Exploiting deep learning and traffic models for freeway traffic estimation

Conference Paper

Author(s):
Genser, Alexander; Makridis, Michail; Kouvelas, Anastasios

Publication date:
2023

Permanent link:
https://doi.org/10.3929/ethz-b-000584466

Originally published in:
Verkehrstelematik 9

Exploiting deep learning and traffic models for freeway traffic estimation

Alexander Genser*[1], Michail A. Makridis*[1], and Anastasios Kouvelas[1]

[1]Traffic Engineering Group, Institute of Transport Planning and Systems, Swiss Federal Institute of Technology (ETH) Zurich, Switzerland

SHORT SUMMARY

Emerging sensors and intelligent traffic technologies provide extensive data sets in a traffic network. However, realizing the full potential of such data sets for a unique representation of real-world states is challenging due to data accuracy, noise, and temporal-spatial resolution. Data assimilation is a known group of methodological approaches that exploit physics-informed traffic models and data observations to perform short-term predictions of the traffic state in freeway environments. At the same time, neural networks capture high non-linearities, similar to those presented in traffic networks. Despite numerous works applying different variants of Kalman filters, the possibility of traffic state estimation with deep-learning-based methodologies is only partially explored in the literature. We present a deep-learning modeling approach to perform traffic state estimation on large freeway networks. The proposed framework is trained on local observations from static and moving sensors and identifies differences between well-trusted data and model outputs. The detected patterns are then used throughout the network, even where there are no available observations to estimate fundamental traffic quantities. The preliminary results of the work highlight the potential of deep learning for traffic state estimation.

Keywords: Traffic state, Traffic prediction, Traffic models, Deep learning, Data assimilation.

1 Introduction

Research interest in understanding and estimating traffic conditions on highway networks has been vivid for over 50 years. Traffic estimation methods aim to derive traffic variables based on modeling, observation, or both. In the era of intelligent transportation systems, efficient traffic monitoring can help develop robust traffic control schemes to facilitate mobility, reduce congestion and emissions, and ensure safety. In addition, the abundance of portable, accurate, and low-cost sensors at the infrastructure, vehicle, or people level is expected to generate large amounts of data that can be used for traffic estimation solutions.

In the literature, methodologies on the topic adopt different approaches, spanning from time-series analysis, see (Ahmed & Cook, 1979), data assimilation with Kalman filtering techniques, see (Mihaylova, Boel, & Hegyi, 2007; Wang & Papageorgiou, 2005; Risso, Bhouri, Rubiales, & Lotito, 2020), fuzzy logic as in (Stathopoulos, Karlaftis, & Dimitriou, 2010), neural networks or multiple linear regression models as in (X. Ma, Tao, Wang, Yu, & Wang, 2015) and (Genser, Hautle, Makridis, & Kouvelas, 2022), statistical approaches, see (Crawford, Watling, & Connors, 2017), and others. Additionally, some studies exploit parametric traffic models such as Cell Transmission Model (CTM) (Daganzo, 1994), and METANET (Papageorgiou, Blosseville, & Hadj-Salem, 1990), non-parametric (X. Ma et al., 2015), or hybrid ones (Boto-Giralda et al., 2010). Finally, several data-driven and machine learning-based frameworks aim to capture data correlations between points that are not directly linked in time and space, see (Treiber & Kesting, 2012; Lint & Hoogendoorn, 2010; T. Ma, Antoniou, & Toledo, 2020).

Parametric traffic models such as CTM or METANET are computationally inexpensive and capture the traffic dynamics in space and time to a reasonable extent. However, they need significant efforts for proper calibration, and their performance drops when applied on large networks due to highly complex dynamics and implementation difficulties.

Recently, the interest in data-driven traffic estimation methodologies has been increasing. Data observations provide accurate estimates if adequately used. At the same time, advanced deep learning strategies can capture higher non-linear traffic dynamics than traffic models. However, a significant problem to date is data scarcity. Observations are usually available at specific parts of the networks, with low penetration rates, and their reporting frequency is often inconsistent.

For the last two decades, a large number of data assimilation studies have most commonly exploited the flexibility of Kalman-based solutions and the low computational complexity of traffic models to fuse heterogeneous data from models and data sources.

