



**HAL**  
open science

## Towards replicable mode choice models for transport simulations in France

Sebastian Hörl

► **To cite this version:**

Sebastian Hörl. Towards replicable mode choice models for transport simulations in France. 9th International Symposium on Transportation Data & Modelling (ISTDM2023), Joint European Research Center, Jun 2023, Ispra, Italy. <hal-04050002>

**HAL Id: hal-04050002**

**<https://hal.science/hal-04050002v1>**

Submitted on 29 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Towards replicable mode choice models for transport simulations in France

S. Hörl<sup>a\*</sup>

<sup>a</sup> IRT SystemX, F-91120, Palaiseau, France

## Abstract

Discrete mode choice models are an important component of (agent-based) transport simulations to assess the impact of new technologies, services and policies on a territory. The present abstract describes a semi-standardized pipeline to estimate such models for French use cases. We introduce step by step the data preparation process towards the final model and sketch a specific implementation for the Île-de-France region around Paris. Furthermore, we show how the process can be adapted to other French use cases with the aim to provide an automated, replicable and case-agnostic estimation pipeline in the future.

Keywords: *transport, demand, simulation, mode choice, replicability, France*

## Introduction

Transport models play a crucial role in understanding the impact of new technologies, services, and policies on the transportation system. Their impact can be studied through simulations of entire territories. The simulations incorporate these new elements with the aim of evaluating their effects on economic, ecological, and social indicators.

Recently, a framework for creating synthetic travel demand data sets in France was developed by (Hörl and Balac, 2021). It creates households, persons, and their daily activities with times and locations in a replicable way. Since the input data sets are standardized in France and the framework is open source, it has been applied for various additional cases such as Nantes (Le Bescond et al., 2021), Lille (Diallo et al., 2021), Lyon (Hörl and Puchinger, 2022), or Rennes (Leblond, Desbureaux, and Bielecki, 2020).

Various modelling frameworks, such as MATSim (Horni, Nagel, and Axhausen, 2016) or SUMO (Lopez et al., 2018) exist which make use of the trips connecting the activities in travel demand data sets. The trips are simulated all at once, leading to phenomena such as congestion and crowding. The experienced travel characteristics can then be used to obtain updated choices for each trip's mode of transport. The choices are mandated by a behavioural model, the Multinomial Logit model (Train, 2009) being the most commonly applied approach in transportation. Extending the logic from (Hörl and Balac, 2021), in this paper, we take another step towards a fully replicable simulation of French territories by focusing on the mode choice component.

## Methodology

We propose a data preparation pipeline in six steps leading to semi-standardized mode choice models for French use cases. The first step is the **cleaning** of the **Household Travel Survey** (HTS) data. They provide the daily activities and connecting tips for a specific sample of the population in a study area. In France, several formats of such data sets exist and they are partly standardized. They are available for most of the cities and regions in France and generally available for research. Some of them are available as open data, for instance, for Nantes and Lille. Often, the surveys are spatially anonymized. However, Euclidean distances between trip origins and destinations are known, and zone identifiers (municipalities or specifically designed shapes) are provided. Based on this information, it has been shown that it is possible to synthesize representative origin and

---

<sup>1</sup> Corresponding author (sebastian.horl@irt-systemx.fr)

destination coordinates from which choice models with high quality can be estimated (Balac, Hörl, and Schmid, 2022). The goal of the cleaning process is to obtain origins and destinations of all trips in the HTS along with additional attributes on the trips and the decision-makers.

The second step is the **generation of car alternatives**, i.e., for each trip in the HTS data, we obtain a representative travel time when using the car. Our goal being to develop an easily adaptable and replicable pipeline, we don't make use of the common approach of using a commercial routing API. Instead, we propose to use data from OpenStreetMap (OSM) which is available anywhere in France to obtain travel times based on the speed limit. The routing is based on the Dijkstra algorithm and is performed using the *osmnx* package (Boeing, 2017). After, the travel times need to be corrected to represent realistic speeds. The correction is performed in two steps. In order to consider congestion, we make use of the TomTom Congestion Index (Cohn, Kools, and Mieth, 2012) which is available online and describes the inflation of travel times by hour of the day with respect to free-flow conditions in a region. On top of these hourly factors, we add another global factor that (1) corrects for the difference between maximum speeds according to the speed limit and regular free-flow speeds and (2) potential shifts in travel times between the HTS survey period and today. This global correction factor is obtained by performing a line search with the objective of minimizing the offset between the reported travel times of HTS trips that have been performed by using the car and the routed values. For validation, the resulting statistical deviation can be examined.

