AUTOMATED IDENTIFICATION OF DELAY-GENERATING LOCATIONS FOR BUS PRIORITY INTERVENTION PLANNING

M. Scháno[*], *J. Nový*[†], *D. Schmidt*[*]

Abstract: Public transport systems face increasing pressure to improve reliability under constrained infrastructure and limited investment capacity. Small-scale bus priority interventions represent a cost-effective tool for improving operational performance, yet their implementation requires reliable identification and prioritisation of delay-generating locations. This paper presents a fully automated method for spatial identification and quantification of delay formation in bus transport systems based exclusively on high-resolution AVL data. The proposed approach reconstructs vehicle trajectories using a high spatial resolution vectorized road network and map-matching, enabling continuous delay estimation along the entire route rather than only at stops. Referential travel times are derived empirically from historical data using a percentile-based approach, allowing delay quantification independently of static timetables and accommodating heterogeneous operating conditions. The method supports aggregation across multiple trips, lines, and corridors, providing a system-wide view of delay accumulation and its infrastructure-related causes. The methodology is demonstrated on regional bus services. An experimental evaluation of machine learning models as substitutes for referential journeys indicates that, given the available data structure, AI-based approaches fail to achieve meaningful predictive performance and cannot reliably replace the proposed statistical reference. The presented method offers a scalable and robust decision-support tool for prioritizing bus priority interventions and improving public transport reliability using operational data already available to most transport authorities.

Key words: *public transport reliability, bus delay analysis, bus priority interventions, automated vehicle location (AVL), map-matching, decision support systems*

Received: July 22, 2025
Revised and accepted: October 23, 2025

DOI: 10.14311/NNW.2025.35.006

[*]Martin Scháno – Corresponding author; Drahomír Schmidt; Czech Technical University in Prague, Faculty of Transportation Sciences, Konviktsk textrm 'a 20, 110 00 Prague 1, Czech Republic, E-mail: schano@fd.cvut.cz, drahomir.schmidt@cvut.cz

[†]Josef Nový; Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Trojanova 13, 120 01 Prague 2, Czech Republic, E-mail: josef.novy@cvut.cz

©CTU FTS 2025

1. Introduction

Transport is an integral component of a rapidly changing human society. For the majority of the population, it offers broad and seemingly unlimited mobility options. However, many cities today are characterized by the dominance of private cars which, by their sheer numbers, collectively constrain mobility, reduce travel speeds and safety, and increase air pollution and noise.

In particular, the problem of limited road infrastructure capacity is usually not solvable, primarily due to existing urban layouts. It is evident that a balanced and long-term financially, technically, and environmentally sustainable transport system requires a symbiosis of private transport, public transport, cycling, and walking.

Unfortunately, the global trend in public transport usage has been declining. [20] report partially positive statistics suggesting a stabilization of public transport modal share at approximately 11% in Western Europe and 16% in Eastern Europe.

Public transport has also been significantly affected in recent years by the COVID-19 pandemic, which led to a critical loss of passengers. [14] reported a decline of up to 80% in public transport ridership in Germany shortly after the pandemic. [3] later presented an overview of passenger recovery trends. By 2024, some countries managed to approach pre-pandemic ridership levels. Germany reached 94% and the United Kingdom 90% while others have been recovering more slowly. For example, Canada reached 83% and the United States only 77%.

Based on current developments, it is evident that public transport, if it is to remain competitive, requires targeted and systematic support. The literature clearly shows that reliability is a critical factor of perceived public transport service quality from the passengers' perspective and that it significantly influences their satisfaction and, indirectly, their loyalty ([2, 8, 11, 17, 19, 24], among others).

Reliability can be effectively improved through the implementation of relatively small-scale priority measures. However, the question of where and how to implement such measures in larger systems has been the subject of many studies whose results are not always transferable. Limited transferability can be illustrated, for example, by [25], who identified systematic accumulation of small delays over time as a key issue in Wroclaw (Poland), while [22] found that the vast majority of delays in Portland (USA) are concentrated upstream of major intersections.

Several authors (e.g., [15, 22, 27]) highlight the lack of studies utilizing high-frequency GPS data for comprehensive and aggregated evaluation of multiple routes (e.g., lines or trips) sharing the same corridor. Such analyses are essential for fully automated evaluation of heterogeneous public transport systems.

This paper presents one such method – specifically, a fully automated approach for monitoring public transport operations with the aim of identifying locations with the highest potential for implementing priority measures and providing a data-driven basis for their prioritization.

2. Literature Review

This paper focuses on bus priority interventions (BPI), i.e., small-scale interventions in street geometry, stop placement, traffic signal plans, and similar measures that support public transport (often at the expense of private car traffic) or minimize interactions between buses and other traffic. Multiple types of BPI can be combined along a single corridor, each targeting a different operational issue. Their limited spatial extent and relatively low costs make them practically implementable measures.

This focus is reflected in the analytical methods used in the literature. Some studies concentrate on inter-stop segments (e.g., [18, 21, 26]), others on bus stops themselves (e.g., [9]), and some specifically on signalized intersections (e.g., [16]).

Alternatively, several authors have examined specific priority measures, such as the implementation of bus lanes [5] or dynamic signal control strategies [7].

Most existing systems are based on AVL (automated vehicle location) and APC (automated passenger counting) data. However, many approaches rely on aggregation or clustering techniques that often fail to achieve the spatial resolution required for precise localization of delay sources.

For example, [6] presented clustering of delay profiles in the Hague (NL) and explicitly acknowledged that limited spatial resolution is a weakness of their approach, making it more suitable for macro-level analyses, identification of problematic routes, or inter-stop segments rather than precise localization. Similar approaches were used by [25], who applied hierarchical agglomerative clustering in Wroclaw (PL) to derive representative delay evolution profiles for individual clusters, and by [13], who clustered hours of the day based on shared characteristics in Istanbul (TR) to identify stops influencing average speeds at different times.

[4] and [1] proposed network-wide analyses with higher spatial resolution at the level of individual city blocks. This was achieved by subdividing standard inter-stop segments at major intersections, thereby increasing granularity for identifying problematic locations.

High-frequency GPS data In recent years, several authors have demonstrated the use of so-called high-frequency GPS data (sampling intervals $\leq 5\,\mathrm{s}$), which enable much more detailed analysis of vehicle movement ([12, 16, 23, 28], among others).

High-frequency GPS data were used, for example, by [23]. Their method for identifying delay sources on urban arterials is based on the analysis of average speeds in the vicinity of predefined points of interest and on visualization tools illustrating speed trends across a large number of trips. This approach allowed to identify congestion locations and speed variability and directly relate them to traffic elements such as intersections, pedestrian crossings, or bus stops.

AVL data were recorded at a fixed 5-second interval, which the authors leveraged by analysing the density of GPS points. Higher point density corresponded to lower speeds, while lower density indicated smoother traffic flow.

Although this method provides very detailed spatial analysis, it also has several limitations. Lower average speed does not necessarily imply the occurrence of delay. A bus running slightly ahead of schedule may intentionally reduce speed to improve passenger comfort and avoid early departure from the next stop. Furthermore, the method requires a fixed GPS sampling frequency, which may not be guaranteed, particularly in rural areas.

Systems based on map-matching The use of high-frequency GPS data from multiple vehicles often requires a shared reference network to which individual records can be related. Combined with map-matching methods, this approach can mitigate common urban effects such as noise, signal reflections, and positional inaccuracies [10].

