Jan, 2025
GRINS DISCUSSION PAPER SERIES DP N° 123/2026 ISSN 3035-5576

GRINS

Estimation of Dynamic Origin-Destination Matrices in a Railway Transportation Network integrating Ticket Sales and Passenger Count Data

DP N° 123/2026

Authors:
Greta Galliani, Piercesare Secchi, Francesca Ieva

Jan, 2025

GRINS DISCUSSION PAPER SERIES DP N° 123/2026 ISSN 3035-5576

Estimation of Dynamic Origin-Destination Matrices in a Railway Transportation Network integrating Ticket Sales and Passenger Count Data

Greta Galliani, Piercesare Secchi, Francesca Ieva

KEYWORDS

Origin-Destination matrices | Railway transport | Passenger counting systems | Realtime data integration

JEL CODE

R41, C63, C81

ACKNOWLEDGEMENTS

This study was funded by the European Union - NextGenerationEU, in the framework of the GRINS - Growing Resilient, INclusive and Sustainable project (GRINS PE00000018). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

CITE THIS WORK

Title: Estimation of Dynamic Origin-Destination Matrices in a Railway Transportation Network integrating Ticket Sales and Passenger Count Data. Publication Date: 2025.

The last two chapters have a more methodological flavor. This study proposes an original method for estimating dynamic Origin-Destination matrices. This method is applied to the railway transportation network in Milan, and integrates data from ticket sales and Automated Passenger Counting (APC) systems. The interesting feature of the methodology proposed is that it is based on data that are potentially available in real time, so that it can be used by operators to immediately detect anomalies in the system and respond quickly.

Estimation of Dynamic Origin-Destination Matrices in a Railway Transportation Network integrating Ticket Sales and Passenger Count Data

Greta Galliani[1], Piercesare Secchi[1], Francesca Ieva[1]

[1]*Polytechnic University of Milan, Department of Mathematics*

January 7, 2025

Abstract

The last two chapters have a more methodological flavor. This study proposes an original method for estimating dynamic Origin-Destination matrices. This method is applied to the railway transportation network in Milan, and integrates data from ticket sales and Automated Passenger Counting (APC) systems. The interesting feature of the methodology proposed is that it is based on data that are potentially available in real time, so that it can be used by operators to immediately detect anomalies in the system and respond quickly.

Keywords: Origin-Destination matrices; Railway transport; Passenger counting systems; Real-time data integration
JEL codes: R41; C63; C81

1. Introduction

The present study is published in Galliani et al (2024) and aims to develop a new methodology for estimating dynamic Origin-Destination (OD) matrices specifically designed for railway transportation networks. This methodology is innovative in its integration of diverse data sources, namely ticket sales and Automated Passenger Counting (APC) systems. This fulfills GRINS goals in developing tools for supporting sustainable mobility, with a focus on green mobility and mobility of densely populated areas.

In this perspective, the core objective of the paper is to enhance the accuracy and reliability of OD matrices, which are essential tools for understanding passenger flow, optimizing transportation schedules, and making informed policy decisions. Additionally, the study seeks to address common challenges in the field, such as data incompleteness and the need for real-time analysis. Also, these latter issues are relevant to GRINS assets in terms of data fusion and gap filling of missing information.

Within the challenge of estimating dynamic Origin-Destination (OD) matrices in a railway network – which represents a critical task for effective transportation management and urban planning - traditional methods often fall short due to limitations in data availability and accuracy, particularly in the context of evolving mobility patterns post-COVID-19. This study seeks to overcome these challenges.

The paper's key objectives include:

1. Developing a robust pipeline for dynamic OD matrix estimation that can adapt to various data environments.
2. Demonstrating the application of this pipeline on the Trenord railway network in Lombardy, Italy.
3. Evaluating the effectiveness of the Iterative Proportional Fitting (IPF) algorithm in refining OD matrices derived from ticket sales and APC data.

The research is driven by the following key questions:

- Integration of Data Sources: How can ticket sales and APC data be effectively integrated to provide a more accurate and comprehensive estimation of OD matrices in railway networks?
- Methodological Challenges: What are the primary challenges in merging these two different types of data, and how can they be overcome?
- Effectiveness of the Proposed Method: To what extent does the proposed methodology improve the detection of anomalies in passenger mobility, and what are the implications for urban transport management?

