SUMO User Conference 2025

Conference paper

https://doi.org/10.52825/scp.v6i.2633

Published: 15 Jul. 2025

SUMO Simulation of DLR's Research Intersection

Yun-Pang Flötteröd[1,*] and Peter Wagner[1,2]

[1]DLR Institute for Transportation Systems, Germany

[2]TU Berlin, Institute for Land- and Sea Traffic, Germany

*Correspondence: Yun-Pang Flötteröd, yun-pang.floetteroed@dlr.de

Abstract. Trajectory data are great data to work with, since they are the most natural data for traffic. However, they provide considerable challenges when tried to put into a micro-simulation framework such as SUMO. This work here gives an example what had to be done to arrive at a simulation that is driven by these data. Succeeding in this, microscopic tools can be much better tested against real data.

Keywords: Trajectory Data, Calibration and Validation, SUMO

1. Introduction

When it comes to traffic data, a lot of improvements could be noted during the past 30 years. Among this, trajectory data have been the most complete data to be used to understand traffic, and to fit models so that they can dutifully replicate traffic phenomena. Arguably one of the earliest work is the one in [1], which can be found also in [2]. One of the projects that broke ground in this respect was the NGSIM project [3] in the early 2000. During the recent 15 years, driven especially by AI-based object recognition, these datasets have become quite common, see [4], [5], [6] for a short list and [7] for a little bit of an overview.

Recently, we have made public one day of trajectory data from DLR's research intersection (RI for short in the following) in Braunschweig (Brunswick), Germany [8]. The visualization of this intersection in SUMO is illustrated in Figure 1. The data contain the trajectory of any object that has been detected by the machinery of the RI (these were motorized vehicles, bicycles, and pedestrians), and in addition, it contains the data of the traffic signals and weather data with different sampling rates. The latter one has been ignored in this research, while the trajectories as well as the traffic signals have been transformed so that they can drive a SUMO [9], [10] simulation. The focus of this work is on investigating the extent to which the above-mentioned data can be used to set up a traffic simulation with SUMO and the extent to which SUMO can reproduce the respective traffic conditions with the given dataset.

2. Dealing with the data

The trajectory data have been recorded by video cameras. Objects in the videos were recognized by applying an object recognition system [11], subsequently classified and

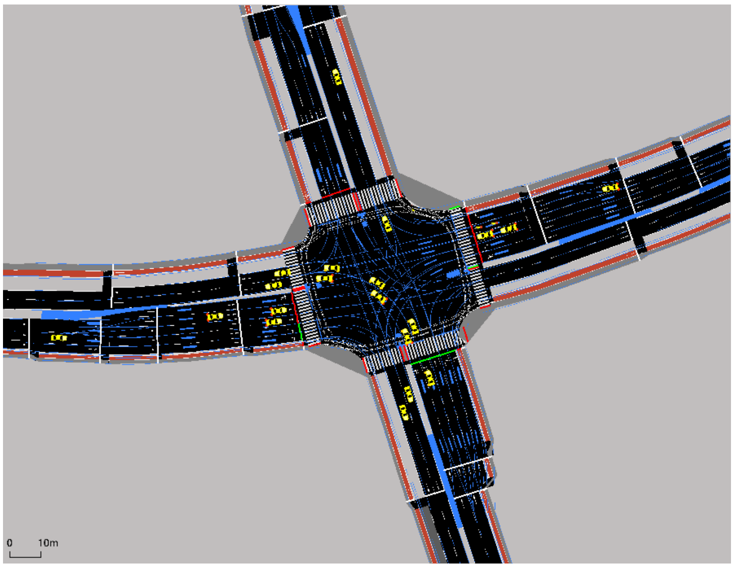


Figure 1. *SUMO screenshot of the RI. (blue: the road geometry lines)*

they were geo-located from the video coordinates into real-world co-ordinates[1]. In subsequent steps, the data were Kalman filtered, to achieve two goals:

- Smoothing the positions $(x(t), y(t))$ of the objects at time t, which runs with the video frequency of 20 Hz ($\Delta t = 0.05$).
- To obtain the speed and acceleration vectors $(v_x(t), v_y(t))$, and $(a_x(t), a_y(t))$, and from this the scalar speed and acceleration $(v(t), a(t))$, respectively.

