Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 37 (2019) 242–249

Transportation Research Procedia

www.elsevier.com/locate/procedia

21st EURO Working Group on Transportation Meeting, EWGT 2018, 17th – 19th September 2018, Braunschweig, Germany

Validation of an agent-based travel demand model with floating car data

Carlos Llorca[a]*, Sasan Amini[a], Ana Tsui Moreno[a], Rolf Moeckel[a]

[a] *Technical University of Munich, Arcisstraße 21, 80333 Munich (Germany)*

Abstract

This paper compares the results of the agent-based travel demand model MITO (Microscopic Travel Demand Orchestrator) with Floating Car Data. MITO is developed using household travel survey data, and uses the traffic assignment model MATSim. The model estimates the travel demand for an average working day and is applied to the metropolitan area of Munich. In contrast to traditional approaches where travel demand models are validated using the local traffic counts, average travel speed from Floating Car Data (FCD) are used in this study. The main advantage of using FCD is that they cover extremely large parts of the network, whereas the local traffic counts are sparse and limited to a few major streets. The average link travel time and average speed between the model estimation and the FCD were compared with the goal of validating travel time calculations within an agent-based transport model.

Keywords: Travel demand modeling; Floating Car Data (FCD); Agent-based model

1. Introduction

Agent-based travel demand models predict the trips made by individuals. Trips are generated and distributed among the destinations and later assigned to the network. These models are usually calibrated with household travel surveys and validated against local traffic counts. The model is presumed to be valid if modeled and observed traffic volumes at the location of the traffic count stations match more or less. However, the distribution of traffic counts stations is

* Corresponding author. Tel.: +49-89-289-28443
E-mail address: carlos.llorca@tum.de

10.1016/j.trpro.2018.12.189

sparse. Furthermore, traffic counts are inherently unreliable due to manual counting errors, vehicle type misclassifications, failure of automatic traffic recorders and errors introduced by interpolating and forecasting of traffic counts (Moeckel & Donnelly, 2016). These shortcomings of count data lead to an inaccurate and biased validation of travel demand models. In other words, no information regarding the quality of the performance of the model is provided at the rest of the network where no count data are available (Bernardin, Trevino, Slater, & Gliebe, 2015). Alternatively, user-generated mobile sensor data, in particular Floating Car Data (FCD), are becoming more relevant as they cover extensive areas of the urban road network. The main advantage of FCD in comparison to local traffic counts is that they provide travel time along travelled paths as well as local speed information on almost all parts of the network, whereas local traffic counts are limited to a few number of links and commonly only provide the number of vehicles but no speed information. However, an important limitation of FCD data is that they do not cover all vehicles on the network. A careful analysis is required to assess how far this technology may be capable of replacing traditional travel count data.

1.1. Literature review

The use of Floating Car Data (FCD) requires, firstly, to convert point-based data to valuable information in the form of trip routes, travel times, or further valuable information. FCD data consists of a set of points registered typically by a GPS tracker, and thus, characterized by coordinates, time, speed and heading. The conversion of this points to actual (or most probable) trajectories of vehicles is carried out by using a map matching and path inference algorithm. Different methods have been proposed to transform FCD into trajectory data, adjusted to the level of detail and accuracy of the sources (Hashemi & Karimi, 2014; Hsueh & Chen, 2018; Lou et al., 2009; Newson & Krumm, 2009; Rahmani & Koutsopoulos, 2013; Rahmani, Koutsopoulos, & Jenelius, 2017).

Once FCD data is matched to a road network, the most probable trajectories of the sampled vehicles are derived which can be used to analyze and even predict travel demand. The immediate result of processing the FCD using map matching algorithms is the calculation of travel times on the network under diverse traffic circumstances. For example, Rahmani et al. (2017) calculated travel times using FCD data and compared them with Google Maps API travel times. Based on that external dataset, the map matching algorithms were improved.

