A GLS approach for Origin-Destination Matrices Estimation accounting for Spatio-Temporal Flexibility and Congestion

Marisdea Castiglione*[1], Guido Cantelmo[2], Ernesto Cipriani[1], Andrea Gemma[1], and Marialisa Nigro[1]

[1]Department of Civil, Computer Science and Aeronautical Technologies Engineering, Roma Tre University, Italy
[2]Department of Technology, Management and Economics, Technical University of Denmark, Denmark

Short summary

This paper introduces an enhanced framework for Origin-Destination Matrices Estimation (ODME) using the Flex-GLS model, which integrates state-specific flexibility parameters to account for varying congestion levels. Building on prior research, the study leverages Floating Car Data (FCD) and Google Popular Times (GPT) to classify travel demand into macro-activities characterized by spatio-temporal flexibility metrics, which are further utilized to enhance the ODME process. The extended model incorporates network state types, predicted using Gaussian Processes, to dynamically adjust flexibility parameters according to prevailing traffic conditions. This methodology is validated through a case study in the EUR district of Rome, utilizing extensive FCD datasets spanning 2020 and 2023. Results demonstrate that Flex-GLS outperforms traditional GLS, offering more accurate demand estimation and link flow reproduction. Moreover, the study highlights the critical relationship between congestion levels and flexibility parameters, emphasizing the model's adaptability to real-world urban mobility challenges.
Keywords: Congestion Sensitivity, Crowd-sourced data, Dynamic OD Matrices Estimation, Floating Car Data, Spatio-temporal Flexibility.

1 Introduction

The evolving dynamics of urband mobility demand ongoing advancements in Origin-Destination Matrices Estimation (ODME) models. While traditional tools, such as loop detectors for monitoring traffic flows, are necessary, they often fall short in capturing the intricacies of travel demand (Carrese et al., 2017). Conversely, crowd-sourced data, such as mobile phone and social media data as well as GPS traces provide high-resolution insights into human mobility. Therefore, integrating these data sources into ODME models can significantly improve the understanding of trip purposes and activities and offer a better overview of urban travel behavior (Timokhin et al., 2020).

We define travel flexibility as the extent to which individuals can adjust the timing and locations of their activities. In previous research (Castiglione et al., 2024), we demonstrated how flexibility is directly influenced by activity types and trip purposes by leveraging Floating Car Data (FCD) and Google Popular Times (GPT). The FCD derived flexibility metrics revealed substantial variability across activity types and time frames. In particular, six demand components, characterized by distinct temporal and spatial flexibility levels, were systematically profiled to obtain aggregated OD matrices FCD observations.

The demand components and their associated flexibility metrics have been integrated into the Flex-Generalized Least Squares (Flex-GLS) model (Castiglione et al., 2024), an enhancement of the Generalized Least Squares (GLS) framework (Cascetta et al., 1993). This extension enables the utilization of spatio-temporal flexibility insights from crowd-sourced data during the estimation process. Notably, the Flex-GLS introduces constraints to the traditional GLS, thereby mitigating the underdetermined nature of ODME. Benchmarking studies further emphasize its effectiveness in improving ODME accuracy, delivering a more refined and detailed representation of urban mobility patterns.

The performance of the Flex-GLS model has been found to vary depending on the state of the network, particularly under differing levels of congestion. This highlights the need to investigate how spatio-temporal flexibility parameters are influenced by congestion to enhance the model's responsiveness. The extent to which demand components react to congestion, in fact, depends on the trip purpose and their inherent flexibility. For instance, trips with rigid purposes, such as commuting to work or school, are relatively unaffected, while more flexible activities, like shopping or leisure, may adapt to congestion through changes in timing or destination.

This study examines how demand components behave under various congestion scenarios, seeking to answer key questions: Which trips are most likely to be canceled during severe congestion? Which trips exhibit adjustments, such as destination changes? Which remain largely consistent? By leveraging detailed FCD data, the Flex-GLS framework is refined to include congestion-aware flexibility parameters. This enhancement provides a more dynamic and versatile model, capable of representing urban travel demand under a wide range of traffic conditions.

