Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

A DEMAND MODELLING PIPELINE FOR AN AGENT-BASED TRAFFIC SIMULATION OF THE CITY OF BARCELONA

Jonas F. Leon

Spindox S.A.
Universitat Oberta de Catalunya
156 Rambla del Poblenou
Barcelona, 08018, SPAIN

Francesca Giancola
Andrea Boccolucci
Mattia Neroni

Spindox S.p.A.
via Bisceglie, 76
Milan, 20152, ITALY

ABSTRACT

The growth of urban population and the proliferation of mobility options in big cities are adding to the complexity of comprehending how people move about and how efficiently they do it. Understanding how traffic patterns change throughout the day is essential for legislators, public administrations, and other stakeholders, as it has a direct impact on citizens' quality of life by, for instance, increasing greenhouse gas emissions and noise pollution. In this context, simulation becomes an essential tool for grasping the emerging dynamics of urban transportation, citizens' mobility patterns, and traffic flow bottlenecks. This work presents a complete data modelling pipeline for generating the population, network and transportation demand that is fed to a multi-modal traffic simulation of the city of Barcelona using MATSim and open-access statistical data sources. The model is calibrated, the results are obtained, and future applications of the developed tool are outlined.

1 INTRODUCTION

In recent years, urban mobility has become increasingly complex due to several factors. Some of these interrelated factors are *(i)* the population growth and its concentration in cities, *(ii)* the evolution of mobility options and modes of transportation, and *(iii)* the change of citizens' habits and behaviors. The first factor listed, i.e. the fact that people gradually concentrating in cities, has straightforward consequences: the number of potential interactions and the aggregated transportation demand increase. The second factor, related to the new types of vehicles (e.g., electric, connected and autonomous vehicles) combined with new services emerging from the so-called 'on-demand economy', are expected to boost the number of transport and mobility operations in cities and metropolitan areas. The final mentioned factor is the change in transport demand, influenced by the launch of new "e-commerce" products and services, car-sharing and ride-sharing platforms, or new work-related practices such as working from home after Covid-19 pandemic (Kuo et al. 2023).

All these technological and social changes arise complex operational challenges for city planners. For instance, when re-designing public spaces or urban roads, decision-makers might wonder: what is the impact of a road network redesign in terms of greenhouse gas emissions or in noise pollution? where should charging stations or shared vehicle fleets be located? In the present study, we explored the combination of two different elements to face the challenge posed by the previous questions. The first is the open-access data from the local government of Barcelona, in particular the "Open Data BCN" public repository and the periodic statistical reports on mobility and sociodemographic data. The second is the open-source simulation tool MATSim. Being a modular agent-based and activity-based simulator, MATSim allows approaching the traffic modeling from an individual traveler perspective, from which aggregated city-level metrics for mobility flow emerge.

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

In terms of academic contributions, the generation of reliable transport demand profiles based on purely open-access data through a reproducible pipeline is lacking in the literature (Hörl and Balac 2021). Furthermore, although partial models of the city of Barcelona exist, they do not cover multi-modal transport for the whole area of the city. Therefore, to the best of the authors' knowledge, this is the first work proposing a pipeline to synthesize transport demand in the city of Barcelona considering a multi-modal network and using purely open-access data and open-source tools.

The rest of the paper is structured as follows: Section 2 presents a short literature review on some of the key topics analyzed in this paper. Section 3 describes the tool and the datasets used in this study, while Section 4 explains the proposed modelling approach. Section 5 describes the calibration of the model and presents the results. Finally, Section 6 draws the main conclusions of this work and points out lines of future research.

2 RELATED WORK

The simulation of the traffic at a city level is a complex challenge that spans across multiple dimensions, such as the spatial coverage of the model, the level of detail to be considered or the interaction between the modeled components (Weyl et al. 2019). Historically, there have been different approaches to study urban mobility, ranging from macroscopic wave-like modelling, to detailed agent-based simulations (Miller et al. 2017). Regardless of the simulation paradigm employed (micro-, meso- or macro-simulation, discrete-events or discrete-time, trip- or activity-based, etc.), researchers identify the same types of problems, namely: *(i)* data acquisition (Bassolas et al. 2019), *(ii)* model calibration and validation (Bowman et al. 2022) and *(iii)* model scaling (Ma and Fukuda 2015). In recent years, activity-based micro-simulation seems to be gaining ground, since some of the aforementioned problems are diminishing due to the improved access to data at agent level or the increase in computational processing power (Moeckel et al. 2020). The data acquisition has been simplified by the effort of state agencies, private companies and volunteers from the open community, that provide tools and data to the general public (Neves et al. 2020). On the other hand, the increase in computational power has allowed the scientific community to exploit the advantages of agent-based traffic micro-simulations, such as their interpretability and adaptability by controlling individual interaction between agents (Gräbe and Joubert 2022).

