Modeling Crowd-Sourced Spatio-Temporal Flexibility Insights in Origin-Destination Matrices Estimation

Marisdea Castiglione[a,*], Guido Cantelmo[b], Ernesto Cipriani[a] and Marialisa Nigro[a]

[a] Department of Civil, Computer Science and Aeronautical Technologies Engineering, Roma Tre University, Rome, Italy
marisdea.castiglione@uniroma3.it, ernesto.cipriani@uniroma3.it, marialisa.nigro@uniroma3.it
[b] Technical University of Denmark, Kongens Lyngby, Denmark
guica.dtu.dk
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 12, 2024

Keywords: OD matrices estimation; Spatio-Temporal Flexibility; Crowd-sourced Data; Floating Car Data; Trip purpose.

1 INTRODUCTION

In response to the challenges posed by the rapidly evolving urban environment, there is an emerging need to supplement traditional Origin-Destination Matrices Estimation (ODME) models with data sources that can offer broader and more insightful perspectives (Cantelmo *et al.*, 2014). While traditional fixed-location data collection tools have been foundational in establishing reliable traffic metrics, they often do not fully encompass the multifaceted nature of travel demand (Carrese *et al.*, 2017). Crowd-sourced data – which includes mobile phone data, GPS-based data, and social media analytics – offer promising avenues for gathering high-resolution information that reflects the actual travel patterns of urban travelers. In particular, the integration of location-based crowd-sourced data (Timokhin *et al.*, 2020) into ODME models can offer significant insights into the activities users engage in at destination or the purpose of their trips.

In an earlier work (Castiglione *et al.*, 2024), the authors investigated the motivations behind people's trips, estimating how travel flexibility varies in relation to the nature of the activity carried out at destination, leveraging crowd-sourced data such as Floating Car Data (FCD) and Google Popular Times (GPT). The findings have shown how flexibility parameters vary among different types of activities and over time, and have allowed detailed estimations of spatio-temporal flexibility for different components of travel demand (e.g. rigid or flexible demand components). In particular, four demand components C have been identified, each associated with a specific degree of temporal and spatial flexibility, represented by a set of sample OD matrices. These matrices are obtained from the aggregation of FCD trips, categorized based on comparable values of spatio-temporal flexibility. The evidence from earlier research highlights the necessity of integrating the insights obtained from crowd-sourced data into the framework of traditional dynamic ODME models, such as the Generalised Least Squares (GLS) model (Cascetta *et al.*, 1993).

This paper introduces the Flex-GLS approach, a novel extension of the GLS model that is designed to account for multiple demand components characterised by spatio-temporal flexibility metrics derived from real-world, crowd-sourced data. This model aims to offer a more accurate representation of travel demand by integrating both temporal and spatial flexibility dimensions to better reflect the complex dynamics of urban travel. The paper is organized as follows: the Methodology section describes the Flex-GLS model, detailing its theoretical foundations based on the concepts of Temporal and Spatial Flexibility. The Results section evaluates the model's effectiveness compared to the standard GLS model, while the Conclusions outline key takeaways and future research venues.

2 METHODOLOGY

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). 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. 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:

$$\begin{cases} d_{C,t,od} & = p_C(t) \times d_{t,od} \\ p_C(t) & = \frac{d_{C,t}}{\sum_t d_{C,t}} \end{cases} \tag{2}$$

Where $p_C(t)$ is obtained from the classified FCD sample OD matrices. 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 $P_C(t)$ and variances σ_C^2 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 constraints can be visualized as a matrix segmented into blocks across OD and t dimensions (Figure 1). Temporal and spatial flexibility constraints influence these blocks' size and overlap, with rigid components depicted as smaller, closely overlapping blocks, and flexible components as larger, spanning multiple intervals, reflecting the varying levels of flexibility in the constraints.

