

A Literature Review of Public Transport OD Matrix Estimation

Joan Buralat ^{1,2}, Gael Pallares ^{1,*}, Myriam Foucras ^{3,*} and Yohan Dupuis ⁴

¹ CESI, CESI LINEACT, Montpellier, Immeuble le Quatrième Zone aéroportuaire de Montpellier Méditerranée 34130 Muguio, France

² ENSAM, Paris, 151 Boulevard de l'Hôpital, 75013 Paris, France

³ CESI, CESI LINEACT, Toulouse, 16 Rue Magellan 31670 Labège, France

⁴ CESI, CESI LINEACT, Paris, 1 Avenue du Général de Gaulle 92800 Puteaux, France; ydupuis@cesi.fr

* Correspondence: gpallares@cesi.fr (G.P.); mfoucras@cesi.fr (M.F.)

Abstract

Origin–Destination matrices (ODms) are a fundamental input for public transport planning and optimization, as they characterize travel demand across a network. Traditionally estimated from user surveys, ODms are now increasingly inferred from large-scale automatically collected data, such as Automated Fare Collection (AFC), Automated Passenger Counting (APC), and Automated Vehicle Location data (AVL). This review focuses on the reconstruction of static ODms in public transport systems, while accounting for studies that exploit dynamic or short-term observations when these are used to infer static or quasi-static demand patterns. We provide a transversal synthesis of OD estimation approaches by jointly analyzing data sources, modeling assumptions, uncertainty handling, and validation strategies. A structured comparative table summarizes representative case studies across different data contexts, objectives, and methodological families. Beyond a descriptive overview, this review identifies key research gaps, including the lack of uncertainty-aware benchmarking frameworks, the limited propagation of uncertainty across modeling stages, and the strong dependence of reported performance on data quality and validation references. These findings highlight that OD estimation performance is context-dependent and that methodological choices should be aligned with data availability, modeling objectives, and acceptable assumptions rather than with reported accuracy alone.

Keywords: OD matrix estimation; public transportation; performance indicators; operational validation; flow estimation

1. Introduction

1.1. Context and Motivation

Efficient management and optimization of public transport systems require a detailed understanding of passenger flows. These flows are commonly represented through Origin–Destination matrices (ODms), which quantify flows between boarding and alighting locations and serve as a fundamental input for network design, service planning, and operational analysis. However, obtaining accurate OD information remains challenging in practice, as direct field surveys are costly, time-consuming, and often limited in spatial and temporal coverage. To overcome these limitations, a growing body of research has focused on estimating ODms from partial or indirect data sources. Traditional household travel surveys are increasingly complemented or replaced by automatically collected

Academic Editor: Xiaobao Yang

Received: 16 December 2025

Revised: 5 February 2026

Accepted: 9 February 2026

Published: 12 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

data, such as automated fare collection (AFC), automated passenger counting (APC), vehicle location (AVL), and household travel survey (HTSD) data. While these sources provide unprecedented data volume and temporal resolution, they also introduce new challenges related to noise, bias, incomplete observations, and uncertainty. As a result, ODM reconstruction has become a complex problem involving data fusion, behavioral assumptions, methodological choices, and validation constraints.

Numerous OD reconstruction methods have been proposed, ranging from trip-chain models and statistical inference techniques to clustering-based analyses and probabilistic approaches. Existing review papers have provided comprehensive overviews of these methods, typically classifying them by data sources or algorithmic families. However, despite this growing literature, reported performance levels vary widely across studies, and it remains difficult to interpret these differences or to determine under which conditions specific approaches are more appropriate. In particular, the respective roles of data quality, modeling assumptions, uncertainty handling, and validation references in shaping reported performance are rarely analyzed in a unified manner.

This review addresses this gap by providing a transversal qualitative synthesis of static ODM reconstruction methods in public transport systems. While the focus is on static ODms as modeling objects, the review includes studies that exploit dynamic or short-term observations when these are used to infer static or quasi-static demand patterns. Rather than ranking methods based on reported accuracy alone, the review explicitly links data availability, modeling assumptions, uncertainty profiles, and validation strategies in order to explain observed performance differences across studies.

In literature, the terms estimation, inference, prediction, and reconstruction are sometimes used interchangeably, although they refer to distinct problem settings in the context of OD matrix analysis. In this review, estimation refers to numerical procedures that assign unknown OD flows or alighting locations by optimizing a cost function under given constraints. Inference denotes probabilistic approaches that infer latent variables (e.g., alighting stops or OD flows) as random variables, typically through posterior distributions conditioned on observed data and behavioral assumptions. Prediction is used when models are trained to forecast future or unseen OD patterns based on historical observations, often in a short-term or dynamic setting. Finally, reconstruction refers to the overall process of assigning missing trip components (such as alighting locations) and aggregating individual trips into a consistent OD matrix, regardless of whether estimation, inference, or prediction techniques are used.

Throughout this paper, the term OD matrix reconstruction is used as an umbrella concept encompassing these complementary approaches, while distinctions are made explicit when necessary.

In open systems, only boarding events are observed, while alighting locations are latent. Numerical estimation methods assign alighting stops by optimizing a cost function, probabilistic approaches infer their posterior distribution given behavioral assumptions, and learning-based models predict them from observed patterns. Once alighting stops are assigned, regardless of the underlying mechanism, the individual trip chain is reconstructed and contributes to the reconstructed ODM.

Based on this analysis, the review seeks to answer the following research questions:

- How do different data sources influence the types of uncertainty affecting OD estimation?
- How do modeling assumptions and validation references shape reported performance?
- Under which data and modeling conditions are specific OD estimation approaches more suitable than others?

1.2. Review Protocol

Review protocol and study selection: This review follows a PRISMA-inspired workflow to ensure transparency and reproducibility. Literature was retrieved using Google Scholar as the primary search engine (chosen for its broad coverage of journal articles, conference proceedings, and theses in transport engineering). Searches were performed between 2010 and 2025, restricted to publications in English, to reflect the widespread adoption of AFC/AVL systems and standardized supply formats such as GTFS.

Search strategy: The search relied on a set of predefined keyword combinations targeting ODM reconstruction in public transport. The exact search strings are reported in Figure 1, e.g., («origin-destination» OR «O-D matrix» OR «OD estimation») AND («public transport» OR «transit» OR «metro» OR «bus») AND («static» OR «single period»).

And the other string was («origin-destination» OR «O-D matrix» OR «O-D estimation») AND («uncertainty» OR «error propagation» OR «imprecision») AND («public transport» OR «transit» OR «metro» OR «bus»).

Backward referencing (reference lists of included papers) and forward snowballing (papers citing key references) were used to reduce omission risk.

Inclusion criteria: Studies were included if they:

- Addressed ODM reconstruction or reconstruction for public transport systems (bus/metro/tram);
- Used survey data (HTSD), automated data collection systems (AFC/APC/AVL) or open supply data (GTFS) as inputs;
- Provided either a methodological contribution, a case study, or a validation framework relevant to ODM reconstruction.

Exclusion criteria: Studies were excluded if they:

- Focused exclusively on private vehicle traffic without public transport applicability;
- Addressed only general travel demand modeling without OD reconstruction;
- Were non-peer-reviewed or lacked sufficient methodological detail, or (iv) duplicated results already reported in a more complete version.

Screening and final corpus. Figure 1 summarizes the literature selection process. The initial search returned 9860 records for the first query string and 1640 records for the second one. While this review primarily focuses on studies published after 2010, a limited number of earlier works were retained when they constitute seminal methodological references that have been widely adopted as baselines or foundational frameworks in subsequent research.

A citation threshold of 150 citations was used solely as a pragmatic indicator of sustained community uptake to identify such foundational contributions. This threshold corresponds to approximately 10 citations per year, reflecting continuous scholarly impact rather than short-term visibility. Applied to the 1680 initially retrieved records, this criterion was used selectively to identify a small set of methodological cornerstones, and not to privilege older or mainstream studies.

Importantly, this citation-based criterion was not used to prioritize results in the synthesis or to rank methods, but only to justify the inclusion of a limited number of pre-2010 references that remain relevant for interpreting post-2010 developments. To mitigate potential bias toward older literature, the core synthesis and gap analysis are exclusively based on the post-2010 corpus, while pre-2010 studies are used solely for historical context and methodological grounding.

Among the pre-2010 references, two categories were considered. First, a small number of review articles were selected when they provided sufficiently general perspectives on data sources or methodological families relevant to ODM reconstruction. Second, a limited set of case studies on ODM reconstruction was included. In these cases, selection

was based on the relevance of the methodological assumptions, the clarity of the reconstruction framework, and the interpretability of the reported results, rather than on reported accuracy alone.

While several recent reviews have provided comprehensive overviews of public transport ODM reconstruction methods, they primarily adopt descriptive classifications based on data sources or algorithmic families. The distinguishing contribution of this study lies in its transversal analytical perspective, which explicitly connects data availability, modeling assumptions, uncertainty handling, validation references, and operational applicability across methods. Rather than proposing performance-based rankings, this review explains observed differences in reported accuracy through method–data–objective alignment and derives specific, method-driven research gaps. This approach provides a critical synthesis that complements existing reviews and offers a structured framework for interpreting and selecting OD estimation methods in practice.

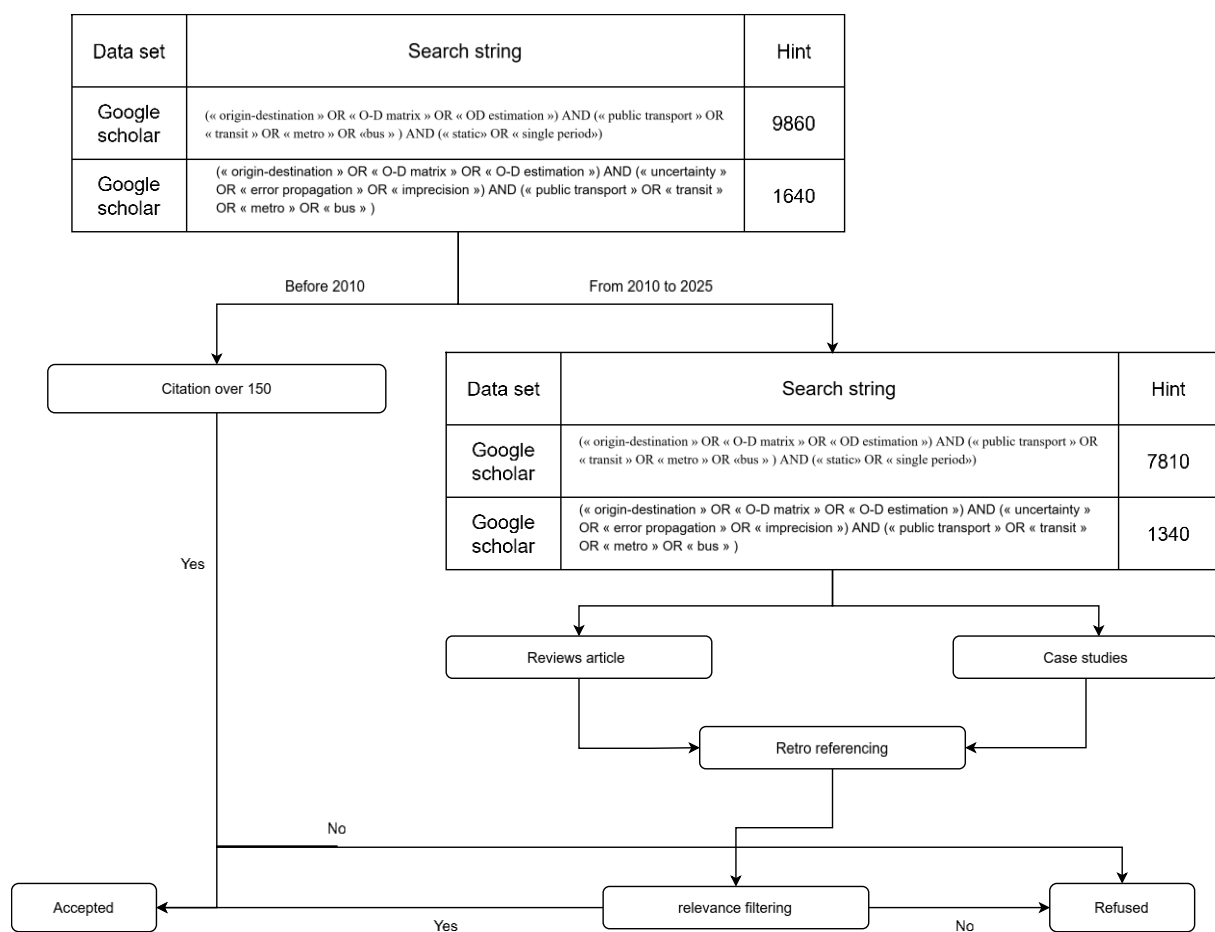


Figure 1. PRISMA diagram.