The data assimilation family of developments inspires the basic idea of the proposed work. It aims to exploit the best of both worlds, models, and data by developing a physics-informed deep learning framework for traffic estimation on large freeway networks. Traditional traffic models capture well the traffic dynamics in space and time. On the other hand, observations provide a significantly better representation of the traffic dynamics, but their estimates are confined to space at the measurement place. At the same time, a deep learning framework can, in theory, study the discrepancies between models and observation and identify error patterns between the two worlds. The assumption in this idea is that error discrepancies between models and data can be identified in a large number of patterns over a network. Therefore, a systematic study of these patterns can help us provide reasonable estimates locally, i.e., with data availability and globally in the network.

The proposed study uses the macroscopic network discretization provided by traffic models and studies triplets of neighboring segments. The idea of using triplets is to capture the error correlations between adjacent parts in the network. A framework that implements a bidirectional Long-Short-Term-Memory (biLSTM) neural network is proposed. The model is trained on network segments with data availability of traffic model estimates as inputs. Consequently, the model can be assessed along the network where data is unavailable. Preliminary results of a case study that utilizes a stretch of the Antwerp ring-road network seem promising compared to data assimilation approaches. However, further investigations are needed to conclude if such a framework can elevate the performance of traditional Kalman-based implementations on a large-scale network level.

2 Methodology

The presented framework aims to estimate traffic states on a freeway network with on- and off-ramps by combining the strengths of traditional traffic models and deep learning techniques. Our approach, therefore, utilizes traffic state estimates from traffic models and measurement data that, e.g., various established and fixed installed sensors or moving Connected and Automated Vehicles (CAV) provide. Figure 1 depicts the proposed framework with its components. As a starting point for our method, traffic states from traditional traffic models serve as input. The utilized model(s) (e.g., CTM, METANET) are members of a model collection $\mathcal{C}$. The set $\mathcal{C}$ can hold estimates from either one or multiple sources (e.g., Kalman Filters, Particle Filters). The generic approach allows for incorporating different traffic model estimate combinations to assess and tune performance.

The provided estimates are denoted as the traffic density $\rho_i(k)$ and space mean speed $v_i(k)$ of a network segment i at discretized time k. A network consists of $\mathcal{I}$ segments and $i \in \mathcal{I}$. $r_i(k)$ denotes the inflow at an on-ramp connected to segment i. $s_i(k)$ denotes the split of traffic flow that exits the freeway via an off-ramp. $s_i(k)$ is based on an exiting ratio denoted as $\beta_i(k)$. Due to the prior estimation of traffic density and space mean speed with a traffic model, the quantities

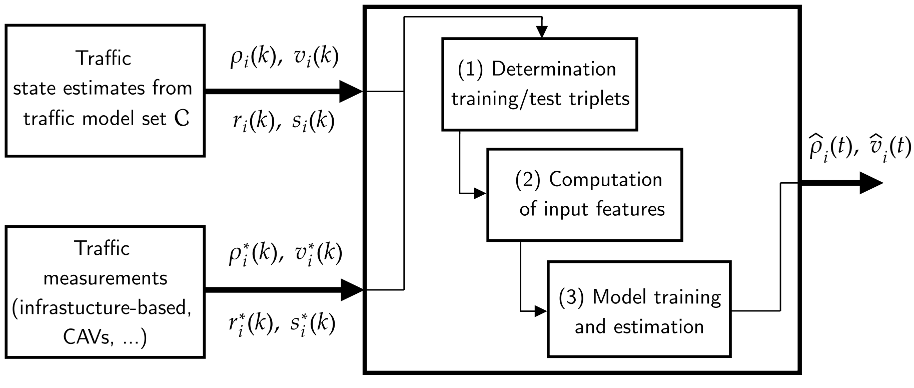


Figure 1: Framework description to determine traffic state estimates.

