then be trained to predict transit travel time, stored
 316 locally and then deployed to feed a public transit simulation model.

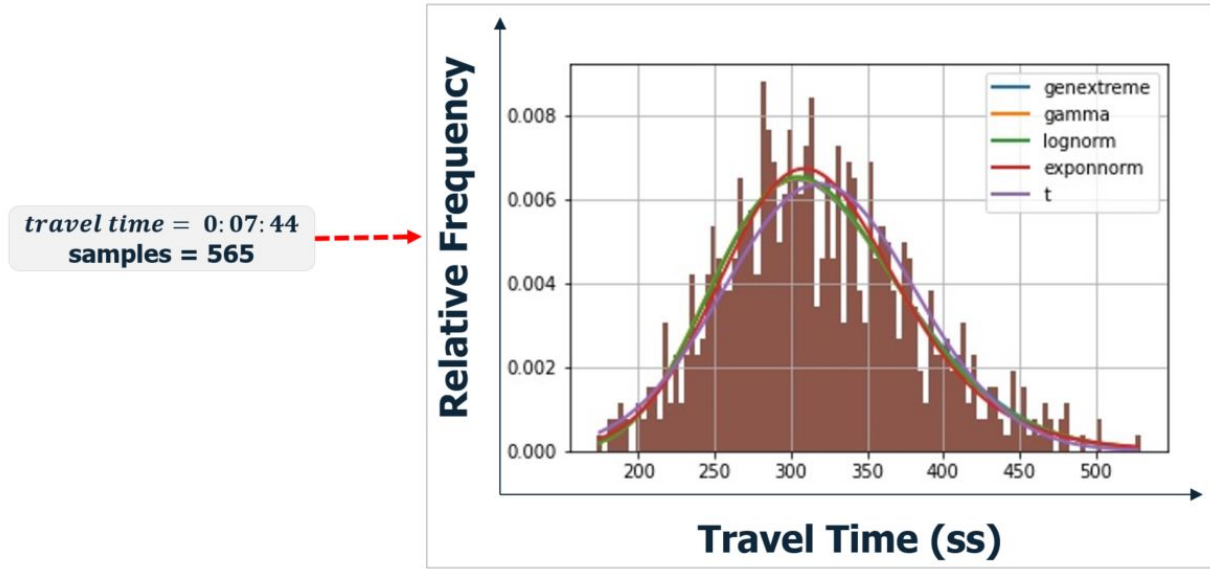


Figure 7: A detailed view of the proposed model output for travel time prediction. Illustration of the resulting day-to-day travel time variability (Left): A regression tree terminal node, including average travel time and sample size. (Right): The sample travel time distribution, fitted to a number of probabilistic distributions, to approximate day-to-day travel time variability with emphasis on probability density functions

key	algorithm	transit line	from stop	to stop	input features	model target	score (R ²)
68dsfnr6	RT	A	STP-1883	STP-982	stop time	real travel time	36%
chhm3moj	RT	A	STP-982	STP-990	stop time	real travel time	54%
wr1w7ecg	RT	A	STP-990	STP-994	stop time	real travel time	47%
z4p0biet	RT	A	STP-994	STP-998	stop time	real travel time	35%
jj8gxgzf	RT	A	STP-998	STP-1005	stop time	real travel time	41%
...
cq5x4ddu	RT	A	STP-1883	STP-982	stop time;stop delay	real travel time	36%
53t58nae	RT	A	STP-982	STP-990	stop time;stop delay	real travel time	53%
1y1givje	RT	A	STP-990	STP-994	stop time;stop delay	real travel time	46%
qvin2kyt	RT	A	STP-994	STP-998	stop time;stop delay	real travel time	34%
yiikt0cl	RT	A	STP-998	STP-1005	stop time;stop delay	real travel time	42%

Table 3: A sample of the deployed models index database. Index information can be grouped into four groups. 1) *key* used to distinguish models with same other characteristics. 2) *transit line*, origin and destination stops, denoted *from stop* and *to stop*, refer to model segment route. 3) *input features* and *target attribute* are the data attributes used to build the model. 4) *score* refers to the evaluation measurement value

3.1.3. Public Transportation Simulation Model

In this section, *Starling* proposed by Leblond et al. [86] will be briefly described. *Starling* combines an agent-based framework and a discrete-event approach, it is a microscopic model for mobility simulation. The goal of this framework is to provide a basis for the development of computer models for the simulation of specific transport systems, consisting of generic simulation classes that can be extended to match the specifications of the simulated system. Among the mobility modules, a public transport bus simulation module is implemented. To work, it requires a set of inputs including: transit schedules feed (e.g. GTFS data) and bus operator settings that define the behaviour of the bus vehicle and the operating rules, such as: minimum bus recovery time and bus holding control strategy. Simulating buses according to schedules, with defined travel times between two successive stops, can be seen as hypothetical. The bus may be delayed due to: traffic jams, traffic accidents and passenger boarding.

327 In order to perform a bus simulation with a more realistic behaviour, travel times can be estimated using ML
 328 models trained on historical observations. The generated travel times are then subject to operating rules, in order
 329 to maintain realistic bus behaviour. Additionally, in order to perform transit simulation, scenario data and operator
 330 parameters will be added to the simulator inputs. An overview of how deployed ML models are used with a generic
 331 framework for public transport simulation is shown in Figure 8.

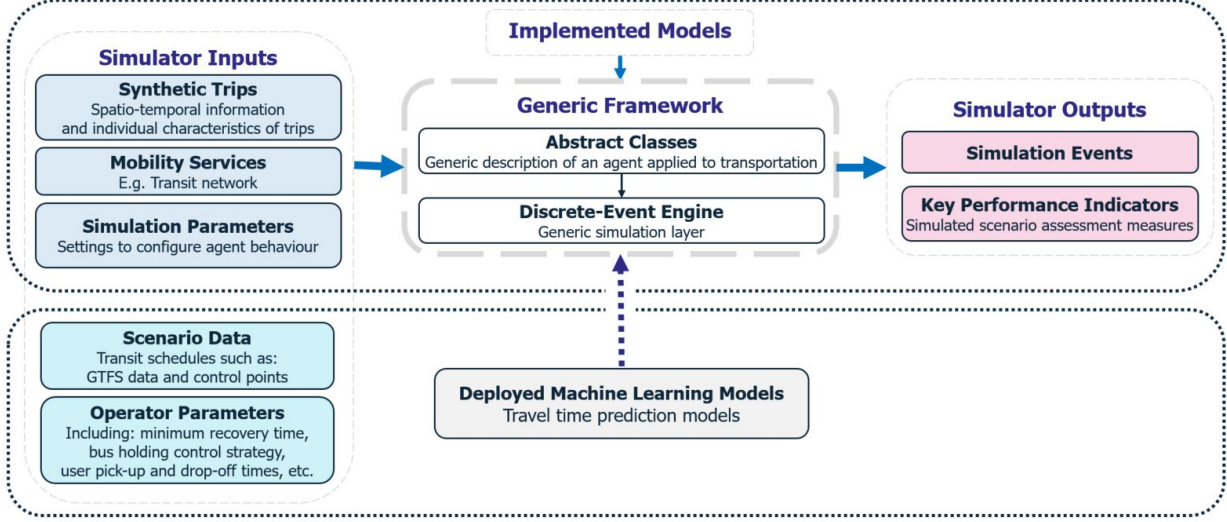


Figure 8: An overview of using a generic framework for public transport simulation based on ML models. It consists of two blocks: the block above refers to the simulation framework; the block below gathers the scenario data and the operator parameters as well as the ML models deployed for the prediction of travel times

332 *Holding Control Strategies.* They are among the most widely used public transport control methods, aiming to
 333 improve the regularity of service by regulating departure time from stops according to predefined criteria [103].
 334 Implementing bus holding strategies involves two key design decisions: selecting the set of time point stops and the
 335 holding criteria [104]. Holding strategies are generally classified into two categories: schedule-based strategies and
 336 headway-based strategies. A schedule-based holding strategy defines the earliest time a bus can depart from a time
 337 point stop relatively to the schedule. This rule can be formulated as:

$$ET_{ijk} = \max(SET_{ijk} - s_{ij}, AT_{ijk} + DT_{ijk}) \quad (2)$$

338 Where:

- 339 • ET_{ijk} : exit (departure) time for line i on trip k from stop j
- 340 • SET_{ijk} : scheduled exit (departure) time for line i on trip k from stop j
- 341 • s_{ij} : non-negative slack size defined for line i at stop j
- 342 • AT_{ijk} : actual arrival time for line i on trip k from stop j
- 343 • DT_{ijk} : dwell time for line i on trip k from stop j

344 Previous studies on the interaction between slack size and generalized passenger travel time have concluded that
 345 the slack size should be set to zero [105; 106]. This implies that buses arriving early have to wait at time point stops
 346 until the scheduled departure time. Schedule-based strategies are useful for low-frequency services when passengers
 347 follow the timetable or when transfer coordination is an important issue [107]. In contrast, headway-based holding
 348 strategies use headways between consecutive vehicles as a criterion to regulate departure times from time point stops.
 349 If the headway-based strategy only takes into account the headway from the preceding vehicle, then the holding
 350 criteria is defined by a minimal headway requirement:

$$ET_{ijk} = \max(AT_{ij,k-1} + \alpha \cdot H_i^{k-1,k}, AT_{ijk} + DT_{ijk}) \quad (3)$$

Where:

- $H_i^{k-1,k}$: planned headway between trips $k-1$ and k on line j
- α : threshold ratio parameter

Stop-Skipping. Also known as expressing is a control measure allowing a vehicle to skip one or a series of stops, if it is running behind schedule. Stop-skipping can correct service inconsistencies due to the inherent variations in travel time and passenger demand, but may result in increased waiting times for passengers waiting at the skipped stops [108]. Thus, most stop-skipping approaches address the problem by considering: 1) passenger-related costs concerns, such as passenger waiting times and their in-vehicle times; and 2) operator cost concerns, such as total vehicle trip travel times [109]. Stop-skipping strategies can be designed either at the tactical planning level or at the operational level (dynamic stop skipping). At the tactical planning level, the emphasis is on developing reliable, resilient or robust strategies that will maintain good performance in the event of disruptions during actual operations. On the other hand, dynamic stop-skipping strategies at the operational level are reactionary, need to be computationally efficient, and in which the skipped stops of a vehicle trip are determined just before dispatch [109; 110; 111]. Stop-skipping strategies have been proposed mainly at the planning level and formulated as optimization problems [108; 112]. When stop-skipping, the vehicle leaves the stop immediately after the alighting time without boarding of passengers [113]. The departure time can be formulate as:

$$ET_{ijk} = AT_{ijk} + DT_{ijk} \quad (4)$$

In this case, the dwell time DT_{ijk} corresponds to alighting only.

3.2. Two-Step Validation Framework

The second proposed framework in this study is the two-step validation framework. It aims to validate the simulation results obtained using performance indicators. The validation is twofold: first, assessing the accuracy of travel times generated by ML models; second, evaluate how well the simulation model reproduces reality using certain reliability metrics. The proposed validation process is detailed below. From a technical point of view, the validation process will be undertaken by comparing the output simulation results with actual observations, including: 1) simulated segment travel times and 2) bus delay at timing points.

The first validation process will take place as follows. First, the theoretical and actual travel times will be extracted from the actual data, while the simulated travel times will be extracted from the simulation results. Next, the theoretical and simulated travel times will be compared to the actual observed travel times in order to: firstly, understand how well the actual and theoretical travel times are fit and secondly, validate the ML models for travel time prediction. We denote by TT_{actual} and TT_{sim} , respectively the actual and simulated average travel times of the route segment r . In order to compare numerically the travel time measurements, a derived metric Δ_r , referring to the relative difference per route segment r between TT_{actual} and TT_{sim} , is defined in Equation 5.

$$\Delta_r = \frac{TT_{actual,r} - TT_{sim,r}}{TT_{actual,r}} \quad (5)$$

The second validation process aims to *investigate how well the simulation model reproduces reality based on certain reliability metrics*, by validating the simulation results against actual observations, using on-time performance measurements. In this work, the punctuality of both operated and simulated buses will be assessed according to departure delay at the bus stop at the control points of each of the bus lines. The validation process will be performed at three delay metrics, defined as follows:

- 388 • Aggregated delay metric: refers to on-time performance within the one-minute early and five-minutes-late
389 arrival range used by several urban transit agencies across the United States [114]. In this metric, the delay of
390 a bus can fall into three categories, in which the bus is considered 1) *on-time travel* if the delay is between *not*
391 *earlier than one minute and not later than five minutes* at the timing point; 2) in *Advance*, if the bus is *more than*
392 *one minute ahead of the scheduled time*; and 3) in *Significant delay*, if the bus is *late by more than five minutes*.
- 393 • Bus severity delay metric: in which, delay at control points is grouped into five categories; while the classifi-
394 cation of *in Advance* and in *Significant delay* remains similar to the first metric, *On-time travel* is divided into
395 three classes including, in addition to *On-time travel*, *minor delay* and *moderate delay*, in which each subclass
396 spans a two-minute range. The last two delay values have been defined according to the classification used in
397 the work proposed in [115].
- 398 • A third bus delay metric is defined in which stop delays are grouped in one-minute increments between *more*
399 *than five minutes early and more than seven minutes late*.

400 The two first delay metrics are given in more detail in Table 4.

Delay metric	Delay class	Range (seconds)
Aggregate Delay Metric	In advance	delay < -60
	On-time	delay ∈ [-60; 300]
	Significant delay	delay > 300
Bus Severity Delay Metric	In advance	delay < -60
	On-time	delay ∈ [-60; 60]
	Minor delay	delay ∈ [60; 180]
	Moderate delay	delay ∈ [180; 300]
	Significant delay	delay > 300

Table 4: A description of the aggregate and severity delay metrics

401 After defining the three sets of bus punctuality metrics, in the following, the defined on-time performance metrics
402 will be applied at the control points on the actual observations and the simulated results. In order to compare the
403 resulting punctuality shares, an offset and a deviation measure denoted δ and Δ , defined respectively, in equations 6
404 and 7, can be then be calculated between two sets of shares. Therefore, the defined measure Δ aims to quantify the
405 part of the share poorly located between the two sets.

$$\delta_c(S_i, S_j) = S_{i,c} - S_{j,c} \quad (6)$$

$$\Delta(S_i, S_j) = \sum_{c \in C} \delta_c(S_i, S_j) \cdot \alpha_{i,j,c} \quad (7)$$

406 Where:

- 407 • i and j are two datasets
- 408 • C : refers to delay values according to the punctuality metric used
- 409 • c : a delay value
- 410 • S_i : corresponds to the punctuality shares of the buses from the i dataset
- 411 • $S_{i,c}$: refers to share of c in S_i
- 412 • $\alpha_{i,j,c}$: 1 if ($S_{i,c} > S_{j,c}$) and 0 otherwise

413 A detailed illustration of the proposed two-phases validation framework is shown in Figure 9.

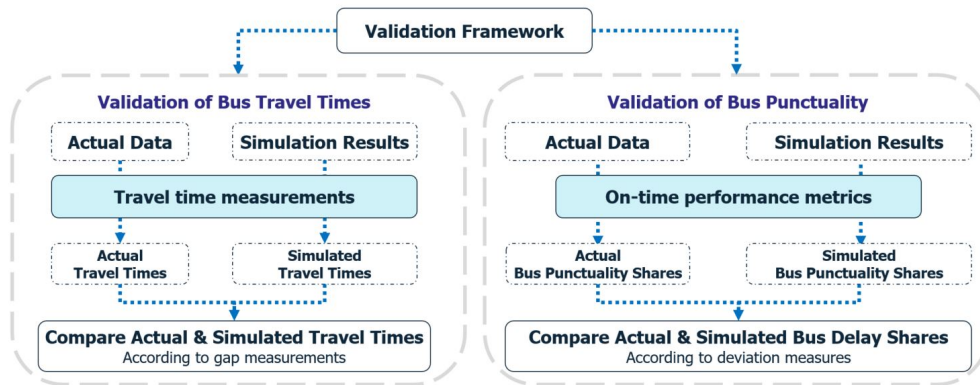


Figure 9: A detailed illustration of the proposed two-phase validation framework. The first phases consist of validating the bus travel times. Actual and simulated travel times can be extracted, from actual data and simulation results, respectively and then compared based on deviation measurements. The second phases consist of validating bus punctuality, a set of on-time performance metrics can be applied to estimate the actual and simulated punctuality share of the buses, then a comparison using deviation measures will take place for the assessment

414 In summary, the proposed ML-based microscopic simulation and validation frameworks can performed together as
 415 follows. First, ML models will be trained on actual data to predict travel times of bus route segments, then deployed
 416 to feed the public transport simulation model. Next, the simulation scenario will be implemented and then run.
 417 Subsequently, the results of the simulated scenario will then be assessed by the two-step validation framework. Finally,
 418 based on the adequacy of the simulated travel times and bus punctuality shares, the ML models selected for bus travel
 419 times and the simulation parameters such as recovery time and bus holding strategy can be updated. An overview of
 420 how the two proposed frameworks perform together is shown in Figure 10.

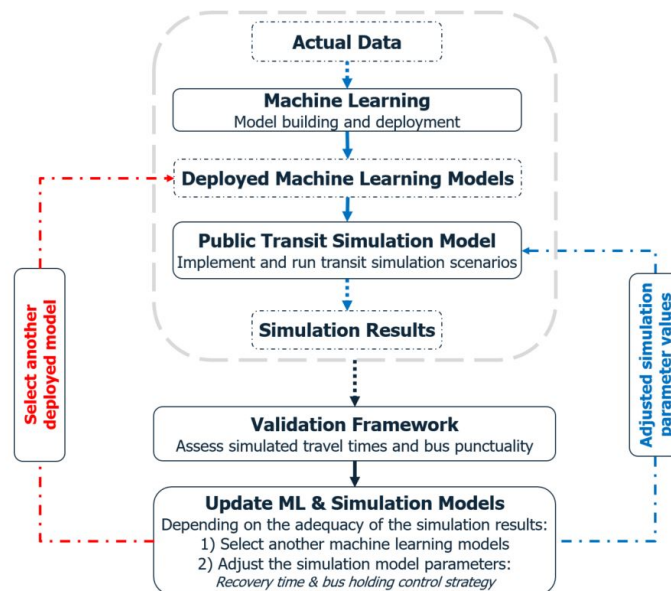


Figure 10: An overview of the two proposed frameworks: ML-based microscopic simulation framework and a validation framework

421 The remainder of the paper is structured as follows: in Section 4, a case study scenario consisting of three bus
 422 lines of a French city will be presented. In Section 5, the simulation results will be analyzed and validated based on a
 423 set of metrics. In Section 6, a discussion of the analyses presented, as well as some concluding remarks will be given
 424 in Section 7.

425 **4. Simulation Scenarios**

426 The simulation scenarios were developed for three bus lines referenced respectively by *A*, *B* and *C*. In this section,
 427 the chosen bus lines and the timing points will be described first. Next, the ML feature selection technique along with
 428 the trained ML models will be given. Finally, the simulation parameters including bus holding strategies as well as
 429 the implemented simulation scenarios will be detailed.

