Open access • Proceedings Article • DOI:10.1109/ITST.2015.7377394

# On the importance of real data for microscopic urban vehicular mobility trace



Marie-Ange Lebre, Frédéric Le Mouël, Eric Menard

**Institutions:** University of Lyon, Valeo

**Published on:** 01 Dec 2015 - International Conference on ITS Telecommunications

# On the Importance of Real Data for Microscopic Urban Vehicular Mobility Trace

Marie-Ange Lèbre, Frédéric Le Mouël, Eric Ménard

# On the Importance of Real Data for Microscopic Urban Vehicular Mobility Trace


Marie-Ange Lèbre[*†], Frédéric Le Mouël[*], Eric Ménard[†]

[*] University of Lyon, INSA-Lyon, CITI-INRIA, F-69621 Villeurbanne, France

Email: {marie-ange.lebre@insa-lyon.fr, frederic.le-mouel@insa-lyon.fr}

[†] VALEO, Advanced Technology Developement, F-94000 Creteil, France

Email: {marie-ange.lebre@valeo.com, eric.menard@valeo.com}



***Abstract*—Vehicular networks reflect user mobility behavior and present complex microscopic and macroscopic mobility patterns. Microscopic mobility is often simplified in macroscopic systems and we argue that its impact is too largely neglected. Notwithstanding improvements in realistically modeling and predicting mobility, few vehicular traces - especially complex microscopic ones - are available to validate such models. In this paper, we present a realistic synthetic dataset of vehicular mobility over two daily traffic peaks in a small area: the Europarc roundabout in the town of Creteil, France. We outline how the description and comprehensive representation of local mobility at an intersection, such as the roundabout chosen here, is important for any interpretation made of it.**



***Keywords*—synthetic trace, real data, microscopic vehicular mobility, roundabout, traffic flows, traffic lights, simulation.**


## I. Introduction

Vehicular networks have attracted increasing attention from the research community over recent years. The reason is that communications between vehicles, infrastructures, users and urban areas will provide new services and applications for safety, traffic fluidity, infotainment, etc [1]. These application functionalities (collision warning, local traffic information), communication performance (range of communication, data saturation), environment interaction (traffic pattern) and the need for experimenting in real contexts therefore require specific and costly field tests including the road infrastructure and management systems to test cooperative communications. However, it is difficult to obtain real traffic situations during Operational Field Tests making it necessary to rely on simulation as a solution. When simulations are used, attention must be paid to faithful representation of reality, i.e. real vehicular dynamics characterized by high peaks, highly variable speeds, road topology, road rules, trips by users, user behaviors, etc. The macroscopic and microscopic aspects must be studied accurately in order to give credible results. Due to the difficulty of obtaining real traces (cost, data confidentiality), simplistic models have been employed to get around the problem [2]. The increasing interest in vehicular networks reflects the growing complexity, progress and quality of reproducing car movements. Some approaches base their models on real traces [3] while others try to reproduce real traffic flows [4]. Nowadays, the challenge is to generate real traffic traces that are realistic and faithful to reality, to identify traffic flows and run realistic models of traffic assignments [5] and allow the correct interpretation of the study performed on these traces [6]. Few large-scale traces on urban macroscopic mobility have been made public and available [7], but real local microscopic traces are difficult to find. We argue that analysing microscopic mobility can help to explain local phenomena such as traffic light control and local bottlenecks, and can be used to test systems for improving the fluidity of traffic locally and globally, and to observe data saturation impact at specific points of the city [8].

In this paper, we present an original roundabout dataset and explain the methodology and improvements applied to the map and all the items of equipment in order to match them with real traffic flows. Then, we show the importance of a real dataset by comparing it to other well-known mobility models used in simulations and we show that common traffic flow models employed in simulators are not realistic. Our paper supports this hypothesis presented in [9] -that were never confirmed by real data- by showing that the distribution of inter-arrival times of vehicular users follows a Pareto-Weibull distribution. Finally, we highlight the importance of fine-grained and realistic traffic lights models.