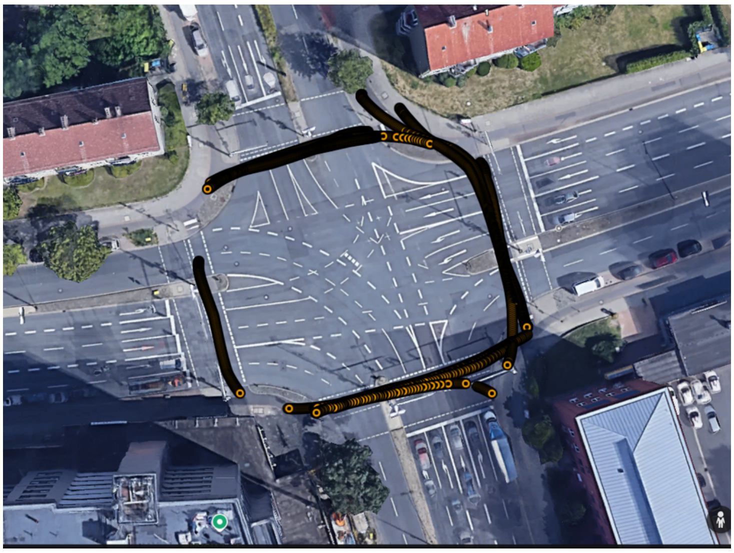
Note, that glitches in the process broke some of the trajectories into small pieces; some of them could be repaired in the post-processing, but one must be aware, that some objects had been there, but had not been recorded, and vice versa (it seems to happen less frequently that the process invented objects). This is mentioned here, since it is important if one wants to reproduce these data with SUMO: if a leading vehicle has vanished in the recorded dataset, then the simulated trajectory of the follower in SUMO will become wrong when compared with the trajectory in the dataset.

What is truly nice with this dataset is that it contained also the trajectories of the bicycles and the pedestrians that do interact with the motorized traffic. In order to get the simulation right, these data are needed as well, since they especially hinder the right-turning motorized traffic. However, during the processing of the data it became clear, that especially the classification of pedestrians and bicyclists is questionable. Therefore, some of the objects had to be reclassified. We have nevertheless indications, that pedestrians are under-represented in these data.

2.1 Data cleaning and correction

The trajectory dataset not only contains positions, but also information on speed, acceleration, yaw and the probabilities of the identified object classes. There are a total of six object classes: car, van, truck, motorbike, bicycle, and pedestrian. These data are the main source to reconstruct the set of traffic flows, the origin-destination (O-D) relations of the traffic flows for each object class and their start positions and speeds for the simulation.

[1]Note, that the detailed algorithm is not published. It is based on stereo cameras and therefore tries to reconstruct objects in 3D, not in 2D only, by joining voxels with a similar optical flow into objects.



Source: map data: Google Earth (©2025); Image: GeoBasis-DE/BKG (©2009); Picture from the following date: August 3, 2022

Figure 2. *An example of incorrect pedestrian classification*

Several issues have been revealed during the data processing, e.g. some interrupted trajectories are only within the RI, in the bike/pedestrian crossings or in the edges approaching/leaving the RI, and the start and end points of some trajectories are very close to each other. There are also some road users with extraordinary speeds, e.g. cyclists and pedestrians with a travel speed higher than 5 m/s or 3 m/s, respectively. Figure 2 shows an example of incorrect pedestrian classification, where the trajectories of the objects are in the bicycle crossings, but they are identified as pedestrians. It is therefore necessary to clean up the data and correct it before generating the traffic demand. Overall, the following measures were implemented with the additionally defined polygons for cycle paths/crossing and vehicle lanes.

- Objects whose start and end points are close to each other (less than 6 meters) are removed, taking into account that a complete path should be at least 60 meters long in order to cross the RI.
- Bicycles with illegal direct u-turn have been removed since there are fully equipped bicycle paths and crossings at the intersection.
- Pedestrians traveling at speeds higher than 2.2 m/s and on the bike lanes/crossings have been reclassified to bicycles.
- Pedestrians traveling at speeds less than 1 m/s and on the pedestrian crossings have been set to travel at a speed between 1 and 1.5 m/s (general average walking speed) to avoid congestion in the simulation.
- Motorbikes with low speed (less than 4 m/s) and on bike paths/crossings have been reclassified to bicycles.
- Bicycles in vehicle lanes with speed higher than 5 m/s have been reclassified to motorbike. When the corresponding speeds are less than 5 m/s, they are removed.

