HAL open science

Pre-calibration of a Discrete Choice Model and Evaluation of Cycling Mobility for Île-de-France

Guoxi Feng, Maxime Jean, Alexandre Chasse, Sebastian Hörl

To cite this version:

Guoxi Feng, Maxime Jean, Alexandre Chasse, Sebastian Hörl. Pre-calibration of a Discrete Choice Model and Evaluation of Cycling Mobility for Île-de-France. Procedia Computer Science, 2021, 184, pp.172-177. 10.1016/j.procs.2021.04.003 . hal-03597432

HAL Id: hal-03597432

https://ifp.hal.science/hal-03597432v1

Submitted on 4 Mar 2022

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 Attribution - NonCommercial - NoDerivatives 4.0 International License

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 184 (2021) 172–177

Procedia Computer Science

www.elsevier.com/locate/procedia

The 12th International Conference on Ambient Systems, Networks and Technologies (ANT)
March 23-26, 2021, Warsaw, Poland

Pre-calibration of a Discrete Choice Model and Evaluation of Cycling Mobility for Île-de-France

Guoxi Feng[a,*], Maxime Jean[a], Alexandre Chasse[a], Sebastian Hörl[b]

[a]*Digital Science and Technology Division, IFP Energies Nouvelles, 1 et 4 Avenue du Bois-préau – 92852 Rueil-Mailmaison, France*
[b]*Institut de Recherche Technologique SystemX, 8Avenue de la Vauve – 91120 Palaiseau, France*

Abstract

An open and reproducible MATSim scenario of Île-de-France has recently been made available. In this paper, first, we aim to customize the discrete choice model for Île-de-France scenario and estimate its initial parameterization via the routing API OpenTripPlanner. The goal of this step is to reproduce the travel patterns observed from the regional Household Travel Survey (HTS) in MATSim without running any time-consuming simulation. Since the modeled cycling behavior could impact strongly the mode choice decision on bike, we evaluate the cycling routing with disaggregated GPS traces obtained dynamically from the mobile application Geovelo. The objective of this pre-calibration and evaluation is to provide a solid base for our future work which aims to calibrate the cycling flow per link in MATSim for Paris with the normalized flow from recent crowdsensing data.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the Conference Program Chairs.

Keywords: Discrete Choice Model, pre-calibration, MATSim, GPS trace, cycling flow;

1. Introduction

An open and reproducible MATSim scenario of Île-de-France [1] has recently been made available. The entire scenario can be generated and simulated via open data and open software. Based on this work, it is possible to customize the baseline scenario according to specific needs, for example, by extending the model to other French cities, modeling the impact of home-office on travel demand, analyzing the travel pattern of bike mobility in the pandemic context, etc. In particular, the mobility patterns evolve dynamically under short-term constraints (e.g. the Covid-19 crisis) and long-term constraints such as infrastructure and economic conditions. The baseline scenario generated from the survey data may not be able to capture dynamic short-term behavior or some specific disaggregated behavioral patterns. Therefore, there is a great need for discussing how to evaluate and calibrate the baseline scenario with dynamic information such as crowdsensing data, measures from count station, and recent mobility surveys.

* Corresponding author.
E-mail address: guoxi.feng.externe@ifp.fr

1877-0509 © 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2021.04.003

In the present study, we aim, first, to customize the Discrete Choice Model (DCM) [2]-[4] used in the base Île-de-France scenario and provide its initial parameterization via the routing API OpenTripPlanner. The goal of this step is to reproduce the travel pattern observed by the regional Household Travel Survey (HTS) in MATSim without running any time-consuming simulation. Since the modeled cycling behavior could impact strongly the mode choice decision on bike, we evaluate, then, the cycling routing with GPS traces obtained dynamically from the mobile application Geovelo [5]. The objective of this pre-calibration and evaluation is to provide a solid base for our future work which aims to calibrate the cycling flow per link in MATSim for Paris with the normalized flow from crowdsensing data e.g. GPS traces of Geovelo and the measurements from eco-counters [6].