## II. The Europarc Roundabout dataset

The vehicular mobility trace presented in our paper focuses on a roundabout in the working area called Europarc located in the town of Creteil, France. It aims at reproducing the traffic mobility on this large roundabout, and the traffic light system included in the real infrastructure, with the highest possible level of realism. In the following section, we describe the features and the methodology used to generate the mobility trace.

### *A. Road topology*

The street layout was obtained from the OpenSteetMap database (OSM). OSM is a collaborative project aimed at creating a free editable map of the world. Maps include roads, railways, buildings, points of interests, parks, shopping centers, leisure areas all described as a function of reality. OSM road information is generated and validated by imaging and GPS traces. Therefore, OSM is commonly regarded as the highest quality road data publicly available today. We exported the roundabout from OSM in an xml file. It represents an area of $1.1km^2$ around the Europarc Roundabout. Then, we processed it using the Java OpenStreetMap Editor (JOSM) to repair the OSM file and make it compatible with the real roundabout (described in section III).

### *B. Vehicular mobility*

The microscopic mobility of vehicles is simulated with the Simulator of Urban MObility software (SUMO version

0.22) an open source microscopic and continuous road traffic simulation package. This simulator is able to model the behavior of individual drivers. SUMO can import a network from OpenStreetMap and faithfully reproduces roads, traffic lights, stop signs, numbers of lanes, etc. The microscopic models implemented by SUMO for moving vehicles is the Krauss car following model. In this model, SUMO allows choosing the safety distance between cars, acceleration, deceleration and driver imperfection. Krajzewicz's lane-changing model is used as it allows creating, regulating and distributing traffic in the case of multiple lanes, and it includes overtaking decisions. These models have been widely validated in the transportation research community. The high scalability of SUMO makes it the most complete and reliable of today's open-source microscopic vehicular mobility generators.

### *C. Traffic data*

The traffic information used for the Europarc Roundabout is obtained from real observations of vehicle flows. Cameras on the site were placed in order to observe the traffic in detail. A manual count was performed to create an O/D (origin/destination) matrix every 900 seconds. The camera observations distinguished between vehicle types: trucks, vehicles, motorcycles and buses. As motorcycles represent 5% of the vehicles in the morning and 3.6% in the evening, we don't consider them. Also, we do not have information on trucks (heavy or light); as they represent 4.5% of the vehicles in the morning and 1.5% in the evening they are not taken into account. The resulting O/D matrix faithfully mimics the daily movements of vehicles in the area for a 2-hour period (7.15 AM to 9.15 AM) on Tuesday morning and 2 hours (5.00 PM to 7.00 PM) on Tuesday evening in a working week in September.

### *D. Roundabout Characteristics*

The roundabout is composed of 6 entries and 6 exits, 2 entries and 2 exits are exclusively reserved for buses. The road for buses is shown in Figure 3 by visible induction loops. Other induction loops -not indicated- exist on external roads to measure the rate of occupancy. 15 traffic lights are included in the map: 12 are located inside it to regulate traffic ; 3 are for pedestrians outside and work even if there are no pedestrians ; 4 of the 12 traffic lights are dedicated for buses. In addition, 4 white lanes inside are intersections without traffic lights -non-stop intersections- that indicate the exits of the roundabout.

### *E. Simulation*

The OSM map of the Europarc roundabout is converted to a format readable by SUMO. The traffic data mentioned later in the article are described in terms of flows in an xml file used by SUMO. This file allows introducing "x" vehicles from a departure lane to an arrival lane and specifying the type of vehicle. This flow description method is that most suitable for the type of data we have though it introduces a small error on the number of vehicles in each data stream of 900 seconds (see table I). However, this error is negligible and the simulated flow is similar to $94.4\%$ of a real flow.

## III. Repairing the dataset

The results obtained by running the vehicular mobility simulation available by the OSM map and traffic data were unusable. We simulated only the first 2700 steps of the traffic flows (45 minutes). Figure 1 shows the evolution over time of the number of vehicles crossing the roundabout. The number of vehicles present in the simulation rapidly grows to exceed the capacity of the roundabout and gives an unrealistic configuration. It can be seen by the fact that the number of vehicles that end their trip increases very slowly over time. Due to congestion, a small fraction of the cars present on the road are able to cross the roundabout. The road topology cannot sustain the volume of cars injected according to the traffic demand model. In the following sections, we discuss the reasons for such a simulation result and present solutions for them.