An interesting direction aimed at increasing spatial resolution is proposed by [1] and [15]. On inter-stop segments, they define free-flow speed (i.e., an ideal vehicle speed unaffected by any negative factors), which allows detection of deviations from this idealized movement with significantly greater detail.

The potential of combining these approaches was demonstrated by [15] using data from the Massachusetts Bay area (USA). His method enables high-resolution speed analysis and classification of bus delays.

Delay on an inter-stop segment was defined as the difference between actual travel time and free-flow travel time, defined as the 5th percentile of travel time outside peak periods. This approach enables precise quantification of operational delays; however, it also requires strict consistency in timetable planning throughout the day. This limits applicability in systems employing adaptive timetable strategies, such as extended running times during peak periods or scheduled dwell time adjustments to stabilize delays. In other words, the method assumes a static system and is therefore not fully compatible with the dynamic planning practices used by many modern public transport operators.

As this review illustrates, existing methods provide highly detailed insights into the operational reality of public transport. However, their application is often limited to partial analytical studies or short-term experimental deployments.

The remaining challenge is therefore the transformation of these methods into universal tools capable of supporting strategic decision-making.

3. Method

The design of an automated system for the spatial identification and quantification of the impacts of individual locations on delay formation, together with its demonstration, was conducted using operational data from regional bus transport in the Ústí nad Labem Region, Czech Republic (hereafter referred to as DÚK).

The method is structured logically according to the system design process, i.e., definition of requirements, data preparation, map-matching, and definition and classification of delays.

3.1 Requirements and Assumptions

Based on the intended purpose of the system and constraints arising from its application to regional bus transport, the following requirements were defined:

- the ability to operate with a variable network of routes and trips, primarily due to request stops generating possible detours from the main route;

- the ability to handle variability in travel times between individual stops, allowing for the inclusion of schedule recovery time or inclusion of extended travel times, as some delay-generating factors (e.g., traffic intensity) are consistent throughout the day and cannot be considered permanent sources of delay, but rather implicit characteristics of the operating environment;

- the ability to cope with inconsistent signal availability (both GNSS and data transmission), as certain locations reduce the bus's ability to report its position due to limited capability to generate or to transmit current data;
- the ability to localize the origin of delays with high spatial resolution in order to avoid false-positive identification, for example, in situations where two intersections are in close proximity, and the generated delay could be assigned incorrectly, thereby degrading the economic evaluation of proposed measures;
- the ability to quantify generated delays for potential conversion into monetary units and possible integration with APC systems;
- the ability to perform joint analysis across different routes and trips, including sections with overlapping routes, in order to adequately evaluate the impacts of individual locations on all buses within the public transport network;
- the ability to perform spatial and temporal analyses to enable the identification of potential additional factor interactions (e.g., linkage to weather data);
- functionality based exclusively on AVL data.

3.2 Data Source

For the system design, data on the current position and operational status of the DÚK vehicle fleet were used, collected directly from the public transport dispatching system. These data represent comprehensive, continuous, and centrally managed information on all vehicles in operation.

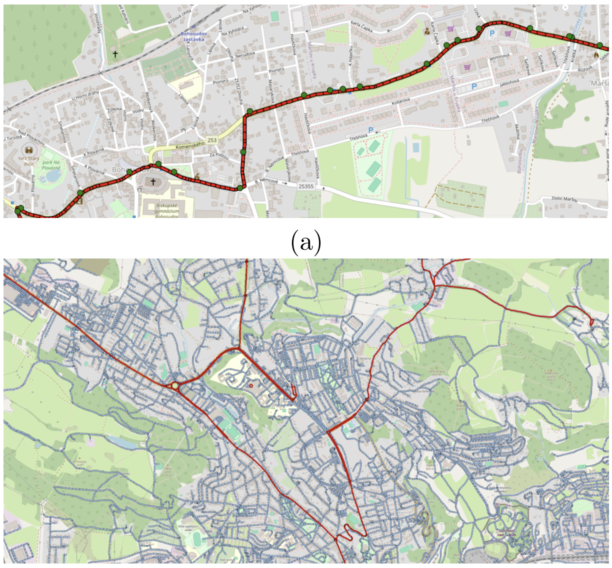
The data were obtained through an API enabling data transfer from the DÚK system to third-party applications. The interface is provided via the TCP protocol, specifically using HTTP/HTTPS at the application level.

Data queries are issued every 5 seconds. However, this frequency cannot be directly translated into a one-to-one temporal spacing between individual position records. It is necessary to account for server load, the ability of buses to detect and transmit their positions at a given location, and validation mechanisms that check data integrity and detect errors. A valid new position record for an average bus is received approximately once every 15–20 seconds.

3.3 Vector Map and Map-Matching

To reconstruct the spatiotemporal trajectories of individual buses, background vector maps (shown in Fig. 1(a)) were used to which individual position coordinates were map-matched. This approach was chosen because of the variable frequency of position detection, to increase the accuracy of derived variables such as segment speed, and to enable aggregation of multiple bus trajectories across the same road network segments.

Since individual map segments (vectors) also inherit performance indicators of buses, the length of these vectors determines the spatial resolution of the analysis results. Within the proposed system, the maximum length of individual vectors was set to 30 meters. On a line approximately 40 km long, this corresponds to roughly 3,000 vectors per direction (with an average vector length of 10–15 meters), enabling sufficiently high spatial resolution analysis.



(a)

(b)

Fig. 1 *Route reconstruction using the map-matching method (a) and regional reference vector map with highlighted vectors of monitored routes (b).*

To enable meaningful analysis of the entire public transport system, it was necessary to generate a unified representation of the entire DÚK network by aggregating data across all routes. This was achieved using a two-level vector map structure (see Fig. 1(b)). The upper level contains all road infrastructure in the region, while lower levels are created separately for each route as subsets of the regional reference map. This structure allows individual route maps to be easily edited, added, or removed, for example, in response to route changes, without requiring the exclusion of previously collected data from the analysis.

The process of map-matching is also designed as a two-level procedure.

First, all unique travel directions in the vicinity of the bus are detected, taking into account position detection uncertainty. Subsequently, for each direction, the closest bus position relative to the vectors in that travel direction is identified. These positions are considered potential vehicle locations, and each enters a pairing process with the subsequent detected valid position of the same vehicle. The most probable position pair is then selected as the actual vehicle position.

This approach ensures that the correct travel segment is identified even in complex parts of the network. Without this procedure, situations such as stopping at an intersection where the route branches could result in incorrect assignment to the wrong branch due to positional uncertainty, leading to incorrect route identification and incorrect calculation of subsequent KPIs.

The output of this step is a list of pairs of bus positions (start and end) and parameters of the selected route between these positions (distance travelled, time, segment speed, etc.). The fully reconstructed trip then serves as the input for subsequent analyses.

For each travel direction, the vector map is enriched with a range of attributes providing additional contextual information, simplifying search processes and enabling more detailed classification of generated delays. These attributes include, among others, the presence and identification of bus stops, pedestrian crossings, intersections (including their influence on bus movement, such as traffic signals or yield requirements), railway crossings, and other monitored infrastructure elements.