The review article is structured as follows: In the first part, we present the data. In the second part, the methods and models for constructing the ODM are presented. The third section zooms in on the output data and performance indicators, both operational from the perspective of the user and the network operator. In the final section, we conclude and present the remaining challenges.

2. Public Transport Data

In this article, we focus on the reconstruction of origin-destination matrices for public transport (bus, metro, tram). We divided data sources into three categories: data provided by surveys, Automatic Data Collection Systems (ADCS), and open-access data.

Figure 2 summarizes the data collection and processing pipeline commonly involved in public transport ODms reconstruction. Data are generated at the passenger, vehicle, and network levels and provide partial and complementary observations of travel demand. ODM are therefore not directly observed but inferred by combining these heterogeneous sources through modeling assumptions and estimation methods described in Section 3.

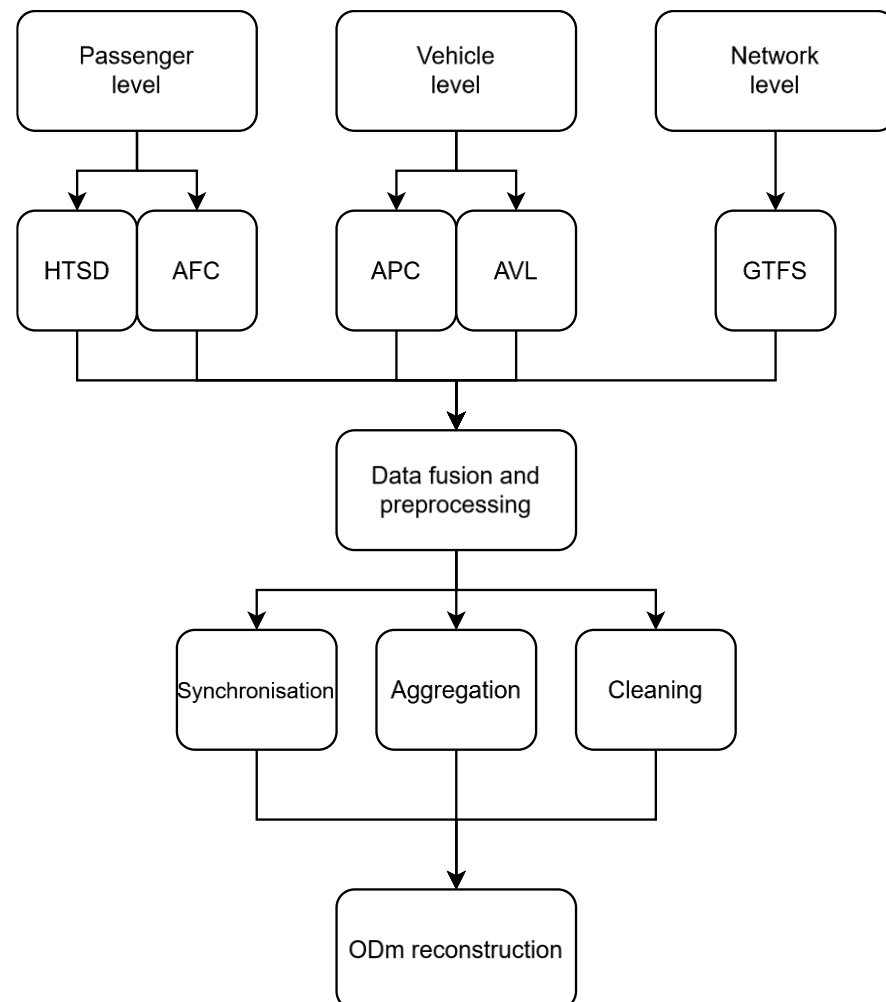


Figure 2. Illustration of the data collection and processing pipeline for public transport OD matrix reconstruction.

2.1. Surveys

Household Travel Survey Data (HTSD) from field surveys are very valuable because they closely reflect on-the-ground reality and are therefore highly representative. However, they have significant disadvantages: high financial and time costs, the need for a sufficient number of participants to be meaningful, and a focus on metropolitan areas and fixed periods.

2.2. Data from ADCS

Several types of data are collected automatically by different ADCS systems:

- APC (Automated Passenger Counting);
- AFC (Automated Fare Collection);
- AVL (Automated Vehicle Location).

APC data corresponds to passenger counting, for example, using infrared sensors at bus entrances. APC data have the advantage of directly measuring the actual number of passengers boarding and alighting [1–4] but the disadvantage of not knowing the trip origin. This type of data is difficult to use alone for ODm reconstruction.

AFC data are derived from ticketing: booking of soft mobility options such as electric bikes, or validating bus, metro, or tramway tickets. However, this data may vary depending on the validation system (smart card, smartphone, etc.). Additionally, the information gathered varies depending on the type of network being studied. In a closed system (where the fare depends on the trip distance), both boarding and alighting information are collected. In open networks (where ticket prices are fixed), only boarding information is retrieved.

An increasing number of public transport vehicles are equipped with sensors such as odometers or Global Navigation Satellite Systems (GNSS) receivers that allow vehicles to be geolocated in real time. AVL data therefore contains the location of a fleet of vehicles. When possible, AFC and AVL data are combined, but unfortunately, due to inaccuracies and poor synchronization, the deduction of stops is not always reliable.

2.3. Open Data

In conjunction with automatically collected data, open-access datasets such as GTFS can be used to describe the public transport supply [5]. The GTFS format provides a standardized representation of network topology, routes, stops, and scheduled timetables, combining spatial and temporal information on the planned service. As such, GTFS data are frequently used as a structural reference to locate stops, reconstruct routes, or support data fusion with AFC, APC, or AVL sources [2,6–10]. However, GTFS represents scheduled rather than realized operations and may differ from actual vehicle trajectories due to delays, disruptions, short turns, or operational adjustments. Consequently, while GTFS is valuable for network representation and consistency checks, it does not constitute ground truth for passenger movements or service performance and should be interpreted with caution when used for ODm reconstruction.

Overall, the literature shows that data availability largely dictates OD estimation strategies. While AFC data dominate recent studies due to their volume and temporal resolution, they systematically introduce destination uncertainty and user-selection bias. Survey-based data remain essential for validation but are increasingly impractical at scale, leading to hybrid approaches that trade representativeness for coverage.

3. Methodology for OD Matrix Reconstitution

The ODm reconstruction can be performed at different spatial and temporal scales. Matrices can be estimated at various spatial scales: macroscopic (city level), mesoscopic (neighborhood level), or microscopic (station level). Generally, this estimation is conducted over time scales long enough to exceed the average duration of a trip. This approach helps mitigate transitional and boundary effects.

Rather than providing a performance-based ranking of OD estimation methods, this section adopts a thematic perspective structured around modeling paradigms and their associated trade-offs. The reviewed methods are compared in terms of data requirements, modeling assumptions, interpretability, scalability, robustness to data imperfections, and operational applicability. This approach reflects the fact that OD estimation performance is inherently context-dependent and that no single method dominates all data configurations and planning objectives.

3.1. The Four Step Model

The Four Step Model (FSM) is a classical model used to predict transport demand in urban planning. The study area is delineated by a boundary, with the inner part divided into different Traffic Analysis Zones (TAZ), where modeling and explicit analysis are carried out. Interactions with zones outside the boundary are defined by external stations, and the models representing these interactions are less complex than those representing interactions within the boundary.

The internal activity system is defined as the set of socio-economic, demographic, and land-use data and characteristics that describe the activities and interactions within the study area for the TAZs. The number of TAZs can vary from a hundred to several thousand, depending on the data available and the model's specifics (age, objectives, etc.). The transport system is represented by network graphs with lines and nodes. The transport system and the activity system are correlated via connectors between the TAZ centroids and existing zones on the physical transport network [11,12]. The role and interest of each stage of the FSM are summarized in the Table 1.

Table 1. Overview of the FSM process.

	Action	Use	Interest
Trip generation	Area is divided in TAZ	Trips are generated for each TAZ based on population density, land-use or household socio-economic data.	The division allows variations in OD trips to be captured at a more granular scale.
Trip distribution	Trip origins are linked to destinations across zones using models	Creation of trip matrix between each OD pairs	To model how journeys are distributed across the transport network, considering the distances and attractions of the different zones.
Mode choice	Determines the proportion of trips using different modes of transport.	To show travelers' modal preferences for each origin and destination zone.	To analyze the mode choices specific to each zone, and to understand the infrastructure requirements for each mode of transport.
Route choice	Travel between zones is assigned to specific routes in the transport network.	Routes are calculated for each mode of transport using principles, and the results show how the different routes are used by journeys between zones.	To forecast traffic volumes and capacity requirements on the different parts of the transport network.

3.2. Trip Chaining Method (TCM)

Now, we present a technique, the Trip Chaining Method (TCM), which is specific to the public transportation context. The TCM is used for ODM reconstruction when information is missing from the collected data on users' journeys. For example, in open systems, we may not know the time or the station where a person alighted. To estimate this information, hypotheses are used to approximate the data in a logical time-space continuum. One disadvantage of this process is that it requires at least two trip records per day, making it challenging to assess the uncertainty of network usage reconstruction [2].

To carry out trip chaining, certain assumptions about individual mobility are made:

- Most passengers start the next trip at or near the point where they got off the previous trip.
- Most passengers end their last journey of the day at or near the boarding point of their first journey of the day.
- There are distance/time thresholds for inter-stop passenger transfers.

For additional TCM hypotheses, refer to [13] or consult [6] for a bibliographical study based on distance/time thresholds.

