A data fusion framework for the estimation of dynamic multimodal OD flows within urban areas

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

October 31, 2024

Keywords: data fusion, modal shares, dynamic OD estimation

1 INTRODUCTION

Quantifying dynamic Origin-Destination flows by mode of transport across various regions of an urban area is essential for understanding mobility patterns within a territory. The Covid-19 pandemic introduced new travel habits, significantly altering how people move through these spaces. Household travel surveys (HTS) are conducted periodically (every 5 to 10 years) to obtain detailed insights into the mobility of a representative sample of people (Heinen & Chatterjee, 2015). However, due to their high cost, HTS are infrequent and unable to provide a dynamic view of modal shares. In contrast, digital data sources (e.g., smart card data, mobile phone data, and sensor measurements) enable continuous, fine-grained analysis across geographic and temporal scales. However, these data taken in isolation are partial and biased, and their ability to capture complex and interrelated phenomena, such as modal shares, is still limited (Ounoughi & Yahia (2023)). Thus, apart from infrequent travel surveys, which quickly become outdated, we lack up-to-date information on modal share for trips in urban areas.

Data fusion appears to be a promising approach that can integrate disparate data sources and highlight complex phenomena such as evolving modal shares. It has already been used to update travel surveys and quantify modal shares for various ODs. Heydari *et al.* (2023) combine data from several sources to obtain evolving inter-regional mobility before and during Covid19 pandemic in Finland. Still, they do not estimate modal shares, and their method does not apply to the scale of cities. In an urban context, Graells-Garrido *et al.* (2023) update the modal shares from an HTS from 2012 with data from 2020 (mobile phone data in particular), using a matrix factorization algorithm, but they do not study the fine temporal evolution of modal shares.

This paper aims to develop a new data fusion framework that addresses the limitations above and apply it to the Lyon metropolitan area in France. We focus on estimating OD flows divided among private vehicles, public transport, and other modes for each 2-hour time slot. Seven origin/destination regions are defined: one region covering Lyon city center, five regions covering the suburbs, and one referring to the outer metropolitan area. These spatiotemporal partitions make it possible to study city-centre/suburb or suburb/suburb exchanges at different times of each day while maintaining a common coverage between our different data sources. We have access to a 2015 HTS and three recent digital data sources: smart card data, traffic counts, and mobile phone data provided as OD matrices across regions. The data share common periods: March and September 2021 and 2022.

2 METHODOLOGY

Our data fusion framework is a three-block pipeline, detailled in figure 1.

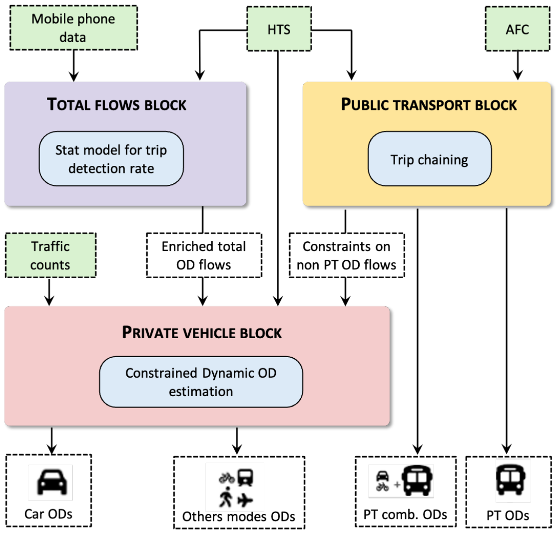

Figure 1 – *Schematic presentation of our data fusion framework for estimating dynamic OD flows split by transport mode between different regions.*

The "total flows block" aims to accurately estimate the total OD flows Z_{ijt}^{tot} between each region i and j at each timestamp t using mobile phone data. This dataset is inherently biased, as trips are recorded only if the regions of origin and destination differ, and if individuals spend at least one hour in each region. To address this bias, this block trains a logistic model on the HTS data to estimate the detection rate of trips that satisfy these two conditions between each region. Subsequently, the biased OD flows from the mobile phone data are adjusted accordingly. Concurrently, the "public transport block" estimates the OD flows associated with public transport, denoted as Z_{ijt}^{pt}. We employ the trip chaining algorithm (Egu & Bonnel, 2020) to infer the alighting stations for each boarding validation in the Lyon public transit system. We can then deduce the overall flows between regions serviced by public transport.

