



**HAL**  
open science

# Benchmark mobility problems using real-world data: the example of bus stops spacing problem for the city of Calais

Valentin Vendi, Cyril Fonlupt, Sébastien Verel

## ► To cite this version:

Valentin Vendi, Cyril Fonlupt, Sébastien Verel. Benchmark mobility problems using real-world data: the example of bus stops spacing problem for the city of Calais. The 7th International Conference on Logistics Operations Management, GOL'24, May 2024, Marrakesh, Morocco. hal-04594899

**HAL Id: hal-04594899**

**<https://ulco.hal.science/hal-04594899>**

Submitted on 30 May 2024

**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 4.0 International License

# Benchmark mobility problems using real-world data: the example of bus stops spacing problem for the city of Calais

Valentin Vendi<sup>1</sup>, Cyril Fonlupt<sup>1</sup> and Sébastien Verel<sup>1</sup>

Univ. Littoral Côte d'Opale, LISIC  
Calais, France, F-62100  
valentin.vendi@univ-littoral.fr,

**Abstract.** This paper proposes a benchmark for the bus stops spacing problem based on real-data and agent based simulation. An instance of the benchmark is created using a standard pipeline in order to have a reproducible scenario that can be adapted to other cities. The pipeline is based on the MATSim simulator and related tools. As an example, we show the details of the pipeline for the city of Calais from open source raw data to the computation of an objective function to optimize. The potential of such approach in terms of fair optimization algorithms comparison is discussed, and the code used to design our benchmark instance is also shared with the goal to later generalize it.

**Keywords:** benchmark, bus stops spacing, simulation, real-world problem, combinatorial optimization

## 1 Introduction

Mobility problems in a city are one of the major target issue to face the climate change. The long term goal of our work is to participate in the development of the cities of tomorrow by focusing on the implementation of innovative mobility solutions for different mobility problems such as car sharing system [1, 2], public transportation management [3], multi-modality transportation system [4], etc. A number of new systems or new optimization algorithms are proposed to solve these mobility problems in a city. However, due to the complexity of designing a relevant mobility scenario for a city, the difficulty of having similar real-world data across different cities in the world and the high level of complexity of agent based simulation themselves (see SUMO [5], MATSim [6], etc.), it is difficult to compare systems and optimization algorithms in a fair way. Moreover, the reproducibility of a mobility scenario is one of the main issues for logistic and related mobility problems [4].

This paper is part of the process of proposing reproducible and standard benchmark scenarios for mobility problems. The work is based on previous effort in that direction. Indeed, from the agent based simulator MATSim [6] and the pipeline Eqasim [7] which create mobility scenario from available real-world data,

we propose an additional tool for this original pipeline to create a scenario of the bus stops spacing problem [8, 9]. As an example, we propose a case study of the bus stops spacing problem in the city of Calais following the work of Leprêtre et al. [10]. Whereas the previous work is based on fully artificial scenarios of mobility, we propose a reproducible pipeline to design a mobility scenario based on real-world data. As a drawback, the computation time of the simulation is increased.

This paper will cover the following topics. The section 2 introduces the components of the original pipeline from the previous works and presents the bus stops spacing problem. Section 3 describes our current pipeline of simulation using MATSim and Eqasim. Section 4 shows the potential of the scenario from preliminary experiments, and shares on the web a code and data for our instance of the bus stops spacing problem in the city of Calais. Finally, last section concludes the paper and presents our future work directions.

## 2 Related Work

### 2.1 Real-World Based Simulation

MATSim [6] is an open-source framework used to create agent-based transport simulations. With its agent-based approach, MATSim tends to reproduce how individuals make travel decisions and interact within a transportation network based on their habits and constraints. MATSim is composed of several modules which can be combined. One important feature of MATSim is the re-planning module which allow to compute the travelling path of each agent. Each agent has a score according to the modality of mobility (car, public transportation, bike, etc.) and the duration of the travel. Then, the re-planning module maximizes iteratively the score of each agent using an evolutionary algorithm until a stable travel plan is found. In that way, MATSim is able to simulate the traffic flow according to the transport scenario.