3.4 Definition and Classification of Delays

To analyse bus movements from the perspective of delay generation, it was necessary to extend the conventional definition of delay considering only temporal deviations at stops relative to scheduled times. To define continuous delay at any point along the entire route, a procedure similar to that proposed by [15] was adopted. He computed reference times using so-called free-flow speed, derived from off-peak runs unaffected by external influences (typically nighttime trips).

This study proposes an alternative approach for generating reference times using the 5th percentile (lowest) of travel times through each vector over the previous six months, calculated separately for each trip of each route. This ensures that each trip has its own optimal travel profile, and that each analysed bus is compared against a near-best achievable travel time for the given time of day.

Each bus is represented by a spatiotemporal trajectory, which is compared at every traversed vector with the corresponding reference time. Two quantities are recorded: (1) the absolute delay value, defined as the difference between the actual time and its reference time, and (2) the change in delay on the vector, defined as the difference in absolute delay between the current vector and the preceding one. Both are computed in three variants: (a) for the full set of buses, (b) for the 50% of the most delayed buses, and (c) for the 25% most delayed buses.

For each increase in delay, the number of stops affected by that delay is also recorded. An illustrative example is shown in Fig. 2 below. This enables a clear distinction between delays that are reduced again within the same inter-stop segment and delays that persist over a longer distance, thus affecting a greater number of passengers. This can be observed in Fig. 2 by comparing the different colours from yellow affecting only 1 bus stop to red affecting 10 bus stops and all passengers getting off at those stops. Using this approach, it is possible to calculate the number of affected stops or, when combined with an APC system, the number of affected passengers for each individual delay increase.

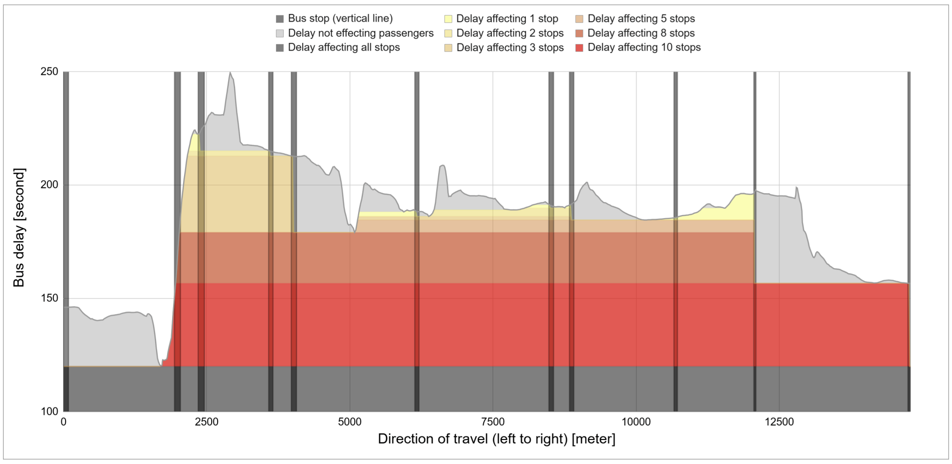


Fig. 2 *Illustrative example of delay propagation along a trip segment; vertical grey lines represent bus stops; yellow-red palette indicates the number of affected stops.*

This method provides high spatial resolution and has strong potential for integration with APC systems (or historical manual passenger count data).

In the final identification of candidate delay-causing locations for corrective measures, it is necessary not only to consider indicators such as generated delay multiplied by the number of affected stops (or passengers), but also to focus on locations that exhibit higher delay increases for already delayed buses. This allows the identification of potentially lower-cost priority interventions.

3.5 Visualization of Results

The visualization of results introduces several additional challenges, particularly related to readability and interpretability.

It is not practical to visualize delay increases using a linear scale, as the majority of map vectors incur almost zero delay, whereas a very small number of vectors account for most delay increases (typically less than 1% of vectors, predominantly in the vicinity of stops or intersections). Low delay increases, which are nevertheless important to visualize because of their cumulative effect, would be suppressed or completely lost when using a linear scale.

For this reason, a logarithmic transformation was applied for data visualization, allowing extremely small delays (on the order of seconds) to be displayed on the same scale as very large delays (often hundreds of seconds).

To further improve clarity, a counterintuitive yet effective reduction of spatial resolution was introduced for visualization purposes. This is motivated by frequent cases where long route sections are mostly unproblematic, except for one or two short vectors exhibiting extreme delay increases. When visualizing the entire route (i.e., thousands of vectors), such isolated problematic vectors are difficult to detect, and their actual impact on delay formation is hard to interpret visually.

To address this issue, a moving average approach was applied, maintaining a constant physical window size (50 meters) for visualization. This limits the minimum displayed unit without introducing undesirable distortion due to the logarithmic scale. As a result, for example, a 30-second delay over a 20-meter segment is visualized with a very similar colour to a 15-second delay over two consecutive 10-meter segments. This ensures a more balanced distribution of colours and line widths and significantly improves interpretability at a higher-level view, without compromising the underlying analytical resolution.

All automated analytical functions are, of course, performed on the original data with higher spatial resolution.

3.6 Using AI as a Substitute for Referential Journey

The proposed method is fully functional; however, it requires a relatively extensive amount of historical data to reliably construct and optimize the referential journey. This requirement may impose limitations in situations such as timetable changes or partial or complete realignments of a bus line's route.

For this reason, an alternative approach based on AI methods was investigated. The idea is to train models on existing lines and trips using sufficient historical data and subsequently transfer the learned patterns to new routes until an adequate database for constructing a referential journey is available.

In this framework, the models estimate either the expected traversal time of individual segments or the expected delay. By monitoring deviations from these estimates (residuals), an equivalent reference trajectory is obtained that automatically adapts to changes in operating conditions.

Several model families were tested, including XGBoost, MLP, CNN, LSTM, GRU, and transformer-based architectures. Each model family exhibits different inductive biases and data requirements:

1. **XGBoost** (gradient boosted trees) – A strong reference baseline for tabular data. It naturally accommodates heterogeneous features (categorical encodings, temporal features, aggregated statistics). Sequential information must be incorporated through feature engineering (e.g., previous delays, rolling statistics over preceding segments, traffic state indicators). The model offers fast training, good interpretability (e.g., via SHAP values), and robustness to missing data.

2. **MLP** (multilayer perceptrons) – Suitable when well-prepared static and derived features are available. Without specialized architectures, MLPs struggle to capture long sequential dependencies; however, they perform well as lightweight inference models or as components of ensemble systems. In this study, MLPs served as the baseline model.

3. **CNN** (1D convolutions over sequences of segments) – Efficient in capturing local patterns and computationally lightweight. Particularly suitable when short-range local dependencies dominate system behaviour.

4. **LSTM/GRU** (recurrent neural networks) – A natural choice for modelling long-range dependencies across multiple segments. GRUs offer a simpler architecture and are often sufficient. Careful handling of masking and padding

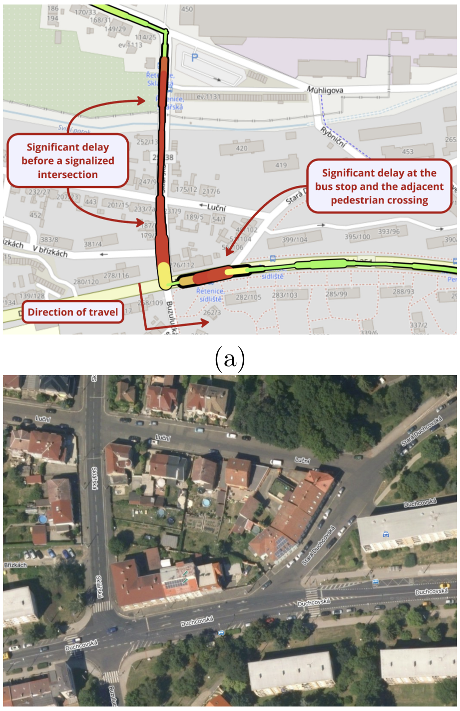
(due to variable-length journeys), normalization, and regularization (e.g., dropout) is required to prevent model drift.