This paper presents two main contributions:

1. **State-Specific Flexibility Parameters:** Expanding upon the Floating Car Data (FCD) classification methodology outlined in Castiglione et al. (2024), this study identifies recurring traffic network "state-types". For each state-type, unique congestion-level-specific flexibility parameters and constraints are designed for each demand component, significantly enhancing the Flex-GLS model's responsiveness to varying traffic conditions;
2. **Congestion-Responsive Flex-GLS:** Gaussian Processes (GPs) are employed to link state-specific flexibility parameters to real-time network congestion levels. This integration allows the Flex-GLS model to predict the prevailing network state-type and dynamically adjust its flexibility parameters accordingly, ensuring a more accurate alignment with real-world traffic scenarios.

The proposed methodology is demonstrated through a case study of the EUR district in Rome, Italy. Utilizing an FCD dataset of over 1.5 million trips recorded between September and December 2020, the study is further complemented by an additional dataset from the same area collected in 2023. This supplementary data facilitates a comprehensive examination of traffic states across different years and conditions. The remainder of the paper is structured as follows: the Methodology section elaborates on the congestion-aware Flex-GLS model, with a focus on its theoretical underpinnings and the interplay between Temporal and Spatial Flexibility. The Results section evaluates the proposed model's efficacy compared to the traditional GLS under varying congestion scenarios. Finally, the Conclusions section highlights the key findings and proposes venues for future research.

2 Methodology

Given $n_t \times n_C$ sample OD matrices, each cell represents trips from origin O to destination D within a specific time interval t, associated with a distinct travel demand component C. These demand components, referred to as macro-activities, group together activities that share similar levels of spatio-temporal flexibility parameters, derived from crowd-sourced data as detailed in (Castiglione et al., 2024). We define Temporal Flexibility (TF) as an individual's capacity to adjust the timing of their activities, and Spatial Flexibility (SF) as the ability to vary activity locations. Each demand component is characterized by a distribution of TF and SF values. In the extended Flex-GLS framework, the spatio-temporal flexibility distribution σ_{C,S^*} is uniquely defined for each demand component C and prevailing network state type S^*, capturing the influence of congestion on travel behavior.

During the Initialization phase, traffic metrics such as observed counts and speeds are used to predict the most likely network state type S^*. A pre-trained Gaussian Process (GP) model is employed to estimate the posterior probability of each state type $P(S|X)$, where X represents the observed traffic data and each state type S corresponds to a specific congestion condition. The most probable state type S^* is identified as:

$$S^* = \arg\max_S P(S|X) \tag{1}$$

Once the most probable state type S^* is determined, the associated spatio-temporal flexibility

parameters σ_{C,S^*} are assigned to each demand component C. This dynamic adjustment ensures that the model aligns with the real state of traffic conditions.

The traditional Generalized Least Squares (GLS) formulation (Cascetta et al., 1993) can be easily extended to multiple demand components C, each characterized by a spatio-temporal flexibility distribution σ_{C,S^*}. The modified objective function is defined as:

$$d^* = \arg\min_{d^*} \left(\sum_t \left(\sum_l w_l \cdot (v_{l,t}(d^*) - \hat{v}_{l,t})^2 + \sum_{od} \sum_C w_C(S^*) \cdot (d^*_{C,t,od} - \hat{d}_{C,t,od})^2 \right) \right) \quad (2)$$

where:

- $v_{l,t}(d^*)$: Simulated traffic flows from estimated demand d^* at time t;
- $\hat{v}_{l,t}$: Observed traffic counts at time t;
- $d^*_{C,t,od}$: Estimated demand for origin-destination pair od, demand component C, and time t;
- $\hat{d}_{C,t,od}$: Seed OD matrix for each demand component C at time t, derived from FCD;
- w_l: Weight for flow discrepancies;
- $w_C(S^*)$: Weight for demand component C under state type S^*.

