A Comparison of Urban Mobility Simulation Strategies for the City of São Paulo

Renata Callado Borges

***Abstract*—Abstract**

***Index Terms*—Simulation, Traffic, Multi-agent simulation**

I. Motivation

Simulation of urban traffic is a key component of digital twins of cities, specially megalopolises such as São Paulo. Using simulation strategies based on hydrodynamics or equivalent, or Machine Learning (ML) modeling currently fails to predict accurately traffic for future times beyond a 20 minutes window [3]. There is no equivalent assessment of the accuracy of multi-agent based modeling of urban traffic flow. We intend to fill this gap in research. It is our understanding that digital twins of megalopolises will make use of traffic simulations as an aspect of more complex simulations and services, and agent-based simulations are best suited for such a multidisciplinary work.

II. Previous Works

Medina-Salgado et al [3] presented a compilation of results using multiple software solutions for the simulation of traffic with a window of 20 minutes. Accuracy of predictions is indicated for a few solutions, that range from 71.93% accuracy to 98.70% accuracy. These accuracy results are available only for a subset of the simulators listed by the authors, and they all are dependent on ML based methods. Irawan et al [2] consider that ML based approaches to traffic flow simulation has particular value when combining long-term and short-term random forest models. Many simulators that are agent-based have been proposed, but the literature lacks a comparison of agent-based solutions, lacking in particular the comparison (from an accuracy standpoint) of different strategies for agent-based simulation of urban traffic. We also note a lack of literature integrating ML based traffic modeling and other dimensions of the urban environment that must be integrated in a comprehensive digital twin of a city.

III. Methodology

Current ML based approaches lack explanatory power, since they rely on the fact that neural networks will select weights appropriate to their training data, and these sets of neural network weights are difficult to interpret by humans. Multi-agent modeling, on the other hand, relies on formal mathematical modeling that is executable by computers. The source code for such models, and their fundamental elements (namely, agents) are much more human-readable than a set of neural network weights, and therefore advances in modeling traffic flow using such an approach can go hand-in-hand with advances in scientific knowledge about this phenomena.

Accuracy of prediction is crucial for this approach, since we can create abstract models of arbitrary complexity and size, but their usefulness as elements in a larger system (a digital twin of a city) is impacted by their accuracy. Our goal is not to model an abstract city, but to model existing cities.

Abundant, standardized data about cities' structures and vehicular data is available by the OpenStreetMap (OSM) project[1]. Being a Free Software project, using widely adopted software and methodologies, OSM was our main source of data.

Historically, the city of São Paulo has been known for its extensive traffic jams. The Companhia de Engenharia de Tráfego[2] (CET) publishes data about the city and its jams, including historical data. We have used this data as an estimator for the real historical values of road occupancy in the city of São Paulo.

The GAMA platform [1] is a Free Software project that provides an agent-based modeling framework, and its documentation includes a road traffic simulation that was used as a basis for some of the simulators presented here.

The simulators that we have developed yield results that can be compared with the CET data, in the following manner: the CET data indicates, for sparsely indicated times, the extent of main roads' congestion (in meters). We infer from this data, the percentage of the total road extent that is congested for these specific times. Since we do not have realworld data for all times of day, we ignore the data points in the simulation that do not match these data points in the realworld data. The CET data has data points with half an hour granularity; so we register, for each elapsed half hour of the simulations, the number of agents that have traveled through the road indicated by the data point. For this paper, we selected the Avenida Mercúrio[3], because it is a chokepoint in the city of São Paulo center, it is often congested, and therefore has a lot of CET data. Also, its extent is small enough that we can easily obtain OSM data using their web interface.

The "relative difference" between a simulation result and the CET data consists in the average difference, in percentage points, between the realword occupancy and the simulated occupancy.

We compare the fitness of the simulations by comparing the hourly profile of the normalized average for a range of

[1]https://www.openstreetmap.org

[2]Possible translation: Traffic Engineering Company.

[3]Possible translation: Hermes Avenue.

TABLE I
COMPARISON OF BASELINE SIMULATION AND CET DATA.

Hora	CET (m)	Baseline v1.0 (#agents)
2010-01-04 14:30:00		0
2010-01-04 15:30:00		-16
2010-01-04 16:30:00		16500
2010-01-04 17:30:00	420.00	21036
2010-01-04 18:30:00	840.00	22998
2010-01-04 19:30:00	840.00	22163
2010-01-04 20:30:00	840.00	21717
2010-01-04 21:30:00	840.00	6421
2010-01-04 22:00:00		369

TABLE II
COMPARISON OF SIMULATIONS AND CET DATA FOR 2010/JAN AND 2017/OCT.

