Network-wide Traffic State Forecast Using Discrete Wavelet Transform and Deep Learning

Mohammad Javad Hassan Zada
Chair of Transportation Systems Engineering
Department of Mobility systems Engineering
Technical University of Munich
Munich, Germany
mj.hassan@tum.de

Iuliia Yamnenko
Chair of Transportation Systems Engineering
Technical University of Munich
Munich, Germany
Department of Electronic Devices and Systems
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”
Kyiv, Ukraine
i.yamnenko@tum.de

Constantinos Antoniou
Chair of Transportation Systems Engineering
Department of Mobility systems Engineering
Technical University of Munich
Munich, Germany
c.antoniou@tum.de

2023 8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS) | 978-1-6654-5530-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/MT-ITS56129.2023.10241683

***Abstract*—Traffic state prediction models are a crucial element with many applications in intelligent transportation systems. Short-term network-wide modeling of traffic states is a challenging task due to the existence of inherent characteristics such as nonlinearity, periodicity and stochasticity in the traffic state time series. This issue was responded by the evolution of advanced machine learning algorithms, e.g. deep learning. Deep neural networks can cope with high dimensionality, and also, are capable of extracting nonlinearity, comovement patterns, and spatiotemporal interdependencies between the traffic state variables from different locations. Nevertheless, they cannot completely capture the location-specific features of traffic information. Therefore, we propose the Discrete Haar Wavelet Transform (DHWT) as a preprocessing scheme prior to Multi-layer Perceptron (MLP) neural networks for one-hour ahead traffic state prediction. DHWT can help MLP to simultaneously learn the network-wide comovement patterns through the trend component time series, and seize the significant characteristics of each unique detector efficiently via the noise component. The results on 20 sensors in Paris indicated that the hybrid DHWT-MLP model with a two-level down decomposition improves the Mean Squared Error (MSE) of a non-preprocessed MLP by 33.73% and 17.58 %, for the six-month and three-month data, respectively. However, the proposed model does not perform well over the one-month period compared to the MLP model. Therefore, it may be helpful to use lower wavelet decomposition levels (higher orders) when dealing with relatively small traffic datasets.**

Keywords—traffic state prediction, preprocessing, discrete Haar wavelet transform, MLP, deep learning, Decomposition level

I. Introduction

In the traffic flow theory, a traffic state is defined by three macroscopic variables: traffic flow, mean speed and density. Sometimes travel time is also added to traffic state variables as an important traffic quality indicator. Also, traffic density is usually replaced by a relevant variable called occupancy because the density is valid only under highly homogeneous traffic conditions with no variations in speeds or dimensions of vehicles [1]. Also, most local detectors e.g. inductive loop detectors measure occupancy directly. Traffic states can be an appropriate representative of the traffic flow behavior and a good criterion to determine the level of service.

Traffic state prediction models are a crucial element of intelligent transportation systems which are used for a variety of purposes; e.g. traffic signaling, and optimal vehicle routing as real-time traffic operations. These models can be localized or network-wide, and can predict for short-term or long-term. The short-term network-wide traffic state prediction is a difficult task due to the existence of inherent characteristics such as nonlinearity, periodicity, and stochasticity in the traffic state time series. To tackle this issue, a wide range of advanced machine learning algorithms have been already applied. A popular example is the deep learning models also known as artificial neural networks. The reasons for why they are a robust choice for traffic state modeling are: (i) It is simple to embed multiple time series into a neural network as inputs, and have a multi-variate response vector as the output prediction (flexibility for high dimensionality). (ii) Artificial neural networks are also capable of capturing the nonlinear spatiotemporal interdependencies between the traffic state variables throughout transportation systems. According to Ma et al. [2], neural networks are appropriate frameworks to seize the traffic cross-network correlations, but they are not able to completely extract the location-specific features of traffic information. Therefore, it is helpful to employ preprocessing and post-processing methods before or after the time series are fed into a neural network. A second challenge of artificial neural networks when working with network traffic state time series is regarding the residuals diagnosis. Although normally distributed in most cases, the neural network residuals for each output neuron almost never result in a white noise signal. As discussed, this is mainly because some information specific to each loop detector is not captured by the neural network. To deal with this issue, we proposed the Discrete Haar Wavelet Transform (DHWT) as a preprocessing scheme prior to Multi-layer Perceptron (MLP), also known as feed-forward Back-Propagation (BP) neural networks. To evaluate the prediction performance of the introduced model, the hourly data for 20 loop detectors in the city center of Paris between January and June 2019 were retrieved. The forecast accuracy of this model was compared to that of a simple MLP. In this way, the MLP neural network is enabled to learn not only the comovement patterns of traffic dynamics on a network scale, but also the location-specific variations via decomposing each input time series into its corresponding trend and noise components.

