TRANSPORT PROBLEMS 2026 Volume 21 Issue 1
PROBLEMY TRANSPORTU DOI: 10.20858/tp.2026.21.1.01

Keywords: public transport; arrival delay; probabilistic analysis; stop clusters

Viktor MARTYANOV[1]*, Andrzej CZEREPICKI[2], Maciej KOZŁOWSKI[3], Weronika ZAKLIKA[4]

PROBABILISTIC-STATISTICAL ANALYSIS OF THE CORRESPONDENCE BETWEEN SCHEDULED AND REAL-TIME ARRIVAL TIMES OF PUBLIC TRANSPORT BUSES

Summary. Although reliable schedule adherence is pivotal for service quality and passenger planning in urban transit, realized arrivals often deviate from timetables in non-Gaussian and time-dependent ways. This study provides a probabilistic–statistical assessment of how actual arrivals accord with schedules, thereby addressing the analytical gap of stop-level comparisons across time of day and weekday strata. Using one month of real-time GNSS records matched to static GTFS for Warsaw bus line 112, we analyzed 8.6 million validated stop events. Delay distributions were characterized by robust statistics (median, interquartile range, skewness, kurtosis) and normality diagnostics (Lilliefors, Anderson–Darling, Jarque––Bera), revealing heavy tails up to ±20 min and invalidating Gaussian assumptions. Non-parametric tests (Kruskal–Wallis; Dunn–Šidák) showed statistically significant but practically small differences in median delay across seven time-of-day bands and weekdays (largest shift ≈ 1.2 min; typical deviations within ±1 min). We then clustered stop-level delay profiles (medians by time band) with K-means to uncover four interpretable punctuality archetypes (punctual, evening catch up, pre-schedule variability, and afternoon peak bottleneck), mapped along both travel directions and examined by weekday. We used these profiles to link delay patterns to infrastructure and operations (e.g., initial segments, intersections, recurring congestion) and furnish actionable outputs: targeted buffer allocation, identification of segments for control, and interpretable features to enhance learning-based arrival time prediction. Although demonstrated on a single line, the methodology is general and transferable to other networks with comparable GNSS and GTFS data.

1. INTRODUCTION

Urban public-transport systems are vital components of city infrastructure, underpinning mobility and spatial accessibility while reducing private vehicle usage, traffic congestion, road-traffic incidents, and environmental impacts [1-4]. By offering safe, efficient, and sustainable mobility to diverse social groups, public transport supports broader objectives of urban development and social equity [5]. Despite growing demand and recognized societal benefits, private car travel often remains the preferred mode in many regions, prompting authorities to introduce policies to improve the public-transport experience and foster pro-transit attitudes among residents [6].

[1] Warsaw University of Technology, Faculty of Transport; Koszykowa 75, 00-662 Warsaw, Poland; e-mail: viktor.martyanov@pw.edu.pl; orcid.org/0009-0000-6538-1690
[2] Warsaw University of Technology, Faculty of Transport; Koszykowa 75, 00-662 Warsaw, Poland; e-mail: andrzej.czerepicki@pw.edu.pl; orcid.org/0000-0002-8659-5695
[3] Warsaw University of Technology, Faculty of Transport; Koszykowa 75, 00-662 Warsaw, Poland; e-mail: maciej.kozlowski@pw.edu.pl; orcid.org/0000-0002-1068-8991
[4] Łukasiewicz Research Network - Automotive Industry Institute (Łukasiewicz - PIMOT); Jagiellońska 55, 03-301 Warsaw, Poland; e-mail: weronika.zaklika@pimot.lukasiewicz.gov.pl; orcid.org/0000-0002-8397-9263
* Corresponding author. E-mail: viktor.martyanov@pw.edu.pl

Punctuality–the degree to which vehicle arrivals adhere to published timetables–is a critical indicator of service quality and reliability in public transport [5, 7]. Schedule adherence directly affects reliability and passengers' trip planning; missed or uncertain connections translate into tangible costs for users and operators. Consequently, agencies require robust, stop-level diagnostics that quantify where and when systematic deviations from the timetable occur, rather than aggregate averages. However, scheduled arrival times frequently diverge from reality owing to the multifaceted dynamics of urban traffic–including factors like street congestion, stochastic weather, passenger-boarding delays, and infrastructural constraints [8]–which produce irregular, asymmetric, and heavy-tailed delay distributions [9]. As a result, classical approaches predicated on Gaussian assumptions lose robustness, necessitating methods that accommodate non-normality, heteroscedasticity, and outliers.

In the literature, considerable effort has been devoted to predicting bus arrival times online and offline. Online models leverage real-time Global Navigation Satellite System (GNSS) feeds, traffic feeds, and passenger-load information to deliver dynamic estimates for operators and users [10, 11]. Offline approaches, in contrast, analyze historical datasets to optimize route planning and timetable design; typical methods include queueing-theoretic models [12], Kalman-filter frameworks [13], Gaussian-process regression [14], and deep learning architectures capable of capturing spatiotemporal dependencies [15]. More recently, attention has been paid to quantifying prediction uncertainty, with studies demonstrating that delay distributions exhibit skewness and heavy tails [16, 17], rendering mean-based forecasts insufficient and motivating the development of alternatives such as quantile regression, Student's distributions, and Bayesian hierarchical models [18-22]. Beyond prediction, service reliability has been examined at the stop, route, and network levels and is shaped by infrastructure and control policies [23]. Operational strategies such as holding or short-turning can materially affect adherence and passenger costs [24, 25]. Szymański et al. [26] analyzed 15 million delay reports and applied clustering to stop-pair delay-change dynamics, focusing on inter-stop delay propagation rather than stop-level punctuality patterns. There are examples of integrating schedule and GNSS trajectory data [27], but their approach emphasized reliability metrics and visualization, not clustering at the stop level. Similarly, Pałys et al. [28] used geopositioning data for the machine-learning-based prediction of Warsaw buses, but they did not study punctuality clustering.

Despite these advances, three key gaps remain. First, few studies have offered a comprehensive, probabilistic comparison of actual versus scheduled arrival times at individual stops and daily intervals. Second, spatially resolved analyses that identify route segments or stops with distinct delay-behavior profiles are scarce. Third, the potential operational value of clustering stop-level delay profiles for targeted timetable adjustment or real-time control has not been fully explored.

