VILNIUS TECH
Vilnius Gediminas Technical University

Proceedings of the 13th International Scientific Conference

ENVIRONMENTAL ENGINEERING

April 23–24, 2026, Vilnius, LITHUANIA

eISSN 2029-7092
eISBN 978-609-476-444-8
Article ID: enviro.2026.2313
https://doi.org/10.3846/enviro.2026.2313

II. ROADS, RAILWAYS AND SMART CITIES *http://vilniustech.lt/enviro*

METHODS OF DETERMINING THE O-D MATRIX IN THE ABSENCE OF DATA SURVEYS

Simonas ČIŽAUSKAS[1], Rasa UŠPALYTĖ-VITKŪNIENĖ[2*]

[1]UAB "Kauno autobusai", Kaunas, Lithuania
[2]Vilnius Gediminas Technical University, Vilnius, Lithuania

Received 18 February 2026; revised 20 February 2026; accepted 22 February 2026

Abstract. Public transport is like a living organism that is constantly changing along with the city and the needs of its residents. Based on the changes in the city, its infrastructure or the habits of its residents, the public transport system must also change. One of the main tools for understanding the public transport in urban areas is passenger origin-destination (OD) matrices. OD matrices are compiled based on mass population surveys, during which the city is divided into zones, and the collected data is used to create travel patterns. Such studies implementation is expensive and repeated infrequently – usually every few or even ten years. Currently, on-board computers collect a significant amount of data about daily citizens' trips, but the main problem is that this data is not enough to create an O-D matrix, because only boarding data is recorded. The second largest city in Lithuania was chosen for the study due to the abundance of data collected. The aim of this study is to develop a methodology based only on boarding data, allowing to reflect the passenger origin-destination matrix. Two methods were later tested during the study: the first, matching sequential entries based on that the same anonymised card identifier, and the second, the time that the passenger gets off at the stop where they get back on. The results of these methods are analyzed and evaluated at the stop and zone level.

Keywords: public transport, origin-destination matrix, passengers' habits, public transport modelling.

1. Introduction

Public transport systems can be viewed as dynamic organisms that continuously evolve together with the city and the mobility needs of its residents. As urban structure, infrastructure, and travel behavior change, public transport systems must adapt accordingly. Therefore, continuous monitoring and evaluation are essential to support timely decisions regarding route configuration, service frequency, and vehicle allocation.

One of the key analytical tools for understanding passenger mobility patterns is the origin–destination (OD) matrix. OD matrices describe the spatial distribution of trips by identifying trip origins and destinations, thereby enabling assessment of travel demand between urban areas and supporting network planning and optimization. Accurate estimation of OD matrices is a fundamental prerequisite for efficient and reliable public transport planning, as it allows forecasting passenger flows and adjusting service supply accordingly.

Traditionally, OD matrices are constructed using household travel surveys, in which cities are divided into zones and collected data are used to develop travel demand models. Although such surveys provide valuable insights, they are expensive and typically conducted at long intervals, often every five to ten years. Moreover, they frequently capture overall population mobility rather than focusing specifically on public transport users, which may limit their applicability for operational public transport planning. Despite the importance of urban transport models for decision-making, there remains a lack of comprehensive guidance on how to effectively integrate available data sources with advanced modelling techniques to obtain reliable demand estimates (Mohammed & Oke, 2023).

Recent studies have explored alternative approaches for OD matrix generation using operational data. Dong et al. (2024) proposed a method based on the relationship between boarding events and travel cost to estimate passenger travel distance and infer trip purpose. However, reliability challenges arise when passengers transfer between routes, particularly in complex bus networks. The accuracy of reconstructed OD matrices may depend on the fare structure applied in a given city. Systems based on time-based or integrated fare schemes tend to provide more consistent trip reconstruction results, whereas single-ride payment systems may generate less reliable data (Tang et al., 2024).

* Corresponding author. E-mail: *rasa.uspalyte@vilniustech.lt*

Another emerging data source for OD estimation involves Wi-Fi sensor data collected from onboard devices. Although such methods offer promising opportunities for passive data collection, limitations remain, as Wi-Fi sensors do not detect all nearby devices and not all passengers carry detectable equipment (Fabre et al., 2025).