2.2 Simulation network

As the network data, we have used the OpenDrive map generated by the Institute for Driver Assistance and Connected Mobility (IFM) of Kempten University of Applied Sciences in the research project VVM [12]. This map is available under the CC-BY-NC license. We also used the respective shape file as a reference background map to further improve the geometry of the simulation network manually with SUMO's tool netedit and bring it as close as possible to reality. The dimensions of some network elements, such as the width of road, land and crossings, were examined accordingly. The accuracy of the matching of the trajectory positions to the network can then be ensured. Figure 1 shows that the road geometry, pedestrian and bike crossings fit well with the geometry lines marked in blue. In addition, the vehicle and bike movement curves within the intersection were also updated according to the trajectory data.

2.3 Traffic demand

The demand for SUMO was extracted from the cleaned trajectory data. The main idea is to decide the O-D relation of each road user with the predefined O-D polygons/zones and the respective start position according to each road user's first and last position. Although simple in principle, a number of work-arounds were needed, mainly due to incomplete trajectories that are not located in the O-D polygons.

In order to rebuild representative traffic states in the simulation, we need to make sure that at least the number of vehicles corresponds to reality. Thus, a 2nd matching of the trajectories to the network has been executed, only for unidentified vehicles and bicycles. This matching is based on the existing markings on each vehicle lane, i.e. right, through and left. Once a lane is matched, the respective departure time will be adjusted according to the distance between the original start point and the matched lane and the recorded start speed. This adjustment is intended to prevent vehicles/bicycles from missing their green phases and waiting longer in the simulation.

In addition, the 6 object classes in the dataset have been matched to SUMO's vehicle/pedestrian classes accordingly, i.e. car: passenger car; van: delivery vehicle; truck: truck; bus: bus; motorbike: motorcycle; bicycle: bicycle and pedestrian: pedestrian.

2.4 Traffic signal data

The traffic light at the RI is an adaptive signal control depending on traffic conditions. The signal control logic is not available. Thus, the recorded traffic light data in the dataset were converted into the SUMO format as the timing plans for the traffic signals. The empirical signal groups have been already defined on the basis of traffic movements. It is therefore easily possible to compare them with the traffic movements in the simulation and set up the corresponding signal groups in the simulation. If competing traffic movements, e.g. westbound through traffic and eastbound left-turn traffic, share the same green phase, one of them will have a major green phase and the other ones will have a minor green phase in the simulation to avoid unrealistic conflicts.

3. Running the simulation

After the data cleaning, correction and preparation, a simulation was executed for the hour from 7 am to 8 am. The number of the road users originally recorded is 1279. After the data processing, the total number of road users is 1168, of which around 92 percent are motorized vehicles and 86 percent are passenger cars. No conspicuously critical situation occurred.

As mentioned before, we have tried to match interrupted trajectories to the network using the lane marking information to determine appropriate traffic demand. This produced SUMO trajectories, that are well different from the original ones, since additional parts of the trajectory had to be invented. These ones will not be used in the subse-