In order to implement this type of traffic micro-simulation, there are numerous possible alternatives available to researchers and practitioners. There are some software dedicated to traffic micro-simulation like SUMO, MATSim or libraries like Python CityFlow. These software packages provide ready-to-use environments for simulating traffic and to handle road networks, vehicles, etc. Equally, powerful commercial options like Vissim, Aimsun or Paramics exist, but in spite of offering free-trial versions, their source code is proprietary, limiting reproducibility and community-based extensibility. Finally, generic agent-based simulation software can also be used to build traffic models, such as NetLogo, Python libraries like Mesa or Pandora, or commercial software like AnyLogic (Mohebbi and Murali 2022). These non-specific agent-based software has the advantage of being more flexible, but they require much more effort in order to be able to describe the road network and the dynamic and interaction of the agents. In (Mahmud and Town 2016), a comprehensive review of different traffic simulation software is provided. For the particular case of MATSim, apart from enjoying the already mentioned micro-simulation advantages, its open modularity contributes also to expand its simulation capabilities. For instance, it allows incorporating the study of multi-modal public transportation (Poletti et al. 2017), or the investigation of new mobility trends in urban areas, such as car-sharing (Ciari et al. 2013).

For the city of Barcelona, there are some publications that have covered the study of its mobility dynamics using a micro-simulation description. In Argota Sánchez-Vaquerizo (2022), researchers approached the modelling of Barcelona similarly as in the present work, but using SUMO as their simulation package and being limited to private cars as the only mode of transport. In Bassolas et al. (2019), researchers employed MATSim for simulating the city of Barcelona, but the main data source for modeling the demand were mobile phone records. The results helped to study different policies in the Metropolitan Area of Barcelona.

Finally, Montero et al. (2017) considers a traffic flow simulation, modeled using AIMSUM, but it is limited to a single Barcelona district. Our simulation model, covering the multi-modal mobility of the entire city of Barcelona, can help public and private organizations to make better and more informed decisions, such as, for instance: where to place electric charging stations, which roads to open or close to traffic, or how to dimension the public bus fleet.

3 DATA SOURCES AND SIMULATION TOOL

MATSim (Horni et al. 2016) is an open-source simulation framework that allows mobility analysis for large-scale scenarios, in the present case, at city level. In MATSim, individual agents (people) are modeled by assigning them a spatial location, some interaction rules and an activity plan (i.e., a series of time-scheduled activities) that they try to undertake. In aggregate, these individual agents constitute the population to be simulated (or a representative percentage of it), and their activity plans constitute the aggregated transport demand. In that sense, it could be considered an agent-based, and also an activity-based, simulation model. It can also be classified within the traffic micro-simulation paradigm, since the interaction of individual vehicles-like agents can be studied directly, following a queue-based approach as opposed to the more computationally expensive car-following behavior. The different agent plans compete with one another following a stochastic co-evolutionary algorithm that reaches an equilibrium after a certain number of iterations. In the equilibrium, every agent is carrying out its preferred plan to the best of its possibilities (e.g., without delays), and cannot further improve it without deteriorating another agent's plan.

This study has been produced using both open-access data and open-source software, allowing the wider scientific community to consult and modify the code, and potentially to reproduce the pipeline presented in the following sections. Table 1 present the main data sources that were employed to feed the demand pipeline. All simulations were run using an Intel Xeon E5-2630 v4 multi-core processor with 32GB RAM.

Table 1: Open data sources.

Source	Description
A. OpenStreetMap	Open geographic database used to extract the nodes and links that compose the network of Barcelona used for the simulation.
B. Open Data BCN	Public repository of data for the city of Barcelona. It contains hundreds of datasets regarding population, transportation and services.
C. ESDB	Sociodemographic survey of Barcelona for the year 2020. It contains comprehensive social and economic data for the city of Barcelona.
D. EMEF	Working day mobility survey for the city of Barcelona (2021). The survey summarises the daily transportation habits of Barcelona's citizens.

4 MODELLING APPROACH

In this section, the modelling approach followed is described in detail. In particular, the generation process for the three main ingredients of the simulation model is described, namely the network, the population and the transportation demand. The demand generation pipeline is graphically described in Figure 1. It is important to stress that the proposed pipeline is merging together very different and huge datasets, that are mostly related to spatial and sociodemographic distributions of people, and, after synthesizing a population of individual agents, trying to infer the mobility dynamic of an entire city.