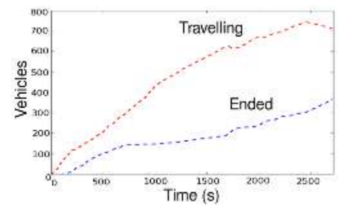


Fig. 1: Simulated Vehicle Status

*1) Inconsistent road information:* The first source of error in the simulation was identified in the OpenStreetMap road topology. The information embedded in the map can sometimes be inconsistent with reality (as the roundabout is a local map, few contributors of OpentStreetMap have worked on it). Consider the situation in Figure 2 (a) which shows the real world aerial photograph of the roundabout (left), matched with the OSM road topology for one intersection (right). We see that the east street for vehicles contains four lanes, although in reality there are only three lanes. As most of the vehicles follow the same trajectories, some lanes are not used, as can be seen in the white circle in Figure 2 (b), causing vehicles to get stuck by waiting indefinitely for the possibility to go to the right lane. Other problems are represented by wrong traffic movement restrictions, of which an example is shown in the same Figure 2, the streets on the right and the left are exclusively for buses but vehicles in the south and east have the right to turn into these streets. Another inconsistency can be identified when observing vehicle traffic. This is represented by wrong or nonexistent connections between lanes in the map file. For instance, in Figure 2 (b), lane 3 is not connected to neighboring lanes 1 and 2, thus it cannot be used by vehicles and congestion occurs. We identify a large number of nonexistent connections, which forbid using the full capacity. We solved these topology problems with JOSM.

*2) Traffic light programs:* The OpenStreetMap road information is imported by SUMO through an automated conversion process that proved not to be error-free. For traffic light deployment, OSM already includes data on the presence or absence of traffic lights at road junctions and SUMO automatically sets the periodicity of the green and red lights according to the priority of the roads entering each junction. They must first be placed correctly to match reality, with the deletion of some of them. When the first errors on road topology, connections, and traffic light positions are solved, simulations still do not present the correct flows. Indeed, traffic light programs are automatic and fixed by OSM, but in reality traffic light programs are complex. First of all, we must add induction loops deployed on the surface of bus lanes in order

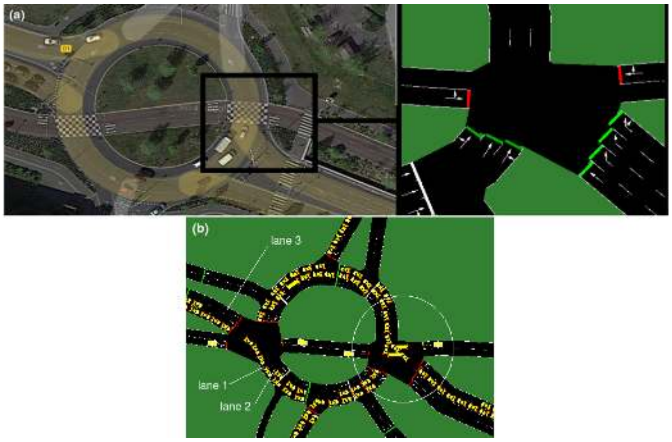


Fig. 2: Example of errors in the road topology conversion (a) and through vehicle traffic (b)

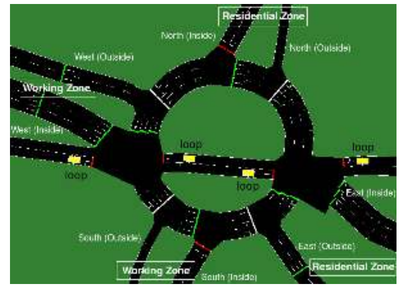


Fig. 3: The Europarc Roundabout Map on the simulator

to detect them and accord their priority. Indeed, when a bus passes over an induction loop, the traffic light facing it turns green. Then, other induction loops are located on each external lane in order to detect whether the rate of occupancy is higher than 15%, and if this is the case, certain traffic light phases are extended according to dynamic traffic regulation schedules provided by the General Department Council.