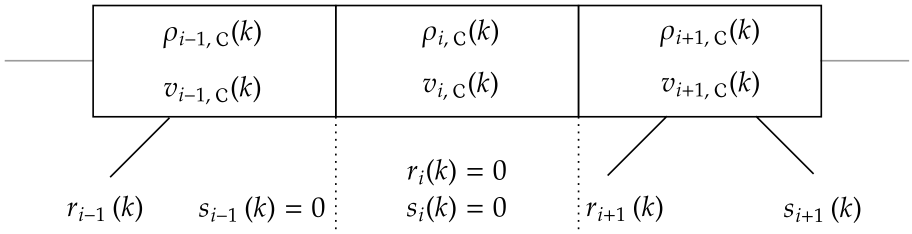


Figure 2: Triplet design with corresponding features that represent traffic state estimates from a traffic model $C \in \mathcal{C}$.

serve in Figure 1 as an input to the estimation framework proposed in the following.

The second set of inputs in Figure 1 constitute measurements from sparsely available sensor data in the network. Such inputs can either be determined from sensors implemented at a fixed location (e.g., loop detectors, Bluetooth sensors, cameras) or, in a Vehicle-to-Infrastructure (V2I) environment, from moving CAVs that broadcast the traffic-related information. In this work, the measurements for traffic density $\rho_j^*(k)$ and space mean speed $v_j^*(k)$ are already aggregated for a segment i. Also, measurements for inflow at an on-ramp and the split of exiting flow, $r_i^*(k)$ and $s_i^*(k)$, are available. The asterisk denotes measurement quantities.

The two different sets of inputs (traffic estimates from traditional traffic models and measurement data) are then utilized in block (1) (Figure 1). To determine an estimation model, we introduce training and testing triplets. The concept of triplets is defined as three contiguous segments in a freeway network denoted with the indices $i-1$, i, $i+1$. The idea of using triplets is to capture the error correlations between neighboring parts in the network; i.e., the spatial component of traffic dynamics is considered during model training. Figure 2 depicts a triplet that can constitute any part of the network with three contiguous segments.

For the training and testing of a model candidate, the input data needs to be split into training and testing data. We utilize the segments with sparse available measurement data to compute training triplets. A training triplet strictly holds the available measurements in segment i (i.e., always the middle segment). Consequently, we utilize all traffic model estimates for $i-1$, i, and $i+1$ as input and define the measurements in segment i as a target. This approach allows for solving a supervised regression problem to mitigate the error from traffic model estimates. All other triplets that do not hold measurements for segment i are denoted as testing triplets. Consequently, evaluating these triplets allows for assessing the model's performance, where the actual traffic state is unknown. Note that the ground-truth of testing triplets is available in our data set but only utilized for model assessment (i.e., given a real-world application, the ground-truth traffic state is also unknown and therefore not available for training a model).

In block (2), the input features are computed for model training. To ensure a consistent input feature vector, all on-ramp and off-ramp features are present, regardless of infrastructure design. The procedure can be explained with the following example: Segment $i-1$ in Figure 2 receives an inflow from on-ramp $r_{i-1}(k)$ but no exit is possible (i.e., no off-ramp is present). Consequently, the feature of this off-ramp is set to zero; $s_{i-1}(k) = 0 \ \forall k$. For segment i both features are set to zero, as no additional infrastructure to enter or exit the freeway segment is available. Additionally, we utilize the fundamental relationship $q_i(k) = v_i(k)\rho_i(k)$ and compute the traffic flow $q_i(k)$ as an additional feature. Finally, if, e.g., one traffic model is considered as an input (e.g., METANET), a triplet feature vector out of 15 features is concatenated (3 x traffic density, 3 x space mean speed, 3 x traffic flow, 3 x on-, and 3 x off-ramps). Also, the input feature vector is scaled as occurring values of traffic density, traffic flow, and space mean speed can vary substantially throughout network segments. Note that the framework is flexible, so the number of input features (number of models or consideration of traffic density and/or space mean speed) and output quantities (traffic density and/or space mean speed) are not fixed.