This paper addresses these gaps by presenting a probabilistic-statistical framework for assessing the compliance of real-world bus arrival times with General Transit Feed Specification (GTFS) schedules. Our main contributions are as follows:

- We constructed a rigorously matched, stop-level dataset for a representative urban bus line by aligning real-time GNSS observations with frequently revised GTFS schedules, yielding 8.6 million validated events over one month; the pipeline is general and applicable wherever GNSS and GTFS are available.
- We performed a detailed descriptive analysis of delay distributions, employing robust measures (median, interquartile range, skewness, kurtosis) and normality diagnostics (Lilliefors, Anderson–Darling, Jarque–Bera) to characterize heavy-tailed behavior.
- We conducted non-parametric hypothesis tests (Kruskal–Wallis, Dunn's post-hoc with Šidák correction) to compare median delays across weekdays and seven daily intervals, thereby quantifying statistically significant yet practically modest temporal effects.
- We introduced a *K*-means clustering of stop-level delay profiles (vectors of median delays over the day) to uncover four distinct punctuality patterns and map their spatial distribution along the route.
- We extended the cluster analysis to incorporate weekday variability and discuss the operational implications for timetable resilience, dynamic recovery strategies, and predictive-model enhancement.

The outputs complement learning-based arrival-time models by providing interpretable stop-type labels and weekday–time features that can be used directly as inputs to improve forecast accuracy, while offering planners explicit guidance on where and when systematic deviations arise.

The remainder of this article is organized as follows. Section 2 describes data acquisition, preprocessing, and the analytical methods. Section 3 presents the results of descriptive statistics, non-parametric tests, and clustering. Section 4 discusses the findings in the context of existing literature and operational practice. Finally, Section 5 presents the conclusions and suggests directions for future research.

2. MATERIALS AND METHODS

2.1. Materials

The probabilistic–statistical analysis of actual bus arrival times against the published timetable requires two distinct datasets: one for static schedules and one for real-time arrival data.

Schedule datasets are made available by the operator via a static Uniform Resource Locator [29] as compressed text files in the GTFS format [30], a de facto standard for storing and exchanging public transport timetables. Each GTFS release becomes effective on publication and remains valid until the next update. In May 2024, the operator issued 12 GTFS releases, with validity periods ranging from 1–4 days between 2024-05-01 and 2024-05-31. Incorporating these revisions is essential to ensure correct matching between realized and scheduled arrivals.

The GTFS feed comprises three principal entities: lines, stops, and trips. These entities are linked via scheduled stop times. A simplified hierarchy of these objects is shown in Fig. 1.

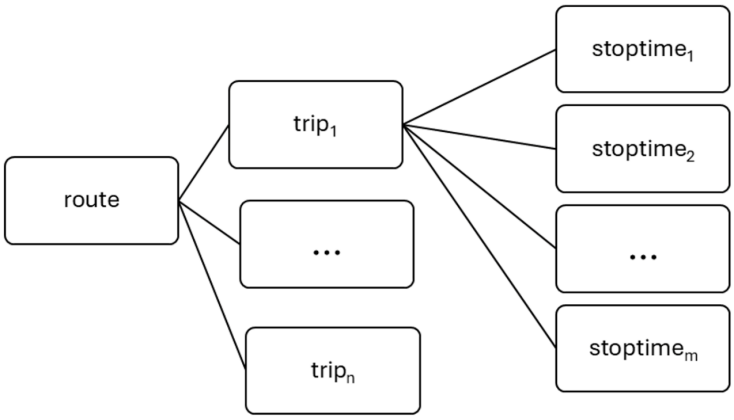

Fig. 1. Simplified structure of static GTFS timetable data

Real-time vehicle-movement data are recorded on board each bus by an on-vehicle recorder, which samples sensors (GNSS position, vehicle time, route identifier) at a frequency of 0.1 Hz and buffers them locally. Each sampled packet is transmitted to the operator's server and archived; the latest packet is also exposed via a public Web Application Programming Interface (API) (31) accessible with a registered token. As no historical real-time datasets are available through the public API, we implemented a proprietary system to download and store all incoming vehicle-position records. The software was developed in C# on the Microsoft .NET Framework 4.8 platform. A Microsoft SQL Server 2019 Express Edition database stores actual vehicle positions, while static schedules are stored in the graph database Neo4j Community Edition (32). Statistical analysis was performed using Python 3.12. The overall data flow is shown in Fig. 2.

Independent studies have reported discrepancies between Intelligent Transport Systems (ITS) reported and ground-truth times due to GNSS error and telemetry latency; this motivated the strict

spatiotemporal filtering described later to mitigate such artifacts [33]. Table 1 illustrates the format of the data frame that the public transport operator provided via the API interface.

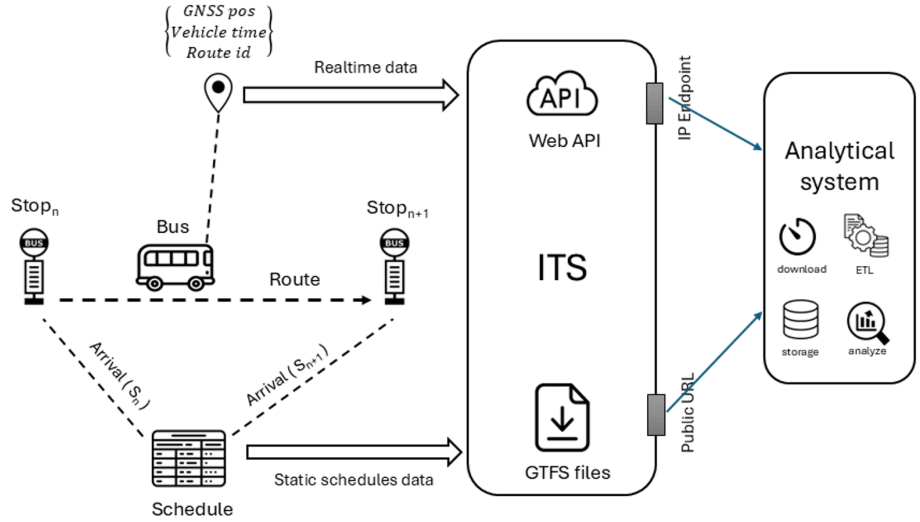

Fig. 2. Architecture of the data collection and processing system

Table 1

Format of the data frame provided by the public transport operator

Field	Description	Data type
line	service line identifier	String
lat	vehicle position (latitude)	Float
lon	vehicle position (longitude)	Float
vehicleNumber	vehicle number	String
brigade	service shift identifier	String
time	vehicle-on-board timestamp	Time

For the present study, we selected May 2024 data for Warsaw bus line 112, a representative c. 24-km route featuring residential streets, dedicated bus lanes, high-speed corridors, bridges, and variable traffic intensities (including peak loads at suburban exits before and after weekends). The line serves 45 stops in one direction and 43 in the other, operating bi-directionally and continuously.