*3) Departure and Arrival of flows:* We have access to the flows in each directions: south to north/east/west/south, north to north/east/west/south, east to north/east/west/south and west to north/east/west/south every 900 seconds. Using the tools in SUMO, we can enter the xml flows. Despite the ease of introducing the flows, we have to modify the entry and exit positions of the flows in order to be realistic. Indeed, we do not know exactly when the vehicles were counted, so to avoid a bottleneck at the entries of the roundabout, it is necessary to introduce them outside it. Therefore, the departure and arrival lanes are outside the roundabout. Despite these corrections, and the entry of the exact vehicle data in the flow files, certain differences subsisted between our data and the real flows. Table I shows the similarity degree every 15 min in the evening. However we are very close to the real traffic flows, the simulated data is similar, in average, to 95.5% of real data every 15 min in the evening.

| Hours | Real data | Simulated data | Similarity |
|---|---|---|---|
| 5.00-5.15 | 658 | 526 | 80% |
| 5.15-5.30 | 643 | 657 | 97% |
| 5.30-5.45 | 660 | 670 | 98% |
| 5.45-6.00 | 636 | 607 | 95% |
| 6.00-6.15 | 670 | 649 | 97% |
| 6.15-6.30 | 636 | 633 | 99% |
| 6.30-6.45 | 661 | 657 | 99% |
| 6.45-7.00 | 541 | 600 | 90% |

TABLE I: Similarity between real data and synthetic data

## IV. Local urban mobility

The resulting dataset comprises around ten thousand trips over a period of two hours in the morning and two hours in the evening. The simulated traffic mimics the normal daily road activity on the Europarc roundabout. These data are interesting because they represent a strategic roundabout between residential and working areas (Figure 3). Evidence of the correct behavior of the simulated mobility is given in Figure 4. They can be compared to Figure 1 which show mobility without any modifications of the map. We now conform to the real flows between 7.15 AM and 9.15 AM and 5.00 PM and 7.00 PM. It is clear that the number of vehicles traveling now follows increasing traffic demand during the peak morning and evening hours. The number of finished trips now increases over time as more and more drivers reach their destinations. The mean travel time and the average speed over time during the morning period confirm the previous conclusions as constant behaviors can be observed over time. We now find that the microscopic traffic matches well with

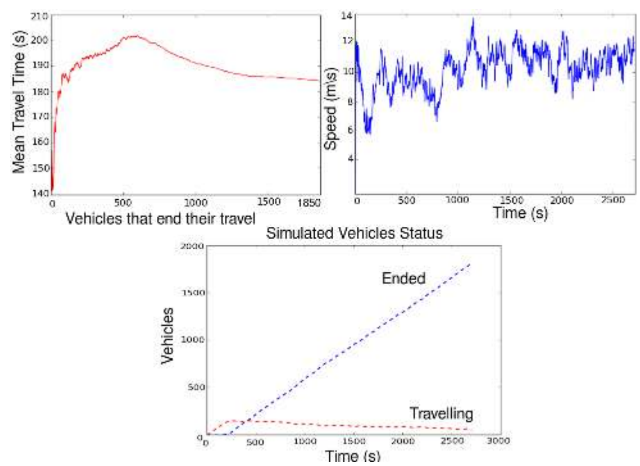


Fig. 4: Mean travel time, speed and Simulated Vehicles Status

reality. Figures 5 and 6 represent the mean rate of occupancy for each edge on the roundabout (see Figure 3 in section V for their locations) during the two peak hours in the morning and the two peak hours in the evening. For each direction, North, South, East, and West, there are significant traffic flows in the morning in one direction, and in the opposite direction in the evening. Indeed, drivers leave their homes to go to work in the morning. Congestion then stems from the working area in the evening as the drivers return back home. These observations show the realism of the dataset and their implementations. The key properties (flow rate, travel time, etc.) of real-world mobility patterns are faithfully reproduced in the dataset.

