



HAL
open science

Generation of Realistic Geolocated Agendas in a Digital Twin using GPS-based Mobility Dataset

Maxime Colomb

► **To cite this version:**

Maxime Colomb. Generation of Realistic Geolocated Agendas in a Digital Twin using GPS-based Mobility Dataset. NetMob 2025 - 9th edition of NetMob, INRIA, Oct 2025, Paris, France. <hal-05471353>

HAL Id: hal-05471353

<https://inria.hal.science/hal-05471353v1>

Submitted on 22 Jan 2026

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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Generation of Realistic Geolocated Agendas in a Digital Twin using GPS-based Mobility Dataset

MAXIME COLOMB^{1,2,*}

¹ASCII, Inria, CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France.

²LASTIG, Univ Gustave Eiffel, ENSG IGN, F-94160 Saint-Mande, France

*maxime.colomb@inria.fr

1. INTRODUCTION

Digital twins of cities consist in very detailed representations of chosen intricated subsystems composing urban environment. The construction of such models depends on a deep understanding of each subsystem mechanisms. It also rely on the availability, quality, consistency and richness of diverse datasets. Realistic, representative, geo-localized and historicized sets of data are then required to ensure the quality and the precision of digital twins. Transportation is a key sub-system as it moves inhabitants within places and allows to simulate multiple externalities (transportation saturation, air pollution, etc.) However, this very dynamic sub-system has an expansive modelling cost in term of computation and size of input datasets. Human mobility modelling includes knowing the reason and the destinations of individual's trips. Many applications which need mobility simulation would only require to model those two components, for example analyses of the quality of access to various services, places saturation, epidemiology, and so on.

When possible, limiting human transportation models to geolocated agendas would save modelling computing time, data availability and reduce number of hypotheses and parameters. Focusing on improving this subsystem should increase digital twins accuracy and reflect on simulations which are applied within. Very detailed knowledge is required about why and where individuals are travelling, even more when their personal characteristics are taken into consideration. Individual socio-demographic classes would help build representative categorisations of individuals in order to extrapolate behaviours to similar individuals. Different contextual elements, when applied to single individuals (telework days vs on-site day of work) or to the whole population (holiday, strike, pandemic-related restrictions...), could help understand how individuals modify their schedules. Digital twins of cities could then adapt dynamically to specific conditions.

Multiple challenges shall be addressed. Datasets, however large they are, should be sufficiently representative to allow diverse extrapolation of behaviours to different classes of population. Such a representativity must apply to population socio-demographic characteristics, location and goal of trips. Probabilistic methods shall be used to transform such representative data into an exhaustive model. Rare events which are not captured by datasets shall also be modelled. Another challenge is to integrate such modelled information into a well described and feature-rich digital twin. Designed methods must be generic and optimized enough to apply to multiple contexts and on very large urban models.

This work aims to simulate this minimal aspect of mobility: generating individual agendas and locate each steps. Such information could then be injected in a digital twin of community and individuals in order to model how places are occupied and how individuals meet. A new model is designed and presented in this poster to perform such tasks in an efficient way.

It uses the well detailed dataset provided by the *NetMob 2025* Data Challenge [1]. Probabilistic tools are generated to provide a framework of agenda generation. Simulation of geo-localised agendas can then be applied to every individuals of the digital twin.

2. METHOD

Few distinct steps are needed to achieve such a goal. A probabilistic mobility framework shall first be generated (section A and B). The simulation of individual agendas can then be generated (section C). Adaptation of NetMob Data Challenge and results about the occupation are presented in a last sections (3 and 4)

A. Markov chain parametrization

We designed a Markov Chain, represented in Figure 1, which models the main types of activities available for any individuals. Probabilities to switch from an activity to another are set using calibration data. This chain is calibrated respectfully to a determined number of socio-demographic categories, for each time lapse of the simulation and for different types of days. A probabilistic list of activities can then be generated for every kind of individual. As specified in [1],

Researchers should always consider (...) whether to disaggregate by day type or geographic unit.

A statistical verification must ensure that each discrete class contains enough values to be sufficiently representative, otherwise categories would be merged with neighbourhood categories.

As Markov Chains are let unconstrained, individuals can switch from an activity to any other one at every time of the simulation. Imposing events (lunch break, get children from school, etc.) could eventually be added, but the law of large numbers should reproduce such event at a satisfying level.

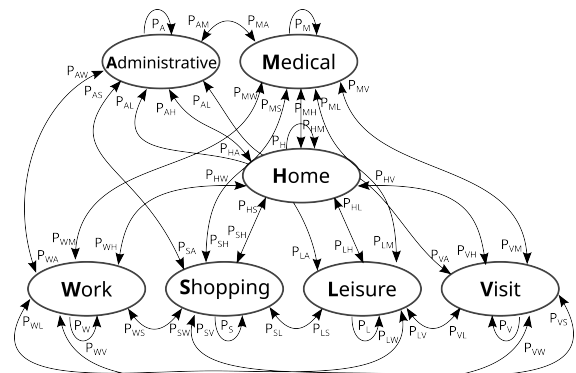


Fig. 1. Markov chain for activity selection

B. Micro-zoning with mobile data

The presence of different types of individuals at specific times and places of the city should be modelled using time-dependant origin and destination traces. A spatial aggregation is performed in order to determine a destination area rather than a precise place. Origin-Destination matrix are calculated, based upon mobility data. For each origin aggregate, the probability of going to any other aggregate is calculated based on all the weighted trips happening at that time and from that place. Such matrix are disaggregated to hours of the day and could also be disaggregated to few population classes if data are representative enough. Aggregates can be merged to meet statistical representativeness requirements. An example of OD Probability matrix can be found in Figure 2.