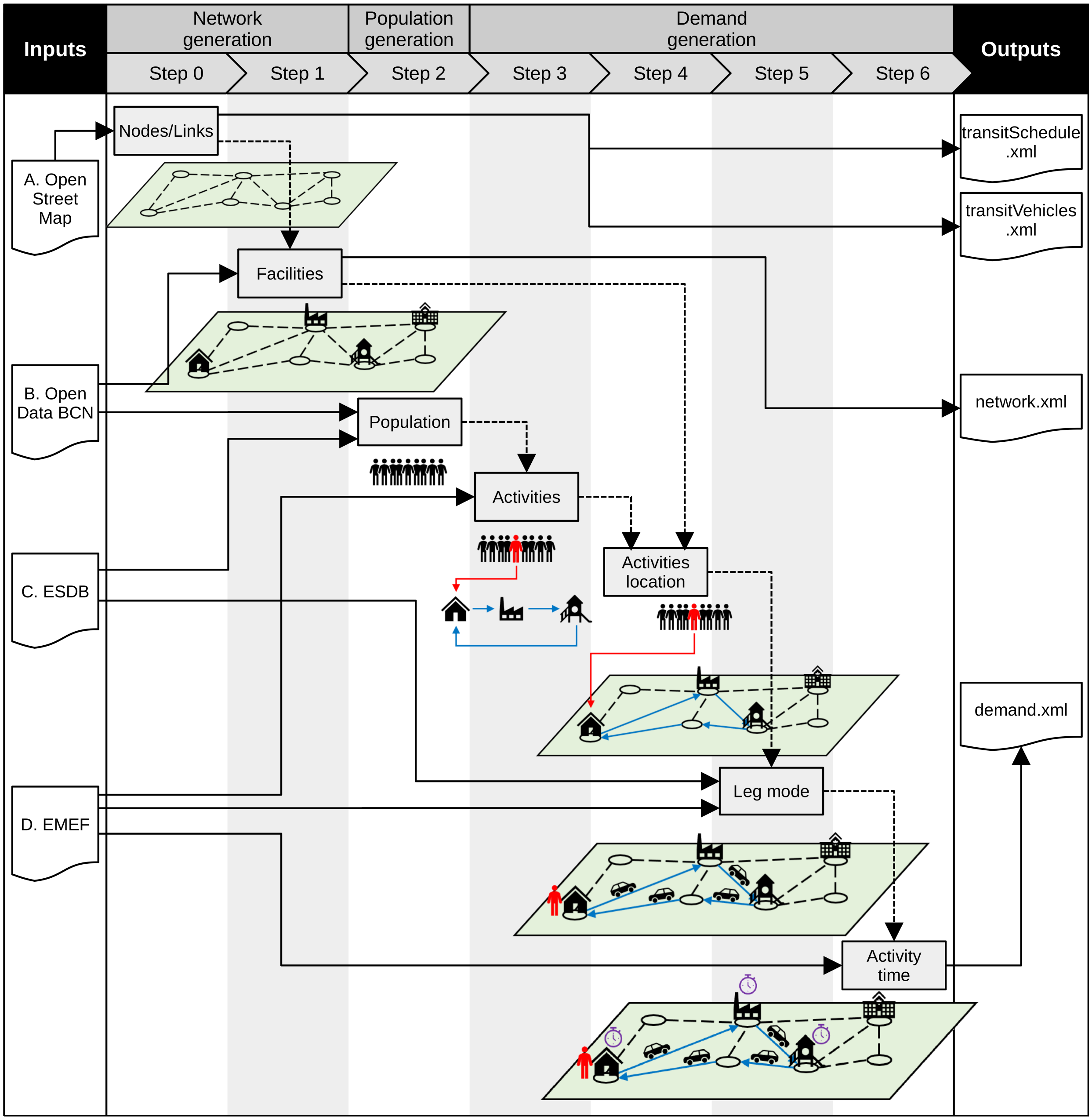

Figure 1: Demand generation pipeline. The steps followed are described in detail in Section 4 and the input data sources are contained in Table 1. Infographics are inspired and adapted from Simunto MATSim tutorials.

4.1 Network Generation

In this section, the geographical scope of the simulation is defined, and the first two steps of the demand pipeline are explained. The steps consist in the generation of the network composed of nodes and links, and the localization of facilities within that network. The modelled area covers the approximately 101 km^2 of the urban area of Barcelona, and includes the 10 districts in which it is divided, with its corresponding 73 neighborhoods. Figure 2 shows a map view of the target extension of the simulation.

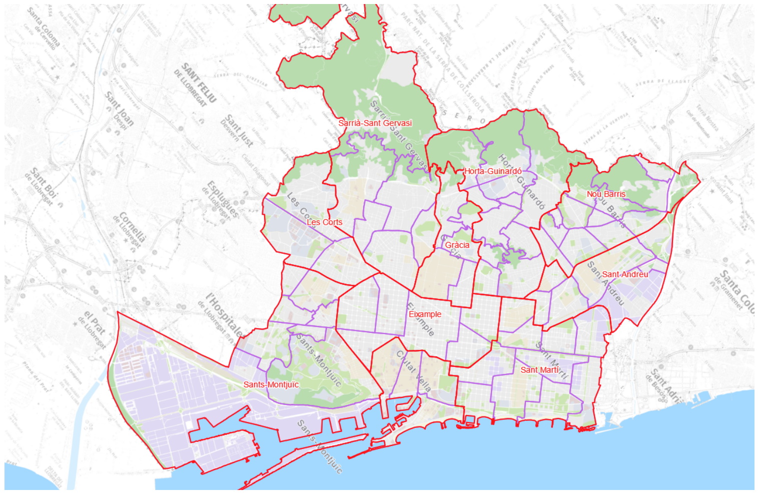

Figure 2: Map of the city of Barcelona, showing the districts and neighborhoods considered in the simulation. *Source*: Interactive map available at Barcelona City Council webpage: https://ajuntament.barcelona.cat/es/

4.1.1 Nodes and Links (Step 0)

The network is modeled as a direct graph, where each arc (i.e., link) represents a road or street, and each node represents a road junction. While a node is described solely by its coordinates (using EPSG:2062 as spatial reference system), the links contain more information. Firstly, the link needs to indicate the two nodes that are connected, indicating the direction. The other characteristics relate to the length of the link, the maximum allowed speed, the maximum number of agents that can leave the link per unit time (called capacity), the number of lanes, the direction of the link, and the allowed means of transport in the link. Regarding the means of transport, an agent can move through a link on foot, by bicycle, by car or by public transport. In order to extract the spatial coordinates of nodes and links, the open-data community-maintained OpenStreetMap tool was employed. Some level of data processing was necessary for properly defining intersections and for classifying the type of road for the purpose of importing the data into MATSim. At the end, more than almost $100,000$ nodes and approximately $260,000$ links were included in the simulation model. The public transport considered, which include buses, underground (metro) and trams, were merged into the model network after obtaining the data from GTFS and employing the MATSim module pt2Matsim to properly format the data.

