NETWORK-LEVEL TRAFFIC FLOW PREDICTION: FUNCTIONAL TIME SERIES VS. FUNCTIONAL NEURAL NETWORK APPROACH

BY TAO MA[1*], FANG YAO[2] AND ZHOU ZHOU[3]

[1]Ingram School of Engineering, Texas State University, 601 University Drive, San Marcos, TX 78666 USA

[2]Department of Probability & Statistics, School of Mathematical Sciences, Centre for Statistical Science, Peking University, 5 Yiheyuan Road, Haidian District, Beijing, P. R. China

[3]Department of Statistical Sciences, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada

Traffic state prediction is an essential component and an underlying backbone of intelligent transportation systems especially in the context of smart city framework. Its significance is mainly twofold in modern transportation systems: supporting advanced traffic operations and management for highways and urban road network to mitigate traffic congestion, and enabling individual drivers with connected vehicles in the traffic system to dynamically optimize their routes to improve travel time. Traffic state prediction with interval-based pointwise methods at 15-minute or hourly interval is common in traffic literature. However, because traffic dynamics are a continuous process over time, the discrete-time pointwise methods for traffic prediction at a fixed time interval hardly meet the advanced demands of continuous prediction in modern transportation systems. To close the gap, we propose functional approaches to intraday and day-by-day continuous-time prediction for traffic volume. This research focuses on network-level traffic flow predictions concurrently for all locations of interest. Two functional approaches are introduced, namely the network-integrated functional time series model and the functional neural network model. With functional approaches, a 24-hour intraday traffic profile is modeled as a functional curve over time, and sequences of historical traffic curves are used to predict traffic curves for near future days in a row and multiple locations of interest. We also include the functional varying coefficient model, Sparse VAR and traditional AR models in the comparative study, and the empirical results show that the network-integrated functional time series model outperforms other approaches in terms of the accuracy of predictions at network-scale.

1. Introduction.

1.1. *Background.*

Traffic state prediction is an essential component and an underlying backbone of intelligent transportation systems, especially in the context of smart city framework. Its significance is mainly twofold in modern transportation systems: (1) supporting advanced traffic operations and management for highways and urban road network to mitigate traffic congestion, such as enabling the traffic signal control system to proactively optimize a control plan based on prediction. (2) enabling individual drivers with connected vehicles in the connected traffic system to dynamically optimize their routes during a trip based on network state prediction to improve travel time. Fig. 1 (b) illustrates the concept of a connected traffic system. With advanced communication technology, connected vehicles can use internet connectivity

* Corresponding author: Tao Ma <tao.ma@txstate.edu>

Keywords and phrases: matrix-variate factor model, functional principal component, network-level traffic flow prediction, functional time series, functional neural network, functional varying coefficient.

to send and receive real-time information via communication between vehicles, transportation infrastructure, mobile devices, cloud computing platforms, and the global positioning system (GPS), etc. Giving vehicles the ability to communicate with each other and beyond brings significant value to safety, traffic efficiency and situational awareness. For instance, the vehicle-to-network (V2N) via cloud services enables real-time receiving network traffic prediction and routing optimization; the vehicle-to-infrastructure (V2I) enables traffic signal timing optimization and prioritization; the vehicle-to-vehicle (V2V) can help to avoid collision; the vehicle-to-pedestrian (V2P) can give safety alerts to pedestrians and bicyclists.

Although methods for traffic state prediction have been extensively studied in the past decades, most are discrete-time pointwise predictive methods for a fixed time interval. Different time intervals require different models to be estimated. Among others, traffic state prediction at 15-minute interval or hourly interval is common in traffic literature. However, because traffic dynamics are a continuous process over time, the discrete-time pointwise methods for traffic prediction at a fixed time interval hardly meet the advanced demands of continuous prediction for any time instant in modern transportation systems. Continuity is important in traffic prediction, because traffic generation by a network-wide population is a continuous random process where it may occur at any location and time of day. As intelligent transportation systems advance, the near-future traffic state is expected to be predicted for any location and time of day. It can be envisioned that, for instance, given the network-level traffic flow profile predicted for 24-hour ahead, the connected traffic system allows network operators to formulate proactive traffic management and control plans to balance network-wide traffic flows, and connected vehicles to automatically calculate an optimal route for their trips.

