Transportation Research Part C 111 (2020) 352–372

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast[☆]

Tao Ma[a,*], Constantinos Antoniou[a], Tomer Toledo[b]

[a] Chair of Transportation Systems Engineering, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany

[b] Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel

ARTICLE INFO

Keywords:
Machine learning
Postprocess residuals
Network-wide traffic prediction
MSVR
Neural network
ARIMA

ABSTRACT

We propose a novel approach for network-wide traffic state prediction where the statistical time series model ARIMA is used to postprocess the residuals out of the fundamental machine learning algorithm MLP. This approach is named as NN-ARIMA. Neural Network MLP is employed to capture network-scale co-movement pattern of all traffic flows, and ARIMA is used to further extract location-specific traffic features in the residual time series out of Neural Network. The experiment results show that the postprocessing the residuals of Neural Network by the ARIMA analysis helps to significantly improve accuracy of traffic state prediction by 8.9–13.4% in term of mean squared error reduction. In order to verify the efficiency of the ARIMA analysis in the postprocessing, Multidimensional Support Vector Regression (MSVR) model is also employed to replace the role of Neural Network in the comparative experiment. Two streams of comparisons, (1) NN vs. NN-ARIMA and (2) MSVR vs. MSVR-ARIMA, are performed and show consistent results. The proposed approach not only can capture network-wide co-movement pattern of traffic flows, but also seize location-specific traffic characteristics as well as sharp nonlinearity of macroscopic traffic variables. The case study indicates that the accuracy of prediction can be significantly improved when both network-scale traffic features and location-specific characteristics are taken into account.

1. Introduction

1.1. Motivation & contributions

Short-term traffic state prediction is an essential component of Intelligent Transportation Systems that serves traffic operations and management in smart city framework. The definition of traffic state prediction primarily refers to forecasting three macroscopic traffic variables, including traffic volume, speed, and occupancy. These traffic variables are the key measurements of roadway performance. Some studies also include travel time as a key indicator of traffic state. This study focuses on prediction of three macroscopic traffic variables.

From the methodological perspective, the predictive methods can be loosely classified as statistical methods, such as ARIMA (autoregressive and moving average), and machine learning methods, such as gradient-descent based Neural Network or kernel-based Support Vector Regression. A great number of methodologies were developed over the past three decades for traffic state prediction.

☆ This article belongs to the Virtual Special Issue on "Traffic flow modeling".

* Corresponding author.
E-mail address: tao.ma@tum.de (T. Ma).

https://doi.org/10.1016/j.trc.2019.12.022
Received 17 March 2019; Received in revised form 14 December 2019; Accepted 27 December 2019
0968-090X/

A more comprehensive summary of literature review in this regard is presented in Section 2. From the perspective of predictive capacity and scope, traffic prediction approximately can be classified into local prediction and network-wide prediction. In the literature, most traffic prediction studies are local prediction. Along with rapid advance of communication technology for large-scale data collection, the researches for traffic state prediction shift from local prediction to network-wide prediction. Most recent studies on network-wide traffic prediction among others include (Dauwels et al., 2014; Laharotte et al., 2015; Ma et al., 2017; Mitrovic et al., 2015; Yu et al., 2017; Hara et al., 2018; Dai et al., 2019; Wang et al., 2019). The methodologies commonly used in those studies for network-wide prediction are machine learning algorithms, such as long short-term memory (LSTM) recurrent neural network (RNN), convolution neural network (CNN), K-nearest neighborhood (KNN), or Multi-dimensional Support Vector Machine (MSVR). Most of those methods preprocess data. Some of them explicitly preprocess the data before feeding the data to machine learning algorithms; the others implicitly embed the preprocess in the machine learning algorithms. For instance, RNN adopts specially designed reverse flow structure to capture dependence in the time series and integrates the structure with gradient-descent algorithm. Most methods basically treat machine learning algorithms as a final stop in data process. The residuals out of the models are usually not checked and guaranteed for whiteness. However, it is found that the traffic features still exist in the residuals.

In this research, we propose a novel approach for network-wide traffic prediction that explicitly and independently postprocesses the residuals out of a machine learning algorithm. It is a combination of a machine learning algorithm with a fundamental statistical time series model. More specifically, we sequentially concatenate a fundamental Neural Network model with an ARIMA model, named as NN-ARIMA. According to the experimental results, the proposed approach is able to significantly reduce the MSE (mean squared error) of predictions, thus improving the accuracy of prediction. The effectiveness of the proposed approach is further verified and confirmed by a combination of MSVR and ARIMA, named as MSVR-ARIMA. More comprehensive details of experimental outcomes are presented in Section 4.

The reason to choose Neural Network algorithm as the component of the proposed approach is because it is able to (i) process high-dimensional multiple time series data all at once; (ii) capture sharply fluctuated nonlinearity of the data; and (iii) implicitly capture the spatial and temporal correlations among network-wide traffic flows. All those challenges exhibit in macroscopic traffic variables and are presented in detail in the next section. Therefore, we employ Neural Network to perform network-wide simultaneous forecast for all locations of interest, by taking all traffic time series all at once as the model input.

There are numerous variants of Neural Networks in the literature. The reason to choose a fundamental neural network in the proposed approach is that most of neural networks and their variants contain data preprocess mechanism explicitly or implicitly. In order to better assess the efficiency of the residual postprocess by ARIMA, we do not want any interference on the data from any external algorithms other than Neural Network algorithm itself, i.e. feed forward pass and gradient-descent backward propagation. Therefore, a simple and fundamental Neural Network algorithm such as MLP is chosen to satisfy this purpose.

1.2. Challenges of methodological development

1.2.1. Sharp nonlinearity

Traffic flows usually exhibit either a free flow state or congested state. The sharp nonlinearity of traffic variables comes from recurrent or non-recurrent congested states, due to bottlenecks, incidents, extreme weather condition or other events. Fig. 1 presents a set of Fundamental Diagrams that are constructed with observed hourly traffic volume, speed, and occupancy collected from a loop detector on the Ayalon Highway in Tel Aviv, Israel. A free flow state in the diagrams is color-coded as green, and a congested state as red. Correspondingly, Fig. 2 shows time series plots of hourly traffic volume, speed, and occupancy, using the same set of traffic data as Fig. 1. Red dots in both figures match one-on-one. Figs. 1 and 2 complement each other. The macroscopic fundamental diagram (MFD) is an important instrument for adaptive traffic control. The construction of an urban network-level MFD may refer to (Buisson and Ladier, 2009). The fundamental diagrams show traffic states and relationships between pairs of traffic variables, but not traffic dynamics over time, whereas, time series plots show traffic evolution over time, but not traffic states.

The red dots on the time series plots represent those traffic events or measurement outliers that are usually difficult to capture by many traffic predictive models.

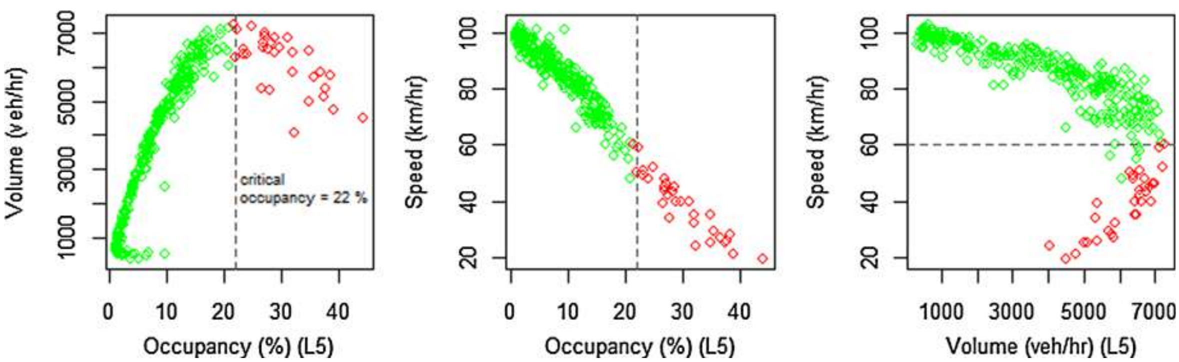

Fig. 1. Fundamental diagrams created with the traffic data from loop detector 5.

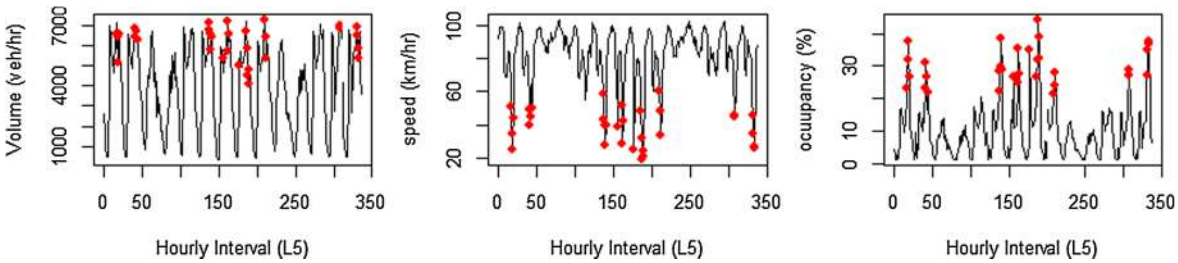

Fig. 2. Time series of hourly traffic volume, speed, and occupancy from loop detector 5.

1.2.2. Network-scale spatial and temporal correlation

From a system perspective, the dynamics of traffic flows across urban road network exhibit network-wide temporal and spatial correlations, and co-movement patterns. As correlation matrix is symmetrical, Fig. 3(a) shows the upper triangular part of a correlogram of 10 traffic volume time series from the highway network. This correlogram combines the correlation coefficient and the *p*-value of the correlation for all pairs of traffic time series. The *p*-values correspond to the significance levels of correlations. If the *p*-value in the matrix is bigger than the specified significant level (e.g. $\alpha = 0.01$), then the corresponding correlation coefficient is regarded as insignificant and its graph is blank. Positive correlations are displayed in blue and negative correlations in red color. The intensity of color and the size of the circle are proportional to the correlation coefficients. The right-hand-side legend shows the scale of correlation coefficients and the corresponding colors. Fig. 3(b) shows the upper triangular part of a correlogram of 10 traffic speed time series from the highway network.

There are different methods for correlation analysis between two random variables including, but not limited to, Pearson, Kendall, and Spearman (Bonett and Wright, 2000). Pearson's correlation is a parametric method that measures the linear dependence between two variables. Kendall and Spearman are non-parametric rank-based correlation analysis. In this study, Pearson's correlation coefficients between all pairs of traffic volume time series are computed to investigate the linear dependence between pairs of network-wide traffic flows. The numerical value of correlation coefficient, the color intensity, and the circle size in Fig. 3(a) and (b) together indicate that there exist strong spatial correlations and co-movement patterns among network-wide traffic flows.

Fig. 4 shows a plot of cross-correlation between two traffic time series from location 5 and 6. The plot indicates a strong temporal correlation between two traffic time series. The cross-correlation function also indicates that repeating patterns or seasonality concurrently appear among two traffic flows.

1.2.3. Summary

The challenge and complexity of developing a methodology for network traffic state prediction lies in the fact that a good approach must have the capacity to: (1) capture sharp nonlinearity of traffic variables; (2) take into account network-wide spatial and temporal correlations and co-movement pattern of traffic flows; (3) capture location-specific characteristics of traffic flows; (4) make sure that all traffic features are extracted by the model and residuals are thoroughly whitened to white noise (Ma et al., 2018).

The paper is organized as follows: Section 2 presents a near comprehensive literature review; Section 3 elaborates the proposed methodologies; Section 4 presents the application of the proposed method to a real case study and model comparison in terms of predictive performance; Section 5 presents conclusions and discussions.

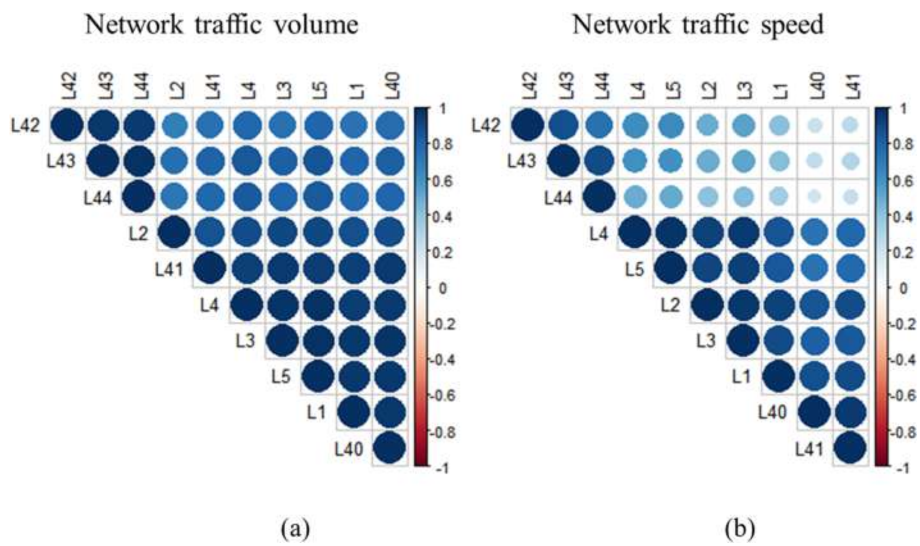

Fig. 3. Correlograms of network-wide traffic flows.