Finally, all utilized triplets for training are concatenated to the input vector X that is fed to a biLSTM neural network (Figure 3) in block (3). A neural network model consists of an input, hidden, and output layer. The layers are composed of neurons taking one/multiple inputs and computing one/multiple outputs. The connections between neurons are modified by weights that scale the computations throughout the network. In this work, we denote $f(\cdot)$ as the model that aims to approximate the functional relationship between the traffic model estimates and the measurements in segment i of training triplets. X denotes the concatenated input features and

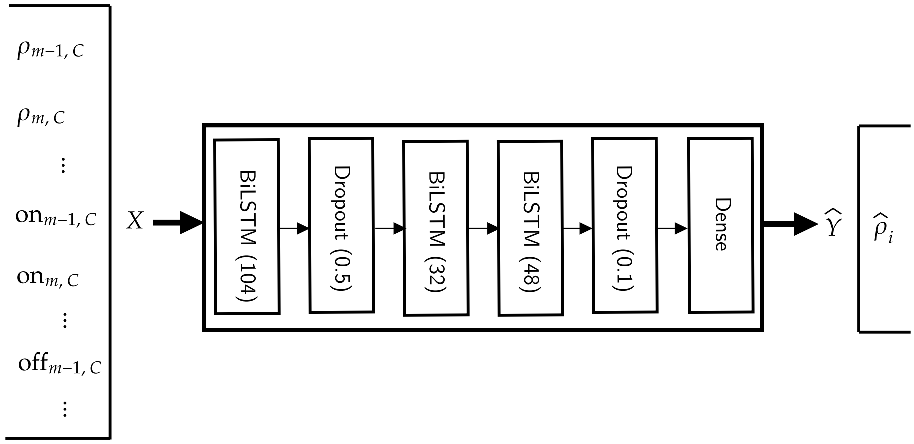


Figure 3: Schematic presentation of biLSTM model. The discrete time steps k are omitted for readability.

W defines a vector that combines all weights w_{ab} of connections from a neuron b to neuron a, and the bias b_a (Genser & Kouvelas, 2022). Consequently, the model computes an estimate of the traffic state (in Figure 3 the traffic density) with $\widehat{p}_i = f(X, W)$. For more specific details about the biLSTM model architecture, the interested reader is referred to (Huang, Xu, & Yu, 2015).

We design a biLSTM with the architecture illustrated in Figure 3: A bidirectional LSTM layer with 104 units, followed by regularization implemented with a dropout layer with a rate of 0.5. Afterward, two additional biLSTM layers with 32 and 48 units are added. Finally, another dropout with a rate of 0.1 and a dense layer for the output is added. The model is trained by utilizing the optimizer Adam. As a loss function, the Mean Squared Error (MSE) is used, and early stopping with the patience of 25 epochs is utilized in combination with model checkpoints to prevent overfitting.

For a final model assessment, the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) are utilized that are defined as follows:

$$\text{MAE} = \frac{1}{K}\sum_{k=1}^{K}\left|\hat{Y}_i(k) - Y_i(k)\right|, \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left|\hat{Y}_i(k) - Y_i(k)\right|^2}. \tag{2}$$

In equation (1) and (2), K represents the total number of samples, $\hat{Y}_i(k)$ the traffic state estimation and $Y_i(k)$ the ground-truth data.

3 Results and Discussion

The case study is based on a stretch from a 120km freeway ring road in Antwerp, Belgium. The data set is based on a calibrated representation of the network in Aimsun. The simulation scenario represents three hours, where the first hour corresponds to the loading phase (20% of peak hour demand) and the last hour to the unloading phase (20% of peak hour demand). The network stretch is represented by 17 multi-lane segments, three on-ramps, and one off-ramp. Four segments hold measurements (CAV measurements with a penetration rate of 50%) of traffic density and space mean speed. For the other 12 segments, no measurements are available. In this work, we utilize the well-known model METANET as the traditional traffic model. First, the candidate is applied to the segmented freeway stretch for the prior traffic state estimation. This step allows the computation of traffic density and space mean speed for all 17 segments.

To prove our concept, we focus this work on estimating traffic density. Note that the framework also allows training a model for space mean speed estimation or both quantities at a time. Consequently, the traffic density estimates from METANET serve as input to the proposed framework. To obtain the training triplets, we select the four segments that hold traffic density measurements

Table 1: Performance assessment of UKF and biLSTM on training (highlighted in green) and testing triplets (highlighted in blue) with RMSE, respectively.