Further research focused on the aggregation of individual FCD trajectories to describe traffic conditions on the network. For instance, Yong-chuan, Xiao-qing, li-ting, & Zhen-ting (2011) identified congested roads based on the analysis of average speeds. Then, the links with low speed values were classified as congested. The authors qualitatively compared these results with observations of traffic conditions. In other cases, FCD have been used to characterize travel demand patterns. Kong et al. (2018) generated travel demand datasets (origin/destination matrices) by using FCD. Similar results were obtained by combing travel time calculation and route choice probabilities estimated using FCD (Nigro, Cipriani, & del Giudice, 2017). However, the use of conventional data sources is still required, given that floating vehicles represent only a small proportion of the total flow and may exhibit their flows need to be expanded. Other examples of FCD applications are found in the literature, such as Chen, Jiang, Liao, Zou, & Zhang (2018), who have used FCD to analyze travelers' behavior to predict future travel demand patterns on a road network or Mannini, Cipriani, Crisalli, Gemma, & Vaccaro (2017), who analyzed on-street parking behavior.

Previous research used FCD to validate or calibrate travel demand models. Jenelius, Kristoffersson, & Fransson (2017) used FCD to validate the results of the dynamic traffic assignment model for Stockholm. The authors used a Macroscopic Fundamental Diagram approach, which relates the space mean density and flow on a network level (Daganzo & Geroliminis, 2008). One of the advantages of FCD is that it includes a higher proportion of the road network, in contrast to traffic counts (Bernardin et al., 2015). Therefore, its potential to validate the results of travel demand models is high. However, the application of FCD to validate travel demand agent-based models is still in an early stage.

1.2. Research contribution

This paper compares the results of an agent-based travel demand model with travel times derived from FCD in Munich, Germany. The main goal is to analyze the suitability of FCD for such task, analyzing its coverage and to compare the modeled average speed against the measured speeds for various road types in the study area.

2. Methodology

2.1. Agent based travel demand

An agent-based travel demand model has been developed within the framework of an integrated land use/ transport model (Moeckel & Nagel, 2016). The travel demand model was applied for the case study of the Munich metropolitan area (Munich, Augsburg, Ingolstadt, Landshut, Rosenheim and surrounding municipalities with relevant commuter shares to those five cities). A synthetic population for the entire study area was generated, including the socio-demographic characteristics and geolocation of households and jobs.

Based on the synthetic population, the travel demand model MITO (Microscopic Travel Demand Orchestrator) generates individual synthetic trips and assigns a destination and a mode to each of them. Trip generation, destination choice and mode choice were estimated using the household travel survey Mobilität in Deutschland (DLR, 2008). Both home-based work trips and home-based education trips are doubly-constrained to housing and workplace locations that are specified in the synthetic population. Other trip purposes are singly-constrained. Once the demand is created, demand adjustment (time of day choice) and traffic assignment (route choice) are performed using MATSim (Multi Agent Transport Simulation). MATSim is an agent-based simulation framework designed for large-scale scenarios (Horni, Nagel, & Axhausen, 2016). The MATSim traffic flow model is a queue-based approach, computationally efficient but omitting car-following or lane-changing interactions. To improve MATSim runtimes, the simulation typically runs a subsample of the entire populating, while scaling down road capacity by the same proportion (Llorca, Moreno, Okrah, & Moeckel, 2017).

The model simulates traffic flows for the year 2011 (base year in which the synthetic population read by MITO was generated) for an average working day. MATSim outputs include the complete path of agents, as well as link and node volumes and link speeds along the simulation period.

2.2. FCD dataset

The FCD dataset was provided by the Allgemeiner Deutscher Automobil-Club (ADAC), which covers the entire urban area of the city of Munich (Germany) for the whole month of October 2017. The raw data consist of timestamp, coordinates, instantaneous speed and vehicle type, among others. The position of individual vehicles is updated every minute, which is not ideal for fast moving vehicles, but it provides sufficient detail for the analysis of average travel speeds. It is worth to mention that the penetration rate of the probe vehicles is not known to the authors but it is believed to be rather low. To derive the average travel time for every origin-destination pair in the network, the raw data must be first map matched to links and then aggregated to obtain the average weekday speed of each link for every minute throughout the day. The map matching has been performed using a Hidden Markov Model that was introduced (Newson & Krumm, 2009). In order to achieve a representative weekday behavior, the school and public holidays as well as Fridays have been excluded from the dataset. Moreover, the time window has been limited to 6 to 10 AM as it is assumed that most of the work and education trips provided by the MITO/MATSim model will occur during this time period.