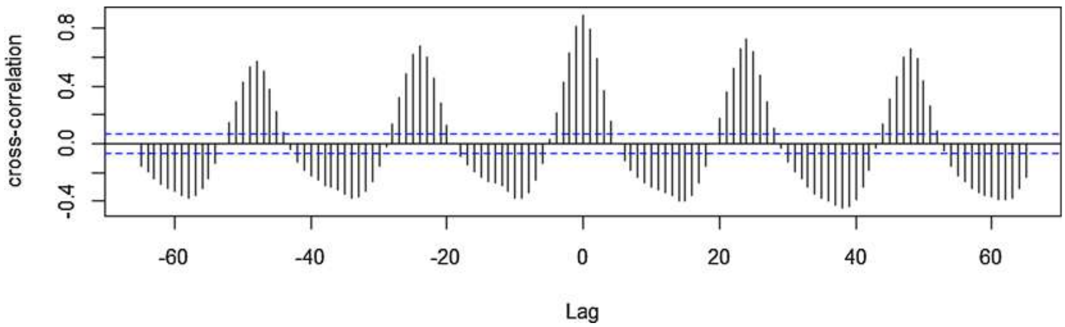

Fig. 4. Spatial cross-correlation between two traffic volume time series (L5 and L6).

2. Literature review

Many methodologies have been proposed to predict macroscopic traffic states over the past three decades. Those methodologies are derived from multiple disciplines, including, but not limited to, statistical sciences, control theory, artificial intelligence, and applied mathematics. There are different ways to classify the methodologies in the literature; however, we are inclined to loosely dividing those methodologies into two major classes: statistical time series and machine learning methods.

2.1. Statistical methods

There are many classes of time series models that have been applied to traffic state prediction. A benchmark model starts from ARIMA (Autoregressive Integrated Moving Average). It had been extensively used for single-location traffic prediction and often chosen by researchers as a yardstick for the purpose of comparison with other models (Lee and Fambro, 1999; Smith et al., 2002). Williams and Hoel (2003) introduce a SARIMA (seasonal ARIMA) to model repeating patterns or seasonality of macroscopic traffic variables. Their results indicate that SARIMA outperforms ARIMA and provides significantly improved prediction accuracy. However, ARIMA and SARIMA models are mostly used for single-location traffic prediction, unable to reflect spatial correlation of topology structure of a transportation network. Kamarianakis and Prastacos (2005) introduce STARIMA (Space-Time ARIMA) to incorporate both spatial correlation between adjacent locations and temporal autocorrelation in one model, where a spatial-correlation matrix is introduced to a standard ARIMA as a coefficient matrix. The success of the model is conditional on the construction of a spatial weight matrix. This model was also investigated by Cheng et al. (2012), and Min et al. (2010). Cheng et al. (2012) argue that it is not always reliable to use a distance function to determine spatial order. An effective spatial order is dynamic, such that it becomes larger under a free flow traffic state and smaller under a congested traffic state. In order to consider multivariate traffic variables, Williams (2001) introduces ARIMAX model that uses transfer functions to include exogenous variables from upstream traffic. Min and Wynter (2011) introduce MSTARMA that is a multivariate spatial-temporal autoregressive moving average model for traffic volume and speed prediction.

These are linear models that cannot deal with nonlinearity or structural changes in the traffic characteristics time series. To be able to consider nonlinearity, D'Angelo et al. (1999) and Liu et al. (2010) develop regime-switching SETAR (self-exciting threshold autoregressive) model for modelling and forecasting hourly traffic volume. Ma et al. (2015) develop nonlinear multivariate time-space threshold vector error correction model for short term traffic state prediction. The regime-switching models involve structural change detection and threshold identification mechanism. Similarly, Sun and Liu (2011) introduce a STAR (smooth transition autoregressive) model for traffic prediction. The model switches between two autoregressive (AR) parts via a smooth transition function. Kamarianakis et al. (2012) propose a temporal regime-switching model for traffic volume prediction. The time-based thresholds are predefined and fixed by partitioning 24 h into five temporal regimes. A linear autoregressive model is constructed for each time-based regime.

Some statistical time series models are variants of simple models via combination. Zhang et al. (2014) propose a hybrid of ARIMA and GARCH model, where ARIMA is used to determine the mean and GARCH models the variance. The ARIMA-EM model is a combination of ARIMA and expectation maximization algorithm proposed by Cetin and Comert (2006). Other statistical methodologies are also proposed for traffic forecast. Castillo et al. (2008) develop a Bayesian Network for predicting traffic flow. Wang et al. (2014) propose a Bayesian combination method with three component predictors including ARIMA, Kalman Filter, and BPNN for traffic flow forecasting. Fusco et al. (2016) propose a hybrid modeling framework that joins a Bayesian network and a neural network for traffic speed predictions.

State-Space models are often classified as a standalone class. However, the essence of a state-space model is a statistical time series type of model that runs two time series in parallel, explicit and implicit. The term "state space" originated in the 1960s from the area of control engineering, and Kalman (1960) developed a Kalman algorithm to estimate state space models that are extensively used for traffic prediction (Antoniou et al., 2007; Okutani and Stephanedes, 1984; Whittaker et al., 1997). Stathopoulos and Karlaftis (2003) employ multivariate time-series state space models for traffic volume prediction for urban arterial streets and report that the state space model reduced into an ARIMA model in some cases. Guo et al. (2014) develop an adaptive Kalman filter approach to convert the SARIMA-GARCH structure into two state space representations for modeling and predicting traffic speed series. Dong et al. (2014) develop multivariate state space models based on macroscopic traffic flow models for network flow rate and time mean speed predictions taking into account congested and non-congested traffic state respectively. The Extended-Kalman-Filter (EKF) method was combined with second order macroscopic traffic flow models for traffic state prediction (Bellemans et al., 2006; Tampere and Immers, 2007; Wang and Papageorgiou, 2005). As both the Kalman recursion structure and the second order macroscopic traffic flow models are first order difference equations, it implies that the future traffic state is only correlated with the current measurement. Other possible correlations between the future state and the past intervals and spatial correlations are not taken into account. Loosely speaking, Kalman recursion is a linear model. Its prediction performance is compromised for a nonlinear traffic state with significant sharp fluctuations. Qi and Ishak (2014) develop a Hidden Markov Model for short-term freeway traffic prediction during peak periods. A hidden Markov model is a specific type of state space model, if those states exhibit the Markov property. Both state space model and hidden Markov model are a kind of "hidden process model" based on probabilistic theory.

Similarly, non-parametric K-Nearest Neighborhood (KNN) regression is also a type of statistical model, but often treated as a standalone class. Non-parametric models mainly refer to non-parametric KNN regression and its variants. The KNN regression is developed based on chaotic system theory rather than stochastic system theory that is the basis of time series models. Disbro and Frame (1989) and Wang et al. (2005) argue with some evidence that traffic flows exhibit chaotic behavior and properties. Shang et al. (2005) argue with phase space techniques that traffic time series have a strong chaotic signature, due to the positive largest Lyapunov Exponent and low correlation dimension. With the KNN method for traffic prediction, K observations are selected from a historical database based on their nearness to the current observation of traffic variable to form a nearest neighborhood. The efficiency of non-parametric KNN regression is directly dependent on the quality and size of the database. However, execution time searching for the nearest neighborhood will be compromised as the database expands. This is a significant issue for online applications of KNN to traffic flow prediction (Smith et al., 2002). Ryu et al. (2018) adopt KNN model with traffic state vectors for traffic flow prediction, where the traffic state vectors consider time delays and spatio-temporal correlations between the road sections in urban road network. Habtemichael and Cetin (2016) use an enhanced K Nearest Neighborhood algorithm for traffic volume forecast. Zheng and Su (2014) use a principle component technique to enhance the performance of the K-Nearest Neighborhood approach in forecasting traffic volume.

2.2. Machine learning

Machine learning techniques for traffic prediction primarily include Neural Network, Deep Learning, Support Vector Regression, and the recently proposed Spinning Network. Many types of Neural Networks have been developed for the purpose of traffic prediction, such as Back Propagation Neural Networks (BPNN) (Guo and Zhu, 2009; Liu et al., 2012; Zhu et al., 2010), Time Delay Neural Networks (TDNN) (Abdulhai et al., 2002), Elman Neural Network (ENN) (Gao et al., 2008) and Radial Basis Function (RBF) Network (Wang and Xiao, 2003; Yang et al., 2010).

Qiao et al. (2001) propose a neural network-based system identification approach to establish an auto-adaptive model for simulating traffic flow dispersion and use it for online traffic flow forecasting. Zheng et al. (2006) develop a Bayesian combined Neural Network approach for freeway traffic flow prediction. Yin et al. (2002) develop a fuzzy-neural model for online rolling training and prediction. This fuzzy-neural model consists of two modules: a gate network is used to classify input data into clusters, whereas, an expert network is used to determine the input-output relationship. Vlahogianni et al. (2005) develop a genetic approach to optimize multilayer Neural Network structure for traffic flow prediction for an urban signalized arterial. Zhu et al. (2014) adopt RBF Neural Networks for traffic volume forecast at a single location while considering traffic flows of the adjacent intersections. Dunne and Ghosh (2012) develop a Neural Network model with predetermined uncongested and congested regimes to perform traffic flow and speed prediction. Laña et al. (2019) propose evolving spiking neural networks for adaptive long-term traffic state estimation.

Polson and Sokolov (2017) develop a deep learning model to predict traffic flows, where the first layer identifies spatio-temporal relations among predictors and other layers model nonlinear relations. It is reported that the deep learning architectures can capture the nonlinear spatial-temporal effects due to recurrent and non-recurrent traffic congestion patterns. Wu et al. (2018) propose a deep Neural Network-based traffic flow prediction model that consists of an attention-based network, convolutional neural network, and recurrent neural network. Three networks are used to determine the importance of past traffic flow, mine the spatial features and temporal features of traffic flow respectively. Do et al. (2019), and Zhang et al. (2019) respectively propose a deep learning neural network for traffic flow prediction with attention mechanism to exploit spatial and temporal dependencies between road segments as well as time steps.

Wu et al. (2004) propose SVR (support vector regression) for travel-time prediction using real highway traffic data. Wang and Shi (2013) propose a wavelet kernel function for SVR model to perform traffic speed forecasting, because the wavelet kernel function could capture both stationary and nonstationary, nonlinear characteristics of traffic speed data. In addition, Huang and Sadek (2009) develop a SPN (spinning network) model that is inspired by and mimics human memory mechanism for interstate highway traffic volume forecast.

In addition to aforementioned two categories of methodologies, some fusion methods are proposed for traffic state prediction. Antoniou et al. (2013) propose a framework that includes a set of machine-learning approaches that perform three functions including clustering and classification, modeling the evolution of traffic states, and state-specific speed prediction. Guo et al. (2018) evaluate fusion-based frameworks using three different fusion strategies: averaged, weighed and k-Nearest Neighbor methods, applied to three different machine learning methods, Neural Networks, Support Vector Regression and Random Forests. It is worth mentioning that there is a class of models for estimating traffic state at any intermediate point from the boundary conditions of a segment of urban roadway or highway based on macroscopic traffic flow theory. Some representative works include, but are not limited to, the LWR partial differential equation (Lighthill and Whitham, 1955; Richards, 1956), the CTM (cell transmission model) (Daganzo, 1994; Munoz et al., 2003), second-order traffic flow model with Kalman filter (Nanthawichit et al., 2003), or with Lagrangian measurements (Herrera and Bayen, 2010), kinematic wave model (Daganzo, 2005; Newell, 1993), stochastic Newell's three-detector method (Deng et al., 2013; Laval et al., 2012).

Overall, most of the aforementioned methodologies in the literature focus on local prediction. Moreover, the residual time series are usually not checked and verified by statistical tests to be a white noise. It is likely that traffic features still exist in the residuals. Hence, it is necessary to introduce a post-process mechanism for the residual time series of machine learning algorithms to make sure no traffic features remain in the residuals. To this end, we select the statistical ARIMA model as a post-processor to achieve this goal. The details of the working principles are described in Section 3.

3. Methodology

This section elaborates the theoretical background and working mechanism of the methodologies used in the study, including Neural Network, MSVR, ARIMA, NN-ARIMA, and MSVR-ARIMA.

3.1. Neural network

Neural Network models have found widespread applications in short-term traffic forecasting as described in the literature review. Many different types of Neural Network models have been developed. A multilayer perceptron (MLP) Neural Network is chosen for traffic state prediction in this study. Apart from the reason stated in the introduction section, it is chosen not only due to its simple architecture, but also mainly because it is a universal approximator that is capable to approximate any nonlinear relationship between pairs of input and output data (Hornik et al., 1989). The Kolmogorov theorem (1957) provides a solid theoretical basis in this regard (Tikhomirov, 1991).

It has been proven that a single hidden layer MLP feed-forward Neural Network can approximate any bounded continuous and multivariate function with arbitrary precision. Networks with more than one hidden layer can be converted to an equivalent network with just one hidden layer (Cybenko, 1989; Hornik et al., 1989). Therefore, one hidden layer architecture is employed for MLP Neural Network in this study.