In recent years, automated passenger counting and entry–exit data collection systems have been increasingly used to improve OD estimation. However, implementing such systems may involve substantial costs and operational constraints. Evidence from London suggests that link flow estimation is more complex than station exit estimation, and that entry–exit data alone are insufficient to accurately determine transfer flows without additional information, such as average trip distance (Ait-Ali & Eliasson, 2022).

Studies conducted in Mashhad, Iran, demonstrated the application of an improved trip-chaining model based on smart card data for OD matrix reconstruction. In that case, 52.4% of recorded transactions corresponded to trip origins, while the remaining records represented transfers (Radfar et al., 2025). Similar to the Mashhad case, the primary challenge in the Kaunas public transport system is the absence of alighting information. The available dataset includes only boarding ("check-in") records, without corresponding "check-out" events. Therefore, a methodological framework is required to infer passenger destinations and construct OD pairs based solely on boarding data.

The aim of this study is to develop a methodology for constructing an origin–destination matrix for Kaunas city and the adjacent suburban areas served by the urban public transport network, using boarding-only data from an automated fare collection system. Two inference methods are examined: the first method pairs consecutive boarding records using the same anonymised card identifier, while the second assumes that passengers alight at the stop where they subsequently board another vehicle. The results are analysed and compared at both stop and zone levels.

2. Methodology of the research

This study applies a data-driven methodological approach to construct a passenger origin–destination (OD) matrix based exclusively on boarding records obtained from an automated fare collection system. The underlying assumption of the research is that passenger trip sequences can be reconstructed using an anonymised card identifier and the chronological order of boarding events, applying the trip chaining principle.

Similar approaches have been widely discussed in the literature when automated fare collection systems do not record alighting events. Studies using smart card data suggest that behavioural assumptions can be applied to infer passenger destinations and construct OD matrices (Alsger et al., 2019; Nassir et al., 2021; Wang et al., 2022). Passenger journey reconstruction methods integrating fare transaction data with vehicle movement information have been successfully implemented in large metropolitan areas (Gordon et al., 2018).

The study area includes Kaunas city and the adjacent suburban territories of Kaunas district that are served by routes operated by UAB "Kauno autobusai." More distant settlements within Kaunas district, primarily served by regional transport operators, are excluded from the analysis. Thus, the research focuses on an urban public transport system with its immediate influence zone, where a unified electronic ticketing system and consistent data collection procedures are applied.

The selection of the study area is justified by data coverage and quality. The electronic ticketing system "Žiogas" records more than 90% of all passenger boarding events, and data have been collected consistently under identical technical conditions for several years. This provides a reliable empirical foundation for OD matrix construction and seasonal mobility analysis.

The research is conducted in three stages: (1) preparation and structuring of boarding data within a database environment; (2) application of two OD reconstruction methods; and (3) aggregation of the resulting matrices at stop and zone levels, followed by spatial analysis.

2.1. Data

This study uses boarding records obtained from the electronic ticketing system "Žiogas," operated by UAB "Kauno autobusai." The system registers every ticket validation performed by passengers onboard public transport vehicles and generates a structured digital record for each boarding event.

The structure of the dataset is presented below in Table 1.

Table 1. The structure of the dataset

Field	Description
Date	Boarding date and time (timestamp)
ID	Anonymised passenger identifier
Line	Route number
Reisas	Scheduled trip number
Reiso_id	Unique trip identifier
Stop_id	Stop identifier

This information enables precise identification of the passenger boarding location, timestamp, and travel context.

The analysis is based on data from two selected periods shown in the Table 2.

Data Representativeness

Electronic boarding records account for more than 90% of all recorded trips, providing a highly representative

Table 2. The main data from the from two selected periods

Month	Electronic boardings ("Žiogas")	Total boardings	Share of electronic boardings (%)
April 2024	3 633 345	3 916 506	92.74
September 2025	4 053 090	4 334 786	93.50

dataset for OD matrix construction. The remaining share consists of single-ride tickets validated onboard and sold directly by drivers. These records are not suitable for OD analysis, as they do not allow identification of alighting locations and represent only 6–8% of total boardings.

Due to the high coverage rate and the consistent data collection procedures applied over several years, the "Žiogas" electronic ticketing system provides a robust empirical basis for reconstructing trip origin–destination pairs at both stop and zone levels.

2.2. Tools and analytical environment

The processing of boarding data and the construction of the origin–destination (OD) matrix were carried out using a sequence of specialised analytical tools, covering data processing, geographic analysis, and visualisation.