5. **Transformers** (attention-based models) – Enable modelling of global dependencies across entire journeys. They handle masking (e.g., missing segments) and contextual information effectively but require larger datasets and careful design of positional or relative encoding.

4. Results

The proposed system enables visualization of individual trips separately as well as their aggregation for arbitrary combinations, ranging from single routes to entire lines and up to the whole system. This section demonstrates the potential of the presented method to identify specific elements of the transport network that contribute to the accumulation of delay changes and thus indicate the need for bus priority interventions.

Fig. 3 below presents an illustrative example of a signal-controlled intersection of Duchcovská, Sklářská, and Buzulucká streets in the city of Teplice. Buses approach this intersection from a minor approach. During the morning peak hours, a queue



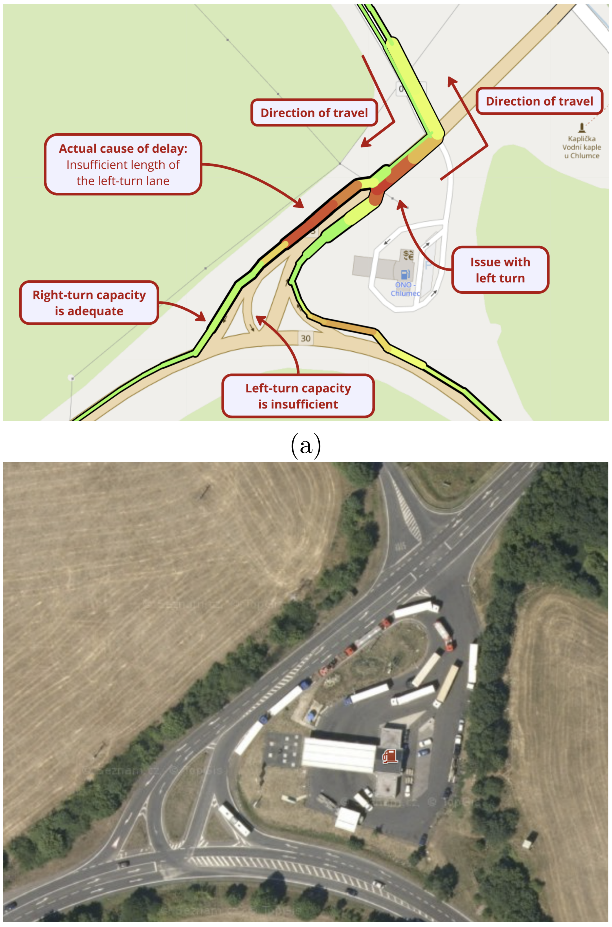
(a)

(b)

Fig. 3 *Visualization of bus passage through a signal-controlled intersection for trip 107 on line 488 during the morning peak period.*

of up to 250 meters regularly forms at this location. In addition to this approach-related delay, independent delay generation can also be observed in the vicinity of a nearby bus stop and a pedestrian crossing on Duchcovská Street.

Fig. 4 below shows another example of a complex intersection generating significant delay increases. This case concerns the intersection of roads I/13 and I/30 near the town of Chlumec. The specific issue at this intersection lies in the insufficient length of the left-turn lane from road I/13 (northeastern approach) towards road I/30 (southeastern approach). The queue of vehicles turning left blocks buses intending to turn right. As a result, bus delays concentrate near the end of the turning lane rather than at the right-turn movement onto road I/13 (southwestern approach).



(a)

(b)

Fig. 4 *Visualization of bus passage through an intersection for trip 106 on line 488 during the morning peak period.*

The figure also illustrates a problem in the opposite travel direction, specifically during a left turn from road I/13 onto a minor road (a smaller intersection located in the upper right part of the Fig. 4).

Fig. 5 presents the morning trip of line 452 passing through the city center of Ústí nad Labem (travel direction from left to right), highlighting several problematic locations (listed in the direction of travel): Divadlo bus stop, Hlavní nádraží bus stop, the intersection of Velká Hradební and Hrnčířská streets, and the signalized intersection of Předmostí and Přístavní streets.

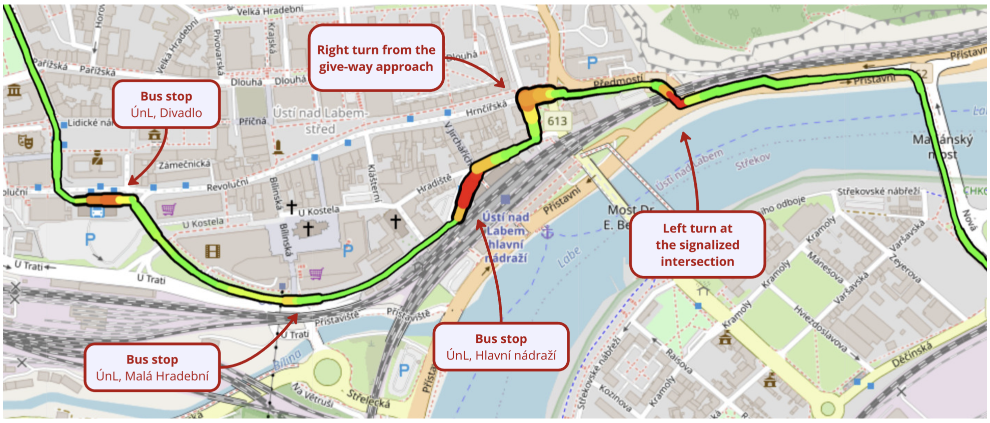


Fig. 5 *Visualization of the passage through the city center of Ústí nad Labem for trip 104 on line 452 during the morning peak period.*

When analysing a larger number of trips and lines, their visualization within a single map is straightforward. However, temporal information is partially lost when multiple lines traverse the same road segment. This does not necessarily pose a problem for analysis, as illustrated by Fig. 6 below, which aggregates delay-related

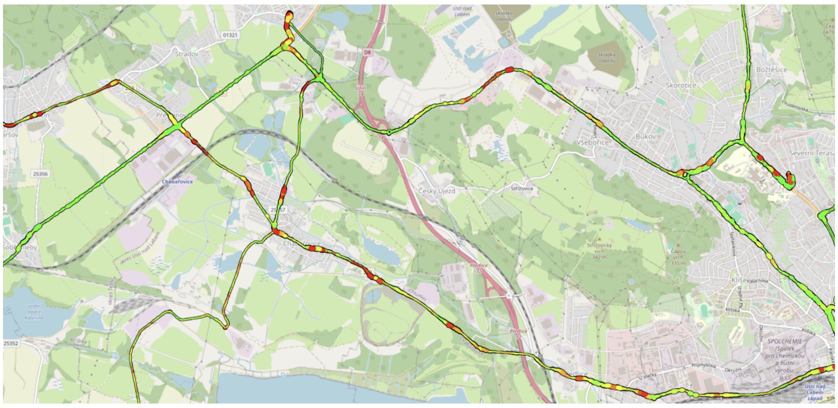


Fig. 6 *Visualization of parameters for lines 450, 452, 453, 454, 458, and 488 during the morning peak period.*

parameters for all trips operating during the morning peak period (approximately 6:00–8:00) across six lines in the Ústí nad Labem area.