Fig. 5 shows a typical MLP Neural Network architecture for traffic time series forecast. It consists of three layers, including input, hidden, and output layer. The specification of the number of neurons in the input layer and output layer are determined by the number of variables required by the subject problem. Similar to a regression model, the number of neurons in the input layer is determined by the number of regressors taken into account, and the output layer contains the number of neurons equivalent to the number of response variables. A sliding time window technique is used to set up pairs of input and output data. The p in the input layer denotes the number of lag time intervals being considered. The number of neurons in the hidden layer is usually chosen from a rule of thumb. This issue and the overfitting problem are elaborated in the following subsection. The thickness and the color of the links indicates the strength and sign of the connection between two neurons.

Corresponding to the MLP Neural Network architecture in Fig. 5, the mathematical form of MLP can be written as Eq. (1).

$$Y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} Y_{t-i} + \beta_{0j}\right) + \varepsilon_t \tag{1}$$

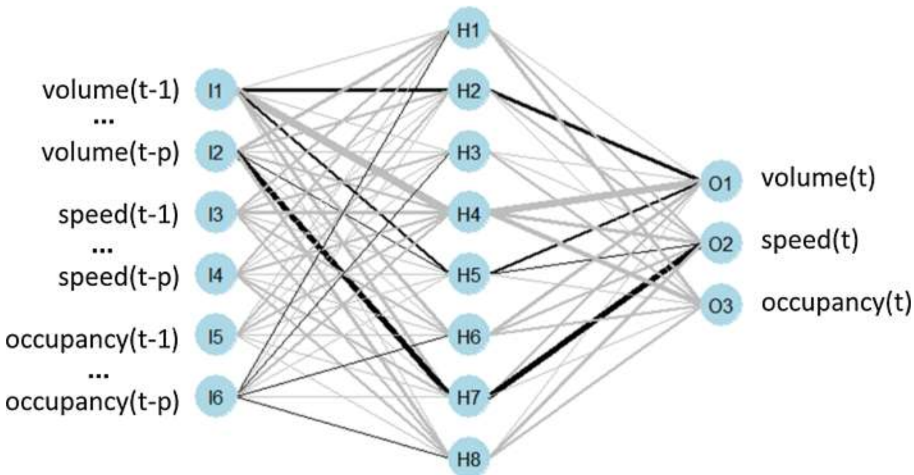

Fig. 5. A typical MLP Neural Network architecture for traffic time series forecast.

where Y_t represents a response variable in the output layer, Y_{t-i}, $i = 1, \cdots, m$ denotes explanatory variables in the input layer, m is the number of input neurons, n is the number of hidden neurons, f is a nonlinear activation function, α_j $j = 1, \cdots, n$ is a vector of weights from the hidden to output neurons and β_{ij}, $i = 1, \cdots, m$; $j = 1, \cdots, n$ are weights from the input to hidden neurons. α_0 and β_{0j} are the threshold values.

3.2. Multi-dimensional Support Vector regression

In order to verify the effectiveness of the proposed approach, Multi-dimensional Support Vector Regression (MSVR) is introduced to replace the role of Neural Network. In the literature, Support Vector Regression (SVR) is widely used in time series prediction (Cortes and Vapnik, 1995; Drucker et al., 1997; Mattera and Haykin, 1999; Müller et al., 1997) and traffic prediction (Yang and Lu, 2010).

SVR is a kernel-based machine learning algorithm. The pairs of input and output data are mapped into a high-dimensional feature space through a nonlinear transformation function, and a functional dependency is determined between the pairs of input and output in the high-dimensional feature space through quadratic optimization technique. However, the nonlinear mapping function usually is not known. Thanks to Mercer's theorem (Mercer, 1909), a kernel function can serve for this purpose. The linear regression function in feature space takes the form as Eq. (2).

$$f(x) = \langle w, x \rangle + b \tag{2}$$

where $w \in \chi$, $b \in R$ are hyperplane regression parameters to be determined. The solution for w and b can refer to Bertsekas (1997), Fletcher (2013), Herbrich (2001), Schölkopf et al. (2002). However, SVR is only limited to 1-dimensional output. Pérez-Cruz et al. (2002) generalized the Support Vector Machines (SVM) to solve multiple-input multiple-output problems for regression estimation and function approximation. The solution of the MSVR for multi-dimensional regression parameters w and b is obtained from an iterative procedure over the Karush-Kuhn-Tucker conditions.

3.3. Residuals of neural network and MSVR

Along with the rapid development of artificial intelligence, machine learning technique is often used as a panacea for large-scale data process, but some useful information hidden in the residuals is neglected. In the literature of traffic prediction, it is seldom to see the residuals out of machine learning algorithms being checked for whiteness. However, it is found that the residuals after machine learning algorithms often contain some valuable information.

Neural Network cannot explicitly identify the temporal correlation in the same way as a statistical time series model does. Hence, many residual series from neural network algorithm often appear to be temporally correlated instead of being a white noise. Taking location 6 of the network as an example, Fig. 6(a) shows ACF (autocorrelation function) plot of the residual time series out of a neural network; Fig. 6(b) shows ACF plot of the same residual time series out of a MSVR model. Both plots indicate that there are still significant correlations presented in the residual series with a repeating pattern (seasonality). Hence, it motivates us to employ ARIMA model to further capture the local traffic features within the residual series that are not yet captured by the network-scale model.

3.4. ARIMA

ARIMA model is a univariate linear model. It has been widely applied to traffic flow forecasting for a single location in urban areas for a few decades, but it has limited accuracy due to the fact that it lacks the capacity to capture nonlinear properties of traffic time series. However, the advantages of ARIMA models are its simple mathematical form and capacity to explicitly identify the temporal

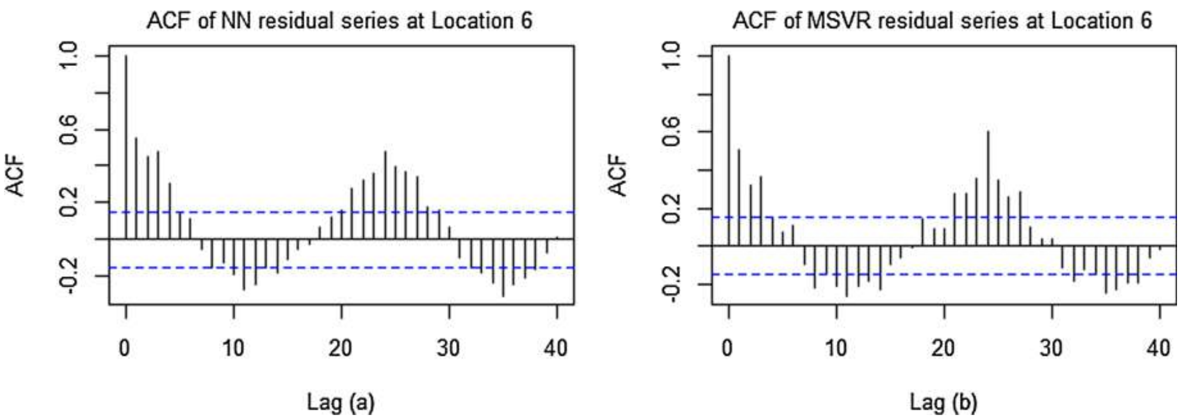

Fig. 6. ACF plot of residual series from Neural Network (a), and MSVR (b).

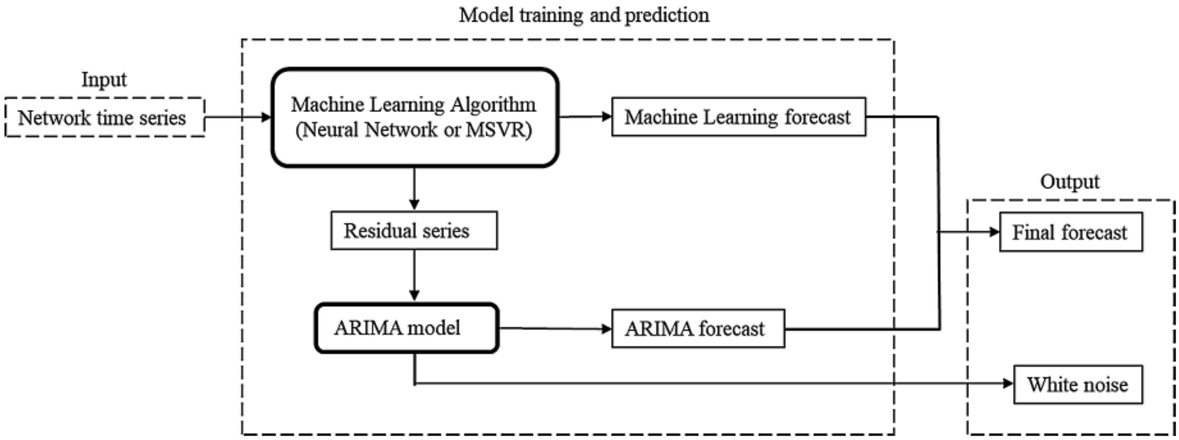

Fig. 7. Flow chart of the proposed approach.

correlation in the time series. The mathematical form of ARIMA(p, d, q) model can be written as Eq. (3) (Brockwell, 2016).

$$\phi(B)(1-B)^d X_t = \theta(B)e_t \tag{3}$$

where X_t is time series variable, $\{e_t\}$~WN(0, σ^2), $\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + ... + \theta_q B^q$ are pth and qth-degree polynomials, ϕ_1, ...,ϕ_p and θ_1, ...,θ_q are coefficients, d denotes the order of difference, B is the backward shift operator, p and q represent the autoregressive and moving average order, respectively.

3.5. Proposed approach: NN-ARIMA and MSVR-ARIMA

The essence of the proposed approach is to sequentially concatenate machine learning algorithms with ARIMA algorithm for network-wide traffic prediction. Machine learning algorithms aim to take care of the network-scale traffic flow features and co-movement patterns, whereas, ARIMA aim to exhaust the rest of traffic features in the residuals of the machine learning algorithms if any exists. As ARIMA examines each individual residual time series, it therefore captures location-specific traffic characteristics.

Fig. 7 shows the flow chart of the proposed approach. The network-wide observed traffic flow time series enter machine learning algorithms as input. The machine learning model is trained and validated through a supervised learning process. The spatial and temporal traffic features are saved in a weight matrix of the model. With future new input, the trained machine learning model is then used to produce network-wide predictions, as well as residuals for all locations. Subsequently, each residual time series is used to fit an ARIMA model, which in turn produces location-specific prediction and a white noise. Finally, the predictions from the machine learning algorithms and ARIMA models are added up to yield the final forecast for each location.

In summary, the proposed approach that fuses a machine learning algorithm with a statistical time series model is expected to be able to mutually complement each other and capture traffic flow complex patterns, sharply fluctuated nonlinearity, and take full advantage of network-scale correlations to improve accuracy of prediction. The hybrid model can be written in a mathematical form as Eq. (4).

$$Z_i(t) = Y_i(t) + X_i(t) + e_i(t) \tag{4}$$

where $Z_i(t)$ represents the observation of a time series at location i at time t, $Y_i(t)$ is the forecast of the machine learning algorithm, $X_i(t)$ is the forecast of the ARIMA model, and $e_i(t)$ is white noise. Respectively, the theoretical basis and working mechanism of the component $Y_i(t)$ from the machine learning algorithms, as well as the component $X_i(t)$ from the ARIMA model, are described in detail in the Sections 3.1, 3.2, and 3.4 respectively. The machine learning algorithms in the studies include both neural network and multiple output support vector regression algorithms, which end up with two hybrid NN-ARIMA and MSVR-ARIMA. Although the MSVR-ARIMA is not the one we finally recommend for network-wide traffic prediction due to its less competitive performance than NN-ARIMA, we keep it in the study so that two models with similar structure can be used for cross verification of the effectiveness of the proposed postprocessing.

4. Case study

4.1. Study area and data collection

Fig. 8 shows the map of the study area. The locations of the loop detectors are color-coded as blue dots on the map. The data set for this study includes hourly traffic volume, speed, and occupancy time series collected from 44 double loop detectors along the Highway Ayalon in Tel Aviv, Israel. There are totally 132 traffic time series (three measurements from each detector).

Fig. 9 includes the plots of traffic volume, speed, and occupancy time series respectively. Each plot contains 44 time series from 44

Fig. 8. Map of study area with 44 locations of loop detectors.

locations. Each time series used in the study is an approximately 6-month-long time series that contains 4512 data points.

4.2. Model estimation

4.2.1. NN-ARIMA

There is no consistent guidance, but many rules of thumb are available for determining the number of hidden layers, neurons in a hidden layer, and training iterations, and preventing overfitting. A smart choice of the number of hidden layers and neurons depends on what form of regularization is being used (SARLE, 1997). In this study, a cross-validation technique with an early stopping policy is employed to find a near optimal number of neurons in the hidden layer and number of training iterations for weight matrix optimization. The dataset is divided into three non-overlapping subsets, including training, validation, and testing set. The training set is used for model estimation. The generalization error is estimated from the validation set. The test set is used for assessing predictive performance. Cross-validation and early stopping policy also prevent Neural Network from overfitting via monitoring the generalization error to decide when to stop the training process. If the generalization error shows no further improvement or increases after a certain number of training iterations, then the training process terminates. The number of neurons in the hidden layer and the number of iterations that yield the minimum generalization error determine the Neural Network architecture. Fig. 10 shows that the training of a Neural Network with 118 neurons in its hidden layer should stop at the 190th iteration.