A relational SQL database environment was used as the primary platform for initial data processing. Within the SQL environment, the following tasks were performed: import of monthly datasets (approximately 3–4 million records per month), data cleaning (removal of duplicates, correction of erroneous records, handling of missing values), grouping of records by anonymised passenger identifier and chronological order, generation of OD pairs using both reconstruction methods, implementation of the initial stop-to-stop OD logic, and aggregation of results at stop and zone levels. SQL served as the central analytical component, enabling efficient processing of large-scale passenger flow data (Pelletier et al., 2011; Zhao et al., 2024).

The reconstructed OD records were subsequently integrated with geographic datasets, including the GTFS stop database, stop coordinates, and route information (linking stop_id to latitude and longitude). This integration allowed the OD matrix to be represented not only in tabular form but also spatially, through maps, flow diagrams, and other visual representations.

Further spatial analysis was conducted using ArcGIS Pro. The software was applied for spatial joins between stops and zoning layers, assignment of stops to selected territorial units, grouping of stops (e.g., by urban districts, or functional zones), and cartographic interpretation of OD flows. ArcGIS Pro enabled visual assessment of spatial flow concentration, directional patterns, and territorial differences.

Finally, PTV Lines was used for network-level visualisation and interpretation of OD results. The software facilitated zonal OD matrix representation, identification of major flow corridors, visualisation of network load patterns, and evaluation of functional route roles (e.g., trunk–feeder structure). Through this process, OD matrices were translated into transport network models and graphical outputs suitable for planning and decision-making purposes.

2.3. Methods

Two passenger trip pairing methods are applied in this study to identify boarding and inferred alighting locations based solely on boarding data. Both methods are analysed within daily boundaries (Date), as passenger travel behaviour across different days is assumed to be independent.

First method – two-trips-per-day approach

The first method applies strict filtering: only passengers who perform exactly two boarding events within a single calendar day are considered. This method is regarded as the most accurate, as it does not require additional behavioural assumptions regarding trip chaining.

Let us assume that for a given passenger (ID) two boarding events are recorded within one day:

$$S=\left\{\left(stop_1,t_1\right),\left(stop_2,t_2\right)\right\}\text{, where } t_1<t_2. \quad (1)$$

Under this condition, the following OD pairs are defined.

The first trip is constructed as:

$$OD_1=\left(stop_1\rightarrow stop_2\right). \quad (2)$$

It is assumed that the passenger starts the journey at $stop_1$ and finishes at $stop_2$. Since these are the only two boarding events recorded on that day, it is logically inferred that the passenger returns from $stop_2$ to $stop_1$. This assumption typically corresponds to a home–work–home travel pattern.

Therefore, the second *OD* pair is defined as:

$$OD_2=\left(stop_2\rightarrow stop_1\right). \quad (3)$$

The resulting *OD* set for this method is:

$$OD=\left\{\left(stop_1\rightarrow stop_2\right),\left(stop_2\rightarrow stop_1\right)\right\}. \quad (4)$$

Only passenger-day combinations with exactly two boarding events are selected from the dataset.

Advantages of the first method:
- High accuracy without additional behavioural assumptions,
- Symmetrical OD pairs reflecting bidirectional flows,
- Suitable as a reference dataset for validating other methods.

Limitations of the first method:
- Includes only passengers performing exactly two trips per day,

- Does not reconstruct the full city-wide OD structure,
- Requires a complementary method to capture all flows.

Second method – trip chaining approach

The second method is applied to the same passenger identifier (ID) but, unlike the first method, includes all boarding events recorded for that passenger within a single day. This approach enables reconstruction of the maximum possible number of trips from the available dataset and serves as the primary method for constructing the full OD matrix.

Passenger trips within one day are treated as a chain, where each trip begins at the stop of the previous boarding and ends at the stop of the subsequent boarding.

If a passenger (ID) has the following sequence of boarding events during one day:

$$S = \{(stop_1, t_1), (stop_2, t_2), \ldots, (stop_n, t_n),\ \text{where}\ t_1 < t_2 < \ldots < t_n \qquad (5)$$

then the following *OD* pairs are constructed:

$$\begin{aligned} &OD_1 = (stop_1 \to stop_2);\\ &OD_2 = (stop_2 \to stop_3);\\ &\ldots\\ &OD_{n-1} = (stop_{n-1} \to stop_n). \end{aligned} \qquad (6)$$

This procedure reconstructs the complete sequence of daily movements. Since boarding-only data do not include alighting information, the destination of the final trip cannot be directly observed. Therefore, a closing assumption is introduced:

$$OD_n = (stop_n \to stop_1). \qquad (7)$$

This assumption creates a closed daily cycle, reflecting a typical travel pattern in which passengers begin and end their day at the same location.