430 *4.1. Bus Lines: Description*

431 Line *A* buses run Monday to Saturday from approximately 4:45 a.m. (first departure) to 9:05 p.m. (last departure)
 432 and Sunday from approximately 5:35 a.m. to 8:05 p.m. Buses on this line serve 31 stops in one direction and 30
 433 stops in the other direction, for approximately 8 km of distance and 30 minutes travel time. This bus line is highly
 434 frequented line with on average one bus every 6 minutes during peak-hours and one bus every 8 minutes during off-
 435 peak hours, for a total of 234 trips on working days. Line *B* operates between approximately 5:42 a.m. and 8:26 p.m.
 436 only on working days. Like line *A*, the buses serve 31 and 30 stops respectively, over a distance of approximately 9
 437 and half km and 23 minutes of travel time. Line *B* operates with an average of one bus every 6 minutes and 8 minutes
 438 respectively, during peak and off-peak hours, for a total of 271 trips per working day. Line *C* buses operate all week
 439 from approximately 5:15 a.m. to 8:30 p.m., serving 10 stops in both directions. Compared to previous lines, line *C*
 440 is much shorter with approximately 3 km distance and a travel time of 10 minutes. This line runs on average with a
 441 bus every 6 minutes and 10 minutes respectively, during peak and off-peak hours, for a total of 198 trips per day. A
 442 summary description of the bus lines studied is presented in Table 5.

transit route ID	bus stops	travel distance (kms)	travel time (minutes)	headway on/off peaks
A_A	31	7.90	30	6/8 minutes
A_R	30	7.87	35	6/8 minutes
B_A	31	9.67	23	6/8 minutes
B_R	30	9.65	23	6/8 minutes
C_A	10	3.05	12	6/10 minutes
C_R	10	3.17	10	6/10 minutes

Table 5: A description of the bus lines studied including: transit line identifier, number of stops per direction, distance and time travel, in addition to the line’s service frequency during peak and off-peak hours. A_A and A_R respectively designate the route in one direction *Aller*, and the route in the opposite direction *Retour* of line *A*

443 The emphasis of this first work is placed on the simulation of a public transport scenario on working days (Monday
 444 to Friday), assumed to be more representative in terms of operating days and, consequently, operational trips for the
 445 public transport operator. The available operational data for the bus lines covered 73, 146 and 41 days for lines *A*, *B*
 446 and *C* respectively. A workday can be divided into peak and off-peak periods, as presented in Table 6.

time period	abbreviation	time interval
Morning off-peak hours	AM off-peak	[00h00 - 07h30]
Morning peak hours	AM peak	[07h30 - 09h30]
Inter-peak hours	–	[09h30 - 16h30]
Evening peak hours	PM peak	[16h30 - 19h30]
Evening off-peak hours	PM off-peak	[19h30 - 23h59]

Table 6: A description of the periods of the day, their abbreviations and associated time intervals

447 The shares of bus line trips, according to the time of day, are reported in Table 7.

448 *4.2. Bus Lines: Timing Points*

449 After selecting the bus lines for the simulation scenario, the next step is to set the timing points for each transit
 450 bus route. The control points are used to split a bus route into route segments in order to: firstly, build ML segment

bus line	morning off-peak hours	morning peak hours	inter-peak hours	evening peak hours	evening off-peak hours
A	12%	18%	41%	22%	6%
B	13%	13%	43%	20%	10%
C	9%	19%	46%	21%	6%

Table 7: Relative frequency of trips of the bus lines studied by period of the day

451 travel time models and, subsequently, deploy them for use in simulation; second, the control points will be used by the
 452 simulator for regularization purposes by applying bus holding strategies. Based on the actual data available, timing
 453 points correspond to bus stops, in which scheduled departure times are rounded to the nearest minute (hh:mm:00).
 454 The control points, along with the derived route segments and theoretical segment travel times, for the chosen bus
 455 lines are shown in Figure 11.

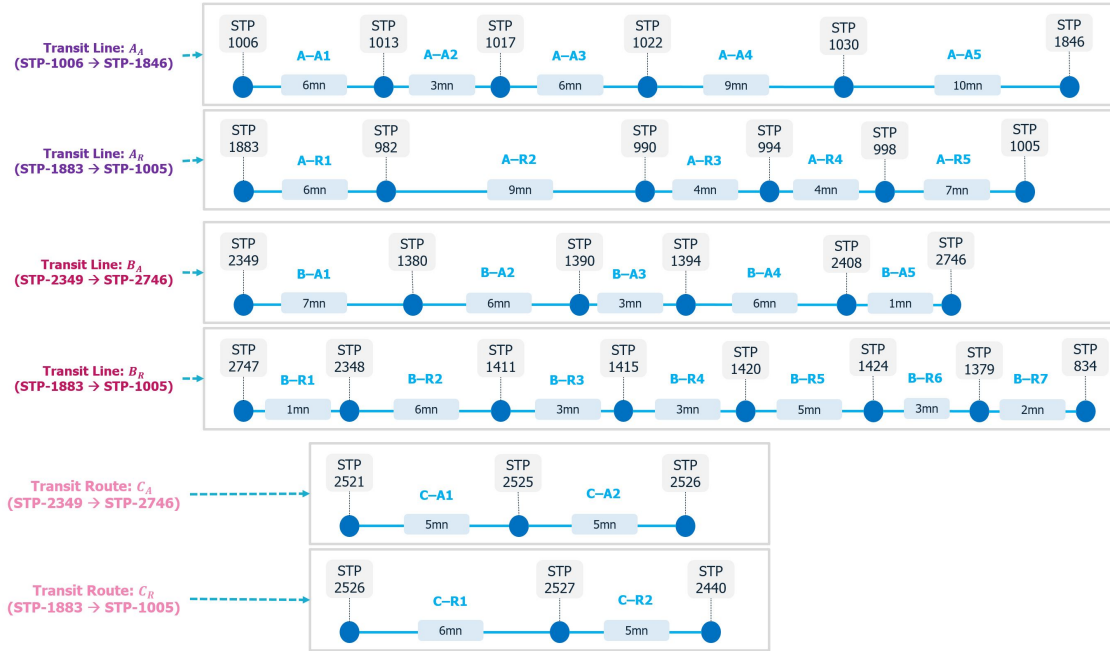


Figure 11: A detailed illustration of the route segments of the three chosen bus lines. We count, six and three control points per direction for lines A and C respectively; for bus line B, six timing points are defined in the *Aller* direction, and eight in the *Retour* direction. The route segment is referenced by: its transit line, its transit route direction (A/R) and its index in the transit route. For instance, A-A2 refers to the *second* segment of the *Aller* route of line A

456 4.3. ML: Algorithm and Feature Selection

457 *ML Algorithm.* In this study, trained ML models are used to feed public transport simulation model with travel times
 458 matching actual travel times, with more emphasis on accurately estimating travel time variability rather only a single
 459 travel time prediction. Thus, ML algorithms returning a single output value are not suitable. On the other hand, ML
 460 algorithms such as RT, kNNs and K-means, in which according to IF-THEN rules, the initial dataset can be split into
 461 samples, are suitable for dealing with travel time variability. The choice of regression tree as the ML algorithm for
 462 this study is motivated by the fact that RT is a rule-based approach, in which the built tree can be easily transformed
 463 into regression rules. Furthermore, to overcome the limitations of the standard RT, in which only one output value is
 464 returned, probabilistic distributions are used to fit travel times of the corresponding samples, in order to estimate the
 465 variability of the segment travel time.

466 *Feature Selection.* Feature importance refers to techniques that assign a score to input features based on their useful-
 467 ness in predicting a target variable. Feature importance scores can provide insight into the dataset, in which relative
 468 scores can highlight which features may be most relevant to the target, and the conversely, which features are least
 469 relevant. For tree-based models, mean decrease in impurity is a measure of how each variable contributes to the ho-
 470 mogeneity of nodes and leaves in the resulting trees. The higher the value of mean decrease Gini score, the higher
 471 the importance of the variable in the model. The resulting feature importance based on mean decrease in impurity,
 472 applied to three bus line datasets studied, is summarized in Table 8.

feature	line A	line B	line C
stop time	43%	49%	41%
stop delay	40%	32%	41%
day of the week	11%	11%	12%
season	6%	8%	6%

Table 8: Feature importance based on mean decrease in impurity

473 From Table 8, it can be seen that *stop time* and *stop delay*, are the most important features compared to the features
 474 *day of the week* and *season*. Moreover, an attempt to extend dataset, by including weather information was carried
 475 out. The mean decrease in Gini score applied to three studied lines is reported in Table 9.

feature	line A	line B	line C
stop time	36%	38%	35%
stop delay	21%	19%	23%
temperature	17%	17%	17%
wind	16%	16%	15%
day of the week	6%	6%	7%
season	2%	3%	3%
rain	1%	0%	1%

Table 9: Feature importance based on mean decrease in impurity, including the weather information

476 From Table 9, *temperature* and *wind* attributes seem to have similar scores, roughly equivalent to *stop delay*, and
 477 therefore can be considered as potential features for predicting bus travel time. On the other hand, *rain* attribute has
 478 almost an almost zero score, this low score may be due to the nominal nature of *rain* attribute grouped into four values
 479 (*none*, *slight*, *moderate* and *shower*). Therefore, ML models will be built with five sets of input features, including
 480 *stop time*, *stop delay* and weather attributes such as temperature, wind and rain. The first set includes only *stop time*,
 481 the second set includes *stop time* and *stop delay*, the third set only includes *stop time*, the fourth set includes only the
 482 weather attributes, while the fifth set includes *stop time* and the weather attributes. The observed actual travel time is
 483 chosen as model target.

484 From Figure 12, ML models built with *stop time* and *stop delay* as features seem to outperform. Moreover, ML
 485 models, in which *stop time* is a feature, appear to score higher compared to those built without *stop time* feature, for
 486 each of the three bus lines studied. Additionally, ML models with only *stop delay* feature score relatively higher than
 487 those built with only weather features. From a technical point of view, integrating weather information is challenging
 488 and requires the implementation of a validated weather generator into the public transport simulation model, adding
 489 an additional layer of complexity to the simulation framework. In the remainder of this paper, we refer to ML models
 490 built with *stop time* as a feature by f_1 , *stop time* and *stop delay* as features by f_2 , *stop delay* as feature by f_3 .

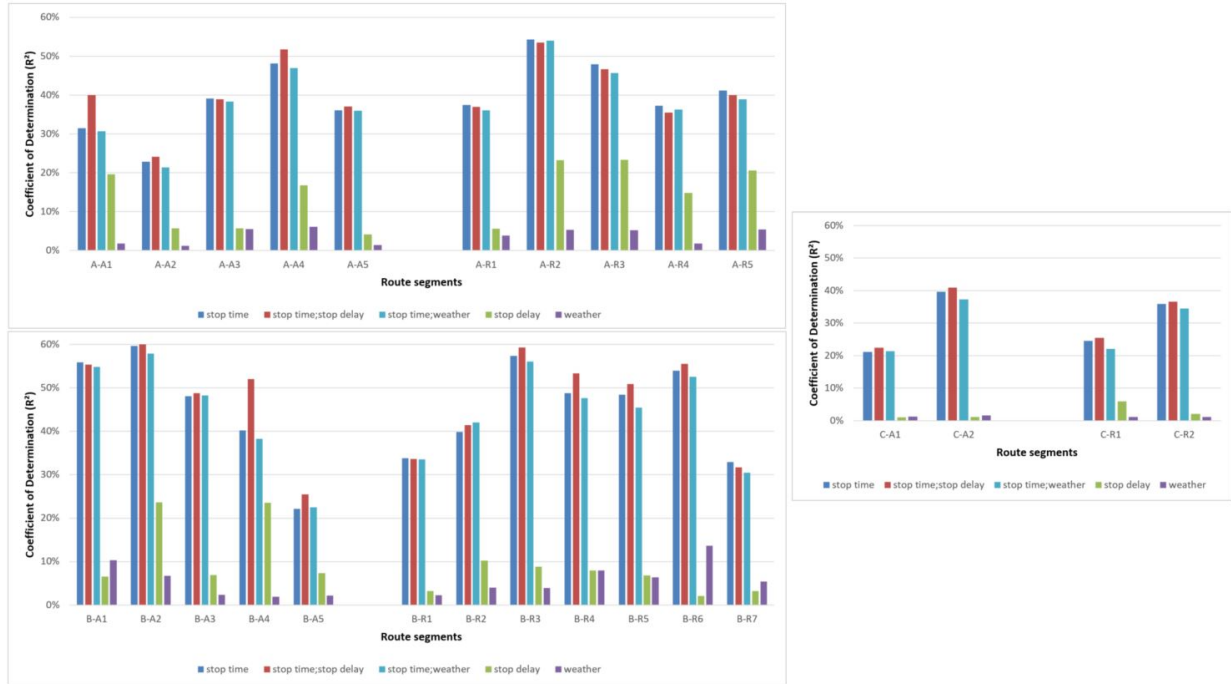


Figure 12: The coefficient of determination for ML models built with different feature sets, for the studied route segments

4.4. Trained Models: Hyperparameters Tuning

In order to tune the ML model parameters, the Grid Search approach, a traditional hyperparameter optimization method was selected. It simply iterates exhaustively through a manually specified subset of a learning algorithm's hyperparameter space [116]. To build the ML model, we opt for k-fold cross-validation (CV), a technique used in ML to evaluate the performance of a model on unseen data. In k-fold CV, the dataset is divided into k subsets (known as folds). A fold is used once in each iteration as test data, while the remaining folds are used as training data [117]. So the process is iterative until the entire dataset is evaluated. In this study, $k = 5$ is selected. In terms of parameters, regression trees are built according to the following parameter values: *tree maximum depth* varies between 2 and 14 levels, while *minimum samples per leaf* takes four values (25, 50, 75 and 100). In total, for each route segment, 156 regression trees are built and deployed. For lines A and C, respectively 780 and 312 trees were generated per route direction. For line B, 780 and 1092 were generated, respectively, for B_A and B_R .

From Figure 13 and Figure 14, *maximum tree depth* parameter seems to have an impact on the quality of the built ML models, with a steady increase in terms of score for models with *maximum tree depth* between 2 to 5 and the highest score recorded for models with *maximum tree depth* between 6 and 8 levels. In contrast, *minimum sample size* parameter appears to have less impact on model scores, with very small differences observed.

4.5. Simulation: Model and Parameters

In order to assess and validate the quality of the proposed ML-based microscopic simulation framework, a simulation model as well as a set of its parameters need to be defined. Firstly, the simulation scenarios will be carried out using the generic framework *Starling* proposed in [86] and briefly described in Section 3.1.3. Secondly, the number of runs to simulate is set to 1000 iterations. Usually, a high number of iterations is required in order to generate enough samples to validate the simulation results. In addition to validating the simulated travel times, this study aims to investigate the impacts bus recovery time strategies and bus holding strategies, on the punctuality of simulated buses.

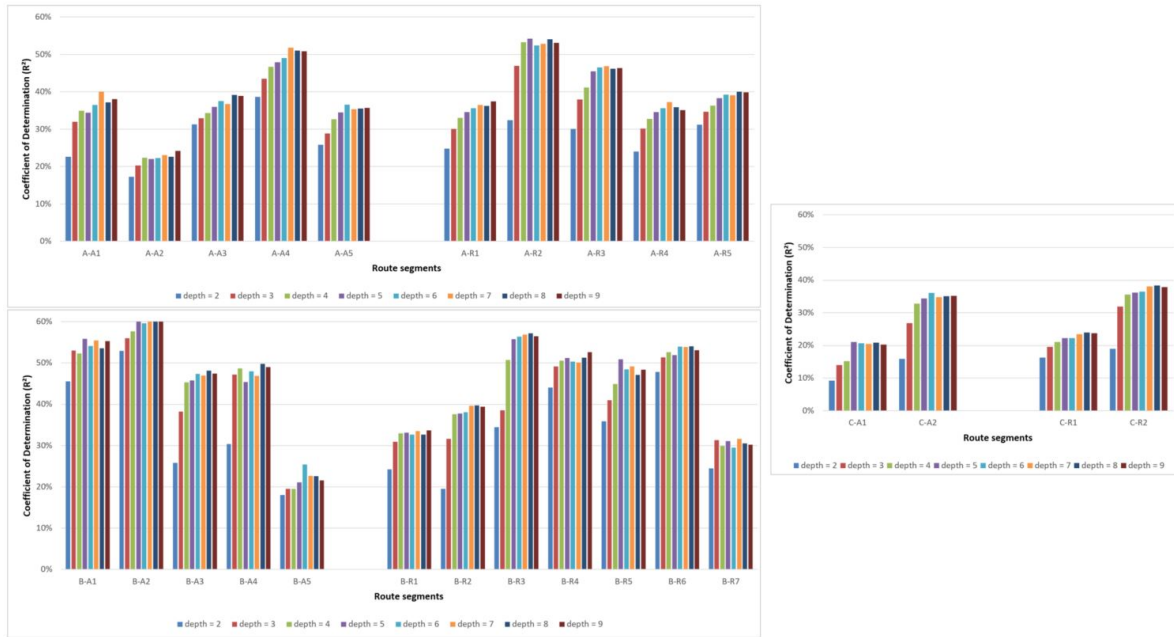


Figure 13: The coefficient of determination for ML models built with different values of maximum tree depth parameters, for the studied route segments

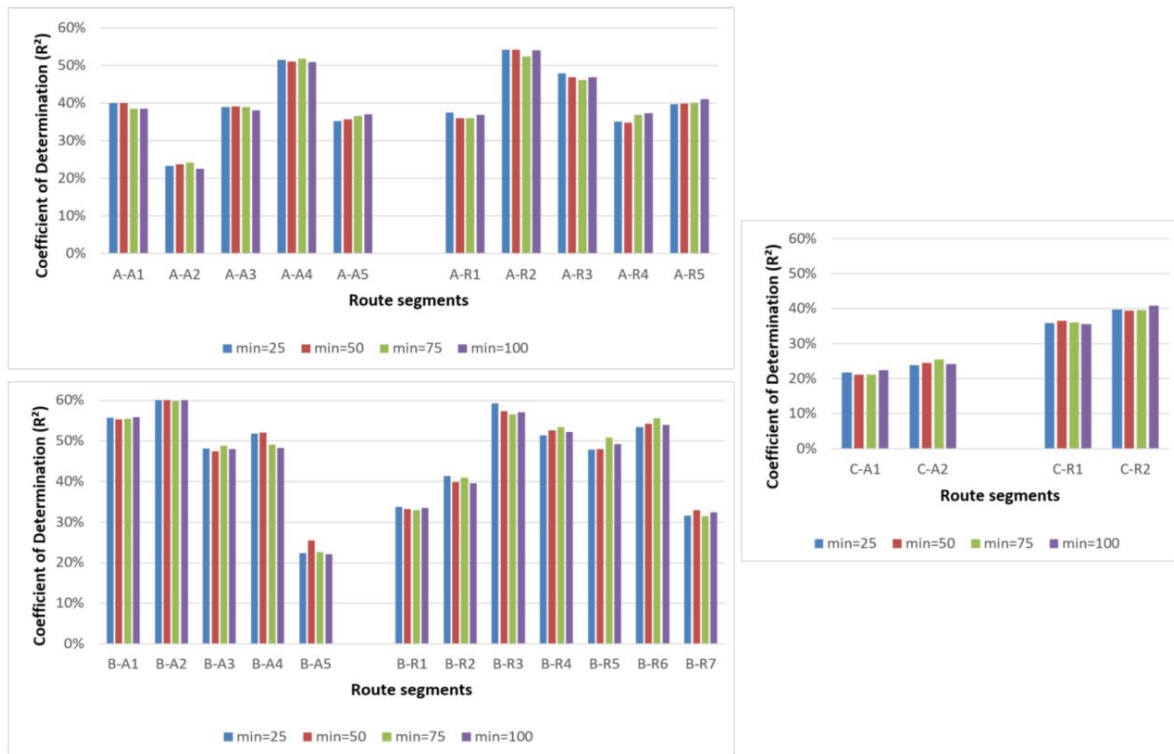


Figure 14: The coefficient of determination for ML models built with different minimum sample size parameters, for the studied route segments

514 *4.5.1. Bus Recovery Time Strategies*

515 Sufficient recovery time should be built into the schedule to ensure that the delays encountered during one trip
 516 do not carry over on subsequent trips. In this study, a first simulation scenario S_1 will be carried out with a fixed
 517 *minimum recovery time* of two minutes. Based on the actual departure delays presented in Figure 15, it appears that
 518 buses are likely departing a minute earlier and up to two minutes later than the scheduled time. Therefore, a second
 519 scenario S_2 , in which recovery times adjusted based on actual departure delays fitted to probabilistic distributions,
 520 will be conducted.

