UDC 656.256: 681.32
DOI: 10.15587/1729-4061.2025.340757

IMPROVING CONTROL OVER OPERATIONAL CHARACTERISTICS OF SUBWAY ROLLING STOCK BY USING RETROSPECTIVE PASSENGER FLOW ESTIMATION

Ivan Siroklyn
Corresponding author
PhD*
E-mail: seroklin@kart.edu.ua
Serhii Zmii
PhD*
Vasyl Sotnyk
PhD*
Olena Shcheblykina
Doctor of Philosophy (PhD)*
*Department of Automation and Computer Telecontrol of Trains
Ukrainian State University of Railway Transport
Oboronny Val (Feuerbach) sq., 7, Kharkiv, Ukraine, 61050

This study investigates the process that forms operational loads on the electric rolling stock within a city's underground railroad system. The task addressed relates to the impossibility to directly measure passenger flows on separate sections of the network because most control systems only record passenger entry and exit events without tracking their full route. This creates analytical gaps and makes it impossible to accurately assess the actual load, which is critical for technical diagnostics as well as maintenance.

To solve this problem, a method for estimating passenger flows has been devised, based on stochastic modelling of passenger movement on an oriented graph of the transport network.

An approach to load estimation has been proposed, the distinctive feature of which is the rejection of determining a single most probable route for each passenger. Instead, the conditional "weight" of a passenger is probabilistically distributed among the set of all possible routes that could lead him/her from the potential entry station to the actual exit station within time constraints.

Method verification by simulation on a conditional network showed high accuracy of the results (the average relative error did not exceed 0.5%). The distribution of errors is symmetrical, close to normal, and concentrated around zero, which indicates the absence of systematic deviations. The accuracy is attributed to the fact that the probability distribution makes it possible to level out the uncertainty of the passenger's choice of a specific route and obtain an objective integrated assessment of the load at the level of individual routes.

The scope of practical application of the results includes technical monitoring systems for rolling stock, adaptive repair planning, optimization of timetables, as well as improvement of transport safety in the context of using anonymous means of payment for travel

Keywords: route reconstruction, subway, adaptive repair planning, optimization of timetables

Received 30.06.2025
Received in revised form 03.09.2025
Accepted 22.09.2025
Published 31.10.2025

***How to Cite:** Siroklyn, I., Zmii, S., Sotnyk, V., Shcheblikina, O. (2025). Improving control over operational characteristics of subway rolling stock by using retrospective passenger flow estimation. Eastern-European Journal of Enterprise Technologies, 5 (3 (137)), 68–74. https://doi.org/10.15587/1729-4061.2025.340757*

1. Introduction

Urban rail transport systems, in particular the subway, are critical infrastructure in large cities, which provide transportation of a significant number of passengers every day. Effective operation of electric rolling stock in such systems requires an accurate understanding of the dynamics of passenger flows and the real load on individual network elements. This becomes especially relevant in the context of technical diagnostics, extension of the service life of electric rolling stock, and prevention of emergencies associated with overloading.

After introducing contactless and impersonal ways to pay for travel (tokens, cards, tickets without personalization), personalized methods for tracking passenger routes become ineffective. The lack of individual identification creates significant analytical gaps, in particular, it makes it impossible to directly record passengers on specific trains or on certain sections of the network.

Thus, the relevance of research is predetermined by the urgent need to devise new methods for assessing operational load, adapted to conditions of incomplete observability and impersonal data. The assessment of operational load is a necessary condition for ensuring the reliability and safety of subway operation. Devising such methods opens up new prospects for automated monitoring of operational loads, adaptive planning of traffic schedules, as well as assessment of the technical condition of electric rolling stock in real time. Moreover, such methods could be used not only for retrospective analysis but also as a tool for operational management and forecasting under conditions of limited access to complete passenger route data.

2. Literature review and problem statement

Paper [1] gives an analysis of the development of methods for studying operations in the field of public transport over the past 50 years, focusing on the efficiency of planning and logistics of transportation. However, most of these models do not take into account the individual trajectory of the passenger and the dynamics of route changes, which limits the application of these approaches in technical diagnostics tasks.

Another area of research is the use of data from electronic ticketing systems. In particular, in [2, 3], methods for estimating OD matrices and passenger exit points using data from "smart" maps were proposed. However, even in cases when the entry and exit stations are known, the problem of uncertainty of the actual route followed by the passenger remains. To solve this problem, in [4], probabilistic approaches to route reconstruction taking into account the traffic schedule and transfer times are studied. An improved version of such an approach is reported in [5], in which a latent cluster model is given taking into account changes in route choice over several days. These approaches allow for the ambiguity of route selection, but they are usually used for planning tasks and are not directly related to the assessment of technical load on rolling stock.