This paper is structured as follows: Section II provides a simple introduction on MATSim and the packages used in this study. Section III presents the pre-calibration of DCM and its integration in the simulation. Section IV evaluates cycling behavior. The last section concludes the paper and discusses the future work in more detail.

2. MATSim

A MATSim simulation requires a *scenario* which consists mainly of a *population*, a *network* and a *public transport supply*. The population includes daily plans for many synthetic persons including activities and connecting trips between these activities. To satisfy the travel demand of the population, the network provides the infrastructure. Once a scenario is prepared, a simulation iteration consists of the following three steps: Mobsim, Utility estimation and Replanning.

The Mobsim executes the selected trips and activities of the population in parallel. The traffic flow dynamics are simulated by a queue model. When too much vehicles are on the same road, traffic slows down and travel time increases. On the other hand, less cars on the network could decrease the traffic load and then reduce travel times in general. This way, the decision of an agent could affect the travel time of others and the quality of their trips.

To qualify the executed trips, a utility is assigned in the Utility estimation stage. When an agent is selected for replanning, the utilities of all available transport modes are estimated for each trip. The estimated utilities serve later for a new decision making in the Replanning stage. The structure of these utility functions is free to be assigned depending on the travel behavior to be reproduced.

In the Replanning stage, a decision of the selected agents is generated via a discrete choice model with the estimated utilities of each available transport mode. For a given trip, all available transport modes are in competition with each other. The agent has a highest chance to select the mode with the highest utility and execute the routed itinerary of the selected mode in simulation. For example, heavy traffic congestion increases the travel time and decreases the utility of using car. Such a situation could encourage an agent to use other alternatives instead of taking the car. At the end of Replanning, the new decision will be simulated in the next iteration. The same procedure is repeated until the travel behavior of the agents stabilizes. Due to the integration of a Discrete Choice Model, the MATSim workflow discussed above does not use *scoring* and *evolutionary decision-making* for Replanning, which is the default approach in MATSim. For details on the original workflow and the one around Discrete Choice Models, we suggest readers to refer to the MATSim book [7] and the work in [2], respectively.

In the coming subsections, we will present the setting up of the Île-de-France simulation. To this end, the Île-de-France scenario generation and the bicycle module integration are briefly introduced.

2.1. Île-de-France scenario

To generate a synthetic population from raw survey data, we use the pipeline presented in [1]. The synthesized agents are sampled from the population census 2017 published by *INSEE* (Insistut National de la Statistique et des Etudes Economiques). According to their socio-demographic attributes such as sex, age and profession, each synthesized agent is associated statistically to an activity chain from the regional Household Travel Survey 2010 (Enquête Globable Transport). An activity chain contains daily activities with duration, connected trips between activities and their travel times. To assign the location of each activity of each synthesized agent, we use the facility census 2017 (Base permanente des équipements) and commuting data from the census 2017. The socio-demographic attributes, activity chains and location assignments are computed via statistical matching such that the synthesized population contains the statistical attributes of the surveys. Finally, 5% of the Île-de-France population are sampled for creating the scenario.

The multimodal network is generated from OpenStreetMap (OSM). However, OSM does not provide any information about elevation which is crucial element for simulating cycling behaviors. For mapping the elevation on the network, a network link is connected to two nodes with elevation data and contains information like length, speed limit, flow capacity, road type, road quality, street name and type of cycleway. To complete the public transit stops and the corresponding schedule on the network, data in General Transit Feed Specification (GTFS) format is used.

2.2. Cyclist model

Cyclists are "teleported" in most MATSim simulations. During teleportation, cyclists are moved directly on a straight line between origin and destination with a constant velocity. Therefore, cyclists do not interact at all with the network. To overcome this drawback, we apply the cyclist module [8]. This MATSim plugin enables synthetized cyclists to move on the network model. The module contains two main components: a speed calculator and a cycling itinerary calculator.

The speed calculator computes the moving speed of cyclists by link in function of road quality, road slopes and cycling infrastructure. This road information was already modeled during the network generation. Basically, the cycling speed in a well-built cycleway should be higher than the one in a cobblestone road without cycling infrastructure. For the same road, cyclists should go faster in the downhill direction than in the uphill direction. The itinerary calculator supplies an itinerary depending on the variables such as travel distance, travel time, cycling infrastructure, slopes and road quality.