4.2.2. MSVR-ARIMA

The code of multi-dimensional support vector regression algorithm developed by Fernando Pérez-Cruz (Sanchez-Fernandez et al.,

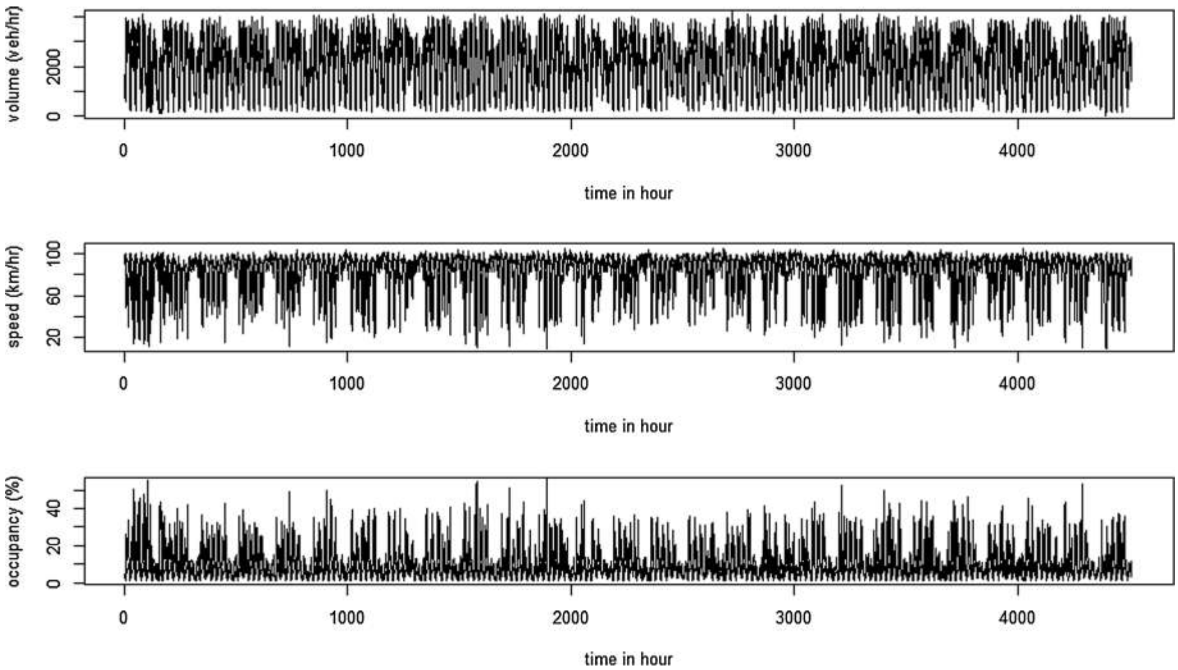

Fig. 9. Time series plots of traffic volume, speed, and occupancy from 44 locations.

2004; Tuia et al., 2011) is used for network-wide prediction. The gamma (γ), epsilon (ε), and cost factor C are the key hyperparameters for the MSVR algorithm. Gamma (γ) is the key kernel parameter that needs to be chosen carefully, as it implicitly defines the structure of the high dimensional feature space and thus controls the complexity of the final solution. The parameter C is a general penalizing parameter indicating the cost of constraints violation. It is referred to as the regularized constant that determines the tradeoff between the empirical error and the regularized term in the Lagrange formulation. The ε in the insensitive-loss function is called the tube size of MSVR and is equivalent to the approximation accuracy placed on the training data. The cross-validation method and a grid search process is executed to find near optimal value for the set of hyperparameters based on training data set. The radial basis function (RBF) is chosen as the kernel function. It is highly effective in mapping nonlinear relationships.

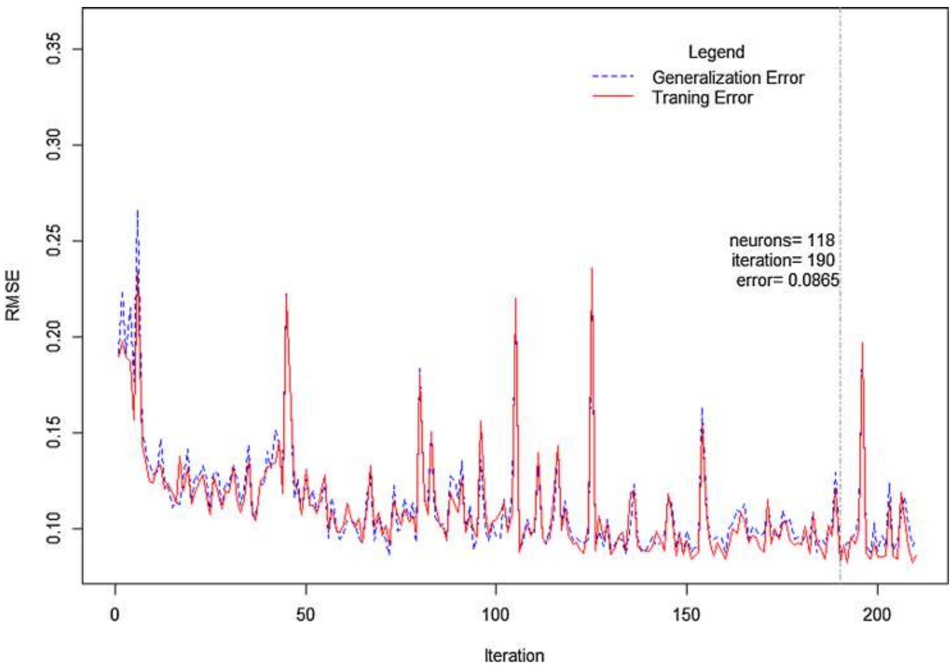

Fig. 10. Cross-validation method to control Neural Network overfitting problem.

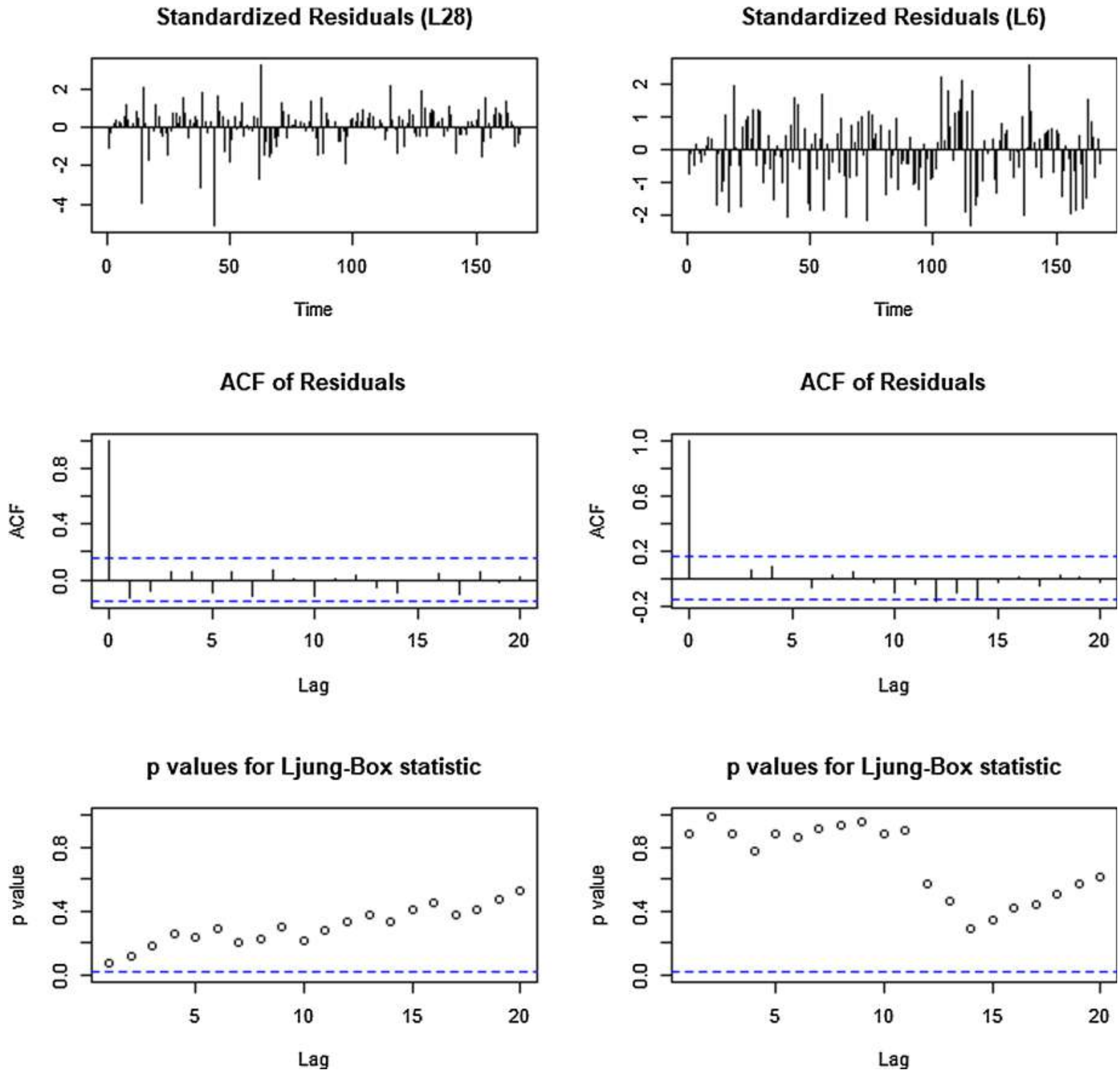

Fig. 11. Residuals, auto-correlogram, and *p*-values of Ljung-Box test.

4.2.3. ARIMA

The ARIMA models for location-specific residual time series are estimated using the package *forecast* in R. The specification of ARIMA models can be determined by the function *auto.arima*. The forecast is performed with the function *forecast*. Fig. 11 shows two residual time series at location 6 and 28 after the ARIMA analysis and their ACF plot. It indicates that no significant correlation exists in the residual series. The result of Ljung-Box test provides further evidence that the residual series after ARIMA analysis is a white noise.

The residual series at each location is verified to be a white noise. Using location 6 and 28 as typical representatives, Fig. 11 provides three plots: standardized residual series, autocorrelogram, and *p*-values of Ljung-Box statistics within 20 lags. From the standardized residual plot, it can be seen that the residuals are randomly and evenly distributed around mean zero, and no apparent patterns appear along time. The ACF plots show no significant correlation among lags, which indicates that traffic features in the time series is adequately captured by the model. The *p*-values of the Ljung-Box test are all greater than 0.05, therefore, the null hypothesis of independence cannot be rejected at 5% significance level, which means that the residual series are white noise.

4.3. Forecast and comparison

In the study, one-step-ahead hourly traffic forecast is performed with the Neural Network, NN-ARIMA, MSVR, MSVR-ARIMA, and ARIMA, respectively. The full data set contains 6-month-long traffic time series including volume, speed, and occupancy. In order to

Table 1
Comparison of network model vs. postprocessed network model, i.e. Neural Network vs. NN-ARIMA, MSVR vs. MSVR-ARIMA.

Time	Index	Neural Network MSE	NN-ARIMA MSE	Error reduction %	MSVR MSE	MSVR-ARIMA MSE	Error reduction %
1-month-long time series							
Jan	1	145271.90	123985.10	14.65	524210.80	373861.50	28.68
Feb	2	161343.30	152089.50	5.74	434354.00	382466.50	11.95
Mar	3	224953.60	194828.40	13.39	541142.80	443739.80	18.00
Apr	4	226831.40	205133.00	9.57	590176.40	453669.20	23.13
May	5	224092.40	193829.40	13.50	603194.10	440682.70	26.94
Jun	6	204024.90	157607.40	22.75	551013.70	414793.00	24.72
	Average	197752.92	171245.47	13.40	540681.97	418202.12	22.65
3-month-long time series							
Jan-Mar	1	153970.60	146397.10	4.92	496003.00	365846.60	26.24
Feb-Apr	2	257422.80	224846.10	12.65	558900.90	487160.90	12.84
Mar-May	3	126888.10	116636.80	8.08	534635.40	371007.80	30.61
Apr-Jun	4	168635.80	156099.80	7.43	591379.20	427471.10	27.72
	Average	176729.33	160994.95	8.90	545229.63	412871.60	24.28
6-month-long time series							
Jan-Jun	1	177693.50	159159.40	10.43	617371.00	505687.70	18.09

test the robustness of the proposed approach, we create 11 scenarios with the 6-month-long full data set. The experiment is repeated for 6 times with monthly data sets, 4 times with 3-month-long rolling data sets, and 1 time with the full data set.

For each scenario we perform forecasts for 168 time points, which is equivalent to 7 days of hourly traffic states in a row. The predictive performance of each model is assessed according to the accuracy of those seven days of predictions. The MSE (mean squared error) and MAPE (mean absolute percentage error) are adopted as the key measurements of prediction accuracy. Statistically, MSE is a second order moment that contains information on both bias and variance.

4.3.1. Network model vs. postprocessed network model

To confirm the effectiveness of the postprocess approach, two comparisons are carried out: (1) a standalone Neural Network vs. NN-ARIMA, (2) a standalone MSVR vs. MSVR-ARIMA in term of the accuracy of their predictions. Table 1 summarizes the results of the MSE values produced by the four models. It can be seen that the MSE values produced by the NN-ARIMA are consistently smaller than the ones by a standalone Neural Network in all scenarios.