Third, we **generate transit alternatives** giving information on the number of transfers, in-vehicle times for individual transit modes (metro, tram, bus, ...), access times, and connection frequencies when using public transport. For that purpose, we make use of digital transit schedules in GTFS format that are generally available in France. While for some regions, like Île-de-France, consolidated GTFS data sets with all operators are available, they need to be collected and merged for other use cases. We provide the respective tools in our framework. After, the routing is performed using the RAPTOR algorithm (Delling, Pajor, and Werneck, 2015) with a specific implementation in Java that is part of the MATSim framework. During routing, waiting times and travel times for the individual transit modes can be weighted differently to represent differing levels of comfort and reliability across the transit modes. For that purpose, we use as ground truth those HTS trips for which public transport was chosen. We perform a black-box optimization using the CMA-ES algorithm (Hansen, Müller, and Koumoutsakos, 2003) to obtain routing weights that minimize the offset between the distribution of transit modes and the distribution of transfers between the HTS and the routed trips. We have found that the resulting itineraries are more realistic than bare time-minimizing routes and that they provide a better model fit in the final choice model.

As the fourth step, a **cost model** needs to be defined, which will be specific to each individual use case. While the car costs may be fixed to a general rate per distance, we also include parking costs that can be defined on a spatial zoning system and may depend on the duration of the activity following the trip. Public transport pricing systems are usually more complex and need to be modelled for every use case.

Fifth, we impute **contextual information** to the data sets. Specifically, for the current version of the generic model, we include an indicator on *parking pressure* which is the quotient of the number of cars registered in a statistical area (available for any place in France) and the total length of the road network in that area (obtained from OSM). The indicator greatly increases the explanatory power of the resulting choice model.

Finally, a **choice model** for the *chosen alternative* in the HTS can be estimated making use of trip attributes for the *car* and *transit* modes. Additional attributes for *walking* and *bicycle* are obtained, but they are modelled simplistically (for the current state of the pipeline) based on Euclidean distances. As an additional feature, we include the *car passenger* mode, which, in the downstream simulations is often a requirement to realistically simulate activity chains, especially for persons without car access (because of

age) and limited public transport supply (in rural and suburban areas). The model is a standard Multinomial Logit Model with mode availabilities based on per-person attributes such as whether the person has a driving license.

## Results

The described pipeline has been applied to the case of Paris and the Île-de-France region. The *Enquête Globale Transport* survey (Île-de-France Mobilités, OMNIL, and DRIEA, 2010) has been used which provides detailed origin and destination coordinates for research purposes. Table 1 shows the choice dimensions for the five defined transport modes. While a full documentation of the model equations and the estimated model parameters exceeds the scope of the present abstract, they will be presented at the symposium. We have obtained a promising  $R^2$  of 0.54 and Values of Time (10.50 EUR/h for car) that are in line with previous research (Meunier and Quinet, 2015).

For Île-de-France, specific developments have been performed to clean the HTS data and define the cost models for the car and public transport modes. All other steps of the pipeline are sufficiently generic so that they can directly be applied to other use cases. Furthermore, during the selection of variables, it was verified that the same choice attributes are available in other relevant HTS data sets.

**Table 1.** Attributes and Value of Time (VOT) for the estimated model for the Île-de-France region

Transport mode	Attributes
Private car	Travel time (min), Monetary cost (EUR), Parking pressure (veh/km)
Car passenger	Driving permit (Y/N), Travel time (min), Parking pressure (veh/km)
Public transport	Driving permit (Y/N), Headway (min), In-vehicle time (min), Only Bus (Y/N), Access/Egress time (min), Transfers (discrete), Monetary cost (EUR)
Bicycle	Travel time (min)
Walking	Travel time (min)

Source: Own elaborations

## Conclusion

In conclusion, we have set up a pipeline to estimate a discrete choice model for mode choice in French territories. The pipeline provides a clean basis for future improvements with a successful application to the case of Paris and Île-de-France. We have available most of the necessary data for other cities, including cleaned HTS for Nantes, Lyon, and Lille, which are based on the above-referenced synthetic population implementations. The only pieces that are missing so far are the pricing structures for these cities.