Simulator	Date range	Relative difference (pp)
Baseline v1.0	2010/01	45.9
Baseline v1.0	2017/10	49.5
Lunchtime v1.0	2010/01	25.4
Lunchtime v1.0	2017/10	39.3

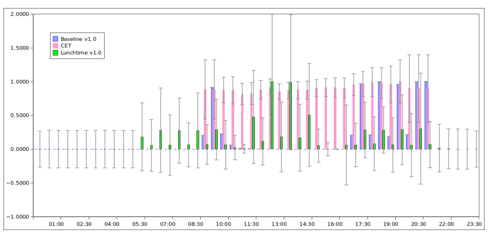

Fig. 1. Hourly averages for 1 year of CET data and 1 month of simulations.

a year of CET data and one month of simulated data. These ranges yield broadly comparable values, in regards to their uncertainties' ranges.

IV. RESULTS

We have developed two simulators, one with the least possible complexity that yields quantitative results comparable to CET data. We have named this simulator "Baseline" and versioned it so that modification to its source can be assessed properly. Table I shows the results for the date of January 4th 2010. This simulator models a pendular movement of agents: during nighttime they remain in residential buildings, during office hours they remain in industrial buildings, and it is their movement from one place to another that yields traffic.

We have also developed a slightly more complex simulator, that extends the Baseline simulator by adding lunchtime traffic. This added simulation strategy corresponds to the notion that traffic around noon is due to movement of workers leaving their places of work and searching for restaurants.

For both simulators, we have produced results for the range of two randomly selected months, January 2010 and October 2017, and compared their results with CET data and among themselves. The data is on Table II.

We proceeded to compare the simulated data for a whole month of simulated data (for both simulations) and a year of CET data. The comparison can be summarized by Figure 1.

After consideration of the previous results, we developed a new version of Baseline that generates traffic more evenly throughout the day, and compared the normalized hourly average occupancy with CET data, as seen on Figure 2.

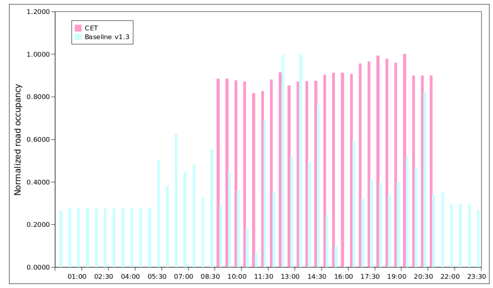

Fig. 2. Hourly averages for 1 year of CET data and 1 month of simulations.

V. DISCUSSION

From Table II, we can infer that the Lunchtime simulator predicts traffic for Av. Mercúrio with an discrepancy to real-word data that is at least 10 percentage points lower than the Baseline simulator.

The comparison of CET data and simulations in Figure 1 shows that the simulations fail to capture the average traffic of this highly congested road; in particular, the simulation models seem to show traffic concentrated around rush hours, while the real averaged data shows a more evenly distributed traffic throughout the working day.

VI. FUTURE WORK

There are many possible improvements for agent-based simulators that would be of impact in a digital twin system (e.g. semaphores, subways, buses, bicicles, car variety, real state information). Another possible improvement is in the simulation structure, which could include bounding box frontier events, such as agents entering or leaving the simulated region.

VII. CONCLUSION

The results in Table II indicate that the strategy of adding a goal for agents to move around noon better models realworld traffic.

ACKNOWLEDGMENTS

I would like to thank my advisor, Prof. Kelly Rosa Braghetto, and Prof. Daniel Cordeiro and the PhD candidate Francisco Wallison Carlos Rocha from USP, without whose help I would not have been able to utilize InterSCSimulator.

REFERENCES

[1] Taillandier, P., Gaudou, B., Grignard, A.,Huynh, Q.-N., Marilleau, N., P. Caillou, P., Philippon, D., and Drogoul, A. (2019). *Building, composing and experimenting complex spatial models with the GAMA platform*. Geoinformatica, (2019), 23 (2), pp. 299-322, [doi:10.1007/s10707-018-00339-6]

[2] Irawan, K., Yusuf, R. and Prihatmanto, A. S.. *A Survey on Traffic Flow Prediction Methods*. 6th International Conference on Interactive Digital Media. 2020. [doi:10.1109/ICIDM51048.2020.9339675].

[3] Medina-Salgado, B., Sánchez-DelaCruz, E., Pozos-Parra, P., and Sierra J. E.. *Urban traffic flow prediction techniques: a review*. Sustainable Computing: Informatics and Systems 35. 2022. https://doi.org/10.1016/j.suscom.2022.100739