II. Literature review

A. Neural Networks

There are many studies which have already implemented different architectures of artificial neural networks for traffic state prediction. The study conducted by Xiaojian and Quan [3], established BP neural networks to make forecast of traffic flows in crossroads. Their results indicated that the model is a suitable tool for short-term prediction with high reliability and accuracy. This finding was confirmed in another study by Liu et al. [4], where the traffic flows at intersections were modeled by a BP neural network model using different optimization algorithms. Also, neural networks have been used with other models or in other architectures. Ma et al. [2] concatenated MLP neural networks with the Auto-Regressive Integrated Moving Average (ARIMA) model to account for location-specific traffic attributes, and therefore, promote the model's predictive accuracy. Ali et al. [5] employed both recurrent and convolutional neural networks in parallel to develop a dynamic traffic flow prediction model, which exploits the spatial and temporal dependencies in the traffic data. Their results signified that this model outperforms the existing state-of-the-art algorithms. Gao et al. [6] worked on Elman recurrent neural networks along with a dissimilation particle swarm optimization technique for short-term traffic flow forecasting, and showed that this model can predict with higher accuracy and lower computational cost compared to BP and Radial Basis Function (RBF) neural networks. However, the RBF neural network could improve the performance of traffic state prediction models when they were used by Zhu et al. [7] with the traffic volume data of adjacent intersections for network-wide prediction. Do et al. [8] used attention-based neural networks to capture both the spatial and temporal relationships between the traffic flows from different road segments and time lags. Wu et al. [9] proposed a hybrid model consisting of the attention-based, Long Short-Term Memory (LSTM) and convolutional neural networks. Their findings portended that the algorithm is capable of predicting traffic flows accurately. Abdulhai et al. [10] introduced a Time Delay Neural Network (TDNN) model which is synthesized by a neuro-genetic algorithm. The model showed superior results to the conventional BP neural network. Wang et al. [11] applied a hybrid neural network including graph, recurrent, attention-based, and MLP layers, and then, carried out precise traffic flow prediction via this hybrid model. Lv et al. [12] introduced the Stacked Auto-Encoders (SAEs) to seize the latent traffic flow features, and improve the forecast accuracy.

B. *Discrete Wavelet Transform*

According to Cui et al. [13], the application of classical wavelet transform can be useful to identify abrupt temporal changes and peaks in traffic signals, and can alleviate the lack of flexibility in the local feature extraction process. In their study, they have implemented graph wavelet transform in the gate units of a graph recurrent neural network, and showed that the sparsity of the proposed methodology can ameliorate the interpretability of graph neural networks [13]. Tian [14] applied Mallat multi-scale wavelet algorithm, and developed ARIMA and also the Least Squares Support Vector Machine (LSSVM) in order to estimate the approximate and detail components of network traffic, respectively. The introduced fusion model turned out to be more accurate than the selected benchmark models. The Mallat multi-scale wavelet algorithm was also proposed by Lu and yang [15] as a preprocessing step for a LSTM recurrent neural network. They found its results more veracious than those of the LSSVM, BP neural network and Elman neural network. Chen et al. [16] utilized particle swarm optimization algorithm for the Morlet wavelet neural network, and concluded that the mentioned optimization algorithm can increase both the accuracy and stability of predictions made by wavelet neural networks. Zhao et al. [17] combined discrete multi-scale wavelet analysis and LSTM Convolutional Neural Network (LSTM-CNN), and reported that this model was superior to other modern deep learning algorithms. Peng and Ziang [18] performed short-term traffic forecast via wavelet denoising and phase space reconstruction before feeding the traffic data to a BP neural network with a genetic algorithm optimizer. The results indicated that the model generated better predictions than the other selected machine learning competitors. Bota-Giralda et al. [19] applied Mallat multi-scale wavelet denoising prior the self-organizing fuzzy neural network for traffic volume prediction. Their algorithm outperformed the MLP, historical average, and Kalman filter, but overcome by the SVM. Jiang and Adeli [20] proposed a novel non-parametric dynamic time-delay wavelet recurrent neural network which was based on a Mexican hat wavelet function to predict traffic flows. Furthermore, the modified Gram–Schmidt algorithm was utilized to drop the redundant wavelet bases when estimating the dynamic system function. Xie and Zhang [21] employed wavelet ANNs with the Morlet mother wavelet for traffic volume forecasting, and the model showed a higher accuracy than BP and RBF neural networks. The study results suggested that the application of discrete wavelet algorithms would enhance the nonlinear approximation capacity of neural networks in real-time traffic flow prediction.

