RSS 2022
8th Road Safety & Simulation International Conference
Road Safety and Digitalization
08-10 June 2022 • Athens, Greece

Towards ML-based real-time traffic simulation

Ferran Torrent-Fontbona[1], Javier Frade[2], Jordi Casas[3]

Aimsun SLU, Ronda Universitat 22B Barcelona, ferran.torrent@aimsun.com
Aimsun SLU, Ronda Universitat 22B Barcelona, javier.fernadez@aimsun.com
Aimsun SLU, Ronda Universitat 22B Barcelona, Jordi.casas@aimsun.com

Abstract

Real-time traffic simulation systems are of great relevance for forecasting traffic and operating traffic infrastructure, but they struggle when simulating huge networks, with a latency of a few minutes; not only do these simulation systems require accurate modelling of traffic demand, which takes up significant time and resources, but the continuous adjustments to rapidly changing traffic demand become unaffordable and often lead to a drop in the accuracy of the simulations. Conversely, machine learning models offer the ability, once trained, to provide quick predictions, and to progressively adapt themselves automatically by using online learning algorithms.

This paper proposes a data-driven approach to building a machine learning model with the capacity to estimate the traffic flow in unobservable sections in real time by using the traffic flow of another set of observable sections in real time. The proposed approach is tested with three real scenarios in Sydney, Bergen and Wiesbaden, and compared against a simulation-based approach. Results demonstrate that the proposed approach outperforms the simulation-based approach because it is better able to generalize when faced with changes in traffic demand.

Keywords: Machine learning; traffic simulation, network clustering, Aimsun Live.

[1] * Corresponding author. Tel.: +34-933-171-693;
E-mail address: ferran.torrent@aimsun.com

1 Introduction

Real-time traffic simulation systems are of great relevance for forecasting traffic and operating traffic infrastructure [1]. These simulation systems are highly complex and require huge computational effort, especially for large-scale models that require dynamic simulations to represent the evolution of congestion over time. The computational requirements depend on the modelling resolution: microscopic models require highly granular inputs with an explicit representation of the vehicles, while macroscopic and mesoscopic models have simplified network representation and vehicle dynamics, although super large-scale networks are still challenging.

Machine learning (ML) models are very efficient to build from data, and once built, to provide predictions from input data. Offering real-time traffic predictions is not a challenge, especially when the throughput frequency is up to about one minute. However, ML models present serious difficulties when they must provide predictions for a target variable that is not observed [2]; for example, extending traffic predictions to the whole network from a set of observation points is not possible with ML alone, while simulation systems per se can offer this as they emulate traffic behavior and network characteristics to infer what happens in a section that has been never observed [2]. Causal inference or causal modelling may solve this ML challenge [3], and the literature exhibits some approaches that make use of causal inference theory and graph neural networks for timeseries forecasting [4]. However, for normal levels of traffic observability, which, even in the best cases, only cover a small portion of the network, causal inference is currently unfeasible.

On the other hand, traffic simulation systems require a great effort in modelling traffic demand and network so they can provide causal inference of what will happen, depending on the estimated demand and the state of the supply [2]. However, modelling the demand requires weeks of work and usually only the main demand patterns are modelled, such as typical workdays, weekends, and holidays, which means that demand is modelled by and in the resolution defined by the set of demand patterns. Any change in the demand that goes beyond these patterns requires the generation of new patterns to ensure that simulation outputs mirror what happens in real life; if demand shifts are frequent, it may become unfeasible to model new demand patterns.

This paper analyzes the ability of a data-driven approach to estimate whole network traffic state from real-time observations of a subset of observable sections. The final objective of this approach would be to support traffic simulations by incorporating such traffic state estimation as the initial conditions of the simulation.

2 Methodology

This paper proposes a data-driven approach for estimating traffic flow in sections without real-time sensor data, i.e. unobservable sections, using data from sections with real-time sensor data, i.e. observable sections. For doing so, we assume that offline data is available for either observable or unobservable sections. Offline data can be real or synthetic, that is, from simulations. Therefore, the approach exploits section correlations to estimate the flow on unobservable sections using a data-driven methodology.

Section-flow cross-correlation is usually high, even between distant sections. The main problem that faces this approach is overfitting, especially in large networks with hundreds or thousands of sensors. To avoid this problem, it is proposed to constrain the model to consider relationships with neighboring sections.