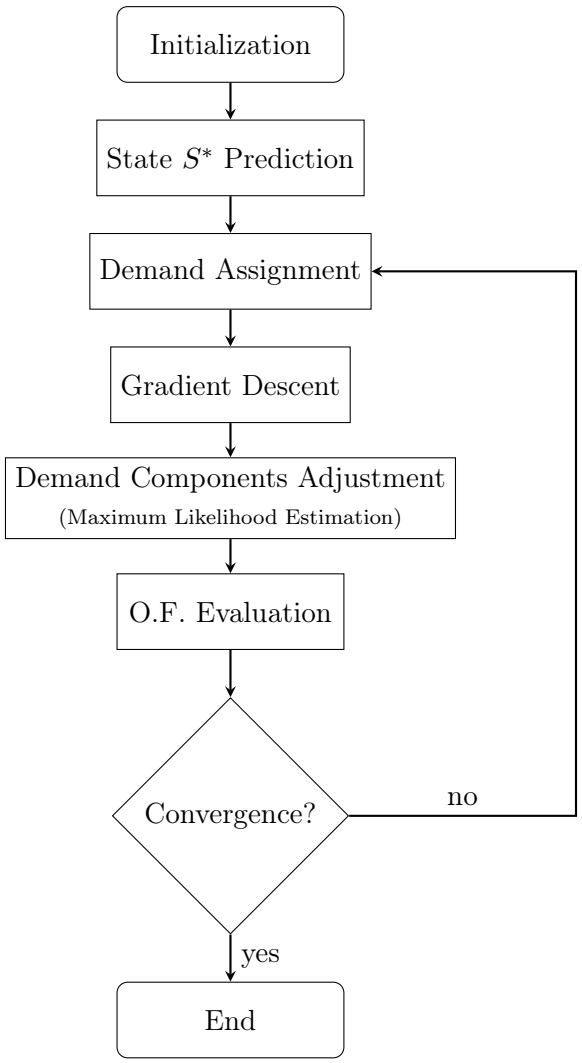

Figure 1: Workflow of the Flex-GLS Model with State-Specific Flexibility Parameters

Incorporating multiple demand components into the GLS framework significantly increases the complexity of the already intricate ODME process by adding a substantial number of variables. To address this issue, the Flex-GLS framework leverages the principle that total demand (d^*_{od}) can

be expressed as the sum of its individual components ($d^*_{od,C}$). Then, by employing conditional probabilities, the model condenses these components into a single variable, effectively reducing dimensionality and maintaining computational efficiency. The process begins with the estimation of total demand for each origin-destination (od) pair and time interval (t), followed by refinement steps that adjust component-specific demands based on temporal and spatial flexibility metrics.

Demand components are defined as:

$$\begin{cases} d_{C,t,od} = p_C(t, S^*) \cdot d_{t,od} \\ p_C(t, S^*) = \frac{d_{C,t}}{\sum_t d_{C,t}} \end{cases} \tag{3}$$

where:

- $d_{C,t,od}$: Demand for component C at time t for od under state S^*,
- $p_C(t, S^*)$: Proportion of total demand at time t for component C under state S^*,
- $d_{t,od}$: Total demand at time t for od,
- $d_{C,t}$: Aggregate demand for component C at time t.

This formulation incorporates the influence of the selected network state S^*, ensuring the disaggregation of total demand reflects prevailing congestion conditions. The objective function is adjusted accordingly:

$$d^* = \arg\min_{d^*} \left(\sum_t \left(\sum_l w_l \cdot (v_{l,t}(d^*) - \hat{v}_{l,t})^2 + \sum_{od} \sum_C w_C(S^*) \cdot (p^*_C(t, S^*) \cdot d^*_{t,od} - \hat{d}_{C,t,od})^2 \right) \right) \tag{4}$$

where:

- $v_{l,t}(d^*)$: Simulated traffic flows from estimated demand d^*;
- $\hat{v}_{l,t}$: Observed traffic counts;
- $d^*_{t,od}$: Estimated total demand for od at time t;
- $\hat{d}_{C,t,od}$: Seed OD matrix for component C, adjusted for state S^*;
- w_l: Weight for flow discrepancies;
- $w_C(S^*)$: Weight for components under S^*, inversely proportional to σ_{C,S^*}.