Eqasim [7] is a pipeline used to obtain a synthetic population from open-source public data. This synthetic population can then be used in a MATSim simulation to depict a real-world mobility scenario. The main goal of the Eqasim pipeline is to provide the reproducibility of an agent-based simulation using a standard way to design synthetic travel demand and transport supply from accessible data. Eqasim uses multiple sources to generate a population, typically including socio-demographic data, population counts, travel surveys, public transportation information, geographic data, and more. The workflow of Eqasim is shown in Figure 1. Briefly, the three main steps of the pipeline are the following :

- Population, activities and their locations are generated by processing the gathered data
- The travel demand is converted to a desired simulator (*i.e.* MATSim) input format
- Run the mobility simulation

The first step is also structured in 5 parts: sociodemographic information, and residence locations of households generation; income information addition to each household; activity chains attachment to the synthetic persons; places of work and education assignment; choice of locations of all other activities in the persons' activity chains. While running the simulation in the last step of the pipeline, Eqasim also replaces the usual scoring method of MATSim used for the re-planning stage by its own discrete mode choice model. By taking into account such large amount of data from the real world, the simulation obtained is quite close to depict a real-world scenario, and have been used successfully in different mobility simulations [11, 12, 13].

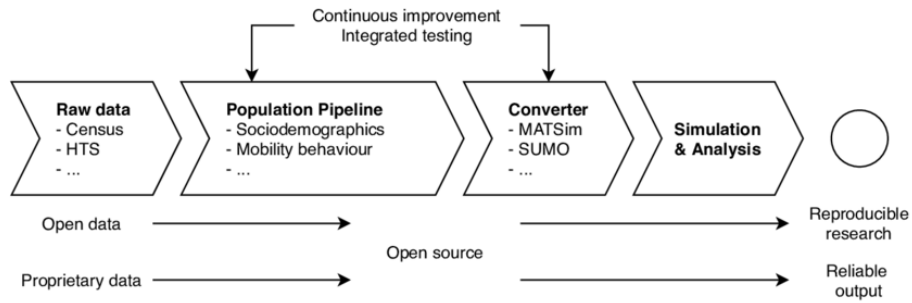


Fig. 1. Eqasim pipeline. Picture from the original article [7].

## 2.2 Bus Stops Spacing Problem

The bus stops spacing problem is a well-known problem in the literature that has been treated in multiple ways. This problem aims to position the bus stops of the public transportation system in order to optimize different possible criteria such as passengers travel time, economic costs or environmental impact [10, 8, 9, 14, 15].

Different approaches exist to solve this problem. One approach is based on the population density and the walking distance to the closest bus stop. In this approach, the authors propose to search the position of stops by dividing the road network in different areas. They combine game theory and, respectively, models based on Thiessen polygons [8] and Voronoi diagrams [9] to solve the problem. Notice that these methods are based on static properties of areas (*e.g.* population density in given area) to search for the best bus stop locations, and not on a simulation that takes into account the possible interactions between individuals choices of travel plan. In [14], the authors propose a static model to compute the optimum bus stops spacing that takes into account multiple constraints (from basic physics applied to the bus such as velocity or acceleration/deceleration to

operating costs). Another approach uses a bi-level optimization method where the bus routes are divided into links of the same distance, that represent potential bus stops locations, with the objective of minimizing the social cost of the overall transport system [15]. Here again this work is not based on a agent-based simulation.

On the contrary in [10], an agent-based simulation is used to compute the global quality of the bus stops locations. From an artificial scenario of mobility plan called SIALAC [16], which can tune the size of the population and activities clusters, the bus stops problem is set in order to minimize travel time of passengers. Given the position of bus stops, the computation time of a single simulation is around 1 minute. So, the authors propose a surrogate assisted optimization approach to optimize the objective given in a SIALAC scenario.

### 3 Bus stop spacing problem model

In this work, following the work of Leprêtre et al. [10], a simulation-based optimization approach is used. The simulation engine is the agent-based simulator MATSim (see Section 2.1). We define all possible locations for bus stops. The optimization problem is then a pseudo-boolean optimization problem. A binary vector represents the bus stop locations: each stop location is associated with a binary variable. When the binary value is equal to one the bus stop is active, and the potential bus stop is not used when the binary value is zero. The objective function to minimize is the average travel time of bus passengers.

One main goal of the work is to provide a pipeline able to create a mobility scenario based on real-world data and allows the reproductibility of the optimization process. We use the Eqasim pipeline to create the synthetic population depicting a real-world scenario of the city of Calais, France, following the indications of [17]. We gathered data from the following sources:

- Institut national de la statistique et des études économiques (INSEE), France : socio-demographic census; origin-destination trips for professional and school activities; population totals; income tax data; service and facility census; zoning registry,
- Ministry of Ecology (France): national household travel survey,
- Institut géographique national (IGN), France: IRIS zoning system; buildings database,
- Open Street Map (OSM): road network,
- Zenbus: GTFS for bus in Calais

This exact data set in fact produces a scenario for the whole region of Hauts-de-France, but Eqasim provides tools to restrict a scenario to a certain area, allowing us to restrict our scenario to the city of Calais and its surroundings up to 50 kilometers away as shown in Fig. 2. We obtain a simulation representing a day of traffic with approximately 240,000 trips for 13,500 agents. About 10% of trips are done by travelling using public transport.