The average error reduction is 13.40% for 1-month-long time series data, 8.90% for 3-month-long time series data, and 10.43% for the full data set. Similar pattern, but in a greater magnitude, can be seen in the comparison of a standalone MSVR vs. MSVR-ARIMA. The average error is reduced by 22.65% for 1-month-long time series data, 24.28% for 3-month-long time series data, and 18.09% for the full data set. According to these experimental results, the postprocess for a machine learning algorithm is necessary.

4.3.2. Individual location vs. network-wide prediction

To verify the advantage of the proposed approach that considers both network-scale traffic co-movement features and location-specific characteristics, two comparisons are carried out: (1) ARIMA vs. NN-ARIMA, (2) ARIMA vs. MSVR-ARIMA. Table 2 summarizes the results of comparisons. It shows that the MSE values produced by the NN-ARIMA are consistently smaller than the ones

Table 2
Comparison of individual location prediction vs. network-wide prediction, i.e. ARIMA vs. NN-ARIMA, ARIMA vs. MSVR-ARIMA.

Time	Index	ARIMA MSE	NN-ARIMA MSE	Improvement %	MSVR-ARIMA MSE	Improvement %
1-month-long time series						
Jan	1	479840.60	123985.10	74.16	373861.50	22.09
Feb	2	511449.00	152089.50	70.26	382466.50	25.22
Mar	3	497644.00	194828.40	60.85	443739.80	10.83
Apr	4	503061.60	205133.00	59.22	453669.20	9.82
May	5	477643.30	193829.40	59.42	440682.70	7.74
Jun	6	536106.10	157607.40	70.60	414793.00	22.63
	Average	500957.43	171245.47	65.82	418202.12	16.52
3-month-long time series						
Jan-Mar	1	382151.40	146397.10	61.69	365846.60	4.27
Feb-Apr	2	531639.30	224846.10	57.71	487160.90	8.37
Mar-May	3	470289.40	116636.80	75.20	371007.80	21.11
Apr-Jun	4	549610.10	156099.80	71.60	427471.10	22.22
	Average	483422.55	160994.95	66.70	412871.60	14.59
6-month-long time series						
Jan-Jun	1	552672.50	159159.40	71.20	505687.70	8.50

Table 3
MAPE (%) of five prediction models.

Time	Index	Neural Network MAPE %	NN-ARIMA MAPE %	MSVR MAPE %	MSVR-ARIMA MAPE %	ARIMA MAPE %
1-month-long time series						
Jan	1	12.26	10.92	23.46	19.92	24.14
Feb	2	9.49	9.78	20.99	18.95	21.26
Mar	3	10.71	9.03	21.42	18.48	19.44
Apr	4	14.27	13.15	27.05	23.80	23.33
May	5	13.10	12.25	26.16	21.58	22.11
Jun	6	11.38	11.38	23.22	21.36	20.96
	Average	11.87	11.08	23.72	20.68	21.87
3-month-long time series						
Jan-Mar	1	9.54	9.05	22.18	19.22	19.26
Feb-Apr	2	11.28	11.49	23.14	21.09	20.53
Mar-May	3	9.50	9.40	22.62	19.03	20.18
Apr-Jun	4	11.30	10.85	22.82	20.46	20.46
	Average	10.41	10.20	22.69	19.95	20.11
6-month-long time series						
Jan-Jun	1	12.41	11.11	22.94	21.74	20.44

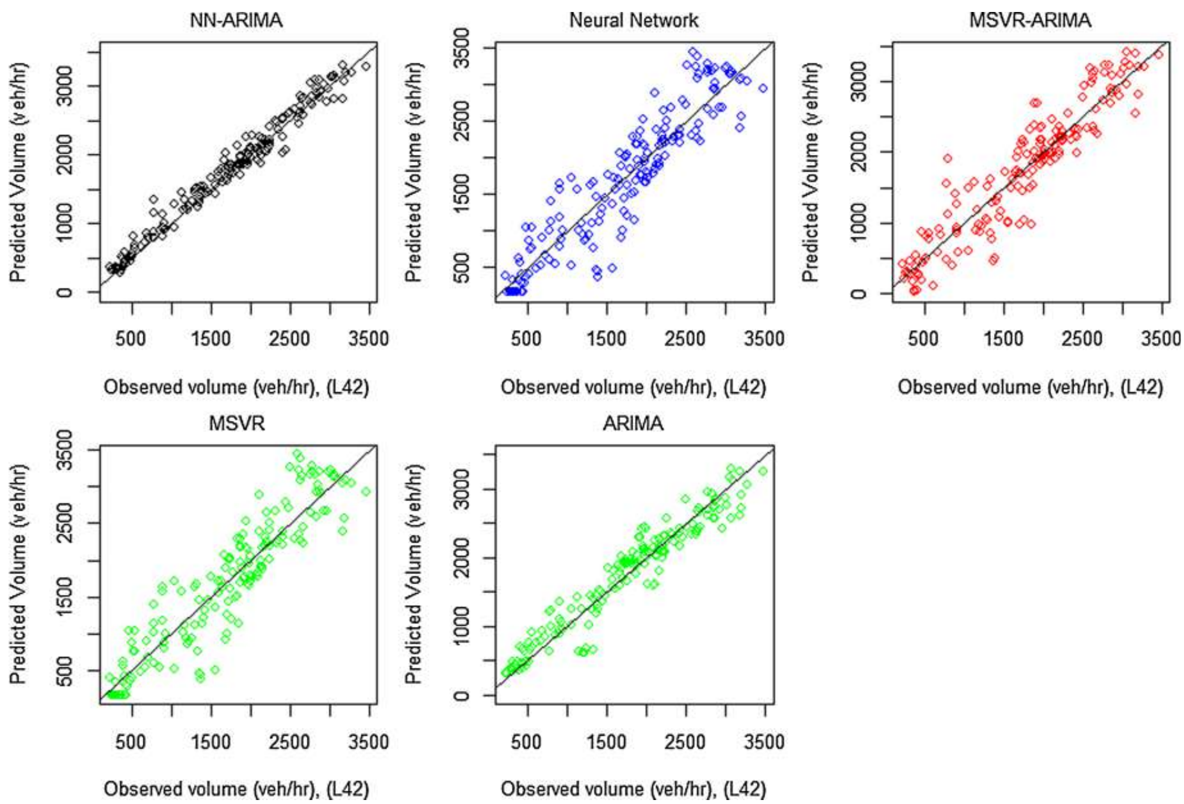

Fig. 12. Observation vs. prediction of traffic volume at location 42.

by ARIMA in all scenarios. The average reduction of the MSE value is 65.82% for 1-month-long data, 66.70% for 3-month-long data, and 71.20% for the full data set. Similar pattern, but in a smaller magnitude, can be found in the comparison of ARIMA vs. MSVR-ARIMA. The average MSE value is reduced by 16.52% for 1-month-long data, 14.59% for 3-month-long data, and 8.5% for the full data set. The results suggest that the approach of treating network-wide traffic flows as a whole provides more accurate prediction with smaller mean squared error.

If we compare NN-ARIMA with MSVR-ARIMA, it can be seen that the MSE values produced by NN-ARIMA are significantly smaller than the ones by MSVR-ARIMA in all scenarios. The average MSE value is improved by 59.05% for 1-month-long data, 61.00% for 3-month-long data, and 68.53% for the full data set. These three numbers are not explicitly shown in the Table 2.

The results of the mean absolute percentage errors produced by five models are presented in Table 3 for further reference. The overall results in Table 1–3 indicate that the NN-ARIMA approach outperforms all other four methods for traffic state prediction.

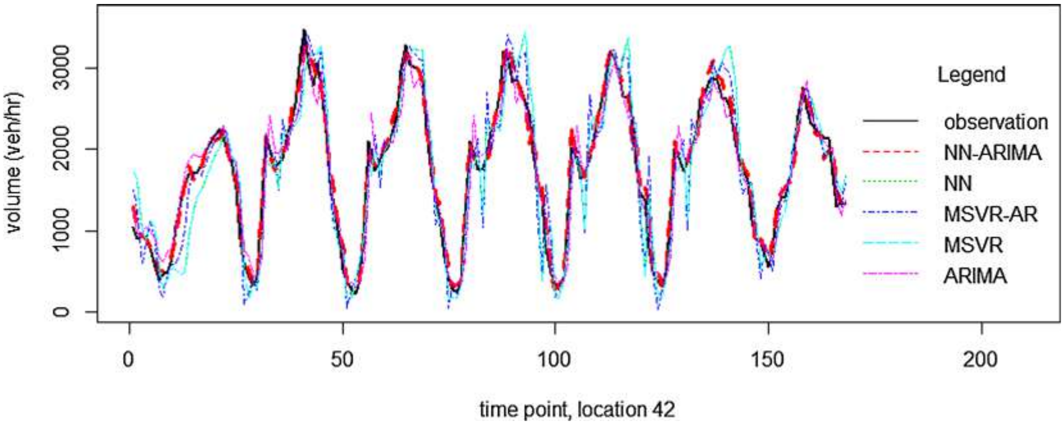

Fig. 13. Observed vs. predicted volume by five models for 168 hourly horizons.

4.3.3. Forecast

The forecast is performed for three macroscopic traffic variables respectively (i.e. volume, speed, and occupancy). Apart from the numerical results in Table 1–3, we use two different ways to visualize the result of prediction. Fig. 12 shows the scatter plots of observed volume vs. predicted volume by five models respectively, using the data from location 42. In each plot, if a forecast value exactly matches the observed value, the dot should be exactly on the 45-degree diagonal line. If a forecast value is greater than the observed value, the dot is above the diagonal line, and vice versa. The further the dots deviate from the diagonal line, the less accurate the forecast value is, and the bigger the mean squared error of the prediction is. If the majority of the dots lie on one side of the line, it means that the model produces forecast with bias. From the plots, it can be seen that the predictions of the NN-ARIMA model are evenly distributed on both sides of the diagonal line and concentrate on the line with very narrow deviation. The predictions by Neural Network, MSVR, or MSVR-ARIMA are almost evenly distributed on both sides of the diagonal line, but with bigger deviation. The predictions by ARIMA model not only have more points on one side of the diagonal line, but also have bigger deviation.

Fig. 13 shows time series plot of observed vs. predicted traffic volume using the outcome from location 42. The lines of predictions show how close each prediction to the observed value and how the predictions evolve over time. It can be seen that the prediction by NN-ARIMA model in the red dot line matches the line of observations better than the other models. Furthermore, the plots of observed traffic volume, speed, and occupancy vs. their predicted values by the proposed approach NN-ARIMA for 44 locations are included in Appendix-A from Figs. A1–A9.

5. Conclusion and discussion

In the study, we proposed a novel approach that sequentially concatenates a machine learning algorithm with a statistical model, i.e. NN-ARIMA. This approach is able to consider both network-scale spatial-temporal correlations among traffic flows and location-specific traffic characteristics. Its postprocessing by the ARIMA analysis can extract traffic features from the residuals of Neural Network, hence, significantly improve the accuracy of prediction. In addition, this approach is also able to capture the sharp non-linearity of traffic flows.

In the context of using 6-month-long highway traffic time series data set, the case study provides numerical evidence that the predictive capacity of the proposed NN-ARIMA model outperforms MSVR-ARIMA, Neural Network, MSVR, and ARIMA. It is an effective and efficient tool for traffic state prediction.

It is noteworthy that the MSE is significantly reduced in the comparison of NN-ARIMA vs. ARIMA. It clearly reflects the important role of network-wide spatial correlations in traffic prediction. It is beneficial if the network traffic flows are treated as a whole in the prediction.

From the comparison of NN-ARIMA vs. Neural Network, it can be seen that postprocessing residuals is necessary and a warrant at least for the situation where the time series data are not sufficiently long. The postprocess significantly improves the accuracy of traffic state prediction.

In the future research, the proposed approach may be applied to traffic prediction for a large urban arterial road network if data is available. It may also be compared with those preprocess machine learning algorithms.

CRediT authorship contribution statement

Tao Ma: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Visualization. **Constantinos Antoniou:** Writing - review & editing, Software, Validation. **Tomer Toledo:** Resources, Writing - review & editing.

Acknowledgment

The authors are sincerely thankful to the municipality of Tel Aviv in Israel for providing traffic data for this research.

Appendix A

See Figs. A1–A9.