The weights $w_C(S^*)$ are defined as the inverse of the variance of each demand component under the selected state S^* ($w_C(S^*) = \frac{1}{\sigma_{C,S^*}}$). This approach assigns greater emphasis to more rigid components, which exhibit smaller variances, within the objective function. Consequently, the model discourages substantial fluctuations in these less flexible components, promoting stability and reliability in the estimation process.

After the estimation phase, the model enters the Demand Components Adjustment phase, where $p^*_C(t, S^*)$ is recalibrated for each demand component using flexibility metrics tailored to the identified state S^*. This adjustment step ensures consistency between the newly estimated demand and the predefined flexibility parameters. If the objective function demonstrates satisfactory convergence, the process terminates; otherwise, the model iteratively returns to the Assignment phase, as illustrated in the workflow diagram in Figure 1.

During each iteration of the Gradient Descent step, the Flex-GLS framework refines individual demand components by solving a constrained Maximum Likelihood Estimation (MLE) problem. This step utilizes prior probabilities $P_C(t, S^*)$ and corresponding distributions σ_{C,S^*}, derived from seed FCD data, ensuring alignment with the total estimated demand d^*. The model evaluates its overall estimation accuracy through the total likelihood L_{total}, calculated as the logarithm of the likelihoods of the individual components.

The MLE constraints are designed to ensure data consistency, treating Temporal and Spatial Flexibility as complementary dimensions influenced by the prevailing network state S^*. Temporal constraints allow for proportional adjustments of demand components within a time interval t,

while ensuring overall consistency across a temporal window $T(S^*)$. This distinction is particularly important for accurately representing travel behavior, such as narrower time windows for commuters versus broader windows for more flexible activities like shopping. Similarly, Spatial Flexibility allows for redistributing demand from a single origin O across multiple destinations within the same time interval, ensuring proportionality under the state S^*.

The spatio-temporal MLE problem constraints are thus formalized as:

$$\begin{cases} \sum_C P_C(t, S^*) = 1 & \forall t \in T(S^*), \forall od \\ \frac{\sum_{t \in T(S^*)} \sum_d d_{C,t,od}}{\sum_{t \in T(S^*)} \sum_d d_{t,od}} \approx \frac{\sum_{t \in T(S^*)} \sum_d d^*_{C,t,od}}{\sum_{t \in T(S^*)} \sum_d d^*_{t,od}} & \forall C, \forall t \in T(S^*), \forall od \end{cases} \tag{5}$$

These constraints enable the model to effectively capture spatio-temporal dynamics while ensuring consistency in the proportions of demand components over time and space, tailored to the specific time windows defined by S^*. The demand component constraints can be visualized as a matrix segmented into blocks across the od and t dimensions (Figure 2). The size and overlap of these blocks are determined by the flexibility constraints and the prevailing network state S^*. For example, rigid demand components, such as commuting trips, are depicted as smaller, closely overlapping blocks, reflecting limited flexibility in departure times or destination choices regardless of S^*. On the other hand, flexible demand components, such as leisure or shopping trips, are represented by larger blocks spanning multiple time intervals and states, highlighting their broader flexibility in both departure times and destination choices.

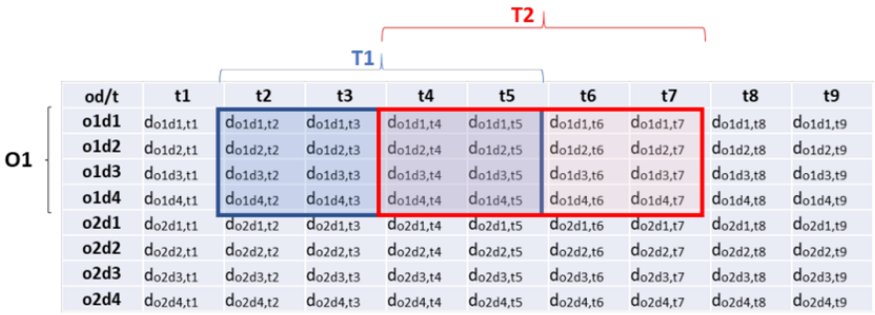

Figure 2: Visualization of constraint blocks within a demand component matrix across od, t, and state S^*