The aforementioned methodologies have used a variety of wavelet transform techniques combined with artificial neural networks or other machine learning algorithms, but there are few studies to implement the discrete Haar wavelet transform, and investigate the contribution that this technique can have to the performance of deep learning models when predicting the network traffic states based on the different sample sizes. Therefore, it was adopted as a preprocessing scheme to be carried out prior the MLP, helping it to capture more location-specific and network-scale information than usual.

III. Data and Methodology

A. *Discrete Haar Wavelet Transform*

The idea of the Discrete Haar Wavelet transform (DHWT) is that non-stationary signals located in different time intervals with various frequency components could be described as a sum of the scaled and shifted basic Haar wavelet functions. DHWT decomposes each single time series into two different components: trend or the low-frequency component (T_n) and noise or the high-frequency component (N_n). The n subscript represents the order of decomposition. If the time series X contains 2^m samples, we have: $n = 0, 1, 2, \ldots, m$. As illustrated by Fig. 1, DHWT does not perform any decomposition at $n = 0$, but once $n \geqslant 1$, the time series is decomposed into one trend and one noise component, and also compressed by half in each order. Note that the decomposition always takes place on the trend component of the previous order.

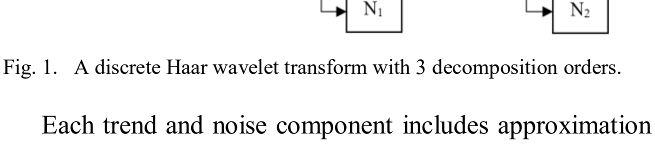

Fig. 1. A discrete Haar wavelet transform with 3 decomposition orders.

Each trend and noise component includes approximation ($S_{j,k}$) and detailing ($d_{j,k}$) coefficients. If we define the decomposition level (j) and consideration interval (k) as [22]:

$$j = j_{max}, j_{max} - 1, \ldots, 0 \qquad j_{max} = m \tag{1}$$

$$k = 0, 1, \dots, 2^j - 1 \tag{2}$$

then, we have for the highest level ($j = j_{max}$):

$$S_{j_{max},k} = \frac{X(\frac{k}{2^{j_{max}}})}{2^{\frac{j_{max}}{2}}} \tag{3}$$

For the lower levels (higher orders), we define $x = \frac{k}{2^{j_{max}}}$, and the coefficients are defined as below:

$$S_{j-1,k} = \frac{1}{\sqrt{2}} \varphi(x)\ S_j \tag{4}$$

$$d_{j-1,k} = \frac{1}{\sqrt{2}} \psi(x)\ S_j \tag{5}$$

where φ and ψ are the Haar scaling function and parent wavelet function, respectively:

$$\varphi(x) = \begin{cases} 1, & 0 \leqslant x < 1 \\ 0, & x < 0 \text{ or } x \geqslant 1 \end{cases} \tag{6}$$

$$\psi(x) = \begin{cases} 1, & 0 \leqslant x < \frac{1}{2} \\ -1, & \frac{1}{2} \leqslant x < 1 \\ 0, & x < 0 \text{ or } x \geqslant 1 \end{cases} \tag{7}$$

To obtain the trend and noise components and reconstruct the original time series, φ and ψ functions are shifted and rescaled as follows:

$$\varphi_{j,k}(x) = 2^{\frac{j}{2}} \varphi(2^j x - k) \tag{8}$$

$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k) \tag{9}$$

The reconstruction of the original time series at each order is done by adding the trend and noise components in that order to all the noise components of the lower decomposition orders.

$$T_{j_{max}-j} = \sum_{k=0}^{2^j-1} S_{j,k} \varphi_{j,k}(x) \tag{10}$$

$$N_{j_{max}-j} = \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x) + \sum_{k=0}^{2^{j+1}-1} d_{j+1,k} \psi_{j+1,k}(x) + \dots + \sum_{k=0}^{2^{j_{max}-1}-1} d_{j_{max}-1,k} \psi_{j_{max}-1,k}(x) \tag{11}$$

$$X = T_{j_{max}-j} + N_{j_{max}-j} \tag{12}$$

In this study, the traffic state time series are decomposed by DHWT for 6 orders, i.e. $n = \{0, 1, 2, 3, 4, 5\}$. The first order transform ($n = 0$) gives the original time series, and this model is equivalent to a MLP neural network without any DHWT conducted. The higher the decomposition order, the smoother the trend component, and the more fluctuating the acquired noise component. Once the optimal decomposition order is identified for the full traffic dataset, the proposed algorithm with this order is applied to smaller datasets for more insights.