## V. Importance of real data

### A. Traffic Flows

We compare two commonly used traffic flow models in order to show the importance of having knowledge of the real

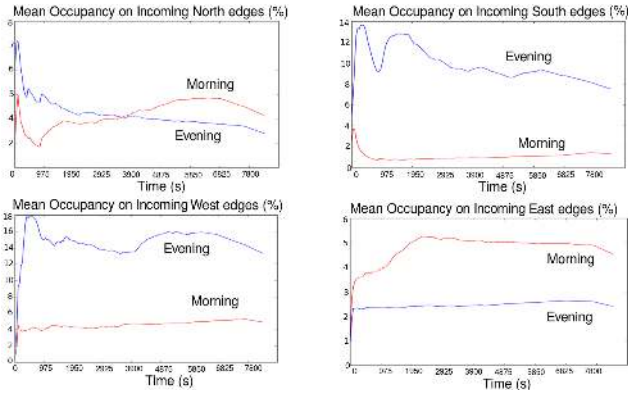


Fig. 5: Occupancy at the Incoming Edges

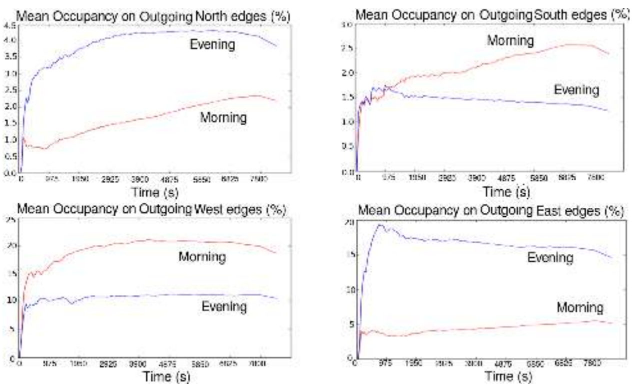


Fig. 6: Occupancy at the Outgoing Edges

data on a local map with our real data set. We compare the rate of occupancy of each outgoing and incoming edge on the roundabout with the rate of occupancy of the two other vehicle flow models.

To use these two models, we hypothesize that in the simulation of the traffic flows, we only know the number of vehicles between 7.15 AM and 9.15 AM (around 5000 trips), and we keep the real traffic flow of the buses (their frequencies can be found easily on the website of the town's public transportation company). The first traffic flow is modeled by a Poisson distribution, as it is a widely used model [10]. Thus 5000 trips are generated according to this distribution. We consider that the topology of the map helps us to determine the approximate flow capacity, so we introduce more vehicles into the West and East roads than into the South and North roads, because they are larger (3 lanes versus 2). The second traffic flow used is a turn algorithm. We estimate the turn probabilities on the incoming and outgoing edges in the roundabout. The turn model is simple: a vehicle entering the roundabout has a 70% chance of taking an exit and 30% chance of continuing its travel on the roundabout. To compare them we evaluate the rate of occupancy at the incoming and outgoing edges with the different vehicle flows (Figures 8 and 7). The model with a Poisson Distribution has a regular rate of occupancy at the outgoing edges. The rate is close to the real flow at the incoming smaller edges (South and North). However, unrealistic congestion is observed at the incoming West edges, which shows that we introduced too many vehicles on this corridor. Based on our approximation of the traffic flows according to the topology of the map using the Poisson Distribution, we obtained traffic flows close to reality for the smaller edges but not for the larger edges.

The model with turn probabilities is close to the real data only for the outgoing West edges. Indeed, in reality, the turn probabilities are not deterministic; they change stochastically over time. In our basic model, they are fixed and the occupancy of the incoming North edges and outgoing and incoming South edges is unrealistic. However, it can be seen that when real information is used for the turn probabilities on the roundabout real flows can be reproduced on the incoming East edges and outgoing West edges. To sum up, Figure 9 shows the