The complete daily OD sequence can therefore be expressed as:

$$\begin{aligned} &(stop_1 \to stop_2);\\ &(stop_2 \to stop_3);\\ &\ldots\\ &(stop_{n-1} \to stop_n);\\ &(stop_n \to stop_1). \end{aligned} \qquad (8)$$

Advantages of the second method:

- Includes all passenger trips without strict filtering,
- Enables reconstruction of the full daily trip chain,
- Suitable for city-wide OD matrix construction,
- Particularly effective when electronic boarding coverage exceeds 90%.

Limitations of the second method:

- Assumes that passengers start and end their day at the same stop, which may not always hold,
- May be less accurate for non-cyclical travel patterns,
- Results should be validated using the first method.

3. Results of the research

This section presents the empirical results obtained from constructing passenger origin–destination (OD) matrices using two different reconstruction methods applied to two separate periods: April 2024 and September 2025. The first stage evaluates the two-trips-per-day method, which considers only passengers performing exactly two boarding events within a single day. This behavioural pattern allows precise identification of both trip origins and destinations. The second stage applies the trip chaining method, enabling reconstruction of all daily passenger movements.

3.1. OD matrix using the first method

In April 2024, the electronic ticketing system "Žiogas" recorded 3,633,345 electronic boardings, representing 92.74% of total monthly boardings. The objective of the first method is to identify passengers who performed exactly two trips within a single day, thereby enabling precise reconstruction of forward and return trip directions. The dataset was grouped by passenger ID and travel date, and only passenger-day combinations with exactly two boarding events were selected. This filtering procedure revealed that more than 650,000 passenger-day combinations met the defined criterion, corresponding to 1,305,688 boarding records (two per passenger-day pair). Consequently, approximately 36% of all electronic boardings in April followed a two-trip-per-day pattern, consistent with a typical weekday mobility structure.

For each selected passenger-day combination, the first boarding event was interpreted as trip A → B and the second as B → A, since only two boarding events were recorded within that day. This approach resulted in two symmetrical OD pairs for each passenger-day observation. The final April 2024 dataset therefore consists of 1,305,688 complete stop-to-stop OD pairs, all with clearly defined origins and inferred destinations. Due to the absence of additional behavioural assumptions, this dataset represents one of the most reliable OD reconstructions derived from boarding-only data and is subsequently used as a reference baseline for evaluating the second method.

In September 2025, 4,053,090 electronic boardings were recorded, approximately 420,000 more than in April 2024. Applying the same filtering logic, 720,604 passenger-day combinations with exactly two trips were identified, corresponding to 1,441,208 reconstructed OD records. This represents approximately 35–36% of total monthly boardings, a proportion nearly identical to that observed in April. The stability of this share suggests that the two-trip-per-day behavioural pattern is largely independent of seasonal variation or short-term fluctuations in service levels. The September dataset includes 1,441,208 complete OD pairs, reflecting an approximately 11% increase compared to April, consistent with the overall growth in boarding volume. This dataset is used for seasonal comparison and for validation of the trip chaining method.

3.2. OD matrix construction using the trip chaining method

The second method, referred to as the trip chaining approach, enables reconstruction of all passenger trip sequences performed within the same day, regardless of the total number of boarding events. Unlike the first method, which includes only passengers performing exactly two trips per day, the trip chaining method incorporates all passengers with at least two boarding events within a single day. This ensures reconstruction of the full public transport trip network based on chronological boarding sequences.

For April 2024, passenger-day combinations containing at least two boarding events were first identified, as a minimum of two boardings is required to generate at least one origin–destination pair. The aggregation process revealed 1,132,260 passenger-day combinations meeting this condition. This represents a substantially broader sample compared to the first method, which identified 652,844 passenger-day combinations with exactly two trips. Consequently, the trip chaining method captures a significantly larger share of overall travel activity.

For each selected passenger-day combination, boarding events were sorted chronologically by date and time. Each boarding was treated as the origin of a trip, while the subsequent boarding within the same day was interpreted as the destination of the previous trip. For example, if a passenger's boarding sequence during a single day was A → B → C → D, the reconstructed OD pairs would be A → B, B → C, and C → D. This procedure reflects the natural progression of passenger movements and allows reconstruction of the entire daily travel sequence.