Model	Seg. 0	Seg. 1	Seg. 2	Seg. 3	Seg. 4	Seg. 5	Seg. 6	Seg. 7	Seg.8
UKF	1.43	1.29	1.04	0.37	0.28	2.45	1.34	1.14	1.04
biLSTM	**1.15**	**1.04**	1.04	0.89	0.77	**1.25**	**1.33**	1.33	1.40

	Seg. 9	Seg. 10	Seg. 11	Seg. 12	Seg. 13	Seg. 14	Seg. 15	Seg. 16
UKF	2.72	0.79	0.86	2.03	2.51	2.25	1.98	1.45
biLSTM	**1.20**	1.10	0.93	**1.43**	**2.31**	**2.22**	2.23	2.48

from sensor data and automatically define the corresponding triplets. Also, we select the remaining segments and compute the triplets that do not have measurements (testing triplets). In other words, only the estimations from METANET are available for defining the traffic states in these segments. The estimates from the traditional traffic model are shown in the results below with black dashed lines and also serve for comparability with our proposed estimation method.

We assess our trained biLSTM model on the training and testing data set against estimation results determined by applying an Unscented Kalman Filter (UKF). The performance metric RMSE is presented for all models and training (highlighted in green) and testing segments (highlighted in blue) in Table 1.

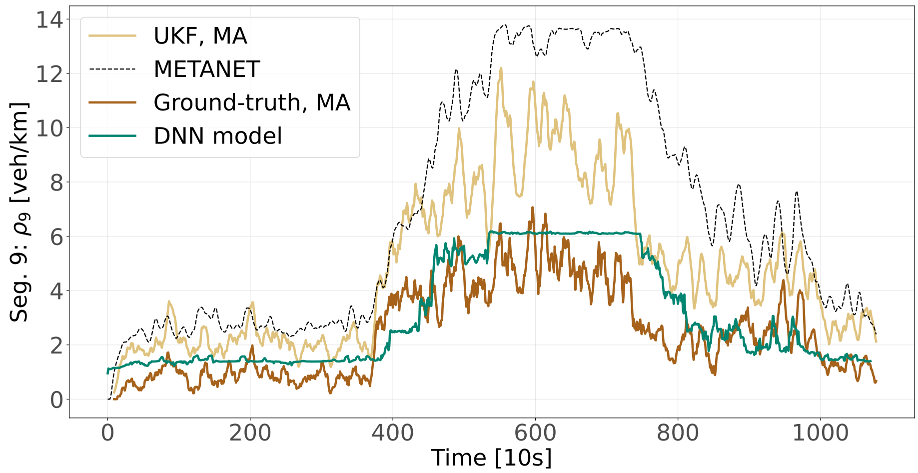
The results show that in the training segments 3, 4, 10, and 11, the second-order model performs better with lower RMSE (0.37 veh/km, 0.28 veh/km, 0.79 veh/km, and 0.86 veh/km, respectively). This highlights that the UKF follows the measurement data in segment i of a training triplet with a low error. Also, the biLSTM model can not outperform the UKF on the training data. Nevertheless, this also means that the proposed model does not overfit the training data.

Finally, we apply the biLSTM model to the unobserved testing triplets to estimate traffic density. Table 1 shows results for all models and testing segments highlighted with blue. Figure 4 showcases estimation results for segments 9 and 12, respectively. A qualitative inspection of METANET estimates (dashed black) shows an overestimated traffic density for the loaded network between 4000s and 8000s in segment 9. The UKF result (in yellow) shows a slight overestimation of traffic density during loading and unloading and a significant overestimation during 4000s and 8000s. The application of the biLSTM model allows for a reduction of the MAE error by 61% from 2.39 veh/km to 0.93 veh/km (the RMSE decreases by 56%). The estimated traffic density shows a good fit throughout the investigated period compared to the ground-truth (in brown).