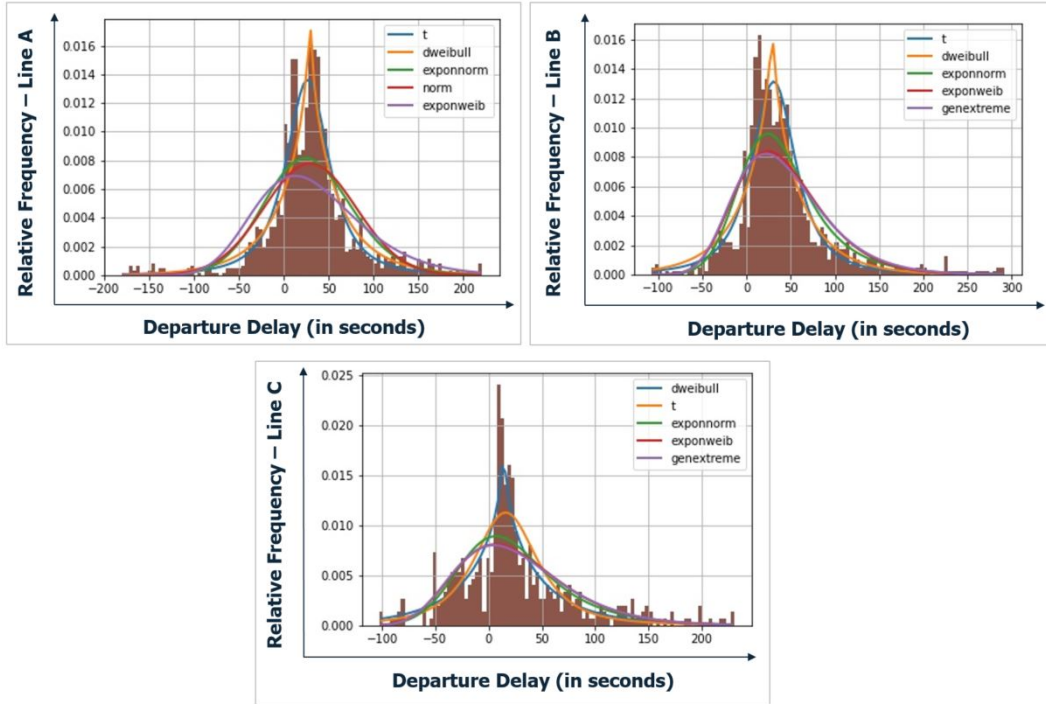


Figure 15: Illustration of the actual departure delay of buses on three lines studied, fitted to a certain number of probabilistic distributions

521 *4.5.2. Bus Holding Control Strategies*

522 In order to investigate the impacts of applying bus holding strategies on bus punctuality, two instances of scheduled-
 523 based holding strategy, respectively named SH_0 and SH_3 , an instance of headway-based holding strategy named HH ,
 524 and a base strategy without control NC , will be implemented. SH_0 and SH_3 , are two variants of the scheduled-based
 525 holding strategy defined in Equation 2, in which s is equals to zero and three minutes, respectively. In NC , the simula-
 526 tion will be conducted without applying a bus holding strategy. HH is a variant of the headway-based holding strategy
 527 defined in Equation 3, in which $\alpha = 0.8$. The strategies implemented are reported in Table 10. Note that NC can be
 528 thought of as an instance of scheduled-based holding strategy, with $s = +\infty$ (SH_∞).

Scenario ID	Scenario Short Name	Bus Holding Strategy	Strategy Parameters
NC	<i>no-control</i>	Scheduled-based	$s = +\infty$
SH_3	<i>3-Min</i>	Scheduled-based	$s = 3 \text{ minutes}$
SH_0	<i>scheduled-based</i>	Scheduled-based	$s = 0 \text{ minutes}$
HH	<i>headway-based</i>	Headway-based	$\alpha = 0.8$

Table 10: A summary of implemented bus holding strategies. s refers to the non negative slack size defined for a bus line, while α is a threshold ratio parameter

4.5.3. Bus Stop-Skipping Strategy

In this section, we aim to investigate to what extent skipping stops impacts bus punctuality. Stop-skipping control measure used to correct service inconsistencies, particularly to eliminate bus delay, is formulated in most studies as an optimization problem. With the interest of this study on bus punctuality, the decision whatever or not to skip a stop will depend primarily on how late is the bus at the control point. The stop-skipping criteria can be defined as follows:

$$D_{ijk} - D_{max} \geq 0 \quad (8)$$

Where:

- D_{ijk} : refers to the delay time for line i , on trip k , at bus stop j
- D_{max} : a non-negative maximum bus delay

If the stop-skipping criteria is met, the set of stops $j + 1, j + 2, \dots, c - 1$ between the current control stop j and the next control stop c will be skipped. Skipping is only permitted if the following conditions are met: 1) the stop is not a control stop and 2) the preceding bus did not skip the stop (to avoid passengers being denied twice). The departure time from a skipped stop is given in Equation 4. In this study, two scenarios SK_3 and SK_5 will be conducted, in which D_{max} is set to three and five minutes, from which a bus is moderately late and significantly late, respectively according to the second punctuality measure presented in Section 3.2.

4.6. Case Study: Challenges

In this study, we faced a number of challenges in setting up this case study, as follows. The first challenge was which bus stops to choose as control points? We first considered each bus stop as a control point. However this choice seemed to be so limited due to: 1) The segment travel times were too short, not exceeding one minute on average, which largely impacted the quality of the trained ML models which seemed to underfit due to low variance in observed travel times; 2) Scheduled stop times (from GTFS) are rounded to the nearest minute, where multiple bus stop times are set to the same rounded minute. As a result, inaccurate punctuality for the simulated buses. To overcome this issue, a set of stops predefined by the bus operator was used.

The second challenge was detecting and handling outliers. With actual data available prone to measurement errors, especially departure time (at first stop) and arrival time (at last stop), in which some stop times were not accurate. In order to clean the dataset, Inter Quartile Range (IQR), a detecting outliers technique was applied. On average, about 5% of observations were detected as outliers and therefore removed.

The third challenge was the choice of a probabilistic distribution to fit the travel time distribution, in order to estimate the travel time variability. To overcome this issue, the sample travel times - constituting according to ML model output, such as regression rules - were fitted to a number of probabilistic distributions, then compared based on the R^2 score and the law with the best-fit score as well as its parameters, will be selected.

5. Experiments and Results

In this section, the proposed two-step validation framework will be performed. The first phase of the proposed validation framework involves validating the simulated travel times while the second involves validating the simulated bus punctuality.

5.1. Validation of Travel Times

In this section, the validation of travel times process will be applied. First, scheduled and actual bus travel times will be compared side by side. Next, the resulting travel times of f_1, f_2 and f_3 will be compared over the entire day and subsequently over the morning and evening peak hours according to the relative difference metric defined in Equation 5. Furthermore, in order to evaluate the quality of ML models, a comparison with four probabilistic distributions, according to two goodness-of-fit measures MAE and $RMSE$ will be carried out.

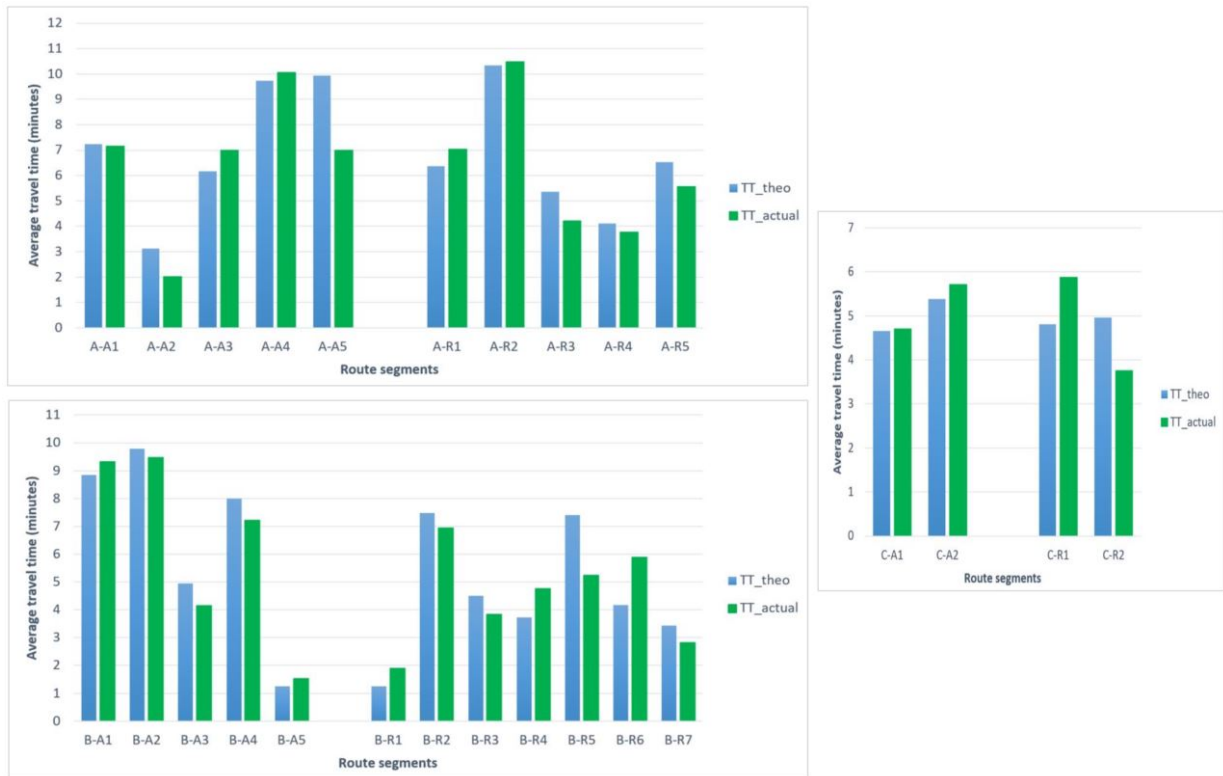


Figure 16: Theoretical and actual average travel times are presented side by side for each segment of lines studied. For line A, overall, larger differences can be observed on certain segments between theoretical and actual travel times. For line B, again, the simulated travel times seem to correspond well to the actual travel times. In contrast, differences can be observed between theoretical and actual travel times, with minor to moderate differences for segments in the *Aller* direction, and more significant for segments in the *Retour* direction. For line C, the simulated travel times appear to match actual travel times very well. Furthermore, the theoretical and actual travel times match well for segments in the *Aller* direction, while moderate differences can be observed for route segments in the *Retour* direction

5.1.1. Validation of Travel Times: Theo vs Actual

Below, Theoretical and actual average travel times are presented side by side for each segment of three lines studied are presented side by side in Figure 16.

At first glance, the average segment travel times for bus line A seem slightly longer for *Aller* than for *Retour*. Furthermore, theoretical travel times are on average a little overestimated compared to the actual travel times, particularly on A-A5 and A-R5, with differences ranging from moderate for segments A-A3, A-R1 and A-R5, to significant for segments A-A2, A-A5 and A-R3. For line B, one can notice scattered differences in travel times between *B-Aller* and *B-Retour*, appearing greater on latter route, especially on segments between B-R4 and B-R6. Overall, differences seem moderate on *Aller* and to a lesser extent on *Retour*. For line C, one can see that the differences are limited between the theoretical travel times and the actual travel times.

5.1.2. Validation of Travel Times: Actual vs Simulation

The resulting travel times of f_1 , f_2 and f_3 as well as the relative differences metric Δ (see Equation 5), for bus lines A, B and C are reported, respectively, in Table 11, 12 and 13.

For line A, the simulated *Aller* travel times estimated using f_1 models, are on average more precise than those estimated using f_2 and f_3 . Furthermore, the simulated *Retour* travel times appear to be more accurate than those of *Aller*, with f_1 and f_2 corresponding well to the actual travel times. Overall, f_1 and f_2 seem to outperform f_3 .

segment ID	segment stops	TT_{actual}	Δ_{f_1}	Δ_{f_2}	Δ_{f_3}
A-A1	STP-1006 → STP-1013	429	-5%	-8%	-6%
A-A2	STP-1013 → STP-1017	121	1%	8%	2%
A-A3	STP-1017 → STP-1022	420	1%	0%	3%
A-A4	STP-1022 → STP-1030	603	2%	3%	5%
A-A5	STP-1030 → STP-1846	419	7%	12%	16%
-	-	-	-	-	-
A-R1	STP-1883 → STP-982	423	0%	0%	-1%
A-R2	STP-982 → STP-990	629	3%	1%	1%
A-R3	STP-990 → STP-994	253	3%	3%	9%
A-R4	STP-994 → STP-998	226	2%	1%	4%
A-R5	STP-998 → STP-1005	333	3%	4%	9%

Table 11: Reported actual average travel times TT_{actual} (in seconds) as well as the relative differences Δ_{f_1} , Δ_{f_2} and Δ_{f_3} , respectively, from f_1 , f_2 and f_3 , for each route segment of bus line A

segment ID	segment stops	TT_{actual}	Δ_{f_1}	Δ_{f_2}	Δ_{f_3}
B-A1	STP-2349 → STP-1380	560	2%	2%	0%
B-A2	STP-1380 → STP-1390	568	3%	2%	3%
B-A3	STP-1390 → STP-1394	249	2%	2%	19%
B-A4	STP-1394 → STP-2408	433	5%	5%	12%
B-A5	STP-2408 → STP-2746	92	-13%	-14%	-13%
-	-	-	-	-	-
B-R1	STP-2747 → STP-2348	114	-4%	-2%	-3%
B-R2	STP-2348 → STP-1411	417	1%	0%	1%
B-R3	STP-1411 → STP-1415	231	3%	3%	4%
B-R4	STP-1415 → STP-1420	286	4%	2%	3%
B-R5	STP-1420 → STP-1424	315	4%	6%	16%
B-R6	STP-1424 → STP-1379	353	4%	5%	2%
B-R7	STP-1379 → STP-834	169	3%	3%	3%

Table 12: Reported actual average travel times TT_{actual} as well as the relative differences Δ_{f_1} , Δ_{f_2} and Δ_{f_3} , for bus line B

586 For line B, the simulated *Aller* travel times of f_1 and f_2 seem to be highly similar. Regarding the travel times of
587 f_3 , more significant deviations can be observed compared to the actual segment travel times, in particular for B-A3
588 and B-A4. Despite the large underestimate of travel times of B-A5, it only amount to around 12 seconds. Concerning
589 the *Retour* route, f_1 and f_2 , provide similar travel times, both seeming to correspond well to actual travel times.
590 Furthermore, f_3 seems to match the actual data, expect for B-R5, in which travel times are greatly overestimated. In
591 sum, the three sets of ML models predict very similar travel times, including f_1 and f_2 , and to a lesser extent f_3 .

segment ID	segment stops	TT_{actual}	Δ_{f_1}	Δ_{f_2}	Δ_{f_3}
C-A1	STP-2521 → STP-2525	282	-3%	-2%	-2%
C-A2	STP-2525 → STP-2526	343	-3%	-3%	-2%
-	-	-	-	-	-
C-R1	STP-2526 → STP-2527	353	-2%	-2%	-1%
C-R2	STP-2527 → STP-2440	226	-1%	-1%	0%

Table 13: Reported actual average travel times TT_{actual} as well as the relative differences Δ_{f_1} , Δ_{f_2} and Δ_{f_3} , for bus line C

592 With regard to Δ in Table 13, the differences seem negligible, averaging a few seconds per route segment.

Thus, from the results illustrated on the three bus lines studied (A , B and C), the simulated travel times provided by ML models are highly resembled and statistically similar to the actual travel times.

To test the reliability the ML models trained for bus travel time prediction, simulated and actual travel times will be compared over morning and evening peak hours. For line A , the results show that the actual segment travel times of *Aller* route are on average relatively longer during morning rush hours than during evening hours, and vice versa for the *Retour* segments. For line B , similar actual travel times are observed at both peak hours, while an increase in segment travel times is observed during the evening hours compared to the morning peak hours, for line C .

In sum, ML models in which features include stop time (f_1 and f_2) perform better than those built by only considering bus delay at the stop. Moreover, ML models with only the stop time as a feature (f_1) relatively match the best actual travel times. In the remainder of this work, the ML models are referred to the f_1 models. An in-depth analysis of actual travel times together along with relative differences from three set of models (f_1 , f_2 and f_3) over peak hours, for bus lines A , B and C can be found in [Appendix B](#).

5.1.3. Validation of Travel Times: ML vs Probabilistic Distributions

To assess the accuracy of the ML models for bus travel time prediction, the simulated travel times provided by ML will be compared to four probabilistic distributions, including: *normal*, *lognormal*, *gamma* and *generalized extreme value* (GEV) distributions, with a focus on morning and evening peak hours. The actual travel times as well as the relative differences resulting from ML and probabilistic distributions over peak hours, for bus lines A , B and C are reported, respectively, in [Table 14](#), [15](#) and [16](#).

segment ID	Morning peak hours						Evening peak hours					
	TT_{actual}	Δ_{ML}	Δ_{norm}	$\Delta_{lognorm}$	Δ_{gamma}	Δ_{GEV}	TT_{actual}	Δ_{ML}	Δ_{norm}	$\Delta_{lognorm}$	Δ_{gamma}	Δ_{GEV}
A-A1	513	-10%	-16%	-16%	-53%	-16%	422	-3%	-3%	-3%	-3%	-4%
A-A2	132	1%	-1%	-3%	33%	-2%	125	1%	1%	0%	1%	0%
A-A3	481	-1%	-10%	-9%	-5%	-8%	433	-5%	-1%	-1%	-1%	-1%
A-A4	653	-1%	-4%	-4%	-3%	-4%	619	-1%	0%	1%	0%	0%
A-A5	500	10%	16%	17%	20%	12%	433	1%	6%	7%	-18%	1%
-	-	-	-	-	-	-	-	-	-	-	-	-
A-R1	418	-3%	-7%	-6%	-7%	-7%	472	1%	-3%	-3%	-3%	-3%
A-R2	628	-2%	-6%	-6%	-5%	-5%	706	4%	2%	2%	2%	2%
A-R3	248	1%	-3%	-3%	-3%	-2%	271	3%	8%	10%	-29%	14%
A-R4	231	-1%	-4%	-4%	-4%	-5%	253	0%	-1%	-1%	0%	-3%
A-R5	349	-5%	-4%	-4%	-3%	395%	349	2%	7%	8%	12%	13%

Table 14: Reported actual average travel times TT_{actual} as well as relative differences Δ_{ML} , Δ_{norm} , $\Delta_{lognorm}$, Δ_{gamma} and Δ_{GEV} , respectively, from *ML*, *norm*, *lognorm* and *GEV* distributions, for each route segment of bus line A , over morning and evening peak-hours

From [Table 14](#), one can see that the ML travel times are on average more precise compared to the travel times generated by the different probabilistic distributions, during morning and evening rush hours. Moreover, the differences between ML and distributions are more significant during morning peak hours than during evening hours.

For line B , a similar trend is observed with more precise ML travel times compared to those estimated by the probabilistic distributions, in the morning and to a lesser extent during the evening rush hours. Furthermore, while the *normal*, *lognormal*, and *gamma* distributions seem to correspond to the actual travel times to a lesser extent, *GEV* distribution appears to strongly overestimate actual travel times.