A number of studies, in particular [6, 7], demonstrate the potential of combining big data and simulation models for load forecasting and technical condition diagnostics. In [6], a method for predictive maintenance of road infrastructure based on historical loads is substantiated, but without detailing temporal patterns. In [7], modeling is carried out based on data from mobile operators, which potentially allows for the assessment of passenger flows but the application of such methods in the subway is limited due to the lack of a signal underground.

Work [8] proposes a passenger flow management model that takes into account the psychological aspect of passengers' readiness to board. However, its main goal is to improve the convenience of service and throughput, not technical diagnostics. The model answers the question "how to best distribute passengers for their convenience?" but not the question "what load falls on a specific train No. X along the Y–Z section?". Papers [9, 10], based on data for analyzing the choice of routes by passengers (Data-Driven), are aimed at optimizing the transport system from the passenger and operator's point of view. Although they allow us to understand how passengers choose their routes, their results do not contain a reference to specific trains at specific points in time, which makes it impossible to calculate the cumulative load for assessing wear.

In [11], an attempt is made to evaluate OD matrices using machine learning. However, the results are reduced to the general distribution of flows without reflecting their detail by routes. This does not allow for an assessment of the load on each specific train or car at a specific time.

In [12], a graph analysis of a transport network using the Madrid subway as an example is proposed to identify structural vulnerabilities. However, such an analysis is not related to the technical load on the rolling stock and does not cover the possibility of partial distribution of the weight of one passenger between several routes.

In [13], a method for route reconstruction based on the hypothesis of minimum travel time is reported. However, the method does not provide for a conditional distribution of passenger weight between several possible trajectories and does not allow for estimating the load in the form of aggregated flows.

Thus, a key unsolved problem has been identified: the lack of a single method that would make it possible to estimate the dynamic load on a separate electric rolling stock and network sections under conditions of partial observability. Existing methods are mostly focused on generalized modeling of passenger flows for strategic planning but do not provide the detail necessary for technical diagnostics tasks. In particular, the question remains unresolved as to how, under conditions of route uncertainty, to objectively distribute the "weight" of one passenger among all possible trajectories of his/her movement to obtain an integrated estimate of the load on each train.

3. The aim and objectives of the study

The aim of our study is to devise a method for estimating the load on electric rolling stock in the subway system based on the reconstruction of passenger routes under conditions of partial observability, when only the entry and exit points of passengers are known. This will make it possible to design data-based technical monitoring systems, which will enable a transfer to adaptive planning of rolling stock maintenance and increasing the safety of its operation.

To achieve the goal, the following tasks have been set:

– to build a mathematical model of passenger movement in the subway transport network as a stochastic process on a directed graph;

– to formalize assumptions about typical passenger behavior patterns (for example, choosing the shortest or fastest route) in the form of probability functions of transition between graph nodes;

– to propose a procedure for probabilistic distribution of the passenger "weight" within a set of permissible routes to estimate the load on network elements;

– to propose a methodology for assessing the load on each elementary route (the movement of electric rolling stock between adjacent stations) as the sum of the fractions of passenger weights with the probability distribution of their movement through the network;

– to simulate passenger traffic under the conditions of a simplified subway network, to obtain reference values of train load based on complete information on passenger traffic.

4. The study materials and methods

The object of our study is the process that forms a load on the electric rolling stock in the subway under conditions of partial observability of passenger traffic. The principal hypothesis assumes that even in the absence of information about the full route of each passenger, it is possible to restore an adequate load distribution on individual trains and runs using a probabilistic model of network movement.

The study assumes that passenger behavior is subject to typical route selection patterns (minimization of time, number of transfers, etc.), and that the topology of the transport network is stationary during the experiment.

Among the simplifications used are aggregated time intervals of movement and generalized coefficients in route selection functions.

The research methods included theoretical modeling based on graph theory and stochastic processes, formalization of probabilistic route selection functions, and simulation of passenger traffic. Specialized software (Python) was used to process data and implement algorithms. The hardware included standard computing equipment (personal computer, Intel i7 processor, 16 GB RAM).

The procedure for checking the adequacy of the models involved comparing the reconstructed passenger flows with reference data generated in the simulation network. Assessment was carried out using the relative error and the distribution of deviations, which allowed us to establish the correspondence of the results to the expected behavior of the system.