Therefore, an approach to intraday and day-by-day continuous-time traffic state prediction is necessary. To close the gap, we propose functional approaches to network-level prediction. The proposed approaches can be easily applied to both highways and an urban road network, and support real-time operations for congestion mitigation, trip planning, route guidance service, travel time saving and emission reduction. On the other hand, the functional approach can provide non-interval-based prediction that relaxes the need to estimate different models for different time intervals.

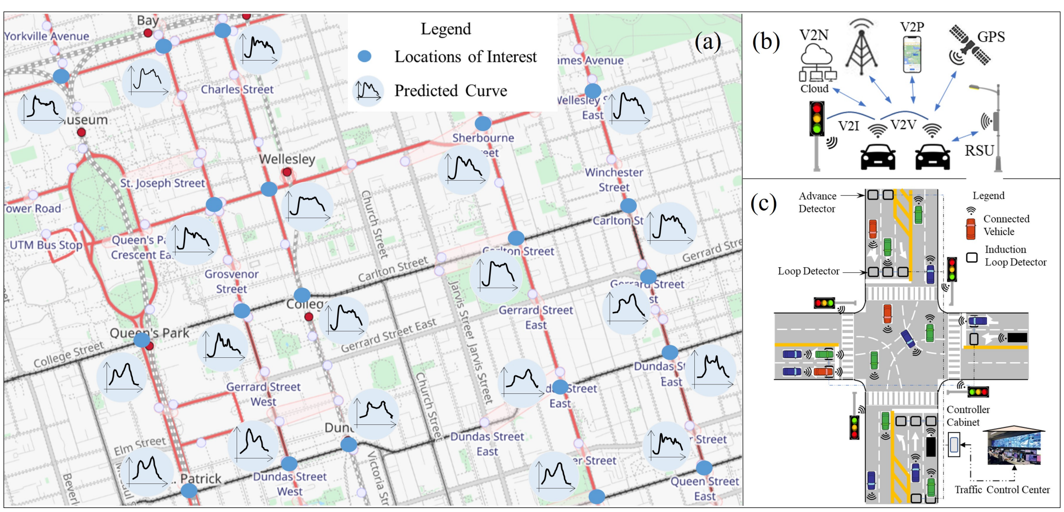


Fig. 1. (a) Illustration of network-level functional traffic predictions concurrently for multi-locations; (b) Connected traffic system; (c) Traffic data collection via loop detectors

1.2. *Characteristics of traffic data.*

1.2.1. *Functional traffic time series.* The functional approach is a powerful modeling tool for continuous data but requires dense and long measurements. Hence, data collection may pose a challenge in many circumstances. Thanks to innovative communication technologies, evolving sensor designs, data storage capacity, and data acquisition processes that made high-resolution and high-frequency traffic data collection across a road network possible. Traffic flow data can be automatically collected via infrared sensor, magnetic sensor, pneumatic tube sensor or inductive loop detector. The sensors are usually mounted on roadside hydro poles. The inductive loop detectors are deployed under road pavement across networks. Fig. 1 (c) illustrates a typical intersection with inductive loop detectors. A loop detector can transmit data at a frequency of 20 seconds via road-side controller cabinet to the computers in the Traffic Control Center. The traffic data can then be aggregated to any scale of dataset to meet the need of functional modeling. As traffic flows are stochastic and dynamic processes where each flow is a function of time, it can be assumed that an intraday traffic flow at a location is generated by an underlying random function. With functional approaches, a 24-hour intraday traffic profile can be approximated with a functional curve. A sequence of functional curves from consecutive days forms a functional time series. Two exemplar plots of functional time series from Highway 401 westbound and eastbound count stations in the city of Toronto are shown in Fig. 2(a) and (b) respectively. The graphs show that different locations might have their own unique features as well as some features in common.