3 Results

This section details the findings from applying the Flex-GLS model to a case study conducted in the EUR district of Rome, Italy. The EUR district, covering an area of 51 km^2 and comprising 54 traffic zones (Figure 3[a]), serves as a complex, real-world testbed for evaluating the model's performance. Figure 3[b] provides an overview of the road network, highlighting the placement of 8 strategic traffic count detectors used for data collection and traffic monitoring.

The primary FCD dataset for this analysis comprises records of 1.5 million car trips collected between September and December 2020 within the Metropolitan City of Rome. A subset of 180,000 trips with destination within the study area was categorized into 'Home,' 'Work,' and 'Other' using rule-based spatial clustering techniques to identify consistent trip patterns. Additionally, activity data was sourced from GPT, capturing dynamic information for 752 Points of Interest (POI) within the area during December 2020. From these datasets, six demand components were identified, each characterized by distinct spatio-temporal flexibility distributions: Home, Work, Services (MA1), Sustenance (MA2), Shopping (MA3), and Drop-Off/Pick-Up (DO-PU) (Castiglione et al., 2024). The analysis revealed that macro-activities demonstrate varying degrees of flexibility based on the time of day and whether it is a weekday or weekend. These differences are closely tied to congestion levels, with peak hours constraining flexibility due to heightened traffic, while off-peak hours allow greater adaptability. This connection between trips flexibility and congestion levels underpins the state-specific enhancements examined in this study.

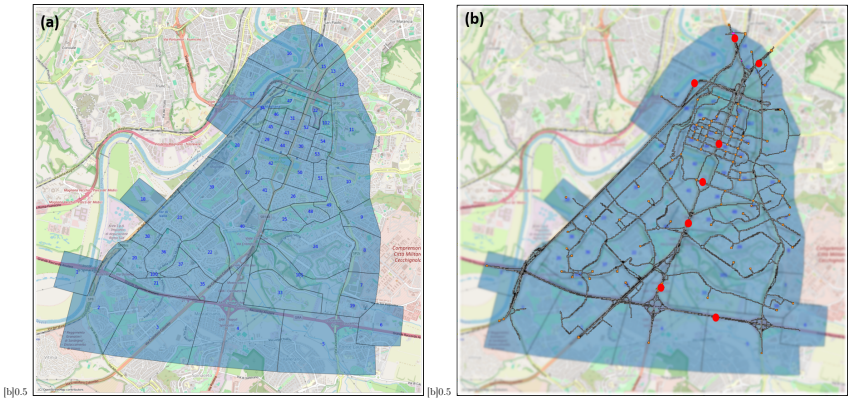

Figure 3: Case study: (a) Traffic Zones (b) Road Network and Detectors of the Eur district of Rome

To expand the scope of the analysis, an additional FCD dataset from 2023 was incorporated. This dataset captures a broader spectrum of network conditions and flexibility measures, encompassing higher congestion states absent in the 2020 data, which largely reflects reduced activity levels due to peculiar conditions. By including the 2023 data, the study accounts for a wider range of congestion scenarios, providing a more comprehensive understanding of the relationship between congestion and spatio-temporal flexibility.

The Flex-GLS and standard GLS models were evaluated across 16 time intervals, each lasting 15 minutes, between 08:00 AM and 12:00 PM. Each demand component was analyzed within specific time windows to capture varying temporal dynamics: a time window $T = 4$ for 'Home' and 'Work'; $T = 8$ for 'Services' (MA1) and 'Drop-Off/Pick-Up' (DO-PU); and $T = 16$ for 'Sustenance' (MA2) and 'Shopping' (MA3). While these time intervals were initially chosen based on general observations, the full paper will report the details of how these windows are influenced by the network's state, demonstrating their dynamic dependence on prevailing congestion conditions rather than being fixed.