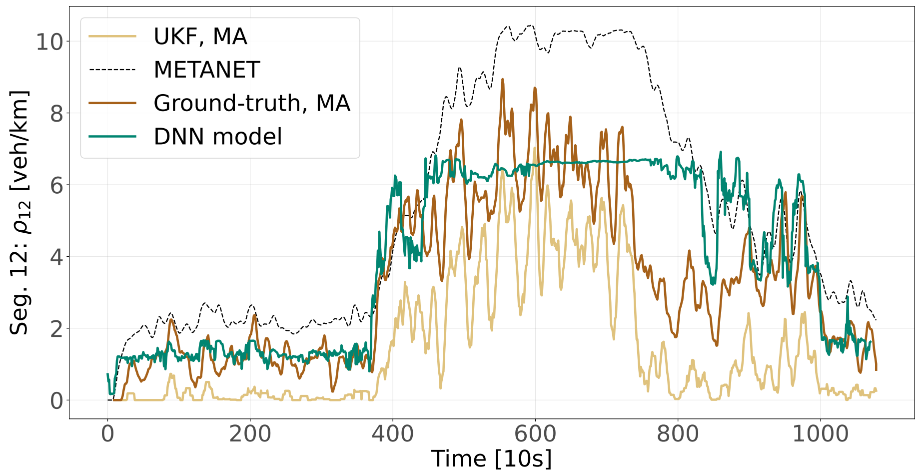
For segment 12, METANET overestimates traffic density throughout the investigated period, and the results of the UKF show an underestimation compared to the ground-truth data. The biLSTM approach shows a good fit with the ground-truth data, although an overestimation between 7000s to 8000s occurs. Results for the MAE show an improvement of 48% from 1.86 veh/km to 0.97 veh/km. Table 1 also shows a reduction of the RMSE for segment 12 from 2.03 veh/km to 1.43 km/veh, corresponding to an improvement of 30%. Out of 13 testing triplets, 8 show similar results for MAE and RMSE, where the biLSTM model outperforms the UKF and the traditional traffic model. Nevertheless, Table 1 also shows testing triplets where the UKF outperforms the biLSTM model. For example, segment 16 shows an RMSE of 1.45 veh/km and 2.48 veh/km, respectively, corresponding to a performance loss of 71%. An analysis of the METANET estimates (i.e., the input for the biLSTM model) shows a substantial overestimation by the traffic model, which is likely to cause modest results for biLSTM.

4 Conclusion

The proposed work studies a flexible framework for traffic state estimation of freeway networks by exploiting deep learning techniques. The work allows for fusing traffic model estimates and available measurement data. The methodology allows one to (a) train and assesses the model on segments where measurements are and (b) estimate the traffic state in segments where no data is available. Results highlight that the framework outperforms traditional traffic models.



(a)



(b)

Figure 4: Estimation results of biLSTM model for (a) traffic density of segment 9 and (b) segment 12, respectively. Comparison with METANET, UKF (moving average with a window of 10), and ground-truth simulation data (moving average with a window of 10).

References

Ahmed, M. S., & Cook, A. R. (1979). ANALYSIS OF FREEWAY TRAFFIC TIME-SERIES DATA BY USING BOX-JENKINS TECHNIQUES. *Transportation Research Record*(722). Retrieved 2021-04-16, from https://trid.trb.org/view/148123 (ISBN: 9780309029728)

Boto-Giralda, D., Díaz-Pernas, F. J., González-Ortega, D., Díez-Higuera, J. F., Antón-Rodríguez, M., Martínez-Zarzuela, M., & Torre-Díez, I. (2010). Wavelet-Based Denoising for Traffic Volume Time Series Forecasting with Self-Organizing Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, *25*(7), 530–545. Retrieved 2021-04-16, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8667.2010.00668.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8667.2010.00668.x) doi: https://doi.org/10.1111/j.1467-8667.2010.00668.x

Crawford, F., Watling, D. P., & Connors, R. D. (2017, January). A statistical method for estimating predictable differences between daily traffic flow profiles. *Transportation Research Part B: Methodological*, *95*, 196–213. Retrieved 2021-04-16, from https://www.sciencedirect.com/science/article/pii/S0191261516300856 doi: 10.1016/j.trb.2016.11.004