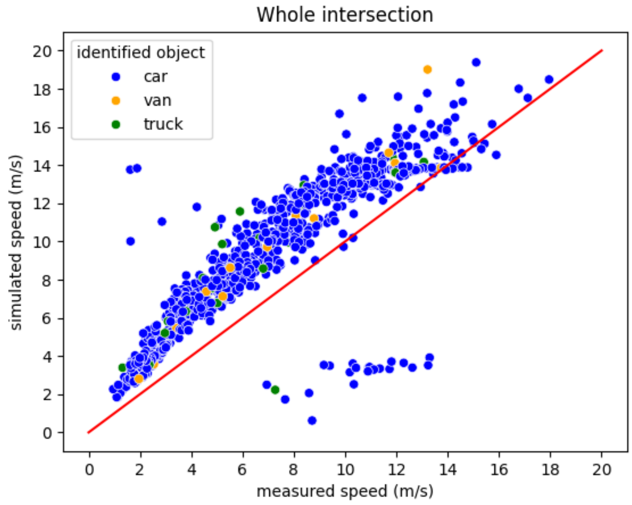


Figure 3. *Comparison of the measured and simulated speed average at the RI.*

quent analysis, only the road users with more complete trajectory sets are considered, which were the result of the first matching work. Figure 3 shows the comparison between the simulated and measured speeds of the motorized vehicles. In general, the simulated speeds are somewhat higher than in reality with the SUMO default setting, and this becomes even clearer when the measured speeds become higher. There are several situations, in which the simulated and measured speeds are opposite, e.g. the measured ones are below than 4 m/s, while the simulated ones are more than 10 m/s. The possible causes for this could be inconsistent synchronization of the traffic signal control and/or different driving decisions during yellow phases. It could also be due to the quality of the input data in relation to departure position and time. Further investigations are needed. In addition, the speed difference between vans, cars and lorries does not seem to be statistically significant. It is therefore questionable whether it is necessary to differentiate between them in the traffic calibration. Further investigations are also required here.

Furthermore, the speed difference were compared for each OD-relation (driving direction). The result in Figure 4 shows the vehicular speed differences and their dispersions for each OD-relation. The average speed difference per OD-relation is between 2 and 3.5 m/s. The smallest speed difference dispersion is found in the direction from north to east. This is probably due to the lower left-turn traffic and the adaptive traffic control for the left-turn traffic. Right-turn traffic shows a greater spread of speed differences in all four directions (E-N, W-S, S-E and N-W). It implies that the right-turn maneuvers in the simulation are smoother than in reality. The respective parking behaviors at the roadside may have to be considered in the simulation.

4. Conclusion

This work has shown that SUMO is able to microscopically reproduce traffic situations using captured trajectory and traffic light data. The default parameter values are currently being used to see to what extent the default setting can reflect reality. The preliminary result with SUMO default setting is quite promising, but also indicates the need for considering other factors in the simulation, e.g. parking and different driving decisions when facing yellow/yellow-red phases and the need for traffic calibration. The coverage of the video cameras at the research intersection is limited. Most of the vehicles

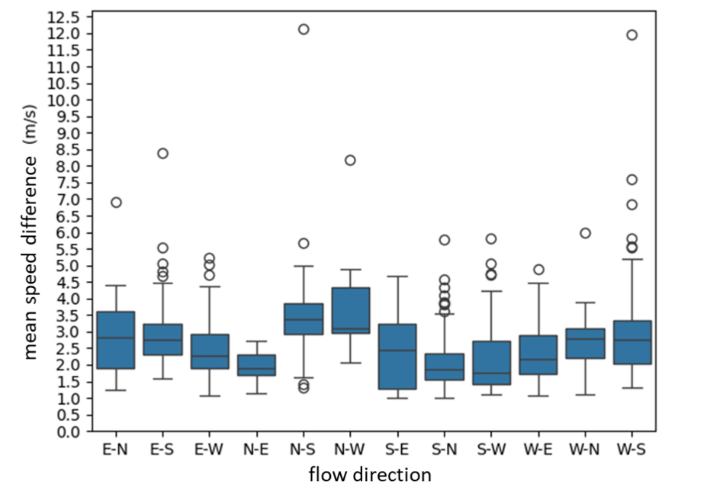


Figure 4. *Comparison of the measured and simulated speed per O-D relation at the RI.*

recorded are already in the lanes with the specified lane markings (right, through and left) and lane changes are rare. The calibration in the next step therefore concentrates on the car-following model parameters.

Several problems were uncovered during data processing, some of which have been addressed in this work. However, further data examination and correction are still required, e.g. in relation to the classification of objects and vehicles traveling at extraordinary speeds, e.g. more than 16 m/s or less than 4 m/s. It should also be worthwhile to check and merge interrupted trajectories if they belong to the same object in order to avoid overestimating traffic flows at the RI.

Moreover, the analysis on the necessity of vehicle classification at the RI in the simulation will be conducted by investigating whether significant differences in speed and acceleration exist between vans, cars, trucks and buses and if so, to what extent. A simulation for the whole day and the corresponding calibration focusing on speed and acceleration for typical traffic conditions, such as peak and off-peak hours, will then be carried out next.

As mentioned above, the signal control logic at the RI is not available. It is then difficult to properly perform scenario analyses with different traffic flows/loadings. Currently, SUMO has two types of adaptive control logic. It is planned to examine the applicability of these two control logic methods with the given dataset.

In addition, the overall data processing and simulation procedure will be further improved, making it not only easier to set up a simulation with any trajectory datasets collected at the RI, but also easier to adapt for use with other trajectory data of similar format.