The use of artificial intelligence tools was limited and concerned exclusively auxiliary tasks, namely, the search for scientific sources in open databases and their primary content analysis. Artificial intelligence technologies were not used when building a mathematical model, developing algorithms, and conducting experiments.

5. Results of investigating a retrospective method for calculating passenger flows

5. 1. Construction of a stochastic model of the passenger route as a process on a directed graph

To formalize the process of passenger movement through the subway network, it is advisable to use graph theory in combination with stochastic modeling. The subway network is described as a directed graph $G=(V, E)$, where V is a set of vertices corresponding to subway stations, and $E \subseteq V \times V$ is a set of arcs, each of which represents a directed segment of the tunnel.

Taking into account the fact that the passenger moves step by step from station to station over a limited period of time, it is advisable to describe such a process in the form of a Markov chain with a variable number of steps. Let $X_t \in V$ be a random variable that denotes the station of the passenger at time t. The process $\{X_t\}_{t \in N}$ models the probabilistic route of the passenger through the subway network. The input condition is the fixation at the input station $v_{entry} \in V$ at time t_0, i.e., $X_{t_0} = v_{entry}$.

For each arc $(u, v) \in E$, the transition probability $P(u \to v)$ is determined. It depends both on the structural characteristics of the network (the presence of transfer stations, topological distance) and on the behavioral characteristics of passengers (the tendency to choose routes with the minimum duration or number of transfers). Thus, the transition probability between stations can be expressed through the function

$$P(u \to v) = \frac{w(u,v)}{\sum_{v' \in Adj(u)} w(u,v')}, \tag{1}$$

where $w(u,v)$ is a weighting function reflecting the "attractiveness" of the transition from u to v; $Adj(u)$ is the set of all vertices accessible from u.

The function $w(u,v)$ may include components that take into account:

– the minimum travel time between stations (physical distance, waiting time for electric rolling stock);

– whether station v belongs to the main or peripheral direction;

– the previous history of passenger movements (in the case of individualized models);

– the presence of typical routes derived based on aggregated statistics.

The stochastic process formalized in this way allows us to model the set of possible passenger routes from the moment of entry to the moment of the end of the trip (or until the permissible time limit is reached). Since the model assumes the absence of data on the starting station, the endpoint of the trajectory is modeled implicitly, either through a restriction on the cumulative duration of stay within the system or through maximizing the posterior probability of the route based on historical data.

As a result of building such a model, a space of permissible routes $P(v_{entry}, T_{\max})$ is constructed, where $T_{\max}$ is the maximum time of the passenger's stay in the system, determined on the basis of access control data. This space serves as the input set for further route reconstruction algorithms, described below.

5. 2. Probabilistic modeling of passenger route selection patterns

The behavior of a passenger in the subway when choosing a route is usually not random but is subject to certain patterns determined by the desire to minimize travel time, the number of transfers, or the number of stations visited. In order to integrate these patterns into the stochastic movement model described in the previous subchapter, it is necessary to formalize them in the form of parameterized functions of the probability of transition between stations.

For each vertex $u \in V$, the set of admissible subsequent vertices $Adj(u) \subseteq V$ determines the possible steps of the passenger's movement. Within this set, a function for evaluating the feasibility of transition to each subsequent vertex $v \in Adj(u)$ is introduced, which takes into account a certain behavioral strategy of the passenger. In particular, the following typical patterns are distinguished:

– the Shortest Path Preference is focused on minimizing the number of arcs in the route;

– the Fastest Path Preference pattern minimizes the total travel time, taking into account the average speed of trains and the transfer time;

– the Minimal Transfers Preference pattern minimizes the number of line changes;

– the Habitual Route Preference pattern reproduces a repeating sequence of movements recorded in the historical data on a particular passenger or a cluster of similar passengers.

To generalize the passenger's choice based on the above strategies, a weighted transition function is introduced

$$P(u \to v) = \frac{e^{-\beta \cdot C(u,v)}}{\sum_{v' \in Adj(u)} e^{-\beta \cdot C(u,v')}}, \tag{2}$$

where $C(u, v)$ is the generalized cost of the transition, defined as a linear combination of distance, time, transfers, etc.; β is the cost sensitivity parameter (adjusts the degree of passenger selectivity).

