Advancing Dynamic Origin-Destination Matrices Estimation Models Using Crowd-Sourced Flexibility Data

Marisdea Castiglione, Guido Cantelmo, Marialisa Nigro, Ernesto Cipriani

Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

February 15, 2025

Keywords: OD Matrices Estimation; Spatio-Temporal Flexibility; Crowd-Sourced Data; Floating Car Data; Trip Purpose.

1 INTRODUCTION

In response to the rapidly evolving challenges of urban environments, there is a growing necessity to enhance traditional Origin-Destination Matrices Estimation (ODME) models with data sources that provide broader and more comprehensive insights (Cantelmo *et al.*, 2014). Traditional fixed-location data collection tools, foundational in developing reliable traffic metrics, often fall short of capturing the complex nature of travel demand (Carrese *et al.*, 2017). Crowdsourced data, which includes mobile phone data, GPS-based data, and social media analytics, offers valuable avenues for obtaining high-resolution information that accurately reflects urban travel patterns and, in particular, it provides insights into the types of activities people engage in throughout the day. Location-based data allows us to assess activity levels at specific venues, providing an indirect measure of when and where urban activity occurs. By capturing these activity patterns, we can better characterize temporal and spatial dynamics of travel demand, recognizing that each trip is inherently associated with performing a specific activity but without making direct assumptions about its purpose.

Previous research (Castiglione *et al.*, 2024) examined the relationship between travel flexibility and the type of activities performed at destinations, leveraging data sources such as Floating Car Data (FCD) and Google Popular Times (GPT). To mitigate potential biases, GPT data is not used in its raw form; rather, it is aggregated and analyzed to extract shared activity patterns, ensuring a more representative and generalizable characterization of spatio-temporal flexibility. These studies demonstrated how flexibility parameters vary across different activities and time periods, providing valuable insights into the nature of travel demand. Specifically, six demand components C were identified, each characterized by distinct levels of temporal and spatial flexibility, and represented through a series of sample OD matrices derived from the aggregation of FCD trips based on their spatio-temporal flexibility. This body of evidence underscores the critical importance of integrating crowd-sourced data into the framework of established dynamic ODME models like the Generalised Least Squares (GLS) model (Cascetta *et al.*, 1993), enhancing their accuracy in modern urban settings.

This paper introduces the Flex-GLS approach, a novel adaptation of the GLS model designed to incorporate multiple demand components characterized by detailed spatio-temporal flexibility metrics derived from real-world, crowdsourced data. This approach aims to more accurately depict travel demand by blending temporal and spatial flexibility measures, thereby more effectively capturing the complex dynamics of urban travel. The practical implementation of the Flex-GLS model is exemplified through a case study in the EUR district of Rome, Italy, utilizing an extensive FCD dataset that recorded over 1.5 million trips between September and December 2020.

2 METHODOLOGY

Temporal Flexibility (TF) and Spatial Flexibility (SF) define an individual's ability to adjust the timing and locations of their activities, respectively. In this context, the traditional GLS model is extended to encompass multiple demand components C, each characterized by its unique spatio-temporal flexibility distribution σ_C. Given $n_{t,C}$ sample OD matrices, where each cell represents trips from an origin O to a destination D within a time interval t for a travel demand component C, derived from the aggregation of crowd-sourced data based on shared spatio-temporal flexibility, as detailed in Castiglione *et al.* (2024), the modified GLS objective function is provided as follows:

$$d^* = \arg\min_{d} \left(\sum_{t} \left(\sum_{l} w_l \cdot (v_l(d) - \hat{v}_l)^2 + \sum_{od} \sum_{C} w_C \cdot (d_{od,C} - \hat{d}_{od,C})^2 \right) \right) \tag{1}$$

Here, $v_l(d)$ denotes simulated traffic flows from a certain demand matrix d, against observed traffic counts $\hat{v}_l$. d^* represents the estimated matrix that minimises the discrepancy between simulated and observed flows. Any generic demand matrix d can then be segmented into C components where $d_{od,C}$ indicates the demand for each OD pair per component, while $\hat{d}_{od,C}$ is the seed matrix for each demand component obtained from the classified FCD in Castiglione *et al.* (2024). The weights w_l and w_C are assigned based on the inverse of traffic counts and demand component variances, respectively. Incorporating multiple demand components, while straightforward conceptually, significantly complicates the estimation process, especially for large urban networks. The Flex-GLS model, however, addresses this complexity by using conditional probabilities to treat demand components as one composite OD variable, ensuring computational efficiency. Demand components are thus defined as:

$$\begin{cases} d_{C,t,od} & = P_C(t, od) \times d_{t,od} \\ P_C(t, od) & = \frac{d^{\text{seed}}_{C,t,od}}{\sum_C d^{\text{seed}}_{C,t,od}} \end{cases} \tag{2}$$