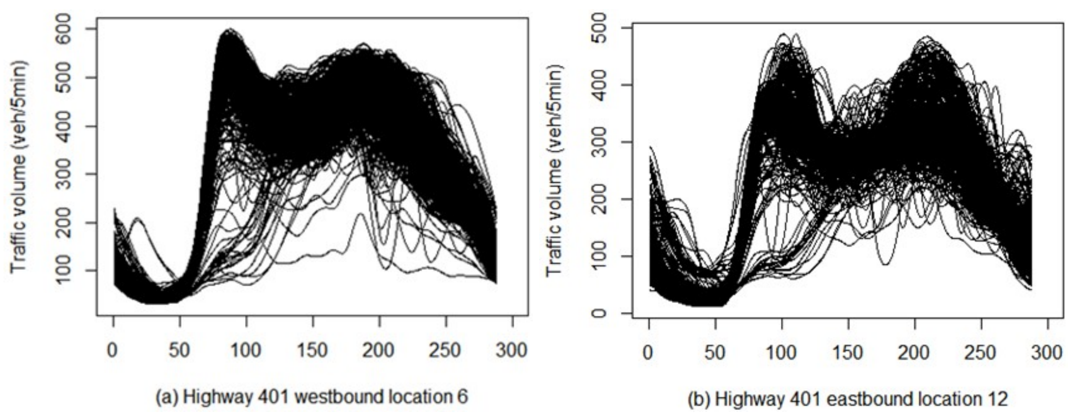
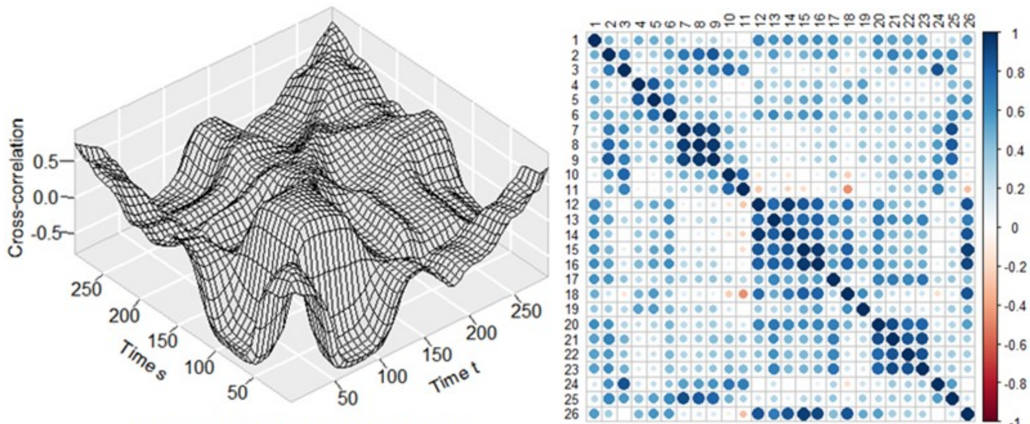