The TCM is based on simplifying assumptions that lead to approximations and generate model uncertainty. Some authors have proposed solutions to improve the performance of TCM in estimating alighting stops. For example, ref. [7] develop a methodology to progressively estimate information related to unrecorded alighting stops, taking into account trip characteristics (continuous travel, non-continuous recurrent travel, non-continuous non-recurrent travel or route pattern) and using temporal travel models along with passenger boarding histories based on transit card data. A similar approach is found in [8] which also proposes a method based on travel histories and AFC data to assign alighting stops. This method generates travel profiles by counting recorded trips over time. These profiles are then grouped into different clusters using k-means clustering, and a neural network is used to associate each cluster with a temporal distribution of boarding times. Then, based on trip records and travel patterns from previous days, months, or years, the alighting stop is estimated. In the same article, ref. [8] states that the existing TCM method succeeded in matching 60% of total trip destinations, while the proposed method improved this match to 74.9%, additionally associating 37.2% of unmatched trip destinations. Mobility indices are quantitative measures based on location or navigation data, used to assess variations in population movement behaviors over time and space. In their article, ref. [9] evaluate the relevance of using aggregated mobility indices, for example Google Community Mobility Reports or Amazon Mobility Trend Reports, along with traditional data in public transport planning. Along the same lines, ref. [10] conducts an analysis in this study on the impact of supplementary data, such as link flows, destination counts, and/or average travel distances, on alighting estimation within the reconstruction of ODms. The TCM is a method for reconstructing ODm when boarding validation data is available. This method has limitations due to public transport data and involves uncertainty related to the model's assumptions. Although it relies on simplified assumptions, these limitations can be addressed by incorporating emerging data sources into traditional transport data. Furthermore, this method can be combined with various tools, such as clustering algorithms or neural networks. We can cite Radfar, in [14] realizes a study that have leveraged smart card data to reconstruct zonal ODmsthrough enhanced trip-chain reconstruction models. By processing a full year of transactions, including unlinked trips, these approaches achieve more accurate origin–destination estimation and reveal detailed patterns of transfers and trip origins. Such data-driven OD reconstruction frameworks provide valuable insights for improving public transport planning and operations. Or Cui in [15] present a short-term OD flow reconstruction in metro systems is challenging due to data lag, high dimensionality, and data malconformation. To address this, the authors propose an ST-LSTM model combining temporal feature extraction, spatial correlation learning, an input-control gate to handle data lag, and a fusion module integrating historical and real-time information. Experiments on large-scale Nanjing and Beijing datasets show that the model outperforms existing methods by more effectively capturing spatio-temporal dependencies.

Despite their widespread adoption, Trip Chaining Methods rely on strong behavioral and threshold-based assumptions that have direct implications for ODm reliability. The assumption of spatial continuity between consecutive trips and of a systematic return to the origin at the end of the day is well suited to regular commuting patterns but becomes increasingly fragile in the presence of non-routine activities, intermediate stops, multi-modal trips, or irregular schedules. As a result, OD estimates derived from TCM tend to be more reliable for habitual users and peak-period travel, while off-peak, discretionary, or occasional trips are more prone to misclassification.

Moreover, TCM performance is highly sensitive to the choice of distance and time thresholds used to link successive trips. Small variations in these parameters can lead to significant changes in inferred alighting locations and OD flows, yet these thresholds are often calibrated heuristically and remain context dependent. In most studies, this sensitivity is not systematically analyzed, and the uncertainty induced by threshold selection is rarely propagated to the final ODm.

Finally, while recent extensions combining TCM with clustering techniques or machine learning models improve destination inference rates, they do not fundamentally remove the dependence on behavioral regularity assumptions. Instead, they tend to shift uncertainty from explicit rules to data-driven representations that remain difficult to interpret and validate. Consequently, TCM-based ODms should be interpreted as behaviorally plausible reconstructions rather than precise ground-truth representations, and their use in planning or optimization should be accompanied by sensitivity analyses or complementary validation sources whenever possible.

3.3. Data Fusion Methods for Public Transport OD Estimation

The growing availability of heterogeneous public transport data sources, such as Automatic Fare Collection (AFC), Automatic Vehicle Location (AVL), Automatic Passenger Counting (APC), and General Transit Feed Specification (GTFS), has led to the development of various data fusion approaches for OD reconstruction. Each source provides only partial and noisy information on passenger movements: AFC captures boardings but rarely alighting, AVL describes vehicle trajectories with positioning errors, APC data are often incomplete, and GTFS reflects planned rather than realized operations. Consequently, OD estimation increasingly relies on data fusion methods to mitigate these limitations (Table 2).

Deterministic or rule-based methods typically combine AFC data with GTFS schedules and network topology to infer passenger trajectories or OD pairs. In these approaches, heuristic rules based on temporal proximity, spatial feasibility, and service consistency are used to assign transactions to routes and alighting stops. Their main advantages are simplicity, interpretability, and low computational cost.

However, several studies report that these methods are highly sensitive to modeling assumptions and data quality. Deviations between scheduled and realized services, as well as noise in transaction timestamps, can significantly affect results. Moreover, uncertainty is not explicitly modeled, limiting robustness when data are incomplete or inconsistent.

Probabilistic and statistical approaches address these issues by explicitly modeling uncertainty and latent passenger movements. OD flows, boarding and alighting events are treated as random variables inferred from AFC, AVL, and APC observations using Bayesian frameworks, expectation–maximization algorithms, or state-space models.

These methods have been shown to be more robust to noisy and missing data, particularly when APC coverage is partial or AVL data are used to correct schedule deviations. Their main drawbacks lie in stronger parametric assumptions, calibration effort, and increased computational complexity, especially for large-scale networks.

Optimization-based methods formulate data fusion as the adjustment of a prior (seed) ODm to satisfy constraints derived from observed data, such as passenger counts, link flows, or capacity limits. Techniques based on maximum entropy or generalized least squares are commonly used.

These approaches are well suited for network-level OD estimation and ensure consistency with system-wide constraints. However, their performance strongly depends on the quality of the initial seed matrix, and uncertainty is often treated implicitly rather than explicitly.

Learning-based approaches, including machine learning and deep learning models, have recently been proposed to integrate heterogeneous transit data sources. Multi-input architectures are used to fuse AFC, AVL, and GTFS(-RT) data, capturing complex non-linear relationships.

While these methods often achieve high predictive performance in data-rich contexts, they generally suffer from limited interpretability. In addition, uncertainty quantification is rarely addressed, which restricts their applicability for planning and policy analysis.

Overall, the literature indicates that no single data fusion approach simultaneously addresses uncertainty modeling, interpretability, scalability, and robustness to heterogeneous data. Deterministic methods favor simplicity, probabilistic approaches explicitly handle uncertainty, optimization-based methods ensure network coherence, and learning-based models excel in prediction. The choice of method should therefore depend on data availability and study objectives.

Table 2. Analytical comparison of public transport ODm reconstruction paradigms.

Method Family	Typical Data Sources	Main Strengths	Main Limitations	References
Deterministic/rule-based	AFC + GTFS ($\pm$ AVL)	Simple, interpretable, low computational cost	Strong assumptions, sensitive to noise, no explicit uncertainty	[16,17]
Probabilistic/statistical	AFC + AVL + APC	Explicit uncertainty modeling, robustness to missing data	Parametric assumptions, calibration complexity	[18,19]
Optimization-based	Seed OD + AFC/APC counts	Network consistency, scalable	Dependence on seed matrix, limited uncertainty handling	[20,21]
Learning-based	AFC + AVL + GTFS(-RT)	Captures non-linear patterns, strong predictive performance	Low interpretability, limited uncertainty modeling	[22,23]

3.4. Generic Estimation Techniques

In the literature, many methods are used to estimate ODms. Classical probabilistic methods such as Entropy Maximization (EM), Maximum Likelihood Estimation (MLE), Constrained Generalized Least Squares (C-GLS), Iterative Proportional Fitting (IPF), and Bayesian Estimation (BE) are commonly used because they yield good results in estimation and can be applied to transport data [13,24–27]. Additionally, algorithms from artificial intelligence, such as Recurrent Neural Networks (RNN) and its variants, and genetic algorithms, can also be employed to compute ODm [28–30]. Various filtering methods, including Kalman filters and particle filters, have also shown good results in ODm reconstruction [31,32]. All these techniques are well-known and used with different kinds of data and for various applications, including ODm reconstruction. All these methods can be used to estimate static or dynamic ODms, for example, Galliani in [33] has proposed a scalable pipelines to estimate dynamic railwaODmsDm by fusing easily accessible data sources such as ticket and subscription sales with APC-based passenger counts. By generating OD seeds from ticket data and refining them through Iterative Proportional Fitting, these methods produce reliable weekly ODm without relying on smart card systems or costly AFC infrastructures. Such approaches enable network-wide monitoring, anomaly detection, and data-driven operational planning, highlighting the growing value of dynamic OD estimation for resilient and adaptive transport systems.

3.5. OD Matrix Estimation Model

We can categorize the models into two distinct types: on one hand, the explicit models, or “white box” models, which are based on explainable physical principles; and on

the other hand, the “black box” models, which are driven by data flows. A structured representation of this classification is shown in Figure 3.

This classification highlights a fundamental trade-off between interpretability and flexibility: white-box models favor explanatory power and behavioral transparency, whereas black-box models prioritize empirical accuracy and adaptability to complex data at the expense of interpretability.

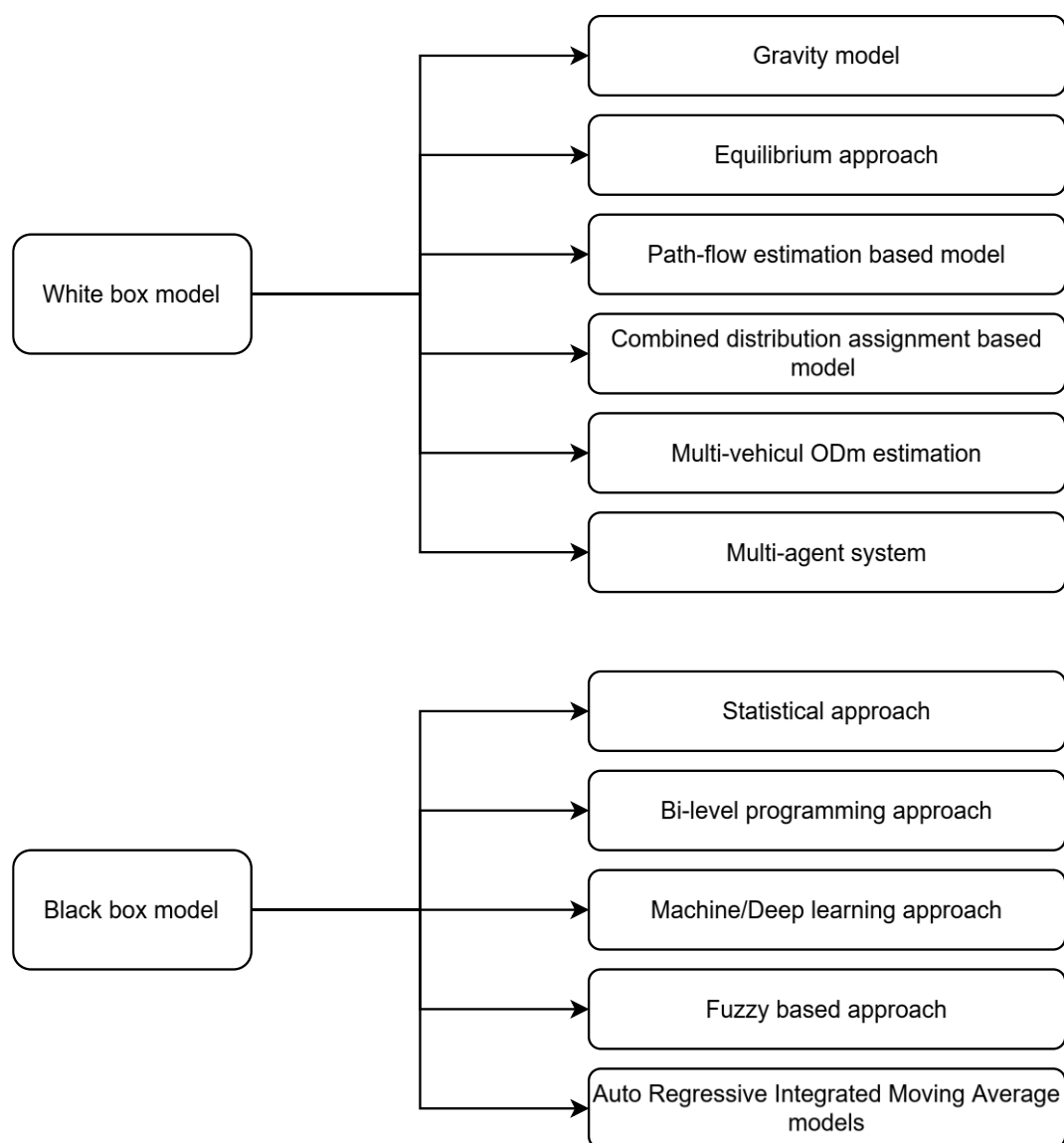


Figure 3. Models Classification.