B. *MLP Neural Networks*

Multi-Layer Perceptron neural networks (MLP) are a class of machine learning algorithms defined as decompositions of nonlinear functions. A MLP neural network model with an input layer of I neurons, one hidden layer of H neurons, activation function g, bias vector δ, and weight matrices w and v is formulated as the following, and visualized in Fig. 2.

$$\hat{f}(x_i) = g(\delta(o) + \sum_{h=1}^{H} (v_h\ g(\delta(h) + \sum_{i=1}^{I} w_h x_i))) \tag{13}$$

The MLP parameters (weight matrices w and v and biases $\delta(h)$ and $\delta(o)$ in our example) are estimated through Adam's algorithm [23] which is a gradient-based optimization method for stochastic objective functions. The hyperparameters (e.g. number of hidden layers, number of neurons in each hidden layer H and the activation function g) are determined using a random search as a diversified heuristic, searching on different predefined sets of candidates to find an optimal combination of hyperparameters.

To prevent from overfitting, reduce the model variance, and increase the generalizability of the MLP neural network, two regularization schemes were implemented: early stopping and drop-out. The early stopping stops the training process once the validation error does not improve after a certain number of epochs. The drop-out scheme drops out a certain portion of hidden neurons at each epoch randomly during training to ensure that the neurons which do not contribute to a high-accuracy prediction are disincentivized by converging to zero, and thus, the neural network is not overparameterized.

In order to reshape the time series data into the inputs and outputs of the neural network (x and $\hat{f}$), we applied a sliding window (forward chaining) method to reconstruct the dataset to an appropriate format for supervised learning with a capture window (embedding dimension) of p, a time lag of h=1 and prediction horizon of a = 1 using the Takens' theorem [24]. A similar example where p = 5, h = 1, and a = 5 is illustrated in Fig. 2.

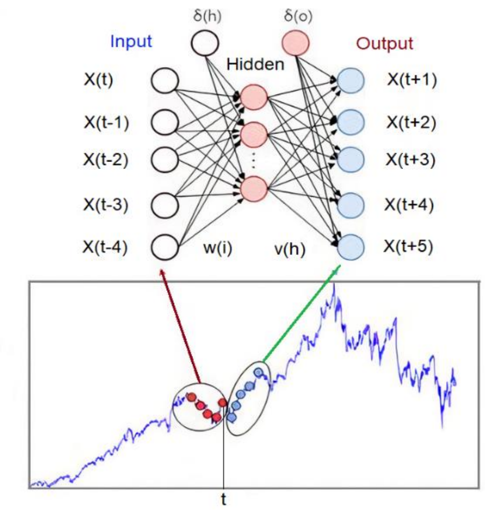

Fig. 2. Construction of a neural network for a time series using a sliding window method, from [25] with edition.

The input values (x) of the neural network were scaled by the min-max normalization. This helps guarantee or speed up the convergence of the Adam's optimization algorithm, and leads to a higher predictive accuracy.

C. Data Description

The hourly traffic data belong to 20 selected loop detectors from an urban corridor in the city of Paris, France between January and June 2019. The study area starts from Quai des Célestins near the city center, continues along the Seine river, and finishes at Cr la Reine in the west as shown by Fig. 3.

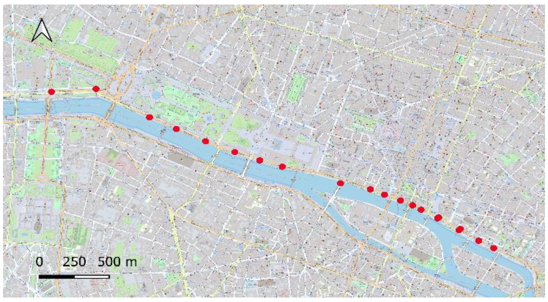

Fig. 3. Study area in Paris, France with 20 inductive loop detectors.