3. Discrete Choice Model Pre-calibration

The motivation of integrating DCM into MATSim has been explained in [2]. One of the main advantages brought by DCM is that more human prior knowledge is taken into account during the mode choice decision. Such prior knowledge could make the agent decision more realistic. Compared with the scoring-based replanning, the application of DCM could reduce the biases linked to the plan exploration in the beginning of the simulation. In this section, we impose first the structure of the utility function and then aim to provide a pre-calibration for its parameter vector via Open Trip Planner.

3.1. Open Trip Planner

Open Trip Planner (OTP) is an open-source API which finds itineraries dedicated to transit, pedestrian, bicycle, and car. The travel time and travel distance of each transport mode is computed via OTP with the origin-destination from Household Travel Survey data. For each trip, the regional household travel survey indicates the departure location and arrival location inside of a 100m × 100m tile while the precise coordinates remain unknown. To assign a precise location to the departure point and the arrival point, the facility census is used. Since the location of facilities are available, these facilities are associated to the tile in which they are located. The medoid location of all the facilities inside of a tile is computed. These median locations are assigned as the departure or arrival point of the corresponding tile. If a tile does not contain any facility, the centroid is considered as its reference point.

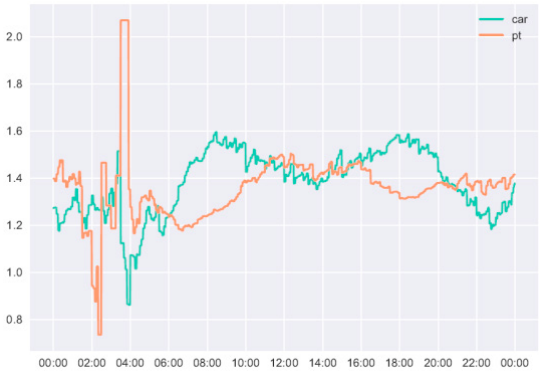

Figure 1.Ratio between the travel time estimated by OTP and the one observed in EGT in function of departure time

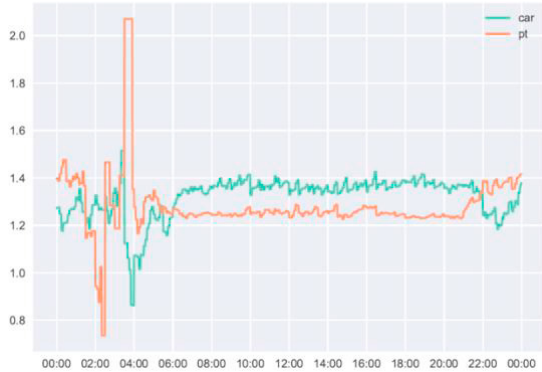

Figure 2. Ratio after the correction

Among 126 366 trips, 107 679 trips have been associated to a precise departure/arrival location. To estimate the travel time, departure time and travel distance, routing requests with the precise departure/arrival location and an available transport mode are send to OTP. The available transport modes car, bike, public transport (pt) and walk are considered. For 3 663 trips, OTP is not able to provide all the itineraries of all available modes. These trips are ignored for the next step. The ratio between the travel time estimated by OTP and the one observed in Enquête Globable Transport (EGT) is shown in Figure 1 and Figure 2. We notice that the ratio is high for car trips which have a departure time during rush hours which is explained by the fact that traffic congestion is not considered by OTP. Since the trend of the ratio is quite obvious to identify between 6:00 and 22:00, the under-

estimated travel time is adjusted only for this period. After adjustment, the impact of underestimation is reduced for the considered period. Although coarse, we expect that this travel time estimation enables an initial guess on the parameters for DCM in MATSim.