4.1 Statistical Overview

In addition to map-based outputs, the data can also be visualized using charts referenced to the cumulative distance along individual lines. Figs. 7 and 8 below present two excerpts illustrating the delay increases on two representative bus lines.

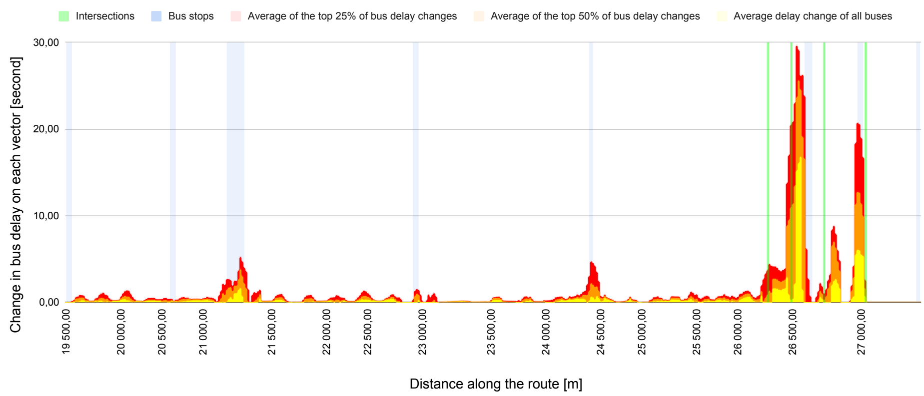


Fig. 7 *Visualization of trip 129 on line 454 during the morning peak period.*

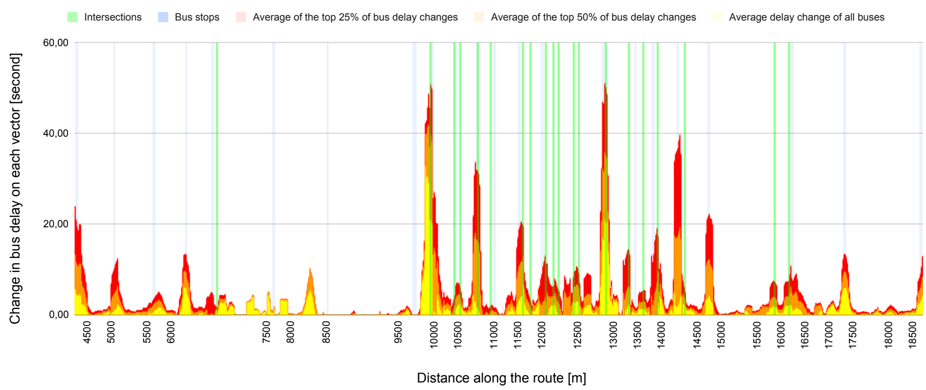


Fig. 8 *Visualization of parameters for trip 107 on line 488 during the period 2023-09 (passage time between 06:19 and 08:08).*

Vertical lines represent individual monitored elements; for the presented lines, these include bus stops shown in blue and intersections shown in green. No railway crossings or other monitored elements are present on the selected segments.

Delay change values are color-coded as follows: yellow represents average values across all buses, orange corresponds to the top 50% of highest delay increases, and red indicates the top 25% of highest delay increases. This representation makes it possible to distinguish whether specific locations generate small delays affecting a large number of buses or, conversely, sporadically generate very high delays affecting only a small subset of vehicles. Such differentiation is useful for understanding the underlying mechanisms of the observed phenomena.

As shown in Fig. 7, the majority of generated delays can be attributed to the monitored elements. These include moderate delays at the 3rd and 5th bus stops, followed by a significant delay occurring around kilometer 26.5, which corresponds to the approach to a shopping center via a signal-controlled intersection.

Fig. 8 presents statistics for a trip with multiple problematic locations. Even in this case, most delays can be observed in the vicinity of monitored elements. One noteworthy phenomenon appears around kilometer 8.25, where an increase in delay occurs outside of the monitored elements (Fig. 9(a)). This corresponds to the identification of frequent stopping outside designated bus stops. A substantial delay increase can be observed around kilometer 10 at a signal-controlled intersection, which was previously presented in Fig. 3. Another significant delay occurs around kilometer 13 near the Teplice Main Railway Station bus stop, the details of which are shown in Fig. 9(b).

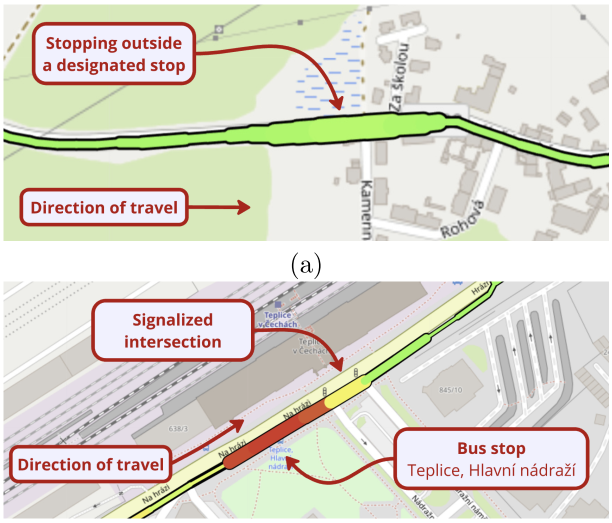


Fig. 9 *Identification of bus stopping outside a designated stop at km 8.25 (a) and visualization of a problematic bus stop at km 13 (b).*

Aggregated analyses across all lines further confirmed the stability of identified delay patterns, but detailed system-level statistics are omitted here for brevity.

When examining aggregated values of DÚK system by individual categories, it is evident that the majority of delays originate either at bus stops (accounting for 43–46% with some lines reaching 67%) or at minor at-grade intersections (accounting for 35–37%). The rest is at signal-controlled intersections (approximately 9.5% of all delays and strongly dependent on the line route).

Notably, across this relatively large sample of bus trips, almost no delays were generated outside the monitored locations (i.e., outside bus stops, intersections, and railway crossings). Only between 2.9% and 4.2% of all delays

4.2 Comparison of Alternative Methods to Substitute for Referential Journey

As described above in Sec. 3.6. Using AI as a substitute for a referential journey, the amount of historical data required to construct a statistically derived referential journey may represent a limitation for the practical deployment.

The objective of this section was therefore to verify whether machine learning methods can be used to replace the statistically constructed referential journey.

The target variable was defined as the change in delay within a given route segment. The models were trained on historical trips and evaluated on a separate validation set, which was split by routes to prevent data leakage between training and validation (e.g., the same trip occurring at different times).

The primary evaluation metric was the coefficient of determination R^2:

$$R^2 = 1 - \frac{\sum_i (y_i - \widehat{y_i})^2}{\sum_i (y_i - \overline{y})^2}, \tag{1}$$

where y_i is the observed value, $\widehat{y_i}$ is the model prediction, and $\overline{y}$ is the mean of the target variable in the validation set.