Fig. 2. Exemplar functional traffic time series from Highway 401

1.2.2. *Characteristics of traffic flows.* Observing how network-wide traffic flows exhibit co-movement phenomenon and repetitive patterns over time and space, it is natural to find that network-wide traffic flows, upstream and downstream, segment adjacent to segment, are not independent but interactive and associated with each other. These network-level spatiotemporal autocorrelation and cross-correlation within and between traffic flows are reflected in Fig. 3 where (a) shows temporal cross-correlation between two functional traffic time series from two locations on Highway 401 westbound, (b) shows a correlogram of traffic flows from 26 locations on Highway 401 westbound. The off-diagonal elements show the spatial cross-correlations between traffic flows. These network-level characteristics of traffic flows must be captured and reflected in prediction. Although functional approaches are becoming much more developed and applied in diverse research fields (Ferraty, 2011; Aneiros et al., 2017), showing encouraging predictive outcomes, they are not sufficiently addressed or with rather sparse attention in traffic state prediction, particularly none for network-level prediction. While a few functional studies now exist in traffic flow prediction (Besse and Cardot, 1996; Chiou, 2012; Klepsch, Klüppelberg and Wei, 2017; Guardiola, Leon and Mallor, 2014; Wagner-Muns et al., 2018), all are limited to a single count station or location. Therefore, to consider network-level spatiotemporal correlation in prediction, this research develops a functional approach to integrate network-level traffic flows from all locations of interest. Fig. 1 (a) illustrates the concept of network-level functional traffic prediction. The blue dots denote the locations of interest. The functional time series from all blue dots are assembled as a tensor to estimate a functional model, then the model is used to predict traffic curves for near future days in a row and all locations with blue dots. The background map in Fig. 1 (a) is an excerpt of the OpenStreetMap® (https://www.openstreetmap.org).



(a) Cross-correlation between two locations (Highway 401 WB location 1 and 2)

(b) Cross-correlations between 26 locations (Highway 401 WB)

Fig. 3. Network-wide spatiotemporal correlations among traffic flows

Two functional approaches are introduced in this research: i.e. the network-integrated functional time series model (NIFTS) and the functional neural network (FNN). The functional varying coefficient autoregressive model (FVCAR), sparse vector autoregressive model (sparseVAR) and traditional autoregressive (AR) model are also included in the comparative study.

A main contribution of this research is that functional principal component analysis (FPCA) with a matrix-variate factor model and a vector autoregressive (VAR) model are seamlessly integrated to capture network-level spatiotemporal co-movement of traffic flows as well as location-specific traffic features, which significantly improve the accuracy of prediction. This is a classic statistical approach. Thanks to FPCA and matrix-variate factor model for its powerful dimension reduction and data representation that made network-level prediction possible. The FNN model and the FVCAR model also provide competitive ways for traffic functional prediction.

The rest of this paper is organized as follows. Section 2 reviews relevant literature for traffic state prediction. Section 3 introduces the methodological basis for functional approaches. Section 4 presents the comparative study with four ground truth datasets. Discussion and conclusion are provided in section 5.

2. Relevant literature.

Over past decades, a wealth of predictive methods can be found in traffic literature for short term traffic state prediction. From early time prediction methods for a single location or a road segment to most recently for network-wide prediction, the methodology development follows a few research trajectories: traffic flow theory and simulation, statistical method, machine learning algorithms, or hybrid approach.

From the perspective of theoretical basis and data representation, the methodology for traffic state prediction may be divided into two classes: interval-based pointwise method and functional approach. With different modeling techniques, two classes of methodologies differ in their capacity to deal with different data forms. The interval-based method uses discrete time series data, whereas, the functional approach takes continuous functional time series data, thus resulting in two forms of prediction.