In Fig. 4, the nonlinear relationships between the traffic state variables are illustrated in the fundamental diagrams. The nonlinearity and randomness can be caused by different phenomena such as shock waves which occur due to many unforeseen reasons such as accidents, bottlenecks, temporary road maintenance operations, harsh weather conditions, etc. Typically, traffic states with high occupancy and low speed (congested states) are harder to predict than the more common free-flow states in which high speed and low occupancy are observed.

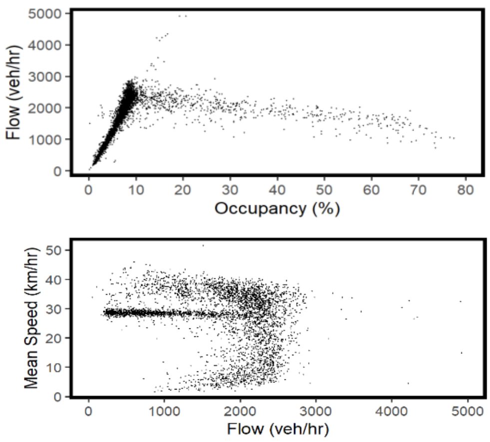

Fig. 4. Fundamental diagrams for the traffic data of loop detector 1 (L1).

Fig. 5 demonstrates the linear correlation coefficients among the traffic flows of some selected loop detectors.

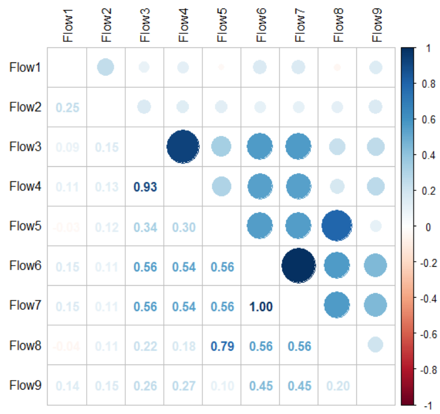

Fig. 5. Correlation matrix for traffic flows of some selected loop detectors.

The correlation matrix reveals that there are positive cross-correlations between all the selected loop detectors, and the numbers indicate high spatial dependencies of the time lag = 0 in some of them.

D. DHWT–MLP Model

The DHWT-MLP model was developed by performing the DHWT, and feeding the results into MLP neural network. The data set was split randomly by 80/20 ratio into the training and validation datasets, respectively. The hold-out validation scheme was utilized to monitor the generalizability of the neural network during the training phase. Also, the training data were divided into mini-batches of 100 records to improve the learning process. The model estimation was repeated for different time periods. Afterwards, hourly forecast was made for the next 7 days (168 time intervals) by the DHWT-MLP and the MLP as the benchmark model.

The Mean Squared Error (MSE) was calculated based on the models' forecasts. It is defined as:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{N} (O_i - A_i)^2}{N} \tag{14}$$

Where O_i is the i^{th} observation value, A_i is the i^{th} actual value, and N represents the number of test samples (in our case, N = 168).

Finally, the configuration of traffic state variables in the neural network are shown by Fig. 6. The traffic state variables include flow, occupancy, mean speed, and speed standard deviation (Sd.). For all the detectors, the previous p traffic states are fed into the MLP, and the state of the upcoming hour is predicted. In case of DHWT, the trend and noise for the past p steps are inserted in the MLP as inputs, and the traffic states are forecasted for the next hour. The model evaluation is done for six, three, and one month of data.

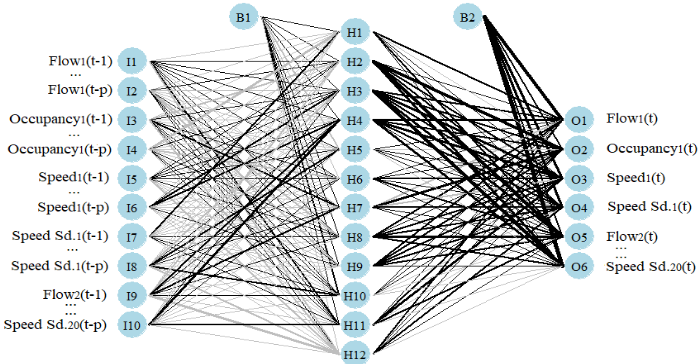

Fig. 6. MLP architecture for network-wide traffic state prediction.