The "private vehicles block" separates the remaining flows into car flows and other modes (bicycle, train, intercity buses). We create a traffic network for each region i, where each node represents an intersection between main roads, serving as a potential origin, destination, or intermediate point, as outlined in Yu *et al.* (2024). The shortest path between each pair of nodes is calculated based on travel time, with each node designated as an entry, exit, or internal point within the region. These networks are constructed to reflect the actual road network, with links between nodes representing roads equipped with one or more traffic count sensors. All traffic counts within a region i at a given time t are stacked in a vector $\mathbf{y}_{it}$. Here we want to quantify the vector of flows $\mathbf{x}_{it}$ between nodes within each region i at each timestamp t. Three elements are to be built:

- Upper bounds on the car travelers OD flows U_{ijt}^{car}, between each regions i and j at time t. They are computed as $U_{ijt}^{car} = Z_{ijt}^{tot} - Z_{ijt}^{pt}$
- Binary link-route incidence matrices for each region i, A_i, that converts the flows $\mathbf{x}_{it}$ into estimated link flows as $A_i\mathbf{x}_{it}$.
- Vector of total car travelers inflows and outflows to/from region i $V_{it} = (V_{it}^{in}, V_{it}^{out})$ computed as $V_{it} = \mathbf{x}_{it}E_i$ with E_i a binary projection matrix that selects and sums up the outgoing and incoming flows of each region.

We then formulate a dynamic OD estimation model (DODE) as a non-negative least squares problem between the estimated link flow and the observed true link flow for all timesteps (H) within a given day. Solving this problem is challenging, as it is significantly under-determined. Therefore, we apply constraints to the model to identify a feasible solution:

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}\left\|\sum_{t\in H}(\mathbf{y}_{it} - A_i\mathbf{x}_{it})\right\|^2 \\
\text{s.t.} \quad & \mathbf{x}_{it} \geq 0, \quad V_{it} \leq \sum_{(i,j)} U^{car}_{ijt}
\end{aligned}
$$

Once the flows between the nodes ($\mathbf{x}_{it}$) are estimated for all regions, it is easy to compute the total car travelers inflows and outflows V_{it}. Furthermore, we assume that proportions of flows leaving each region i and that reach each region j at timestamp t, can be approximated from the bounds by: $M^{out}_{ijt} = U^{car}_{ijt} / \sum_{j'} U^{car}_{ij't}$ (we can also compute the proportion of flows reaching region j that come from region i at timestamp t as $M^{in}_{jit} = U^{car}_{ijt} / \sum_{i'} U^{car}_{i'jt}$). Then, we can compute OD flows made by cars between each region i, j in two ways: one for the flow from i to j ($V^{out}_{it} M^{out}_{ijt}$) and one for the flow to j from i ($V^{in}_{jt} M^{in}_{jit}$). We average the two quantities to get our final estimate of the car flows:

$$
Z^{car}_{ijt} = \frac{1}{2}(V^{out}_{it} M^{out}_{ijt} + V^{in}_{jt} M^{in}_{jit})
$$

3 RESULTS and DISCUSSION

Validating our model is difficult, as the ground truth about modal shares only comes from the 2015 survey. We can, however, compare what we obtain with the modal shares from 2015 to observe if any notable changes took place. For this abstract, we focus on the OD flow to Lyon from region *69091*, which is a suburban region, richly equipped with public transportation. The region has several attractive centers (universities, many tertiary and industrial activities, etc.).