Figure 1 – *Representation of the constraints blocks within a demand component matrix*

3 RESULTS

For benchmarking purposes, a systematic comparison is conducted to assess the performance of the Flex-GLS model against the conventional GLS model across various scenarios. The underlying hypothesis posits that Flex-GLS, as a generalization of GLS, can offer improved accuracy in OD estimation, particularly when the demand components' proportions obtained through crowd-sourced data align closely with the real network conditions. The benchmarking methodology assesses the impact of Temporal Flexibility, Spatial Flexibility, and their joint effect on estimation accuracy. Each scenario provides insights into the models' performance and adaptability, using a "real" demand as a baseline for comparison. Although not directly observable, the "real" demand represents the actual demand generating the traffic counts on the network and offers a solid foundation for evaluating estimation accuracy. These scenarios are evaluated on both a toy network, with one origin and two destinations, and a more complex real-world network. The latter is based on the EUR district of Rome, Italy, which includes 54 traffic zones (2916 OD variables), over four 15-minute time intervals. The toy networks were designed to numerically demonstrate the conditions under which the proposed model is outperforming the traditional GLS, when the two models are the same, and when the GLS is to be preferred. Results on the toy network are presented to illustrate the model's capabilities in a simplified context before applying it to the more complex real-world scenario. The investigated scenarios are as follow:

- **Scenario 1: High congruence between seed and real demand matrices, including component ratios.** This scenario assumes a close match between the seed and real demand ($\text{RMSE}_{\text{seed}} = 5.98$), introducing minor perturbations: up to $\pm 3\%$ for the rigid component and $\pm 10\%$ for the flexible component to reflect their respective stability and variability. It demonstrates the Flex-GLS model's enhanced ability to replicate traffic flows and estimate demand components ($\text{RMSE}_{\text{F-GLS}} = 3.31$) more accurately than the traditional GLS model ($\text{RMSE}_{\text{GLS}} = 4.12$), highlighting its superiority in conditions of high data reliability.

- **Scenario 2: Divergence of total seed demand from real demand with similar component proportions.** This scenario tests the Flex-GLS model under conditions where the overall seed demand significantly diverges from the real demand, yet the components structure remains consistent. After perturbing the real demand components with random noise (±3% for rigid, ±10% for flexible components), the total seed demand is inflated by 30% while preserving component ratio integrity ($\text{RMSE}_{\text{seed}} = 29.96$). The Flex-GLS model shows robustness in reconciling substantial discrepancies between seed and real demands ($\text{RMSE}_{\text{F-GLS}} = 4.31$), outperforming GLS ($\text{RMSE}_{\text{GLS}} = 11.97$) in estimation accuracy. It also achieves improved RMSE for detected versus simulated flows and proves its effectiveness in component-wise demand estimation despite significant seed demand variances.

- **Scenario 3: High congruence between seed and real demand matrices but significant component ratio differences.** This scenario evaluates the Flex-GLS model performance with reliable total seed demand ($\text{RMSE}_{\text{seed}} = 1.75$), but inaccurate component distributions. This reveals the model limitations, indicated by high RMSE values in demand estimations ($\text{RMSE}_{\text{F-GLS}} = 12.12$), signaling considerable inaccuracies. As an extreme test case, it highlights the importance of reliable data on individual demand components for Flex-GLS effectiveness. For unreliable component structure data, a hybrid GLS and Flex-GLS approach may yield better outcomes in less extreme conditions.

4 Conclusions

The paper introduces the Flex-GLS model, a significant advancement to the traditional GLS framework for ODME, by incorporating crowd-sourced data insights. The benchmarking analysis of Flex-GLS across three scenarios reveals the model's adaptability and potential limitations. In Scenario 1, when both the seed demand and the demand component ratios are congruent with the real demand, Flex-GLS demonstrates high estimation accuracy. Scenario 2 highlights the model's robustness in adjusting demand estimations amidst significant seed demand fluctuations, however, Scenario 3 points out the model's dependence on reliable demand component data, with estimation reliability decreasing with component proportions misalignment. This outcome cautions against the Flex-GLS's use in the absence of reliable data on the structure of the individual demand components. However, it is important to note that the insights obtainable from crowd-sourced data align with these requirements. Initial tests on a simplified network underscore the Flex-GLS's potential, yet its full utility is expected in real-world applications using comprehensive data. The application of the model to the EUR district of Rome, Italy, is actually ongoing.Flexibility parameters for this application will be derived from crowd-sourced data collected between September and December 2020 in the EUR district.

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*. Under review.

Timokhin, Stanislav, Sadrani, Mohammad, & Antoniou, Constantinos. 2020. Predicting Venue Popularity Using Crowd-Sourced and Passive Sensor Data. *Smart Cities*, **3**(Aug.), 818–841.