2.3. Merging FCD and MATSim outputs

The study area is selected taking into account the limitations of both data sources: the traffic assignment outputs and the FCD dataset. In order to achieve a comparable traversal speed on different links, first, the two types of output data needed to be reconciled, as MATSim outputs and FCD are not represented at the same resolution. Although MATSim and the FCD datasets are referenced to the OpenStreetMaps (OSM) road network, their structures are slightly different. The FCD are matched originally to the actual OSM network, which contains links with two or more nodes that can be one-way or two-way. The MATSim traffic assignment, in contrast, is performed in a transformed version of that network, which contains solely one-way links that connect only two nodes (the from-node and the to-node). Figure 1 shows the different link specifications.

Therefore, MATSim and FCD were merged by aggregating MATSim links that share a common OSM link. The values of the MATSim-related variables at the resolution of OSM links was the weighted average of the MATSim link corresponding variables, as shown in the equation 1.

$$speed_i = \frac{speed_1 \cdot v_1 \cdot l_1 + speed_{1'} \cdot v_{1'} \cdot l_{1'} + speed_2 \cdot v_2 \cdot l_2 + speed_{2'} \cdot v_{2'} \cdot l_{2'}}{v_1 \cdot l_1 + v_{1'} \cdot l_{1'} + v_2 \cdot l_2 + v_{2'} \cdot l_{2'}} \quad (1)$$

Where:

- i is an index of OSM link (depicted in gray in Figure 1), 1, 2, 1' and 2' are indexes of MATSim links (depicted by black arrow lines in Figure 1)
- l is length of each link
- v is the traffic volume in the analyzed period of each link
- speed is the average speed of each link

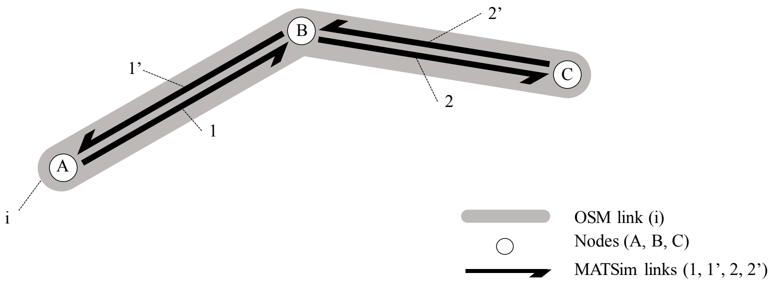

Figure 1. MATSim links and OSM links

After this matching procedure, the following variables are known for each link: link id, one-way or two-way, link length, road type (motorway, arterial and other), MATSim traffic volume during the selected time interval, MATSim average speed in the selected time interval, FCD average speeds and its standard deviation in the selected time interval, FCD number of observed vehicles by in the selected time interval.
The study is focused on comparing observed and simulated speed distributions on the links of the reconciled network. The selected time interval from which the FCD was obtained was the morning 6 to 10 AM period of working days in October, 2017. The model simulated a working day in the base year 2011. While the years are not matching very well, it is assumed that traffic volumes have changed only slightly over these six years, and the impact on links speeds is small.

3. Analysis

First, the coverage of the FCD data is analyzed. After running the model and merging MATSim outputs with FCD, the analysis focused on link speeds between the two data sources.

3.1. Coverage of the network by FCD data

Since FCD are a user-generated dataset, it is assumed that the number of observations is proportional to the number of vehicles on the network. However, the employed FCD are obtained from a relatively small subsample of vehicles. Therefore, the number of observations on secondary roads may be too small to provide useful speed measurements. This subsection analyzes the coverage of FCD with respect of the OSM network, as a function of the minimum number of observations by OSM link. Figure 2 shows the percentage of links with the number of FCD vehicles shown on the horizontal axis. The red lines represent the percentages with 30 vehicles (15%), 50 vehicles (11%) and 100 vehicles (6.5%). Percentages are calculated with respect of the 37,809 OSM links represented by the MATSim network. Even considering the links that have at least one observation, no more than 27% of all links have records of the FCD data.