An R^2 value of 0 corresponds to a trivial baseline model that always predicts the mean, while negative values indicate performance worse than this baseline.

The optimization loop (Optuna) minimized the validation loss (mean squared error). After each run, the corresponding R^2 value was additionally computed to improve interpretability.

For each model architecture (XGBoost, MLP, CNN, LSTM, GRU, Transformer), a separate Optuna study was conducted. Within each study, combinations of hyperparameters were randomly sampled from predefined ranges, including the number of layers, hidden representation sizes, dropout rates, window size, learning rate, weight decay, batch size, gradient clipping, and model-specific parameters (e.g., XGBoost parameters). Each combination defined one trial consisting of:

1. generation of a specific hyperparameter configuration,
2. training of the model on the training dataset,
3. evaluation of the validation loss and computation of R^2,
4. logging of results in Optuna and optional pruning of underperforming trials.

Approximately 1,200 trials were executed for each architecture. The resulting score for each trial, as well as the optimization trajectory, can be seen in Fig. 10.

Optuna also computed hyperparameter importance (feature importance in the hyperparameter space), providing an estimate of how strongly individual hyperparameters influenced the final validation loss.

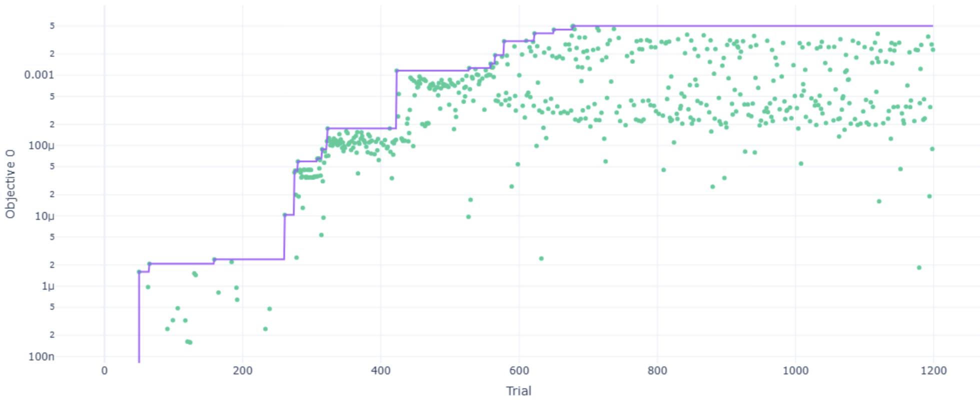


Fig. 10 *Optimization trajectory of the hyperparameter tuning process for the tested models.*

The importance of individual hyperparameters can be effectively illustrated using importance plots, which reveal several common trends across models. For architectures based on sequential models (LSTM, GRU, Transformer, and partially also MLP and CNN), the most important hyperparameter was almost always num_layers, i.e., the number of layers. For LSTM models, their relative importance reached approximately 0.4, while for GRU and Transformer models it ranged between 0.2 and 0.4.

The second key parameter was grad_clip, i.e., the gradient clipping threshold, particularly for LSTM and GRU models. This confirms that training these models is prone to exploding gradients and that optimization stability has a substantial impact on model performance. The window_size parameter (length of the temporal window or sequence) also played a significant role: windows that were too short failed to capture sufficient context, while excessively long windows reduced training stability and increased memory requirements. Across all neural models, sensitivity to learning_rate and weight_decay (L2 regularization) was also observed, although their relative importance was lower than that of the number of layers and gradient clipping.

For convolutional neural networks (CNNs), the dominant hyperparameter was cnn_layers, i.e., the number of convolutional blocks, with a relative importance of approximately 0.38. Additional influential parameters included base_channels, defining the width of convolutional layers, as well as learning_rate and weight_decay. This indicates that for CNNs, model capacity—represented by the number of layers and channels—is critical. Models that are too small fail to capture sufficiently complex patterns, while excessively large models tend to suffer from poorer generalization and reduced training stability.

For the tree-based XGBoost method, the most important hyperparameter was window_size, which determines how large a portion of the trip is included in a single sample. Other key parameters included xgb_max_depth, xgb_learning_rate, and xgb_subsample, reflecting the well-known sensitivity of XGBoost to tree depth, learning rate, and subsampling rate. These parameters strongly influence whether the model can adequately capture data structure without overfitting. The overall score for all tested architectures can be seen in Fig. 11.

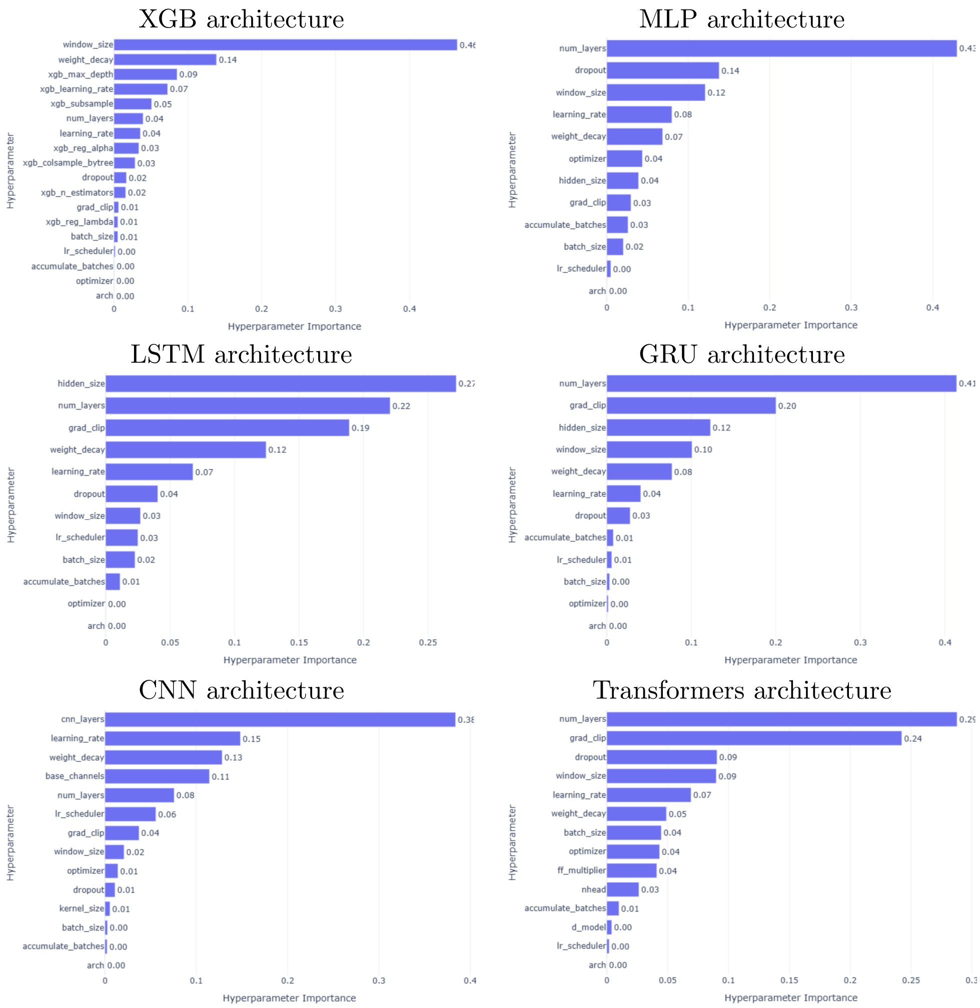


Fig. 11 *The importance of individual hyperparameters for each model architecture.*

Achieved results of alternative methods The achieved results indicate that none of the tested architectures reached a clearly positive value of the coefficient of determination R^2 on the validation dataset. The best performance was achieved by XGBoost, with $R^2 \approx 0.0051$, i.e., only marginally better than the trivial baseline corresponding to predicting the mean value of the target variable. Sequential neural models GRU, LSTM, and Transformer reached their best R^2 values in the range of 10^{-4} to 10^{-5} (GRU ≈ 0.00053, LSTM ≈ 0.00007, Transformer ≈ 0.00007) and thus were practically unable to capture a substantial portion of the variability of the target variable.