4.1.2 Facilities (Step 1)

Given Barcelona's network of nodes and arcs representing the city's streets, the first step is to enrich this network by adding “facilities", which represents the places where agents perform activities. The different types of facilities can be divided into six categories: (1) *Home*: Represents the places where the agents live, from where they start their daily activities and where they return at the end of the day; (2) *Work*: Represents the places where the agents work (office buildings, shops, factories...), normally places visited once a day, during a certain number of hours, by those agents that represent adult citizens that are employed; (3) *Education*: Represents the places (schools, universities, etc.) where younger agents, and those whose main occupation is studying, spend a number of hours every day, normally travelling there only once; (4) *Leisure*: Represents shopping centers, sport centers, parks or other facilities where all type of agents go from time to time to spend a variable amount of time for amusement; (5) *Pharma*: Represents hospitals, pharmacies, and other facilities where agents go to take care of their physical or mental health; (6) *Other*: This category was introduced to handle the parking procedure that is explained in Step 5 (Section 4.3.3).

For defining the agent homes, random coordinates were taken within each residential area, according to the polygonal representation of the city of Barcelona. To define the facilities in which work, education, leisure and health-related activities take place, the Open Data Barcelona dataset reporting the economic activities census of the city of Barcelona was considered (see Table 1). At the end of this step, the list of facilities for each neighborhood, classified by the type of activity that can be carried out in it, is available.

4.2 Population Generation

In practical terms, generating the population means setting up a .xml file containing all the information regarding the agents to be considered in the simulation. The characteristics of each agent, such as their home location, their age and gender, have to be provided to the simulator. The process of generating the "synthetic" population to be used in the MATSim simulator is explained in this section, making also reference to the data sources employed. The city of Barcelona had approximately 1.6 million inhabitants in 2021. Simulating this amount of agents is not feasible if reasonable computational resources are to be employed, and for that reason, only the 1 % of the total population was used. Using a percentage of the total population is a common approach since the aim is only to simulate the mobility flow and to study the congestion patterns in the city of Barcelona. In other words, since the model is interested in the global behavior of the agents, rather than in the interaction details in a specific area (e.g., a roundabout or a neighborhood), a subset of the population can be used as proxy of the total mobility configuration.The flow capacity of the links was reduced accordingly to compensate for the reduction in the population simulated.

4.2.1 Sociodemographic Profile (Step 2)

The second step in the pipeline is the creation of groups of agents per neighborhood, with basic sociodemographic attributes. These agents were generated by combining information contained in different datasets from Open Data Barcelona and from the “ESDB" Barcelona sociodemographic survey (see Table 1), specifically, the datasets related to: *(i)* Number and composition of households per neighborhood; *(ii)* Number of people by address; *(iii)* Population per neighborhood (according to gender and age); *(iv)* Level of employment and education. Having defined the agents and the neighborhood in which they live, it is possible to assign a home location (generated in Step 1 - see Section 4.1.2) to each agent, considering the household composition. At the end of this step, a list of agents with sociodemographic characteristics (gender, age, home location, occupation and education), is available.

4.3 Demand Generation

The last piece of information required to run the simulation is the travel demand. This demand represents the sequence of activities (and their associated movements through the network) that each agent wants to perform during the day. Therefore, the location, the start and end-time, and the mean of transport employed must be specified for each activity. Obviously, the agents’ demand should reflect the mobility habits of the inhabitants of the city of Barcelona. It should be noted that the agents’ plans can be subject to variations according to the traffic state of the network, which evolves following the stochastic co-evolutionary algorithm of MATSim, until an equilibrium state is reached. In fact, the activities’ start times are set as the ideal start time according to agents’ wishes, but the duration of the travel between two consecutive activities’ locations may vary over time.

4.3.1 Activity Chain (Step 3)

The third step consists in assigning an activity chain (i.e., a sequence of activities to be performed during the day) to each agent. The activity chains are defined using the “EMEF” Working Day mobility survey (see Table 1) as a baseline, on top of which more elaborated activity chains were added. All activity chains are a sequential combination of the six types of activities defined by their associated facilities, as described in Step 1 (Section 4.1.2). In general, an activity can be repeated more than once throughout the day. Some examples of activity chains are: *"Home-Work-Home"*; *"Home-Education-Home"*; *"Home-Work-Leisure-Home"*; *"Home-Education-Home-Leisure-Home"*; *"Home-Pharma-Home"*. It is important to note that the activity-chain-to-agent assignment considers the sociodemographic characteristics of the agent (e.g., an activity chain containing the activity "Work" will not be assigned to an agent without employment), and also respects the percentage of population carrying out each type of activity as included within the "EMEF" Working Day mobility survey. At the end of this step, the activity-chain has defined for the agents, which represents the skeleton of their daily plan. In the following steps, this initial plan will be enriched with detailed information on each activity and with the movements between the different activities in the plan.