2.1. *Method for pointwise prediction.*

The class of interval-based pointwise methods provides discrete prediction, i.e. h-step-ahead pointwise prediction for a fixed time interval such as 5, 15, 30 or 60 minutes which depends on the resolution of training data used to estimate the model. Because there is an overwhelming volume of literature in this regard, only a few examples are listed in this review due to limited space. More thorough reviews may refer to Ermagun and Levinson (2018); Karlaftis and Vlahogianni (2011); Lana et al. (2018); Vlahogianni, Karlaftis and Golias (2014). In early times, many predictive methods were developed for a single location or a few segments of roadway due to technical constraint in data collection. Some examples may refer to: traffic flow theory based approaches (Deng, Lei and Zhou, 2013; Munoz et al., 2003; Nanthawichit, Nakatsuji and Suzuki, 2003; Wang and Papageorgiou, 2005), statistical time series methods (Kamarianakis and Prastacos, 2005; Min and Wynter, 2011; Billy M. Williams and Lester A. Hoel, 2003), machine learning algorithms (Abdulhai, Porwal and Recker, 2002; Guo and Zhu, 2009; Qiao, Yang and Lam, 2001), simulation approach (Abadi, Rajabioun and Ioannou, 2015), regression or hybrid methods (Cetin and Comert, 2006; Dell'Acqua et al., 2015; Oh, Kim and Hong, 2015; Wang, Deng and Guo, 2014; Zhang, Zhang and Haghani, 2014; Zheng Weizhong, Lee Der-Horng and Shi Qixin, 2006; Zhu et al., 2020). In recent years, the availability of big data makes network-wide prediction possible. Most solutions largely depend on machine learning algorithms such as long short-term memory (LSTM) recurrent neural network, convolution neural network (CNN), multidimensional support vector regression (MSVR), or tensor-based hybrid method. Most recent studies in this regard may refer to Dai et al. (2019); Feng et al. (2019); Gu et al. (2020); Jia and Yan (2020); Laharotte (2016); Liu et al. (2020); Lv et al. (2015); Ma et al. (2017); Mitrovic et al. (2015); Tan et al. (2016); Wang et al. (2011); Wang, Chen and He (2019); Yu et al. (2017). However, the interval-based method lacks flexibility if one wants to produce a prediction at and for any random time point that does not coincide with the interval window used for model estimation.

2.2. *Method for functional prediction.*

In contrast, the class of functional approach can provide functional prediction. Because the outcome of functional prediction is expressed as a continuous functional curve over time, it can therefore satisfy the need that prediction can be provided for any time instance. The relevant theory and techniques can be found in Ferraty and Vieu (2006); Ramsay and Silverman (2002, 2005) and comprehensive reviews are available via Müller (2005, 2008); Rice (2004); Zhao, Marron and Wells (2004).

There are two lines of approaches for functional prediction that are determined by the structure of the functional data set. If it is longitudinal functional data without time stamps, a functional linear regression model may be considered where both the response and covariates are functions. Recent studies along this line include Chiou (2012); Chiou and Li (2007); Chiou and Müller (2007); Müller and Zhang (2005); Yao, Müller and Wang (2005a,b); Yao, Lei and Wu (2015); Ramsay and Dalzell (1991), among others. If it is a functional time series data, a functional autoregressive (FAR) model may be considered where temporal dependence between lagged curves can be captured. The theoretical elaboration of the FAR model and the functional best linear predictor for a general linear process may refer to Bosq (2000); Hörmann and Kidziński (2015); Hörmann and Kokoszka (2010); Horváth and Kokoszka (2012); Horváth, Kokoszka and Rice (2014). A surrogate approach is proposed by Aue, Norinho and Hörmann (2015); Hörmann, Kidziński and Hallin (2015); Horváth and Kokoszka (2012) to approximate the FAR model using functional principal component analysis (FPCA).

Although functional approach and relevant theory have been developed for more than three decades since it was proposed by Bosq (1991) and successfully applied to many research areas, its research and applications in the area of traffic state prediction appear to be very sparse. Only a few studies relevant to traffic flow data can be found from statistical literature (Besse and Cardot, 1996; Chiou, 2012; Klepsch, Klüppelberg and Wei, 2017) and traffic literature (Guardiola, Leon and Mallor, 2014; Wagner-Muns et al., 2018). Following the modeling process defined by Bosq (1991), Besse and Cardot (1996) estimated a FAR(1) model with one-year long hourly traffic volume from the motorway tollbooth of Saint-Arnoux near Paris and compared it with a SARIMA model. Given a partial traffic flow trajectory of a day up to the current time, Chiou (2012) employed a functional mixture prediction approach to predict the unobserved partial flow trajectory for the rest of the day. The algorithm is validated with 15-min time interval traffic volume data from a dual loop detector near Shea-San Tunnel on National Highway 5 in Taiwan. Extending previous work in Besse and Cardot (1996), Klepsch, Klüppelberg and Wei (2017) approximated a functional ARMA(p, q) model with aid of FPCA and investigated the prediction algorithm with highway traffic speed data from a fixed point on a highway (A92) in Southern Bavaria, Germany. Guardiola, Leon and Mallor (2014) employed the FPCA approach to analyze and monitor the pattern of daily traffic flow profile using 1-min interval traffic data from the I-94 Freeway in Twin Cities, Minnesota. Wagner-Muns et al. (2018) performed functional forecast with aid of FPCA using 5-year long functional traffic volume time series collected from inductive loop detectors located at station (S110) on I-94 eastbound in the Minneapolis-St. Paul area and compared it with the traditional seasonal ARIMA model. Crawford, Watling and Connors (2017) used functional linear models to analyze the distribution of flows throughout a day, identified the predictable variability in daily traffic flow profiles, and output an average flow profile for each different day type. Two years of data from a loop detector on a key arterial route connecting Stockport to the city of Manchester is used to validate the proposed method. Nonetheless, the functional approaches aforementioned relevant to traffic data are based on loop detectors from and for a single location. A functional approach is missing at network-level prediction for multiple locations accounting for spatiotemporal correlations. To this end, we propose a network-integrated functional approach to close the gap.