where $P_C(t, od)$ represents the probability that a trip from origin O to destination D at time t belongs to demand component C, derived from the seed OD matrices obtained from the classified Floating Car Data (FCD). The formulation ensures that the proportion of each demand component is consistent with prior demand structure information. The Flex-GLS model then utilizes a gradient descent algorithm to estimate the demand for each OD pair and time interval t. After each gradient descent step, the Flex-GLS refines the individual demand components through a constrained Maximum Likelihood Estimation (MLE) problem, leveraging prior probabilities $P_C(t, od)$ and variances σ^2_C based on seed FCD data. The MLE constraints in the model aim to ensure data consistency, with Temporal and Spatial Flexibility treated as complementary. Temporal constraints allow adjustments in demand component proportions within a time interval t, while ensuring overall consistency across a temporal window T. This is critical for accurately capturing variations in travel behavior, considering narrower time windows for commuters versus broader windows for other, more flexible activities (e.g. shopping). Similarly, Spatial Flexibility enables the redistribution of demand from one origin O to various destinations within the same time interval, maintaining, however, the proportionality in demand components. The spatio-temporal MLE problem constraints are thus formalized as:

$$\begin{cases} \sum_C P_C(t) = 1 & \forall t \in T, \forall od \\ \frac{\sum_{t \in T} \sum_d d_{C,t,od}}{\sum_{t \in T} \sum_d d_{t,od}} \approx \frac{\sum_{t \in T} \sum_d d_{C,t,od}}{\sum_{t \in T} \sum_d d_{t,od}} & \forall C, \forall t \in T, \forall d \end{cases} \tag{3}$$

The second equation ensures that the proportion of demand allocated to each component remains stable across the estimation period. Specifically, the fraction of demand components originating from the same origin should remain consistent within a given temporal window T. This means that while individual trips within a component may shift departure times or destinations, the overall proportion of each demand component at a given origin remains stable. This constraint prevents excessive shifts in demand allocation during estimation, preserving the relative importance of each demand component over time and ensuring that temporal and spatial adjustments occur in a controlled and realistic manner.

3 RESULTS

For benchmarking purposes, initial tests of the Flex-GLS model were conducted on a toy network with one origin and two destinations to evaluate its performance across different scenarios of data reliability. These preliminary tests explored various alignments and deviations of seed data from real conditions, preparing for a more robust application. The findings inform strategic choices of modeling approaches, advocating for the Flex-GLS when detailed component data is available and for standard GLS in scenarios where data reliability may be compromised. This section presents the results from applying the Flex-GLS model to a comprehensive case study of the EUR district in Rome, Italy. The EUR district, encompassing 51 km^2 with 54 traffic zones (Figure 1[a]), provides a complex real-world environment for testing the model's effectiveness. Figure 1[b] shows the road network along with the locations of 8 traffic count detectors, reflecting the real conditions of data collection and traffic monitoring in the EUR district. Due to their placement, some detectors capture traffic from the same primary road segments, while other areas remain less covered. By incorporating crowdsourced flexibility data, the model improves the observability of OD pairs, particularly in under-monitored regions.

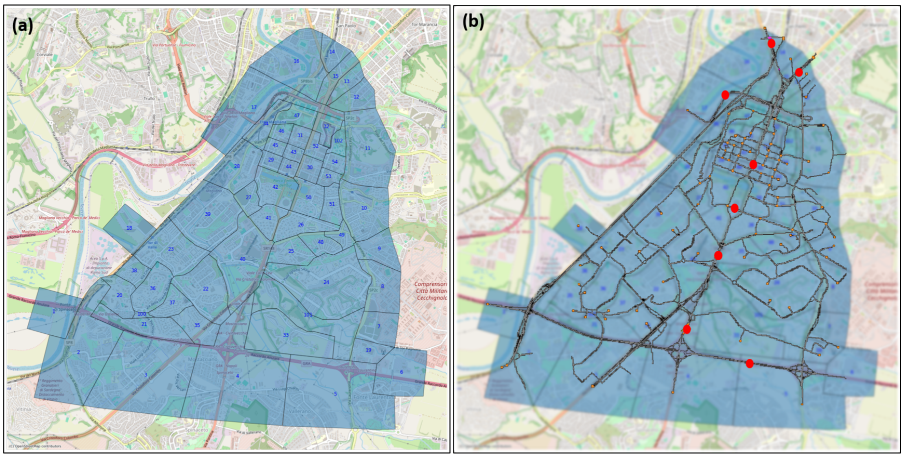

Figure 1 – *Case study: (a) Traffic Zones (b) Road Network and Detectors of the Eur district of Rome*