4.3.2 Activity Location (Step 4)

The fourth step determines where the agents will perform each activity of their plans. The activity-to-location assignment was done following the procedure detailed in Algorithm 1, where different sources of statistical data were combined. This intricate data processing was required because no direct source of information is publicly available for linking particular agents to locations. In fact, it would be difficult to obtain this type of information without compromising the privacy of agents, since the residence, workplace, or regularly visited venues will be linked to every agent, together with their typical schedule. The information is extracted mainly from the "ESDB" Sociodemographic survey of Barcelona (see Table 1), taking into account: *(i)* the probability of changing area/district within a trip depending on the origin, *(ii)* the average distances traveled per trip, and *(iii)* the average distances traveled by activity type. At the end of this step, the agents have each activity of their plan associated to a location.

Algorithm 1 Activity-to-location assignment

```
Input: population of agents          ▷ Each agent has its Home location and set of activities
Output: location for every agent activity
for every agent in population do
    for every agent activity do
        origin ← Home(agent)
        P ← probability of changing district(origin)
        d ← distance travelled(activity)
        if random(0,1) > P then              ▷ Decide whether the agent changes district or not
            neighborhoods ← filter districts ≠ origin
        for every remaining neighborhood do
            if distance(origin, neighborhood) > d then
                neighborhoods ← filter out of reach
        for every remaining neighborhood do
            {neighborhood_i, p_i} ← assign probability based on number of facilities dedicated to (activity)
        destination neighborhood ← select random considering probabilities ({neighborhood_i, p_i})
        location ← select random in (destination neighborhood)
end
```

4.3.3 Leg Mode (Step 5)

The fifth step consists in choosing the mode of transport for each trip (or leg) that the agents must take to perform the activities of their plan. The modes considered were introduced in Section 4.1.1. For assigning the mode of transport to each leg, information extracted from "ESDB" Sociodemographic survey of Barcelona (see Table 1) was used. In particular, the statistical data referring to: *(i)* the average traveled distances by mode of transportation, *(ii)* the probability of using a mode as a function of the agent sociodemographic characteristics (mainly age and gender), *(iii)* the probability of using a mode as a function of the type of activity. To continue with the details regarding the choice of transportation mode, it is necessary to introduce the concept of "round-trip". A round-trip is defined as a subset of activities in a plan where the first and last activity is a *"Home"* type activity. For instance, the plan *"Home-Work-Home-Leisure-Home"*, contains two round-trips, namely: *"Home-Work-Home"* and *"Home-Leisure-Home"*. This is an important concept, since it was assumed that the agents can change their transportation mode only at the end of a round-trip. This means, for instance, that if the agents leave home by car, they cannot return home by public transport; but if they return home and leave again, the second trip can be made by public transport. This implies that the mode of transportation will be defined at the round-trip level. The procedure is detailed in Algorithm 2.

Algorithm 2 Mode-to-leg assignment

Input: population of agents ▷ Each agent has a plan (set of activities)
Output: mode for every round-trip

for every agent in population **do**
    for every agent activity plan **do**
        round-trip ← split(plan)
        for every round-trip **do**
            d_{max} ← calculate maximum distance(round-trip)
            modes ← remove modes based on(d_{max}) ▷ E.g., not reasonable to walk 25km daily
            for every remaining mode **do**
                {mode$_i$, p_i} ← assign probability based on(agent, activity)
                mode ← select random considering probabilities ({mode$_i$, p_i})
end

A computational problem arises when the selected mode of transportation is "car" but the facility to be reached is on a link of the network that cannot be traveled by car. In order to overcome this problem, a "parking" procedure is invoked, introducing additional activities of type "Other" which represent parking locations and allow the route to be broken up into several parts. With this logic, the extreme links of the route can be covered on foot while the central links can be covered by car. An example of this is shown in Figure 3. At the end of this step, the agents' plans have a location associated to every activity and also the travel mode between each pair of activities.

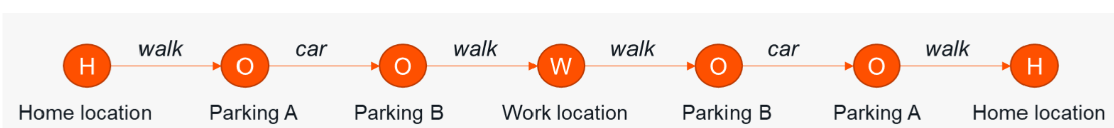

Figure 3: Result of applying the "parking" procedure to a *Home-Work-Home* activity plan, where the Work facility is not reachable by car.

4.3.4 Activity Time (Step 6)

In this final step of the pipeline, the timing is added to each activity in an agent's plan. Technically, for each activity it is possible *(i)* to indicate its duration or *(ii)* to indicate an activity end time. This means that, after the agents arrive at a node and start an activity, they will remain at the node for the timespan indicated (the activity duration), or until the activity end time is reached, whichever comes first. Moreover, activities cannot overlap in time. The "EMEF" Working Day mobility survey (see Table 1) was used to assign an end time for every activity, considering the survey data and the total number of activities to be performed within the plan, i.e., time windows were defined to make sure that each activity's end-time leaves time to perform subsequent activities. For the "parking" activities explained in Step 5 (Section 4.3.3), a fixed activity duration was assumed. At the end of this step, the complete information for every agent's plan is available, namely the activities, with their associated location, mode of transport and timing.