IV. Results

In this section, the results of the developed models are presented. The error residuals of both deep learning models followed a normal distribution. The residuals for the MLP had a negative skewness which disappeared once the DHWT was conducted on the traffic time series in DHWT-MLP. The residuals for the DHWT-MLP model were also hugely, but yet not completely whitened confirmed by the Auto-Correlation Function (ACF) plots and the Ljung-Box statistical hypothesis test. When the corresponding methods were performed on the MLP residuals, though, it was revealed that the obtained residuals are by far different from a white noise. The ACF plot included a spike in the lag of 24.

Fig. 7 shows the MSE dynamics for both training and validation of DHWT-MLP model. It indicates that within 165 epochs, the training error has a monotonically decreasing trend, and the Adam's algorithm has obviated the fluctuations of the validation error, reaching it to a small value close to the training error value before the training is stopped.

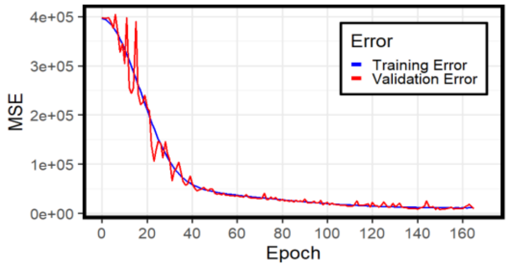

Fig. 7. Change of training and validation mean squared error (MSE).

Table I evaluates the performance of the proposed model when the data for different time periods are provided. The six-month-long DHWT-MLP with n = 1 and p = 3 improved the MSE by 33.73%, while the preprocessed model for three-month data reduced the MSE of the MLP by 17.58% on average. Nevertheless, the DHWT aggravated the predictive accuracy of the MLP neural network for one-month-long data by an average of 14.14%.

TABLE I. COMPARISON OF MLP AND DHWT-MLP MODELS FOR DIFFERENT TIME PERIODS (N=1).

Time period	Index	MLP MSE	DHWT-MLP MSE	Error change %
1-month-long models				
Jan	1	73146.71	84216.39	+15.13
Feb	2	39915.47	41289.48	+3.44
Mar	3	50873.63	58894.41	+15.77
Apr	4	66747.41	79733.74	+19.46
May	5	29807.48	34202.88	+14.75
Jun	6	25105.36	27633.52	+10.07
	Average	47599.34	54328.40	+14.14
3-month-long models				
Jan-Mar	1	24548.76	22618.27	-7.86
Apr-Jun	2	15299.91	10226.98	-33.16
	Average	19924.33	16422.63	-17.58
6-month-long model				
Jan-Jun	1	11915.62	7897.00	-33.73

Fig. 8 illustrates the traffic flow forecasts made by both models against the actual observations for the loop detector 1 (L1). It can be noticed that the DHWT-MLP improves the prediction at both large and small traffic flow values while the MLP overestimates the smaller traffic flows, and also slightly underestimates the larger flow values. Moreover, The DHWT-MLP algorithm has yielded fewer outliers than the MLP. This explains that the proposed model becomes more familiar with the physics of the traffic flow once the traffic time series are preprocessed by the DHWT algorithm. Lastly, Table I shows that as the forecast MSE values and the amounts of error improvement are highly dependent on the dataset size.

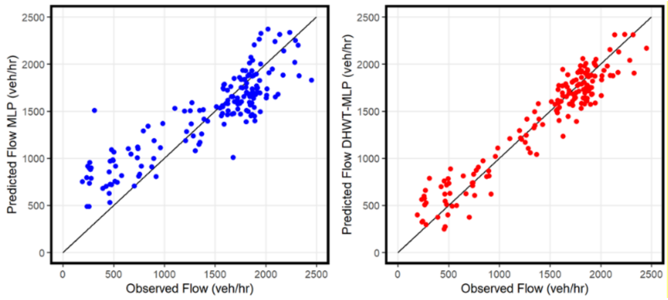

Fig. 8. Observed vs. predicted traffic flows for one-week horizon of L1.