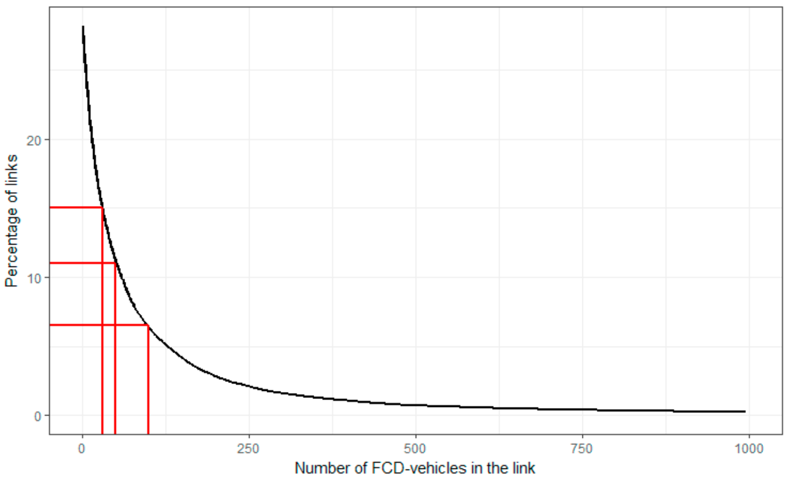

Figure 2. Percentage of links with respect of the MATSim network as a function of the number of FCD vehicles (red lines highlight the values with number of vehicles equal to 30, 50 and 100)

(a) urban zone, at least one FCD observation/link

(b) suburban zone, at least one FCD observation/link

(c) urban zone, at least 30 FCD observation/link

(d) rural zone, at least 30 FCD observation/link

Figure 3. Spatial distribution of FCD (Color legend: blue: motorways, red: arterial roads, pink: other)

Although these percentages are rather low, it is worthwhile to analyze the spatial distribution of the FCD coverage on the network. Figure 3 shows two parts of the area, the left one (3.a and 3.c) in a dense urban location (Munich-Schwabing) and the right one in the suburbs (3.b and 3.d) (City of Germering). The coverage of FCD is practically limited to arterials and motorways and not representative of residential and service roads (Figure 3). Considering only links with more than 30 vehicles, no continuous road is covered.

3.2. Comparison of link speeds

The following comparison is limited to the links were the number of observations from FCD is higher than 30 to ensure a more reliable estimation of average speeds. Figure 4 directly compares the MATSim simulated speeds and the FCD observed speeds. The comparison is segmented by road type (in columns) and congestion level (in rows). The sample size (number of links) is written in each plot. The red band represent the range where the absolute error in speed is within ±20 km/h. Lastly, the blue lines are linear regression models for MATSim speed vs. FCD speed. Generally, the plots show a significant dispersion of the speeds from the two different sources. The first column, which shows arterial roads, presents a higher sample and shows a significantly higher correlation between MATSim speeds and FCD speeds. The speeds are particularly close for congested roads (based on traffic volumes modeled with MATSim). On the contrary, the FCD observed speeds on motorways are not as well represented by MATSim. The

MATSim values are significantly higher compared to FCD observations. MATSim vehicles have a higher tendency to travel at free-flow speed (particularly visible at 50 km/h for arterials and 120 km/h for motorways), which is not achieved by observed vehicles as frequently. On Arterials, this difference is likely to be caused by the insufficient representation of traffic lights in MATSim. On Motorways, we assume that MATSim vehicles tend to travel "too smoothly," whereas observed vehicles have to deal with erratic driver behavior, such as sudden breaking or sudden lane change, unknown to MATSim. It should also be noted that each link is represented by one dot, regardless of link length or link relevance.

The same data were represented in form of maps shown in Figures 5 and 6 to describe the spatial distribution of the speed error. Here, error was defined as the difference between the FCD measured speed and the average speed obtained from MATSim. Thus, negative values mean that MATSim overestimates the FCD speeds. Figure 5 shows only motorways and Figure 6 shows only arterial roads, both in Munich and its surroundings. Figure 5 suggests that MATSim simulates speeds significantly higher than the FCD measurements for motorways. For arterials, however, Figure 6 shows a disperse error distribution. Most of areas with errors under -20 km/h are concentrated in the city center or at network nodes, which suggests that the delay at intersections and interchanges is underrepresented in MATSim. There is a significant number of links with a small error of ±20 km/h, mainly located in arterial roads surrounding the city center.