Figure 1 shows the training workflow of the data-driven approach. In order to force using neighboring relationships, using offline data from observable and unobservable sections and the graph of the network, a clustering algorithm divides the network in zones. In this paper we propose using affinity propagation [5] with a combination of personalized PageRank [6] and flow correlation as affinity function. Personalized PageRank offers the relative importance of graph nodes according to graph structure emulating random walks. This combined with flow correlation between sections offer to the affinity propagation clustering algorithm an affinity metric that considers spatial affinity, according to graph structure, and flow correlation affinity. Other affinity (or similarity) functions and clustering algorithms ca be used to divide the network in zones as long as all zones contain at least one observable section. Therefore, every zone must have observable sections and, optionally, unobservable sections.

Given the division of the network in zones, then a ML model is trained so it can estimate the flow in unobservable sections using data of observable sections in the same zone. The ML model can be trained to minimize the error between the estimated flow and the ground truth. Note, that it is assumed that offline data is available for observable and unobservable sections. As depicted in Figure 1, we propose a linear regression model with L2 regularization (ridge model) for each zone.

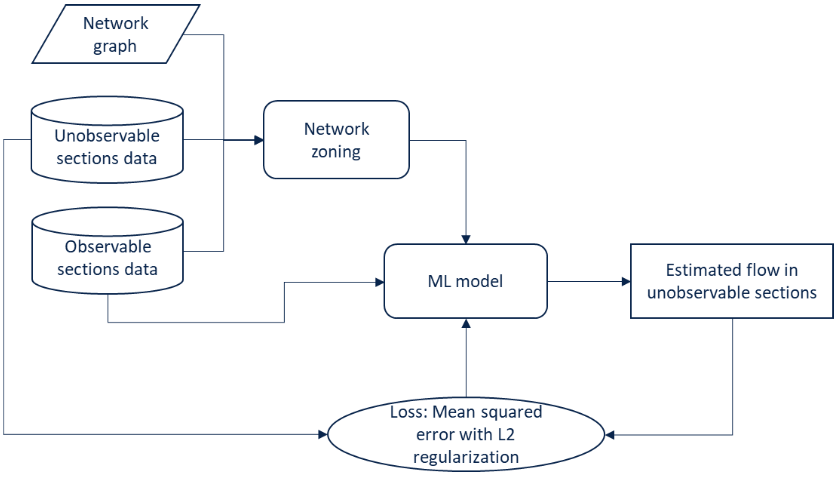
Network graph
Unobservable sections data
Observable sections data
Network zoning
ML model
Estimated flow in unobservable sections
Loss: Mean squared error with L2 regularization

Figure 1. Training workflow.

Once the model is trained, it can be used to estimate the flow in unobservable sections. Figure 2 illustrates the prediction workflow.

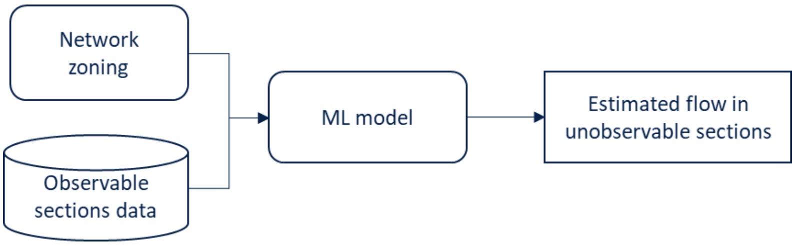

Figure 2. Prediction workflow

3 Analysis and Results

This section presents the results of the proposed methodology with simulated and real datasets from three different cities: Bergen, Wiesbaden and Sydney (city center between Silverwater Bridge and Sydney Harbour Bridge). Figure 3**,** Figure 4 and Figure 5 show the networks for the three cities and additional features are described in **Table 1**.

Table 1. Description of the cities

	Bergen	Wiesbaden	Sydney
Linear km	1111	1446	17465
Surface (km^2)	51.59	63.31	63.31
Number of intersections	4897	4552	1274

Simulations were performed using Aimsun Next [7] traffic modeling software, which forms the core of the Aimsun Live decision support solution. The data used is described in Table 2**:** the period of training and validation data; the number of traffic patterns (24-hour-long traffic patterns that describe recurrent traffic behavior, according to the training set); the number of observable sections (in the simulated dataset and in the real datasets), and the number of unobservable sections (in the simulated dataset and in the real datasets). In the simulated dataset, observable sections are those with real loop detectors, and unobservable sections are those without real loop detectors. In the real datasets, observable and unobservable sections have been chosen randomly from real loop detectors with data in the training and validation datasets. Moreover, training and validation dates for the real dataset have been chosen to follow the real example of three Aimsun Live projects. For the three of them, we used as training data the available data in the model calibration phase. Testing data was gathered with real-time the real-time data pre-processing module in Aimsun Live. Testing dates were chosen to be the same for the three scenarios.