2. Outline of literature background

The paper provides a thorough review of the existing literature on OD matrix estimation methods. It delves into the historical and contemporary approaches to OD matrix estimation, highlighting the evolution from static to dynamic matrices. Traditional methods like travel surveys and AFC systems are discussed, with an emphasis on their limitations in modern, complex transportation networks. In fact, pursuit of estimating OD matrices in transportation networks dates back to the late 1970s [Low, 1972, Robillard,1975, Wilson, 1970]. Early efforts heavily relied on surveys and manual data collection, posing challenges due to the sporadic availability of passenger surveys, mainly because of their cost and relevance limitations [Ben-Akiva and Morikawa, 1989].

Then the literature review covers the impact of recent advancements in data collection technologies, such as APC systems, mobile data, and smart cards, which have transformed the field of OD estimation. These methods, while valuable, often suffer from limitations such as incomplete data, the high cost of data collection, and delayed availability of results. The review includes a discussion of the benefits and challenges of these technologies, especially in terms of data accuracy, cost, and privacy concerns.

In particular, advancements in Automatic Data Collection Systems (ADCS) and mobile communication data collection have transformed the landscape. An extensive literature review on the topic may be found in [Mohammed and Oke, 2023]. Technologies like Automatic Fare Collection (AFC), Automatic Vehicle Location (AVL), and Automatic Passenger Counting (APC) have significantly enhanced the accuracy and efficiency of OD estimation. Examples of the use of such technology for OD estimation can be found in [Ait-Ali and Eliasson, 2019, Cui, 2006, Gordon, 2012, Liu et al., 2021, Torti et al., 2021, Wang, 2010, Zhao et al., 2007].

Despite not initially designed for integrated OD estimation, these technologies offer rich data, reduce collection costs, provide larger sample sizes than surveys, and automate the estimation process [Zhao, 2004]. As a result, the availability of OD matrices in the future is expected to significantly increase [Cui, 2006].

Among these technologies, AFC systems, particularly smart card data, have become the focal point of recent research [Wu et al., 2021, Yang et al., 2020, Chen and Fan, 2020]. Smart cards are widely used in transportation networks across major cities such as Milan, Paris, London, New York, Boston, Beijing, and Hong Kong [Torti et al., 2021, Zannat and Choudhury, 2019, Hussain et al., 2021], offering substantial and almost real-time insights into people's movements. Consequently, smart cards are increasingly replacing surveys in tackling OD estimation challenges. However, estimating accurate OD data for bus or railway networks without AFC systems remains challenging. A potential solution involves leveraging mobile phone data (e.g., Wi-Fi traces, call data records, or global positioning system for mobile communication data) combined with AVL to infer boarding and alighting passengers at stops [Ge and Fukuda, 2016, Jafari Kang et al., 2020, Munizaga and Palma, 2012].

The section also explores the challenges of data integration, particularly when dealing with incomplete or heterogeneous data sources. The concept of data fusion, where multiple data sources are combined to fill gaps and enhance accuracy, is a key focus. In fact, data fusion combines multiple data sources to enhance the robustness of OD estimates. To do this, the literature outlines various approaches, including the use of smart card data, mobile communication data, and advanced computational algorithms like neural networks, regression models, and gravity models. These studies underscore the benefits and limitations of each method, highlighting the potential for integrated data approaches to overcome specific challenges related to data sparsity and real-time applicability.

Our work develops an approach to estimate, with high time frequency, OD matrices. It exploits data readily available to most transportation operators equipped with APC systems but not disposing of AFC ones. Specifically, we leverage ticket and subscription data from a railway network, developing appropriate assumptions to convert this data into OD seeds. These seeds are then combined with passenger counts collected by the APC system to correct the partial estimates accounting for the totality of trips happening in the network.