2.2. Data preprocessing

To compute arrival delays at individual stops, we applied the following steps:

- Partition real-time records into daily files for each date in May 2024.
- Extract the set of unique fleet numbers (vehicle identifiers) operating line 112.
- Within each day and for each vehicle, sort GNSS samples in ascending order of timestamp.
- For every scheduled half-trip (i.e., a single traversal in one direction), match the sequence of real-time GNSS points to GTFS stop_times and compute the actual arrival time at each stop.
- Combine scheduled and matched real-time data into a synthetic output dataset organized by trip.

Trip assignment and stop matching used the GTFS calendar and calendar_dates to select the correct schedule version for each observation date. An ‘arrival’ at a stop was registered when the GNSS position entered a 50-m proximity of the stop centroid and the stop sequence advanced consistently along the route; this prevented false detections in stationary congestion. All timestamps were reconciled to local time using a single time base.

The structure of the resulting dataset is summarized in Table 2. Each row comprises scheduled attributes, real-time measurements, or derived variables. Note: delay_min is defined as (vehicle_time − schedule_time) in minutes (positive = late, negative = early).

Table 2

Schema of the pre-processed dataset

Variable	Description	Source	Data type
date	date of operation	scheduled	Date
hour	hours of operation	real-time	Integer
day_of_week	day of the week	scheduled	Integer
route	bus line number	scheduled	String
vehicle	fleet number	real-time	String
stop_index	sequential index of the stop along the route	scheduled	Integer
stop_id	stop identifier as per the operator's system	scheduled	String
stop_name	stop name in the public transport network	scheduled	String
stop_lat	stop latitude	scheduled	Float
stop_lon	stop longitude	scheduled	Float
stop_distance	The distance between the recorded vehicle position at dwell and the stop's coordinates	scheduled	Float
schedule_time	scheduled arrival time at the stop	scheduled	Time
vehicle_time	actual arrival time at the stop	real-time	Time
delay_min	difference between vehicle time and schedule time expressed in minutes (positive = delay, negative = early)	derived	Float
vehicle_brigade	service shift identifier	real-time	String
trip_id	GTFS trip identifier	scheduled	String
trip_headsign	trip destination headsign	scheduled	String
direction	direction of the trip	scheduled	String

2.3. Data cleaning and filtering

The matched dataset was further cleaned to remove erroneous or incomplete records. Specifically, we excluded:

- trips with implausible time overlaps or clear misassignments between scheduled and actual runs.
- records lacking at least one recorded dwell event at a stop (i.e., incomplete stop time sequences).
- runs in which any arrival deviated by more than Δmax minutes from the schedule, indicative of extraordinary incidents (heavy congestion, vehicle breakdowns, route diversions, etc.).
- entries with missing values.
- stop events for which the recorded stopping position was located more than 50 m from the stop coordinates.

Thresholds were selected to control known ITS artefacts (telemetry latency, urban GNSS error) while preserving typical operations. The 50-m proximity reflects conservative dwell detection in dense urban areas. Δmax excludes extraordinary incidents rather than routine variability; sensitivity checks confirmed that key summary statistics (medians, the interquartile range (IQR)) were stable to reasonable threshold changes. After filtering, 8,600,181 of 9,537,389 recorded events were retained. The principal reasons for exclusion were missing values and out-of-proximity detections; misassignments were rare. The resulting dataset formed the basis for all subsequent analyses.

2.4. Methods

Extraction of descriptive statistics.

Let $\Delta = t_{vehicle} - t_{schedule}$ denote the deviation between actual and scheduled arrival times (in minutes). We computed the following descriptive measures, stratified by day of week and seven daily intervals:

- arithmetic mean of $\bar{\Delta}$
- median of $\tilde{\Delta}$
- standard deviation s
- skewness γ_1
- kurtosis γ_2

Visual diagnostics comprised histograms and quantile-quantile (Q–Q) plots against the standard normal. We tested for normality using the Lilliefors, Anderson–Darling, and Jarque–Bera procedures, all three of which rejected Gaussianity. Heavy tails are further indicated by excess kurtosis (4.66 > 3) and nonlinear departures in the Q–Q tails. Given the considerable sample size ($10^6..10^7$ observations), the standard normality tests possess very high power. Thus, even slight deviations were statistically significant. The combination of tests ensures sensitivity to departures in the center (Jarque–Bera) and in the tails (Anderson–Darling, Lilliefors).

Non-parametric dependency analysis.

The dataset was stratified by weekday (Monday–Sunday) and by seven time-of-day bands defined as 04-06, 07-09, 10-12, 13-15, 16-18, 19-21, and 22-23. To assess whether central tendencies differed across categories, we applied the non-parametric Kruskal-Wallis H test (robust to non-normality and heteroscedasticity), which yielded $p < 0.001$. As Kruskal-Wallis tests rank distribution differences, we also reported median shifts (Hodges–Lehmann estimates) to quantify practical significance; the most significant shift across time-of-day bands was ≈ 1.2 min. Pairwise comparisons were based on Dunn's test with Šidák correction to control the family-wise error rate.

Clustering of stop-delay profiles.

Each of the 95 stops was characterized by a delay-profile vector,

$$d = \left(\tilde{\Delta}_1, \tilde{\Delta}_2, \tilde{\Delta}_3, \cdots, \tilde{\Delta}_7\right), \quad (1)$$

where $\tilde{\Delta}_i$ denotes the median delay in time-of-day band i. To emphasize the temporal shape rather than the absolute level, we centered each profile by subtracting its median, yielding d'. We then applied the K-means algorithm (Euclidean distance) with $K = 4$ to these centered profiles. K was chosen for interpretability and was consistent with elbow/silhouette diagnostics. The four archetypes are referred to consistently as "punctual", "evening catch up", "pre-schedule variability," and "afternoon peak bottleneck." To convey spatial patterns, for each weekday, we presented the four cluster centroids and a continuous coloured bar aligned with the route sequence, with each stop colored according to its cluster membership. These cluster labels and weekday–time medians can subsequently be incorporated into learning-based arrival time models, improving predictive accuracy while retaining interpretability.

3. RESULTS

3.1. Data acquisition, preprocessing, and formal data verification

The measurement campaign yielded a total of 9,537,389 recorded events. We then enforced strict quality-control criteria according to which only records indicating that the bus was dwelling within the designated stop area were retained. A dwell event was registered when the GNSS-derived vehicle position entered a 50-m proximity of the stop centroid and the stop sequence advanced consistently along the route, thus avoiding false detections in stationary congestion. Furthermore, all records containing missing values or manifest registration errors (e.g., loss of connectivity) were removed. After filtering, 8,600,181 stop events remained for subsequent analyses.

Threshold choices reflect the need to mitigate known ITS artefacts (telemetry latency, urban GNSS error) while preserving typical operations. Sensitivity checks showed that key summaries (medians, IQR) were stable under reasonable variations of the proximity threshold (±20 m) and Δmax (percentile-based cut-offs).