The simulation datasets are in 15-minute intervals. Bergen and Sydney real data are in 5-minute intervals but smoothed with a 15-minute rectangular sliding window. Wiesbaden real data are in 15-minute intervals. Despite simulated traffic patterns are in 15-minute intervals, the simulation predictions of the validation period are in the same sampling frequency than real data.

The three scenarios used the same data-driven methodology and the same parameters, however, each scenario has its personalized model, trained only with data of its city.

Table 2: Real data description

City	Training dataset	Validation dataset	Number of patterns (training set)	Observable sections	Unobservable sections
Bergen	2018, 2019 and March 2021	Sep and Oct 2021	8	110	3732
				54	34
Wiesbaden	From Aug 2020 to June 2021	Sep and Oct 2021	12	203	4403
				186	81
Sydney	From June 2020 to Dec 2020 and Aug 2021	Sep and Oct 2021	8	-	-
				2061	514

The experiment has two parts:

- Simulation emulation: this consists of training a ML model with the simulated traffic patterns to use only data from observable sections (those with loop detectors) to estimate the flow in unobservable sections. Traffic patterns were generated using training dataset described in Table 2. Observable sections are the ones with real loop detectors. The analysis uses 5-fold cross-validation, meaning that in every fold the model is trained 80% of patterns and validated with the remaining 20%. The results related to this experimentation are the average throughout the 5 folds. The traffic patterns represent the main traffic patterns of the offline dataset, so they compact the whole offline real dataset into a set of N synthetic days, N being the number of patterns.
- Real-time unobserved section estimation: this consists of training a ML model with offline real data in such a way that from a set of (randomly chosen) observable sections it can estimate the flow in unobserved sections. The model is trained with the real offline dataset (training dataset in Table 2) and validated with the online dataset (validation set in Table 2). Results are compared with the predictions obtained with simulation using the right traffic pattern for each day.

The %GEH<5 and the RMSE have been chosen as error metrics to give and absolute and relative notion of the error. MAPE has not been chosen due to its bias with low flows that occur at night or early morning. Note that in this experimentation, error in unobservable sections can be calculated because the ground truth is available.