Exponential smoothing avoids hard determinism and ensures stochasticity of choice. The specific form of function $C(u,v)$ can be adapted depending on the context

$$C(u,v) = \alpha_1 \cdot Time(u,v) + \alpha_2 \cdot Transfers(u,v) + \\ + \alpha_3 \cdot Distance(u,v), \tag{3}$$

where α_1, α_2, α_3 are weighting factors that can be determined empirically or based on the classification of the passenger type (for example, "regular user", "guest passenger", etc.).

Thus, the model of transition between the vertices of a directed graph taking into account behavioral strategies allows us not only to generate a set of plausible routes but also specify a gradation of their probability. This is critically important for further ranking of possible passenger trajectories based on limited data.

This probabilistic mechanism becomes the basis for implementing the procedure for reconstructing the most probable passenger route in a given time interval.

5. 3. Procedure for the probability distribution of passengers to estimate the load

Given the described stochastic model of passenger movement, reconstruction of a probable route with limited access control data is reduced to the problem of choosing the most probable trajectory $k=(v_0, v_1, \ldots, v_N)$ in a directed graph $G=(V, E)$. This corresponds to time constraints and entry-exit events.

Let the time instants of entry (v_{in}, t_{in}) and exit (v_{out}, t_{out}) of a passenger in the network be fixed. Since information about the starting station v_0 is missing, the task is to determine the probable initial node of the graph and the sequence of transitions that explain the observations.

For each possible route k between stations v_0 and $v_n=v_{out}$ with a travel time not exceeding $t_{out}-t_{in}$, the trajectory probability is calculated as the product of the transition probabilities

$$P(k)=\prod_{i=0}^{n-1} P(v_i \to v_{i+1}), \tag{4}$$

where $P(v_i \to v_{i+1})$ is determined by the formula for probabilistic transition selection taking into account the generalized cost $C(v_i, v_{i+1})$

$$P(v_i \to v_{i+1})=\frac{e^{-\beta \cdot C(v_i, v_{i+1})}}{\sum_{v' \in Adj(u)} e^{-\beta \cdot C(v_i, v')}}. \tag{5}$$

The cost of transition $C(u, v)$ takes into account the travel time between stations, the number of transfers, the distance, and other relevant parameters.

Finding the most probable route can then be reduced to an optimization problem

$$k^*=\arg \max_{k \in P(v_0, v_{out}, T)} P(k), \tag{6}$$

where $P(v_0, v_{out}, T)$ is the set of routes connecting stations v_0 and v_{out} with a duration not exceeding $T=t_{out}-t_{in}$.

In the context of analyzing aggregated passenger flow data, where there is no individual identification of passengers, this procedure is performed for a set of initial observations simultaneously, taking into account the statistical distribution of the probability of entries at each station in the corresponding time intervals.

To solve the problem in practice, algorithmic approaches based on priority search (for example, search in the state space taking into account weights), Monte Carlo methods, or heuristic algorithms are used. These approaches make it possible to effectively find routes with maximum probability. The obtained route reconstruction results can be used for further assessment of the passenger load on each car of the electric rolling stock, superimposing the reconstructed trajectories on the train schedule and the car identification system.

However, in the tasks of assessing the technical condition of electric rolling stock, in particular when calculating the load on each train on each of the runs, there is no need to determine the single most probable route of each passenger. It is more expedient to use the conditional probability distribution approach, within which the "weight" of one passenger is distributed among a set of permissible routes taking into account logical restrictions (direction of movement, permissible lines, duration of transfers, etc.).

In practical implementation, after fixing the exit of a passenger at a certain station S_{exit} at time t_{exit}, the system generates a set of permissible routes $R_{possible}$ that could lead this passenger to the exit point. Each of such routes starts from a certain potential entry station S^k_{entry} at corresponding time t^k_{entry}, and includes a sequence of elementary routes – that is, trips by one train between neighboring stations in a certain time interval.

For each potential entry station, the number of passengers N^k_{entry}, who entered at the corresponding time and could be "companions" of the passenger is determined. The sum of all such potential fellow travelers across all entry stations

$$N_{total}=\sum_k N^k_{entry}, \tag{7}$$

serves as a normalization coefficient, on the basis of which the probability P_k that the passenger entered through S^k_{entry} is calculated

$$P_k=\frac{N^k_{entry}}{N_{total}}. \tag{8}$$

Thus, a passenger is not assigned to a single route, but his/her weight shares are distributed to the corresponding routes, proportional to probabilities P_k.