Daganzo, C. F. (1994, August). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, *28*(4), 269–287. Retrieved 2021-04-16, from https://www.sciencedirect.com/science/article/pii/0191261594900027 doi: 10.1016/0191-2615(94)90002-7

Genser, A., Hautle, N., Makridis, M., & Kouvelas, A. (2022). An experimental urban case study with various data sources and a model for traffic estimation. *Sensors*, *22*(1). Retrieved from https://www.mdpi.com/1424-8220/22/1/144 doi: 10.3390/s22010144

Genser, A., & Kouvelas, A. (2022). Dynamic optimal congestion pricing in multi-region urban networks by application of a multi-layer-neural network. *Transportation Research Part C: Emerging Technologies*, *134*, 103485. Retrieved from https://www.sciencedirect.com/science/article/pii/S0968090X2100471X doi: https://doi.org/10.1016/j.trc.2021.103485

Huang, Z., Xu, W., & Yu, K. (2015). *Bidirectional lstm-crf models for sequence tagging.* arXiv. Retrieved from https://arxiv.org/abs/1508.01991 doi: 10.48550/ARXIV.1508.01991

Lint, J. W. C. V., & Hoogendoorn, S. P. (2010). A Robust and Efficient Method for Fusing Heterogeneous Data from Traffic Sensors on Freeways. *Computer-Aided Civil and Infrastructure Engineering*, *25*(8), 596–612. Retrieved 2020-08-25, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8667.2009.00617.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8667.2009.00617.x) doi: 10.1111/j.1467-8667.2009.00617.x

Ma, T., Antoniou, C., & Toledo, T. (2020, February). Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transportation Research Part C: Emerging Technologies*, *111*, 352–372. Retrieved 2021-04-16, from https://www.sciencedirect.com/science/article/pii/S0968090X19303821 doi: 10.1016/j.trc.2019.12.022

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015, May). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, *54*, 187–197. Retrieved 2021-04-16, from https://www.sciencedirect.com/science/article/pii/S0968090X15000935 doi: 10.1016/j.trc.2015.03.014

Mihaylova, L., Boel, R., & Hegyi, A. (2007, February). Freeway traffic estimation within particle filtering framework. *Automatica*, *43*(2), 290–300. Retrieved 2021-04-16, from https://www.sciencedirect.com/science/article/pii/S0005109806003761 doi: 10.1016/j.automatica.2006.08.023

Papageorgiou, M., Blosseville, J.-M., & Hadj-Salem, H. (1990, September). Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part I: Modelling. *Transportation Research Part A: General*, *24*(5), 345–359. Retrieved 2021-04-16, from https://www.sciencedirect.com/science/article/pii/019126079090047A doi: 10.1016/0191-2607(90)90047-A

Risso, M. A., Bhouri, N., Rubiales, A. J., & Lotito, P. A. (2020, February). A constrained filtering algorithm for freeway traffic state estimation. *Transportmetrica A: Transport Science*, *16*(2), 316–336. Retrieved 2021-04-16, from https://doi.org/10.1080/23249935.2018.1549618 (Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/23249935.2018.1549618) doi: 10.1080/23249935.2018.1549618

Stathopoulos, A., Karlaftis, M. G., & Dimitriou, L. (2010, January). Fuzzy Rule-Based System Approach to Combining Traffic Count Forecasts. *Transportation Research Record*, *2183*(1), 120–128. Retrieved 2021-04-16, from https://doi.org/10.3141/2183-13 (Publisher: SAGE Publications Inc) doi: 10.3141/2183-13

Treiber, M., & Kesting, A. (2012, April). Validation of traffic flow models with respect to the spatiotemporal evolution of congested traffic patterns. *Transportation Research Part C: Emerging Technologies*, *21*(1), 31–41. Retrieved 2021-04-16, from https://www.sciencedirect.com/science/article/pii/S0968090X11001252 doi: 10.1016/j.trc.2011.09.002

Wang, Y., & Papageorgiou, M. (2005, February). Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. *Transportation Research Part B: Methodological*, *39*(2), 141–167. Retrieved 2021-04-16, from https://www.sciencedirect.com/science/article/pii/S0191261504000438 doi: 10.1016/j.trb.2004.03.003