3.2. Parameter Estimation

The structure of the utility function U_i for a transport mode i is as follows:

$$U_i = C_i + \beta_{Traveltime,i} \cdot \mathcal{X}_{Traveltime,i} + \beta_{Velocity} \cdot \mathcal{X}_{Velocity,i} + \beta_{Traveldistance,i} \cdot \mathcal{X}_{Traveldistance,i} + \beta_{CrossingParis,i} \cdot \mathcal{X}_{CrossingParis,i} + \beta_{InsideParis,i} \cdot \mathcal{X}_{InsideParis,i} + \beta_{Drivinglicense,i} \cdot \mathcal{Y}_{Drivinglicense,i} + \beta_{Navigo,i} \cdot \mathcal{Y}_{Navigo,i} + \beta_{Dispo_car,i} \cdot \mathcal{Y}_{Dispo_car,i} + \beta_{Dispo_bike,i} \cdot \mathcal{Y}_{Dispo_bike,i}$$

$$i \in \{car, bike, public\ transport, walk\}$$

where C is mode-specific constant utility, $\mathcal{X}$ are travel variables e.g. $\mathcal{X}_{Traveltime}$ travel time, $\mathcal{X}_{Traveldistance}$ distance and $\mathcal{X}_{Veolcity}$ speed, $\mathcal{X}_{CrossingParis}$ and $\mathcal{X}_{InsideParis}$ indicate respectively if a trip is cross-urban or only inside of Paris. $\mathcal{Y}$ are indicative attributes of agents e.g. $\mathcal{Y}_{Drivinglicense}$ driving license, $\mathcal{Y}_{Navigo}$ public transport subscription, $\mathcal{Y}_{Dispo_car}$ car availability and $\mathcal{Y}_{Dispo_bike}$ bike availability. The objective of the pre-calibration is to search a parameter vector β which enables the DCM to predict correctly the transport mode for a given trip. The travel variables are estimated from OTP. The quality of the prediction is evaluated by the data of the Household Travel Survey considered as the reference in this study. Note that the model formulation (without cost and with highly correlated variables such as travel time, distance, and speed) is not suited for an econometric interpretation, but, rather, aims at providing a good fit with reference data for a first version of the model. A transport mode is selected by choosing the highest utility for a given trip. The predicted total mode shares are close to the mode shares observed in EGT 2010 as depicted in Figure 3.

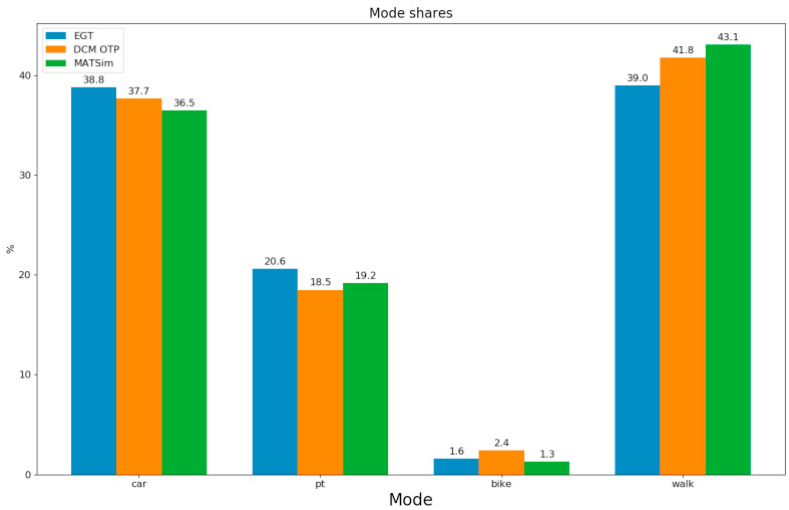

Figure 3. Mode shares of EGT (blue), mode shares obtained by the pre-calibration (orange), and mode shares MATSim (green)

To evaluate the performance of the pre-calibration in the simulation, the obtained DCM is integrated into MATSim. The mode shares in function of Euclidian distance and the mode shares in function of travel time in MATSim are close to the reference as shown in Figure 3, Figure 4, and Figure 5. Even though the pre-calibrated DCM is identified with the travel time coarsely estimated from OTP, it performs unexpectedly well in MATSim. For the trips where several modes provide a similar utility, the coarse estimation of the pre-calibrated DCM may be not precise enough. To increase the precision of DCM and reduce the error between the simulated results and the reference, a fine-tuning operation is required. However, the time-consuming MATSim simulations make such fine-tuning operations difficult. A further discussion on searching an appropriate calibration framework is needed.