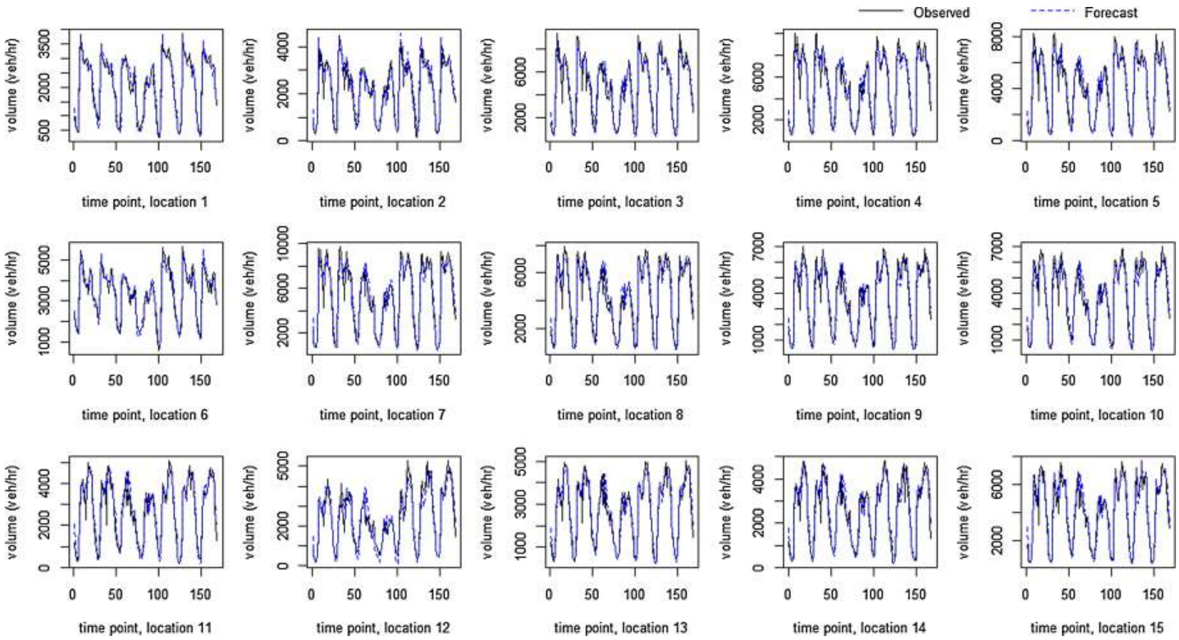

Fig. A1. Observed hourly traffic volume vs. predicted values for 168 hourly horizons at the Ayalon Highway in Tel Aviv, Israel (1–15).

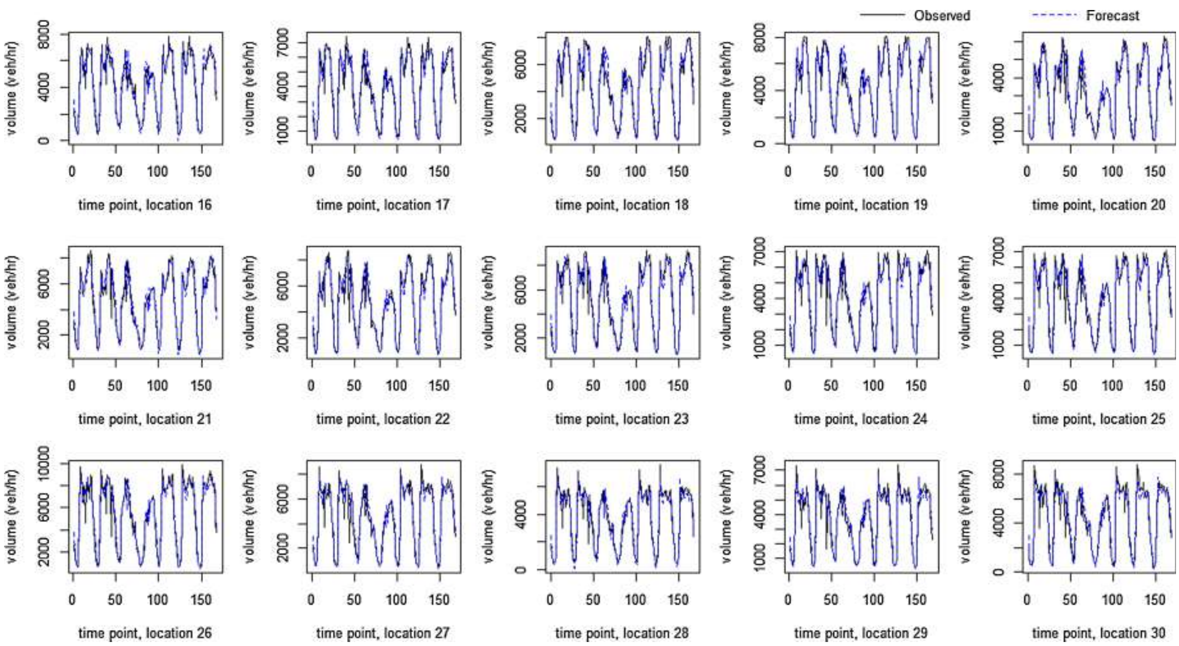

Fig. A2. Observed hourly traffic volume vs. predicted values for 168 hourly horizons at the Ayalon Highway in Tel Aviv, Israel (16–30).

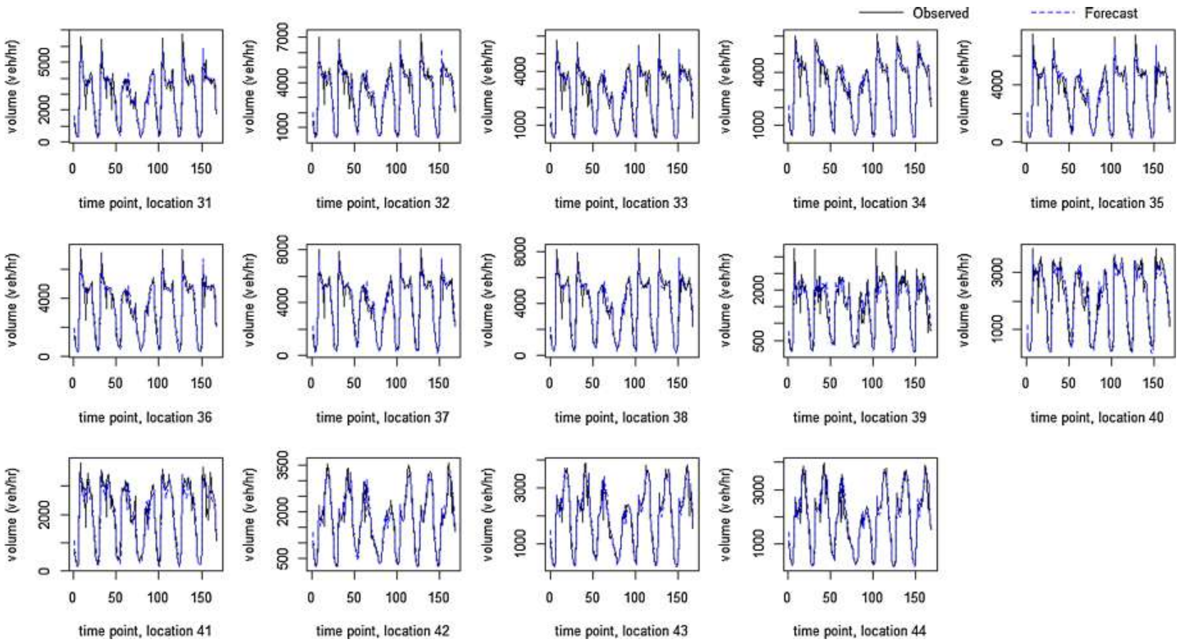

Fig. A3. Observed hourly traffic volume vs. predicted values for 168 hourly horizons at the Ayalon Highway in Tel Aviv, Israel (31–44).

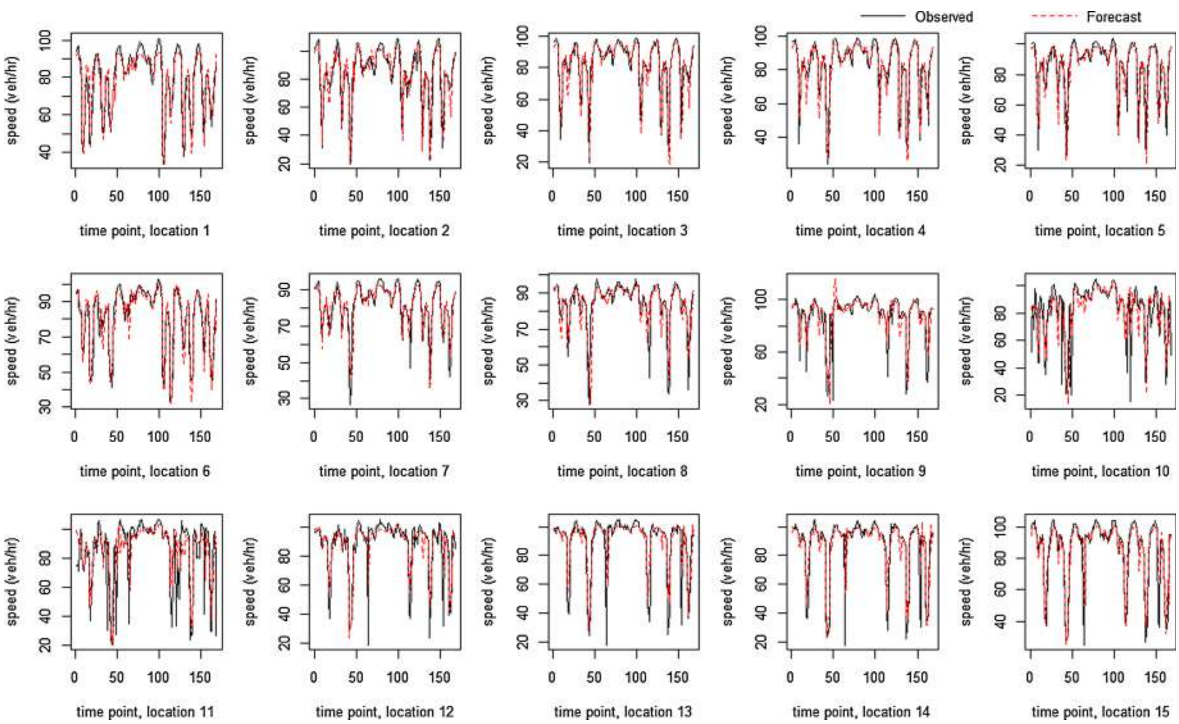

Fig. A4. Observed hourly traffic speed vs. predicted values for 168 hourly horizons at the Ayalon Highway in Tel Aviv, Israel (1–15).

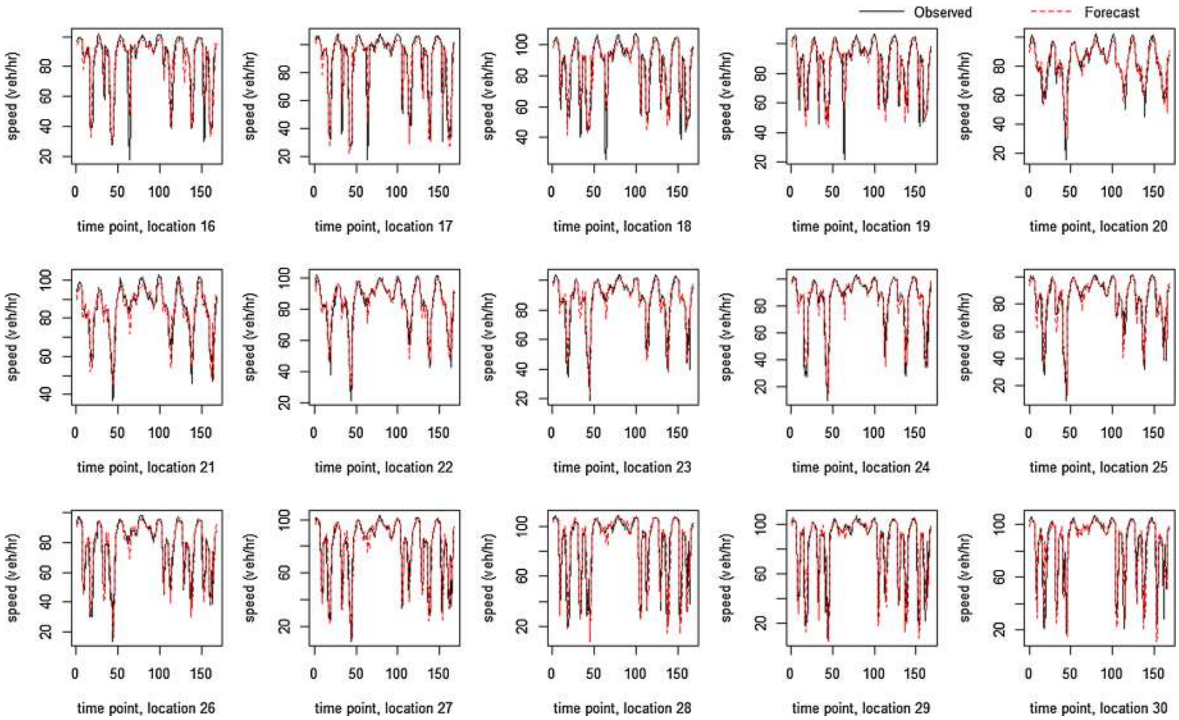

Fig. A5. Observed hourly traffic speed vs. predicted values for 168 hourly horizons at the Ayalon Highway in Tel Aviv, Israel (16–30).