Data availability statement

The data of the RI are available via Zenodo [8]. The data finally used in the SUMO simulation will be made public later during the runtime of the project KoFeMo[13].

Author contributions

- Conceptualization: Y.-P. Flötteröd, P. Wagner
- Data curation: Y.-P. Flötteröd
- Formal Analysis: Y.-P. Flötteröd
- Investigation: Y.-P. Flötteröd, P. Wagner
- Methodology: Y.-P. Flötteröd, P. Wagner
- Software: Y.-P. Flötteröd
- Validation: Y.-P. Flötteröd, P. Wagner
- Visualization: Y.-P. Flötteröd
- Writing – original draft: Y.-P. Flötteröd, P. Wagner
- Writing – review and editing: Y.-P. Flötteröd, P. Wagner

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the mFUND project KoFeMo, and therefore by the German Ministry of Digitalization and Transport. In addition it used some basic funding of the DLR (German Aerospace Center).

Acknowledgements

This work here profited from discussions from our colleagues responsible for the managing of the RI: Lars Klitzke and Clemens Schicktanz, who are gratefully acknowledged here.

References

[1] J. Treiterer and J. Myers, "The hysteresis phenomenon in traffic flow," *Transportation and traffic theory*, vol. 6, pp. 13–38, 1974.

[2] B. Coifman, L. Li, and W. Xiao, "Resurrecting the lost vehicle trajectories of treiterer and myers with new insights into a controversial hysteresis," *Transportation Research Record*, vol. 2672, no. 20, pp. 25–38, 2018. DOI: 10.1177/0361198118786473.

[3] V. Alexiadis, J. Colyar, J. Halkias, R. Hranac, and G. McHale, "The next generation simulation program," *ITE Journal*, vol. 74, no. 8, pp. 22–26, 2004.

[4] E. Barmpounakis and N. Geroliminis, "On the new era of urban traffic monitoring with massive drone data: The pneuma large-scale field experiment," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 50–71, 2020. DOI: 10.1016/j.trc.2019.11.023.

[5] D. Gloudemans et al., "I-24 motion: An instrument for freeway traffic science," *Transportation Research Part C: Emerging Technologies*, vol. 155, p. 104 311, 2023. DOI: 10.1016/j.trc.2023.104311.

[6] A. Kutsch, M. Margreiter, and K. Bogenberger, "TUMDOT-MUC: Data collection and processing of multimodal trajectories collected by aerial drones," *Data Science for Transportation*, vol. 6, no. 2, p. 15, 2024. DOI: 10.1007/s42421-024-00101-5.

[7] M. Berghaus, S. Lamberty, J. Ehlers, E. Kalló, and M. Oeser, "Vehicle trajectory dataset from drone videos including off-ramp and congested traffic – analysis of data quality, traffic flow, and accident risk," *Communications in Transportation Research*, vol. 4, p. 100 133, 2024. DOI: 10.1016/j.commtr.2024.100133.

[8] C. Schicktanz et al., *DLR Urban Traffic dataset (DLR UT)*, version 1.2.0, Zenodo, 2025. DOI: 10.5281/zenodo.14773161.

[9] P. Alvarez Lopez et al., *Simulation of Urban Mobility (SUMO) (version 1.22.0)*, Feb. 2025. [Online]. Available: https://elib.dlr.de/212503/.

[10] P. A. Lopez et al., "Microscopic traffic simulation using sumo," in *The 21st IEEE International Conference on Intelligent Transportation Systems*, IEEE, 2018. [Online]. Available: https://elib.dlr.de/124092/.

[11] I. of Transportation Systems of DLR. "'research intersection' – a hub for data collection in the field," Accessed: Apr. 1, 2025. [Online]. Available: https://www.dlr.de/en/ts/research-transfer/research-infrastructure/test-areas/acquisition-technology/research-intersection.

[12] VVM. "VVM - Verification Validation Methods," Accessed: Mar. 7, 2025. [Online]. Available: https://www.vvm-projekt.de/en/project.

[13] KoFeMo. "Kombinierte Untersuchung von Feinstaub und Mobilität – KoFeMo," Accessed: Mar. 7, 2025. [Online]. Available: https://bmdv.bund.de/SharedDocs/DE/Artikel/DG/mfund-projekte/kofemo.html.