There are two interesting avenues for future research on the model itself. The first is to estimate the same or a similar model for all available French cities and compare the results. The second is to estimate a pooled model for these use cases to have a general model for France, which could have high potential for generalization.

We, furthermore, have combined the model for Île-de-France with our synthetic travel demand data for the region and integrated both into an agent-based simulation based on the MATSim framework. This activity has already provided fruitful results, although further optimization steps for network optimization will be necessary. The final goal in this line of research will be to publish a fully replicable and reproducible agent-based transport simulation for a large territory like Île-de-France in a future publication.

## Acknowledgements

This work has been supported by the French government under the “France 2030” program, as part of the SystemX Technological Research Institute.

## References

- Balac, M., S. Hörl, and B. Schmid, ‘Discrete Choice Modeling with Anonymized Data’, *Transportation*, September 15, 2022.
- Boeing, G., ‘OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks’, *Computers, Environment and Urban Systems*, Vol. 65, September 1, 2017, pp. 126–139.
- Cohn, N., E. Kools, and P. Mieth, ‘The TomTom Congestion Index’, *19th ITS World Congress*, 2012.
- Delling, D., T. Pajor, and R.F. Werneck, ‘Round-Based Public Transit Routing’, *Transportation Science*, Vol. 49, No. 3, August 2015, pp. 591–604.
- Diallo, A.O., A. Doniec, G. Lozenguez, and R. Mandiau, ‘Agent-Based Simulation from Anonymized Data: An Application to Lille Metropolis’, *Procedia Computer Science*, Vol. 184, 2021, pp. 164–171.
- Hansen, N., S.D. Müller, and P. Koumoutsakos, ‘Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES)’, *Evolutionary Computation*, Vol. 11, No. 1, March 2003, pp. 1–18.
- Hörl, S., and M. Balac, ‘Synthetic Population and Travel Demand for Paris and Île-de-France Based on Open and Publicly Available Data’, *Transportation Research Part C: Emerging Technologies*, Vol. 130, September 1, 2021, p. 103291.
- Hörl, S., and J. Puchinger, ‘From Synthetic Population to Parcel Demand: Modeling Pipeline and Case Study for Last-Mile Deliveries in Lyon’, Lisbon, 2022.
- Horni, Andreas, Kai Nagel, and Kay W. Axhausen, eds., *The Multi-Agent Transport Simulation MATSim*, Ubiquity Press, 2016.
- Île-de-France Mobilités, OMNIL, and DRIEA, *Enquête Globale Transport (2010)*, 2010.
- Le Bescond, V., A. Can, P. Aumond, and P. Gastineau, ‘Open-Source Modeling Chain for the Dynamic Assessment of Road Traffic Noise Exposure’, *Transportation Research Part D: Transport and Environment*, Vol. 94, May 2021, p. 102793.
- Leblond, V., L. Desbureaux, and V. Bielecki, ‘A New Agent-Based Software for Designing and Optimizing Emerging Mobility Services : Application to City of Rennes’, *European Transport Conference 2020*, 2020, p. 17.
- Lopez, P.A., E. Wiessner, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flotterod, R. Hilbrich, L. Lucken, J. Rummel, and P. Wagner, ‘Microscopic Traffic Simulation Using SUMO’, *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, Maui, HI, 2018, pp. 2575–2582.
- Meunier, D., and E. Quinet, ‘Value of Time Estimations in Cost Benefit Analysis: The French Experience’, *Transportation Research Procedia*, Vol. 8, 2015, pp. 62–71.
- Train, K., *Discrete Choice Methods with Simulation*, 2nd ed., Cambridge University Press, Cambridge ; New York, 2009.