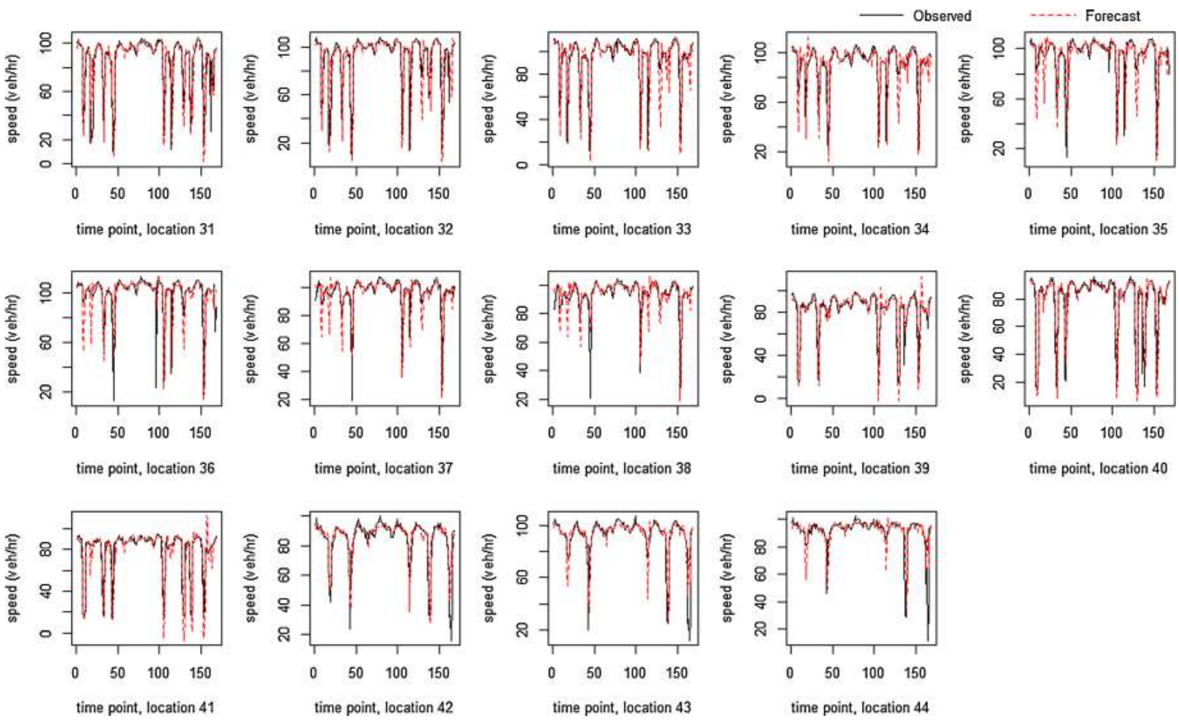

Fig. A6. Observed hourly traffic speed vs. predicted values for 168 hourly horizons at the Ayalon Highway in Tel Aviv, Israel (31–44).

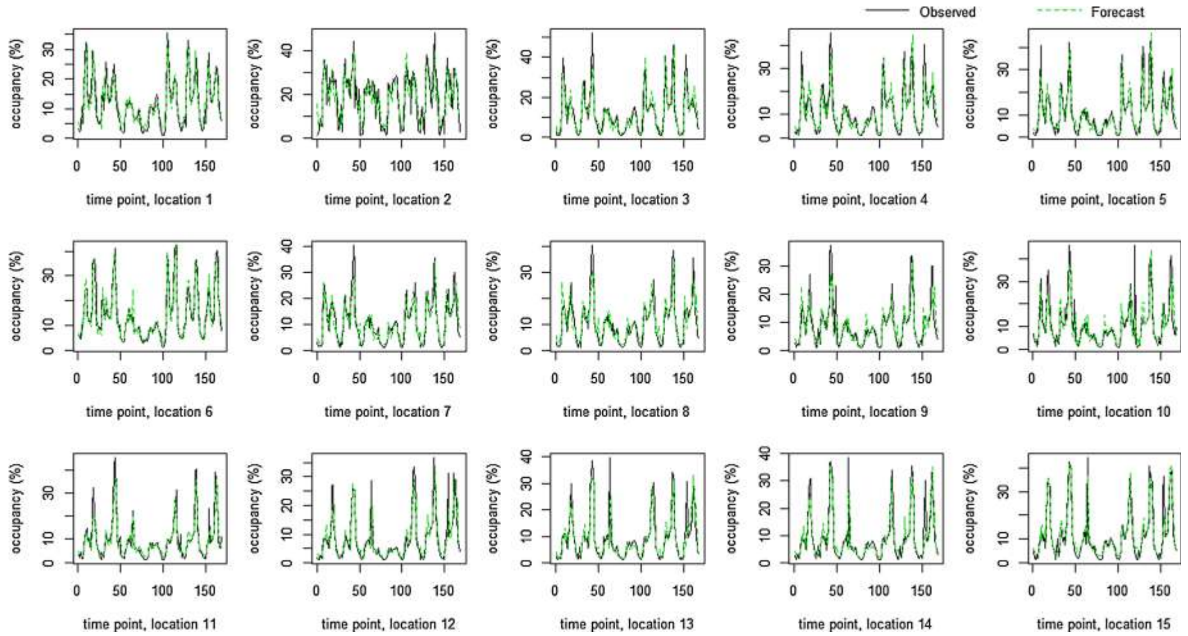

Fig. A7. Observed hourly traffic occupancy vs. predicted values for 168 hourly horizons at the Ayalon Highway in Tel Aviv, Israel (1–15).

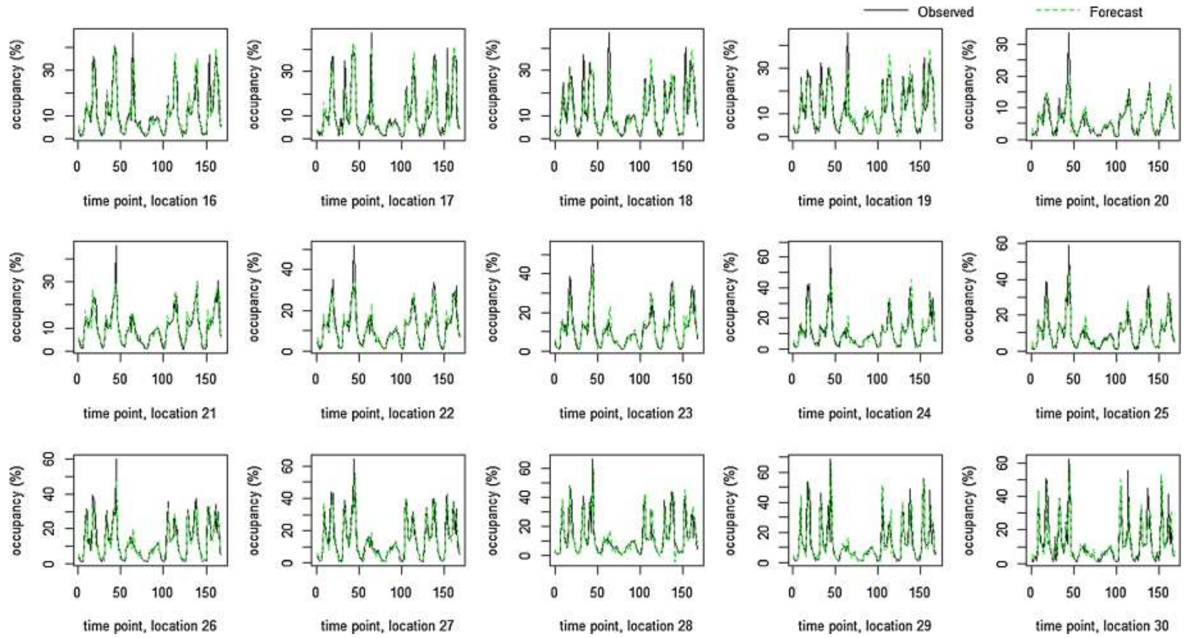

Fig. A8. Observed hourly traffic occupancy vs. predicted values for 168 hourly horizons at the Ayalon Highway in Tel Aviv, Israel (16–30).

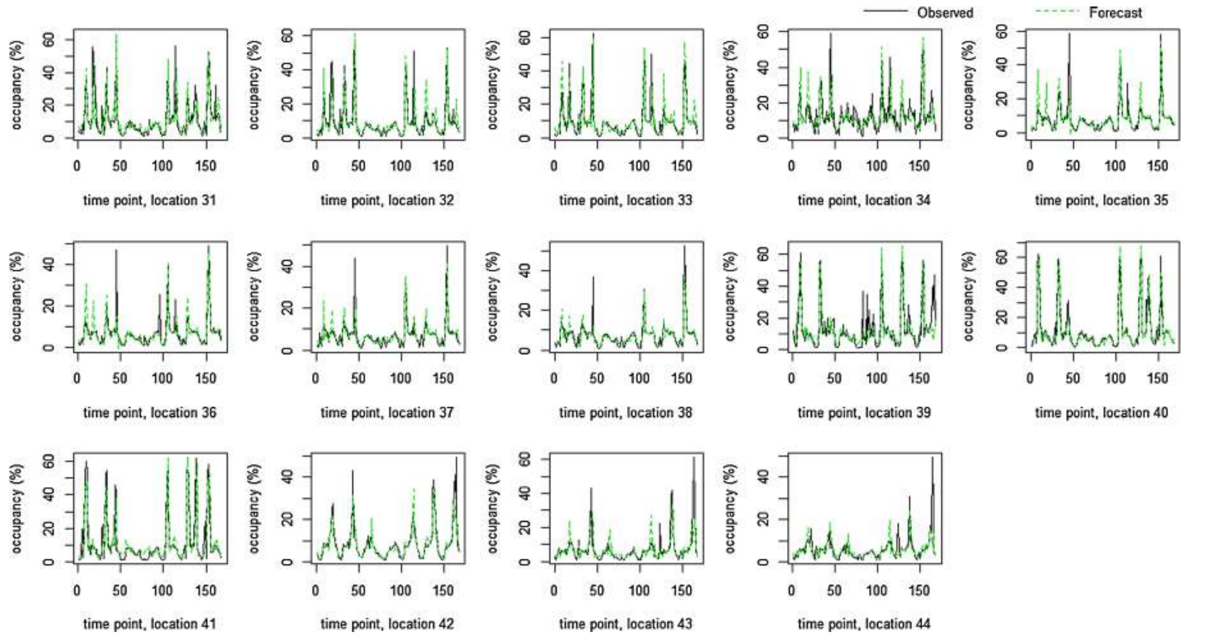

Fig. A9. Observed hourly traffic occupancy vs. predicted values for 168 hourly horizons at the Ayalon Highway in Tel Aviv, Israel (31–44).

References

Abdulhai, B., Porwal, H., Recker, W., 2002. Short-term traffic flow prediction using neuro-genetic algorithms. J. Intelligent Transp. Syst. 7, 3–41.

Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N., 2007. Nonlinear Kalman filtering algorithms for on-line calibration of dynamic traffic assignment models. IEEE Trans. Intelligent Transp. Syst. 8, 661–670.

Antoniou, C., Koutsopoulos, H.N., Yannis, G., 2013. Dynamic data-driven local traffic state estimation and prediction. Transp. Res. Part C: Emerg. Technol. 34, 89–107.

Bellemans, T., Schutter, B.D., Wets, G., Moor, B.D., 2006. Model predictive control for ramp metering combined with extended Kalman filter-based traffic state estimation. In: 2006 IEEE Intelligent Transportation Systems Conference, pp. 406–411.

Bertsekas, D.P., 1997. Nonlinear programming. J. Operat. Res. Soc. 48, 334.

Bonett, D.G., Wright, T.A., 2000. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. Psychometrika 65, 23–28.

Brockwell, J.P., 2016. Introduction to Time Series and Forecasting. Springer Science + Business Media, New York, NY.

Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transp. Res. Rec. 2124, 127–136.

Castillo, E., Menéndez, J.M., Sánchez-Cambronero, S., 2008. Predicting traffic flow using Bayesian networks. Transp. Res. Part B: Methodol. 42, 482–509.

Cetin, M., Comert, G., 2006. Short-term traffic flow prediction with regime switching models. Transp. Res. Rec.: J. Transp. Res. Board 1965, 23–31.

Cheng, T., Haworth, J., Wang, J., 2012. Spatio-temporal autocorrelation of road network data. J. Geogr. Syst. 14, 389–413.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Control Signal Syst. 2, 303–314.

Daganzo, C.F., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transp. Res. Part B: Methodol. 28, 269–287.

Daganzo, C.F., 2005. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transp. Res. Part B: Methodol. 39, 187–196.

Dai, X., Fu, R., Zhao, E., Zhang, Z., Lin, Y., Wang, F.-Y., Li, L., 2019. DeepTrend 2.0: a light-weighted multi-scale traffic prediction model using detrending. Transp. Res. Part C: Emerg. Technol. 103, 142–157.

D'Angelo, M.P., Al-Deek, H.M., Wang, M.C., 1999. Travel-time prediction for freeway corridors. Transp. Res. Rec. 1676, 184–191.

Dauwels, J., Aslam, A., Asif, M.T., Zhao, X., Vie, N.M., Cichocki, A., Jaillet, P., 2014. Predicting traffic speed in urban transportation subnetworks for multiple horizons. In: 2014 13th International Conference on Control Automation Robotics Vision (ICARCV), pp. 547–552.

Deng, W., Lei, H., Zhou, X., 2013. Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach. Transportation Research Part B: Methodological 57, 132–157.

Disbro, J.E., and Frame, M. (1989). Traffic flow theory and chaotic behavior. New York State Department of Transportation Report FHWA (NY/SR-98/91, New York).

Do, L.N.N., Vu, H.L., Vo, B.Q., Liu, Z., Phung, D., 2019. An effective spatial-temporal attention based neural network for traffic flow prediction. Transportation Research Part C: Emerging Technologies 108, 12–28.