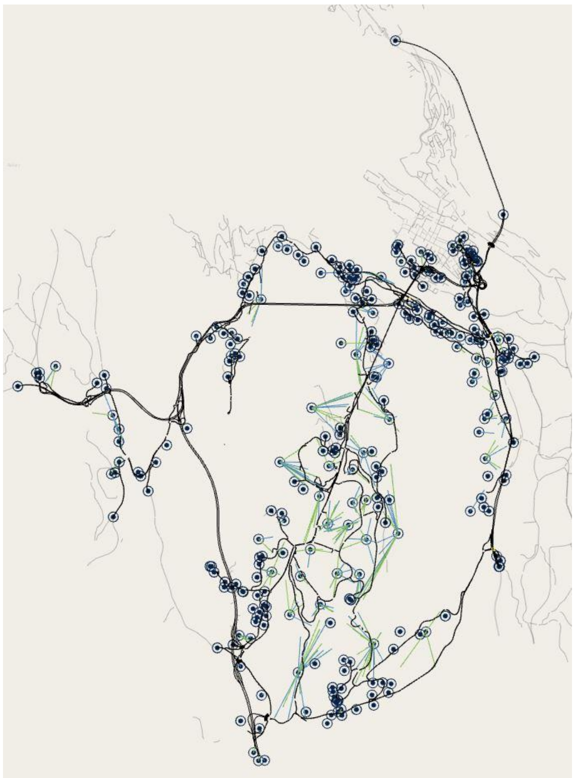

Figure 3. Aimsun model of Bergen network

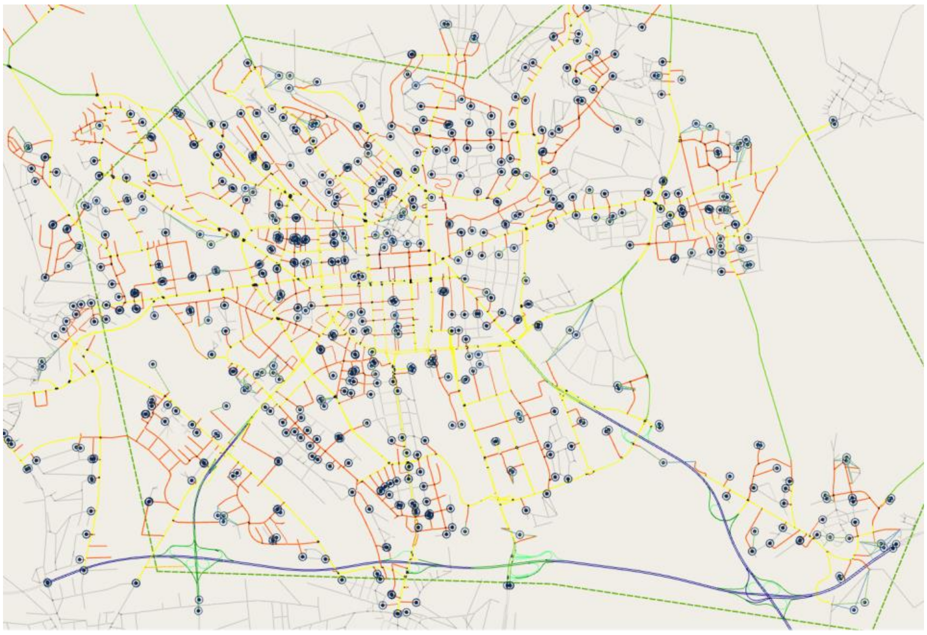

Figure 4. Aimsun Model of Wiesbaden network

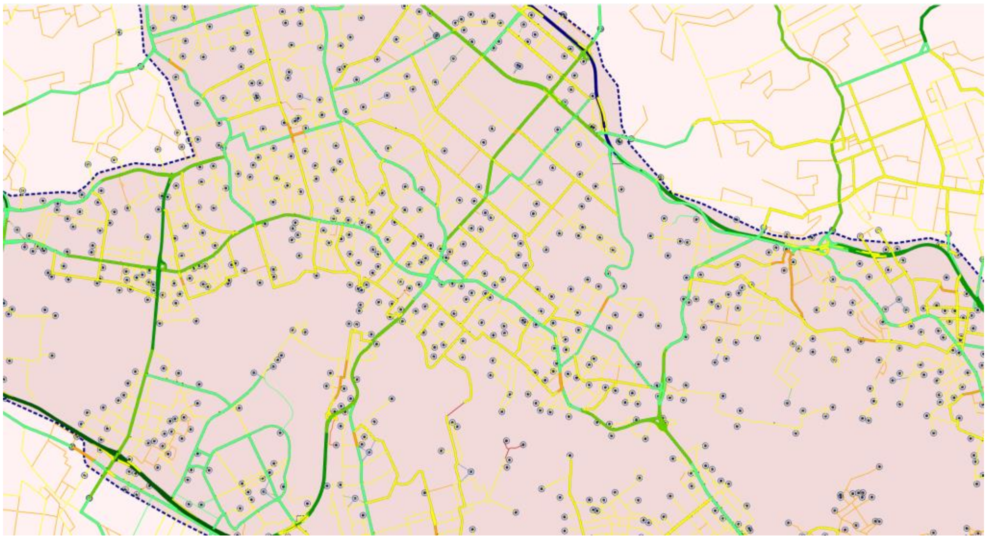

Figure 5. Aimsun model of Sydney network

3.1 Simulation emulation

Figure 6 and Figure 7 show the accuracy, in %GEH<5 and root mean square error (RMSE), of estimating the simulated flow in unobserved sections (see Table 2) by training a ML model for which the input is the measured flow in observable sections. The boxplots represent:

- The upmost horizontal line (whisker): 3rd quartile plus 1.5 times the interquartile range $Q3 + 1.5IQR$
- The box represents the 1st and 3rd quartiles
- The downmost horizontal line (whisker): 1st quartile minus 1.5 times the interquartile range $Q3 - 1.5IQR$
- The horizontal orange line: median
- Green triangle: average

Results show the model's impressive ability to emulate simulated flow behavior even when it is trained with a set of simulated traffic patterns and tested with other traffic patterns. To show the variability of these traffic patterns, Figure 6 and Figure 7 also show the accuracy of estimating the flow using the best possible traffic pattern available in the training set. Traffic patterns must be understood as types of days e.g., workdays, weekends, or holidays. Therefore, the results show the ability of the proposed methodology to emulate simulated traffic patterns, under normal conditions (without incidents, roadworks, etc.), and under unseen traffic patterns.