Next, for each elementary route e^t_{ij} (train movement between stations i and j at time t), the number of transported passengers is calculated as the sum of the corresponding weight shares for all passengers and all routes, where this elementary route is part of

$$W^t_{ij}=\sum_p \sum_{r \in R^p_{possible}} \delta^t_{ij}(r) \cdot w_p \cdot P^{(p)}_r, \tag{9}$$

where $\delta^t_{ij}(r)=1$, if $e^t_{ij} \in r$, otherwise – $\delta^t_{ij}(r)=0$; w_p – conditional mass of passenger p; $P^{(p)}_r$ – probability that passenger p traveled along route r; $R^p_{possible}$ – set of potential routes of passenger p.

This procedure allows us not only to calculate the average load on each train with reference to the run but also apply the results for long-term analysis of wear of electric rolling stock. In addition, it makes it possible to apply the results for optimization of maintenance planning and assessment of the efficiency of the timetable.

5. 4. Methodology for assessing the load of electric rolling stock based on the probability distribution of passenger routes

Given that in the framework of assessing the load on electric rolling stock there is no need to establish a unique route for each individual passenger, a method of conditional probability distribution of the mass of passengers over the set of permissible routes is used. This allows us to build a weighted model of rolling stock load, which takes into account the possibility of alternative trajectories of passenger movement within logical constraints.

To implement this approach, each passenger's exit at a certain station at a certain point in time is interpreted as a point of fixation of the route result. For this point, a set of potential entry stations is determined, taking into account the possible entry time and available combinations of trains and transfers that could lead to a fixed exit. Each of such combinations sets a permissible route, divided into a sequence of elementary routes. that is, movements between neighboring stations by a specific train at a certain time.

The probability of following a specific elementary route for one passenger is defined as the sum of two components:

1. The conditional probability of the passenger entering at the corresponding station (determined based on the ratio of the number of fellow travelers at this station to the total number of potential fellow travelers from all permissible entry stations for this exit);

2. Probabilities that this passenger used a specific route, which includes a given elementary segment.

Thus, mathematically, the mass of one passenger is decomposed into particles, each of which is added to the corresponding elementary route. The total value of all such particles generated by all passengers for a fixed simulation period forms a train load table containing:

– train identifier (by double indexing T<line><number>);

– direction identifier (for example, T11′, T11″, etc.);

– time period during which the train was moving;

– section number between neighboring stations;

– number of weighed passengers transported by this train along this section.

This method avoids false aggregation along fixed routes and provides a more accurate assessment of the dynamics of loading of individual elements of the transport network. The method is important for further diagnostics of the technical condition of rolling stock, optimization of the traffic schedule, and assessment of transportation efficiency.

5. 5. Simulation of passenger flows for a conventional subway network

To verify the accuracy of the proposed method, an experimental modeling of passenger flow in a simplified subway transport network consisting of 12 stations was conducted (Fig. 1). The stations are combined into three separate lines with four stations along each line. The lines do not intersect, but the possibility of transfers between them is provided. This structure provides reasonable complexity of routes, including direct, transfer, and multi-stage passenger movements.

Within the framework of simulation, 60,000 passengers were generated, each of whom traveled along a given route, described as a sequence of elementary routes with a clear time stamp corresponding to the passage of each segment. In total, 366 unique elementary routes with a time reference (i.e., unique combinations of "leg – time") were generated and analyzed. This means that the experimental base included 366 independent observation units for which it is possible to compare the number of transported passengers obtained in two ways:

– by direct counting according to the given routes (Stage 1);

– by reconstruction based on probabilistic modeling and known passenger entry/exit points (Stage 2).

A comparison of the results of direct counting and reconstruction is shown in Fig. 2 for three time stamps from the start of the simulation (at 1200, 1500, and 1800 seconds).