3.6. White Box Model

Model 1: Gravity Model. The gravity model is based on a spatial approach that links activity zones of varying attractiveness, separated by a given distance [32]. The model allows for the assessment of interactions between these zones: interaction tends to decrease with increasing distances between zones and with decreasing attractiveness. While the model is simple enough to capture the fundamental principles of the interaction, it assumes that the travel path is homogeneous. To calibrate the model and estimate the ODm, data on inter-zone movements are required [34,35].

Model 2: Equilibrium Approach. In the equilibrium approach, the estimation of ODm is based on Wardrop’s first principle, which states: “Travel times on all routes are equal (to each other) and less than the time it would take a vehicle to switch to another route.”

This implies that each user aims to minimize their transportation cost by choosing the route that best suits their needs. Equilibrium is achieved when no user can reduce their journey time unilaterally. This approach is particularly well-suited for accounting for congestion effects [36].

Model 3: Path-Flow Estimation based models. Models based on the trip flow estimator relate the ODM of flows to the proportion of route choices, considering congestion effects. The central component of this approach is a path choice model, which interacts with the link cost functions to produce a stochastic equilibrium traffic model for users. The procedure involves solving a shortest path network flow programming sub-problem, with a column generation technique used to generate paths. These alternative paths are then used to determine the optimal solution of the linear programming model [37,38].

Model 4: Combined Distribution and Assignment-based problem. The Combined Distribution and Assignment (CDA) problem is based on network equilibrium. Specifically, it involves simultaneously determining the distribution of trips between origins and destinations—often approximated using a gravity model—and assigning these trips to routes for each OD pair according to the user's (deterministic or stochastic) equilibrium principle [26,35,39].

Model 5: Multi-Agent Systems. A simulation tool can be employed to obtain spatio-temporal information at various scales from the actions of agents (e.g., people, cars, buses) that follow specific behavioral rules (tasks). Typically, a cost function (or score) is associated with the different tasks performed by each agent. Through iterative perturbation of the system, different configurations can be explored, and convergence is achieved when the overall score stabilizes. It is also possible to modify behavioral rules during the simulation, such as inter-agent interactions. In dynamic environments, this approach is referred to as adaptive multi-agent systems [40].

Beyond OD reconstruction, behavioral information extracted from automatically collected data—particularly AFC data—has been increasingly used to infer user preferences and travel regularities. Repeated smart-card observations enable the identification of habitual patterns such as preferred boarding times, recurrent trajectories, or route usage, which are commonly exploited through trip-chain reconstruction, clustering, or probabilistic inference. In OD estimation models, this behavioral information is typically incorporated implicitly via assumptions of regularity, stationarity, or continuity of travel chains, while in more advanced frameworks it is used to calibrate route choice models or agent cost functions. Multi-agent systems make this integration explicit by simulating individual agents whose decisions follow behavioral rules inferred from observed data, allowing the reproduction of aggregate flows and responses to network changes or disruptions. However, behavioral inference remains largely indirect, and associated uncertainty is rarely propagated to OD estimates or simulation outcomes. Table 3 provides a synthesis of all white box models, highlighting their respective limitations.

Table 3. Presentation of the main limitations for white box models.

Models	Main Limitations	White Box Nature
Gravity model	Assumes homogeneous travel conditions and simplified impedance functions; limited ability to capture congestion effects and route choice behavior; calibration strongly depends on the quality of inter-zonal flow data.	Explicit functional form linking attractiveness and distance; parameters have clear physical and behavioral interpretations, making the influence of each factor on OD flows transparent.
Equilibrium-based models	Relies on strong assumptions of rational user behavior and perfect information; may fail to represent heterogeneous preferences or	Directly grounded in Wardrop's principles; travel behavior is explicitly modeled through cost minimization and equilibrium conditions, allowing clear interpretation of OD-network interactions.

	bounded rationality; computational complexity increases for large-scale congested networks.	
Path-flow estimation models	Requires detailed knowledge of path choice sets and link cost functions; sensitive to path enumeration and model specification; computationally intensive due to column generation and shortest-path subproblems.	OD flows are explicitly linked to route choice probabilities and congestion effects; path-level representation provides transparent insight into how OD demand translates into network flows.
Multi-vehicle ODm estimation	Strong coupling between trip distribution and assignment increases model complexity; relies on simplifying assumptions (e.g., gravity-based distribution); convergence and calibration can be challenging.	Simultaneous modeling of trip distribution and route assignment offers an explicit and interpretable representation of travel demand generation and network loading mechanisms.
Multi-agent systems	High computational cost; results depend on behavioral rules and parameter calibration; difficult to scale and validate at large network levels.	Individual agent behaviors and interactions are explicitly defined; emergent OD patterns can be traced back to well-identified behavioral rules and decision processes.

3.7. Black Box Model

Model 6: Statistical Approach. Regardless of whether the effects of congestion on networks are considered, statistical methods such as Bayesian Estimation (BE), Maximum Likelihood Estimation (MLE), or (Constrained) Generalized Least Squares ((C-)GLS) can be employed to estimate the ODm from an initial raw estimation matrix. The BE method utilizes past data and observations to produce posterior estimates, which is useful in transport planning. The MLE method incorporates the assumption that flows follow a specific distribution, such as a multivariate normal distribution or a Poisson distribution. This method aims to maximize the probability of correspondence between the ODm and the observed count data. The (C-)GLS method, on the other hand, avoids making specific assumptions about the data and integrates both count data and survey data to refine the estimation [32,35].

Model 7: Bi-level programming approach. The two-level programmed approach is based on a two-tier programming framework [41] and references therein. The lower level involves solving the equilibrium assignment problem, which can be derived from field data, numerical simulations, or analytical solutions. The upper level focuses on estimating the flow ODm. The goal of these methods is to minimize the difference between the observed counts and the counts estimated using the Spiess formulation [42].

Model 8: Machine/Deep Learning (ML/DL) based approach. As transport problems become increasingly complex, it is important to have tools that provide solutions as close to reality as possible, whether in terms of the scale probed, its resolution, or the dynamic nature of the phenomena. Methods based on neural networks are proving to be increasingly effective in addressing these systems. The operating principle involves training the neural network on annotated data, i.e., known input and output data. Estimation of a given configuration is then based solely on introducing a set of input data and applying the model. There are many subtypes of machine learning models, including Artificial Neural Networks, Hopfield Neural Networks [35], Multi-Layer Perceptrons, Backpropagation Neural Networks (BPNNs), Radial Basis Function Neural Networks, Reinforcement Learning, Evolutionary Computation, and Distributed Intelligence [32]. We can cite [43], who developed a multi-fused residual neural network using AFC data from the Shenzhen metro to perform short-term forecasting on the ODm. Using AFC data and a trip-chaining algorithm, the ODm is reconstructed and used as input for the model. The model then converts the complexity of relationships between OD pairs into spatial dependencies based on graphs. It can robustly capture multiple complex dependencies and

surpasses traditional methods in forecasting accuracy. Additionally, in the article by [44], he identifies the issue of mode changes during trips due to infrastructure and bus network services. He proposes a graph-based deep learning model to predict the ODm while explicitly distinguishing between passenger behaviors in transfers and direct trips.

Model 9: Fuzzy based approach. User behavior is particularly difficult to model with a mathematical function due to its inherent complexity. To address this challenge, the fuzzy approach can be employed. This approach is based on a probabilistic framework and is widely used in mobility studies, including estimating ODms, transport planning, and traffic control. For example, the fuzzy approach allows for the fine-tuning of factors influencing user choices through weighting [35]. Moreno-Palacio proposes in his article [45] a synthesis of public transport trips based on entropy, using fuzzy logic driven by entropy maximization. He proposes a method using fuzzy parameters to estimate travel flows based on traffic counts, solved using entropy maximization techniques for public transport (buses). In this work, the fuzzy parameters are bus tour production, the total cost or impedance in the transit system, and the volume or bus traffic counts. The uncertainty related to the variability of these parameters is integrated into the model through the inclusion of fuzzy logic.

Model 10: ARIMA Model. The Auto Regressive Integrated Moving Average (ARIMA) statistical model is a forecasting tool that relies on the analysis of previously observed time series data. It combines auto-regressive and moving average components, making it optimal for systems with regular patterns in the data. ARIMA is commonly used for traffic forecasting, utilizing historical time series data to make predictions [32].

Table 4 provides a synthesis of all black box models, highlighting their respective limitations.

Table 4. Presentation of the main limitations for black box models.

Models	Main Limitations	Black Box Nature
Statistical approaches (BE, MLE, (C-)GLS)	Strong dependence on prior assumptions and distributional hypotheses; sensitivity to noisy or inconsistent data; may produce unrealistic OD patterns without strong regularization or constraints.	The mapping between observed data and OD estimates is implicit and driven by optimization criteria; the influence of priors, weights, and error structures is difficult to interpret.
Bi-level programming	High computational cost; nonconvexity and convergence to local optima; error propagation from the traffic assignment model to OD estimates; sensitivity to initialization.	Nested optimization obscures causal relationships between observations and OD flows; the contribution of each level to the final ODm is not directly interpretable.
Machine/Deep Learning	Requires large volumes of labeled data; limited transferability across networks; prone to overfitting; difficulty integrating physical or behavioral constraints.	Learned nonlinear mappings lack interpretability; internal representations do not provide explicit behavioral or network-level explanations.
Fuzzy-based approaches	Subjective definition of membership functions; heuristic calibration; limited reproducibility and scalability for large networks.	OD estimates emerge from aggregated fuzzy rules and membership functions; individual parameter effects are difficult to trace or explain.
ARIMA models	Assumes stationarity and linearity; ignores spatial dependencies; poor performance under structural breaks or disruptions.	Temporal correlations are modeled without explicit behavioral or network representation, limiting interpretability of OD dynamics.

3.8. Comparative Synthesis of OD Estimation Paradigms

Across the reviewed literature, OD estimation methods exhibit clear trade-offs rather than a hierarchy of intrinsic performance. White-box approaches, such as gravity,

equilibrium, and path-flow models, offer strong interpretability and explicit behavioral or physical foundations, making them suitable for planning, scenario analysis, and policy evaluation. However, they rely on simplifying assumptions that limit their flexibility in data-rich but behaviorally complex contexts.

Conversely, black-box approaches, including statistical inference, machine learning, and deep learning models, prioritize empirical fit and adaptability to heterogeneous and noisy data. These methods often achieve strong predictive performance in large-scale networks but suffer from limited interpretability and weak uncertainty propagation, which constrains their use for explanatory or decision-support purposes.

Hybrid and data fusion approaches attempt to balance these paradigms by combining structural constraints with probabilistic or learning-based inference. Nevertheless, the literature indicates that no single framework simultaneously optimizes interpretability, uncertainty handling, scalability, and behavioral realism. Method selection therefore reflects a compromise driven by data availability, computational constraints, and study objectives rather than absolute methodological superiority.

3.9. Uncertainty

Referring to [46], uncertainty can be expressed in three different ways:

- Discord or dispersion, due to evidence supporting mutually exclusive alternatives, for example inconsistent data;
- Non-specificity or imprecision, due to evidence supporting nested alternatives, for example the spatio-temporal variability of data;
- Fuzziness or vagueness, due to evidence not supporting sharp definition of alternatives for example model uncertainty.

In the validation process, it is important to consider the uncertainty of input data, which can propagate to the output data, as well as the uncertainty of the model itself.