The primary data source for this analysis has been a FCD dataset, which includes 1.5 million car trips recorded between September and December 2020 in the Metropolitan City of Rome. A subset of 180,000 trips with destinations within the study area has been classified into 'Home', 'Work', and 'Other' categories using rule-based, spatial clustering techniques to identify regular trip patterns. Additionally, activity data was obtained from Google Popular Times for 752 Points of Interest (POI) located within the study area, collected in December 2020. From this data, six demand components were identified, each with its respective spatio-temporal flexibility distributions: Home, Work, Services (MA1), Sustenance (MA2), Shopping (MA3), and Drop-Off/Pick-Up (DO-PU). These components were analysed according to the procedure detailed in (Castiglione *et al.*, 2024). To assign the demand on the network and simulate traffic conditions, we utilize the dynamic traffic simulator Dynasmart, which employs a mesoscopic simulation framework. Both the Flex-GLS and the standard GLS models were tested across 16 time intervals from 08:00 AM to 12:00 PM, with each interval lasting 15 minutes. Each of the six components was evaluated over specific time windows to reflect varying temporal dynamics: a time window $T = 4$ time intervals for 'Home' and 'Work'; $T = 8$ time intervals for 'Services' (MA1) and 'Drop-Off/Pick-Up' (DO-PU); and $T = 16$ time intervals for 'Sustenance' (MA2) and 'Shopping' (MA3). The performance of both models in terms of RMSE is highlighted in two subsequent tables. Table 1 below presents the comparative results of both the Flex-GLS and standard GLS models from the initial to the last iteration in terms of RMSE of demand estimation accuracy and link flow reproduction. Following the initial comparison, Table 2 provides detailed performance results discretized into the six demand components, highlighting their respective abilities to handle the spatio-temporal flexibility inherent in each category.

Table 1 – *RMSE comparisons between Flex-GLS and GLS models*

RMSE	**GLS**	**Flex-GLS**
Detected Counts vs. Simulated Flows (Initial)	84.8	84.8
Detected Counts vs. Simulated Flows (Final)	33.0	25.3
Real vs. Seed Demand	19.1	19.1
Real vs. Estimated Demand	13.2	7.6

Table 2 – *Performance comparison across demand components*

RMSE	**GLS - Real**	**Flex-GLS - Real**
Home	4.3	3.5
Work	1.4	0.6
DO-PU	3.6	1.2
MA1	5.8	4.2
MA2	5.1	2.7
MA3	7.2	3.3

4 Conclusions

This paper presents the Flex-GLS model, an enhancement to the traditional GLS framework for ODME, demonstrating the significant potential of integrating crowd-sourced data. This study validates the model in a real-world setting within the EUR district of Rome, Italy, employing crowd-sourced flexibility data obtained from Floating Car Data and Google Popular times in previous research. The Flex-GLS model successfully utilizes this data to estimate six distinct demand components — Home, Work, Drop-Off/Pick-Up, Services (MA1), Sustenance (MA2), and Shopping (MA3), each characterized by unique spatio-temporal flexibility patterns. The results highlight the model's capability to outperform traditional GLS models by leveraging nuanced insights on the demand structure. This approach not only enhances accuracy in estimating traffic flows but also adapts to the variability inherent in urban travel. Particularly in the EUR district, where traffic data coverage spans less than 1% of links, the Flex-GLS model demonstrates robustness in its estimates, even with minimal traffic data. This feature is invaluable for extensive urban networks where comprehensive data collection poses significant challenges. Moreover, scalability and integration with other data sources represent further strengths of the Flex-GLS model. The model's design allows for parallel processing of the demand components adjustment during the assignment phase, ensuring that runtime remains efficient even when scaled to larger urban networks. Future developments could include integrating additional data sources such as mobile phone records or travel diaries to further enrich the estimation of flexibility and adaptability across various geographic contexts and data scenarios. Furthermore, ongoing research will extend the testing of the Flex-GLS model across varying levels of congestion to assess how traffic conditions influence the flexibility of demand components. This analysis will explore the substitution effects in response to changing travel costs, aiming to refine the model for even more dynamic and adaptive demand estimation.

References

Cantelmo, Guido, Cipriani, Ernesto, Gemma, Andrea, & Nigro, Marialisa. 2014. An Adaptive Bi-Level Gradient Procedure for the Estimation of Dynamic Traffic Demand. *IEEE Transactions on Intelligent Transportation Systems*, **15**(3), 1348–1361. Conference Name: IEEE Transactions on Intelligent Transportation Systems.

Carrese, Stefano, Cipriani, Ernesto, Mannini, Livia, & Nigro, Marialisa. 2017. Dynamic demand estimation and prediction for traffic urban networks adopting new data sources. *Transportation Research Part C: Emerging Technologies*, **81**(Aug.), 83–98.

Cascetta, Ennio, Inaudi, Domenico, & Marquis, Gérald. 1993. Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts. *Transportation Science*, **27**(4), 363–373.

Castiglione, M., Cantelmo, G., Cipriani, E., & Nigro, M. 2024. From trip purpose to space-time flexibility: A study using floating car data and Google Popular Times. *Transportmetrica B: Transport Dynamics*, **13**(1).