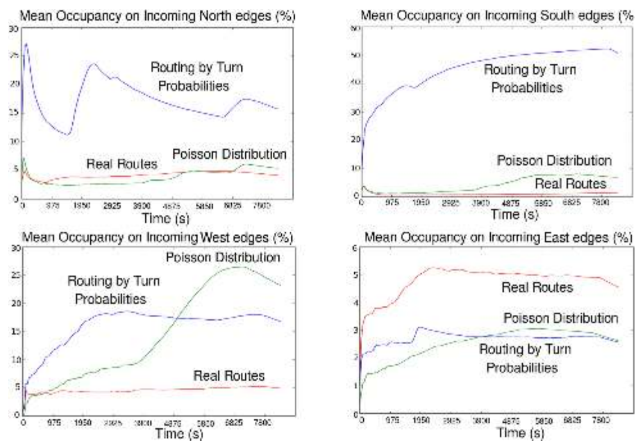


Fig. 7: Occupancy at the Incoming Edges

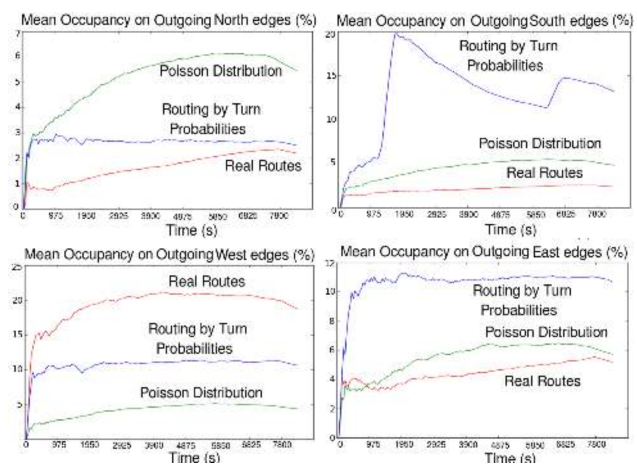


Fig. 8: Occupancy at the Outgoing Edges

difference in travel time and waiting time for each traffic flow. The routes with the Poisson Distribution can be realistic in terms of traffic flows but they do not match with reality. It can be seen that the traffic flow increases logically as a function of the Poisson Distribution and that the turn probability model creates congestion due to wrong knowledge of the flows. This difference can lead to wrong interpretations of whatever type of test (wireless communication, traffic control, etc.) might be performed for this roundabout.

Our results are sufficient to invalidate the common assumption that user arrivals follow a Poisson process. To complete our study, we take an interest to the article [9] which assumes the

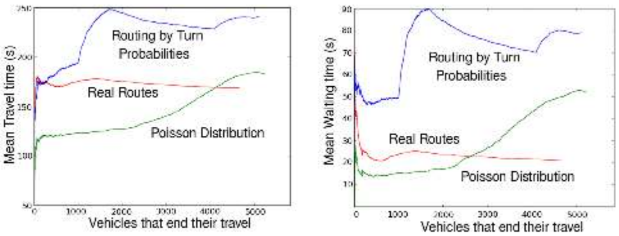


Fig. 9: Mean Travel Time and Mean Waiting Time with the different traffic flows

hypothesis that the inter-arrival times of vehicular users follow a hybrid Weibull-Pareto [11], the hypothesis was nevertheless not confirmed by real data. The figure 10 which represents the Complementary Cumulative Distribution Function (CCDF) for inter-arrival times in our map let us also conclude that the hybrid Weibull-Pareto distribution fits very well the CCDFs for inter-arrival times in the considered scenario. All vehicular users enter within 10 seconds or less of each other. The interval drops to 3 seconds or less for $50\%$ of the vehicular users.

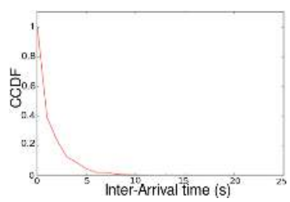


Fig. 10: Distribution of the inter-arrival times

*B. Traffic lights Programs*

We now show the importance of the traffic light program in a simulation. We consider the real programs which are optimized and adapt themselves in real-time according to the rate of occupancy detected on their controlled links. Moreover, induction loops detect the presence of buses which causes the lights to switch to the green phase. We compare it with the situation where the phases are fixed -OSM default fixed traffic light- and with an adaptive model available in SUMO where phase prolongation are based on time gaps between vehicles. In Figure 11, fixed traffic light model clearly leads

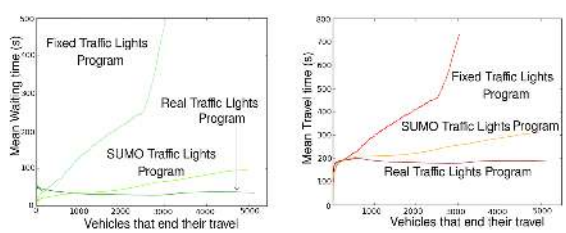


Fig. 11: Mean Waiting Time and Travel Time with different traffic light programs

to a complete congestion. SUMO adaptive traffic light model is more close to reality, but however is not stable and increases linearly and would probably reaches capacity problems when scaling. These results show the impact and the importance of the traffic light model that should fit the topology reality.