3.2. Descriptive statistics

Fig. 3a shows the estimated probability density function of Δ. Fig. 3b displays a Q–Q plot against the standard normal.

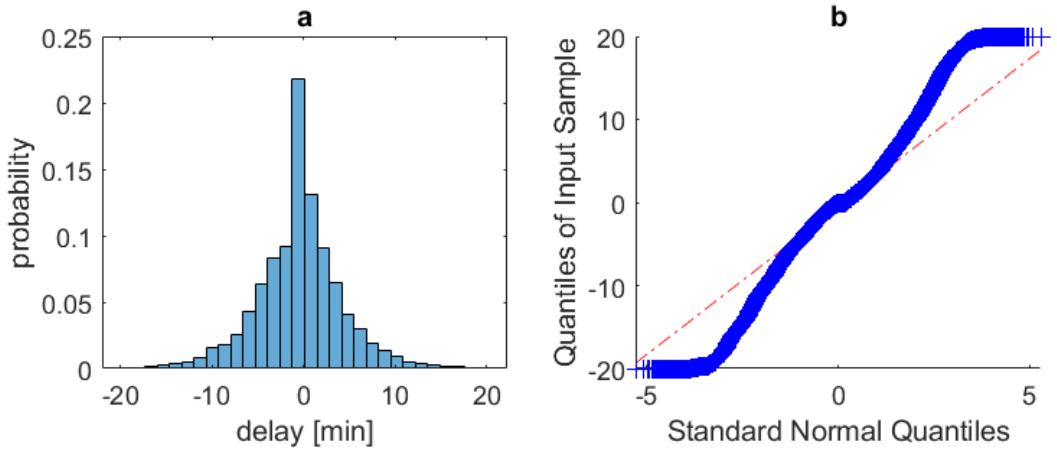

Fig. 3. Features of the probability distribution of the arrival-delay random variable: a) a histogram of the distribution and b) a Q–Q plot against the standard normal

Summary statistics (Table 4) indicate several salient properties. The negative mean (-0.42 min) shows a slight tendency to arrive early on average, while the median of ~0.0 min suggests a balanced split between early and late arrivals. The standard deviation (4.67 min) implies that ≈68% of observations lie within ±4.67 min and ≈95% lie within ±9.34 min of the schedule. Skewness close to zero (-0.04) indicates near symmetry, whereas excess kurtosis (4.66 > 3) signals heavy tails with more extreme early/late arrivals approaching the ±20 min bounds. A visual inspection of Fig. 3b corroborates this outcome, as the central quantiles track the reference line, with pronounced departures in both tails.

Given the large sample (≈8.6 million observations), classical normality tests are overly powerful, meaning that even minor departures from Gaussianity are flagged as significant. Accordingly, subsequent group comparisons rely on non-parametric procedures (Kruskal–Wallis with Dunn–Šidák pairwise tests) that are robust to heavy tails and variance heterogeneity.

To aid interpretation, we report practical, distribution-based summaries alongside several significance tests: medians, interquartile ranges, and, where relevant, Hodges–Lehmann estimates of median differences with confidence intervals.

Table 3

Descriptive statistics of the Δ variable

Parameter	Value [min]
Arithmetic mean $\bar{\Delta}$	-0.42
Median $\tilde{\Delta}$	0.00
Standard deviation s	4.67
Minimum/maximum	[-20.00, 20.00]
Skewness γ_1	-0.04
Kurtosis γ_2	4.66

3.3. Analysis of delays by time of day and day of week

We categorized time of day into seven bands (04-06, 07-09, 10-12, 13-15, 16-18, 19-21, 22-23), and, for each, we computed the number of observations, the median delay, the first (Q1) and third quartiles (Q3), and the interquartile range (IQR = Q3 − Q1). These summaries are presented (in minutes) in Table 4. Box plots for each band are shown in Fig. 4.

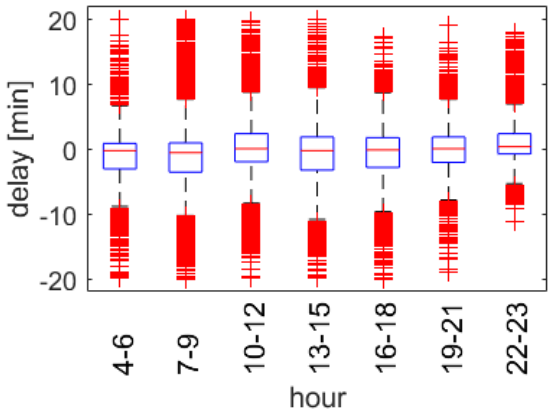

Fig. 4. Box plots of arrival delays (minutes) by time-of-day interval

Table 4

Classification of time-of-day intervals and delay-distribution characteristics

Time interval	Hours	No. of records	Q1, [s]	Median, [s]	Q3, [s]	IQR, [s]
Early morning	4–6	1.0197e+06	-2.05	-0.4	1.65	4.1
Morning peak	7–9	1.5632e+06	-3.00	-0.6	1.80	4.8
Mid-morning	10–12	1.4168e+06	-2.45	0.0	2.45	4.9
Afternoon peak	13–15	1.5876e+06	-3.51	0.0	2.31	5.8
Afternoon	16–18	1.5340e+06	-2.55	0.0	2.55	5.1
Evening	19–21	1.2539e+06	-2.10	0.0	2.10	4.2
Late evening	22–23	2.2501e+05	0.95	0.6	2.15	3.1

In the afternoon-peak band (13:00–15:00), the median delay is 0.0 min, with Q1 = -3.5 min and Q3 = +2.3 min (IQR = 5.8 min), indicating a modest asymmetry around the median and substantial dispersion. Across all seven bands (≈8.6 million records), the most considerable median shift is ≈1.2 min between the morning peak (-0.6 min) and late evening (+0.6 min). The Kruskal–Wallis (KW) test results confirm that there are differences among bands ($p < 0.001$). Because KW tests rank-distribution differences, we report practical effects as Hodges–Lehmann estimates of median differences; these remain small in absolute terms (≤ ≈1.2 min), consistent with the uniformly high punctuality observed throughout the day.

Fig. 4 uses consistent axis limits and marks outliers explicitly. Table 4 reports all summaries in minutes with decimal points to avoid ambiguity.

3.4. Interaction between time of day and day of week

We investigated the joint influence of weekday and time-of-day using the Scheirer-Ray-Hare (SRH) extension of the Kruskal-Wallis test, which does not assume normality or homoscedasticity. Fig. 5 shows a heatmap of the median delay for each 7 × 7 combination. Across all 49 cells, medians range from -1.2 min (Thursday, 22–23) to +1.0 min (Tuesday, 19–21). The SRH test indicates highly significant main effects and an interaction ($p < 0.001$). Because millions of observations were considered, minute differences are detectable statistically, yet practical magnitudes remain small. Even in the most extreme cells, the absolute median deviation is within ±1 min.