Few studies have examined the uncertainty of input data or model uncertainty, and prior research has primarily been limited to four-step models [47]. Various authors have used tools to address this uncertainty in data processing, which can be divided into three categories: deterministic modeling, the Bayesian approach with ML/DL integration, and operations research. In operations research, uncertainty is addressed by [48], who defines uncertainty as Spatio-temporal variability of data and use it for route recommendation on the Chicago Transit Authority under real disturbance cases. Uncertainty is integrated through robust optimization, reducing the average travel time for passengers on the affected line by 2.91% compared to the nominal model. In another study, ref. [49] defines uncertainty as Spatio-temporal variability of data and proposes a three-step hybrid coding method, to optimize bus routes under uncertain conditions. The study is conducted on part of the road network in the An'ning district of Lanzhou city. In his article, Ma shows that the results save 42.11% of operating time compared to the results from the hybrid solution based on K-means and the multi-objective genetic algorithm. For Bayesian approaches with learning tools, we find [50], who defines uncertainty as Spatio-temporal variability of data and designs a zero-inflated negative binomial spatio-temporal graph neural network to quantify uncertainty in sparse ODM reconstruction, with results showing the superiority of this network over reference models in terms of accuracy and tight confidence intervals. Additionally, ref. [51] defines spatio-temporal variability of data and use a day-to-day learning process to evaluate user behavior in relation to travel uncertainty on the network. In [52], Manzo defines spatio-temporal variability of data and proposes a Monte Carlo simulation and scenario analysis approach to quantify the impact of data uncertainty on ODM reconstruction in applying the four-step model to transportation in the city of Næstved, Denmark. Moreover we can cite [53] which present in a recent

work the integration of uncertainty quantification into ODm reconstruction using advanced deep learning models. For instance, graph-based spatial-temporal neural architectures combined with Tweedie distributions have been proposed to jointly learn multiple uncertainty parameters, enabling more robust modeling of zero-inflated and highly variable flow data. These approaches highlight a growing interest in probabilistic frameworks that explicitly capture spatial-temporal dependencies and inherent uncertainty in ODm reconstruction. Or we can cite [54] which present in his work a nonparametric probabilistic frameworks to quantify uncertainty in travel demand prediction. For example, STGCN-VAE combines spatio-temporal graph convolutions with a variational autoencoder to learn the full distribution of demand through sampling and kernel density estimation, enabling both point forecasts and calibrated uncertainty estimates. These contributions illustrate a growing trend toward generative, distribution-aware models that explicitly capture spatial-temporal variability and data sparsity challenges in mobility forecasting. And Fan present in [55] a counterfactual deep learning frameworks to improve ODm reconstruction under operational disruptions. A dual-channel model combining factual forecasting with counterfactual inference has been shown to enhance robustness during incidents by leveraging KL-based propensity matching and incident-type embeddings to capture heterogeneous disruption effects. These approaches highlight the growing importance of causal modeling for understanding spatiotemporal differences between normal and disrupted passenger flows.

Across methodological families, no approach consistently outperforms others across contexts. Instead, methods reflect trade-offs between interpretability, robustness, scalability, and uncertainty handling. The literature reveals a fragmentation of approaches, with limited cross-method comparisons under comparable validation settings.

4. Outputs and Performance

After estimating the ODm, it is essential to assess its quality to determine the confidence level we can place in the results. This assessment step is crucial for evaluating the performance of the estimation.

4.1. Statistical Indicators

We can use statistical indicators to evaluate the performance of a method. The main statistical indicators, which evaluate a difference, are Mean Absolute Error (MAE), Root Mean Square Error (Root Mean Square Error RMSE) and Mean Absolute Percentage Error (MAPE). But it exists a lot of statistical indicators.

While statistical analysis tools are versatile, choosing the ‘right’ estimator for a given measurement and interpreting its results can be challenging. According to [56] MAE and RMSE are generally more stable than MAPE for evaluating discrepancies between estimated and observed data. They note that RMSE values are typically higher than MAE values and following Kleijnen’s work cited by [56], an acceptable limit for MAPE values is 10%. Ref. [57] offers an interpretation of these indicators through function minimization analysis. The zero-derivative approach highlights that optimizing RMSE tends to provide a correct average response, while optimizing MAPE focuses on the median and is less sensitive to outliers in the input data. Similar findings are reported by [58]. Vandeput emphasizes that selecting the appropriate estimator is often dependent on the dataset and is not straightforward.

4.2. Similarity Indicators

Certain indicators are used to evaluate the similarity between the true and estimated ODm.

- **The Spearman correlation:** This method calculates a correlation coefficient based on the ranks rather than the actual values of the two variables. Instead of comparing the numerical values directly, it compares the order of the values of the flows. For the same origin, descending flows are ranked in ascending order, and the degree to which these rankings match between the two matrices is assessed [59]. The Spearman method is computationally efficient but should be used alongside hypothesis testing to validate results [60].
- **Structural SIMilarity (SSim):** Originally developed to compare two images and measure their structural similarity, the SSim method has been adapted by some researchers to assess the similarity between two ODms. This index evaluates similarity by considering the mean, variance, and structure of the origin and destination distributions [61]. While SSim is not a stand-alone solution, it complements statistical approaches effectively. As noted by [61], statistical indicators alone do not capture the structural aspects of the matrix, which SSim provides.
- **Jensen–Shannon Divergence (JSD)** is an information-theoretic measure used to quantify the dissimilarity between two probability distributions. When applied to Origin–Destination (OD) matrices, it evaluates differences in the relative distribution of flows rather than their absolute magnitudes. By symmetrizing and smoothing the Kullback–Leibler divergence, JSD is bounded and well-defined even in the presence of zero flows, making it particularly suitable for sparse ODms. This metric highlights discrepancies in the spatial structure and proportional allocation of trips across OD pairs. However, since it operates on normalized distributions, JSD is insensitive to global scaling effects and does not capture errors in total demand levels. Consequently, it is best suited for assessing how well a reconstructed ODm reproduces the pattern of mobility flows, rather than their absolute volumes [62–64].
- **Normalized Root Mean Square Error (NRMSE)** is a scale-independent error metric that quantifies the average magnitude of deviations between estimated and observed OD flows. By normalizing the Root Mean Square Error with respect to a reference value (such as the mean or the range of observed flows), NRMSE enables comparisons across datasets with different demand levels. This metric is particularly sensitive to large absolute errors, as squaring amplifies the contribution of high-flow OD pairs. As a result, NRMSE effectively reflects the model’s ability to reproduce major traffic flows, which are often of primary operational interest. However, this sensitivity also constitutes a limitation, as errors in low-flow OD pairs contribute marginally to the metric. NRMSE therefore provides a robust measure of global accuracy but may obscure structural mismatches in the distribution of smaller flows [62].
- **Common Part of Commuters (CPC)** metric measures the degree of overlap between two ODms by quantifying the proportion of trips that are consistently represented in both matrices. Defined as the ratio of shared flows to the total observed demand, CPC ranges from 0 (no overlap) to 1 (perfect match). This indicator provides an intuitive assessment of how much of the observed commuting structure is correctly captured by the estimated ODm. CPC is particularly effective at highlighting the overall similarity of flow allocations and is easy to interpret from a policy or planning perspective. Nevertheless, CPC does not penalize compensating errors and is insensitive to the spatial redistribution of mismatches. As such, it may indicate high similarity even when significant localized errors persist and should therefore be complemented with metrics that explicitly account for flow magnitudes and spatial discrepancies [62,63].

4.3. Results

In Table 5. OD matrix estimation performance crossed with references. Table 5, we present the case studies referenced in this paper. This table is not exhaustive, and readers

are encouraged to consult additional sources such as [39,65] for more examples. It is important to note that some references do not always provide explicit details on validation methods and performance. According to the data type and estimation methods used, the validation performance reported in the literature ranges from approximately 60% to 90%.

Table 5. OD matrix estimation performance crossed with references.

Reference	Data	Methods	Tools	Data Set and Size Sample	Source of Uncertainty	Goal	Validation Out Come
[38]	APC/AFC/AVL	PFE—LP (TT)	RMSE/MAE	Northern Virginia Network 164 links/31 zones	Assumptions: deterministic user equilibrium, homogeneous behavior Data: measurement uncertainty in link counts, absence of direct OD observations Models: non-uniqueness of solutions, sensitivity to penalties and constraints	Test the efficiency of LP (TT) model	Validation type: internal/consistency validation Validated against: observed link flow counts Metric/order: low discrepancies in reproduced link volumes Interpretation: ensures network flow consistency only; no OD ground truth
[1]	APC/AFC/AVL	MLE/IPF	Percentage comparison	CTA Chicago 2,736,454 records	Assumptions: regularity of travel chains and return to starting point Data: behavioral uncertainty of users and collection errors (AFC/APC/AVL) Models: uncontrolled propagation of errors between stages, dependence on IPF vs. MLE choice	Estimate OD matrices using ADCS	Validation type: internal/consistency validation Validated against: boarding/alighting marginals (APC) and prior OD estimates Metric/order: improved consistency using MLE vs. IPF Interpretation: pipeline coherence verified; no independent OD ground truth
[31]	APC	Combining Kalman filter and Gibbs sampler	MAE	Taipei Metro 9 stations	Assumptions: linear dynamics of OD flows, local stationarity Data: noise and partial coverage of time counts Models: dependence on covariance matrices and MCMC convergence	Comparisons of real and estimated O-D flows	Validation type: internal/consistency validation Validated against: time-dependent link flow counts Metric/order: reduced temporal flow errors compared to baseline Interpretation: context-dependent validation; OD not directly validated
[66]	AFC/AVL	TCM on AFC and AVL data	Comparison between surveys	HTSD/Oyster ~14,000 boardings recorded	Assumptions: continuity of travel chains and fixed rules for inferring destinations Data: behavioral uncertainty of users and ticketing collection errors Models: dependence on spatio-temporal thresholds, absence of uncertainty propagation	Validation of ODm buses	Validation type: survey-based validation Validated against: on-board passenger OD surveys Metric/order: ~60–70% inferred destinations correctly matched Interpretation: partial validation; sensitive to survey bias and coverage
[67]	AFC	Clustering Algorithm		Rennes metro 53,195 tickets	Assumptions: time profiles representative of actual behavior Data: absence of destinations and anonymisation of ticketing data Models: dependence on the number of clusters, unpropagated classification uncertainty	Evaluate the efficiency of combining SCD and socioeconomics data	Validation type: internal/consistency validation Validated against: internal clustering coherence criteria Metric/order: interpretable and stable passenger profiles Interpretation: exploratory behavioral validation; OD not evaluated
[68]	AFC	Algorithm with reducing distance	Compared with complete OD	Translikn Australia 161,446 records	Assumptions: regular behaviors exploitable by trip-chaining	Improved OD estimation algorithm	Validation type: survey-based validation