In addition, it is noteworthy that all functional approaches mentioned above are statistical models. To the best of our knowledge, there is no functional traffic data analysis and prediction done in the line of machine learning in literature so far. Although, as early as 2005, Rossi and Conan-Guez (2005) and Rossi et al. (2005) proposed an approximate projection method making functional data possible to be taken by multi-layer perceptron neural network and radial-basis neural network. However, there has been no open-source implementation available thus far. In 2020, Thind, Multani and Cao (2020) developed an open-source package that makes functional neural network available. Both Thind et al. and Rossi et al. use projection method to represent functional data through a basis expansion in the functional space and work directly on numerical coefficients of basis expansion in neural network. A subtle difference is that Rossi et al. suggest a correction to those coefficients via the Choleski decomposition if the basis functions are not orthonormal. We introduce functional approach in both statistical and machine learning lines, i.e. functional time series and functional neural network approach, in this research for network-level traffic state prediction. The meaning is twofold: (1) it aims to introduce new ways to model and predict traffic data; (2) compare functional prediction between two lines and with other approaches at network scale.

3. Methodology.

In this section, we introduce theoretical background and mathematical form of functional time series approach and functional neural network, respectively.

3.1. *Network-integrated functional time series approach.*

The network-integrated functional time series approach consists of four modules. With the aid of functional principal component analysis (FPCA) as a dimension reduction tool, we transform the traffic functional data to FPC scores. The network-level functional time series can be transformed into its frequency domain which allows reassembling of the features (i.e. the FPC scores) in a form of tensor-based matrix-variate time series. Under this framework, we use matrix-variate factor model to decompose the networked tensor-based traffic time series into latent common factors and idiosyncratic components, where the former reflects the co-movement pattern of network traffic flows and the latter reflects local traffic features. The latent common factors incorporate spatial correlation among traffic flows via covariance structure of functional data from all locations of interest. The dynamics of co-movement and local traffic features are captured by autoregressive models respectively. The vector autoregressive model of the latent common factors not only accounts for network-wide temporal dependence, but also spatial cross-correlation in prediction. In theory, Aue, Norinho and Hörmann (2015) had shown that the predictions by the FPCA method are consistent with the ones by Bosq's p-order Functional Autoregressive (i.e. FAR(p)) model. Klepsch, Klüppelberg and Wei (2017) also quantified that the difference between the functional best linear prediction and the best linear prediction of the approximating vector model tends to zero if the dimension of the vector model approaches infinity.