In general, information provided by ADCS may be incorporated into a model for OD estimation in a twofold way. On one hand, AFC or APC counts may be integrated into iterative algorithms, in a data fusion flavour, to generate seeds of matrices estimation and/or adjust marginal counts. This approach is becoming more and more popular due to the growing availability of data coming from automated digital counting systems. On the other hand, a more traditional approach to tackle the OD estimation challenge comes from the field of trip distribution modeling. A broad review of models for human mobility is presented in [Barbosa et al., 2018]. Among them, growth factor, genetic algorithms and gravity models are some of the most widely applied techniques [Mohammed and Oke, 2023]. Growth factor models involve updating a seed OD matrix based on growth rates calibrated to match alighting and boarding counts. In this class of models, the Iterative Proportional Fitting (IPF) [Evans, 1970, Macgill, 1977], also known as the Furness method, is the state-of-the-art approach for estimating OD flows [Ji et al., 2014] and has been employed in several works in the field [Cui, 2006, Gordon, 2012, Liu et al., 2021, Torti et al., 2021, Zhao et al., 2007]. Gravity models, on the other hand, estimate OD matrices using a gravitational attraction model, incorporating population masses and distance or travel cost measures. They assume a positive association between flow volume and population size and factors in the effects of distance, space, cost, or travel time on interactions [Wheeler, 2005]. Other works rely on entropy maximization, proven to be an equivalent approach to gravity models [Ait-Ali and Eliasson, 2019, Ge and Fukuda, 2016, Wong, 2005].

Lastly, it is important to notice that many studies focus on static OD matrix estimation, describing trips between zones within a fixed reference period. However, the increasing abundance of data provided by the introduction of ACDS prompts interest in dynamic OD matrix estimation, introducing a temporal dimension which shifts the focus on monitoring mobility patterns instead of estimating them only [Khoshkhah et al., 2022].

Literature in this area is still quite limited [Ait-Ali and Eliasson, 2019, Bierlaire and Crittin, 2004, Cerqueira et al., 2022, Zeng et al., 2015, Fujita et al., 2017] and divides into dynamic a posteriori estimates [Ait-Ali and Eliasson, 2019] and real-time estimation for short-term prediction [Bierlaire and Crittin, 2004, Zeng et al., 2015].

Our proposed dynamic procedure estimates weekly OD matrices over six months, thus placing itself in the a posteriori estimates field.

3. Data and descriptive statistics

The study is geographically focused on the Trenord railway network in Lombardy, Italy. This network is crucial for regional mobility, connecting key urban centers and serving as a lifeline for daily commuters. This region represents a complex and densely populated area, making it an ideal case study for testing the proposed OD matrix estimation method. Obviously this case study focuses on highly populated urban areas, but the methodology developed is flexible enough to be adapted also in other contexts where different conditions are present.
The data covers a substantial time period from May 1, 2022, to December 31, 2022, with a focus on the months from June to December to capture seasonal variations in passenger behavior.
The study integrates three primary data sources:

- Ticket and Subscription Sales Data: Covering transactions from May 1 to December 31, 2022, detailing journeys involving 46 stations within the study scope, the dataset includes detailed records of ticket and subscription sales, encompassing 2,676,629 transactions over the study period. This data provides critical information on origin and destination stations but lacks details on the direction of travel and intermediate stops, necessitating further data integration.
- Automated Passenger Counting (APC) Data: Automated Passenger Counting (APC) data is collected using sensors that monitor passenger boardings and alightings at each station, for each train of the selected lines. This data is vital for validating and refining the OD matrices, providing a ground truth against which ticket data can be calibrated.
- Timetable Data: This data provides actual departure and arrival times for each train ride within the study period.

In the paper, detailed statistics on the data are presented, including a) visual representations of the Trenord network and the lines/stations under study, b) graphical displays of ticket sales trends, passenger counts, and train operations over time, and c) summaries of key statistics such as the number of transactions, train rides, and passenger movements, stratified by time period and station. These descriptive statistics provide a foundational understanding of the passenger flow patterns and the scope of data available for the OD matrix estimation.

4. Methodology

The paper employs a multi-step methodology centered on the Iterative Proportional Fitting (IPF) algorithm to estimate dynamic OD matrices. The approach begins by converting ticket and subscription sales data into initial OD matrices, which are then refined using APC data to account for actual passenger movements, then a gravity model is applied to estimate missing data within the OD matrices, based on factors such as the number of passengers boarding and alighting and the distance between stations.

A detailed explanation of the IPF algorithm is provided, including its role in iteratively adjusting the OD matrices to ensure consistency with observed passenger counts. The methodology is designed to be adaptable, allowing it to be applied to different transportation networks and varying data availability.