## VI. Conclusion and Future Work

The simulation of the Europarc roundabout using real collected data shows that wrong or artificial vehicular flow, topological and traffic light coordination data can lead to very different traffic flows and vastly different traffic properties. In particular, tests on mobility traces based on commonly used models appear to be far from real flow rates and fail to reflect real daily traffic flows. However we must clarify that some models, even not realistic, can nevertheless adress specific characteristics targeted and depend on the research goals. The hybrid Pareto-Weibull model fits very well the real vehicular inter-arrivals time [9]. The trace generated from real collected data is available to the community (http://vehicular-mobility-trace.github.io) under a Creative Commons Attribution-NonCommercial 4.0 International License. In future work, we will compare our real trace with more elaborate models capturing: optimal particle swarms for traffic lights [12]. We will also study decentralized online stop-free strategies to remove traffic lights and consider fully-connected autonomous cars in a complex roundabout, by optimizing the fluidity property.


## VII. Acknowledgments

We thank the General Departmental Council of Val de Marne (94) in France for its precious help, especially Matthieu Perigaud, responsible for traffic studies, for providing all the data needed for this study.


## References


[1] M.-A. Lèbre, F. Le Mouël, E. Ménard, J. Dillschneider, and R. Denis, "Vanet applications: Hot use cases," *arXiv preprint :1407.4088*, 2014.

[2] A. M. Mohammed and A. F. Agamy, "A survey on the common network traffic sources models," *IJCN*, vol. 3, pp. 103–116, 2011.

[3] K. Minkyong, D. Kotz, and K. Songkuk, "Extracting a mobility model from real user traces," in *IEEE INFOCOM*, vol. 6, 2006, pp. 1–13.

[4] X. G. Li, B. Jia, Z. Y. Gao, and R. Jiang, "A realistic two-lane cellular automata traffic model considering aggressive lane-changing behavior of fast vehicle," *Physica A: Statistical Mechanics and its Applications*, vol. 367, pp. 479–486, 2006.

[5] J. Harri, F. Filali, and C. Bonnet, "Mobility models for vehicular ad hoc networks: a survey and taxonomy," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 4, pp. 19–41, 2009.

[6] S. Y. Lin, A. Vinel, C. Englund, and L. Chen, "Poster: Adaptive wavelength adjustment (awla) for cooperative speed harmonization," in *Vehicular Networking Conference (VNC), 2014 IEEE*. IEEE, 2014, pp. 113–114.

[7] H. Huang, Y. Zhu, X. Li, M. Li, and M. Y. Wu, "META: A mobility model of MEtropolitan TAxis extracted from GPS traces," in *IEEE WCNC*, 2010, pp. 1–6.

[8] M.-A. Lebre, F. Le Mouël, and E. Ménard, "Partial and local knowledge for global efficiency of urban vehicular traffic," in *IEEE 82nd Vehicular Technology Conference (VTC'2015-Fall)*, 2015.

[9] S. Uppoor and M. Fiore, "Characterizing pervasive vehicular access to the cellular ran infrastructure: an urban case study," *Vehicular Technology, IEEE Transactions on*, vol. 64, pp. 2603–2614, 2014.

[10] J. Wu, A. Abbas-Turki, and A. El Moudni, "Discrete methods for urban intersection traffic controlling," in *IEEE VTC Spring*, 2009, pp. 1–5.

[11] S. Karpinski, E. M. Belding-Royer, and K. C. Almeroth, "Living on the edge: The distribution of flows across mobile nodes in large wireless networks," Technical Report, Tech. Rep., 2005.

[12] J. Garcia-Nieto, A. Olivera, and E. Alba, "Optimal cycle program of traffic lights with particle swarm optimization," *Evolutionary Computation, IEEE Transactions on*, vol. 17, no. 6, pp. 823–839, Dec 2013.