From [Table 16](#), one can see that on average the ML travel times of line C seem to be more precise than the statistical distribution travel times, both during morning and evening peak hours. Additionally, the resulting travel times of the four probabilistic distribution are highly resembled, with on average, generalized extreme value distribution appearing to provide more accurate travel times than the normal, lognormal, and gamma distributions.

segment ID	Morning peak hours						Evening peak hours					
	TT_{actual}	Δ_{ML}	Δ_{norm}	$\Delta_{lognorm}$	Δ_{gamma}	Δ_{GEV}	TT_{actual}	Δ_{ML}	Δ_{norm}	$\Delta_{lognorm}$	Δ_{gamma}	Δ_{GEV}
B-A1	573	-2%	-7%	-7%	-6%	2%	625	-2%	-4%	-5%	-4%	38%
B-A2	617	0%	-5%	-5%	-5%	74%	635	-1%	-2%	-2%	-2%	118%
B-A3	305	-9%	-16%	-16%	-56%	738%	313	-1%	-7%	-8%	-7%	123%
B-A4	548	2%	-3%	-4%	0%	44%	450	1%	2%	2%	2%	183%
B-A5	105	-14%	-15%	-16%	-16%	-19%	103	-27%	-27%	-26%	-27%	-29%
-	-	-	-	-	-	-	-	-	-	-	-	-
B-R1	99	-5%	-5%	-6%	-5%	0%	131	-4%	-4%	-3%	-3%	-3%
B-R2	413	-1%	-4%	-4%	-4%	-1%	490	-2%	-6%	-6%	-6%	-6%
B-R3	284	1%	-7%	-7%	-7%	58%	260	1%	-4%	-4%	-4%	-6%
B-R4	310	0%	-5%	-5%	-6%	40%	312	0%	-2%	-2%	-2%	-4%
B-R5	342	8%	8%	8%	8%	844%	352	2%	2%	2%	2%	104%
B-R6	375	-2%	-8%	-7%	-7%	2%	384	-1%	-3%	-4%	-4%	-3%
B-R7	188	-3%	-7%	-8%	-7%	21%	179	-1%	-4%	-4%	-4%	6%

Table 15: Reported actual average travel times as well as relative differences resulting from ML and probabilistic distributions, for each route segment of bus line B , over morning and evening peak-hours

segment ID	Morning peak hours						Evening peak hours					
	TT_{actual}	Δ_{ML}	Δ_{norm}	$\Delta_{lognorm}$	Δ_{gamma}	Δ_{GEV}	TT_{actual}	Δ_{ML}	Δ_{norm}	$\Delta_{lognorm}$	Δ_{gamma}	Δ_{GEV}
C-A1	284	-3%	-6%	-6%	-6%	-5%	300	-6%	-9%	-9%	-9%	-9%
C-A2	342	-3%	-9%	-10%	-9%	-8%	395	-5%	-14%	-14%	-15%	-8%
-	-	-	-	-	-	-	-	-	-	-	-	-
C-R1	373	-1%	-10%	-10%	-10%	-10%	396	-4%	-10%	-10%	-10%	-3%
C-R2	231	-3%	-6%	-6%	-6%	-6%	233	0%	-1%	-1%	-1%	-1%

Table 16: Reported actual average travel times as well as relative differences resulting from ML and probabilistic distributions, for each route segment of bus line C , over morning and evening peak-hours

5.1.4. Validation of Travel Times: Goodness of fit

In order to thoroughly evaluate the quality of the simulated travel times provided by ML and the probabilistic distributions, mean absolute error (MAE) and root mean square error (RMSE), two goodness-of-fit measures (GOF), will be estimated at morning and evening peak hours, for each route.

In essence, segment travel times will be grouped into half-minute slot. MAE and RMSE measures, are respectively formulated in Equations 9 and 10. The estimated values of MAE and RMSE, applied to the three bus lines, are reported in Tables 17 and 18, respectively.

$$MAE = \frac{1}{|S|} \cdot \sum_{s \in S} \sum_{t \in T} |P_{d,s}^t - P_{a,s}^t| \quad (9)$$

$$RMSE = \sqrt{\frac{1}{|S|} \cdot \sum_{s \in S} \sum_{t \in T} (P_{d,s}^t - P_{a,s}^t)^2} \quad (10)$$

Where:

- s : a route segment
- S : a set of route segments
- t : a travel time increment.

- T : a set of travel time increments
- $P_{a,s}^t$: refers to the actual relative frequency for segment s with $TT \in t$
- $P_{d,s}^t$: refers to relative frequency of a distribution d for segment s with $TT \in t$

Route ID	Morning peak hours					Evening peak hours				
	ML	norm	lognorm	gamma	GEV	ML	norm	lognorm	gamma	GEV
A-Aller	19%	23%	24%	54%	21%	11%	15%	13%	17%	14%
A-Retour	9%	10%	9%	9%	11%	13%	19%	15%	27%	20%
-	-	-	-	-	-	-	-	-	-	-
B-Aller	13%	17%	15%	29%	28%	10%	12%	11%	11%	42%
B-Retour	10%	14%	13%	13%	27%	9%	10%	8%	8%	10%
-	-	-	-	-	-	-	-	-	-	-
C-Aller	5%	15%	17%	17%	14%	6%	15%	20%	20%	19%
C-Retour	5%	20%	21%	22%	23%	5%	8%	10%	9%	9%

Table 17: Mean Absolute Error (MAE) between actual and simulated travel times, by route, during morning and evening peak-hours

At first glance at Table 17, ML seems to have the lowest MAE on most bus routes. Additionally, the differences between ML and other distributions tend to be smaller, to some extent, during evening peak hours, with the exception of line C routes, on which ML appears to significantly outperform.

Route ID	Morning peak hours					Evening peak hours				
	ML	norm	lognorm	gamma	GEV	ML	norm	lognorm	gamma	GEV
A-Aller	11%	15%	17%	40%	14%	7%	10%	9%	13%	9%
A-Retour	8%	8%	7%	7%	8%	9%	13%	12%	23%	14%
-	-	-	-	-	-	-	-	-	-	-
B-Aller	9%	11%	10%	25%	18%	7%	10%	10%	9%	36%
B-Retour	8%	11%	11%	11%	24%	8%	8%	7%	6%	8%
-	-	-	-	-	-	-	-	-	-	-
C-Aller	4%	12%	13%	13%	11%	4%	10%	14%	14%	13%
C-Retour	4%	15%	17%	17%	18%	4%	6%	7%	7%	6%

Table 18: Root Mean Square Error (RMSE) between actual and simulated travel times, by route, during morning and evening peak-hours

From Table 18, ML models show the lowest RMSE error on most line routes. Similar to MAE, ML appears to outperform probabilistic distributions during morning peak hours and, to a lesser extent, during evening peak hours.

Furthermore, with exception to A-Retour and to a lesser extent C-Aller routes, the resulting travel times during the evening rush hours appear more accurate than those during morning hours.

5.2. Validation of Simulation Results: Impacts of Recovery Time Strategies on Bus Punctuality

In order to investigate the impacts of bus recovery time on simulated bus punctuality, two simulation scenarios S_1 using a fixed *minimum recovery time* of two minutes and S_2 using an adjusted *recovery time*, will be conducted. In the following, aggregated delay and bus delay severity, two of the on-time performance metrics presented in Section 3.2, will be applied at the control points on the actual observations and the two simulated scenarios (S_1 and S_2), for each of the three bus lines (A, B and C). The results obtained on each dataset will then be statistically analyzed and compared side by side.

5.2.1. Impacts of Recovery Time: Aggregated Delay Metric

Figure 17 provided first insights at how bus punctuality is distributed. It illustrates the resulting bus status share according to the first punctuality metric, applied to the 9 datasets. Further on, Table 19 illustrates in more detail the differences between the shares resulting from simulation scenarios and the actual shares.

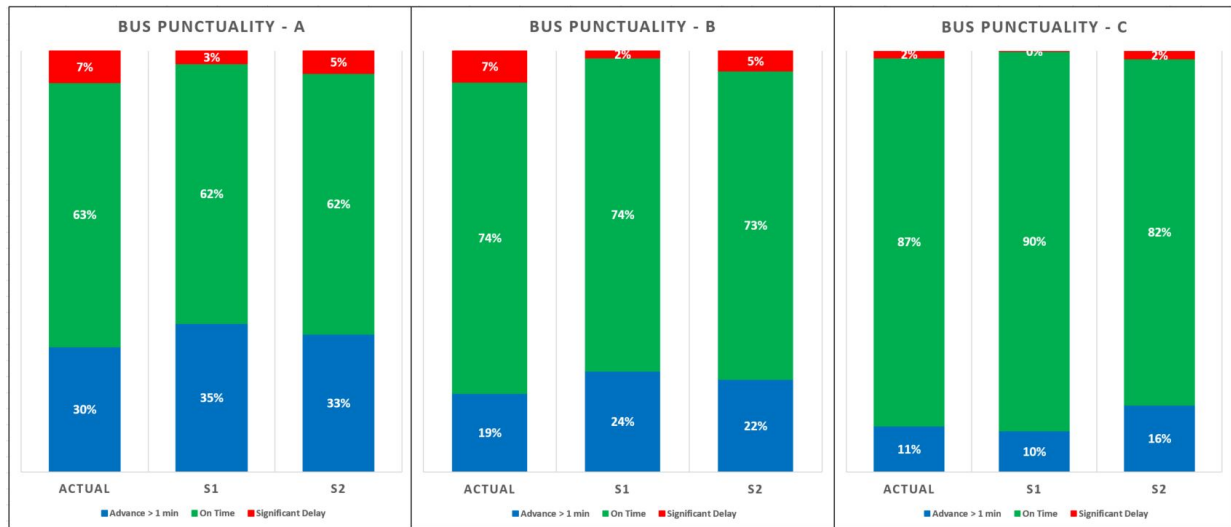


Figure 17: Bus punctuality shares for the bus lines studied, based on the first on-time measurement. For lines A and B, a strong similarity of shares is observed between S_1 and S_2 , while compared to actual shares, the differences are minors, with S_2 appearing to slightly outperform S_1 . For line C, the resulting shares of S_1 and S_2 are different, compared to the referenced shares, S_2 seems to slightly outperform S_1

	S_1 (A)	S_2 (A)	S_1 (B)	S_2 (B)	S_1 (C)	S_2 (C)
Advance > 1 min	+5%	+3%	+5%	+3%	-1%	+5%
On-time travel	-1%	-1%	0%	-1%	+3%	-5%
Significant delay	-4%	-2%	-5%	-2%	-2	0%
$\Delta(S, actual)$	5%	3%	5%	3%	3%	5%

Table 19: Illustration of the differences and gaps between the resulting simulation shares and the actual shares according to the first punctuality metric, for lines A, B and C

656 At first glance, one can see some differences in the shares between the different lines, line C seems to be the most
657 regular on-time bus line among the lines studied, followed by line B then line A. Based on *in advance* shares, line A
658 buses are much further ahead than compared to lines B and C. Furthermore, line C, with its shortest bus routes, has the
659 least late buses, in which the share of *Significant delay* is negligible, while lines A and B show statistically an almost
660 identical ratio for significantly delayed buses. The analysis process will proceed as follows: the bus status shares of
661 the actual observations will be analyzed first, followed by those from the two simulated scenarios. Secondly, the share
662 of the actual observations will be compared to the shares resulting from the simulation.

663
664 For bus line A, the results show that about two thirds of the actual buses are on time and about one third are early
665 at the timing points. On the other hand, the results of the simulated scenarios indicate a very similar trend for S_1
666 and S_2 , with an identical share for *on-time travel* (with 62%) and minor differences ($\pm 2\%$) between *in advance* and
667 *significantly delayed* buses. Compared to the actual data, minor differences were observed with 5% for S_1 and 3% for
668 S_2 respectively.

669
670 For bus line B, the shares indicate that approximately three thirds and one fifth of actual buses are on-time and
671 early, respectively, while significantly late buses represent less than a tenth of the total buses. Furthermore, the results
672 also show that S_1 and S_2 follow a similar trend, with minor differences (3%) for the *in advance* and *significantly late*
673 buses. As for bus line A, the differences between the actual and simulated shares are minor and almost identical, with
674 a difference of approximately 5% and 3%, respectively, for S_1 and S_2 .

675

676 For bus line *C*, it can be observed that almost all actual buses travel *on-time*, while that a tenth are *ahead*. On the
 677 other hand, the share of *significantly delay* is very negligible. Unlike the two previous bus lines (*A* and *B*), the results
 678 of the two simulated scenarios (S_1 and S_2) show different trends. S_2 overestimates the share of buses *in advance*
 679 (+5%) compared to actual observations and to S_1 , while on the other hand, S_1 completely underestimates the buses
 680 with a *significant delay*, estimated to be almost zero.

681
 682 In total, the results showed significant similarities between S_1 and S_2 , with minor differences on bus lines *A* and
 683 *B*, and to a lesser extent on line *C* where the differences appear more significant. Compared to the actual bus delays,
 684 the delays shares resulting from the simulation, using the adjusted recovery time in S_2 scenario, are more accurate
 685 than those in S_1 , in which a fixed minimum recovery time of two minutes was implemented. The results obtained in
 686 S_2 slightly overestimate the ratio of buses ahead of schedule. In short, the resulting simulation distributions for the
 687 three lines studied are with a good fit to the actual reference distributions according to this studied metric.

688 5.2.2. *Impacts of Recovery Time: Bus Delay Severity Metric*

689 In order to investigate in more details the delay shares obtained previously according to the first metric used, the
 690 bus severity delay metric defines buses *ahead of schedule* and *significantly delayed* as in the first metric, in contrast,
 691 distinguishing the *on-time travel* into three sub-classes with a 2 minute horizon, as follows: on-time travel, minor
 692 delay and moderate delay, respectively, if the delay is between -1 and +1, +1 and +3 and +3 and +5 minutes.

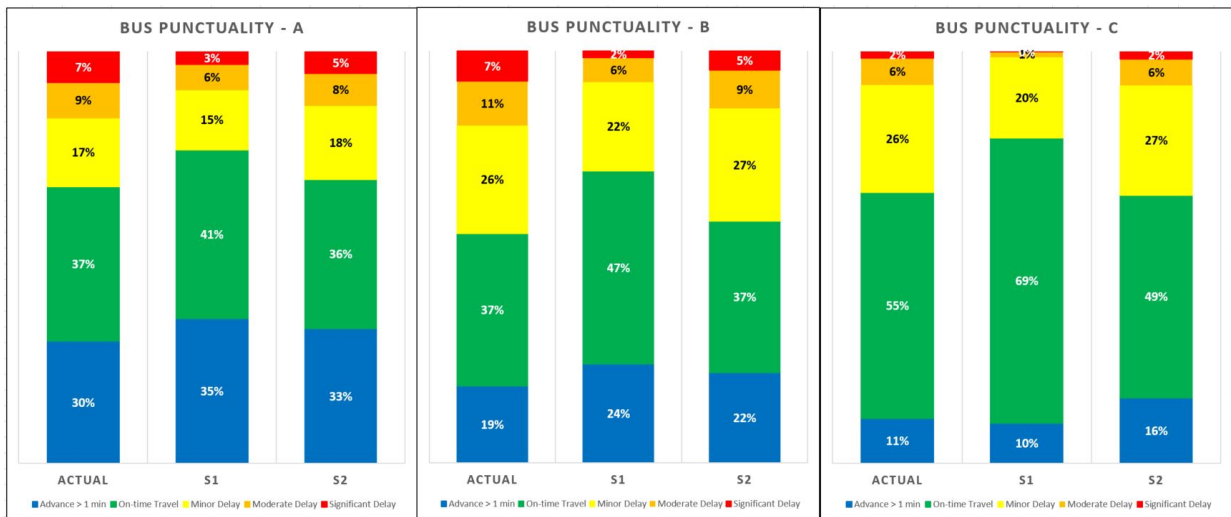


Figure 18: Bus punctuality shares for lines *A*, *B* and *C*, based on the second on-time metric with five levels of delay: *in advance*, *on-time travel*, *minor*, *moderate* and *significantly delayed*. Concerning the resulting shares, the differences between S_1 and S_2 are much more significant, particularly for the *on-time travel* share. The S_2 shares appear to match the actual data well for lines *A* and *B*, and to a lesser extent for line *C*. In contrast, S_1 seems to considerably overestimate the share of *on-time travel* and underestimate all the remaining shares

693 From Figure 18 and Table 20, one can observe a strong resemblance between the shares of bus punctuality both
 694 for the actual observations and S_2 , and that for the three studied bus lines. An exception to note for bus line *C*, part
 695 of the share (around 5%) of *in advance* and *on-time travel* buses seems to be poorly located. Furthermore, the share of
 696 minor, moderate and significant delays is almost identical between S_2 and actual observations. In sum, S_2 is slightly
 697 overestimates the shares of early buses, well adjusted to the shares of *minor*, *moderate* and *significant* delays, with an
 698 approximate deviation of 4% to 6% for the lines studied. In contrast, the S_1 shares show a significant overestimation
 699 of *in advance* and *on-time*, and therefore, an underestimation of *minor*, *moderate* and *significant* shares, with an
 700 approximate deviation of 9% for line *A*, 15% for line *B* and 14% for line *C*.

701
 702 In short, the results of the second on-time performance metric show that applying an adjusted recovery time
 703 strategy improved the quality of results compared to the fixed recovery time strategy.

	S_1 (A)	S_2 (A)	S_1 (B)	S_2 (B)	S_1 (C)	S_2 (C)
Advance > 1 min	+5%	+3%	+5%	+3%	-1%	+5%
On-time travel	+4%	-1%	+10%	0%	+14%	-6%
Minor delay	-2%	+1%	-4%	+1%	-6%	+1%
Moderate delay	-3%	-1%	-5%	-2%	-5%	0%
Significant delay	-4%	-2%	-5%	-2%	-2%	0%
$\delta(S, actual)$	9%	4%	15%	4%	14%	6%

Table 20: Illustration of the differences and gaps between the resulting simulation shares and the actual shares according to the second on-time metric, for lines studied

5.3. Validation of Simulation Results: Impacts of Bus Control Strategies on Bus Punctuality

In this section, the emphasis is placed on investigating the impacts of different bus control strategies on bus punctuality. The strategies implemented are reported in Table 10. In the following, the resulting shares will be analyzed and compared to the actual shares, according to a number of one-time performance metrics.

5.3.1. Impacts of Bus Holding: Aggregated Delay Metric

The bus punctuality shares based on aggregate delay metric for bus lines A, B and C are shown in Figure 19. The differences between the resulting simulation shares and the actual shares according to the first punctuality metric are reported in Table 21.