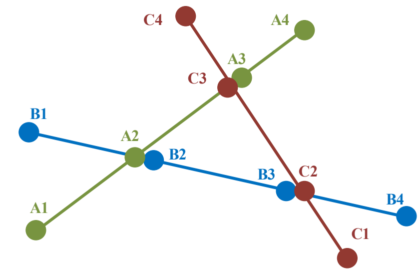

Fig. 1. Conditional subway network for modeling

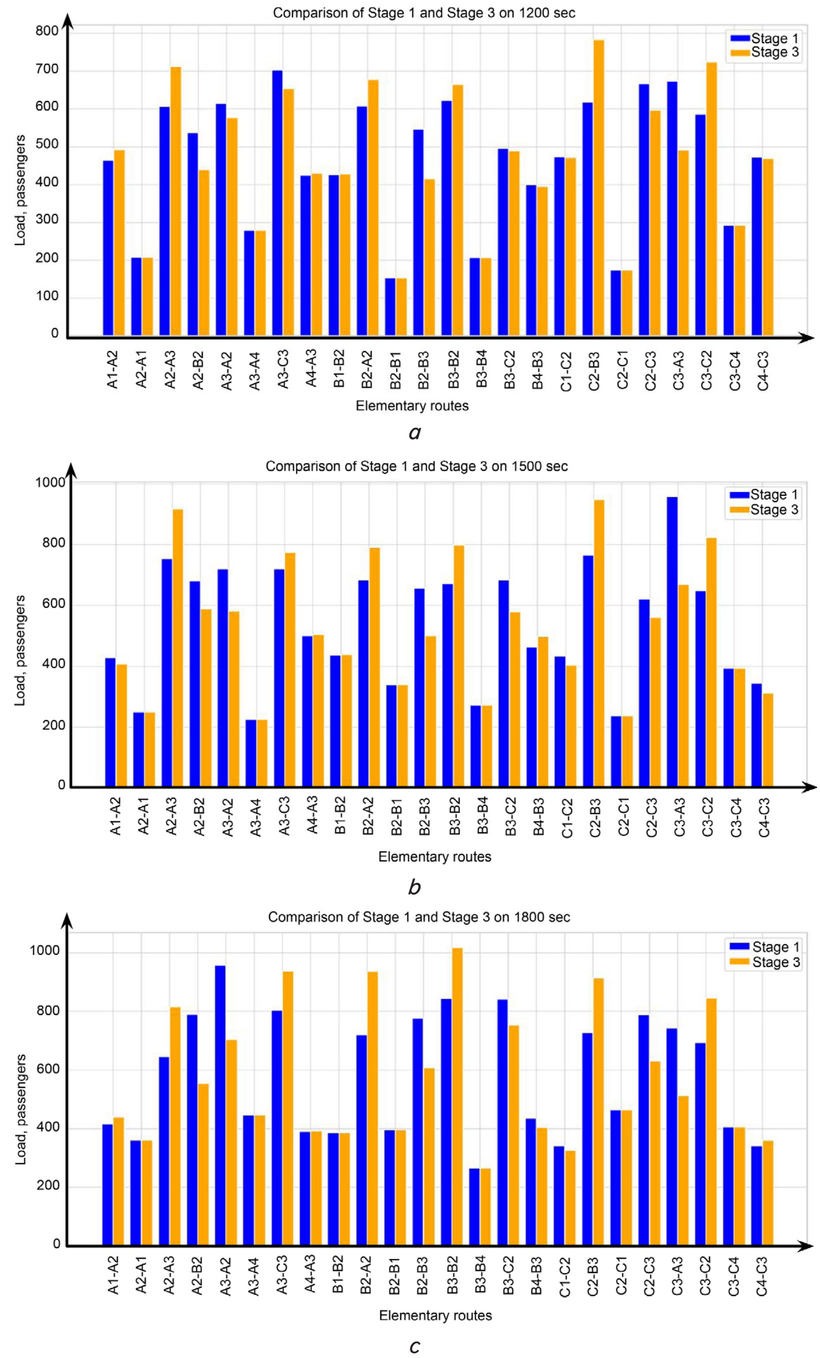

Fig. 2. Comparison between direct and retrospective passenger counting for different timestamps: *a* – 1200 s; *b* – 1500 s; *c* – 1800 s

For each technique, the absolute and relative errors in determining the total number of passengers transported were calculated:

– absolute error $\Delta w_i = | w_i^{(1)} - w_i^{(2)} |$;

– relative error $\delta_i = \Delta w_i / w_i^{(1)}$.

Here $w_i^{(1)}$ is the actual load (in passengers) according to the results of direct modeling (Stage 1); $w_i^{(2)}$ – load reconstructed using the probabilistic approach (Stage 2).

As a result of our analysis of the obtained data, a histogram of the distribution of the relative error was constructed, which takes into account the sign of the deviation. The comparison results show that the average relative error in most cases does not exceed 0.5%. The distribution of errors is symmetrical and concentrated near zero. No systematic shifts were detected; the distribution is close to normal (Fig. 3).