**Fig. 2.** Bus stops in Calais, France, and its surroundings.

For the bus stops spacing problem, an additional step is added prior the eqasim pipeline. Indeed, according to the active bus stops, the time schedule of each bus must be modified. The configuration files, *i.e.* GTFS files, used by the Eqasim pipeline need to be updated by following this process:

- All stops marked as 0 in the binary string are deleted from the files.
- The bus schedule is updated using the Manhattan distance between every pair of consecutive stops (*i.e.* the distance between stops using the road) and a speed of 5m/s.

To speed up the process, all Manhattan distances between each pair of stops were calculated once in python<sup>1</sup>, and saved in a matrix that is read to update the bus schedule according to the set of bus stops used. The whole pipeline to reproduce the simulation part of our work on the bus stops problem in the city of Calais is available to git repository at the following URL: <https://gitlab.com/vvendi/eqasimoptim>. The repository also gives the different sources for the data that can be modified to generate another scenario.

## 4 Experimental Setup

This section gives the precise first scenario of the bus stop spacing problem for the city of Calais in France. Using the same pipeline, it is possible to adapt this scenario to another city.

<sup>1</sup> using <https://github.com/nathanrooy/taxicab>

As mentioned in section 3, we consider this problem as a pseudo-boolean problem where the goal is to find the best combination of bus stops locations to keep to minimize passenger average travel time using bus. To keep this problem at a practicable dimension, we chose to focus only on the bus line n°2 in Calais, containing 76 stops. We later arbitrarily created and added 32 additional bus stops to the existing route to create potential alternatives to existing stops for a total of 108 bus stops. The whole set of bus stops in the bus line n°2 is shown in figure 3. Some of these stops perfectly fit the initial route of the bus while some change the path and create detours. The reason behind these "out of the route" stops is to add heterogeneous choices that the later algorithm will have to detect and not pick in the final set of bus stops. Our problem search space is thus  $\{0, 1\}^{108}$  (76 existing stops + 32 additional ones) and the evaluation function to minimize is the average travel time using bus expressed in seconds. Notice that this pseudo-boolean black-box is large compare to black-box problems in the literature for which the dimension is below 50 binary variables [18, 19, 20].

An example of the simulation & optimization process is described in Algorithm 1. Notice that the optimization algorithm used in line 4 is not given as the goal of this work is to provide the reproducible optimization scenario. Running the Eqasim pipeline, and running the simulator MATSim in lines 6, and 8 are the computational expensive parts of the process. The computation time is approximately 30 minutes on AMD Ryzen 5 4600H 6 Cores, 12 Threads, 3 GHz (turbo 3.9 GHz). The computation time is much larger than the artificial scenario of the SIALAC benchmark [21] where the computation time was at most 1 minute for the same number of agents, but our scenario is based on raw data and re-planning module of MATSim.

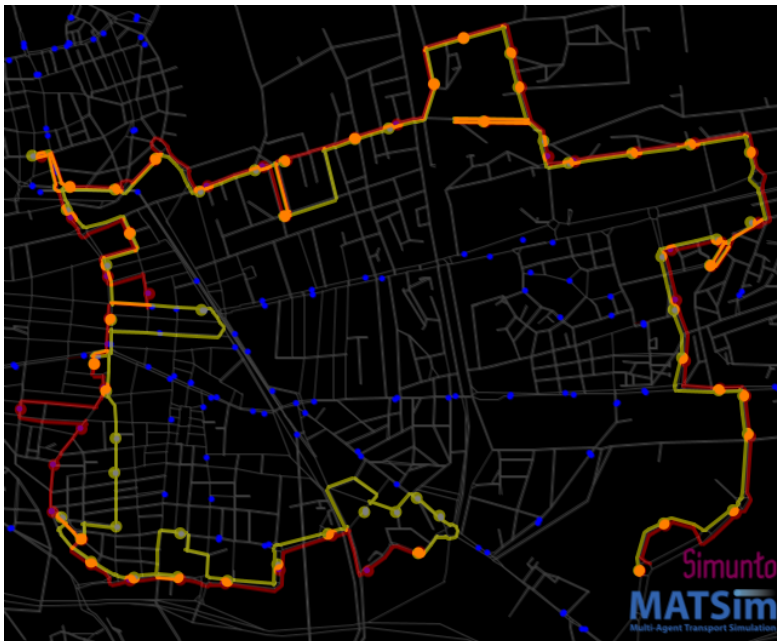
Although we have not conducted experimental analysis of optimization algorithms for this scenario, we evaluate the relevance of our scenario by evaluating random solutions, *i.e.* random selection of bus stops: each bus stop has a probability of 0.5 to be open. The distribution of objective value, *i.e.* the average travel time, of 1000 random solutions is shown in Fig. 4. On average, the travel time is 2746.26 seconds, and the standard deviation is 12.48. The minimum travel time is 2713.10 which shows the potential improvement for future optimization algorithms.