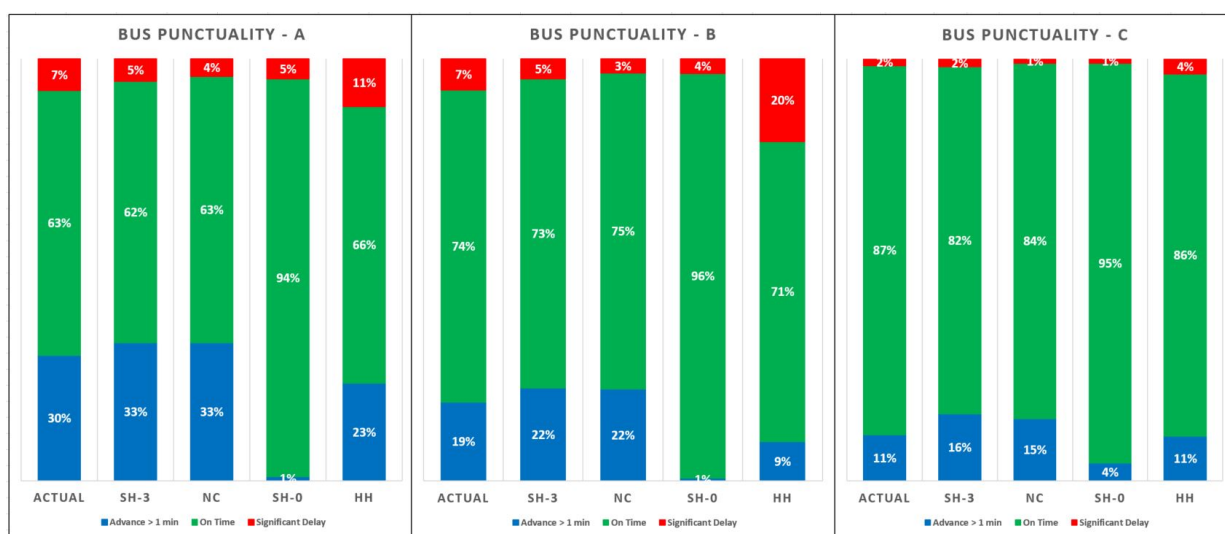


Figure 19: Impacts of bus holding strategies - Bus punctuality shares for the bus lines studied, based on the first on-time measurement

$\Delta(S, actual)$	no-control NC	scheduled-based SH_3	scheduled-based SH_0	headway-based HH
Line A	3%	3%	31%	7%
Line B	4%	3%	22%	12%
Line C	4%	5%	7%	2%

Table 21: Impacts of bus holding - Illustration of the differences between the resulting simulation shares and the actual shares according to the first punctuality metric, for bus lines A, B and C

713 For bus line *A*, results obtained from actual observations show that about two thirds of buses on time and about
 714 one third are early at the timing points. On the other hand, the results indicate a very similar trend for scenarios *NC*
 715 and *SH₃*, with an identical share for *on-time travel* and minor differences ($\pm 3\%$) between *in advance* and *significantly*
 716 *delayed* buses. The results show that *SH₀* appear to completely underestimate the share of ahead bus, and strongly
 717 overestimate the share of on-time buses, the share of *significantly delayed* buses, on the other hand, is slightly underes-
 718 timated. Furthermore, the *HH* results indicate an underestimation of the share of early buses, and an underestimation
 719 to a lesser extent of the share of on-time and significantly late buses. In short, for line *A*, the shares of *NC* and *SH₃*
 720 appear more similar to the actual shares compared to the shares of *HH*.

721
 722 For bus line *B*, the shares shown indicate that approximately three thirds and one fifth of actual buses are *on-*
 723 *time travel* and *in advance*, respectively, while the buses with a significant delay represent less than a tenth of the total
 724 buses. Furthermore, the results show that *NC* and *SH₃* follow a similar trend, with minor differences for the *ahead* and
 725 *significantly late* buses. For *SH₀*, a trend similar to that of line *A* is observed, including a complete underestimation
 726 of the share of early buses. Moreover, the *HH* results indicate an underestimation and overestimation of the share
 727 of *in advance* and *significantly delayed* buses, respectively. Overall, for line *B*, the differences between the real and
 728 simulated shares are minor for *NC* and *SH₃* and, to a greater extent, for *HH* and *SH₀*.

729
 730 For bus line *C*, one can observe that almost all actual buses travel *on-time*, while that a tenth are *ahead*. On the
 731 other hand, the share of *significantly delay* is very negligible. The results show a similar trends for *NC* and *SH₃* with a
 732 slightly overestimation and underestimation of the share of buses ahead and on-time, respectively, and match well with
 733 the actual share of significantly delayed buses. On the other hand, *SH₀* follows a different trend by overestimating and
 734 underestimating, respectively, the share of buses early and on-time. Unlike lines *A* and *B*, the shares of *HH* appear to
 735 correspond well to the actual shares, outperforming those resulting from the scheduled-based scenarios.

736
 737 In total, the analyzed results show significant similarities between *NC* and *SH₃*, corresponding well to the actual
 738 shares of bus punctuality. In contrast, *SH₀* seems to strongly underestimate the share of early buses, while *HH* tends
 739 to overestimate the share of late buses.

740 5.3.2. *Impacts of Bus Holding: Bus Delay Severity Metric*

741 The bus punctuality shares based on bus severity delay metric for bus lines *A*, *B* and *C* are shown in Figure 20. The
 742 differences between the resulting simulation shares and the actual shares according to the second punctuality metric
 743 are reported in Table 22.

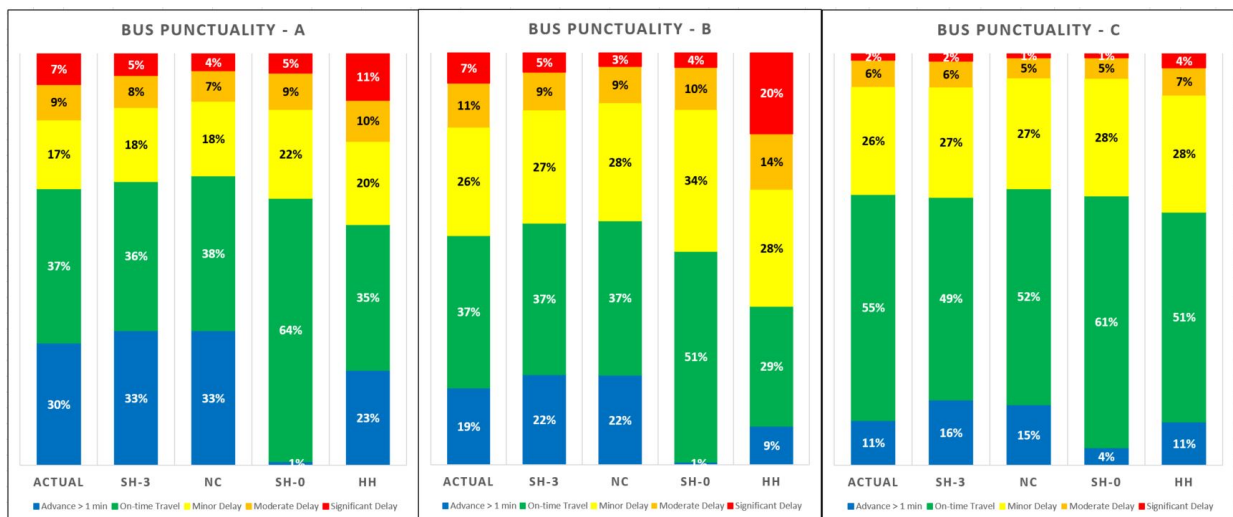


Figure 20: Impacts of bus holding strategies - Bus punctuality shares for the bus lines studied, based on the second on-time measurement

$\Delta(S, actual)$	no-control <i>NC</i>	scheduled-based <i>SH₃</i>	scheduled-based <i>SH₀</i>	headway-based <i>HH</i>
Line A	4%	4%	31%	9%
Line B	6%	4%	22%	17%
Line C	4%	6%	9%	4%

Table 22: Impacts of bus holding - Illustration of the differences and gaps between the resulting simulation shares and the actual shares according to the second punctuality metric, for bus lines A, B and C

744 From Figure 20 and Table 22, one can observe a resemblance between the shares of bus punctuality for the
745 actual observations, *NC* and *SH₃*, and that for the three studied bus lines. Exception to note for line C, part of the
746 share (around 5%) of *in advance* and *on-time travel* buses seems is poorly located. Furthermore, the share of minor,
747 moderate and significant delays is almost identical between *SH₃*, *NC* and actual observations. For *SH₀*, the resulting
748 early and on-time bus shares are heavily neglected and overestimated for lines A and B respectively, while the slightly
749 late bus shares are also overestimated, but to a lesser extent. On the other hand, for line C, one can observe a similar
750 trend but to a lesser extent to that of lines A and B, with an underestimation and an overestimation, respectively for
751 the share of buses *in advance* and *on-time*. For *HH*, the resulting shares of line A indicate an underestimate of the
752 share of early and on-time buses, but also a slight overestimation of share of late buses, a similar trend is observed
753 for line B but to a greater extent, while for line C, minor differences with the actual shares are observed, including an
754 underestimation of the share of on-time buses.

755 In short, the shares obtained according to this second bus punctuality measure follow similar trend to that of
756 the first measure. Compared to actual bus punctuality, the shares resulting from *NC* and *SH₃*, are on average more
757 accurate than those from *SH₀* and *HH*.

758 5.3.3. Impacts of Bus Holding: a Third Bus Delay Metric

759 After having presenting in detail the shares resulting from the application of the two bus punctuality metrics, for
760 the bus lines studied. The simulation results showed some differences between the actually observed delays and the
761 delays resulting from the simulation. To better understand the those resulting differences in more detail, the actual
762 shares will be compared over intervals of one minute increments, alongside the shares resulting from the simulated
763 scenarios.

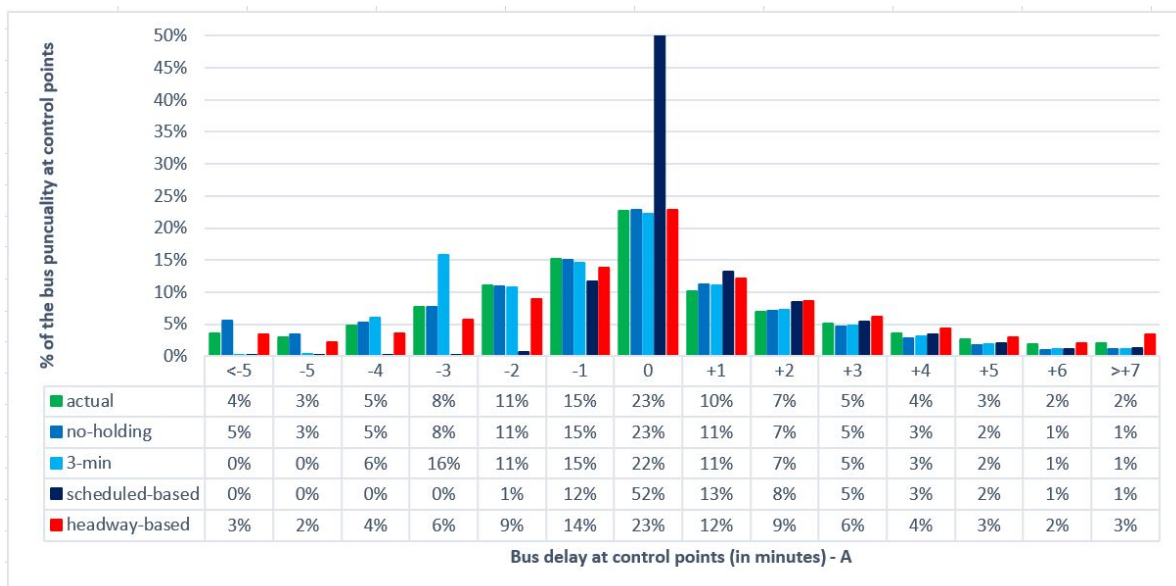


Figure 21: A comparison of actual and simulated bus punctuality shares at control points grouped in one-minute increments, for bus line A

764 In Figure 21 the relative frequencies of bus punctuality at the control points of bus line A are illustrated. The
 765 shares resulting of the *no-control* strategy (*NC*) correspond well to the actual shares with a negligible difference
 766 to note. One can observe that the shares of *SH₃* follow the same trend as those of *NC*, with the exception of the share
 767 of buses more than four minutes ahead, where the share is completely ignored. On the other hand, for buses two to
 768 three minutes ahead, the share is overestimated (+8). For *SH₀*, the share of early buses is completely underestimated,
 769 while *HH* appears to slightly underestimate the share of early buses, slightly overestimate the share of late buses, and
 770 match well with the share of on-time buses.

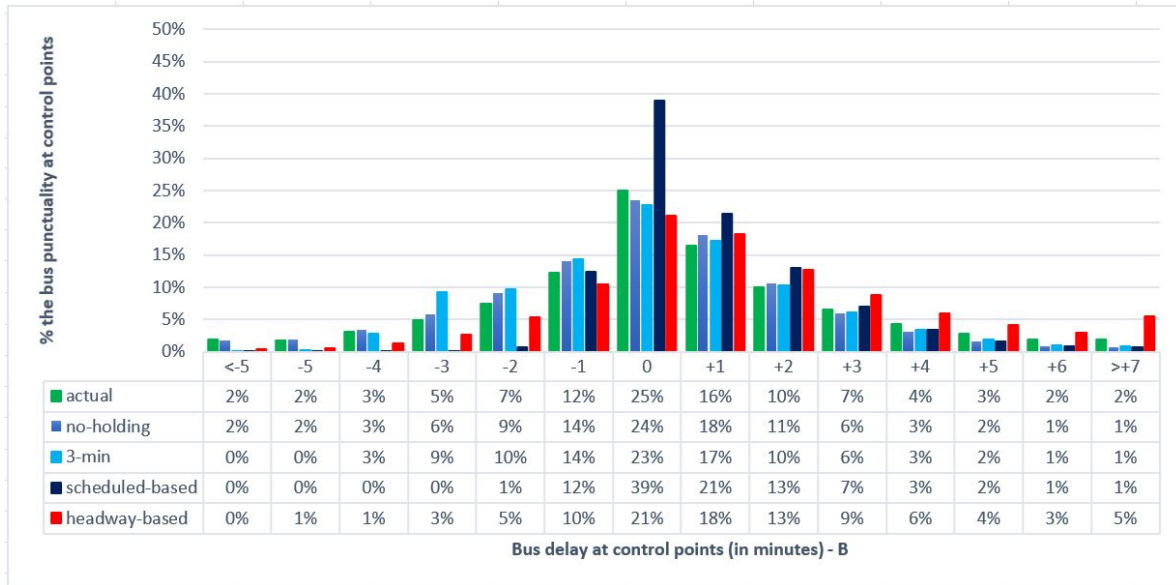


Figure 22: A comparison of actual and simulated bus punctuality shares at control points grouped in one-minute increments, for bus line B

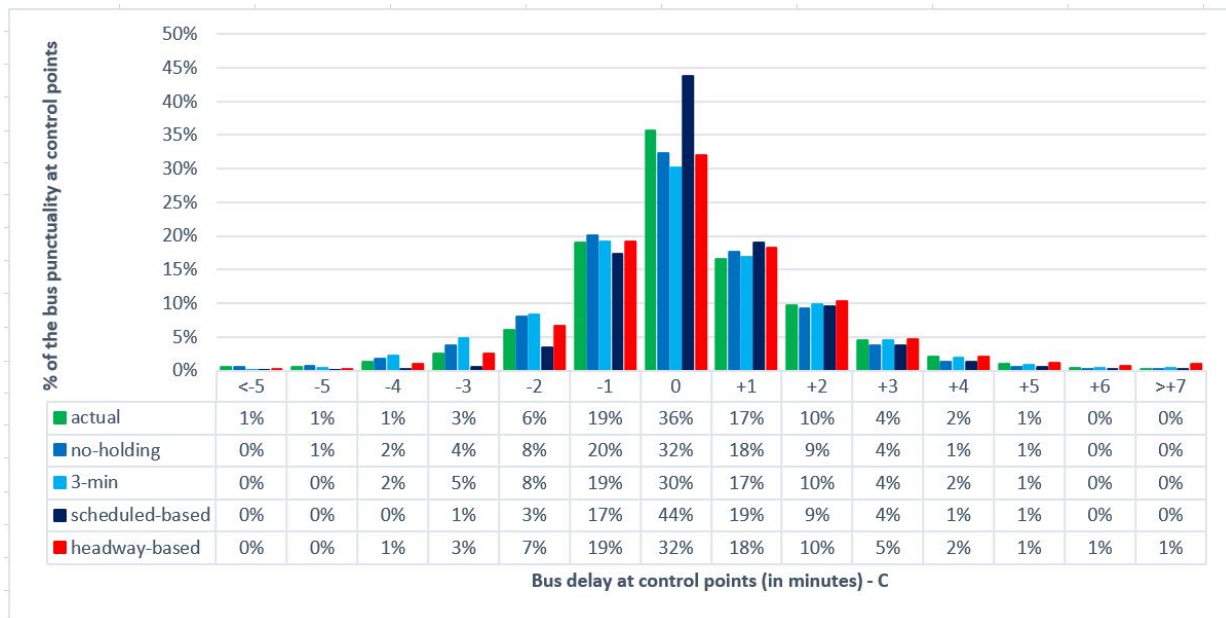


Figure 23: A comparison of actual and simulated bus punctuality shares at control points grouped in one-minute increments, for bus line C

771 Figure 22 illustrates the relative frequency of bus delays for line *B*. The results show that *NC* and to a lesser extent
 772 *SH₃*, ranging from good fit to slightly overestimate the share of buses ahead and on-time, respectively. Furthermore,
 773 *SH₀* and *SH₃* seem to strongly underestimate the share of buses in advance, with more than three minutes and one
 774 minute ahead, respectively. Moreover, the resulting shares of the scheduling-based holding scenarios follow the same
 775 pattern by slightly underestimating the share of buses with moderate to significant delay. On the other hand, *HH*
 776 appears to underestimate the share of early buses, to a lesser extent slightly underestimate and overestimate the share
 777 of on-time and late buses, respectively.

778 From Figure 23, it can be observed that the bus line *C* is the most punctual line among the lines studied. The results
 779 show that the resulting shares of scheduled-based holding scenarios, correspond well to the actual share of late buses.
 780 *NC* and *SH₃* slightly overestimate and underestimate the share of early buses and on-time buses, respectively. In
 781 contrast, *SH₀* significantly overestimates the share of buses on time and underestimates the share of buses in advance.
 782 On the other hand, *HH* seems to underestimate the share of on-time buses and to a lesser extent the share of buses too
 783 early, while it slightly underestimates the share of buses significantly delayed.

784 The differences between the resulting simulation shares and the actual shares according to the third punctuality
 785 metric are reported in Table 23.

$\Delta(S, actual)$	no-control <i>NC</i>	3-min <i>SH₃</i>	scheduled-based <i>SH₀</i>	headway-based <i>HH</i>
Line A	4%	11%	35%	8%
Line B	7%	10%	22%	14%
Line C	6%	6%	11%	5%

Table 23: Illustration of the differences and gaps between the resulting simulation shares and the actual shares according to the third punctuality metric, for bus lines *A*, *B* and *C*

786 Overall, *NC* seems to adequately match the actual shares of lines *A* and *B*, and on average outperform *SH₃* and
 787 *HH*. On the other hand, *HH* shares correspond well to the actual shares of line *C*, slightly outperforming *NC* and *SH₃*.
 788 In contrast, the *SH₀* corresponds the least to the actual shares, for the three studied lines.

789 5.3.4. Impacts of Bus Stop-Skipping on Bus Punctuality

790 In this section, the focus is on studying the impacts of bus stop-skipping approach on bus punctuality. The resulting
 791 bus punctuality shares - based on the second metric - of the two scenarios (*SK₃* and *SK₅*) in which stop-skipping is
 792 implemented, as well as the actual shares and the resulting shares of no control scenario, are shown in Figure 24.