Even worse results were obtained by the MLP and CNN models, which exhibited extremely negative R^2 values (CNN $\approx -8,101$; MLP $\approx -293,892$). Such strongly negative R^2 values indicate that even the best-found configuration performed significantly worse than the trivial mean-based model. These points either to a systematic issue in model setup (e.g., inappropriate scaling of the target variable, an unsuitable loss function, or an error in the inference pipeline) or to a strong tendency toward divergent training (predictions drift far outside a realistic range). Regarding practical applicability, none of the tested models achieved a level of performance that would allow the referential journey to be reliably replaced.

Evaluation of the use of AI The results suggest that, given the current form of the data and their representation, the signal explaining delay formation at individual route segments is very weak compared to noise and unobserved factors. One of the main reasons is likely the absence of key explanatory variables such as real-time traffic conditions, incidents and sudden road closures, current weather conditions, vehicle occupancy, operational dispatch interventions, or driver identification. These factors may account for a substantial portion of delay variability, yet they are not reflected in the available data at all.

The situation is further complicated by the heterogeneity of routes, which a single shared model attempts to cover across different lines, operating regimes (peak vs. off-peak, weekdays vs. weekends), seasonal effects, and segment-specific characteristics. This significantly hinders the learning of stable and transferable patterns. In the case of models with extremely negative R^2 values, an additional contributing factor may be a combination of aggressive training dynamics (e.g., excessively high learning rates or insufficient gradient clipping) and inappropriate scaling or transformation of the target variable.

The hyperparameter analysis further indicates that the models are highly sensitive to architectural settings, particularly the number of layers and the length of the input sequence. Despite systematic exploration of the hyperparameter space using optimization techniques, no substantial improvement in performance was achieved. This suggests that the limiting factor is not merely the choice of architecture or its parameters, but primarily the quality and informational content of the input data.

The experiment involving intensive hyperparameter tuning with Optuna demonstrated that the tested models (XGBoost, MLP, CNN, LSTM, GRU, Transformer) are not capable of achieving useful predictive accuracy under the current setup. The best-performing model, XGBoost, improves upon the trivial mean-based baseline only by a few per mille, which is negligible from a practical standpoint. More complex neural architectures additionally exhibit highly unstable behaviour and, without the inclusion of more informative input variables and more careful control of the training process, provide no tangible benefit for this task.

It therefore follows that, in their current form, artificial intelligence methods are not capable of replacing the referential journey. Their realistic application at this stage is rather supplementary. They may theoretically serve, for example, as support for simpler statistical models (e.g., average segment-level delays over a defined period) or as tools for detecting coarse anomalies.

5. Discussion

This article demonstrated the potential of the proposed method to simply and efficiently identify problematic locations and to provide adequate evidence for their prioritized mitigation. The reader may have noticed that, in contrast to some of the systems discussed in the literature review, trajectory smoothing techniques were not applied. Given that the identification of problematic locations is primarily based on aggregating data from dozens of vehicle passages and from broader areas of influence of the monitored road network elements, an adequate level of high spatial resolution was nevertheless achieved even without smoothing individual trajectories, and with a significantly simpler algorithm.

It should be noted, however, that the potential for further improvement remains. Trajectory smoothing can be applied both in the distance–time domain and in the spatial domain, considering the vertical and directional characteristics of the vehicle route based on the underlying vector map. Such an approach offers the potential to substantially increase spatial resolution and to enable evaluation at the level of individual trips. At the same time, it is necessary to consider the variability in position detection accuracy at certain specific locations within the network.

With this limitation in mind, one of the next steps in the development of the platform will be the interpretation of position detection frequency. It must be acknowledged that fluctuations in position detection frequency are not spatially random and tend to cluster at specific locations, which affects the local validity of the results. Future inclusion of spatial trajectory smoothing is expected to significantly mitigate this issue.

In terms of quantifying impacts, the presented system currently operates at the level of delay per vehicle. This limitation arises because automatic passenger counting (APC) is not deployed system-wide within the DÚK network, and thus the system cannot yet be linked to universally available occupancy data. However, historical data from manual passenger counts are available for selected demonstration lines and can serve as a substitute for a network-wide APC system.

The system can also be integrated with a range of additional data sources. A particularly promising extension would be the attribution of delays to individual bus drivers, which would allow for a comprehensive assessment of individual performance and the identification of a potentially small subset of drivers disproportionately contributing to higher levels of delay.

The experimental application of machine learning tools proved unsuccessful for the purpose of replacing the referential journey. As a result, the method continues to rely on a robust historical database. From the authors' perspective, this does not constitute a significant barrier to deployment, as priority interventions should not be implemented on segments with recent changes. Waiting for sufficient data naturally coincides with the period required for traffic conditions to stabilize.

Alternatively, the method may be applied to segments with a more limited database, while progressively monitoring improvements in the accuracy of the results over time. Given the stability observed in most transport systems of this scale, even a smaller historical data set (e.g., 3 or 2 months instead of the 6 months used here) is expected to function adequately.

6. Conclusion

This paper introduced a fully automated method for the spatial identification and quantification of delay formation in bus transport systems, designed to support the prioritisation of bus priority interventions. The approach operates exclusively on AVL data and is applicable to heterogeneous regional bus networks.

By combining a vectorized road network, map-matching, and empirically derived referential travel times, the method enables continuous delay estimation along the entire route rather than only at stops. This allows both precise localization of delay-generating infrastructure elements and aggregation of their impacts across multiple trips, lines, and corridors. The results demonstrate that the proposed system consistently identifies critical locations, such as stops and intersections, that account for a substantial share of accumulated delays.

An experimental evaluation of machine learning models as substitutes for statistically derived referential journeys showed that none of the tested architectures achieved sufficient predictive performance. These findings indicate that, given the current data structure, delay variability is largely driven by unobserved operational factors and system heterogeneity, making reference-based statistical approaches more robust and reliable for this task.

Overall, the proposed method provides a practical, scalable, and data-efficient tool for transforming high-resolution operational data into actionable insights for strategic planning and prioritisation of public transport interventions.

Acknowledgement

Parts of the manuscript (translation, proofreading, and linguistic refinement) were supported by the use of AI-based tools, specifically OpenAI's ChatGPT. All substantive content decisions remain the responsibility of the authors.