---

**Algorithm 1** Simulation & Optimization Process
 

---

- 1:  $D_{all} \leftarrow$  Gather open-source data
  - 2:  $x^* \leftarrow$  Best known solution, initially  $\emptyset$
  - 3: **repeat**
  - 4:  $x \leftarrow$  Solution, *i.e.* the set of bus stops to keep, selected by an opt. algorithm
  - 5:  $D \leftarrow$  Edit the GTFS files from  $D_{all}$  according to  $x$
  - 6:  $S_{npdc} \leftarrow$  Run the Eqasim pipeline to create the population according to  $D$
  - 7:  $S_{calais} \leftarrow$  Cut the scenario to restrict it to the city of Calais
  - 8:  $f(x) \leftarrow$  Run the MATSim simulation on the scenario  $S_{calais}$
  - 9:  $x^* \leftarrow x$  if  $f(x) < f(x^*)$
  - 10: **until** time limit
-



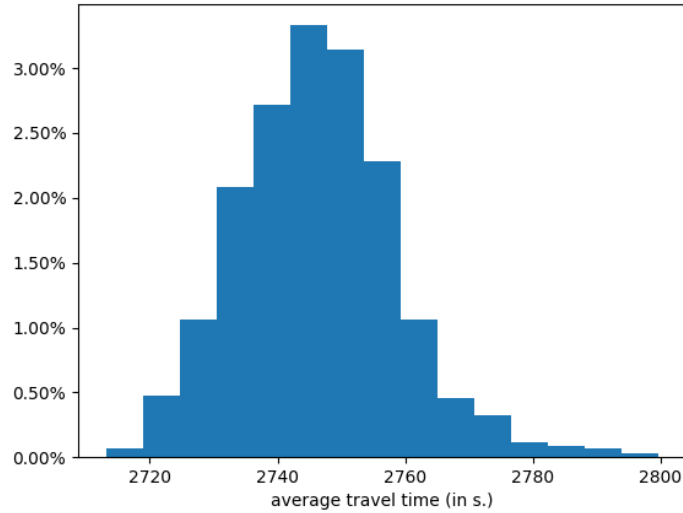
**Fig. 3.** Bus stops: in yellow stops 1-56, in red stops 57-108.

## 5 Conclusion and Future Work

In this paper, we propose a pipeline for reproducible scenario of mobility problem aiming to solve bus stops spacing problems. As an example, we show a usage of the pipeline to create an instance of the bus stops spacing problem using the city of Calais as a play-field. The MATSim simulator with Eqasim pipeline based on real-world available open source data ensures the reproducibility. In this work, the model of the bus stops spacing problem is a black-box pseudo-boolean problem where the goal is to find the best combination of bus stops to keep to minimize the average travel time using bus. So, we add to the initial pipeline the automatic modification of the bus schedule. This scenario is available on a git repository. At a first step, with random solutions, we show the relevance of this scenario to test future optimization algorithms for this expensive evaluation function.

Future work will propose to analyze the performance of existing optimization algorithms in a fair way by using this large scale black-box problem and to design dedicated optimization algorithms for the City of Calais or larger cities. Another step will consist of adding new evaluation criteria such as the number of passengers per mode and pollutant emissions [22, 23]. By doing so, the problem





**Fig. 4.** Objective values, *i.e.* average travel time in second, distribution of 1000 random solutions.

of bus spacing will become multi-objective, and methods will need to be adapted to solve it.

## 6 References

- [1] Milos Balac et al. “Modeling competing free-floating carsharing operators—A case study for Zurich, Switzerland”. In: *Transportation Research Part C: Emerging Technologies* 98 (2019), pp. 101–117.
- [2] Rolando Armas, Hernan Aguirre, and Daniel Orellana. “Evolutionary bi-objective optimization for the electric vehicle charging stand infrastructure problem”. In: *Proceedings of the Genetic and Evolutionary Computation Conference*. 2022, pp. 1139–1146.
- [3] Rolando Armas, Hernan Aguirre, and Kiyoshi Tanaka. “Multi-objective optimization of level of service in urban transportation”. In: *Proceedings of the Genetic and Evolutionary Computation Conference*. 2017, pp. 1057–1064.
- [4] Grace O Kagho, Milos Balac, and Kay W Axhausen. “Agent-based models in transport planning: Current state, issues, and expectations”. In: *Procedia Computer Science* 170 (2020), pp. 726–732.
- [5] Pablo Alvarez Lopez et al. “Microscopic traffic simulation using sumo”. In: *2018 21st international conference on intelligent transportation systems (ITSC)*. IEEE. 2018, pp. 2575–2582.