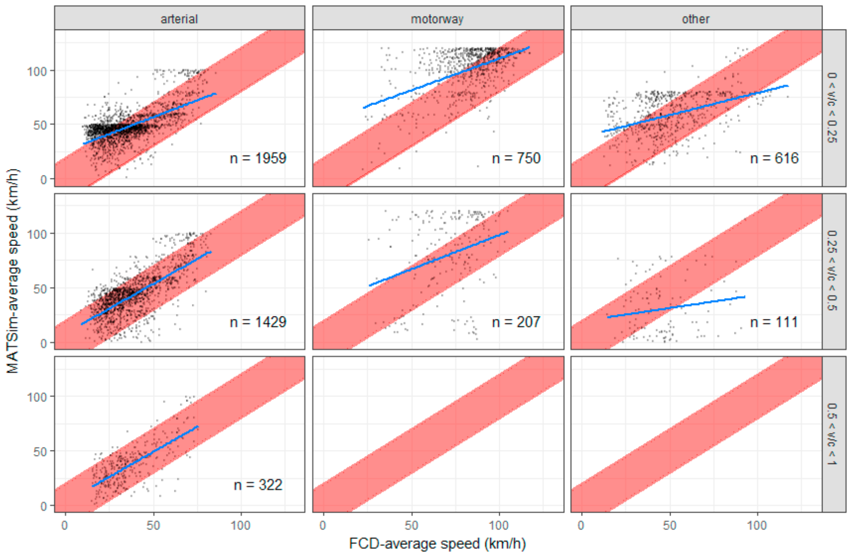

Figure 4. MATSim-speed vs. FCD-speed by road type (columns) and congestion level (rows)

4. Discussion and conclusions

In this paper, we assessed firstly the suitability of FCD data to validate the results of an agent-based traffic assignment model. Assuming that FCD data are adequate for that purpose, we compared the speeds obtained from the MATSim traffic assignment model with those measured by the FCD data. The difference between speeds from FCD and MATSim were significant. Though, there were some patterns in the distribution of errors that are worth investigating. The errors were lower for arterial roads (tagged as *primary*, *secondary* and *trunk* in OpenStreetMaps) compared to motorways. Moreover, the errors were lower for congested parts of the network (measured by volume-to-capacity rates over 0.5 in the selected 4 h time period). In the case of motorways, the results showed that MATSim simulation model is overestimating the speeds (or ignoring congestion of the network). We assume that the insufficient representation of delays at intersections is the main reason for differences observed between FCD and MATSim data. We are investigating the implementation of a node delay for arterials that shall represent traffic lights and turning delays. Several issues can explain the larger errors at motorways. Firstly, the travel demand MITO and the traffic assignment in MATSim did not include heavy vehicles. At the selected time interval, the percentage of heavy vehicles in some of the motorways that were compared is higher than in the arterial roads, where it is almost zero (in the Munich ring motorway A99 the percentage of heavy vehicles is around 10%, while in the analyzed arterial urban roads heavy traffic is restricted during daytime). Secondly, MITO omits external flows (traffic from outside of the study area, to outside of the study area or through the study area), which may be more relevant for motorways compared to other arterial roads. Lastly, there is an inherent source of errors in the comparison, as explained as follows. MATSim speeds are the result of a simplified traffic flow model, which assumes a unique speed for all vehicles within a link, while FCD data is an observation of spot speed every minute, which may not be representative of the average speed on the link. An idea for future research is to compare the travel time along selected routes, and not only the travel time at specific links. This may reduce the errors caused by the different speed definition in both sources, and may be a better indicator on how accurate is the travel demand model represents travel times from A to B. Despite the mentioned limitations, FCD data are a valuable tool to validate the results of traffic assignment models. Even with a relatively low coverage, the number of roads with reliable observations is higher than the available traffic counting stations, especially those that are able to measure speeds.