V. DISCUSSION

The results of this study indicate that the application of DHWT as a preprocessing scheme performed prior to deep learning algorithms can significantly improve the prediction accuracy over the 6-month and 3-month datasets. The findings from Fig. 8 point out a clear decrease in the outlying forecasts of the DHWT-MLP algorithm. Hence, the DHWT helps MLP to simultaneously learn network-wide comovement patterns through the trend (approximation) component, and meanwhile exploit the significant characteristics of each unique detector efficiently via the noise component. From Fig. 8, it is apparent that the DHWT-MLP has mitigated both the overestimation of traffic free flows at midnights and their underestmation during the peak hours for the loop detectors. This location-specific learning can fortify neural networks to avoid making outlying predictions. This finding is aligned with the results from the residual diagnosis (Ljung-Box tests and ACF plots), where the residuals of the introduced model turned out to be relatively similar to white noise time series. To fully whiten the residuals of the proposed hybrid model, it can be advantageous to apply post-processing methods, as well.

The spikes in the ACF plots of the model residuals signify that MLP neural networks did not capture the local seasonality of the traffic state time series. This can well justify for why the literature has introduced hybrid algorithms composed of deep learning and statistical time series models. Although it is common to include time-related variables in machine learning models in order to exploit cyclic features, some time series models can explicitly encapsulate the seasonal traffic patterns. The lag of 24 accounts for the daily periodicity.

As the dataset size (time series length) becomes smaller, the forecast accuracy declines, and the MSE values increase. On the other hand, the improvement in the MSE decreases up until the benchmark MLP error overtakes that of the DHWT-MLP. This can suggest that when the DHWT-MLP model is trained using short time series, a higher predictive accuracy can be obtained in lower decomposition levels (higher orders) where the trend component is much smoother.

VI. CONCLUSIONS

This research introduced the Discrete Haar Wavelet Transform (DHWT) as a preprocessing scheme prior to Multi-Layer Perceptron (MLP) neural network for one-hour ahead network-wide traffic state prediction. Six-month traffic data for an urban corridor in the city center of Paris, France were used to develop the preprocessed DHWT-MLP model, and compare its predictive accuracy with that of MLP. The results indicated that the proposed model with a capture window of p = 3 and decomposition order of n = 1 improved the forecasts significantly. It was revealed that the time period for which the models are estimated is a critical factor in the predictive accuracy of the DHWT-MLP model, and the model with a two-level down decomposition benefits only when the hourly data for at least multiple months are available. By adding the DHWT before the neural network, the MLP was enabled to extract the location-specific information of the loop detector time series, as well as the nonlinear comovement patterns, and spatial and temporal interdependencies between the traffic state variables from different locations. Thus, the existing problems with MLP e.g. the underestimation, overestimation, prediction of outlying values, and the lack of sequential randomness in error residuals were obviated completely or mitigated to a desirable degree. The study also provided evidence that the DHWT can alleviate the need to have deeper MLP neural networks.

One important limitation of most studies on the network-wide traffic state forecast is that the considered traffic data collection points do not cover the entire network, but a sample from the correlated points of interest is taken. This is the case in our study where all 20 loop detectors are located on a similar corridor. Therefore, it should be taken into account that the existence of higher correlations is more likely between these loop detectors, and thus, it can lead to exaggerating results for network-wide traffic forecasting. Consequently, there is need for robust clustering models which can help to wisely classify the inductive loop detectors and other cross-sectional sensors throughout transport networks. Therefore, a neural network model can be developed for each cluster of traffic sensors.

For the future work, the authors are working to explore higher wavelet decomposition orders for data preprocessing, and combine different seasonal adjustment methods with the discrete Haar wavelet analysis in order to feed the neural networks even with more simplified inputs, and increase their predictive accuracy. The preliminary results showed that this hybrid scheme can improve the MSE values to a higher degree than what has been achieved in this study. Finally, there is a gap in the literature about comparing the performance of various discrete wavelet transform techniques when applied to the MLP and other architectures of artificial neural networks.

ACKNOWLEDGMENT

This research was funded by the Technical University of Munich – Institute for Advanced Study (TUM-IAS), Philipp Schwartz-Initiative – Alexander von Humboldt-Stiftung.

REFERENCES

[1] V. T. Arasan and G. Dhivya, “Measurement of occupancy of heterogeneous traffic using simulation technique,” IFAC Proceedings Volumes, vol. 42, no. 15, pp. 19-24, 2009. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] T. Ma, C. Antoniou, and T. Toledo, “Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast,” Transportation Research Part C: Emerging Technologies, vol. 111, pp. 352-372, 2020.

[3] G. Xiaojian and Z. Quan, "A traffic flow forecasting model based on BP neural network," in 2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS), 2009, vol. 3: IEEE, pp. 311-314.