The performance of GLS and Flex-GLS is analyzed under two distinct congestion scenarios. The results for a low-congestion scenario are presented in Tables 1 and 2, where the historical seed matrix was reduced by 30% to reflect the reduced congestion levels observed during the FCD data collection period. In this scenario, the Flex-GLS model consistently outperforms the standard GLS, particularly in its ability to capture the spatio-temporal flexibility of various demand components.

Subsequently, the analysis extends to a high-congestion scenario, where the flexibility parameters are calibrated based on a lower congestion state than that prevailing during the estimation procedure. While the Flex-GLS model still demonstrates superior performance compared to the standard GLS, leveraging its enhanced capacity to improve the observability of OD variables, the accuracy of individual demand component estimations decreases when flexibility parameters are misaligned with the network's current state. This underscores the critical importance of dynamically adapting flexibility parameters to accurately reflect real-time congestion conditions, ensuring optimal performance of the estimation process.

Table 1: RMSE Comparisons Between Flex-GLS and GLS Models (Low-Congestion Scenario)

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 (Low-Congestion Scenario)

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

Table 3: RMSE Comparisons Between Flex-GLS and GLS Models (High-Congestion Scenario)

RMSE	**GLS**	**Flex-GLS**
Detected Counts vs. Simulated Flows (Initial)	87.2	87.2
Detected Counts vs. Simulated Flows (Final)	42.3	35.8
Real vs. Seed Demand	21.5	21.5
Real vs. Estimated Demand	15.6	10.8

Table 4: Performance Comparison Across Demand Components (High-Congestion Scenario)

RMSE	**GLS - Real**	**Flex-GLS - Real**
Home	5.1	4.4
Work	2.1	1.2
DO-PU	4.5	2.8
MA1	6.9	5.6
MA2	6.2	4.3
MA3	8.4	5.8

4 Conclusion

This study presents a congestion-aware enhancement of the Flex-GLS model, underscoring the significance of incorporating state-specific flexibility parameters to improve demand estimation accuracy under varying network conditions. The initial results highlight the model's effectiveness in capturing the spatio-temporal flexibility of demand components, consistently outperforming the standard GLS across both low and high-congestion scenarios. However, the analysis also emphasizes the necessity of dynamically aligning flexibility parameters with the current congestion state, as misalignment can degrade performance, particularly for flexible components.

The full paper will delve deeper into the intricate relationship between flexibility parameters and network congestion states, addressing both global variations and localized disruptions. This analysis will offer a comprehensive framework for understanding the impact of congestion on the spatio-temporal flexibility of different demand components. Furthermore, the study will provide a detailed assessment of the Flex-GLS model's performance when flexibility parameters are precisely tailored to real-time network conditions. This evaluation will highlight the model's adaptability and accuracy, demonstrating its potential as a robust tool for dynamic origin-destination matrix estimation in modern urban networks experiencing diverse congestion scenarios.

References

Carrese, S., Cipriani, E., Mannini, L., & Nigro, M. (2017, August). Dynamic demand estimation and prediction for traffic urban networks adopting new data sources. *Transportation Research Part C: Emerging Technologies*, *81*, 83–98. Retrieved 2024-01-17, from `https://linkinghub.elsevier.com/retrieve/pii/S0968090X17301432` doi: 10.1016/j.trc.2017.05.013

Cascetta, E., Inaudi, D., & Marquis, G. (1993, November). Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts. *Transportation Science*, *27*(4), 363–373. Retrieved 2024-01-17, from `https://pubsonline.informs.org/doi/10.1287/trsc.27.4.363` doi: 10.1287/trsc.27.4.363

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). Retrieved from `https://doi.org/10.1080/21680566.2024.2440596` doi: 10.1080/21680566.2024.2440596

Timokhin, S., Sadrani, M., & Antoniou, C. (2020, August). Predicting Venue Popularity Using Crowd-Sourced and Passive Sensor Data. *Smart Cities*, *3*, 818–841. doi: 10.3390/smartcities3030042