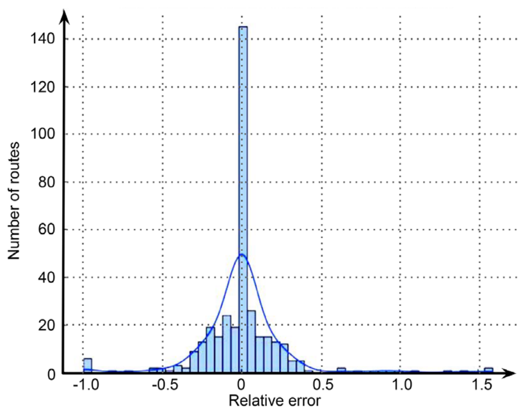

Fig. 3. Distribution of relative error between direct and retrospective passenger counting

Therefore, our results allow us to consider the proposed method acceptable for use in tasks of retrospective assessment of the load on electric rolling stock and for generating data within the framework of technical monitoring systems for the condition of transport infrastructure.

6. Discussion of results based on assessing the load on the electric rolling stock according to the reconstruction of passenger flows

The load assessment method proposed in our work is based on the representation of passenger movements in the form of a stochastic process on a directed graph, which is described by formula (1). Such modeling makes it possible, unlike aggregated models, to take into account the dynamics of movement at the level of individual sections, which is a fundamental prerequisite for detailed technical diagnostics. The integration of probabilistic passenger behavior patterns into the transition weights of graph (2) allowed us to take into account the variability in human decision-making, which favorably distinguishes this method from deterministic models in which all passengers are assigned a single optimal route.

The key element of the devised method is the probabilistic distribution of the "weight" of one passenger between the set of admissible trajectories, which is formalized in formula (9). This method directly resolves the issue of route uncertainty under conditions of partial observability. Instead of making a potentially false assumption about a single path, the method integrates all possibilities, which, when averaged over a large number of passengers, provides a high accuracy of the integral estimate. Thus, the proposed method solves the problem related to the lack of tools for detailed load assessment for individual sections, in contrast to existing approaches that focus on generalized OD matrices.

The results of simulation (Fig. 2) quantitatively confirmed the high accuracy of the proposed method. High accuracy (average relative error $< 0.5\%$) is attributed to the fact that the model does not try to determine a single route of an individual passenger but distributes its "weight" between a set of possible trajectories. Averaging over a large sample (60,000 passengers) eliminates individual deviations and ensures that the reconstructed flows match the reference data. The symmetrical, close to normal distribution of errors (Fig. 3) indicates the absence of systematic shifts in the calculations.

The advantages of our method are due to its specific features. Unlike approaches [2, 3, 11], which focus on OD matrices and do not make it possible to assess the load on individual sections, the proposed algorithm provides detailed values for each leg and train. Unlike deterministic models of the type reported in [13], which assign all passengers the optimal route, stochastic modeling (2) takes into account the variability of real behavior. Unlike methods [7], which require mobile data and are not suitable for the subway, the proposed method relies on existing information from access control systems without the need for additional equipment.

The limitations of the study are that it is retrospective and based on synthetic data: the adequacy of the results is guaranteed only under conditions of stable network operation and complete accounting of entry/exit events. When applied in practice, it is necessary to take into account that real passengers can change routes, miss trains, and be delayed at stations, which will affect the accuracy of the reconstruction.

The disadvantage of the current implementation is the use of generalized coefficients in the route selection functions (3) without calibration on real data, as well as significant computational complexity when scaling to large systems with hundreds of stations and millions of trips. This may limit the possibility of obtaining results in an acceptable time.

Future research should be linked to three areas:

1) calibration of behavioral parameters of the model based on real data from automated access control systems;

2) integration of additional factors of passenger behavior variability (delays, train misses, sudden route changes);

3) adaptation of the algorithm for operational management and predictive maintenance tasks.

These areas will allow the method to be transferred from modeling on a conventional network to practical application in subway monitoring systems.

7. Conclusions

1. A stochastic model of passenger movement on a directed graph has been built, which, unlike aggregated models, allows for the analysis of individual trajectories under the conditions of depersonalized data. This provides the basis for a detailed assessment of the load at the level of individual runs, which is a key advantage for technical diagnostics tasks.

2. Typical passenger behavior patterns have been formalized in the form of parameterized probability functions of transition between stations. A feature is the integration of several criteria (time, distance, transfers) into a single generalized cost of transition. That has made it possible to make

the stochastic model more realistic since it takes into account the rational motives of passengers when choosing a route.

3. A procedure for assessing the load has been proposed, which is based on the probability distribution of the passenger's "weight" between a set of permissible routes. The key difference from known approaches is the refusal to determine a single most probable path, which makes it possible to circumvent the problem of uncertainty. This approach provides an objective integrated load assessment by averaging over all possible trajectories.