4. Cyclist Model Evaluation

It is worth mentioning that the routing parameters of the cyclist model have been tuned manually for a realistic

trade-off between travel time, distance, road quality, road slopes and cycling infrastructure. To evaluate the cycling behavior produced by this model, the daily cycling flow simulated by MATSim is compared with the mean daily flow of November 2019 measured by the eco-counters of Paris [6]. Knowing that the daily cycling travel demand of November 2019 could be different from the one modeled with EGT 2010, the objective of this comparison is to verify if the simulated and observed flows have a similar trend. As shown in Figure 6, red indicates larger flows. The spatial distribution of the simulated flow (links) is quite different from the measured ones (dots). As an example in Figure 7, the trend of the simulated flow at Rue de Rivoli during the day does not follow the measured one, especially in the evening. It is, therefore, valid to ask whether agents are routed to the correct links. The Household Travel Survey 2010 provides less than 2000 cycling trips with origin-destination at departmental level. Furthermore, the few records could have been registered already ten years ago. Yet, the fragmented measurements from the counters also do not provide a flow of the whole cycling network. Therefore, these datasets are not sufficient to render a detailed picture of cycling behavior in Paris.

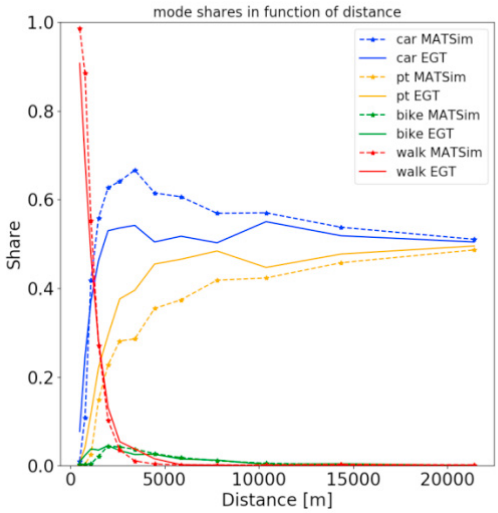

Figure 4. Mode shares in function of Euclidean distance

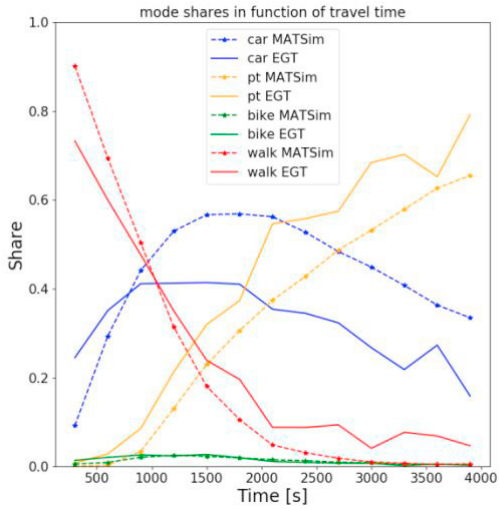

Figure 5. Mode shares in function of travel time

Figure 6. Simulated daily flow represented by lines and daily flow measured from eco-counters represented by points

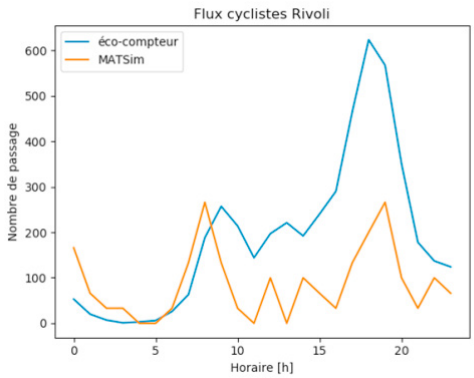

Figure 7. Flows of Rue de Rivoli during the day

The available data of Geovelo is now newer and more disaggregated than the one of EGT. We notice that the trip-mean cycling speeds follow Gaussian distributions centered at 4km/h, 8km/h and 10km/h for the data of EGT, the simulation, and Geovelo, respectively. This difference between the simulation and the observations of Geovelo could be due to the representativeness of the crowdsensing data, the modeled cycling network and the routing.