					Data: behavioral uncertainty of users and AFC validation errors Models: high sensitivity to thresholds (walking, transfer), uncertainty not propagated		Validated against: independent on-board surveys with known alightings Metric/order: ~65–73% zonal OD match; mean error distance ~500–800 m Interpretation: strong trip-chaining validation; context-specific
[69]	HTSD	Sequential update on EM principle	Pearson's coefficient/Cosine similarity	Metropolitan area Tokyo survey 24 measurement intervals	Assumptions: principle of maximum entropy as a behavioral proxy Data: loss of information due to GPS aggregation and coverage bias Models: dependence on entropy constraints and non-uniqueness of solutions	using aggregated data of mobile phone traces to estimate work-related trips.	Validation type: aggregate/indirect validation Validated against: synthetic ground truth and aggregated OD statistics Metric/order: moderate RMSE/MAPE; high OD correlation Interpretation: informative at aggregate level only; no individual OD validation
[70]	AFC	Gaussian Mixture Model/EM		SCD transport Gatineau 3,492,310 records	Assumptions: inter-annual stability of individual behavior Data: survival bias of maps and absence of OD information Models: dependence on longitudinal clustering, unpropagated uncertainty	Tracking and Spatial analysis by cluster and card type	Validation type: internal/consistency validation Validated against: likelihood-based clustering criteria (ICL) Metric/order: stable cluster structure across parameter settings Interpretation: methodological validation only; no external reference
[71]	AFC	Clustering technique on similar boarding		AFC Rennes metro 5,404,096 transactions	Assumptions: conditional independence of temporal events Data: behavioral uncertainty and incomplete ticketing data Models: choice of number of clusters and sensitivity to generative model parameters	Cluster population using SCD	Validation type: internal/consistency validation Validated against: longitudinal stability of smart-card user profiles Metric/order: low KL divergence between yearly clusters Interpretation: validates relative behavioral dynamics; not OD accuracy
[34]	AFC	EM	Compare aggregated statistics with data from transport agency	SCD Stockholm 51 train station	Assumptions: daily symmetry and aggregate attractiveness of stations Data: absence of direct observation of destinations (entry-only AFC) Models: dependence on additional constraints, uncertainty not propagated	Test model performance	Validation type: aggregate/indirect validation Validated against: official aggregate statistics (person-km, exits) Metric/order: ~95–97% fit on aggregate indicators Interpretation: excellent macro-level agreement; no pairwise OD validation
[72]	AFC/HTSD	Combination of Mixed logit SCD-HTDS regression		Nanjing metro SCD 21 million HTSD 12,000 person	Assumptions: regularity of home-work travel and fixed behavioral thresholds Data: ticketing uncertainty and HTS sampling bias Models: dependence on GMM clustering and non-propagation to the explanatory model	Identify factors influencing modal change	Validation type: survey-based validation Validated against: household travel survey (HTS) Metric/order: ~90% commuter identification agreement; pseudo-R ² ≈ 0.3 Interpretation: behavioral validation; OD not explicitly assessed
[73]	AFC	Comparison of algorithm for data processing	Percentage comparison	AFC transit system Seoul 1,101,544,931 logs	Assumptions: fixed chaining rules and representativeness of AFC validations Data: AFC measurement uncertainty and low traffic in certain segments	Obtaining system-level transit information	Validation type: internal/consistency validation Validated against: alternative AFC-based estimation variants

					Models: dependence on reliability parameters (quantiles, DBSCAN)		Metric/order: consistent reliability indicators (IQR, congestion patterns) Interpretation: robust system-level indicators; outside OD scope
[74]	AFC/HTSD	WOAFC- WIAFC/ODS/H TDS	sMAPE/RMSE	Public Transport Lyon 18 zones/324 OD pairs	Assumptions: inter-source comparability with equal aggregation Data: uncertainties specific to each source (AFC, ODS, HTS) Models: dependence on scaling rules and aggregation level	Measure difference in data collections	Validation type: survey-based validation Validated against: OD survey and household travel survey Metric/order: R ² up to ~0.97; SMAPE/RMSE comparisons Interpretation: strong inter-source comparability; not an OD estimation method per se
[75]	AFC	CCD/DTW		SCD transport STO Gatineau 26 time series	Assumptions: temporal similarity captured by distance metrics Data: incomplete temporal profiles and absence of OD Models: strong dependence on the chosen metric (DTW vs. CCD)	Compare CCD to DTW to classify time series	Validation type: internal/consistency validation Validated against: clustering stability and homogeneity criteria Metric/order: clear separation and stable temporal clusters Interpretation: methodological validation only; no external benchmark
[76]	AFC/AVL/ HTSD	Clustering technique	MAE	Ticket recycling survey 10,551 records	Assumptions: regular behaviors exploitable via trip-chaining Data: ticketing uncertainty and GPS/time synchronization errors Models: sensitivity to DBSCAN parameters and spatio-temporal thresholds	Estimate OD matrices	Validation type: survey-based validation Validated against: full-sample field survey (recycled ticket experiment) Metric/order: >80% station accuracy; >90% OD with error <2 passengers Interpretation: strong empirical validation; highly context-dependent
[77]	APC/HTSD/ Demographic Data	PCA	MAPE	Santander network 4205 links/1630 nodes	Assumptions: regularity of the supply–demand relationship captured by the data Data: uncertainty about measured speeds and flows, partial coverage of meters Models: sensitivity to the number of paths, linear PCA and ML errors	Evaluate the efficiency of PCA method	Validation type: aggregate/indirect validation Validated against: synthetic OD ground truth and network-level measures Metric/order: OD MAPE ~20–30%; low network RMSE Interpretation: good global performance; sensitive to modeling assumptions
[2]	AFC	TCM based on MER	Logistic regression	AFC/AVL bus Nanjing 52 million SCD record	Assumptions: stationarity and informational regularity of trajectories Data: dependence on the length of the ticketing history Models: sensitivity to the entropy threshold, indirect propagation of errors	Test new framework based on MER criterion	Validation type: survey-based validation Validated against: observed metro alighting data (ground truth for single trips) Metric/order: +15–20% improvement in single-trip destination inference Interpretation: effective targeted correction; partial OD impact
[27]	AFC	ODIP/EM	OD analysis with EM	Seoul metro ≈1 billion records	Assumptions: Single, stable OD per individual, homogeneous preferences Data: Ticketing uncertainty and indirect reconstruction of segments outside the system	Test inference framework for distilling multi-modal routing preferences	Validation type: aggregate/indirect validation (β -only) Validated against: masked AFC data and route-choice model fit Metric/order: stable β estimates; consistent pseudo-R ²

					Models: Dependence on EM initialisation and attribute selection (β)		Interpretation: preferences validated, OD quality assessed indirectly only
[78]	AFC	BE	MSE	New York metro 1315 records	Assumptions: steady state and conservation of nodal flows Data: noise and overdispersion of entry/exit counts Models: dependence on priors and time window (but quantified uncertainty)	Test the efficiency of the model	Validation type: probabilistic/uncertainty-aware validation Validated against: synthetic OD ground truth and nodal flow counts Metric/order: posterior OD distributions with 95% credible intervals Interpretation: robust uncertainty-aware validation; assumption-dependent
[79]	AFC/AVL/GTFS	Zonal Attraction Clusters	Percentage comparison	Public Transport multimodal Stockholm 10,000 km/800,000 passengers/5700 stops	Assumptions: weekly stability of temporal profiles Data: selection bias of frequent users and absence of OD Models: dependence on the number of clusters and temporal aggregation	Use attractiveness to identify and classify urban areas	Validation type: internal/consistency validation Validated against: longitudinal persistence of individual profiles Metric/order: stable weekly temporal patterns across months Interpretation: strong behavioral consistency; no OD validation

Reported performance and validation metrics are study-specific and not directly comparable due to differences in data sources, spatial aggregation levels, modeling assumptions, and validation procedures.

The comparative analysis summarized in Table 5 highlights a strong dependence between OD estimation methodologies and the nature of the available data. Rather than exhibiting systematic performance dominance, methodological choices appear primarily driven by data informativeness and by the type of outputs targeted. Trip Chain Models (TCM) and their extensions, including MER-based approaches, rely almost exclusively on AFC data, occasionally complemented by AVL information to ensure spatio-temporal consistency. These methods explicitly exploit individual-level regularity in travel behavior and therefore require sufficiently long user histories to remain identifiable.

In contrast, statistical inference approaches such as EM, IPF, MLE, or PCA are more commonly applied to aggregated observations derived from APC, aggregated AFC, or HTSD data. These methods prioritize global flow consistency and network-level coherence rather than the reconstruction of individual trajectories. Exploratory and clustering-based approaches predominantly rely on AFC data alone, leveraging their high temporal resolution and large volume, but inherently lack direct access to destinations, which limits their applicability to descriptive and behavioral analyses rather than ODM reconstruction. Bayesian approaches remain marginal in the literature and are mostly applied to nodal or aggregated flow counts, where assumptions of conservation and stationarity are acceptable and where uncertainty can be formally modeled.

Beyond methodological diversity, several strong assumptions recur across studies regardless of the chosen approach. These include the regularity of user behavior, often associated with stable home–work travel patterns, temporal stationarity over the analysis period, continuity of trip chains with frequent return to an origin location, and partial behavioral homogeneity in aggregated models. Such assumptions are rarely tested explicitly and are generally accepted as necessary conditions for model identifiability.

At the same time, some assumptions are structurally tied to specific families of methods. TCM and MER-based approaches depend critically on individual-level stationarity and on fixed spatial and temporal thresholds governing trip chaining, making their

performance highly sensitive to behavioral deviations. EM-, IPF-, and MLE-based methods rely on assumptions of conditional independence, convexity, or entropy maximization, and are often sensitive to initialization choices and imposed constraints. Clustering approaches assume temporal similarity and separability of user groups, with results strongly dependent on the choice of clustering parameters and the number of clusters. Bayesian methods make their assumptions explicit through prior distributions and conservation constraints, offering a principled framework for uncertainty quantification at the cost of stronger structural assumptions and higher computational complexity.

Further analysis shows that reported performances are not directly comparable across studies, as they strongly depend on the validation reference employed. Internal or consistency-based validations, such as flow conservation or cluster stability, tend to yield favorable results but provide limited guarantees regarding OD accuracy. Survey-based validations typically report more modest correspondence rates, reflecting higher behavioral plausibility but also survey bias and limited coverage. Aggregate-level validations can exhibit very high goodness-of-fit values while concealing substantial errors at the OD-pair level. Probabilistic approaches offer a more rigorous validation framework through uncertainty-aware inference, though their conclusions remain conditional on model assumptions.

Distinct uncertainty profiles emerge across methods and data sources. For TCM-based approaches, uncertainty is dominated by behavioral assumptions and chaining thresholds, with limited propagation to final OD estimates. Statistical inference methods are primarily affected by constraint selection, initialization, and solution non-uniqueness. Clustering-based analyses concentrate uncertainty at the parameter and model-selection level, which is rarely propagated to downstream OD or forecasting tasks. Bayesian approaches explicitly quantify uncertainty, but their results remain sensitive to prior specification and stationarity assumptions. Data sources further shape uncertainty characteristics: AFC data introduce substantial uncertainty in destination and transfer inference and selection bias toward frequent users; APC data suffer from measurement noise and partial coverage; HTSD data are affected by reporting bias and limited temporal resolution; and hybrid datasets reduce some biases while introducing uncertainty related to data fusion and synchronization.

This transversal synthesis reveals several structural gaps in the literature. First, the absence of a unified comparison framework prevents meaningful cross-study performance assessment, as validation references and metrics remain heterogeneous and often implicit. Second, most approaches provide point estimates without formal uncertainty quantification or propagation. Third, OD estimation results are frequently dominated by strong behavioral assumptions that are rarely tested or discussed. Fourth, uncertainties arising from exploratory analyses and clustering are almost never propagated to ODm reconstruction or predictive models. Finally, the increasing availability of large-scale AFC data creates an illusion of precision, as data volume alone does not guarantee improved OD quality without appropriate modeling assumptions and validation strategies.

No method consistently achieves higher accuracy across all settings. Instead, TCM-based approaches tend to outperform others when individual-level AFC/AVL data with long histories are available, whereas aggregate statistical and Bayesian approaches are more suitable for sparse or underdetermined settings. Clustering-based methods achieve robust descriptive performance but are not designed for OD accuracy. Thus, reported accuracy should be interpreted as method–context compatibility rather than universal superiority.

5. Discussion

Rather than proposing a performance-based ranking of OD estimation methods, this review adopts a unifying analytical framework based on five comparative dimensions: (i) data requirements, (ii) modeling assumptions, (iii) uncertainty handling, (iv) validation reference, and (v) operational applicability. Within this framework, methods are not compared in terms of intrinsic superiority, but in terms of contextual suitability. This perspective allows reported performance differences to be interpreted as the outcome of method–data–objective alignment rather than as universal methodological dominance.