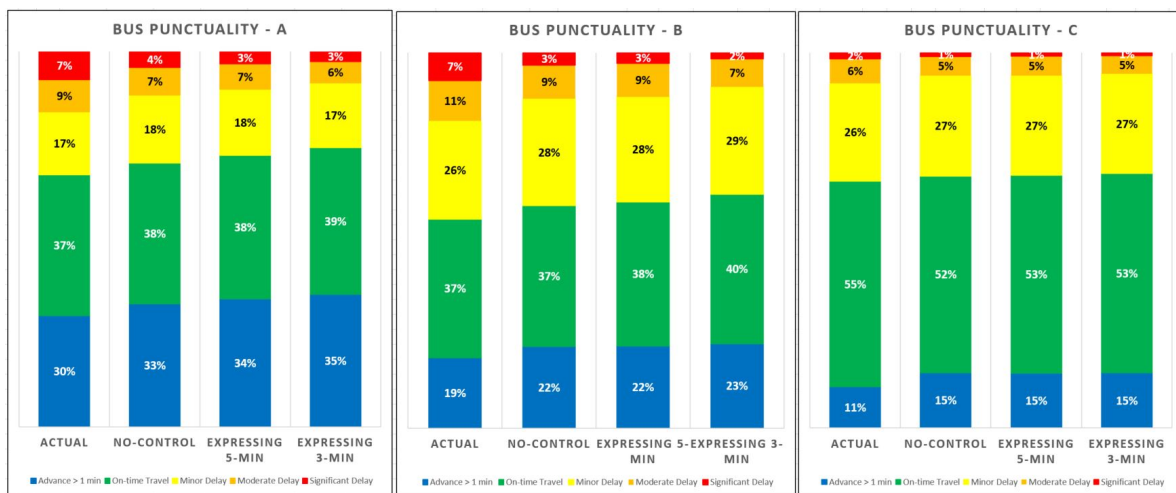


Figure 24: Bus punctuality shares for the bus lines studied, based on the second on-time measurement. Expressing 3-min and Expressing 5-min refer respectively to the resulting shares of scenarios *SK₃* and *SK₅*

793 For bus line *A* and *B*, we can observe a slight increase in the share of early and on-time buses, and at same time
 794 a slight decrease in share of moderately and significantly late buses, for SK_3 compared to NC . In contrast, for SK_5 ,
 795 negligible changes were observed. For line *C*, the shares of NC , SK_3 and SK_5 were almost the same. Compared to the
 796 actual shares, SK_3 and SK_5 seem to diverge, with a greater overestimation of the shares of early buses compared to
 797 NC . In sum, the applied stop-skipping strategy appears to have a slight impact on bus punctuality shares, in particular
 798 by decreasing the share of late buses.

799 **5.4. Simulation vs Actual: Off-Peak and Peak Hours**

800 In addition to the overall validation of bus travel times, previous analyzes using the three bus punctuality metrics
 801 have shown that the simulation results correspond well to the actual bus shares. In the following, a validation process
 802 will be conducted, in which the simulation results of no control scenario will be compared side-by-side with the
 803 actual observations, using the bus severity metric based on defined periods of the day (see Table 6). The simulation
 804 and actual bus punctuality shares, for the three bus lines, are given in Figure 25.

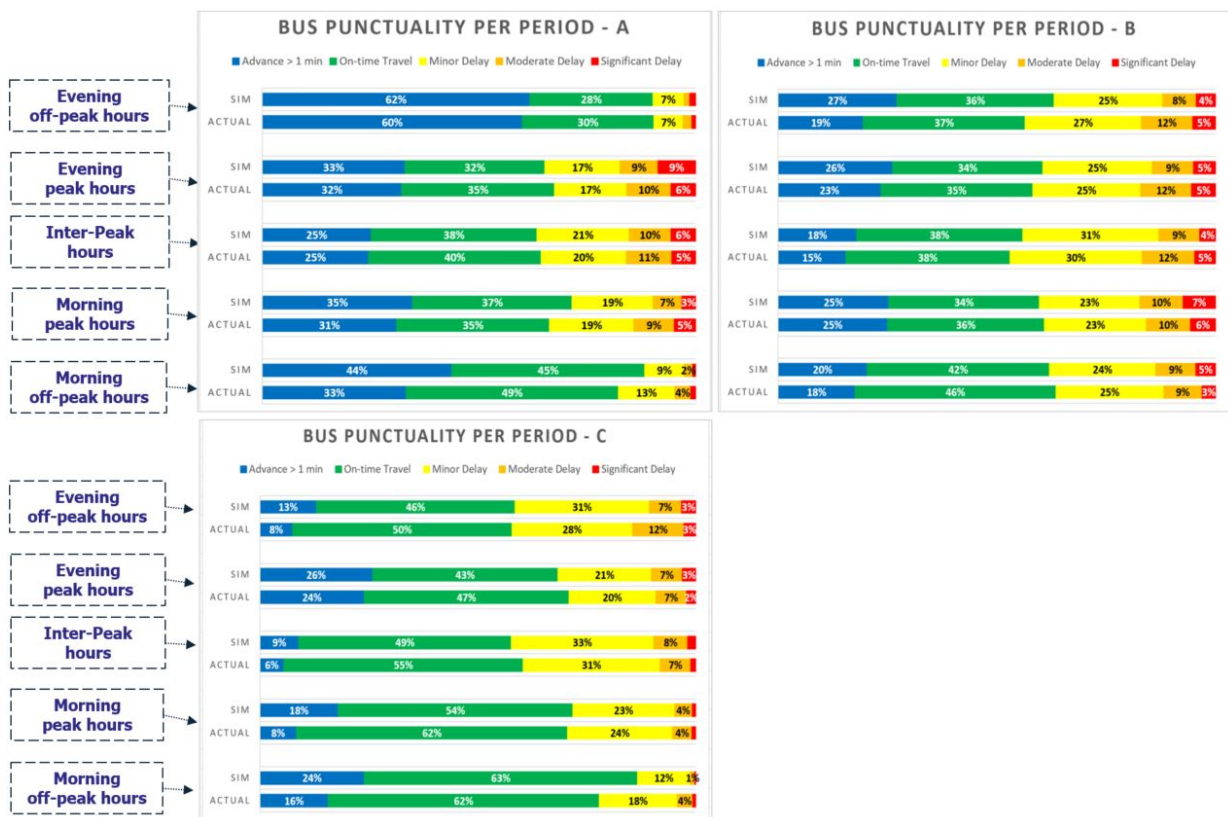


Figure 25: Punctuality shares of buses, per time slot of the day for the three lines studied

805 For bus line *A*, the actual and simulation delay shares appear to follow similar pattern over a number of periods.
 806 During morning peak hours, the share of *in advance* and *on-time travel* buses is slightly overestimated. Next, a similar
 807 trend is observed during inter-peak, evening peak and off-peak hours, which account for more than two-thirds of total
 808 trips. The biggest difference is observed during AM off-peak hours, with a significant overestimation of the share of
 809 *in advance* buses (+11%), and consequently a slight underestimation of the shares of *on-time* and *delayed* buses.

810
 811 For bus line *B*, moderate differences are observed compared to line *A*. A very similar trend is observed, between
 812 the two shares during the morning peak hours, and to a lesser extent during inter-peak and evening peak hours.
 813 Concerning the late evening shares, one can observe a significant overestimation of *in advance* buses (+8%). On the
 814 other hand, few buses run during this late period with approximately less than a tenth of total trips.

815 For bus line *C*, the simulation and actual shares essentially follow a similar trend during evening rush hours and, to
816 a lesser extent during inter-peak hours. Furthermore, larger differences can be observed during AM peak and off-peak
817 hours, and PM off-peak hours, with the share of buses *in advance* being overestimated by +5% in evening off-peak
818 hours, by +8% in morning off-peak hours, and strongly underestimated in morning rush hours with around +10%.

819
820 In short, the results presented in this section give us some insights into the extent to which the simulation is able
821 to reproduce reality. For the two main bus lines studied (*A* and *B*), the shares obtained show a similar trend during
822 the two peak hours and during the inter-peak hours, which represents more than three quarters of the total trips, for
823 the two bus lines. On the other hand, more significant differences are observed during off-peak hours in the morning
824 for line *A*, and during off-peak hours in the evening for line *B*. Concerning bus line *C*, it can be observed that the
825 differences in sharers are considerably greater compared to lines *A* and *B*, due to a certain overestimation on the share
826 of *in advance* buses, particularly during morning periods.

827 6. Discussion and Perspectives

828 In this section, a discussion of the obtained results, including actual and simulated travel times and bus punctu-
829 ality, will be discussed in depth. Next, the feature selection process along with an in-depth discussion of simulation
830 parameters will be presented. Finally, the limitations and perspectives of the proposed work will be given.

831 6.1. Discussion: Results

832 In this first section, the resulting travel times and bus punctuality will be analyzed. Furthermore, a discussion on
833 the impacts of bus recovery time, holding and stop-skipping strategies will be detailed.

834 6.1.1. Analysis of Travel Times

835 From the results at hand, the theoretical travel times seem to differ slightly from the observed travel times, these
836 differences can be explained by: on the one hand, the theoretical arrival times at control points are rounded to nearest
837 minute (*hh:mm:00*); on the other hand, on the last route segments, bus drivers will often tend to reach the terminal
838 stop as quickly as possible. Furthermore, differences travel times between routes on the same line can be explained by
839 a different level of demand and traffic in each direction. For line *B*, another explanation for the observed travel times
840 differences is plausible, with the *Retour* route has more segments than *Aller* route. Furthermore, route segments *B-A1*
841 and *B-A2*, together with *B-R4* to *B-R7* follow the same road but in opposite directions, this latter split of the *Retour*
842 route, resulting in a higher number of control points, generates some of the significant differences due to shorter
843 segments and therefore shorter travel times. For line *C*, the differences are insignificant in both directions, can be
844 explained by the fact that the bus routes of line *C* are much shorter compared to *A* and *B*.

845 The simulated travel times provided by ML models (f_1 , f_2 and f_3), are highly resembled and statistically similar
846 to the actual travel times. For line *C*, the differences across all route segments are negligible, averaging a few seconds
847 per route segment. These similarities can be explained by the nature of travel time variability, which to some extent
848 corresponds to probabilistic distributions regardless of the features selected to train ML models. In sum, ML models
849 with stop time as feature (f_1 and f_2) perform better than those built by only considering bus delay at the stop (f_3).
850 Moreover, ML models with only the stop time as a feature (f_1) relatively match the best actual travel times. As
851 expected, bus travel times are highly dependent on the time of the day. In contrast, travel times seem less sensitive to
852 the bus delay, whether the bus is late or not.

853
854 Overall, ML travel times are on average more accurate compared to the travel times estimated by probabilistic
855 distributions, according to both goodness of fit metrics *MAE* and *RMSE*, during morning and evening peak hours.
856 Furthermore, the travel times obtained during the evening rush hours appear more accurate than those during morning
857 hours, which may be explained by a better fitting of evening travel times using probabilistic distributions compared to
858 the morning travel times.

859 *6.1.2. Analysis of Bus Punctuality*

860 *Impacts of Bus Recovery Time Strategies.* The share of bus punctuality, resulting from the first two simulated sce-
861 narios (S_1 and S_2) appear to closely follow the actual shares according to the first metric, S_2 seems to significantly
862 outperform S_1 according to the second metric, in which S_1 , greatly overestimates the share of buses on time. This
863 could be explained by the applied recovery time strategy, in which the latter simulation parameter was set to a min-
864 imum of two minutes in S_1 , while it was adjusted in S_2 . The choice of recovery time strategy will be discussed in
865 depth below.

866 *Impacts of Bus Holding Strategies.* The second set of scenarios aims to investigate the impacts of bus holding strate-
867 gies on bus punctuality (see Table 10). The shares obtained according to the first and second metric, show significant
868 similarities between NC and SH_3 , with both on average more accurate than those from SH_0 and HH . In contrast, SH_0
869 seems to strongly underestimate the share of early buses, while HH tends to overestimate the share of late buses. Ac-
870 cording to third metric, NC seems to adequately match the actual shares of lines A and B , and on average outperform
871 SH_3 and HH . On the other hand, HH shares correspond well to the actual shares of line C , slightly outperforming NC
872 and SH_3 . In contrast, the SH_0 corresponds the least to the actual shares, for the three studied lines. The bus holding
873 strategies implemented, including holding criteria will be discussed in detail below.

874 *Impacts of Stop-Skipping Strategy.* The third set of implemented scenarios (SK3 and SK5) aims to investigate the
875 impacts of bus stop-skipping strategy on bus punctuality. The observed slight decrease of the shares of moderately
876 and significantly late buses is the result of applying the stop-skipping strategy. Overall, this strategy has a negligible
877 impact on bus punctuality shares, which can be explained by: 1) a low share of buses moderately and significantly
878 late, and 2) the impact of the defined skipping criteria, which will be discussed in detail below.

879 *Simulation vs Actual: Off-Peak and Peak Hours.* Besides, the simulation results fit well with the actual data, in terms
880 of bus punctuality during the two peak hours and during the inter-peak hours, for lines A and B and, to a lesser extent,
881 for line C , in which the differences are more significant, due to an overestimation of the share of *in advance* buses,
882 particularly during morning periods. These larger differences may be explained by the nature of line C , with a low
883 number of route segments (two segments per route) and by relatively short travel times with on average 10 minutes
884 for line C compared to 30 minutes for line A and 23 minutes for line B .

885 On the other hand, the simulation model shows a certain difficulty in adjusting the shares of bus delay during
886 off-peak hours in the morning and evening. This lack might be caused by an overestimation of the share of early buses
887 mainly due to a low number of trips during off-peak hours, and therefore a smaller dataset to train and build more
888 accurate ML models. An interesting observation needs to be pointed. For line A , a very high share of early buses
889 is observed during evening off-peak hours, which can be explained by a lower traffic level, fewer passengers, which
890 implies short dwell times and therefore shorter travel times.

891 *6.2. Discussion: Feature Selection*

892 The training dataset is mainly derived from actual bus AVL data, with bus vehicle and trip information such as
893 bus *stop time*, bus *stop delay* and segment *travel time*, in addition to other derived information such as *day of the*
894 *week* and *season*. Analyzing the resulting mean decrease in Gini score, *stop time* and *stop delay* appear to be the
895 two most relevant features. [Additionally, an in-depth analysis of the input features was then conducted, showing a](#)
896 [higher \$R^2\$ score for ML models built with *stop time* and *stop delay* compared to those built with weather features.](#)
897 The selected features corresponding to the information of a bus during a trip, which make the ML models relatively
898 simple to implement. On the other hand, training a ML model with few features will likely result in an underfit model.
899 Additionally, the ML models implemented in this study are designed to predict the travel time of a bus independently
900 on its fleet. By including information about buses ahead will potentially enhance the quality of trained models. In
901 contrast, some features may be highly correlated with each other due to the temporal nature of the data, which can
902 lead to multicollinearity issues.

903 *6.3. Discussion: Simulation Parameters*

904 [In this section, the simulation parameters including bus recovery time strategies, scheduled-based, headway-based](#)
905 [and stop-skipping holding criteria will be discussed in depth.](#)

906 6.3.1. Bus Recovery Time

907 Another point to discuss is the choice of bus recovery time strategy. Sufficient recovery time should be built into
908 the schedule to ensure that the delays encountered during one trip do not carry over on subsequent trips. In this study,
909 two strategies S_1 and S_2 were implemented, with a fixed minimum recovery time up to two minutes in S_1 and an
910 adjusted recovery time estimated from the observed recovery times in S_2 .

911 Applying a fixed recovery time, as with S_1 , is relatively simple to implement without requiring other correspond-
912 ing data. On the other hand, the implementation of a short duration straightforward at the end of the trip will likely
913 have an impact on the bus departure time, only if the arriving bus was very late, otherwise the next bus departure will
914 be at the predefined time. From the departure delays observed in Figure 15, one can see that buses will likely depart
915 one minute early and up to two minutes late than the scheduled time. S_2 attempted to overcome the limitations of S_1 ,
916 by adjusting the departure delay based on those observed, which improved the quality of simulated bus punctuality.
917 However, the application of this second approach has a major drawback, as it will require historical data, which not
918 always available, limiting its application to other cases studies.

919 6.3.2. Holding Criteria

920 *Scheduled-Based Holding.* In this section, the value of holding criteria s will be discussed with emphasis on the
921 three variants of the scheduled-based holding strategy. With $s = 0$ (SH_0), buses were held whenever were ahead
922 of schedule. This strategy seems too theoretical and does not reflect actual bus trends, leading to near ignorance on
923 the share of early buses. Bus drivers tend to tolerate being to some extent ahead of schedule and will therefore not
924 systematically hold up the bus each time they are early. In fact, bus operators considered a bus up to a minute early
925 to be on time [114]. With $s = +\infty$ (NC), no holding control strategy was applied. Surprisingly, the obtained bus
926 punctuality shares seem to correspond best to the actual shares, with a slight overestimation of buses ahead. Once
927 again, this trend may be explained by a certain tolerance of buses ahead to a certain extent on the lines studied. In
928 order to understand to what extent the holding criteria implemented correspond to the actual criteria, in other words,
929 how early does a bus to be to be held? A third holding strategy was implemented with $s = 3$ minutes (SH_3). The
930 results show that even with more than three minutes ahead, some of *in advance* buses were not held.

931 *Headway-Based Holding.* The headway-based holding strategy HH was implemented with $\alpha = 0.8$, which implies
932 that a bus will only be held if its headway is more than 20% shorter than the scheduled headway. For lines A and B , HH
933 seems to slightly underestimate and overestimate the shares of *ahead* and *late* buses respectively. Explanations may
934 be as follows: 1) HH aims to maintain the predefined headway between running buses and, with the holding control of
935 a bus heavily dependent on the preceding bus, an early or delayed departure of the bus will have a subsequent impact
936 on the following buses; 2) In this study, service reliability is evaluated according to punctuality measures, more suited
937 to scheduled-based strategies than to headway-based strategies, since it evaluates one vehicle at a time, and in which
938 maintaining the headway may not be enough to guarantee the punctuality of buses. For line C , HH presents the best
939 fit with actual shares, which can be explained by: 1) short travel times, on average 10 to 12 minutes; 2) The holding
940 control is carried out at a single timing point, since the first and last points correspond to the origin and destination
941 stops respectively; 3) Line C has a relatively large headway compared its route travel time (6/10 minutes versus 12
942 minutes), and therefore buses will mostly not be held.

943 6.3.3. Stop-Skipping Criteria

944 In this section, the value of stop-skipping criteria D_{max} will be discussed. The stop-skipping will only occur if
945 the bus delay at the control point is greater than D_{max} , and if not, the skipping will not performed. This can lead to
946 additional delays, particularly if the bus continues to run slower than than expected, and subsequently further skipped
947 stops, which likely explains the relative slight variation in the share of moderately and significantly late buses, for line
948 A and B . For line C , with only one skipping stop and relatively short travel times, stop-skipping may not occur, which
949 may explain the no change in terms of SK_3 and SK_5 shares compared to NC .

950 Another factor that we need to highlight is the performed dwell time, in which alighting is allowed, while boarding
951 is denied. Besides, from the available data, the dwell time is considered as part of the running time, which may to
952 some extent overestimate the simulated travel times and thus generate a slight delay for buses.

6.4. Limitations and Perspectives

This study has potential weaknesses and limitations. First, relying on data from just three bus lines in the same city might not fully capture the complexity and variability of urban public transportation systems. Additionally, the model's effectiveness in other contexts, cities, or transportation modes may not have been thoroughly explored or validated. Although this study attempted to validate the simulation framework based on ML results, it will be interesting to extend its scope to other cities by conducting simulation scenarios with a larger number of bus lines.

The dataset for bus travel time prediction is somewhat poor, including only trip information. Therefore, it will be interesting to enrich the training dataset, with information about traffic and road characteristics such as: speed limit and traffic lights, in order to: on the one hand, to improve the accuracy of ML prediction and, on the other hand, to extend the scope of the application to other bus lines. Another limitation needs to be pointed out regarding the resilience and adaptability of the framework. The simulation scenarios were carried out with emphasis on working day, it will be interesting to carry out other scenarios with emphasis on week-ends and public holidays. In addition, incorporating more diverse dataset, including special event, will test the resilience and adaptability of the framework.