The main steps of the procedure can be summarized as follows:

- Data Preparation: This involves cleaning the ticket and APC datasets to address inconsistencies and missing values. Techniques such as imputation and smoothing are employed to ensure that the data is suitable for integration.
- OD Seed Matrix Generation: The initial step in generating the OD matrices involves using ticket sales data to create a "seed" matrix. This matrix provides a preliminary estimate of passenger flows, which is then adjusted using APC data to account for trips not captured by ticket sales alone.
- Iterative Proportional Fitting (IPF): The study uses the IPF algorithm to iteratively adjust the OD matrices. The IPF process ensures that the final matrices are consistent with both the ticket sales data and the APC counts, thus enhancing the accuracy of the passenger flow estimates.
- Anomaly Detection: The methodology includes advanced techniques for detecting anomalies in passenger flow. By continuously updating the OD matrices with real-time data, the system can identify unusual patterns that may indicate operational issues or external events affecting travel behavior.

The paper also discusses common challenges in OD matrix estimation, such as dealing with zero cells in the data and ensuring consistency in marginal totals, and the strategies employed to address these issues.

The chosen methodology is justified based on its ability to leverage the strengths of both ticket sales and APC data while mitigating their respective weaknesses. By integrating these data sources, the approach provides a more comprehensive and accurate picture of passenger movements. Additionally, the use of the IPF algorithm allows for dynamic updates to the OD matrices, and the estimation of a time-series of matrices allows the set up of an almost-real-time control system of the anomalies that may be employed for both surveillance and scenario analyses.

5. Findings and discussion

The findings of the study demonstrate the effectiveness of the proposed methodology in accurately estimating OD matrices for the Trenord railway network.
Key results include:

- Enhanced Accuracy: The integration of ticket and APC data provides a more accurate estimation of passenger flows compared to traditional methods that rely on a single data source. The refined OD matrices capture the complexity of passenger movements, including variations by time of day, day of the week, and season.
- Anomaly Detection: The study successfully identifies several anomalies in passenger flow patterns, such as unexpected spikes or drops in passenger numbers. These anomalies are linked to specific external events (e.g., strikes, special events) or operational issues (e.g., delays, cancellations). The ability to detect such anomalies in real time is highlighted as a key benefit of the proposed methodology.

A comparison is made between the proposed approach and traditional OD estimation methods, with a focus on the improvements in accuracy, timeliness, and applicability to dynamic transportation systems.
More in general, the dynamic OD matrices generated for the study period are analyzed to reveal interesting patterns that may be included in a post-hoc investigation. In the considered case, anomalies highlighted specific strikes or extraordinary intervention on the networks, confirming the ability of the actual pipeline to properly and timely identifying the disruptions and the related consequences in terms of mobility.
The discussion highlights the strengths of the proposed pipeline, such as its ability to integrate diverse data sources and its adaptability to different contexts. The accuracy of the results, as compared to traditional methods, is also discussed. The paper also acknowledges the limitations of the approach, such as potential inaccuracies in data inputs and the complexity of implementing the IPF algorithm in real-time scenarios.
Specifically, Limitations and Challenges may be resumed in the following:

- the reliance on accurate APC data,
- the need for sophisticated data integration techniques,
- computational and expert knowledge effort in scaling the methodology to other networks or regions with different data characteristics

Suggestions for improving the methodology, particularly in terms of enhancing data quality and refining the gravity model, are provided.

6. Policy implications and recommendations

Based on the findings, the paper outlines several policy implications and recommendations for urban transportation planning and management. In fact, the study provides actionable insights for transportation operators, emphasizing the importance of using dynamic OD matrices for more responsive and efficient network management. The ability to monitor and respond to real-time changes in passenger flows is highlighted as a key benefit.

Recommendations are offered for policymakers and urban planners on the use of dynamic OD matrices in long-term infrastructure planning and sustainable transport development. The paper suggests that such matrices can provide crucial data for optimizing network design, improving service quality, and reducing environmental impact.

Specifically:

- Improved Transportation Planning: The dynamic OD matrices provide valuable insights that can be used to optimize train schedules, manage passenger congestion, and plan for future infrastructure investments. This can lead to more efficient use of resources and improved service quality for passengers.
- Data Integration Policies: The study advocates for enhanced data integration policies among transportation authorities. By promoting the sharing and integration of diverse data sources (e.g., ticket sales, APC data, smart card data), authorities can achieve more accurate and comprehensive insights into passenger behavior and network performance.
- Real-Time Monitoring and Response: The paper recommends adopting real-time monitoring tools to detect and respond to anomalies in passenger flow. This can help transportation operators quickly identify and address issues, such as delays or service disruptions, minimizing their impact on passengers and improving overall network reliability.