The comparative analysis conducted in this review reveals that the main challenges in public transport ODm reconstruction are not methodological in isolation, but structural. Across the literature, three fundamental trade-offs repeatedly emerge:

- individual-level reconstruction versus aggregate robustness;
- model interpretability versus predictive performance;
- data richness versus validation feasibility.

These trade-offs explain why methodological advances do not translate into systematic performance gains and why ODmsms estimated under different data and modeling contexts remain difficult to compare. These trade-offs directly correspond to the gaps identified in the review. The absence of a unified comparison framework stems from heterogeneous validation references and incompatible accuracy metrics. The limited treatment of uncertainty is closely tied to strong behavioral assumptions, which are rarely tested or propagated. Finally, the separation between exploratory behavioral analysis and ODm reconstruction results in ODm that are behaviorally plausible but operationally fragile.

Within this framework, method selection reflects a set of fundamental trade-offs rather than incremental performance improvements. Methods are typically preferred in contexts where their data requirements, behavioral assumptions, and computational complexity align with the intended operational use. For instance, trip-chain-based approaches are favored when long individual histories are available and behavioral regularity can be assumed, whereas aggregate statistical or Bayesian methods are preferred in data-sparse or underdetermined settings. Over time, methodological evolution has primarily been driven by increased data availability and computational capacity, enabling more complex models, but not necessarily resolving identifiability or uncertainty issues. As a result, higher model complexity or richer data does not systematically translate into higher OD accuracy, highlighting persistent trade-offs between interpretability, robustness, scalability, and accuracy.

A key insight from the reviewed literature is that behavioral assumptions constitute one of the dominant sources of uncertainty in ODm reconstruction. Regularity, stationarity, and continuity assumptions are not merely modeling simplifications; they condition identifiability and directly shape inferred OD flows. When behavioral heterogeneity increases—due to non-routine trips, disruptions, or multimodality—uncertainty increases accordingly, even when data volume grows. Consequently, integrating user behavior without explicit uncertainty modeling may increase apparent precision while masking structural errors.

Recent literature highlights two major trends in the estimation and reconstruction of public transport ODms.

On the one hand, there is now a widespread use of AFC data as the primary source of information. These datasets are perceived as an optimal compromise between availability, low cost, temporal resolution, and representativeness of real passenger flows. Their use enables finer-grained estimations, more accurate calibration of dynamic models, and, in some cases, the integration of behavioral components. However, this growing reliance also leads to strong heterogeneity: the quality and structure of AFC data vary significantly across networks, making direct comparison between approaches difficult.

On the other hand, the literature reveals a persistent difficulty in quantifying the uncertainty associated with reconstructed ODM. Most studies propose reconstruction methods (MLE, IPF, BE, etc.), but few explicitly address the uncertainty inherent to the estimates. This lack of a unified framework hampers the assessment of the robustness of existing methods and limits scientific reproducibility. Another recurring aspect is methodological fragmentation. The literature offers a wide range of models—probabilistic, optimization-based, statistical, machine learning, and multi-agent simulation—yet no standard protocol exists to compare methods against one another, particularly regarding accuracy, spatial–temporal scale, data requirements, or uncertainty.

Finally, several authors implicitly emphasize that ODM reconstruction should evolve toward an integrated approach combining AFC data, AVL data, network information (GTFS), external mobility datasets, and explicit uncertainty quantification.

6. Conclusions and Opportunities for Further Research

Decision-making and planning for public transport networks rely heavily on precise knowledge of network usage. With the advent of intelligent transport systems, field transport data can now be collected automatically, complementing or even replacing traditional methods such as surveys and manual counting, which are time-consuming and costly. According to Novačko, combining ODMs with network data inputs into the second step of the FSM helps rebalance the ODM [80]. In this review article, we present data, methods for reconstructing the ODM, and their performance in several concise tables.

Building on the cross-cutting synthesis and gaps identified in Section 5, the following research directions are not presented as generic trends, but as targeted opportunities directly derived from the limitations observed across data sources, estimation methods, validation strategies, and uncertainty treatment.

Rather than identifying generic research directions, this review highlights several targeted research needs directly derived from the observed gaps. First, there is a need for benchmarking frameworks that enable comparison across OD estimation methods under comparable data and validation conditions, explicitly accounting for uncertainty. Second, behavioral information extracted from AFC data should be systematically linked to uncertainty quantification, rather than implicitly embedded in reconstruction rules. Third, future research should focus on propagating uncertainty across the full OD reconstruction pipeline, from data preprocessing to operational indicators, to ensure that ODMs used for planning and optimization reflect both demand and confidence levels.

Integrating user motivations and accounting for uncertainty should not be treated as independent research directions. Behavioral heterogeneity is a primary driver of uncertainty in OD reconstruction, particularly in AFC-based methods relying on regularity assumptions. Future work should therefore aim to jointly model behavior and uncertainty, for instance, by embedding behavioral classes within probabilistic OD estimation frameworks or by coupling behavioral inference with sensitivity and scenario analysis.

Advances in artificial intelligence and multi-scale representations offer promising tools for OD reconstruction, provided that training data, interpretability, and uncertainty quantification are jointly addressed. Similarly, spatial and temporal aggregation strategies should be viewed as analytical tools rather than accuracy improvements, enabling the identification of robust demand patterns across scales.

Author Contributions: Conceptualization, J.B., G.P., M.F. and Y.D.; methodology, J.B., G.P., M.F. and Y.D.; investigation, J.B., G.P., M.F. and Y.D.; resources, J.B., G.P., M.F. and Y.D.; data curation, J.B.; writing-original draft preparation, J.B.; writing-review and editing, J.B., G.P., M.F. and Y.D.; visualization, J.B.; supervision, Y.D.; project administration, Y.D.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

Abbreviations

The following abbreviations are used in this manuscript:

Acronym	Meaning
ADCS	Automatic Data Collection System
AFC	Automatic Fare Collection
APC	Automatic Passenger Counter
AVL	Automatic Vehicle Location
BE	Bayesian Estimation
CCD	Cross Correlation Distance
C-GLS	Constrained Generalized Least Squares
DTW	Dynamic Time Warping
EM	Entropy Maximisation
FSM	Four Step Model
GTFS	General Transit Feed Specification
HTSD	Household Travel Survey Data
IPF	Iterative Proportional Fitting
LP (TT)	Linear Programming Trip Table
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MER	Minimum Entropy Rate
MLE	Maximum Likelihood Estimation
MRT	Mass Rapid Transit
ODm	Origin-Destination Matrix
ODS	Origin-Destination Survey
PCA	Principal Component Analysis
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SCD	Smart Card Data
TAZ	Traffic Analyse Zone
TCM	Trip Chaining Method
WOAFC	Weighted Origin Automatic Fare Collection
WIAFC	Weighted Itinerary Automatic Fare Collection

References

1. Cui, A. Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.
2. Lei, D.; Chen, X.; Cheng, L.; Zhang, L.; Wang, P.; Wang, K. Minimum entropy rate-improved trip-chain method for origin-destination estimation using smart card data. *Transp. Res. Part C Emerg. Technol.* **2021**, *130*, 103307.
3. Mohammed, M.; Oke, J. Origin-destination inference in public transportation systems: A comprehensive review. *Int. J. Transp. Sci. Technol.* **2022**, *12*, 315–328.
4. Vanderwaart, C.C.E. Planning Transit Networks with Origin, Destination, and Interchange Inference. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2016.
5. Bok, J.; Kwon, Y. Comparable Measures of Accessibility to Public Transport Using the General Transit Feed Specification. *Sustainability* **2016**, *8*, 224.

6. Hussain, E.; Bhaskar, A.; Chung, E. Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. *Transp. Res. Part C Emerg. Technol.* **2021**, *125*, 103044.
7. Lee, S.; Lee, J.; Bae, B.; Nam, D.; Cheon, S. Estimating Destination of Bus Trips Considering Trip Type Characteristics. *Appl. Sci.* **2021**, *11*, 10415.
8. Lee, I.; Cho, S.H.; Kim, K.; Kho, S.Y.; Kim, D.K. Travel pattern-based bus trip origin-destination estimation using smart card data. *PLoS ONE* **2022**, *17*, e0270346.
9. Lizana, M.; Choudhury, C.; Watling, D. Investigating the potential of aggregated mobility indices for inferring public transport ridership changes. *PLoS ONE* **2024**, *19*, e0296686.
10. Ait-Ali, A.; Eliasson, J. The value of additional data for public transport origin–destination matrix estimation. *Public Transp.* **2022**, *14*, 419–439.
11. McNally, M. The Four Step Model. In *Handbook of Transport Modelling*; Emerald Group Publishing Limited: Leeds, UK, 2000.
12. Yadav, G.; Ghodmare, S. Transportation Planning Using Conventional Four Stage Modeling: An Attempt for Identification of Problems in a Transportation System. *Int. J. Sci. Res. Sci. Technol.* **2021**, *8*, 556–567.
13. Hora, J.; Dias, T.G.; Camanho, A.; Sobral, T. Estimation of Origin-Destination matrices under Automatic Fare Collection: The case study of Porto transportation system. *Transp. Res. Procedia* **2017**, *27*, 664–671.
14. Radfar, S.; Koosha, H.; Gholami, A.; Amindoust, A. Improved Public Transport OD Matrix Estimation Using an Enhanced Trip Chain Model with Smart Card Data. *Int. J. ITS Res.* **2025**, *23*, 1341–1356 <https://doi.org/10.1007/s13177-025-00513-9>.
15. Cui, H.; Si, B.; Wang, J.; Zhao, B.; Pan, W. Short-term origin–destination flow prediction for urban rail network: A deep learning method based on multi-source big data. *Complex Intell. Syst.* **2024**, *10*, 4675–4696.
16. Kumar, P.; Khani, A.; He, Q. A robust method for estimating transit passenger trajectories using automated data. *Transp. Res. Part C Emerg. Technol.* **2018**, *95*, 731–747.
17. Barry, J.J.; Freimer, R.; Slavin, H. Use of Entry-Only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City. *Transp. Res. Rec.* **2009**, *2112*, 53–61.
18. Munizaga, M.A.; Palma, C. Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transp. Res. Part C Emerg. Technol.* **2012**, *24*, 9–18.
19. Gordon, J.B.; Koutsopoulos, H.N.; Wilson, N.H.M. Estimation of population origin–interchange–destination flows on multi-modal transit networks. *Transp. Res. Part C Emerg. Technol.* **2018**, *90*, 350–365.
20. Toledo, T.; Kolehkina, T. Estimation of Dynamic Origin–Destination Matrices Using Linear Assignment Matrix Approximations. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 618–626.
21. Van Zuylen, H.J.; Willumsen, L.G. The most likely trip matrix estimated from traffic counts. *Transp. Res. Part B Methodol.* **1980**, *14*, 281–293.
22. Abdelhalim, A.; Zhao, J. Computer Vision for Transit Travel Time Prediction: An End-to-End Framework Using Roadside Urban Imagery. *arXiv* **2022**, arXiv:2211.12322.
23. Rostami Shahrbabaki, M.; Safavi, A.A.; Papageorgiou, M.; Papamichail, I. A data fusion approach for real-time traffic state estimation in urban signalized links. *Transp. Res. Part C Emerg. Technol.* **2018**, *92*, 525–548.
24. Gordon, J.B. Intermodal Passenger Flows on London’s Public Transport Network: Automated Inference of Full Passenger Journeys Using Fare-Transaction and Vehicle-Location Data. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2012.
25. Ji, Y.; You, Q.; Jiang, S.; Zhang, H.M. Statistical inference on transit route-level origin–destination flows using automatic passenger counter data. *J. Adv. Transp.* **2015**, *49*, 724–737.
26. Michau, G. Link Dependent Origin-Destination Matrix Estimation: Nonsmooth Convex Optimisation with Bluetooth-Inferred Trajectories. Ph.D. Thesis, Université de Lyon, Lyon, France; Queensland University of Technology, Brisbane, Australia, 2016. Available online: <https://theses.hal.science/tel-01417805> (accessed on 1 December 2025).
27. Wu, L.; Kang, J.E.; Chung, Y.; Nikolaev, A. Inferring origin-Destination demand and user preferences in a multi-modal travel environment using automated fare collection data. *Omega* **2021**, *101*, 102260.
28. Boateng, E.Y.; Abaye, D.A. A review of the logistic regression model with emphasis on medical research. *J. Data Anal. Inf. Process.* **2019**, *7*, 190–207.
29. Lamini, C.; Benhlila, S.; Elbekri, A. Genetic Algorithm Based Approach for Autonomous Mobile Robot Path Planning. *Procedia Comput. Sci.* **2018**, *127*, 180–189.
30. Medsker, L.R.; Jain, L. Recurrent neural networks. *Des. Appl.* **2001**, *5*, 64–67.