4. A method for calculating the load of each elementary route as the sum of the fractions of passenger weights has been developed. Unlike methods that calculate only general OD flows, this method allows you to obtain detailed data on the load for each specific train on each section. This creates an information base for the transition to adaptive planning of rolling stock maintenance.

5. We have performed simulation that confirmed the operability of the proposed method. Its key feature is the achievement of high accuracy in estimating passenger flows without the need to determine a single route for each passenger, which is explained by the effect of averaging uncertainties on a large data sample (interpretation). According to the results of modeling on a network of 12 stations and 60,000 passengers, the average relative error in the load assessment did not exceed 0.5%.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

The study was conducted without financial support.

Data availability

All data are available, either in numerical or graphical form, in the main text of the manuscript.

Use of artificial intelligence

The authors used artificial intelligence technologies within acceptable limits to provide their own verified data, which is described in the research methodology section.

References

1. Cats, O., Gkiotsalitis, K., Schöbel, A. (2025). 50 years of Operations Research in public transport. EURO Journal on Transportation and Logistics, 14, 100160. https://doi.org/10.1016/j.ejtl.2025.100160
2. Hussain, E., Bhaskar, A., Chung, E. (2021). Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. Transportation Research Part C: Emerging Technologies, 125, 103044. https://doi.org/10.1016/j.trc.2021.103044
3. Sari Aslam, N., Barros, J., Lin, H., Murcio, R., Bei, H. (2024). Alighting location estimation from public transit data: a case study of Shenzhen. Transportation Planning and Technology, 48 (5), 937–952. https://doi.org/10.1080/03081060.2024.2382247
4. Tiam-Lee, T. J., Henriques, R. (2022). Route choice estimation in rail transit systems using smart card data: handling vehicle schedule and walking time uncertainties. European Transport Research Review, 14 (1). https://doi.org/10.1186/s12544-022-00558-x
5. Mo, B., Ma, Z., Koutsopoulos, H., Zhao, J. (2023). Passenger Path Choice Estimation Using Smart Card Data: A Latent Class Approach with Panel Effects Across Days. arXiv. https://doi.org/10.48550/arXiv.2301.03808
6. Nagy, R., Horvát, F., Fischer, S. (2024). Innovative Approaches in Railway Management: Leveraging Big Data and Artificial Intelligence for Predictive Maintenance of Track Geometry. Tehnički vjesnik – Technical Gazette, 31 (4), 1245–1259. https://doi.org/10.17559/tv-20240420001479
7. Ou, Y., Mihăiţă, A.-S., Ellison, A., Mao, T., Lee, S., Chen, F. (2025). Rail Digital Twin and Deep Learning for Passenger Flow Prediction Using Mobile Data. Electronics, 14 (12), 2359. https://doi.org/10.3390/electronics14122359
8. Su, G., Li, P., Lian, D., Mo, P. (2025). A coordinated passenger flow control model for urban rail transit considering willingness to board. Multimodal Transportation, 4 (3), 100225. https://doi.org/10.1016/j.multra.2025.100225
9. Liu, J., Huang, W., Zhang, N., Ma, Z., Qian, Z. (2024). A Data-Driven Multi-Modal Fusion Method for Path Choice Estimation in Urban Rail Systems. https://doi.org/10.2139/ssrn.4994725
10. Wen, D., Lv, H., Yu, H. (2025). Data-Driven Approach for Passenger Assignment in Urban Rail Transit Networks: Insights From Passenger Route Choices and Itinerary Choices. Journal of Advanced Transportation, 2025 (1). https://doi.org/10.1155/atr/6620828
11. Makridis, M. A., Kouvelas, A. (2023). An adaptive framework for real-time freeway traffic estimation in the presence of CAVs. Transportation Research Part C: Emerging Technologies, 149, 104066. https://doi.org/10.1016/j.trc.2023.104066
12. Bernal, E., Rey, Á. (2019). Study of the Structural and Robustness Characteristics of Madrid Metro Network. Sustainability, 11 (12), 3486. https://doi.org/10.3390/su11123486
13. Tian, X., Zheng, B., Wang, Y., Huang, H.-T., Hung, C.-C. (2021). TRIPDECODER: Study Travel Time Attributes and Route Preferences of Metro Systems from Smart Card Data. ACM/IMS Transactions on Data Science, 2 (3), 1–21. https://doi.org/10.1145/3430768