5 RESULTS: CALIBRATION AND VALIDATION

The simulation pipeline described in previous sections contains a number of parameters that could be modified in order to calibrate the simulation. This calibration is an important activity that aims to adjust the simulation in order to reproduce the actual mobility behavior observed in the city as accurately as possible. The first pipeline calibration parameter considered was the possible set of activity chains, as indicated in Step 3 (Section 4.3.1). This means that, in order to better approximate the right number of movements per hour, some activity chains were modified, always maintaining the statistics at a macro-level. The second pipeline parameter tuned was the maximum distance d_{max}, used for defining the mode of transport in Step 5 (Section 4.3.3). This parameter was explicitly introduced in line 6 of Algorithm 2. On the other hand, some of the critical MATSim parameters to be tuned are the parameters related to the flow through the links and their storage capacity. These parameters, were adjusted to compensate for the reduction in the percentage of the simulated population, as detailed in Section 4.2. Also, the number of iterations considered for the co-evolutionary algorithm has an effect in the quality of the results. The results of the calibrated model can be seen in Figure 4.

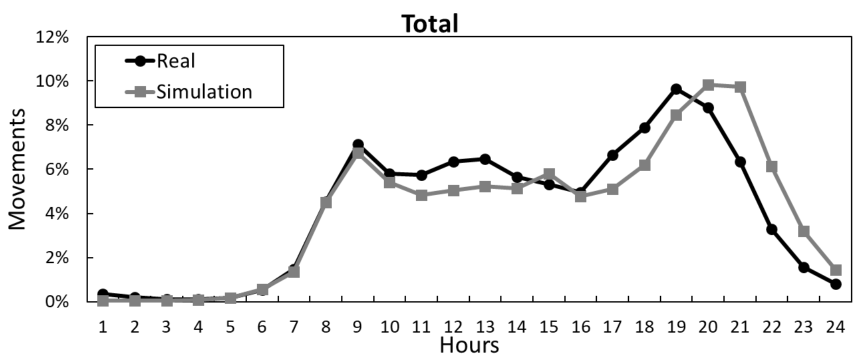

Figure 4: Comparison between the results of the Barcelona MATSim simulation against the real data provided in "EMEF" (see Table 1). The percentage of population that is travelling over the total is shown for each time bucket. Time bucket "1" covers from 00:00 until 00:59, bucket "2" covers from 01:00 until 01:59, and so on.

In terms of the mobility patterns, it can be observed that around 07:00 the transportation activity begins in the city of Barcelona, with a first traffic peak around 09:00, coinciding with the start of the activities in schools and work places. Throughout the central hours of the day, people move about in a somehow constant manner, probably reflecting the dynamism of a metropolis of the size of Barcelona, where different activities contribute to maintain a stable condition. Although the simulation captures this behavior, it underestimates the amount of movements in the city during this period. Around 16:00 there is a change in trend and the amount of journeys increases steadily up until around 20:00, when it starts to decrease at approximately the same rate. Once again, the simulation captures perfectly the dynamic of the mobility demand, but a slight delay can be observed in the second peak, which occurs later. In terms of intensity of the plateau area and the peaks and valleys, it can be seen that the simulation also matches the actual intensity levels.

Since one of the contributions of this study is the multi-modal capability of the MATSim implementation, the same comparison for population movements can be made segregating by mode of transport, as presented

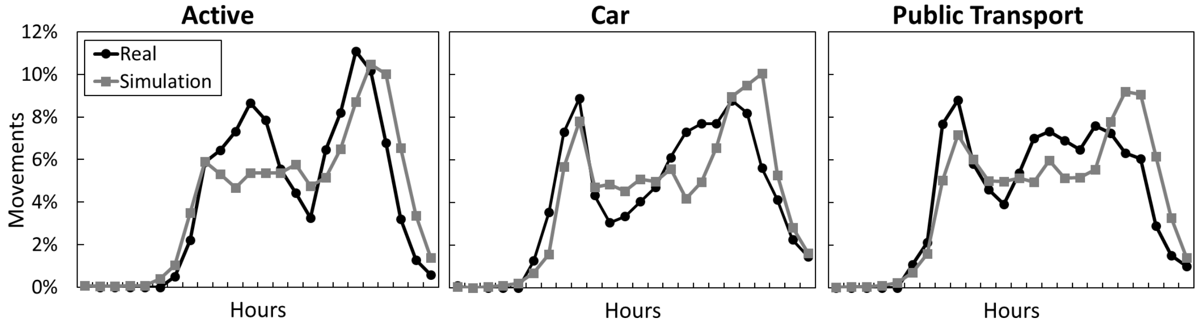

Figure 5: Comparison between the results of the Barcelona simulation against the real data provided in "EMEF" (see Table 1). The percentage of population that is travelling over the total is shown for each time bucket, and split by mode of transport. The time buckets are the same as in Figure 4.