In practical terms, this implies that half of all trips arrive within 1 min of the time indicated on the timetable across all weekdays–time combinations, while the heatmap pinpoints when modest systematic shifts occur, guiding the placement of timing points and peak-only buffers.

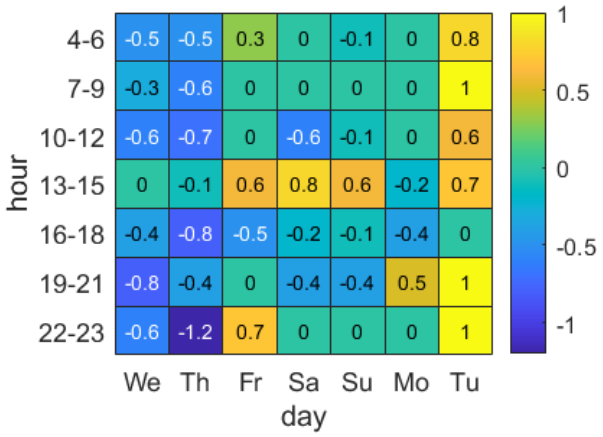

Fig. 5. Median arrival delay (min) by weekday and time-of-day interval

In practical terms, although formally detectable variations are observed between specific day-time combinations, the operator achieves high punctuality: half of all trips arrive within 1 min of the scheduled time, irrespective of the day of the week or time of day.

3.5. Delay patterns – clustering of stop profiles

Each stop's delay profile–defined as the vector of median delays across the seven time-of-day bands–was clustered using K-means (Euclidean distance) with K = 4 used to identify recurrent temporal patterns of deviation from the timetable. Fig. 6 depicts the cluster centroids by time band. K was selected for interpretability and consistency with elbow/silhouette diagnostics (details are available upon request). Cluster labels serve as stable, stop-specific priors that can be used as features in learning-based arrival time models and as operational cues for buffer allocation and timing point placement.

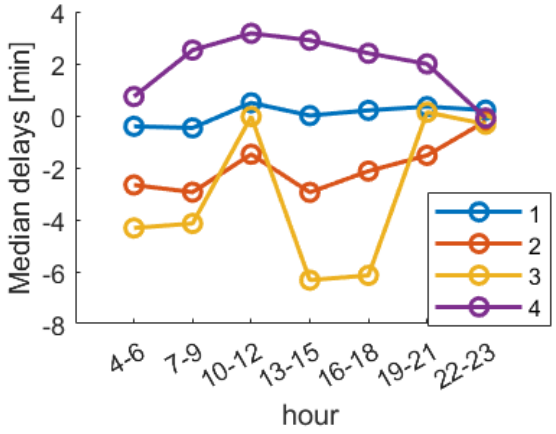

Fig. 6. Centroid delay-profile curves for the four clusters of stops, shown across the seven time-of-day intervals

Cluster 1 ("punctual") contains 52 stops with medians within ±1 min all day, indicating reliable on-time performance. Cluster 2 ("evening catch-up") includes 20 stops with early running (≈ -2 min) in the morning that diminishes towards 0 min by late evening, consistent with recoverable daytime slowdowns. Cluster 3 ("pre-schedule variability") contains 14 stops with strong early arrivals (down to ≈ -6 min) and fluctuation around mid-morning, which suggests excess slack near the origin or deliberate early acceleration. Cluster 4 ("afternoon-peak bottleneck") contains nine stops with pronounced positive delays from 13:00–15:00 (median ≈ +4 min), flagging locations for peak-only buffers or priority measures. These clusters expose punctuality patterns at individual stops that would be obscured by aggregation across stops or time bands.

3.6. Extended cluster analysis including day of week

We retained the original cluster assignments and stratified delays by weekday and time-of-day band. Thus, clusters were not re-estimated; instead, we computed, for each cluster, the seven medians across the seven weekdays. The resulting grid of cluster × weekday × time-band medians is plotted in Fig. 7.

Across clusters, Mondays exhibited the broadest range of medians (approximately −10 to +10 min), whereas Sundays were the most stable (about −4 to +4 min).

Operationally, buffer and control policies can be tuned by weekday. For example, Cluster 4 locations merit peak-only running-time adjustments on working days, while Cluster 3 near origins call for tighter early-day regulation to avoid excessive earliness. These insights complement predictions by making the temporal regularities explicit for planners.