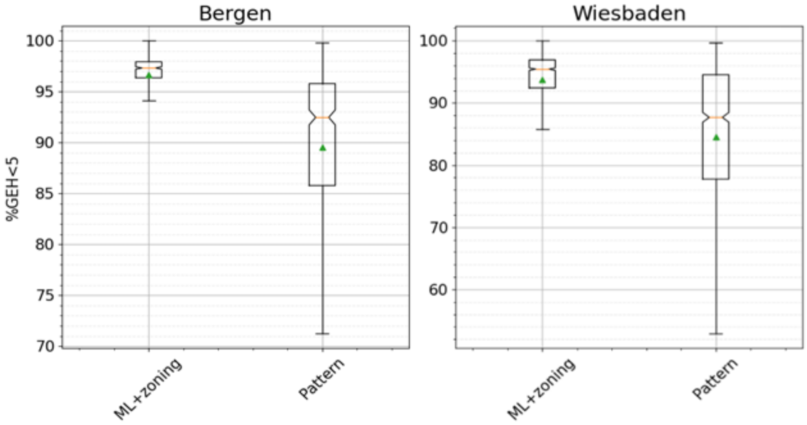

Figure 6. %GEH<5 obtained in unobservable sections comparing simulation outputs with the proposed methodology (ML + zoning) and a simulated pattern in the training set.

It is also remarkable that from a small subset of observable sections (110 in Bergen and 203 in Wiesbaden), the proposed approach can estimate the flow in a number of sections 20-30 times greater. This ability is due to the fact that small cities behave very similarly in terms of demand patterns. In other words, city-wide traffic state in normal supply conditions can be estimated from measurements of a small subset of sections, without the need for full-network observability.

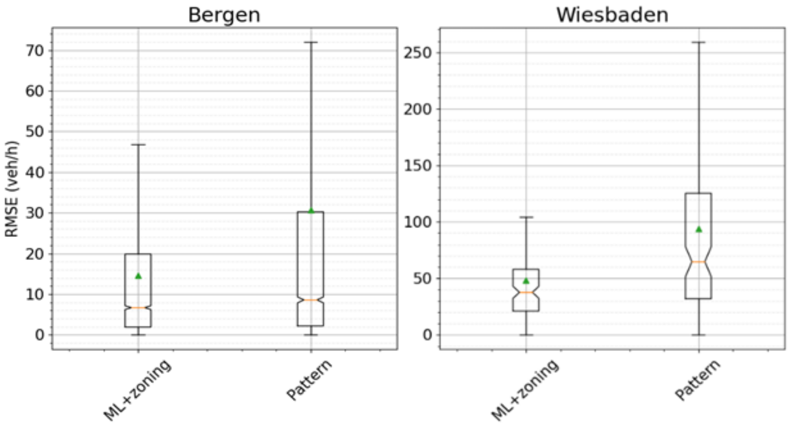

Figure 7. Root mean square error obtained in unobservable sections comparing simulation outputs with the proposed methodology (ML + zoning) and a simulated pattern in the training set.

3.2 Real-time unobserved section estimation

The previous section proved the ability of a ML model to emulate simulation under normal supply conditions, i.e. normal network conditions. But which system better emulates reality? Figure 8 and Figure 9 show the accuracy, in %GEH<5 and RMSE, of the proposed model trained with the training set described in Table 2 and a simulation model with traffic patterns (traffic demand) described in Table 2**.** Both approaches are tested with the validation dataset also described in Table 2. Note that, the proposed approach is limited to receiving data only from observable sections. Similarly, traffic demand selection in the simulation model also uses the subset of observable sections to choose the demand to simulate.