Dong, C., Shao, C., Richards, S.H., Han, L.D., 2014. Flow rate and time mean speed predictions for the urban freeway network using state space models. Transportation Research Part C: Emerging Technologies 43, 20–32.

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A.J., Vapnik, V., 1997. Support vector regression machines. In: Mozer, M.C., Jordan, M.I., Petsche, T. (Eds.), Advances in Neural Information Processing Systems 9. MIT Press, pp. 155–161.

Dunne, S., Ghosh, B., 2012. Regime-based short-term multivariate traffic condition forecasting algorithm. J. Transp. Eng. 138, 455–466.

Fletcher, R., 2013. Practical Methods of Optimization. John Wiley & Sons.

Fusco, G., Colombaroni, C., Isaenko, N., 2016. Short-term speed predictions exploiting big data on large urban road networks. Transp. Res. Part C: Emerg. Technol. 73, 183–201.

Gao, H., Zhao, J., Jia, L., 2008. 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, pp. 1305–1309.

Guo, X., Zhu, Q., 2009. A traffic flow forecasting model based on BP neural network. In: In 2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS), pp. 311–314.

Guo, F., Polak, J.W., Krishnan, R., 2018. Predictor fusion for short-term traffic forecasting. Transp. Res. Part C: Emerg. Technol. 92, 90–100.

Guo, J., Huang, W., Williams, B.M., 2014. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. Transp. Res. Part C: Emerg. Technol. 43, 50–64.

Habtemichael, F.G., Cetin, M., 2016. Short-term traffic flow rate forecasting based on identifying similar traffic patterns. Transp. Res. Part C: Emerg. Technol. 66, 61–78.

Hara, Y., Suzuki, J., Kuwahara, M., 2018. Network-wide traffic state estimation using a mixture Gaussian graphical model and graphical lasso. Transp. Res. Part C: Emerg. Technol. 86, 622–638.

Herbrich, R., 2001. Learning Kernel Classifiers: Theory and Algorithms. MIT Press.

Herrera, J.C., Bayen, A.M., 2010. Incorporation of Lagrangian measurements in freeway traffic state estimation. Transp. Res. Part B: Methodol. 44, 460–481.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Huang, S., Sadek, A.W., 2009. A novel forecasting approach inspired by human memory: The example of short-term traffic volume forecasting. Transp. Res. Part C: Emerg. Technol. 17, 510–525.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82, 35–45.

Kamarianakis, Y., Prastacos, P., 2005. Space–time modeling of traffic flow. Comput. Geosci. 31, 119–133.

Kamarianakis, Y., Shen, W., Wynter, L., 2012. Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. Appl. Stoch. Models Bus. Indust. 28, 297–315.

Laharotte, P.-A., Billot, R., El Faouzi, N.-E., Rakha, H.A., 2015. Network-wide traffic state prediction using bluetooth data. In: TRB 94th Annual Meeting Compendium of Papers, (Washington DC, United States), pp. 15–3022.

Laña, I., Lobo, J.L., Capecci, E., Del Ser, J., Kasabov, N., 2019. Adaptive long-term traffic state estimation with evolving spiking neural networks. Transp. Res. Part C: Emerg. Technol. 101, 126–144.

Laval, J., He, Z., Castrillon, F., 2012. Stochastic extension of Newell's three-detector method. Transp. Res. Rec.: J. Transp. Res. Board 2315, 73–80.

Lee, S., Fambro, D.B., 1999. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. Transp. Res. Rec. 1678, 179–188.

Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. Pro R. Soc. London A: Math., Phys. Eng. Sci. 229, 317–345.

Liu, T., Sun, X., Zhong, X., 2010. Short term traffic flow forecasting based on a three-regime SETAR model. J. Highway Transp. Res. Dev. China 27, 122–127.

Liu, Z., Yang, Z., Gao, P., 2012. Research on the short-term traffic flow prediction method based on BP neural networks. In: World Automation Congress 2012, pp. 1–4.

Ma, T., Zhou, Z., Abdulhai, B., 2015. Nonlinear multivariate time–space threshold vector error correction model for short term traffic state prediction. Transp. Res. Part B: Methodol. 76, 27–47.

Ma, T., Zhou, Z., Antoniou, C., 2018. Dynamic factor model for network traffic state forecast. Transp. Res. Part B: Methodol. 118, 281–317.

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., Wang, Y., 2017. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. Sensors 17, 818.

Mattera, D., Haykin, S., 1999. Advances in Kernel Methods. In: Schölkopf, B., Burges, C.J.C., Smola, A.J. (Eds.) (Cambridge, MA, USA: MIT Press), pp. 211–241.

Mercer, J., 1909. XVI. Functions of positive and negative type, and their connection the theory of integral equations. Philos. Trans. R. Soc. London Series A, Containing Papers Math. Phys. Charact. 209, 415–446.

Min, W., Wynter, L., 2011. Real-time road traffic prediction with spatio-temporal correlations. Transp. Res. Part C: Emerg. Technol. 19, 606–616.

Min, X., Hu, J., Zhang, Z., 2010. Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model. In: 13th International IEEE Conference on Intelligent Transportation Systems, pp. 1535–1540.

Mitrovic, N., Asif, M.T., Dauwels, J., Jaillet, P., 2015. Low-dimensional models for compressed sensing and prediction of large-scale traffic data. IEEE Trans. Intelligent Transp. Syst. 16, 2949–2954.

Müller, K.-R., Smola, A.J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., Vapnik, V., 1997. Predicting time series with support vector machines. In: Gerstner, W., Germond, A., Hasler, M., Nicoud, J.-.D. (Eds.), Artificial Neural Networks — ICANN'97. Springer, Berlin Heidelberg, pp. 999–1004.

Munoz, L., Sun, X., Horowitz, R., Alvarez, L., 2003. Traffic density estimation with the cell transmission model. In: Proceedings of the 2003 American Control Conference, vol.5, 2003, pp. 3750–3755.

Nanthawichit, C., Nakatsuji, T., Suzuki, H., 2003. Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. Transp. Res. Rec.: J. Transp. Res. Board 1855, 49–59.

Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic, part I: General theory. Transp. Res. Part B: Methodol. 27, 281–287.

Okutani, I., Stephanedes, Y.J., 1984. Dynamic prediction of traffic volume through Kalman filtering theory. Transp. Res. Part B: Methodol. 18, 1–11.

Pérez-Cruz, F., Camps-Valls, G., Soria-Olivas, E., Pérez-Ruixo, J.J., Figueiras-Vidal, A.R., Artés-Rodríguez, A., 2002. Multi-dimensional function approximation and regression estimation. In: Dorronsoro, J.R. (Ed.), Artificial Neural Networks — ICANN 2002. Springer, Berlin Heidelberg, pp. 757–762.

Polson, N.G., Sokolov, V.O., 2017. Deep learning for short-term traffic flow prediction. Transp. Res. Part C: Emerg. Technol. 79, 1–17.

Qi, Y., Ishak, S., 2014. A Hidden Markov Model for short term prediction of traffic conditions on freeways. Transp. Res. Part C: Emerg. Technol. 43, 95–111.

Qiao, F., Yang, H., Lam, W.H.K., 2001. Intelligent simulation and prediction of traffic flow dispersion. Transp. Res. Part B: Methodol. 35, 843–863.

Richards, P.I., 1956. Shock Waves on the Highway. Operat. Res. 4, 42–51.

Ryu, U., Wang, J., Kim, T., Kwak, S., Juhyok, U., 2018. Construction of traffic state vector using mutual information for short-term traffic flow prediction. Transp. Res. Part C: Emerg. Technol. 96, 55–71.

Sánchez-Fernández, M., de-Prado-Cumplido, M., Arenas-García, J., Pérez-Cruz, F., 2004. SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems. IEEE Trans. Signal Process. 52, 2298–2307.

SARLE, W., 1997. Neural network FAQ, part 1 of 7 : Introduction, periodic posting to the usenet newsgroup comp. ai. neuralnets. Ftp://Ftp.Sas.Com/Pub/Neural/FAQ.Html.

Schölkopf, B., Smola, A.J., Scholkopf, M.D. of the M.P.I. for B.C. in T.G.P.B., and Bach, F., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (MIT Press).

Shang, P., Li, X., Kamae, S., 2005. Chaotic analysis of traffic time series. Chaos, Solitons & Fractals 25, 121–128.

Smith, B.L., Williams, B.M., Keith Oswald, R., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. Transp. Res. Part C: Emerg. Technol. 10, 303–321.

Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. Transp. Res. Part C: Emerg. Technol. 11, 121–135.

Sun, X., Liu, T., 2011. A STAR model for urban short-term traffic flow forecasting. In: 7th Advanced Forum on Transportation of China (AFTC 2011), (Beijing, China: IET), pp. 185–190.

Tampere, C.M.J., Immers, L.H., 2007. An extended Kalman filter application for traffic state estimation Using CTM with Implicit Mode Switching and Dynamic Parameters. In: 2007 IEEE Intelligent Transportation Systems Conference, pp. 209–216.

Tikhomirov, V.M., 1991. On the Representation of continuous functions of several variables as superpositions of continuous functions of one variable and addition. In: Selected Works of A. N. Kolmogorov: Volume I: Mathematics and Mechanics, V.M. Tikhomirov, ed. (Dordrecht: Springer Netherlands), pp. 383–387.

Tuia, D., Verrelst, J., Alonso, L., Perez-Cruz, F., Camps-Valls, G., 2011. Multioutput support vector regression for remote sensing biophysical parameter estimation. IEEE Geosci. Remote Sensing Lett. 8, 804–808.

Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transp. Research Part C: Emerg. Technol. 13, 211–234.

Wang, J., Shi, Q., 2013. Short-term traffic speed forecasting hybrid model based on Chaos-Wavelet Analysis-Support Vector Machine theory. Transp. Res. Part C: Emerging Technologies 27, 219–232.

Wang, X., Xiao, J., 2003. A radial basis function neural network approach to traffic flow forecasting. In: Proceedings of the 2003 IEEE International Conference on

Intelligent Transportation Systems, vol. 1, pp. 614–617.
Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. Transp. Res. Part B: Methodol. 39, 141–167.
Wang, J., Shi, Q., Lu, H., 2005. The study of short-term traffic flow forecasting based on theory of chaos. In: IEEE Proceedings. Intelligent Vehicles Symposium, 2005, pp. 869–874.
Wang, J., Deng, W., Guo, Y., 2014. New Bayesian combination method for short-term traffic flow forecasting. Transp. Res. Part C: Emerg. Technol. 43, 79–94.
Wang, J., Chen, R., He, Z., 2019. Traffic speed prediction for urban transportation network: a path based deep learning approach. Transp. Res. Part C: Emerg. Technol. 100, 372–385.
Whittaker, J., Garside, S., Lindveld, K., 1997. Tracking and predicting a network traffic process. Int. J. Forecast. 13, 51–61.
Williams, B., 2001. Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling. Transp. Res. Rec.: J. Transp. Res. Board 194–200.
Williams, B.M., Hoel, L.A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. J. Transp. Eng. 129, 664–672.
Wu, C.-H., Ho, J.-M., Lee, D.T., 2004. Travel-time prediction with support vector regression. IEEE Trans. Intelligent Transp. Syst. 5, 276–281.
Wu, Y., Tan, H., Qin, L., Ran, B., Jiang, Z., 2018. A hybrid deep learning based traffic flow prediction method and its understanding. Transp. Res. Part C: Emerg. Technol. 90, 166–180.
Yang, Y., Lu, H., 2010. Short-term traffic flow combined forecasting model based on SVM. In: 2010 International Conference on Computational and Information Sciences, pp. 262–265.
Yang, W., Yang, D., Zhao, Y., Gong, J., 2010. Traffic flow prediction based on wavelet transform and Radial Basis Function network. In: 2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM), pp. 969–972.
Yin, H., Wong, S.C., Xu, J., Wong, C.K., 2002. Urban traffic flow prediction using a fuzzy-neural approach. Transp. Res. Part C: Emerg. Technol. 10, 85–98.
Yu, H., Wu, Z., Wang, S., Wang, Y., Ma, X., 2017. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. Sensors 17, 1501.
Zhang, Y., Zhang, Y., Haghani, A., 2014. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. Transp. Res. Part C: Emerg. Technol. 43, 65–78.
Zhang, Z., Li, M., Lin, X., Wang, Y., He, F., 2019. Multistep speed prediction on traffic networks: a deep learning approach considering spatio-temporal dependencies. Transp. Res. Part C: Emerg. Technol. 105, 297–322.
Zheng, Z., Su, D., 2014. Short-term traffic volume forecasting: a k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. Transp. Res. Part C: Emerg. Technol. 43, 143–157.
Zheng, W., Lee, D.-H., Shi, Q., 2006. Short-term freeway traffic flow prediction: bayesian combined neural network approach. J. Transp. Eng. 132, 114–121.
Zhu, C., Xu, X., Yan, C., 2010. The research of method of short-term traffic flow forecast based on GA-BP neural network and chaos theory. In: The 2nd International Conference on Information Science and Engineering, pp. 1617–1620.
Zhu, J.Z., Cao, J.X., Zhu, Y., 2014. Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections. Transp. Res. Part C: Emerg. Technol. 47, 139–154.