Since boarding-only data do not include alighting events, the final trip of the day lacks a directly observable destination. Therefore, a commonly applied behavioural assumption was adopted: the passenger is assumed to return to the initial boarding location at the end of the day. Thus, for the sequence A → B → C → D, an additional OD pair D → A is generated, forming a closed daily trip chain. This assumption is particularly consistent with typical weekday mobility patterns, where passengers return to their point of origin.

Applying the trip chaining method to April 2024 resulted in a total of 3,202,355 reconstructed OD pairs. Each pair contains a clearly defined origin and inferred destination, and all trips were generated according to the chronological movement sequence principle. Compared to the first method, this approach enables reconstruction of the complete monthly passenger mobility structure, independent of the number of trips performed per day.

The same methodology was applied to September 2025 in order to evaluate the effect of seasonality and increased travel demand during the autumn period. In September, 1,256,345 passenger-day combinations with at least two boarding events were identified, representing an increase of approximately 11% compared to April (1,132,260). This growth corresponds to the typical rise in travel activity associated with the beginning of the academic year and full work-season dynamics.

Using the same trip linking logic, 3,557,963 complete OD pairs were reconstructed for September 2025. Compared to April, this represents an increase of 11.1% (from 3,202,355 to 3,557,963), reflecting higher passenger mobility levels. The September dataset indicates that a larger proportion of passengers performed three or more trips per day, meaning that the trip chaining method captures an even broader spectrum of actual travel behaviour during the autumn period. The overall stability of the proportional growth between boarding volumes and reconstructed OD pairs suggests that the method remains robust and consistent across seasonal variations.

The September OD dataset represents one of the most detailed data blocks within this study and provides a strong empirical basis for analysing seasonal and structural mobility patterns within the Kaunas public transport system.

3.3. Aggregation of stops into territorial zones using spatial join

In public transport analysis, stops represent the most detailed spatial unit, comprising approximately 1,000 point locations within Kaunas city and its surrounding service area. Although this level of granularity enables precise identification of micro-level passenger movements, it is not suitable for analysing broader mobility patterns or structural travel flows. Therefore, in practical applications, stops are aggregated into larger territorial zones.

In this study, zones are defined as spatial units to which stops can be assigned for analytical purposes. These may include traffic analysis zones (TAZ), administrative districts (such as municipal subdivisions used in this case), neighbourhood units, functional areas (e.g., residential, employment, or service zones), buffer-based influence areas (e.g., 500-metre catchment zones around stops), or any other geographic partition relevant to the research objectives. The choice of zoning system does not alter the OD reconstruction methodology itself but affects only the level of spatial aggregation and interpretation.

Stop-to-zone assignment was performed using a spatial join procedure in ArcGIS Pro. In this process, the stop layer (point features) was spatially intersected with the selected zone layer (polygon features), ensuring that each stop was attributed to the corresponding territorial unit. This procedure enabled transformation of stop-level OD matrices into zone-level OD matrices suitable for strategic mobility analysis and network planning.

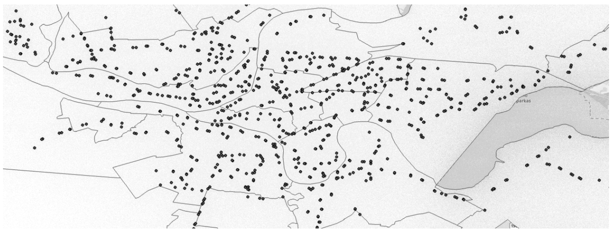

Figure 1. Public transport stops (points) and boundaries of municipal districts (polygons)

The Figure 1 demonstrates how public transport stops are distributed across different zones and justifies the need for spatial aggregation.

Each OD pair (from all four constructed matrices) is supplemented with zonal attributes:

- *origin_zone* – the zone in which the origin stop is located,
- *destination_zone* – the zone in which the destination stop is located.

After aggregation, a zone-to-zone OD matrix is obtained. This enables:

- identification of the main inter-zonal passenger flows,
- analysis of the functional structure of the city,
- comparison of the applied methods (pair-based vs. trip-chaining),
- visualization of flows in PTV Lines or other transport modelling environments.