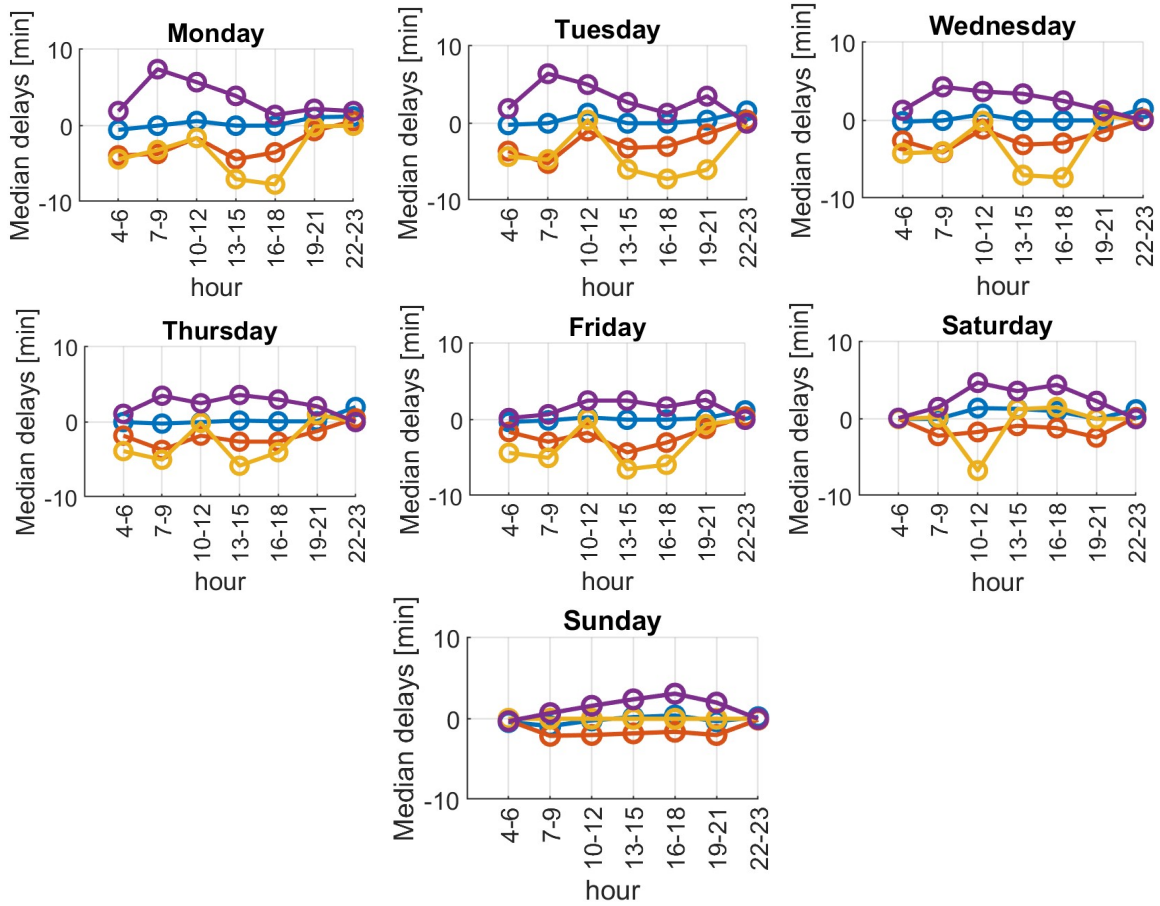

Fig. 7. Delay profiles for four stop clusters, depending on the day of the week

3.7. Distribution of stop clusters along the route

Retaining Section 3.5 assignments, we mapped the four stop profile clusters along both directions of line 112 (Karolin → CH Marki and CH Marki → Karolin). To enhance interpretability, we present Fig. 8 in two aligned panels per direction: one is a colored bar with one bar segment per stop (cluster identity), plotted against the percentage of route length; the other is an along-route performance profile showing the share of trips with a positive delay increment exceeding 2 min (Δdelay > 2 min) for each inter-stop segment. This second panel serves as an empirical proxy for where drivers are most likely to lose time and where congestion is most severe.

In both directions, origin and terminal stops belong to Cluster 1 (“punctual”), indicating stable adherence at endpoints. Immediately beyond the termini, sequences of Cluster 3 (“pre-schedule variability”) appear, consistent with early segment acceleration to build recovery buffers; this is particularly pronounced on Karolin → CH Marki. Further along the route, many stops transition to Cluster 2 (“evening catch up”), and central sections frequently revert to Cluster 1, signaling stable pacing.

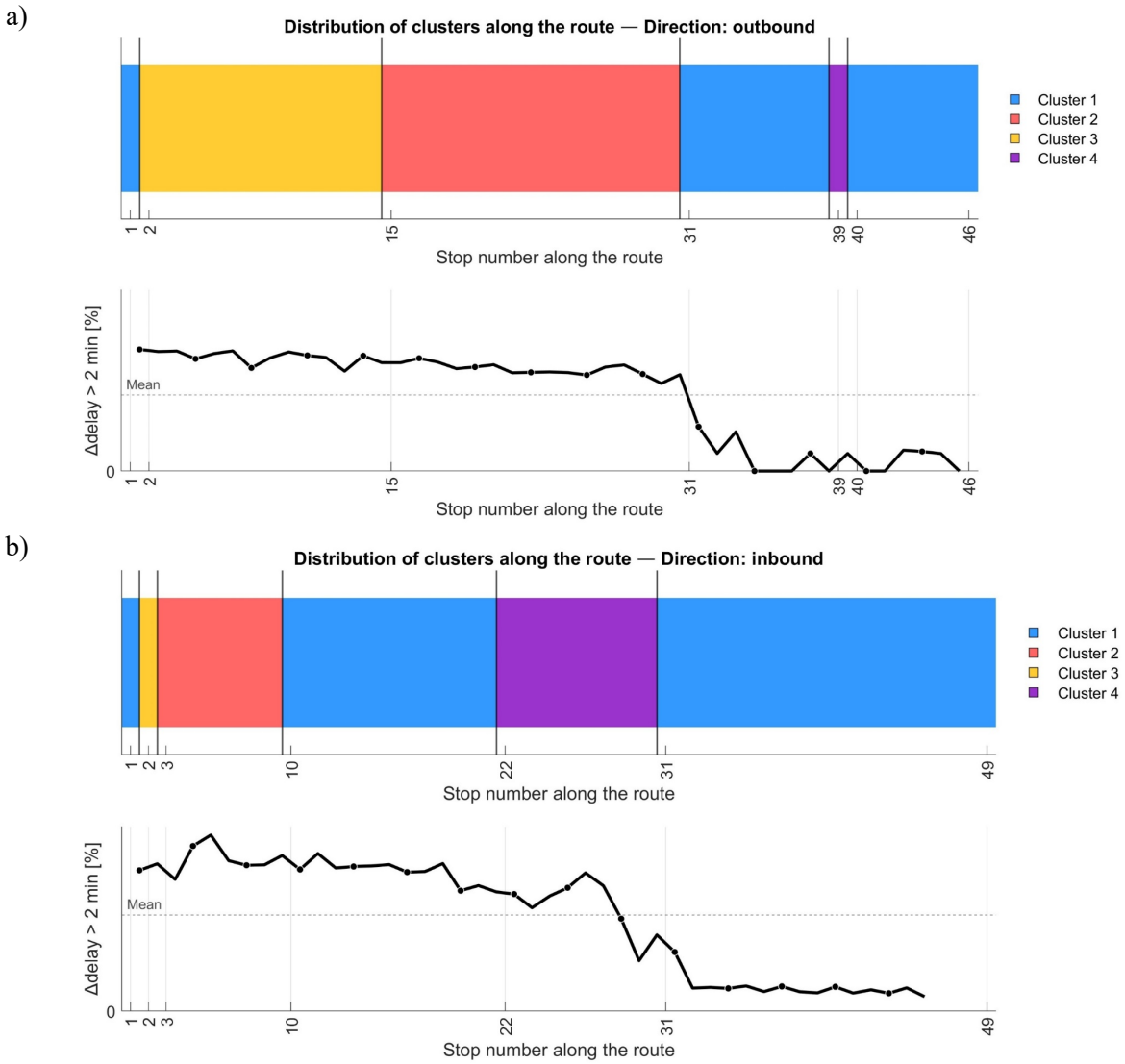

Fig. 8. Spatial distribution of stop profile clusters along the route (top) and segment level share of trips with Δdelay > 2 min (bottom) in the two travel directions: a) Karolin → CH Marki and b) CH Marki → Karolin

The segment level *Δ*delay profile corroborates these patterns. Elevated shares (typically 4–6%) are observed on the initial segments, indicating a higher probability of time loss where acceleration and merging occur. Mid-route segments exhibit lower and flatter shares (≈2–3%), consistent with Cluster 1 stability. On the return leg, local maxima occur around the Cluster 4 section (“afternoon peak bottleneck”), pinpointing the specific portions of the line where afternoon rush congestion systematically induces ≥ 2-min time losses. Together, the two panels identify where the operating regime changes (cluster transitions) and where time is most frequently lost (*Δ*delay hotspots), guiding the placement of timing points and targeted buffers.