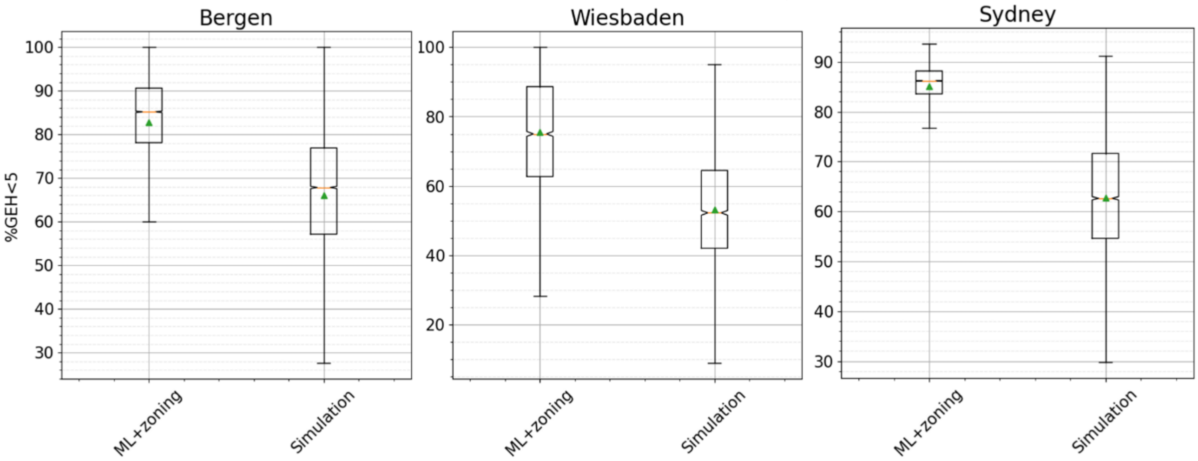

Figure 8. %GEH<5 obtained in unobservable sections comparing real data with the proposed model and simulation.

The results show that the proposed approach achieves better accuracy in estimating the flow in unobserved sections than simulation. Moreover, the variance of the boxplots is smaller for the proposed approach, meaning that it achieves better accuracy consistently throughout all (or most) of the unobserved sections. Simulation results show that modelled traffic demand patterns cannot emulate traffic conditions in the validation period. Thus, there has been a traffic demand shift that requires the calibration of further patterns. Indeed, this is due to mobility shifts related to COVID-19 since demand patterns were generated from data with different traffic demand conditions.

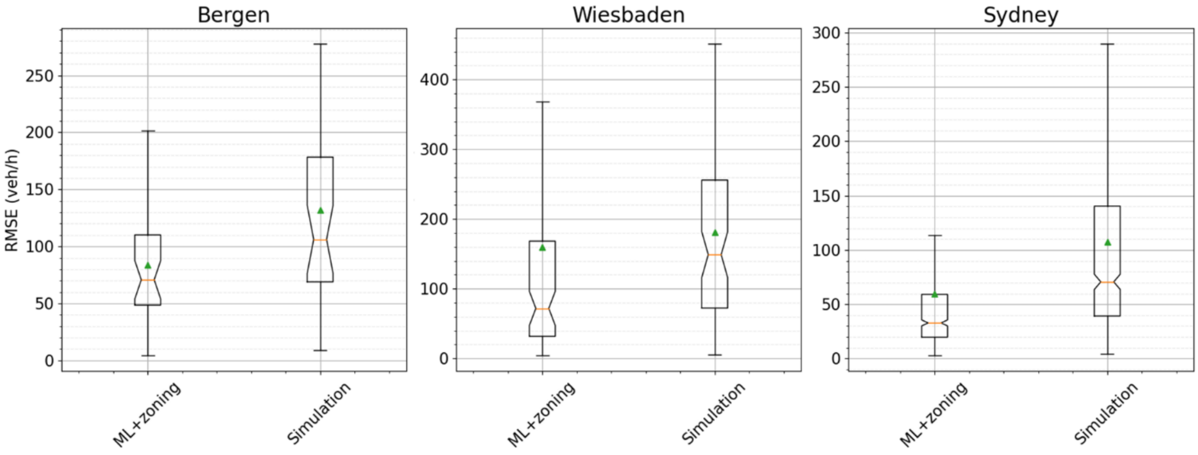

Figure 9. Root mean square error obtained in unobservable sections comparing real data with the proposed model and simulation.

4 Discussion

The results presented in the previous section show the ability of the proposed approach to emulate the simulation of different traffic pattern demands under normal supply conditions, using only data from a reduced subset of observable sections.

Moreover, the proposed approach outperforms the simulation model when estimating the real measured flow in the set of unobservable sections. Therefore, the results show the superior accuracy of the proposed model in making predictions under new traffic demand conditions compared to using the simulation of previous traffic demand patterns.