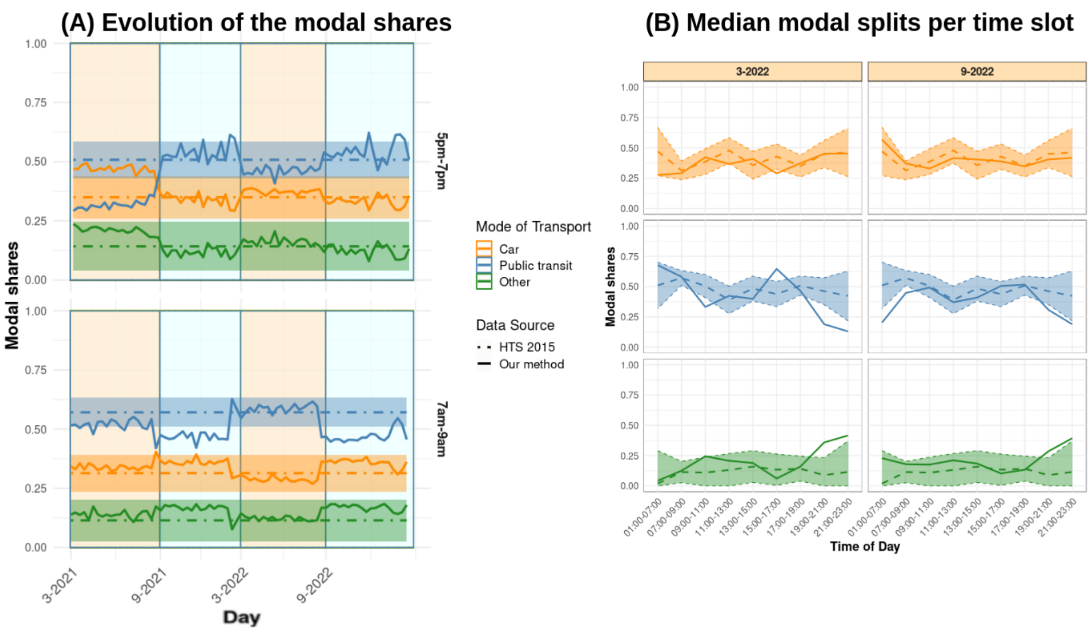

Figure 2 – *(A) Evolution of the modal shares over the four months of our study for the morning (7am-9am) and evening (5pm-7pm) peaks. The modal shares are compared to those of the HTS 2015 (fixed) with 95% confidence intervalle. (B) Comparison of median flows per mode and time slot between what we obtain with our method (march 2021 and september 2022) and estimations from the HTS 2015 with 95% confidence intervalle.*

In figure 2.A, we show the evolutions of modal shares for morning and evening peaks over the four months of our study (march and september 2021 and 2022). Since 2015, some notable changes have emerged, particularly during the evening peak in March 2021, when a curfew was imposed in response to the Covid-19 pandemic. Public transportation usage dropped significantly, with a corresponding increase in cars and other modes of transport. Public transport usage subsequently rose and stabilized in the following months. We also observe variations in modal shares between months and years. Public transport shares are higher in March than in September during the morning peak but lower during the evening peak. We attribute this difference to the varying day lengths: nightfall occurs around 6:30 p.m. in March and around 8:30 p.m. in September. In September, people may tend to return home later, concentrating public transport flows within the 5–7 p.m. time slot.

In Figure 2.B, we compare the median modal shares from March 2021 and September 2022, computed using our method, with those from the HTS. In March 2021, the curfew's impact is evident, with the evening peak for public transport occurring earlier (3–5 p.m.) compared to the HTS data (5–7 p.m.). Public transport shares have either decreased or remained stable between 2015 (HTS) and 2021/2022, with the decline most pronounced in March 2021 due to the curfew but still visible in September 2022. Overall, car shares have remained steady. Other modes (e.g., walking, bicycling) gained share in the evening between 2015 and 2021/2022, likely due to the COVID-19 pandemic, with these elevated modal shares persisting in September 2022.

By estimating dynamic OD flows by mode across different regions in a densely populated urban area, we aim to address the limitations of HTS.Our method provides valuable insights for studying mobility dynamics within urban spaces. The proposed blocks of the framework are adaptable, allowing for generalization or modification for other case studies whether data is limited or enriched by new sources. This work offers several avenues for extension. Certain assumptions within the different blocks may warrant reevaluation, as they could impact the robustness of the findings. A primary limitation is the absence of validation against an independent data source that is not used within the framework. For the Lyon case, planned surveys in 2025 present an opportunity to assess the methodology rigorously. Furthermore, our approach currently lacks a mechanism for quantifying uncertainty or estimating confidence intervals, which would enhance the reliability of our estimates.

ACKNOWLEDGEMENT

This research is conducted within the MobiTIC research project (grant number ANR-19-CE22-0010), funded by the Agence Nationale de la Recherche in France.

References

Egu, Oscar, & Bonnel, Patrick. 2020. How comparable are origin-destination matrices estimated from automatic fare collection, origin-destination surveys and household travel survey? An empirical investigation in Lyon. *Transportation Research Part A: Policy and Practice*, **138**, 267–282.

Graells-Garrido, Eduardo, Opitz, Daniela, Rowe, Francisco, & Arriagada, Jacqueline. 2023. A data fusion approach with mobile phone data for updating travel survey-based mode split estimates. *Transportation research part C: emerging technologies*, **155**, 104285.

Heinen, Eva, & Chatterjee, Kiron. 2015. The same mode again? An exploration of mode choice variability in Great Britain using the National Travel Survey. *Transportation Research Part A: Policy and Practice*, **78**, 266–282.

Heydari, Sara, Huang, Zhiren, Hiraoka, Takayuki, de León Chávez, Alejandro Ponce, Ala-Nissila, Tapio, Leskelä, Lasse, Kivelä, Mikko, & Saramäki, Jari. 2023. Estimating inter-regional mobility during disruption: Comparing and combining different data sources. *Travel Behaviour and Society*, **31**, 93–105.

Ounoughi, Chahinez, & Yahia, Sadok Ben. 2023. Data fusion for ITS: A systematic literature review. *Information Fusion*, **89**, 267–291.

Yu, Han, Zhang, Suyanpeng, Suen, Sze-chuan, Dessouky, Maged, & Ordonez, Fernando. 2024. Extending Dynamic Origin-Destination Estimation to Understand Traffic Patterns During COVID-19. *arXiv preprint arXiv:2401.10308*.