4. DISCUSSION

This study aimed to quantify how realized arrivals align with the timetable at the stop level and convert raw variability into interpretable patterns with operational value. The principal distributional finding is that delays are nearly centered on zero but are markedly heavy-tailed, a combination that undermines Gaussian assumptions often implicit in classical analyses. Because 8.6 million matched events were considered, estimates of medians and interquartile ranges are precise, and the non-parametric inference framework appropriately accommodates non-normality and heteroscedasticity. Although formal tests detect statistically significant differences across time-of-day bands and weekdays, the practical magnitudes of the median shifts are small (typically within ±1 min), indicating consistently high punctuality under routine conditions.

The cluster analysis adds value by linking temporal deviation profiles to the route topology and to predictable operating regimes. The "punctual" type corroborates sections where existing timing points and buffers are well calibrated. "Pre-schedule variability" near origins indicates segments with excess slack or early acceleration; moderating early running or relocating timing points downstream can mitigate passenger penalties from early departures. "Evening catch-up" indicates recoverable slowdowns that are mainly absorbed later in the day and that can be supported with light downstream buffers rather than uniform padding. The "afternoon peak bottleneck," confined to one direction, isolates specific stops and a narrow time band that warrants peak-only running time adjustments or priority measures. Because these archetypes persist across weekdays–with the widest spread occurring on Mondays and the narrowest occurring on Sundays–they provide a stable basis for resilient timetable design that reflects recurrent temporal patterns rather than averages alone.

The along-route Δdelay profile (share of trips with delay change greater than 2 min per inter-stop segment) provides an incident-oriented complement to cluster identities. Peaks of this share on early segments confirm where recovery buffers are actively consumed, while troughs on central sections corroborate deliberate pace control. On the inbound afternoon peak, a localized share rise coincides with the Cluster 4 stretch, isolating recurrent congestion rather than random fluctuations. This dual evidence (profile + clusters) helps distinguish early running due to overly generous initial slack from genuine traffic-induced slowdowns, informing whether timing points should be relocated or whether peak-only priority and control should be deployed. Regarding causes, for Cluster 3 near the origins, we observed a high rate of on-time or early departures and elevated speeds on the first kilometers, implying that faster running, rather than early gate departure control failures, is the primary driver of early arrivals. For Cluster 4, the Δdelay peaks align with signalized junctions and merge areas on the inbound leg from 13:00-15:00, consistent with recurrent afternoon congestion.

The results complement prediction models, including deep learning models. Cluster identities and weekday–time medians offer low variance, stop specific priors that improve forecast accuracy when input histories are short, and reveal where and when systematic deviations arise–insights that pure prediction does not expose. From a data quality perspective, the strict spatiotemporal filtering and explicit trip–schedule alignment mitigate known ITS artefacts; nevertheless, a limitation is the absence of independent, local video validation. Prior studies have shown that camera-based ground truth helps calibrate Automatic Vehicle Location latency and dwell detection, and a targeted campaign at selected stops would further strengthen inference and refine thresholds.

Overall, the evidence suggests that timetable resilience can be improved not through indiscriminate slack but through targeted, cluster-informed interventions (i.e., tightening early segments prone to earliness, placing timing points at cluster transitions, and applying peak-only buffers and control where bottlenecks persist). These actions can be implemented straightforwardly within existing operational practices and monitored using the same stop-level profiles.

5. CONCLUSIONS

We presented a probabilistic–statistical framework that evaluates the concordance between realized and scheduled arrival times at the stop level and extracts interpretable temporal profiles with direct operational relevance. The framework was applied to a month of GNSS and GTFS data for a

representative urban line, and the analysis shows near-zero central tendency, moderate dispersion, and heavy tails extending to ±20 min. Non-parametric tests confirm statistically detectable but practically minor differences across time-of-day bands and weekdays. At the same time, clustering revealed four robust archetypes: "punctual," "evening catch up," "pre-schedule variability," and "afternoon peak bottleneck," whose spatial placement and weekday modulation map recurrent dynamics along the route. Beyond mapping clusters, an along-route, segment-level profile of the share of trips with delay change greater than 2 min highlights specific locations where time is most frequently lost, providing a practical diagnostic to reposition timing points, reallocate buffers, and prioritize peak-only interventions. Expressing the *x*-axis as route percentage ensured comparability across directions and lines.

The chief implication of the present findings is that timetable resilience benefits from precision rather than padding. Cluster-based profiles indicate where buffers should be reallocated, where timing points should be placed or moved, and where peak-only priority or control is justified. At the same time, the profiles provide interpretable features that enhance learning-based arrival time prediction without sacrificing transparency. Although demonstrated for a single line, the methodology relies on standard GNSS/GTFS feeds and is transferable to other routes and cities with comparable data availability. Future work will extend the analysis to multiple lines under diverse operating regimes. It will also incorporate independent video validation at selected stops to calibrate ITS timing and dwell detection, which will improve the reliability assessment and the design of targeted interventions.