[4] Z. Liu, Z. Yang, and P. Gao, “Research on the short-term traffic flow prediction method based on BP neural networks,” in World Automation Congress 2012, 2012: IEEE, pp. 1-4.

[5] A. Ali, Y. Zhu, Q. Chen, J. Yu, and H. Cai, “Leveraging spatio-temporal patterns for predicting citywide traffic crowd flows using deep hybrid neural networks,” in 2019 IEEE 25th international conference on parallel and distributed systems (ICPADS), 2019: IEEE, pp. 125-132.

[6] H. Gao, J. Zhao, and L. Jia, “Short-term traffic flow forecasting model of Elman neural network based on dissimilation particle swarm optimization,” in 2008 IEEE International Conference on Networking, Sensing and Control, 2008: IEEE, pp. 1305-1309.

[7] J. Z. Zhu, J. X. Cao, and Y. Zhu, “Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections,” Transportation Research Part C: Emerging Technologies, vol. 47, pp. 139-154, 2014.

[8] L. N. Do, H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung, “An effective spatial-temporal attention based neural network for traffic flow prediction,” Transportation research part C: emerging technologies, vol. 108, pp. 12-28, 2019.

[9] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, “A hybrid deep learning based traffic flow prediction method and its understanding,” Transportation Research Part C: Emerging Technologies, vol. 90, pp. 166-180, 2018.

[10] B. Abdulhai, H. Porwal, and W. Recker, “Short-term traffic flow prediction using neuro-genetic algorithms,” ITS Journal-Intelligent Transportation Systems Journal, vol. 7, no. 1, pp. 3-41, 2002.

[11] X. Wang et al., "Traffic flow prediction via spatial temporal graph neural network," in Proceedings of the web conference 2020, 2020, pp. 1082-1092.

[12] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865-873, 2014.

[13] Z. Cui, R. Ke, Z. Pu, X. Ma, and Y. Wang, “Learning traffic as a graph: A gated graph wavelet recurrent neural network for network-scale traffic prediction,” Transportation Research Part C: Emerging Technologies, vol. 115, p. 102620, 2020.

[14] Z. Tian, “Network traffic prediction method based on wavelet transform and multiple models fusion,” International Journal of Communication Systems, vol. 33, no. 11, p. e4415, 2020.

[15] H. Lu and F. Yang, “A network traffic prediction model based on wavelet transformation and LSTM network,” in 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), 2018: IEEE, pp. 1-4.

[16] Q. Chen, Y. Song, and J. Zhao, “Short-term traffic flow prediction based on improved wavelet neural network,” Neural Computing and Applications, vol. 33, no. 14, pp. 8181-8190, 2021.

[17] J. Zhao, H. Qu, J. Zhao, and D. Jiang, “Spatiotemporal traffic matrix prediction: a deep learning approach with wavelet multiscale analysis,” Transactions on Emerging Telecommunications Technologies, vol. 30, no. 12, p. e3640, 2019.

[18] Y. Peng and W. Xiang, “Short-term traffic volume prediction using GA-BP based on wavelet denoising and phase space reconstruction,” Physica A: Statistical Mechanics and its Applications, vol. 549, p. 123913, 2020.

[19] D. Boto-Giralda et al., "Wavelet - based denoising for traffic volume time series forecasting with self - organizing neural networks," Computer - Aided Civil and Infrastructure Engineering, vol. 25, no. 7, pp. 530-545, 2010.

[20] X. Jiang and H. Adeli, "Dynamic wavelet neural network model for traffic flow forecasting," Journal of transportation engineering, vol. 131, no. 10, pp. 771-779, 2005.

[21] Y. Xie and Y. Zhang, "A wavelet network model for short-term traffic volume forecasting," Journal of Intelligent Transportation Systems, vol. 10, no. 3, pp. 141-150, 2006.

[22] I. Yamnenko and V. Levchenko, "Video-data compression using wavelet analysis," in 2019 IEEE 20th International Conference on Computational Problems of Electrical Engineering (CPEE), 2019: IEEE, pp. 1-4.

[23] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[24] F. Takens, “Detecting strange attractors in turbulence,” in Dynamical systems and turbulence, Warwick 1980: Springer, 1981, pp. 366-381.

[25] R. Chandra and Y. He, “Bayesian neural networks for stock price forecasting before and during COVID-19 pandemic,” Plos one, vol. 16, no. 7, p. e0253217, 2021.