The results of this study put into perspective certain points on which it will be interesting to work. 1) Running a simulation model with ML models can be time-consuming, especially for bus lines with a large number of segments. Thus, in order to overcome this drawback, several ML-Simulation coupling strategies, can be implemented. 2) Another direction for future research concerns the findings of this study. Although previous analyzes have shown that the simulation model produces good results according to a number of bus on-time metrics, it is also interesting to investigate the extent to which operational recovery time influences bus on-time performance. 3) One further step could be the application of different strategies to eliminate delay occurrences, which can be useful to bus operators to improve service regularity.

7. Conclusions

This paper sheds light on the evaluation of public transportation performance, which constitutes an important issue facing operators. In this paper, we aim to propose a validation framework integrating microscopic simulation model with ML techniques. We introduced a hybrid approach combining regression trees trained with three feature sets for travel times prediction and probabilistic distributions to accurately estimate travel time variability to feed simulation model. Furthermore, a case study on three bus lines was carried out, in which two scenarios (S_1 and S_2) and six scenarios (SH_0 , SH_3 , NC , HH , SK_3 and SK_5) were implemented, respectively, to study the impacts of recovery times and bus control strategies on bus punctuality. The results show that the travel times simulated using ML were very resembled, with f_1 adequately matching the best actual travel times with minor differences. They also show that the proposed hybrid approach combining ML with probabilistic distributions can better better estimate travel time variability than probabilistic distributions on average.

A first analysis of bus punctuality was carried out, showing a good match between the actual and simulated simulated shares resulting of S_1 with a fixed minimum recovery time and S_2 with an adjusted recovery time derived from actual data, with overall S_2 more precise than S_1 . A second analysis was carried out. Surprisingly NC with no holding control seems to correspond best to actual shares, while SH_0 seems too theoretical and does not reflect actual bus trends, leading to high underestimation of bus shares ahead. HH with headway-based holding strategy, presented the best fit with actual shares for line C . A third analysis was carried out, showing a limited impact of the implemented bus stop-skipping strategy on the bus punctuality, due to a low share of late buses and the limitation of stop-skipping to buses very late. We also conduct an error analysis on the punctuality of simulated buses depending on the time of the day. The simulated buses follow a similar trend during inter-peak and peak hours, and to a lesser extent during the off-peak hours, in which lines A and B seem to correspond better to actual shares compared to C . The results show that bus travel times seem strongly dependent on the time of the day and, on the other hand, seem less sensitive to bus delays. Furthermore, the departure delay at the originating terminal as well as the applied holding control strategy are the two most influential factors on bus punctuality.

Further research should involve further testing with a larger number of lines and different sets of control points to evaluate the performance of the proposed framework. Additionally, actual data can be enriched by including information on traffic and road characteristics. This will improve the quality of the input data and improve the effectiveness and robustness of the ML model. The results of this study benefit bus operators to assist them in decision-making, by

1002 providing appropriate simulation parameters that correspond to a large extent to the actual situation, and thus realize
 1003 new scenarios, in actual and similar contexts, based on validated results. We believe this paper is of interest especially
 1004 for the practitioners in the field, as it provides in-depth insight into how ML and public transport simulation model
 1005 can be integrated into the same framework, to improve quality simulation results.

1006
 1007 **Author Contributions:** Investigation, Y.D.; Methodology, Y.D. and O.C; Supervision, O.C. and M.N.; Writing—
 1008 original draft, Y.D.; Writing—review and editing, O.C, M.N. and M.H.; Funding acquisition, M.H. All authors have
 1009 read and agreed to the published version of the manuscript.

1010 **Conflicts of Interest:** The authors declare no conflict of interest.

1011
 1012 **Acknowledgement:** The authors thank the A1 Statistiques company and its operator for making the data available.
 1013 They also thank Vincent Leblond of the Tellae company and François Queyroi and Gabriel Ferretini of DUKe, a
 1014 research team of the Laboratory of Digital Sciences of Nantes (LS2N), for their help for structuring and analysing the
 1015 data. They also thank the region of Pays de la Loire for SIMULBUS grant.

1016 **Appendix A. Data Section**

date	vehicle id	trip id	stop id	stop sequence	scheduled arrival time	scheduled departure time	real arrival time	real departure time
20200916	VCL-255	TRP-21022	STP-1883	1	07:41:00	07:41:00	07:41:22	07:41:22
20200916	VCL-255	TRP-21022	STP-1884	2	07:42:20	07:42:20	07:43:21	07:43:21
...
20200916	VCL-255	TRP-21022	STP-982	7	07:47:00	07:47:00	07:48:10	07:48:10
20200916	VCL-255	TRP-21022	STP-983	8	07:47:53	07:47:53	07:49:12	07:49:12
...
20200916	VCL-255	TRP-21022	STP-990	16	07:56:00	07:56:00	07:58:15	07:58:15
20200916	VCL-255	TRP-21022	STP-991	17	07:57:06	07:57:06	07:58:52	07:58:52
...
20200916	VCL-255	TRP-21022	STP-994	20	08:01:00	08:01:00	08:02:07	08:02:07
20200916	VCL-255	TRP-21022	STP-995	21	08:01:50	08:01:50	08:02:49	08:02:49
...
20200916	VCL-255	TRP-21022	STP-998	23	08:04:00	08:04:00	08:05:38	08:05:38
20200916	VCL-255	TRP-21022	STP-2105	24	08:05:14	08:05:14	08:06:14	08:06:14
...
20200916	VCL-255	TRP-21022	STP-1004	30	08:10:20	08:10:20	08:10:59	08:10:59
20200916	VCL-255	TRP-21022	STP-1005	31	08:11:00	08:11:00	08:11:10	08:11:10

Table A.24: A sample of combined data, including scheduled and actual time attributes

from stop	to stop	section index	scheduled from stop time	scheduled to stop time	scheduled section travel time	real from stop time	real to stop time	real section travel time
STP-1883	STP-1884	1	07:41:00	07:42:20	00:01:20	07:41:22	07:43:21	00:01:59
...
STP-982	STP-983	7	07:47:00	07:47:53	00:00:53	07:48:10	07:49:12	00:01:02
...
STP-990	STP-991	16	07:56:00	07:57:06	00:01:06	07:58:15	07:58:52	00:00:37
...
STP-994	STP-995	20	08:01:00	08:01:50	00:00:50	08:02:07	08:02:49	00:00:42
...
STP-998	STP-2105	23	08:04:00	08:05:14	00:01:14	08:05:38	08:06:14	00:00:36
...
STP-1004	STP-1005	30	08:10:20	08:11:00	00:00:40	08:10:59	08:11:10	00:00:11

Table A.25: A sample dataset aggregated at section level, including: section stops and index, scheduled and real times, and derived travel times

from stop	to stop	segment index	scheduled from stop time	scheduled to stop time	scheduled segment travel time	real from stop time	real to stop time	real segment travel time
STP-1883	STP-982	1	07:41:00	07:47:00	00:06:00	07:41:22	07:48:10	00:06:48
STP-982	STP-990	2	07:47:00	07:56:00	00:09:00	07:48:10	07:58:15	00:10:05
STP-990	STP-994	3	07:56:00	08:01:00	00:05:00	07:58:15	08:02:07	00:03:52
STP-994	STP-998	4	08:01:00	08:04:00	00:03:00	08:02:07	08:05:38	00:03:31
STP-998	STP-1005	5	08:04:00	08:11:00	00:07:00	08:05:38	08:11:10	00:05:32

Table A.26: A sample dataset aggregated at segment level

1017 **Appendix B. Validation of Travel Times**

segment ID	segment stops	TT_{theo}	TT_{actual}	TT_{f1}	TT_{f2}	TT_{f3}	$\Delta_{t,a}$	Δ_{f1}	Δ_{f2}	Δ_{f3}
A-A1	STP-1006 → STP-1013	434	429	408	395	402	-1%	-5%	-8%	-6%
A-A2	STP-1013 → STP-1017	188	121	123	131	124	-35%	1%	8%	2%
A-A3	STP-1017 → STP-1022	371	420	422	421	432	13%	1%	0%	3%
A-A4	STP-1022 → STP-1030	584	603	617	619	636	3%	2%	3%	5%
A-A5	STP-1030 → STP-1846	597	419	450	470	487	-30%	7%	12%	16%
-	-	-	-	-	-	-	-	-	-	-
A-R1	STP-1883 → STP-982	383	423	423	421	420	10%	0%	0%	-1%
A-R2	STP-982 → STP-990	620	629	648	637	636	1%	3%	1%	1%
A-R3	STP-990 → STP-994	322	253	262	261	275	-21%	3%	3%	9%
A-R4	STP-994 → STP-998	247	226	230	229	235	-8%	2%	1%	4%
A-R5	STP-998 → STP-1005	392	333	342	346	362	-15%	3%	4%	9%

Table B.27: Reported average travel times (in seconds) as well as derived relative metric Δ , for each route segment of bus line A

segment ID	segment stops	TT_{theo}	TT_{actual}	TT_{f1}	TT_{f2}	TT_{f3}	$\Delta_{t,a}$	Δ_{f1}	Δ_{f2}	Δ_{f3}
C-A1	STP-2521 → STP-2525	280	282	275	276	277	1%	-3%	-2%	-2%
C-A2	STP-2525 → STP-2526	323	343	333	334	336	6%	-3%	-3%	-2%
-	-	-	-	-	-	-	-	-	-	-
C-R1	STP-2526 → STP-2527	288	353	347	345	349	22%	-2%	-2%	-1%
C-R2	STP-2527 → STP-2440	297	226	223	223	225	-24%	-1%	-1%	0%

Table B.28: Reported average travel times (in seconds) as well as derived relative metric Δ , for each route segment of bus line C

segment ID	segment stops	TT_{theo}	TT_{actual}	TT_{f1}	TT_{f2}	TT_{f3}	$\Delta_{t,a}$	Δ_{f1}	Δ_{f2}	Δ_{f3}
B-A1	STP-2349 → STP-1380	530	560	569	568	560	6%	2%	2%	0%
B-A2	STP-1380 → STP-1390	588	568	587	578	586	-3%	3%	2%	3%
B-A3	STP-1390 → STP-1394	297	249	254	255	296	-16%	2%	2%	19%
B-A4	STP-1394 → STP-2408	480	433	454	453	486	-10%	5%	5%	12%
B-A5	STP-2408 → STP-2746	75	92	80	79	80	22%	-13%	-14%	-13%
-	-	-	-	-	-	-	-	-	-	-
B-R1	STP-2747 → STP-2348	75	114	110	112	110	52%	-4%	-2%	-3%
B-R2	STP-2348 → STP-1411	449	417	422	417	420	-7%	1%	0%	1%
B-R3	STP-1411 → STP-1415	270	231	238	237	240	-15%	3%	3%	4%
B-R4	STP-1415 → STP-1420	224	286	296	291	295	28%	4%	2%	3%
B-R5	STP-1420 → STP-1424	444	315	327	334	365	-29%	4%	6%	16%
B-R6	STP-1424 → STP-1379	251	353	366	370	359	41%	4%	5%	2%
B-R7	STP-1379 → STP-834	206	169	174	223	174	-18%	3%	32%	3%

Table B.29: Reported average travel times (in seconds) as well as Δ values, for each route segment of bus line B

segment ID	Morning peak hours				Evening peak hours			
	TT_{actual}	Δ_{f1}	Δ_{f2}	Δ_{f3}	TT_{actual}	Δ_{f1}	Δ_{f2}	Δ_{f3}
A-A1	513	-10%	-14%	-22%	422	-3%	-6%	-3%
A-A2	132	1%	2%	-3%	125	1%	-1%	-1%
A-A3	481	-1%	0%	-8%	433	-5%	-4%	1%
A-A4	653	-1%	0%	-3%	619	-1%	0%	4%
A-A5	500	10%	11%	13%	433	1%	3%	13%
-	-	-	-	-	-	-	-	-
A-R1	418	-3%	-4%	0%	472	1%	1%	-10%
A-R2	628	-2%	-4%	0%	706	4%	4%	-7%
A-R3	248	1%	-3%	6%	271	3%	8%	13%
A-R4	231	-1%	-4%	1%	253	0%	2%	-9%
A-R5	349	-5%	-2%	0%	349	2%	3%	14%

Table B.30: Reported actual average travel times TT_{actual} as well as the relative differences Δ_{f1} , Δ_{f2} and Δ_{f3} , for bus line A, over morning and evening peak-hours

segment ID	Morning peak hours				Evening peak hours			
	TT_{actual}	Δ_{f1}	Δ_{f2}	Δ_{f3}	TT_{actual}	Δ_{f1}	Δ_{f2}	Δ_{f3}
C-A1	284	-3%	-3%	-3%	300	-6%	-6%	-7%
C-A2	342	-3%	-5%	-2%	395	-5%	-4%	-14%
-	-	-	-	-	-	-	-	-
C-R1	373	-1%	-2%	-7%	396	-4%	-4%	-10%
C-R2	231	-3%	-3%	-4%	233	0%	-1%	-2%

Table B.31: Reported actual average travel times TT_{actual} as well as the relative differences Δ_{f1} , Δ_{f2} and Δ_{f3} , for bus line C, over morning and evening peak-hours

segment ID	Morning peak hours				Evening peak hours			
	TT_{actual}	Δ_{f1}	Δ_{f2}	Δ_{f3}	TT_{actual}	Δ_{f1}	Δ_{f2}	Δ_{f3}
B-A1	573	-2%	-3%	-2%	625	-2%	-2%	-10%
B-A2	617	0%	-2%	-5%	635	-1%	-2%	-7%
B-A3	305	-9%	-9%	-3%	313	-1%	-1%	-4%
B-A4	548	2%	-3%	-6%	450	1%	1%	9%
B-A5	105	-14%	-15%	-24%	103	-27%	-28%	-24%
-	-	-	-	-	-	-	-	-
B-R1	99	-5%	0%	10%	131	-4%	-2%	-16%
B-R2	413	-1%	-3%	1%	490	-2%	-3%	-14%
B-R3	284	1%	0%	-14%	260	1%	1%	-4%
B-R4	310	0%	-2%	-5%	312	0%	-1%	-6%
B-R5	342	8%	9%	14%	352	2%	2%	9%
B-R6	375	-2%	-1%	-4%	384	-1%	-2%	-8%
B-R7	188	-3%	4%	-8%	179	-1%	4%	-4%

Table B.32: Reported actual average travel times TT_{actual} as well as the relative differences Δ_{f1} , Δ_{f2} and Δ_{f3} , for bus line B, over morning and evening peak-hours

1018 **References**

- 1019 [1] A. Jindal, G. S. Aujla, N. Kumar, R. Chaudhary, M. S. Obaidat, I. You, Sedative: Sdn-enabled deep learning architecture for network traffic
1020 control in vehicular cyber-physical systems, *IEEE network* 32 (6) (2018) 66–73.
- 1021 [2] T. Reich, M. Budka, D. Hulbert, Bus journey simulation to develop public transport predictive algorithms, *Soft Computing Letters* 3 (2021)
1022 100029.
- 1023 [3] N. Van Oort, *Service reliability and urban public transport design*, Citeseer, 2011.
- 1024 [4] A. Ceder, *Public transit planning and operation: Theory, Modeling and practice*. Oxford: Elsevier.
- 1025 [5] R. B. Noland, J. W. Polak, Travel time variability: a review of theoretical and empirical issues, *Transport reviews* 22 (1) (2002) 39–54.
- 1026 [6] B. Büchel, F. Corman, Modelling probability distributions of public transport travel time components, in: 18th Swiss Transport Research
1027 Conference (STRC 2018), STRC, 2018.
- 1028 [7] M. Taylor, Travel time variability—the case of two public modes, *Transportation Science* 16 (4) (1982) 507–521.
- 1029 [8] E. B. Emam, H. Al-Deek, Using real-life dual-loop detector data to develop new methodology for estimating freeway travel time reliability,
1030 *Transportation research record* 1959 (1) (2006) 140–150.
- 1031 [9] N. Uno, F. Kurauchi, H. Tamura, Y. Iida, Using bus probe data for analysis of travel time variability, *Journal of Intelligent Transportation*
1032 *Systems* 13 (1) (2009) 2–15.
- 1033 [10] L.-M. Kieu, A. Bhaskar, E. Chung, Public transport travel-time variability definitions and monitoring, *Journal of Transportation Engineering*
1034 141 (1) (2015) 04014068.
- 1035 [11] E. Durán-Hormazábal, A. Tirachini, Estimation of travel time variability for cars, buses, metro and door-to-door public transport trips in
1036 Santiago, Chile, *Research in Transportation Economics* 59 (2016) 26–39.
- 1037 [12] Z. Dai, X. Ma, X. Chen, Bus travel time modelling using gps probe and smart card data: A probabilistic approach considering link travel
1038 time and station dwell time, *Journal of Intelligent Transportation Systems* 23 (2) (2019) 175–190.
- 1039 [13] A. Polus, A study of travel time and reliability on arterial routes, *Transportation* 8 (2) (1979) 141–151.
- 1040 [14] W. C. Jordan, M. A. Turnquist, Zone scheduling of bus routes to improve service reliability, *Transportation science* 13 (3) (1979) 242–268.
- 1041 [15] H. Al-Deek, E. B. Emam, New methodology for estimating reliability in transportation networks with degraded link capacities, *Journal of*
1042 *intelligent transportation systems* 10 (3) (2006) 117–129.
- 1043 [16] S. Susilawati, M. A. Taylor, S. V. Somenahalli, Distributions of travel time variability on urban roads, *Journal of Advanced Transportation*
1044 47 (8) (2013) 720–736.
- 1045 [17] M. A. Taylor, Fosgerau’s travel time reliability ratio and the burr distribution, *Transportation Research Part B: Methodological* 97 (2017)
1046 50–63.
- 1047 [18] M. Harsha, R. H. Mulangi, H. D. Kumar, Analysis of bus travel time variability using automatic vehicle location data, *Transportation*
1048 *Research Procedia* 48 (2020) 3283–3298.
- 1049 [19] Z. Ma, L. Ferreira, M. Mesbah, S. Zhu, Modeling distributions of travel time variability for bus operations, *Journal of Advanced Transporta-*
1050 *tion* 50 (1) (2016) 6–24.
- 1051 [20] V. J. M. Low, H. L. Khoo, W. C. Khoo, Quantifying bus travel time variability and identifying spatial and temporal factors using burr
1052 distribution model, *International Journal of Transportation Science and Technology* 11 (3) (2022) 563–577.
- 1053 [21] R. Padmanaban, L. Vanajakshi, S. C. Subramanian, Estimation of bus travel time incorporating dwell time for apts applications, in: 2009
1054 *IEEE Intelligent vehicles symposium, IEEE, 2009*, pp. 955–959.
- 1055 [22] J. Wang, L. Sun, Dynamic holding control to avoid bus bunching: A multi-agent deep reinforcement learning framework, *Transportation*
1056 *Research Part C: Emerging Technologies* 116 (2020) 102661.
- 1057 [23] D. J. Morgan, A microscopic simulation laboratory for advanced public transportation system evaluation, Ph.D. thesis, Massachusetts
1058 *Institute of Technology* (2002).
- 1059 [24] M. Errampalli, M. Okushima, T. Akiyama, Microscopic simulation model considering public transport policy, *Journal of the Eastern Asia*
1060 *Society for Transportation Studies* 6 (2005) 2718–2733.
- 1061 [25] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, E. Wießner, Micro-
1062 *scopic traffic simulation using sumo*, in: 2018 21st international conference on intelligent transportation systems (ITSC), IEEE, 2018, pp.
1063 2575–2582.
- 1064 [26] L.-M. Kieu, D. Ngoduy, N. Malleon, E. Chung, A stochastic schedule-following simulation model of bus routes, *Transportmetrica B:*
1065 *Transport Dynamics* 7 (1) (2019) 1588–1610.
- 1066 [27] Z. Wall, D. Dailey, An algorithm for predicting the arrival time of mass transit vehicles using automatic vehicle location data, Master’s
1067 *thesis*, Citeseer (1998).
- 1068 [28] W.-H. Lin, J. Zeng, Experimental study of real-time bus arrival time prediction with gps data, *Transportation Research Record* 1666 (1)
1069 (1999) 101–109.
- 1070 [29] R. Jeong, R. Rilett, Bus arrival time prediction using artificial neural network model, in: *Proceedings. The 7th international IEEE conference*
1071 *on intelligent transportation systems (IEEE Cat. No. 04TH8749)*, IEEE, 2004, pp. 988–993.
- 1072 [30] Y. Ramakrishna, P. Ramakrishna, V. Lakshmanan, R. Sivanandan, Use of gps probe data and passenger data for prediction of bus transit
1073 *travel time*, in: *Transportation Land Use, Planning, and Air Quality*, 2008, pp. 124–133.
- 1074 [31] A. Shalaby, A. Farhan, Bus travel time prediction model for dynamic operations control and passenger information systems, *Transportation*
1075 *Research Board* 2.
- 1076 [32] J. Patnaik, S. Chien, A. Bladikas, Estimation of bus arrival times using apc data, *Journal of public transportation* 7 (1) (2004) 1–20.
- 1077 [33] S. I.-J. Chien, Y. Ding, C. Wei, Dynamic bus arrival time prediction with artificial neural networks, *Journal of transportation engineering*
1078 128 (5) (2002) 429–438.
- 1079 [34] F. Cathey, D. J. Dailey, A prescription for transit arrival/departure prediction using automatic vehicle location data, *Transportation Research*
1080 *Part C: Emerging Technologies* 11 (3-4) (2003) 241–264.