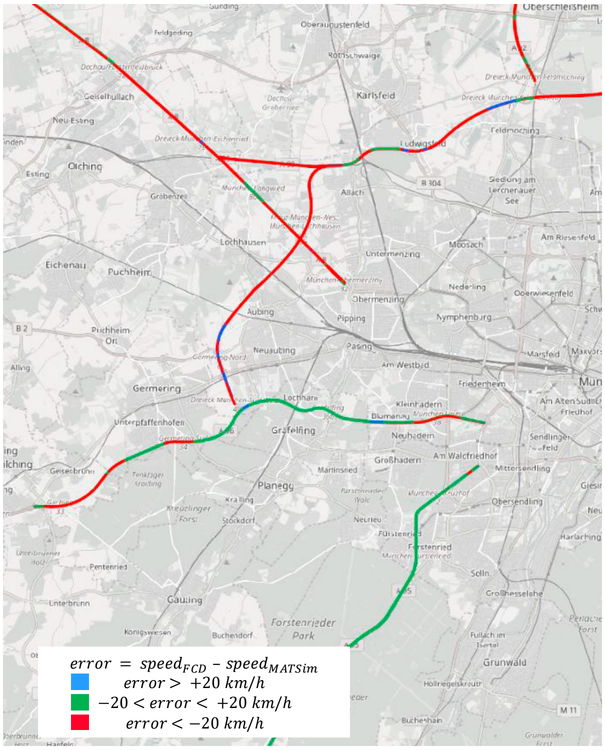

Figure 5. Map of speed absolute error (motorways)

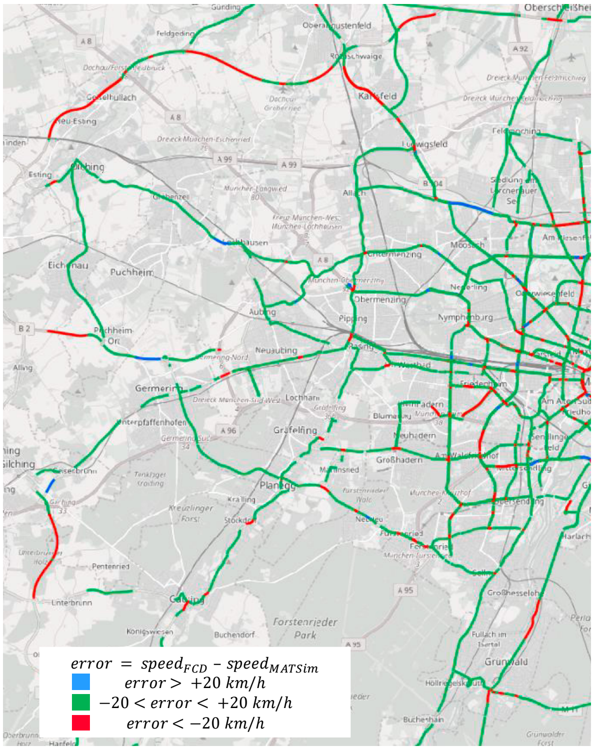

Figure 6. Map of speed absolute error (arterial roads)

5. Acknowledgement

The research was completed with the support of the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement n° 291763. This work is part of the TUM Living Lab Connected Mobility (TUM LLCM) project. The acquisition of FCD data has been partially funded by the Bavarian Ministry of Economic Affairs, Energy and Technology (StMWi) through the Bavarian Digitalization Center, an initiative of the Bavarian State Government.