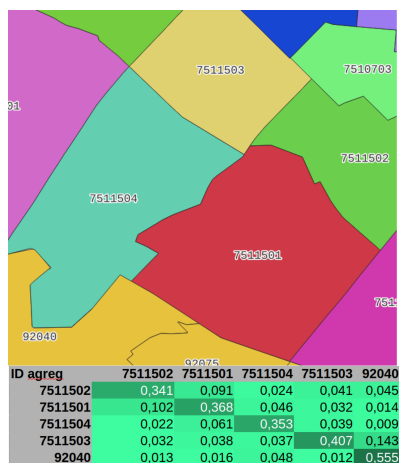


Fig. 2. Probability of going to different neighbourhood of the 15th arrondissement of Paris

C. Agenda generation

An agenda can be generated for every individual of the simulation by using the appropriate Markov Chain parameters. Location of some of the activities are linked to each individual by the digital twin (home and work location), some others are fluctuant and need to be spatialized. An IRIS is chosen using the OD probability matrix and within it, a random place with a type corresponding to the activity is attributed. The respect of the digital twin characteristics (type of activities open to the public, opening hours, level of frequentation, etc.) should reproduce places occupation in a realistic way.

3. INTEGRATION OF NETMOB DATA CHALLENGE

Data gathered for the NetMob Data Challenge [1] were the only mobility data used in this work. In order to improve speed of all requests on data that would happen latter on, the Net-Mob datasets have been loaded on a PostGIS database. This model have initially been designed with survey for Markov chain parametrization and mobile data for OD matrix calibration. The latter dataset provides a high number of different traces, and the size of statistical aggregates can be small. Nevertheless, the fact that the same dataset calibrate every probabilistic model brings a unique consistency to the results.

4. RESULTS

Figure 3 shows the frequentation of a driving school in the 11th arrondissement of Paris. Frequentations of every activity can be found here: <https://ici.saclay.inria.fr/simu/netmobresult.php>. This

digital twin, described in ([2] - submitted), cover the entire Île-de-France and includes 2 070 444 activities, 473 348 being open to the public. For now, frequentation for each arrondissement is presented separately.

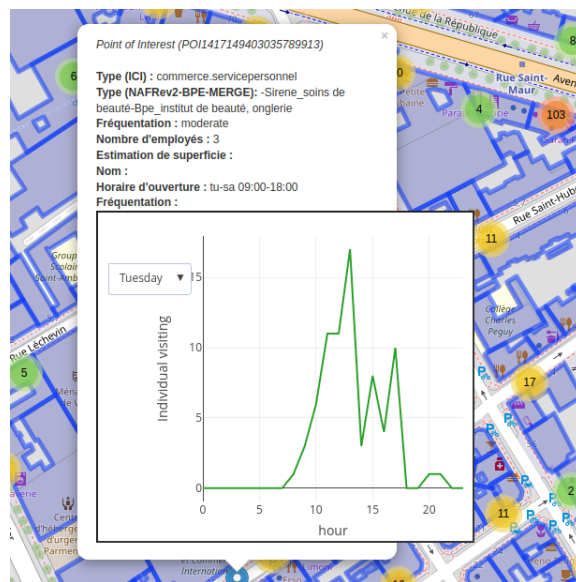


Fig. 3. Example of visits in an activity open to the public in an Île-de-France digital twin

5. CONCLUSION AND PERSPECTIVE

This model is actively used by the ICI project [4], which couples a digital twin with a contagion model. Sensitivity analysis should qualify uncertainty of this agenda generator and how its variability is affecting ICI epidemiological simulations. The model of the effect of particular events (strike, telework) on mobility behaviour necessitate the use of complementary datasets. For example, [5] could be used to model the influence of COVID-19 lockdowns on individual mobility.

6. BACK MATTER

Funding. This work will benefits from the Mob Sci-Dat Factory project [6].

Acknowledgment. This work have been developed as a part of the ICI project, involving the ASCII team of INRIA, LASTIG of IGN, CMAP of École Polytechnique, CRESS of INSERM.

Disclosures. The author declares no conflicts of interest.

REFERENCES

1. A. Chasse, A. J. Kouam, A. C. Viana, *et al.*, arXiv (2025).
2. M. Colomb and J. Perret, IJGIS (2025).
3. C. Lecomte, G. Le Roux, H. Commenges, and J. Vallée, "Le Mobiliscope. Un outil de géovisualisation de la composition sociale 24h/24 des territoires," in *Spatial Analysis and GEomatics 2017*, (INSA de rouen, Rouen, France, 2017).
4. M. Colomb, Q. Cormier, J. Garnier, *et al.*, "ICI Project: Individual-based epidemic propagation simulator based on a Digital Twin," in *Health GIS: Spatial Thinking in Applied research 2024*, (O'Brien Institute for Public Health (OIPH) and University of Calgari, Calgari, Canada, 2024).
5. H. C. Fantecele, S. Rabenjamina, A. Carneiro Viana, *et al.*, "SafeCityMap (1st phase) – COVID INRIA mission: Investigating population mobility habits in metropolitan zones and the lockdown impact using mobile phone data," Research report, Inria (2021).
6. "Projet Cibl  Mob Sci-Dat Factory du PEPR MOBIDEC," (2025). [Online; accessed 15. Jul. 2025].