- 1081 [35] L. Chu, S. Oh, W. Recker, Adaptive kalman filter based freeway travel time estimation, in: 84th TRB Annual Meeting, Washington DC,
1082 2005.
- 1083 [36] B. A. Kumar, L. Vanajakshi, S. Subramanian, Pattern-based bus travel time prediction under heterogeneous traffic conditions, Transportation
1084 Research Record, Transportation Research Board, National Research Council, Washington, DC.
- 1085 [37] M. Altinkaya, M. Zontul, Urban bus arrival time prediction: A review of computational models, International Journal of Recent Technology
1086 and Engineering (IJRTE) 2 (4) (2013) 164–169.
- 1087 [38] U. Mori, A. Mendiburu, M. Álvarez, J. A. Lozano, A review of travel time estimation and forecasting for advanced traveller information
1088 systems, Transportmetrica A: Transport Science 11 (2) (2015) 119–157.
- 1089 [39] M. Yang, C. Chen, L. Wang, X. Yan, L. Zhou, Bus arrival time prediction using support vector machine with genetic algorithm, Neural
1090 Network World 26 (3) (2016) 205.
- 1091 [40] A. Abdi, C. Amrit, A review of travel and arrival-time prediction methods on road networks: classification, challenges and opportunities,
1092 PeerJ Computer Science 7 (2021) e689.
- 1093 [41] M. Soysal, E. G. Schmidt, Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison,
1094 Performance Evaluation 67 (6) (2010) 451–467.
- 1095 [42] K. Ellis, S. Godbole, S. Marshall, G. Lanckriet, J. Staudenmayer, J. Kerr, Identifying active travel behaviors in challenging environments
1096 using gps, accelerometers, and machine learning algorithms, Frontiers in public health 2 (2014) 36.
- 1097 [43] G. Fusco, C. Colombaroni, L. Comelli, N. Isaenko, Short-term traffic predictions on large urban traffic networks: Applications of network-
1098 based machine learning models and dynamic traffic assignment models, in: 2015 International Conference on Models and Technologies for
1099 Intelligent Transportation Systems (MT-ITS), IEEE, 2015, pp. 93–101.
- 1100 [44] H. Nguyen, C. Cai, F. Chen, Automatic classification of traffic incident's severity using machine learning approaches, IET Intelligent
1101 Transport Systems 11 (10) (2017) 615–623.
- 1102 [45] T. Thianniwet, S. Phosaard, W. Pattara-Atikom, Classification of road traffic congestion levels from gps data using a decision tree algorithm
1103 and sliding windows, in: Proceedings of the world congress on engineering, Vol. 1, 2009, pp. 1–3.
- 1104 [46] M. S. B. Othman, S. L. Keoh, G. Tan, Efficient journey planning and congestion prediction through deep learning, in: 2017 International
1105 Smart Cities Conference (ISC2), IEEE, 2017, pp. 1–6.
- 1106 [47] P. Balaji, X. German, D. Srinivasan, Urban traffic signal control using reinforcement learning agents, IET Intelligent Transport Systems
1107 4 (3) (2010) 177–188.
- 1108 [48] S. El-Tantawy, B. Abdulhai, H. Abdelgawad, Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers
1109 (marlin-atsc): methodology and large-scale application on downtown toronto, IEEE transactions on intelligent transportation systems 14 (3)
1110 (2013) 1140–1150.
- 1111 [49] W. Huang, G. Song, H. Hong, K. Xie, Deep architecture for traffic flow prediction: Deep belief networks with multitask learning, IEEE
1112 Transactions on Intelligent Transportation Systems 15 (5) (2014) 2191–2201.
- 1113 [50] M. Fouladgar, M. Parchami, R. Elmasri, A. Ghaderi, Scalable deep traffic flow neural networks for urban traffic congestion prediction, in:
1114 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, 2017, pp. 2251–2258.
- 1115 [51] N. G. Polson, V. O. Sokolov, Deep learning for short-term traffic flow prediction, Transportation Research Part C: Emerging Technologies
1116 79 (2017) 1–17.
- 1117 [52] Z. Zhao, W. Chen, X. Wu, P. C. Chen, J. Liu, Lstm network: a deep learning approach for short-term traffic forecast, IET Intelligent
1118 Transport Systems 11 (2) (2017) 68–75.
- 1119 [53] Y. Yang, A. Heppenstall, A. Turner, A. Comber, Using graph structural information about flows to enhance short-term demand prediction in
1120 bike-sharing systems, Computers, Environment and Urban Systems 83 (2020) 101521.
- 1121 [54] Q. Cheng, Y. Liu, W. Wei, Z. Liu, Analysis and forecasting of the day-to-day travel demand variations for large-scale transportation networks:
1122 A deep learning approach, Transportation Analytics Contest, Tech. Rep.
- 1123 [55] J. Baek and, K. Sohn, Deep-learning architectures to forecast bus ridership at the stop and stop-to-stop levels for dense and crowded bus
1124 networks, Applied Artificial Intelligence 30 (9) (2016) 861–885.
- 1125 [56] J. Ke, H. Zheng, H. Yang, X. M. Chen, Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep
1126 learning approach, Transportation research part C: Emerging technologies 85 (2017) 591–608.
- 1127 [57] L. Liu, R.-C. Chen, A mrt daily passenger flow prediction model with different combinations of influential factors, in: 2017 31st International
1128 Conference on Advanced Information Networking and Applications Workshops (WAINA), IEEE, 2017, pp. 601–605.
- 1129 [58] W. Genders, S. Razavi, Using a deep reinforcement learning agent for traffic signal control, arXiv preprint arXiv:1611.01142.
- 1130 [59] L. Li, Y. Lv, F.-Y. Wang, Traffic signal timing via deep reinforcement learning, IEEE/CAA Journal of Automatica Sinica 3 (3) (2016)
1131 247–254.
- 1132 [60] K. Dwivedi, K. Biswaranjan, A. Sethi, Drowsy driver detection using representation learning, in: 2014 IEEE international advance comput-
1133 ing conference (IACC), IEEE, 2014, pp. 995–999.
- 1134 [61] W. Dong, J. Li, R. Yao, C. Li, T. Yuan, L. Wang, Characterizing driving styles with deep learning, arXiv preprint arXiv:1607.03611.
- 1135 [62] R. Hadsell, A. Erkan, P. Sermanet, M. Scoffier, U. Muller, Y. LeCun, Deep belief net learning in a long-range vision system for autonomous
1136 off-road driving, in: 2008 IEEE/RJSJ International Conference on Intelligent Robots and Systems, IEEE, 2008, pp. 628–633.
- 1137 [63] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, et al., An
1138 empirical evaluation of deep learning on highway driving, arXiv preprint arXiv:1504.01716.
- 1139 [64] J. Pan, X. Dai, X. Xu, Y. Li, A self-learning algorithm for predicting bus arrival time based on historical data model, in: 2012 IEEE 2nd
1140 International Conference on Cloud Computing and Intelligence Systems, Vol. 3, IEEE, 2012, pp. 1112–1116.
- 1141 [65] J. Li, Bus arrival time prediction based on random forest, in: 2017 5th International Conference on Frontiers of Manufacturing Science and
1142 Measuring Technology (FMSMT 2017), Atlantis Press, 2017, pp. 867–872.
- 1143 [66] T. Yin, G. Zhong, J. Zhang, S. He, B. Ran, A prediction model of bus arrival time at stops with multi-routes, Transportation research procedia
1144 25 (2017) 4623–4636.
- 1145 [67] B. Yu, W. H. Lam, M. L. Tam, Bus arrival time prediction at bus stop with multiple routes, Transportation Research Part C: Emerging

- Technologies 19 (6) (2011) 1157–1170.
- [68] N. Shalit, M. Fire, E. Ben-Elia, A supervised machine learning model for imputing missing boarding stops in smart card data, *Public Transport* 15 (2) (2023) 287–319.
- [69] M. Chen, X. Liu, J. Xia, S. I. Chien, A dynamic bus-arrival time prediction model based on apc data, *Computer-Aided Civil and Infrastructure Engineering* 19 (5) (2004) 364–376.
- [70] X. Zhang, Z. Liu, Prediction of bus arrival time based on gps data: Taking no. 6 bus in huangdao district of qingdao city as an example, in: 2019 Chinese Control Conference (CCC), IEEE, 2019, pp. 8789–8794.
- [71] C. Chen, H. Wang, F. Yuan, H. Jia, B. Yao, Bus travel time prediction based on deep belief network with back-propagation, *Neural Computing and Applications* 32 (2020) 10435–10449.
- [72] P. He, G. Jiang, S.-K. Lam, D. Tang, Travel-time prediction of bus journey with multiple bus trips, *IEEE Transactions on Intelligent Transportation Systems* 20 (11) (2018) 4192–4205.
- [73] Z. Junyou, W. Fanyu, W. Shufeng, Application of support vector machine in bus travel time prediction, *Int. J. Syst. Eng* 2 (1) (2018) 21–25.
- [74] B. Yu, H. Wang, W. Shan, B. Yao, Prediction of bus travel time using random forests based on near neighbors, *Computer-Aided Civil and Infrastructure Engineering* 33 (4) (2018) 333–350.
- [75] B. Ashwini, R. Sumathi, H. Sudhira, Bus travel time prediction: a comparative study of linear and non-linear machine learning models, in: *Journal of Physics: Conference Series*, Vol. 2161, IOP Publishing, 2022, p. 012053.
- [76] F. Serin, Y. Alisan, M. Erturkler, Predicting bus travel time using machine learning methods with three-layer architecture, *Measurement* 198 (2022) 111403.
- [77] C. Bai, Z.-R. Peng, Q.-C. Lu, J. Sun, Dynamic bus travel time prediction models on road with multiple bus routes, *Computational intelligence and neuroscience* 2015 (2015) 63–63.
- [78] B. A. Kumar, R. Jairam, S. S. Arkatkar, L. Vanajakshi, Real time bus travel time prediction using k-nn classifier, *Transportation Letters* 11 (7) (2019) 362–372.
- [79] M. Shoman, A. Aboah, Y. Adu-Gyamfi, Deep learning framework for predicting bus delays on multiple routes using heterogenous datasets, *Journal of Big Data Analytics in Transportation* 2 (2020) 275–290.
- [80] J. Wu, Y. Wang, B. Du, Q. Wu, Y. Zhai, J. Shen, L. Zhou, C. Cai, W. Wei, Q. Zhou, The bounds of improvements toward real-time forecast of multi-scenario train delays, *IEEE Transactions on Intelligent Transportation Systems* 23 (3) (2021) 2445–2456.
- [81] M. Behrisch, L. Bieker, J. Erdmann, D. Krajzewicz, Sumo–simulation of urban mobility: an overview, in: *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation, ThinkMind*, 2011.
- [82] J. Hueper, G. Dervisoglu, A. Muralidharan, G. Gomes, R. Horowitz, P. Varaiya, Macroscopic modeling and simulation of freeway traffic flow, *IFAC Proceedings Volumes* 42 (15) (2009) 112–116.
- [83] L. Smith, R. Beckman, K. Baggerly, Transims: Transportation analysis and simulation system, Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States) (1995).
- [84] Q. I. Yang, H. N. Koutsopoulos, A microscopic traffic simulator for evaluation of dynamic traffic management systems, *Transportation Research Part C: Emerging Technologies* 4 (3) (1996) 113–129.
- [85] M. Balmer, M. Rieser, K. Meister, D. Charypar, N. Lefebvre, K. Nagel, Matsim-t: Architecture and simulation times, in: *Multi-agent systems for traffic and transportation engineering*, IGI Global, 2009, pp. 57–78.
- [86] V. Leblond, L. Desbureaux, V. Bielecki, A new agent-based software for designing and optimizing emerging mobility services: Application to city of rennes, in: *European Transport Conference 2020*, 2020.
- [87] M. Fellendorf, P. Vortisch, Validation of the microscopic traffic flow model vissim in different real-world situations, in: *transportation research board 80th annual meeting*, Vol. 11, 2001.
- [88] C. Corporation, TransModeler Traffic Simulation Software, <https://www.caliper.com/transmodeler/default.htm>, [Online; accessed 30-August-2023] (2023).
- [89] H. Shafizadeh-Moghadam, A. Asghari, A. Tayyebi, M. Taleai, Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth, *Computers, Environment and Urban Systems* 64 (2017) 297–308.
- [90] H. Tongal, M. J. Booi, Simulation and forecasting of streamflows using machine learning models coupled with base flow separation, *Journal of hydrology* 564 (2018) 266–282.
- [91] X. Yan, K. Xu, W. Feng, J. Chen, A rapid prediction model of urban flood inundation in a high-risk area coupling machine learning and numerical simulation approaches, *International Journal of Disaster Risk Science* 12 (6) (2021) 903–918.
- [92] S. Chabanet, H. B. El-Haouzi, P. Thomas, Coupling digital simulation and machine learning metamodel through an active learning approach in industry 4.0 context, *Computers in Industry* 133 (2021) 103529.
- [93] M. Shahhosseini, G. Hu, I. Huber, S. V. Archontoulis, Coupling machine learning and crop modeling improves crop yield prediction in the us corn belt, *Scientific reports* 11 (1) (2021) 1–15.
- [94] H. Abdelaty, A. Al-Obaidi, M. Mohamed, H. E. Farag, Machine learning prediction models for battery-electric bus energy consumption in transit, *Transportation Research Part D: Transport and Environment* 96 (2021) 102868.
- [95] M. S. B. Othman, G. Tan, Machine learning aided simulation of public transport utilization, in: 2018 IEEE/ACM 22nd International Symposium on Distributed Simulation and Real Time Applications (DS-RT), IEEE, 2018, pp. 1–2.
- [96] A. Gal, A. Mandelbaum, F. Schnitzler, A. Senderovich, M. Weidlich, Traveling time prediction in scheduled transportation with journey segments, *Information Systems* 64 (2017) 266–280.
- [97] V. Kumar, B. A. Kumar, L. D. Vanajakshi, S. C. Subramanian, Comparison of model based and machine learning approaches for bus arrival time prediction, Tech. rep. (2014).
- [98] C.-T. Lam, B. Ng, S. H. Leong, Prediction of bus arrival time using real-time on-line bus locations, in: 2019 IEEE 19th International Conference on Communication Technology (ICCT), IEEE, 2019, pp. 473–478.
- [99] D. Panovski, T. Zaharia, Real-time public transportation prediction with machine learning algorithms, in: 2020 IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2020, pp. 1–4.
- [100] B. Büchel, F. Corman, Review on statistical modeling of travel time variability for road-based public transport, *Frontiers in Built Environ-*

1211 ment 6 (2020) 70.

1212 [101] Google, General Transit Feed Specification, <https://gtfs.org/>, [Online; accessed 31-August-2023] (2023).

1213 [102] P. Geurts, A. Irrthum, L. Wehenkel, Supervised learning with decision tree-based methods in computational and systems biology, *Molecular*

1214 *Biosystems* 5 (12) (2009) 1593–1605.

1215 [103] M. D. Abkowitz, M. Lepofsky, Implementing headway-based reliability control on transit routes, *Journal of Transportation Engineering*

1216 116 (1) (1990) 49–63.

1217 [104] O. Cats, A. N. Larijani, H. N. Koutsopoulos, W. Burghout, Impacts of holding control strategies on transit performance: Bus simulation

1218 model analysis, *Transportation Research Record* 2216 (1) (2011) 51–58.

1219 [105] G. Liu, S. Wirasinghe, A simulation model of reliable schedule design for a fixed transit route, *Journal of Advanced Transportation* 35 (2)

1220 (2001) 145–174.

1221 [106] U. Vandebona, A. J. Richardson, Effect of checkpoint control strategies in a simulated transit operation, *Transportation Research Part A:*

1222 *General* 20 (6) (1986) 429–436.

1223 [107] J. G. Strathman, K. J. Dueker, T. Kimpel, R. Gerhart, K. Turner, P. Taylor, S. Callas, D. Griffin, J. Hopper, Automated bus dispatching,

1224 operations control, and service reliability: Baseline analysis, *Transportation Research Record* 1666 (1) (1999) 28–36.

1225 [108] X. Chen, B. Hellinga, C. Chang, L. Fu, Optimization of headways with stop-skipping control: a case study of bus rapid transit system,

1226 *Journal of advanced transportation* 49 (3) (2015) 385–401.

1227 [109] K. Gkiotsalitis, O. Cats, At-stop control measures in public transport: Literature review and research agenda, *Transportation Research Part*

1228 *E: Logistics and Transportation Review* 145 (2021) 102176.

1229 [110] Y. Li, J.-M. Rousseau, M. Gendreau, Real-time scheduling on a transit bus route: a 0-1 stochastic programming model.

1230 [111] L. Fu, Q. Liu, P. Calamai, Real-time optimization model for dynamic scheduling of transit operations, *Transportation research record*

1231 1857 (1) (2003) 48–55.

1232 [112] Z. Liu, Y. Yan, X. Qu, Y. Zhang, Bus stop-skipping scheme with random travel time, *Transportation Research Part C: Emerging Technologies*

1233 35 (2013) 46–56.

1234 [113] J. Rodriguez, H. N. Koutsopoulos, S. Wang, J. Zhao, Cooperative bus holding and stop-skipping: A deep reinforcement learning framework,

1235 *Transportation Research Part C: Emerging Technologies* 155 (2023) 104308.

1236 [114] S. A. Arhin, E. C. Noel, O. Dairo, Bus stop on-time arrival performance and criteria in a dense urban area, *International Journal of Traffic*

1237 *and Transportation Engineering* 3 (6) (2014) 233–238.

1238 [115] M. Chen, X. Liu, E. Bachman, Analyze bus delay severity using ordered probit model.

1239 [116] Pe. Liashchynskiy, Pa. Liashchynskiy, Grid search, random search, genetic algorithm: a big comparison for NAS.

1240 [117] Z. Xiong, Y. Cui, Z. Liu, Y. Zhao, M. Hu, J. Hu, Evaluating explorative prediction power of machine learning algorithms for materials

1241 discovery using k-fold forward cross-validation.