6. References

Bernardin, V. L., Trevino, S., Slater, G., & Gliebe, J. (2015). Simultaneous Travel Model Estimation from Survey Data and Traffic Counts. *Transportation Research Record: Journal of the Transportation Research Board*, *2*, 69–

76. https://doi.org/10.3141/2494-08

Chen, Y., Jiang, X., Liao, L., Zou, F., & Zhang, M. (2018). Driving Behavior Motivation Model Research Based on Vehicle Trajectory Data (pp. 36–44). Springer, Cham. https://doi.org/10.1007/978-3-319-70730-3_5

Daganzo, C. F., & Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, *42*(9), 771–781. https://doi.org/10.1016/j.trb.2008.06.008

DLR. (2008). *Mobilität in Deutschland*. Retrieved from http://www.mobilitaet-in-deutschland.de/pdf/MiD2008_Abschlussbericht_I.pdf

Hashemi, M., & Karimi, H. A. (2014). A critical review of real-time map-matching algorithms: Current issues and future directions. *Computers, Environment and Urban Systems*, *48*, 153–165. https://doi.org/10.1016/j.compenvurbsys.2014.07.009

Horni, A., Nagel, K., & Axhausen, K. W. (2016). *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press. https://doi.org/http://dx.doi.org/10.5334/baw

Hsueh, Y. L., & Chen, H. C. (2018). Map matching for low-sampling-rate GPS trajectories by exploring real-time moving directions. *Information Sciences*, *433–434*, 55–69. https://doi.org/10.1016/j.ins.2017.12.031

Jenelius, E., Kristoffersson, I., & Fransson, M. (2017). Validation of Traffic Simulation Models Based on the Macroscopic Fundamental Diagram. In *Transportation Research Procedia* (Vol. 27, pp. 561–568). https://doi.org/10.1016/j.trpro.2017.12.073

Kong, X., Xia, F., Ning, Z., Rahim, A., Cai, Y., Gao, Z., & Ma, J. (2018). Mobility Dataset Generation for Vehicular Social Networks Based on Floating Car Data. *IEEE Transactions on Vehicular Technology*, 1–1. https://doi.org/10.1109/TVT.2017.2788441

Llorca, C., Moreno, A. T., Okrah, M. B., & Moeckel, R. (2017). Traffic assignment for an integrated land use and transportation model in a large metropolitan area: case study of Munich. In *6th Symposium of the European Association for Research in Transportation (hEART 2017)*. Haifa.

Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W., & Huang, Y. (2009). Map-matching for low-sampling-rate GPS trajectories. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09* (p. 352). New York, New York, USA: ACM Press. https://doi.org/10.1145/1653771.1653820

Mannini, L., Cipriani, E., Crisalli, U., Gemma, A., & Vaccaro, G. (2017). On-Street Parking Search Time Estimation Using FCD Data. In *Transportation Research Procedia* (Vol. 27, pp. 929–936). https://doi.org/10.1016/j.trpro.2017.12.149

Moeckel, R., & Donnelly, R. (2016). A model for national freight flows, distribution centers, empty trucks and urban truck movements. *Transportation Planning and Technology*, *39*(7), 693–711. https://doi.org/10.1080/03081060.2016.1204091

Moeckel, R., & Nagel, K. (2016). Maintaining Mobility in Substantial Urban Growth Futures. In *International Scientific Conference on Mobility and Transport Transforming Urban Mobility, mobil.TUM*. Munich, Germany.

Newson, P., & Krumm, J. (2009). Hidden Markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09* (p. 336). New York, New York, USA: ACM Press. https://doi.org/10.1145/1653771.1653818

Nigro, M., Cipriani, E., & del Giudice, A. (2017). Exploiting floating car data for time-dependent Origin–Destination matrices estimation. *Journal of Intelligent Transportation Systems*, 1–16. https://doi.org/10.1080/15472450.2017.1421462

Rahmani, M., & Koutsopoulos, H. N. (2013). Path inference from sparse floating car data for urban networks. *Transportation Research Part C: Emerging Technologies*, *30*, 41–54. https://doi.org/10.1016/j.trc.2013.02.002

Rahmani, M., Koutsopoulos, H. N., & Jenelius, E. (2017). Travel time estimation from sparse floating car data with consistent path inference: A fixed point approach. *Transportation Research Part C: Emerging Technologies*, *85*, 628–643. https://doi.org/10.1016/J.TRC.2017.10.012

Yong-chuan, Z., Xiao-qing, Z., li-ting, Z., & Zhen-ting, C. (2011). Traffic Congestion Detection Based On GPS Floating-Car Data. *Procedia Engineering*, *15*, 5541–5546. https://doi.org/10.1016/J.PROENG.2011.08.1028