- [6] Andreas Horni, Kai Nagel, and Kay Axhausen, eds. *Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press, Aug. 2016, p. 618.
- [7] Sebastian Hörl and Milos Balac. “Introducing the eqasim pipeline: From raw data to agent-based transport simulation”. In: *Procedia Computer Science* 184 (2021). The 12th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 4th International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops, pp. 712–719.
- [8] Changjiang Zheng, Shukang Zheng, and Genghua Ma. “The Bus Station Spacing Optimization Based on Game Theory”. In: *Advances in Mechanical Engineering* 7.2 (2015), p. 453979. DOI: 10.1155/2014/453979.
- [9] Zhenjun Zhu et al. “Optimization of Urban Mini-bus Stop Spacing: A Case Study of Shanghai (China)”. In: *Tehnicki Vjesnik* 24 (June 2017), pp. 949–955.
- [10] Florian Leprêtre et al. “Combinatorial Surrogate-Assisted Optimization for Bus Stops Spacing Problem”. In: *Biennial International Conference on Artificial Evolution (EA 2019)*. Mulhouse, France, Oct. 2019.
- [11] Milos Balac and Sebastian Hörl. “Simulation of intermodal shared mobility in the San Francisco Bay Area using MATSim”. In: *2021 IEEE international intelligent transportation systems conference (ITSC)*. IEEE. 2021, pp. 3278–3283.
- [12] André Maia Pereira et al. “An Advanced Travel Demand Synthesis Process for Creating a MATSim Activity Model: The Case of Ústí nad Labem”. In: *Applied Sciences* 12.19 (2022), p. 10032.
- [13] Biao Yin et al. “Evaluation of Low-Traffic Neighborhoods and Scale Effects: The Paris Case Study”. In: *Transportation Research Record* (2023), p. 03611981231170130.
- [14] Anthony A Saka. “Model for determining optimum bus-stop spacing in urban areas”. In: *Journal of Transportation Engineering* 127.3 (2001), pp. 195–199.
- [15] Ángel Ibeas et al. “Optimizing bus stop spacing in urban areas”. In: *Transportation Research Part E: Logistics and Transportation Review* 46.3 (2010), pp. 446–458. ISSN: 1366-5545.
- [16] Florian Leprêtre et al. “Walsh functions as surrogate model for pseudo-boolean optimization problems”. In: *The Genetic and Evolutionary Computation Conference (GECCO 2019)*. Proceedings of the Genetic and Evolutionary Computation Conference. Prague, Czech Republic: ACM, July 2019, pp. 303–311.
- [17] Sebastian Hörl and Milos Balac. “Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data”. In: *Transportation Research Part C: Emerging Technologies* 130 (2021), p. 103291. ISSN: 0968-090X.
- [18] Sébastien Verel et al. “A surrogate model based on Walsh decomposition for pseudo-boolean functions”. In: *PPSN 2018 - International Conference*

- on Parallel Problem Solving from Nature*. Vol. 11102. Lecture Notes in Computer Science. Coimbra, Portugal, Sept. 2018, pp. 181–193.
- [19] Changyong Oh et al. “Combinatorial bayesian optimization using the graph cartesian product”. In: *Advances in Neural Information Processing Systems* 32 (2019).
  - [20] Hamid Dadkhahi et al. “Combinatorial black-box optimization with expert advice”. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020, pp. 1918–1927.
  - [21] Florian Leprêtre. “Méta-modélisation, simulation et optimisation de flux urbains”. 2020DUNK0571. PhD thesis. 2020.
  - [22] Kickhöfer Benjamin. “Emission Modeling”. In: *Multi-Agent Transport Simulation MATSim*. Ed. by Andreas Horni, Kai Nagel, and Kay Axhausen. London: Ubiquity Press, Aug. 2016. Chap. 36. DOI: 10.5334/baw.
  - [23] Friederike Hülsmann et al. “Towards a multi-agent based modeling approach for air pollutants in urban regions”. In: Jan. 2011, pp. 144–166. ISBN: 978-3-941790-77-3.
  - [24] Shuai Guo. “An Introduction To Surrogate Optimization: Intuition, illustration, case study, and the code”. In: Dec. 2020.