The conditions of the training and validation sets strengthen this affirmation. The Bergen training set comprises 2018, 2019 and March 2021, so the traffic demand patterns used by the simulation cover 24 months before the COVID-19 outbreak and mobility restrictions, and 1 month under COVID-19 mobility restrictions. However, the validation set mobility context is different as there are no mobility restrictions, the vaccination rate is high, and we can think of it as a middle ground between the pre-COVID period and March 2021.

The Sydney training set comprises the second half of 2020 and August 2021, so the context is one of partial mobility restrictions and severe lockdowns like the one in August 2021. On the other hand, the context for the validation set is one of progressive normalization of mobility. The current state is between the demand patterns of 2020 and those of August 2021.

For Bergen and Sydney, the proposed approach greatly outperforms simulation predictions, meaning that it better fits new demand contexts. Note that traffic simulation systems require accurate modelling of the demand as input; they are also rigid to the input demand unless a dynamic demand adjustment algorithm adjusts the demand to the observed traffic state.

Wiesbaden is different: the training set goes from August 2020 to May to June 2021. But at the end of June 2021, an important bridge located at the southern border of the model was closed. This bridge closure represents a mix of a supply change and demand change because of its border location. Moreover, traffic demand also suffered a small shift related to COVID-19. The approach taken for Wiesbaden achieves overall better accuracy than the simulation, but due to the bridge closure and its impact on the relationships between sections, the variability of the accuracy is greater than in Bergen and Sydney. Indeed, average RMSE is similar to that obtained by the simulation.

Thus, the results demonstrate the ability of the proposed approach in estimating the flow in unobservable sections and its robustness in front of demand changes. However, the results also glimpse the weakness of the proposed approach, as it is a data-driven approach, in front of supply changes. Nevertheless, accurate insights of traffic state are a key requirement for other downstream tasks such as traffic management or road safety analysis.

5 Conclusions

This paper presents a data-driven approach for estimating traffic flow in unobservable sections from another subset of observable sections. The results show its ability to emulate different simulated traffic demands. Moreover, the approach has been tested with three real datasets and compared with predictions obtained with Aimsun Next traffic modeling software. These results also show that the proposed data-driven approach outperforms simulation predictions under new traffic demand patterns. Therefore, the estimated traffic flow could be a useful way to enrich or update initial conditions for simulation. Future work could analyze the performance of the proposed approach when trained with the simulation of many different demand patterns to see how it generalizes to new demand contexts.

The proposed approach takes advantage of the network graph to limit the sections that can be used to estimate the flow of unobservable sections to regularize the trained model. However, the proposed approach is based on the correlations between sections and does not model causal relationships, which means it can't make predictions under supply changes that significantly change the local relationships between sections. However, this does serve to emphasize the need for deeper integration of data-driven approaches in simulation tools.

Considering the overall results, the proposed approach opens the door to a data-driven system that returns full-network traffic state estimation under recurrent supply conditions from partial observability of the network with sensors such as loop detectors.

Acknowledgment

Thanks to Aimsun Live team for the support and material for conducting this research.

References

1. Pell, A., Meingast, A., & Schauer, O. (2017). Trends in Real-time Traffic Simulation. Transportation Research Procedia, 25, 1477–1484. https://doi.org/10.1016/j.trpro.2017.05.175
2. Shafiei, S., Mihăiţă, A. S., Nguyen, H., & Cai, C. (2021). Integrating data-driven and simulation models to predict traffic state affected by road incidents. Transportation Letters, (July). https://doi.org/10.1080/19427867.2021.1916284
3. Hernán, M., J.M. Robins, Causal Inference, 2010.
4. Xu, H., Y. Huang, Z. Duan., X. Wang, J. Feng, and P. Song, Multivariate Time Series Forecasting with Transfer Entropy Graph. arXiv preprint arXiv:2005.01185, 2020
5. Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. Science (New York, N.Y.), 315(5814), 972–976. https://doi.org/10.1126/science.1136800
6. Bahmani, B., Chakrabarti, K., & Xin, D. (2011). Fast personalized PageRank on MapReduce. Proceedings of the ACM SIGMOD International Conference on Management of Data, 973–984. https://doi.org/10.1145/1989323.1989425
7. Aimsun, Aimsun Next 20 User´s Manual. Aimsun Next Version 20.0.2, 2020