References

1. Huk, K. & Bajda, N. & Bednarek, W. & Górak, A. Transport miejski w koncepcji logistyki miasta a zrównoważony rozwój. *Zeszyty Naukowe Małopolskiej Wyższej Szkoły Ekonomicznej w Tarnowie.* 2021. Vol. 51(3). P. 107-125. [In Polish: Urban transport in the concept of city logistics and sustainable development].
2. Wróbel, I. & Bartosik, B. & Gondek, P. & Piwowar, B. Transport solutions and indicators in smart cities – Part I. *Probl Kolejnictwa – Railw Report*s. 2023. Vol. 67(198). P. 147-159.
3. Kuo, Y.H. & Leung, J.M.Y. & Yan, Y. Public transport for smart cities: Recent innovations and future challenges. *Eur J Oper Res*. 2023. Vol. 306(3). P. 1001-1026.
4. Czerepicki, A. & Krukowicz, T. & Górka, A. & Szustek, J. Traffic light priority for trams in Warsaw as a tool for transport policy and reduction of energy consumption. *Sustainability.* 2021. Vol. 13(8). No. 4180.
5. Brzeszczak, A. & Imiołczyk, J. & Czuma-Imiołczyk, L. Zrównoważony transport publiczny - społeczna ocena transportu zbiorowego w Częstochowie. *Stud Miej.* 2018. Vol. 30. P. 85-98. [In Polish: Sustainable public transport - social assessment of public transport in Częstochowa].
6. Dubiel, K. Analiza wyników badań oceny i oczekiwań pasażerów gminnego transportu pasażerskiego w Wieliczce. *Transp Miej i Reg.* 2022. Vol. 11-12. [In Polish: Analysis of the results of the survey on the assessment and expectations of passengers of municipal passenger transport in Wieliczka].
7. Czerepicki, A. & Kozłowski, M. Probabilistic analysis of public bus transport in Warsaw. In: Gitolendia, B. & Sladkowski, A. & Natriashvili, T. & Gogiashvili, F. (eds.) In: 8*th Georgian-Polish International Scientific Conference "Transport Brigde Europe-Asia.*" Tbilisi: Georgian Technical University. 2024. P. 330-334.
8. Šojat, D. & Slavulj, M. & Sikirić, M. & Ćosić, M. Using minimum travel time to determine factors influencing travel time discrepancy and variability in tram transit. *Appl Sci*. 2024. Vol. 14(24). No. 11599.
9. Ma, Z. & Ferreira, L. & Mesbah, M. & Zhu, S. Modeling distributions of travel time variability for bus operations. *J Adv Transp.* 2016. Vol. 50(1). P. 6-24.
10. Rashvand, N. & Hosseini, S.S. & Azarbayjani, M. & Tabkhi, H. Real-time bus arrival prediction: A deep learning approach for enhanced urban mobility. *arXiv Prepr arXiv*. 2023. No. 230315495.
11. Shanthi, N.V.E.S. & Upendra Babu, K. & Karthikeyan, P. et al. Analysis on the bus arrival time prediction model for human-centric services using data mining techniques. *Comput Intell Neurosci*. 2022. Vol. 2022. P. 1-13.

12. Luo, T. & Liu, X. & Jin, H. Bus queue time estimation model for a curbside bus stop considering the blocking effect. *Sci Rep*. 2022. Vol. 12(1). No. 11576.
13. Ding, H. & Xu, D. & Xu, S. et al. Bus travel time prediction based on time-varying adaptive Kalman filter method. In: Ma, C. (ed.) *International Conference on Frontiers of Traffic and Transportation Engineering (FTTE 2022).* SPIE. 2022. DOI: 10.1117/12.2652414
14. Vidya, G.S. & Hari, V.S. Prediction of bus passenger traffic using gaussian process regression. *J Signal Process Syst*. 2023. Vol. 95(2-3). P. 281-292.
15. Jin, G. & Yan, H. & Li, F. et al. Spatio-temporal dual graph neural networks for travel time estimation. *ACM Trans Spat Algorithms Syst.* 2024. Vol. 10(3). P. 1-22.
16. Kieu, L.M. & Bhaskar, A. & Chung, E. Empirical evaluation of public transport travel time variability. In: *Australasian Transport Research Forum 2013 Proceedings.* 2013. P. 1-18.
17. Fosgerau, M. & Fukuda, D. Valuing travel time variability: Characteristics of the travel time distribution on an urban road. *Transp Res Part C Emerg Technol.* 2012. P. 83-101.
18. Ma, Z. & Zhu, S. & Koutsopoulos, H.N. & Ferreira, L. Quantile regression analysis of transit travel time reliability with automatic vehicle location and farecard data. *Transp Res Rec J Transp Res Board*. 2017. No. 2652. P. 19-29.
19. Chen, X. & Cheng, Z. & Schmidt, A.M. & Sun, L. Conditional forecasting of bus travel time and passenger occupancy with Bayesian Markov regime-switching vector autoregression. T*ransp Res Part B Methodol*. 2025. Vol. 192. No. 103147.
20. Chen, X. & Cheng, Z. & Jin, J.G. et al. Probabilistic forecasting of bus travel time with a bayesian gaussian mixture model. *Transp Sci*. 2023. Vol. 57(6). P. 1516-1535.
21. Huang, Y.P. & Chen, C. & Su, Z.C. et al. Bus arrival time prediction and reliability analysis: An experimental comparison of functional data analysis and Bayesian support vector regression. *Appl Soft Comput*. 2021. Vol. 111. No. 107663.
22. Czerepicki, A. & Kozłowski, M. Method for determining the urban bus cycle using bayesian inference. In: *2024 3rd International Conference on Problems of Logistics, Management and Operation in the East-West Transport Corridor (PLMO).* Baku: IEEE. 2024. P. 1-5.
23. Chen, X. & Yu, L. & Zhang, Y. & Guo, J. Analyzing urban bus service reliability at the stop, route, and network levels. *Transp Res Part A Policy Pract.* 2009. Vol. 43(8). P. 722-734.
24. Cats, O. & Larijani, A.N. & Koutsopoulos, H.N. & Burghout, W. Impacts of holding control strategies on transit performance. *Transp Res Rec J Transp Res Board.* 2011. No. 2216. P. 51-58.
25. Tirachini, A. & Cortés, C.E. & Jara-Díaz, S.R. Optimal design and benefits of a short turning strategy for a bus corridor. *Transportation.* 2011. Vol. 38(1). P. 169-189.
26. Szymański, P. & Żołnieruk, M. & Oleszczyk, P. et al. Spatio-temporal profiling of public transport delays based on large scale vehicle positioning data from GPS in Wrocław. *IEEE Transactions on Intelligent Transportation Systems.* 2017. Vol. 19(11). P. 3652-3661.
27. Nguyen, K. & Yang, J. & Lin, Y. et al. Los Angeles Metro bus data analysis using GPS trajectory and schedule data (demo paper). *arXiv.* 2019. No. 1909.00955v1.
28. Pałys, Ł. & Ganzha, M. & Paprzycki, M. Applying machine learning to predict behavior of bus transport in Warsaw, Poland. *arXiv*. 2022. No. 2204.04515v1.
29. Warsaw Public Transport Department. *Warsaw public transport timetables.* 2024. Available at: https://www.ztm.waw.pl/pliki-do-pobrania/dane-rozkladowe/.
30. *General Transit Feed Specification*. Available at: https://gtfs.org/documentation/overview/.
31. Warsaw City Council. *Warsaw Open Data Repository*. Available at: https://api.um.warszawa.pl/.
32. Czerepicki, A. Application of graph databases for transport purposes*. Bull Polish Acad Sci Tech Sci*. 2016. Vol. 64(3). P. 457-466.
33. Żarski, K. & Oskarbski, J. & Bliszko, K. Accuracy analysis of public transport vehicle travel time data from ITS services. In: Rosiński, A. & Siergiejczyk, M. (ed.) *Transport of the 21st Century. Conference proceedings*. 2025. P. 333.

Received 05.04.2024; accepted in revised form 16.02.2026