31. Cho, H.J.; Jou, Y.J.; Lan, C.L. Time Dependent Origin-destination Estimation from Traffic Count without Prior Information. *Netw. Spat. Econ.* **2009**, *9*, 145–170.
32. Shahgholian, M. Advanced traffic management systems: An overview and a development strategy. *arXiv* **2018**, arXiv:1810.02530.
33. Galliani, G.; Secchi, P.; Ieva, F. Estimation of dynamic Origin–Destination matrices in a railway transportation network integrating ticket sales and passenger count data. *Transp. Res. Part A Policy Pract.* **2024**, *190*, 104246.
34. Ait-Ali, A.; Eliasson, J. Dynamic Origin-Destination Estimation Using Smart Card Data: An Entropy Maximisation Approach. *arXiv* **2019**, arXiv:1909.02826.
35. Bera, S.; Rao, K.V.K. Estimation of origin-destination matrix from traffic counts: The state of the art. *Eur. Transp.* **2011**, *49*, 2–23.
36. Cascetta, E.; Postorino, M.N. Fixed Point Approaches to the Estimation of O/D Matrices Using Traffic Counts on Congested Networks. *Transp. Sci.* **2001**, *35*, 134–147.
37. Nie, Y.; Zhang, H.M.; Recker, W.W. Inferring origin–destination trip matrices with a decoupled GLS path flow estimator. *Transp. Res. Part B Methodol.* **2005**, *39*, 497–518.
38. Sherali, H.D.; Sivanandan, R.; Hobeika, A.G. A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes. *Transp. Res. Part B Methodol.* **1994**, *28*, 213–233.
39. Abrahamsson, T. *Estimation of Origin-Destination Matrices Using Traffic Counts—A Literature Survey*; IR-98-021; IIASA: Laxenburg, Austria, 1998. Available online: <https://pure.iiasa.ac.at/id/eprint/5627/> (accessed on 1 December 2025).
40. Guastella, D.A.; Campss, V.; Gleizes, M.P. A Cooperative Multi-Agent System for Crowd Sensing Based Estimation in Smart Cities. *IEEE Access* **2020**, *8*, 183051–183070.
41. Behara, K.N.S.; Bhaskar, A.; Chung, E. A Novel Methodology to Assimilate Sub-Path Flows in Bi-Level OD Matrix Estimation Process. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 6931–6941.
42. Spiess, H. A gradient approach for the OD matrix adjustment problem. *Semantic Scholar* **1990**, *1*.
43. He, Y.; Zhao, Y.; Tsui, K.L. Short-term forecasting of origin-destination matrix in transit system via a deep learning approach. *Transp. A Transp. Sci.* **2023**, *19*, 2033348.
44. Tang, T.; Mao, J.; Liu, R.; Liu, Z.; Wang, Y.; Huang, D. Origin-Destination Matrix Prediction in Public Transport Networks: Incorporating Heterogeneous Direct and Transfer Trips. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 7645–7659.
45. Moreno-Palacio, D.P.; Gonzalez-Calderon, C.A.; Posada-Henao, J.J.; Lopez-Ospina, H.; Gil-Marin, J.K. Entropy-Based Transit Tour Synthesis Using Fuzzy Logic. *Sustainability* **2022**, *14*, 14564.
46. De Jong, G.; Daly, A.; Pieters, M.; Miller, S.; Plasmeijer, R.; Hofman, F. Uncertainty in traffic forecasts: Literature review and new results for The Netherlands. *Transportation* **2007**, *34*, 375–395.
47. Rasouli, S.; Timmermans, H. Probabilistic forecasting of time-dependent origin–destination matrices by a complex activity-based model system: Effects of model uncertainty. *Int. J. Urban Sci.* **2013**, *17*, 350–361.
48. Mo, B.; Koutsopoulos, H.N.; Shen, Z.J.M.; Zhao, J. Robust path recommendations during public transit disruptions under demand uncertainty. *Transp. Res. Part B Methodol.* **2023**, *169*, 82–107.
49. Ma, C.; Wang, C.; Xu, X. A Multi-Objective Robust Optimization Model for Customized Bus Routes. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 2359–2370.
50. Zhuang, D.; Wang, S.; Koutsopoulos, H.; Zhao, J. Uncertainty Quantification of Sparse Travel Demand Prediction with Spatial-Temporal Graph Neural Networks. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022; pp. 4639–4647. <https://doi.org/10.1145/3534678.3539093>
51. Cats, O.; Gkioulou, Z. Modeling the impacts of public transport reliability and travel information on passengers’ waiting-time uncertainty. *EURO J. Transp. Logist.* **2017**, *6*, 247–270.
52. Manzo, S.; Nielsen, O.A.; Prato, C.G. How uncertainty in input and parameters influences transport model output A four-stage model case-study. *Transp. Policy* **2015**, *38*, 64–72.
53. Jiang, X.; Zhuang, D.; Zhang, X.; Chen, H.; Luo, J.; Gao, X. Uncertainty Quantification via Spatial-Temporal Tweedie Model for Zero-inflated and Long-tail Travel Demand Prediction. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, Birmingham, UK, 21–25 October 2023; pp. 3983–3987. <https://doi.org/10.1145/3583780.3615215>
54. Peng, T.; Gao, J.; Cats, O. Uncertainty-aware probabilistic travel demand prediction for mobility-on-demand services. *Transp. Res. Part C Emerg. Technol.* **2025**, *181*, 105383.
55. Fan, Q.; Yu, C.; Zuo, J. Predicting Urban Rail Transit Network Origin–Destination Matrix Under Operational Incidents with Deep Counterfactual Inference. *Appl. Sci.* **2025**, *15*, 6398.

56. Mayer, D.G.; Butler, D.G. Statistical validation. *Ecol. Model.* **1993**, *68*, 21–32.
57. Vandeput, N. *Data Science for Supply Chain Forecasting*; Walter de Gruyter GmbH & Co KG: Berlin, Germany, 2021.
58. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623.
59. Fekih, M.; Bellemans, T.; Smoreda, Z.; Bonnel, P.; Furno, A.; Galland, S. A data-driven approach for origin–destination matrix construction from cellular network signalling data: A case study of Lyon region (France). *Transportation* **2021**, *48*, 1671–1702.
60. Wied, D.; Dehling, H.; Kampen, M.V.; Vogel, D. A fluctuation test for constant Spearman’s rho with nuisance-free limit distribution. *Comput. Stat. Data Anal.* **2014**, *76*, 723–736.
61. Djukic, T.; Hoogendoorn, S.; Van Lint, H. Reliability assessment of dynamic OD estimation methods based on structural similarity index. In Proceedings of the Transportation Research Board 92nd Annual Meeting, Washington, DC, USA, 13–17 January 2013.
62. Yuan, Z.; Chen, X.; Chen, B.; Luo, Y.; Zhang, Y.; Teng, W.; Zhang, C. Generating Large-Scale Origin–Destination Matrix via Progressive Growing Generative Adversarial Networks Model. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 172.
63. Rong, C.; Wang, H.; Li, Y. Origin-Destination Network Generation via Gravity-Guided GAN. *arXiv* **2023**, arXiv:2306.03390.
64. Alayón-Martínez, J.F.; Riascos, A.P. From Motifs to Lévy Flights: Modeling Urban Mobility in Bogotá’s Public Transport System. *arXiv* **2025**, arXiv:2510.02621.
65. Willumsen, L.G. *Estimation of an O-D Matrix from Traffic Counts? A Review*; Report No.: Working Paper 99; Institute of Transport Studies, University of Leeds: Leeds, UK, 1978. Available online: <https://eprints.whiterose.ac.uk/2415/> (accessed on 1 December 2025).
66. Wang, W.; Attanucci, J.P.; Wilson, N.H.M. Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems. *J. Public Transp.* **2011**, *14*, 131–150.
67. El Mahrsi, M.; Côme, E.; Baro, J.; Oukhellou, L. Understanding Passenger Patterns in Public Transit Through Smart Card and Socioeconomic Data: A case study in Rennes, France. In Proceedings of the 2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014.
68. Alsger, A.; Assemi, B.; Mesbah, M.; Ferreira, L. Validating and improving public transport origin–destination estimation algorithm using smart card fare data. *Transp. Res. Part C Emerg. Technol.* **2016**, *68*, 490–506.
69. Ge, Q.; Fukuda, D. Updating origin–destination matrices with aggregated data of GPS traces. *Transp. Res. Part C Emerg. Technol.* **2016**, *69*, 291–312.
70. Briand, A.S.; Côme, E.; Trépanier, M.; Oukhellou, L. Analyzing year-to-year changes in public transport passenger behaviour using smart card data. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 274–289.
71. El Mahrsi, M.K.; Côme, E.; Oukhellou, L.; Verleysen, M. Clustering Smart Card Data for Urban Mobility Analysis. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 712–728.
72. Ji, Y.; Cao, Y.; Liu, Y.; Guo, W.; Gao, L. Research on classification and influencing factors of metro commuting patterns by combining smart card data and household travel survey data. *IET Intell. Transp. Syst.* **2019**, *13*, 1525–1532.
73. Wu, L.; Kang, J.E.; Chung, Y.; Nikolaev, A. Monitoring Multimodal Travel Environment Using Automated Fare Collection Data: Data Processing and Reliability Analysis. *J. Big Data Anal. Transp.* **2019**, *1*, 123–146.
74. Egu, O.; Bonnel, P. How comparable are origin-destination matrices estimated from automatic fare collection, origin-destination surveys and household travel survey? An empirical investigation in Lyon. *Transp. Res. Part A Policy Pract.* **2020**, *138*, 267–282.
75. He, L.; Agard, B.; Trépanier, M. A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method. *Transp. A Transp. Sci.* **2020**, *16*, 56–75.
76. Huang, D.; Yu, J.; Shen, S.; Li, Z.; Zhao, L.; Gong, C. A method for bus OD matrix estimation using multisource data. *J. Adv. Transp.* **2020**, *2020*, 5740521.
77. Krishnakumari, P.; Lint, H.V.; Djukic, T.; Cats, O. A data driven method for OD matrix estimation. *Transp. Res. Part C Emerg. Technol.* **2020**, *113*, 38–56.
78. Blume, S.O.P.; Corman, F.; Sansavini, G. Bayesian origin-destination estimation in networked transit systems using nodal in- and outflow counts. *Transp. Res. Part B Methodol.* **2022**, *161*, 60–94.
79. Cats, O.; Ferranti, F. Voting with one’s feet: Unraveling urban centers attraction using visiting frequency. *Cities* **2022**, *127*, 103773.
80. Novačko, L.; Šimunović, L.; Krasić, D. Estimation of origin-destination trip matrices for small cities. *Promet* **2014**, *26*, 419–428.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.