References

[1] AEMMER Z., RANJBARI A., MACKENZIE D. Measurement and classification of transit delays using GTFS-RT data. *Public Transport*, 2022, 14, pp. 263–285, doi: 10.1007/s12469-022-00291-7.

[2] BÖRJESSON M., RUBENSSON I. Satisfaction with crowding and other attributes in public transport. *Transport Policy*, 2019, 79, pp. 213-222, doi: 10.1016/j.tranpol.2019.05.010.

[3] BUEHLER R., PUCHER J., WHITE P., CURRIE G. Public transport and the COVID-19 pandemic: A comparative analysis of trends and policies in Great Britain, Germany, the USA, Canada, and Australia. Transportation Research Part A: Policy and Practice, 2025, 199, 104549, doi: 10.1016/j.tra.2025.104549.

[4] CAROS N. Leveraging spatial relationships and visualization to improve public transit performance analysis. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2021, available online: https://hdl.handle.net/1721.1/139338.

[5] CHEN Q. An Optimization Model for the Selection of Bus-Only Lanes in a City. PloS one 10, e0133951, 2015, doi: 10.1371/journal.pone.0133951.

[6] CHENG Y., KRISHNAKUMARI P. Unsupervised Learning for Public Transport Delay Pattern Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2024, 2678, pp. 358–373, doi: 10.1177/03611981231215333.

[7] CHOW A.H.F., LI S., ZHONG R. Multi-objective optimal control formulations for bus service reliability with traffic signals. Transportation Research Part B: Methodological, 2017, 103, pp. 248-268, doi: 10.1016/j.trb.2017.02.006.

[8] DE OÑA J. DE, OÑA R., EBOLI L., MAZZULLA G. Perceived service quality in bus transit service: A structural equation approach. *Transport Policy*, 2013, 29, pp. 219-226, doi: 10.1016/j.tranpol.2013.07.001.

[9] EL-GENEIDY A., HORNING J., KRIZEK K. Analyzing Transit Service Reliability Using Detailed Data From Automatic Vehicular Locator Systems. *Journal of Advanced Transportation*, 2011, 45, pp. 66–79, doi: 10.1002/atr.134.

[10] ELHAJJ M., OCHIENG W. Impact of New GPS Signals on Positioning Accuracy for Urban Bus Operations. *Journal of Navigation*, 2020, 73, pp. 1284–1305, doi: 10.1017/S0373463320000272.

[11] FELLESSON M., FRIMAN M. Perceived Satisfaction with Public Transport Service in Nine European Cities. *Journal of the Transportation Research Forum*, 2012, 47, doi: 10.5399/osu/jtrf.47.3.2126.

[12] FIGLIOZZI M.A., STOLL N.B. A Study of Bus High-Resolution GPS Speed Data Accuracy. *Transportation Research Record: Journal of the Transportation Research Board*, 2018, 2672(8), pp. 187–198, doi: 10.1177/0361198118793273.

[13] YAŞAR I., CETINEL Y., ARISOY A.A. Analysis of the Effect of Time of Day on Bus Speeds near a Bus Stop using GPS Data. 13th International Congress on Advances in Civil Engineering, 2018. Available from: https://www.researchgate.net/publication/332441115_Analysis_of_the_Effect_of_Time_of_Day_on_Bus_Speeds_near_a_Bus_Stop_using_GPS_Data

[14] HELFERS A., REISERER M., SCHNEIDER N., EBERSBACH M., SOMMER C. Should I Stay or Should I Go? Risk Perception and Use of Local Public Transport During the COVID-19 Pandemic. *Frontiers in Psychology*, 2022, 13, 926539, doi: 10.3389/fpsyg.2022.926539.

[15] HUANG Y. Understanding Bus Operations Using High-Resolution Vehicle Location Data. Master Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2023, available online: https://hdl.handle.net/1721.1/152466.

[16] MARTIN I., GREEN T. Using high resolution transit geolocation data to analyze performance and identify intersections that contribute to transit delay, 2019. Available from: https://westernite.org/annualmeetings/19_Monterey/Presentations/6A/6A-Green-Paper.pdf

[17] MOUWEN A. Drivers of customer satisfaction with public transport services. Transportation Research Part A: Policy and Practice, 2015, 78, pp. 1–20, doi: 10.1016/j.tra.2015.05.005.

[18] OSMAN O., RAKHA H., MITTAL A. Application of Long Short Term Memory Networks for Long- and Short-term Bus Travel Time Prediction, preprint, 2021, doi: 10.20944/preprints202104.0269.v1.

[19] PETROVIĆ D., IVANOVIC I., JOVIĆ J., DJORIC V. Customer perception of service quality in public transport. *Transport*, 2014, 29, doi: 10.3846/16484142.2014.951685.

[20] POLIAK M., POLIAKOVÁ A., MRNIKOVA M., SIMURKOVA P., JASKIEWICZ M., JURECKI R. The Competitiveness of Public Transport. *Journal of Competitiveness*, 2017, 9, pp. 81–97, doi: 10.7441/joc.2017.03.06.

[21] QIU B., FAN W. Machine Learning Based Short-Term Travel Time Prediction: Numerical Results and Comparative Analyses. *Sustainability*, 2021, 13, 7454, doi: 10.3390/su13137454.

[22] REDELMEIER P., VICTORIANO-HABIT R., CRUMLEY M., EL-GENEIDY A. Bit by bit: A method for using bus data to develop plan bus priority interventions in Portland, Oregon, USA. *Journal of Public Transportation*, 2025, 27, 100135, doi: 10.1016/j.jpubtr.2025.100135.

[23] STOLL N.B., GLICK T., FIGLIOZZI M.A. Using High-Resolution Bus GPS Data to Visualize and Identify Congestion Hot Spots in Urban Arterials. *Transportation Research Record: Journal of the Transportation Research Board*, 2016, 2539, pp. 20–29, doi: 10.3141/2539-03.

[24] SUSILO Y.O., CATS O. Exploring key determinants of travel satisfaction for multi-modal trips by different traveler groups. Transportation Research Part A: Policy and Practice, 2014, 67, pp. 366–380, doi: 10.1016/j.tra.2014.08.002.

[25] SZYMAŃSKI P., ŻOŁNIERUK M., OLESZCZYK P., GISTEREK I., KAJDANOWICZ T. Spatio-temporal profiling of public transport delays based on large scale vehicle positioning data from GPS in Wroclaw. IEEE Transactions on Intelligent Transportation Systems PP, 2017, doi: 10.1109/tits.2018.2852845.

[26] THAKKAR S., SHARMA S., ADVANI C., ARKATKAR S., BHASKAR A. Comparative analysis of travel time prediction algorithms for urban arterials using Wi-Fi Sensor Data. International conference on communication systems & networks (COMSNETS), Bangalore, India, 2021, pp. 697–702, doi: 10.1109/COMSNETS51098.2021.9352845.

[27] ZHANG H., LIU Y., SHI B., JIA J., WANG W., ZHAO X. Analysis of Spatial-Temporal Characteristics of Operations in Public Transport Networks Based on Multisource Data. *Journal of Advanced Transportation*, 2021, pp. 1–15, doi: 10.1155/2021/6937228.

[28] ZOU Q., XIONG W., WANG X., QIN F. Research on Real-Time Anomaly Detection Method of Bus Trajectory Based on Flink. *Electronics*, 2023, 12, 3897, doi: 10.3390/electronics12183897.