The paper concludes with recommendations for future research, particularly in the areas of data integration and real-time OD matrix estimation. The potential for extending the methodology to other modes of transport and different geographical contexts is also discussed.

7. Key references

[1] Akbar, P., Couture, V., Duranton, G., Storeygard, A. (2023) "Mobility and congestion in urban India", American Economic Review, 113(4), 1083-1111.

[2] Ben-Akiva, M., Lerman, S.R. (1979) "Disaggregate travel and mobility choice models and measures of accessibility", In Hensher, D.A., Sopher, P.R. (Eds.), Behavioural Travel Modelling, Croom Helm, Andover, Hants, 654–679.

[3] Bunten, D.M., Fu, E., Rolheiser, L., Severen, C. (2024) "The problem has existed over endless years: Racialized difference in commuting, 1980–2019", Journal of Urban Economics, 141, 103542.

[4] Burns, L.D. (1979) "Transportation, temporal and spatial components of accessibility", Lexington Books, Lexington/Toronto.

[5] Christodoulou, A., Dijkstra, L., Christidis, P., Bolsi, P., Poelman, H. (2020) "A fine resolution dataset of accessibility under different traffic conditions in European cities", Sci Data 7, 279.

[6] Couture, V., Duranton, G., Turner, M.A. (2018) "Speed", Review of Economics and Statistics, 100(4), 725–739.

[7] Dalvi, M.Q., Martin, K.M. (1976) "The measurement of accessibility: Some preliminary results", *Transportation*, 5, 17–42.

[8] Fuchs, M., Jost, R., Weyh, A. (2024) "How many gaps are there? Investigating the regional dimension of the gender commuting gap", Papers in Regional Science, 103(1, 100005.

[9] Geurs, K.T., van Wee, B. (2004) "Accessibility evaluation of land-use and transport strategies: Review and research directions", Journal of Transport Geography, 12, 127-140.

[10] González, M.C., Hidalgo, C., Barabási, AL. (200), "Understanding individual human mobility patterns, Nature, 453, 779–782.

[11] Hansen, W.G. (1959) "How accessibility shapes land use", Journal of the American Planning Institute, 25, 73-76.

[12] Harris, C.D. (1954) "The market as a factor in the localization of industry in the United States", Annals of the Association of American Geographers, 44(4), 315-348.

[13] Ingram, D.R. (1971) "The concept of accessibility: A search for an operational form", Regional Studies, 5(2), 101–107.

[14] Jacob, N., Munford, L., Rice, N., Roberts, J. (2019) "The disutility of commuting? The effect of gender and local labor markets", Regional Science and Urban Economics, 77, 264-275.

[15] Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E. (2008) "Hole-filled seamless SRTM data V4", International Centre for Tropical Agriculture (CIAT), http://srtm.csi.cgiar.org.

[16] Kreindler, G. (2024) "Peak-hour road congestion pricing: Experimental evidence and equilibrium implications", Econometrica, 92(4), 1233-1268.

[17] Matas, A., Raymond, J.L., Ruiz, A. (2018) "Regional infrastructure investment and efficiency", Regional Studies, 52(12), 1684-1694.

[18] Miyauchi, Y., Nakajima, K., Redding, S.J. (2021) "The economics of spatial mobility: Theory and evidence using smartphone data", Tech. rep., National Bureau of Economic Research WP 28497.

[19] Nelson, A. (2008) “Travel time to major cities: A global map of accessibility”, Office for Official Publications of the European Communities, Luxembourg.

[20] Nunn, N., Puga, D. (2012), Ruggedness: The blessing of bad geography in Africa, Review of Economics and Statistics, 94(1), 20–36.

[21] Wachs, M., Kumagai, T.G. (1973) “Physical accessibility as a social indicator“, Socio-Economic Planning Science, 6, 357-379.

[22] Weiss, D.J., Nelson, A., Vargas-Ruiz, C.A. et al. (2020) "Global maps of travel time to healthcare facilities", Nature Medicine, 26, 1835-1838,

[23] Weiss, D.J., Nelson, A., Gibson, H.S. et al. (2018) "A global map of travel time to cities to assess inequalities in accessibility in 2015", Nature, 553, 333-336.

[24] Wickstrom, G.V. (1971) “Defining balanced transportation– a question of opportunity”, Traffic Quarterly, 25(3), 337–349.