Regardless of whether municipal districts, TAZ zones, or another zoning system is applied, the aggregation logic remains identical. Therefore, the methodology is fully transferable to different spatial classification systems.

3.4. Visualization of OD flows

After constructing origin–destination (OD) matrices at both stop and zone levels, the results were visualized using the PTV Lines platform. This software enables spatial representation of inter-zonal connections and allows the intensity, direction, and distribution of passenger flows to be analysed across the entire Kaunas public transport network.

The zonal OD matrix was imported into PTV Lines as:

- a zone layer (municipal district boundaries or any other selected territorial unit layer),
- a directional flow layer (origin → destination pairs),
- an intensity attribute (number of trips).

PTV Lines automatically generates a spatial visualization in which:

- zones are displayed as coloured polygons,
- inter-zonal flows are represented as connecting lines,
- line thickness and colour correspond to the number of trips.

Figures 2–5 allows immediate identification of the main mobility corridors and enables comparison of differences among all four datasets generated in the study.

In total, four zonal OD matrices were constructed and visualized:

Figure 2. Visualization of zonal trip flows for April 2024 (two-trip pairing method)

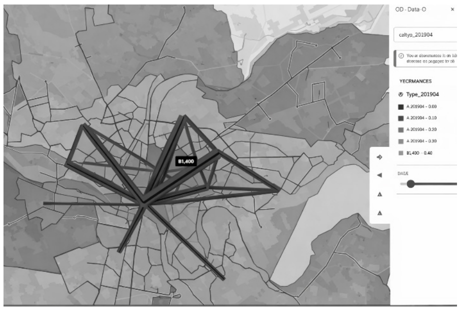

Figure 3. Visualization of zonal trip flows for September 2025 (two-trip pairing method)

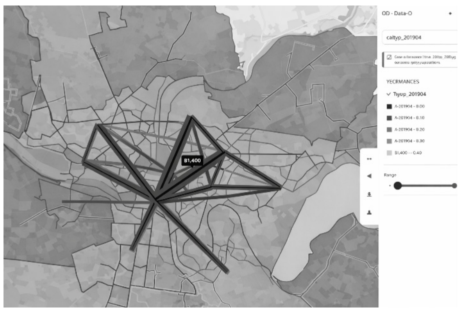

Figure 4. Visualization of zonal trip flows for April 2024 (trip chaining method)

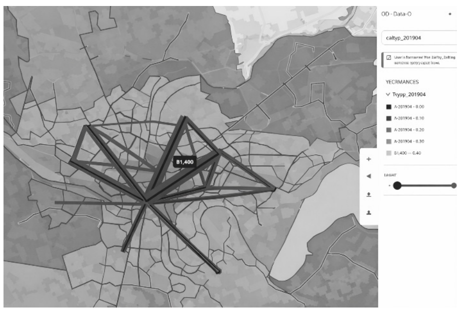

Figure 5. Visualization of zonal trip flows for September 2025 (trip chaining method)

In these figures, the following patterns can be observed:

- the main connections between zones generating the highest travel flows, a pronounced dominance of the city centre as both an origin and a destination.

The visualisations produced in PTV Lines provide an additional analytical layer to the study:

- they enable rapid identification of the main mobility corridors,
- help to interpret the intensity of inter-zonal travel flows,
- illustrate seasonal differences between April and September,
- and allow comparison of the results obtained using the two-trip pairing and trip chaining methods.

The visualisation stage completes the entire OD reconstruction process:

1. boarding data,
2. stop-level OD matrix,
3. zone-level OD matrix,
4. spatial visualisation.

Through this process, the analysis of all four datasets becomes fully interpretable in a spatial context, and the results can be applied both in academic research and in practical transport planning and decision-making.