To evaluate the cycling routing, the origin-destinations are extracted from the 73 737 GPS traces which were validated in a pre-processing step. The invalidated traces such as car trips or inconsistent records at the end of trips are filtered based on the criteria of moving speed, detour ratio, and others. The extracted origin-destinations are simulated. The idea is to see if MATSim is able to reproduce the pattern observed by Geovelo on condition that the travel demand is the same for both cases. The mean traveled distance is 4.9 km in MATSim and 5 km observed by Geovelo. The mean traveled time is 16 minutes and 17 minutes respectively for MATSim and Geovelo.

Figure 8. The simulated cycling flow (left) and the measured cycling flow (right)

As shown on the right side of Figure 8, the main flow marked by the red color is located at the well-known major cycling routes such as Rue de Rivoli and Boulevard de Sebastopol. The simulated flow is less concentrated and dispersed into different routes. The reason could be that the infrastructure factors of each link on the network and the parameters of routing may not be attributed correctly. This could lead to the differences in simulated flow and observed flow. To reproduce the behavior, the parameters of routing and the cycling network should be adjusted or more factors such as security, noise and pollutant could be taken into account.

5. Conclusion and Future Work

In this paper, it was illustrated that the travel patterns observed in the Household Transport Survey are coarsely reproduced by the pre-calibrated DCM in MATSim. The evaluation of the cyclist model with GPS data showed that the cycling routing in our MATSim model for Paris needs to be improved.

The objective of this pre-calibration and evaluation is to provide a solid base for our future work which aims to calibrate the cycling flow per link with the normalized flow from crowdsensing data e.g. GPS traces of Geovelo and the measures from eco-counters. For this future work, the approach is two-folded. The first step aims to provide a realistic itinerary to each cycling trip. To this end, the behavioral cyclist model will be adjusted with indicators generated dynamically from Geovelo GPS data such as road quality per link and moving speed per link. Then, a scenario created from the cycling demand observed by GPS traces of Geovelo will serve to calibrate the cycling routing parameters. With such a scenario, we know which itinerary was used by a person in the reality and how this person should behave in MATSim.

Once cycling trips are correctly routed, the second step is to estimate what cycling travel demand is able to reproduce the normalized cycling flow. However, the cycling travel demand observed in the Household Transport Survey 2010 is obsolete and cannot be considered as reference. In this step, we consider that the cycling travel demand is marginal compared with other transport modes in IdF. Then, the estimation of the cycling travel demand can be done by finely tuning the DCM from the initial parameterization obtained in this study. Such calibration with the crowdsensing data should allow to track dynamically cycling mobility and better understand its evolution.

References

[1] Hörl, S., & Balac, M. (2020). Open data travel demand synthesis for agent-based transport simulation: A case study of Paris and Île-de-France.

[2] Hörl, S., Balac, M., & Axhausen, K. W. (2018). A first look at bridging discrete choice modeling and agent-based microsimulation in MATSim. Procedia computer science, 130, 900-907.

[3] Hörl, S., Balać, M., & Axhausen, K. W. (2019). Pairing discrete mode choice models and agent-based transport simulation with MATSim. In 2019 TRB Annual Meeting Online (pp. 19-02409). Transportation Research Board.

[4] Ilahi, A., Balac, M., Li, A., & Axhausen, K. W. (2019). The first agent-based model of greater Jakarta integrated with a mode-choice model. Procedia Computer Science, 151, 272-278.

[5] https://www.geovelo.fr/.

[6] https://www.eco-compteur.com/.

[7] W Axhausen, K., Horni, A., & Nagel, K. (2016). The multi-agent transport simulation MATSim (p. 618). Ubiquity Press.

[8] Ziemke, D., Metzler, S., & Nagel, K. (2019). Bicycle traffic and its interaction with motorized traffic in an agent-based transport simulation framework. Future Generation Computer Systems, 97, 30-40.