in Figure 5. Since these graphs are the decomposition of Figure 4, the conclusions that can be extracted are very similar. If we analyze them one by one, it is interesting to observe that the simulation underestimates the use of active mobility (walking and biking) in the morning period. This underestimation is compensated with the use of cars and public transport, and for all of them it is clear the delay in the last part of the graph. That fact, namely that in the simulation there is a concentration of movements at the end of the day, could be explained by the way the activities are slotted in the pipeline. As explained in Step 6 of the pipeline (Section 4.3.4), activities can be of fixed duration or have a fixed start/end-time. Even considering well-defined times for the activities does not ensure that these take place as scheduled. In fact, the duration of travel between two consecutive activities is affected by at least two factors: *(i)* the traffic conditions on the network, that vary as a result of the co-evolutionary algorithm of MATSim; and *(ii)* the adjustment to the plan introduced by the parking procedure, which was necessary to ensure consistency in the data. Delays accumulate and this is reflected as a consequence on the latest activities of agents' plans, leading to more trips to be performed in the last hours of the day. This fact could be improved in a subsequent version of the model in order to distribute the activities end-times in such a way that reflects with more realism the actual mobility demand dynamic. Finally, in order to quantitatively evaluate the overall adjustment of the MATSim model to the real data, in Table 2, a summary of the gap between the obtained results and the real mobility data is provided for groups of time buckets and per mode of transport. Additionally, the Mean Average Percentage Error (MAPE) is calculated for each mode and also in total.

Table 2: Quantitative numerical results of the simulation model.

Mode of transport	Gap in time bucket group				MAPE
	1-6	7-12	13-18	19-24	
Active %	1	5	3	7	37
Car %	0	1	6	8	24
Public Transport %	0	3	9	12	27
Total %	0	3	5	8	31

As explained above, the gap increases towards the end of the day bucket groups, with values between 7 % to 12 %. The MAPE was used to quantify the difference between the actual and simulated values, obtaining figures close to 30 %, which could be interpreted as a model accuracy level of roughly around 70 %. This accuracy varies depending on the transport mode, being greater for the car than for the public transport and active mobility, probably reflecting the fact that the software logic was conceived initially for road traffic simulation. The presented model could greatly benefit from further efforts in order to increase its validity, specially when looking at street level traffic comparison. The match between simulation and reality at this maximum granularity level were difficult to obtain given the current state of the model.

6 CONCLUSIONS AND FUTURE WORK

In this work, a detailed step-by-step pipeline for generating the demand for mobility for the city of Barcelona has been defined. The pipeline was entirely developed considering open-access data (mostly from the Barcelona City Council open-access initiatives) and open-source tools (MATSim simulator). The generated demand was represented by a population of agents with sociodemographic characteristics and a plan of activities to be carried out during the day. This population of agents was feed into MATSim, which simulates the interaction between them, as they to move throughout the network of the city, using multiple modes of transportation, i.e. not only cars, but also bikes and public transport. From this agent-based micro-level interaction emerges the mobility patterns that characterize the city of Barcelona. The simulation model has been calibrated in order to reproduce the real behavior as faithfully as possible, with the final results showing a remarkable correspondence between the model and the reality.

A calibrated simulation model that replicates the dynamics of citizen mobility at the level of the city of Barcelona is an extremely useful tool on its own for researchers and decision-makers, and opens many future lines of research and application. One direct use of the simulator could consist in studying the impact of different public policies on the traffic profile of the city, such as the changes in people mobility as a consequence of restricting certain types of vehicles in some areas, changing the availability or the routes of some public transport lines, or moving the car-sharing service stations. Furthermore, health and sustainability aspects can be considered and the effects of different policies on CO2 emissions reductions evaluated. One very interesting line of research is considering the use of this tool in combination with other developed scientific fields, like optimization and operations research, where the simulator could be an essential part in advanced simulation-optimization algorithms. As concluding remarks, it is worth mentioning that the details provided for the demand pipeline construction, and the use of publicly available tools and dataset, should encourage other researchers to try and replicate this modelling methodology in other cities, where similar data is available. Also, the study's result has demonstrated that a synergetic alliance between public administrations and the research community can help towards creating powerful tools that empower citizens and companies towards smart city with the final goal of increase efficiency and reduce carbon emissions.

ACKNOWLEDGMENTS

This work has been partially funded by Spindox and the Industrial Doctorate Program of the Catalan Government (2020 DI 116), the Spanish Ministry of Science (PID2019-111100RB-C21/ AEI/ 10.13039/ 501100011033), and the Barcelona City Council – "la Caixa" Foundation under the framework of the Barcelona Science Plan 2020-2023 (grant 21S09355-001).

REFERENCES

Argota Sánchez-Vaquerizo, J. 2022. "Getting Real: The Challenge of Building and Validating a Large-Scale Digital Twin of Barcelona's Traffic with Empirical Data". *ISPRS International Journal of Geo-Information* 11(1):24.

Bassolas, A., J. J. Ramasco, R. Herranz, and O. G. Cantú-Ros. 2019. "Mobile Phone Records to Feed Activity-Based Travel Demand Models: MATSim for Studying a Cordon Toll Policy in Barcelona". *Transportation Research Part A: Policy and Practice* 121:56–74.

Bowman, C., J. A. Miller, and Y. Wang. 2022. "Microscopic Vehicular Traffic Simulation: Comparison of Calibration Techniques". In *Proceedings of the 2022 Winter Simulation Conference*, edited by B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. G. Corlu, L. H. Lee, E. P. Chew, T. Roeder, and P. Lendermann, 2234–2245. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ciari, F., N. Schuessler, and K. W. Axhausen. 2013. "Estimation of Carsharing Demand Using an Activity-Based Microsimulation Approach: Model Discussion and Some Results". *International Journal of Sustainable Transportation* 7(1):70–84.