4. Conclusions

1. Based on the literature review, traditional origin–destination (OD) matrix construction methods relying on household travel surveys are costly, infrequently updated, and do not always accurately reflect the actual movement of public transport passengers. Therefore, data from automated fare collection systems represent a promising and operational alternative for urban mobility analysis.
2. The results of the study demonstrate that even in the absence of alighting ("check-out") data, it is possible to construct a reliable OD matrix by applying trip pairing and trip chaining methods based on anonymised card identifiers and chronological boarding sequences.
3. The first method, applied to passengers who perform exactly two trips per day, enables the construction of a highly accurate and symmetrical OD dataset, which can serve as a reference base for evaluating the reliability of more complex reconstruction methods.
4. The trip chaining method allows reconstruction of the complete daily travel sequence and the formation of a comprehensive city-level OD matrix. In April 2024, a total of 3,202,355 OD pairs were generated, while in September 2025, 3,557,963 OD pairs were formed, confirming the stability and applicability of the method across different seasons.
5. The electronic ticketing system "Žiogas", covering more than 90% of all passenger boardings, provides a sufficiently representative and statistically robust dataset for analysing the structure of urban public transport without the need for additional large-scale surveys.
6. Aggregation of stops into territorial zones and spatial analysis of OD matrices enable the identification of major mobility corridors, the dominance of the city centre, and seasonal variations in travel flows, thus providing a practical basis for route network optimisation and strategic planning.
7. The integration of GPS and electronic ticketing technologies makes it possible not only to collect large volumes of passenger movement data, but also to use them analytically for decision-making. Although such analysis does not replace comprehensive population-wide mobility surveys, it ensures timely, data-driven insights that are highly valuable for transport operators and municipalities when planning public transport development.

References

Ait-Ali, A., & Eliasson, J. (2022). The value of additional data for public transport origin–destination matrix estimation. *Public Transport*, *14*, 419–439. https://doi.org/10.1007/s12469-021-00282-0

Alsger, A., Assemi, B., Mesbah, M., & Ferreira, L. (2019). Validating and improving methods for estimating transit passenger destination locations using smart card fare data. *Transportation Research Part C*: *Emerging Technologies*, *104*, 180–198. https://doi.org/10.1016/j.trc.2019.05.003

Dong, N., Li, T., Liu, T., Tu, R., Lin, F., Liu, H., & Bo, Y. (2024). A method for short-term passenger flow prediction in urban rail transit based on deep learning. *Multimedia Tools and Applications*, *83*(22), 61621–61643. https://doi.org/10.1007/s11042-023-14388-z

Fabre, L., Bayart, C., Kone, Y., Manout, O., & Bonnel, P. (2025). A machine learning approach to estimate public transport ridership using Wi-Fi data. *IEEE Transactions on Intelligent Transportation Systems*, *26*(1), 906–915. https://doi.org/10.1109/TITS.2024.3485802

Gordon, J. B., Koutsopoulos, H. N., Wilson, N. H. M., & Attanucci, J. (2018). Automated inference of transit passenger journeys using fare transaction and vehicle location data. *Transportation Research Part C*: *Emerging Technologies*, *96*, 289–307. https://doi.org/10.1016/j.trc.2018.09.012

Mohammed, M., & Oke, J. (2023). Origin-destination inference in public transportation systems: A comprehensive review. *International Journal of Transportation Science and Technology*, *12*(1), 315–328. https://doi.org/10.1016/j.ijtst.2022.03.002

Nassir, N., Khani, A., & Hickman, M. (2021). Trip chaining and destination inference in automated fare collection systems. *Transportation Research Record*, *2675*(9), 1–12.

Pelletier, M.-P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transport: A literature review. *Transportation Research Part C: Emerging Technologies*, *19*(4), 557–568. https://doi.org/10.1016/j.trc.2010.12.003

Radfar, S., Koosha, H., Gholami, A., & Amindoust, A. (2025). Improved public transport OD matrix estimation using an enhanced trip chain model with smart card data. *International Journal of Intelligent Transportation Systems Re-*

search, *23*, 1341–1356. https://doi.org/10.1007/s13177-025-00513-9

Tang, T., Mao, J., Liu, R., Liu, Z., Wang, Y., & Huang, D. (2024). Origin-destination matrix prediction in public transport networks: Incorporating heterogeneous direct and transfer trips. *IEEE Transactions on Intelligent Transportation Systems*, *25*(12), 19889–19903. https://doi.org/10.1109/TITS.2024.3447611

Wang, H., Chen, C., Ma, J., & Wang, Y. (2022). Inferring public transport origin–destination matrices from smart card data with limited information. *Transportation Research Part C*: *Emerging Technologies*, *135*, Article 103512.

Zhao, D., Mihăiţă, A. S., Ou, Y., Grzybowska, H., & Li, M. (2024). Origin-destination matrix estimation for public transport: A multi-modal weighted graph approach. *Transportation Research Part C*: *Emerging Technologies*, *165*, Article 104694. https://doi.org/10.1016/j.trc.2024.104694