Gräbe, R. J., and J. W. Joubert. 2022. "Are We Getting Vehicle Emissions Estimation Right?". *Transportation Research Part D: Transport and Environment* 112:103477.

Horni, A., K. Nagel, and K. Axhausen. 2016. *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press.

Hörl, S., and M. Balac. 2021. "Synthetic Population and Travel Demand for Paris and Île-De-France Based on Open and Publicly Available Data". *Transportation Research Part C: Emerging Technologies* 130:103291.

Kuo, Y.-H., J. M. Y. Leung, and Y. Yan. 2023. "Public Transport for Smart Cities: Recent Innovations and Future Challenges". *European Journal of Operational Research* 306(3):1001–1026.

Ma, Z., and M. Fukuda. 2015. "A Multi-Agent Spatial Simulation Library for Parallelizing Transport Simulations". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, V. W. Chan, I.-C. Moon, T. M. Roeder, C. Macal, and M. D. Rossetti, 115–126. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Mahmud, K., and G. E. Town. 2016. "A Review of Computer Tools for Modeling Electric Vehicle Energy Requirements and their Impact on Power Distribution Networks". *Applied Energy* 172:337–359.

Miller, J. A., H. Peng, and C. N. Bowman. 2017. "Advanced Tutorial on Microscopic Discrete-Event Traffic Simulation". In *Proceedings of the 2017 Winter Simulation Conference*, edited by V. W. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. H. Page, 705–719. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Moeckel, R., N. Kuehnel, C. Llorca, A. T. Moreno, and H. Rayaprolu. 2020. "Agent-Based Simulation to Improve Policy Sensitivity of Trip-Based Models". *Journal of Advanced Transportation* 2020:1902162.

Mohebbi, S., and P. S. Murali. 2022. "Spatial Agent-Based Simulation of Connected and Autonomous Vehicles to Assess Impacts on Traffic Conditions". In *Proceedings of the 2022 Winter Simulation Conference*, edited by B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. G. Corlu, L. H. Lee, E. P. Chew, T. Roeder, and P. Lendermann, 2487–2498. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Montero, L., M. P. Linares, O. Serch, and J. Casanovas-Garcia. 2017. "A Visualization Tool Based on Traffic Simulation for the Analysis and Evaluation of Smart City Policies, Innovative Vehicles and Mobility Concepts". In *Proceedings of the 2017 Winter Simulation Conference*, edited by V. W. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. H. Page, 3196–3207. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Neves, F. T., M. de Castro Neto, and M. Aparicio. 2020. "The Impacts of Open Data Initiatives on Smart Cities: A Framework for Evaluation and Monitoring". *Cities* 106:102860.

Poletti, F., P. M. Bösch, F. Ciari, and K. W. Axhausen. 2017. "Public Transit Route Mapping for Large-Scale Multimodal Networks". *ISPRS International Journal of Geo-Information* 6(9):268.

Weyl, J., U. A. Lenfers, T. Clemen, D. Glake, F. Panse, and N. Ritter. 2019. "Large-Scale Traffic Simulation for Smart City Planning with MARS". In *Proceedings of the 2019 Summer Simulation Conference*, July 22nd-24th, Berlin, Germany, 1–12. San Diego, CA, USA: Society for Computer Simulation International.

AUTHOR BIOGRAPHIES

JONAS F. LEON is a predoctoral researcher at the Computer Science Dept. at the Universitat Oberta de Catalunya. He holds an M.Sc. in Industrial Engineering. With almost a decade of experience in the industry, at present, he works as Analyst for optimization and artificial intelligence projects at Spindox (www.spindox.it). His main research interests include advanced simulation-optimization methods, such as simheuristics, and complex systems analysis. His email address is jofule@uoc.edu.

FRANCESCA GIANCOLA is currently attending the ABRO Ph.D. Program in Operations Research at the Department of Computer, Automatic and Management Engineering "Antonio Ruberti" (DIAG) at Sapienza University of Rome, in collaboration with the Institute of Systems Analysis and Informatics (IASI) of the National Research Council (CNR). She has a Master's degree in Management Engineer and she works as Modeling & Simulation Research Engineer at Spindox. Her main research interests include Mixed Integer Programming, Optimization under Uncertainty and Simulation. Her email address is francesca.giancola@ahead-research.com.

ANDREA BOCCOLUCCI is a Simulation and Modeling Research Engineer at Spindox. At Sapienza University of Rome, he obtained a Master's degree in Management Engineering with a thesis on simulation to support logistical decisions. His main research interests include all the new techniques and technologies to be applied in industrial sector. His email address is andrea.boccolucci@ahead-research.com.

MATTIA NERONI is a postdoctoral researcher currently working as Senior Research Engineer at Spindox. He has a Master's degree in Management Engineering and a Ph.D. in Industrial Systems Engineering. He worked as Adjunct Professor at University of Parma (Italy) and as postdoctoral researcher in several European universities. He is also co-founder of a startup called Kheperer (www.kheperer.it). He is currently co-author of 25 scientific publications on international Scopus indexed journals. His main research interests include optimization, simulation, machine learning, logistics, and computer science (email: mattia.neroni@ahead-research.com, website: https://mattianeroni.github.io).