3.1.1. *Mathematical form.*

The mathematical form of the proposed functional time series approach is written as follows.

$$x_{i,t}(\cdot) = \bar{x}_{i,\cdot}(\cdot) + \sum_{j=1}^{\infty} \phi_{i,j}\xi_{i,t,j} \tag{3.1}$$

$$\xi_{\cdot,t,\cdot} = \left(\xi_{i,t,j}\right)_{\substack{1\le i\le N\\ 1\le j\le J}} = Q_1 F_t Q_2^T + e_{\cdot,t,\cdot} \tag{3.2}$$

$$vec(F_t) = A_1 vec(F_{t-1}) + A_2 vec(F_{t-2}) + \cdots + A_p vec\left(F_{t-p}\right) + u_t \tag{3.3}$$

$$e_{i,t,\cdot} = B_{i,1}e_{i,t-1,\cdot} + B_{i,2}e_{i,t-2,\cdot} + \cdots + B_{i,q_i}e_{i,t-q_i,\cdot} + \varepsilon_{i,t,\cdot} \tag{3.4}$$

where Eq.(3.1) is the functional principal component model (Ramsay and Silverman, 2005). Based on the Karhunen-Loève expansion theorem, a functional curve can be expressed by a linear combination of the FPC scores. The functional principal components (i.e. eigenfunctions) represent the dominant modes of variation and covariation, and form an orthogonal basis. The FPC scores are the projection coordinates of the functional curve to this orthogonal system. The scores can be treated as a dual form of the functional curve in Hilbert space. For all equations, subscript $i = 1, \cdots, N$ denotes the location index and $t = 1, \cdots, T$ denotes the index of day. To simplify, i and t are omitted in the explanation of variables. $x_{i,t}(\cdot)$ denotes a functional curve and $\bar{x}_{i,\cdot}(\cdot)$ denotes the mean functional curve where $\bar{x}_{i,\cdot}(\cdot) = 1/T\sum_{t=1}^{T} x_{i,t}(\cdot)$. In the linear combination of Eq.(3.1), the coefficient $\phi_{i,j}$ is the j^{th} functional principal component where $j = 1, 2, \cdots$. Any two functional coefficients satisfy $\left\|\phi_{i,j}\right\|^2 = \int \phi_{i,j}(t)^2 dt = 1$ and $\langle \phi_{i,l}, \phi_{i,m}\rangle = \int \phi_{i,l}(t)\phi_{i,m}(t)dt = 0$ where $l \ne m, j = 1, 2, \cdots$ because they are orthonormal functions, i.e. the eigenfunctions of the covariance of the functional data set. $\xi_{i,t,j}$ denotes the j^{th} FPC score of a functional curve at location i on day t which is given by $\xi_{i,t,j} = \left\langle \phi_{i,j}, x_{i,t} - Ex_{i,t}\right\rangle = \int \phi_{i,j}(s)\left[x_{i,t}(s) - Ex_{i,t}(s)\right]ds$. In functional principal component analysis, the dimension reduction can be achieved by truncating the number of functional principal components up to J based on the percentage of variance explained, i.e. $j = 1, 2, \cdots, J$.

Eq.(3.2) is a matrix-variate factor model due to Wang, Liu and Chen (2019). $\xi_{\cdot,t,\cdot}$ is a $N \times J$ matrix that is a reassembled form of $\xi_{i,t,j}$ from Eq.(3.1). $\xi_{\cdot,\cdot,\cdot}$ is a $N \times J \times T$ tensor that represents the networked functional traffic time series. $\xi_{\cdot,t,\cdot}$ is one slice of $\xi_{\cdot,\cdot,\cdot}$ representing a matrix of location by FPC score on the day t. The matrix-variate factor model decomposes $\xi_{\cdot,t,\cdot}$ into latent common factor component F_t and idiosyncratic component $e_{\cdot,t,\cdot}$. F_t is a $k_1 \times k_2$ factor matrix. As k_1 and k_2 are significantly smaller than N and J respectively, i.e. $k_1 << N$, $k_2 << J$, therefore, Eq.(3.2) performs a further dimension reduction in both row and column directions of $\xi_{\cdot,t,\cdot}$ so that the prediction is carried out in a lower dimension factor space. Q_1 is a $N \times k_1$ front-loading matrix, and Q_2^T is a $k_2 \times J$ back-loading matrix. According to Wang, Liu and Chen (2019), the front-loading matrix $Q_1 = [q_{1,1}, q_{1,2}, \cdots, q_{1,k_1}]$, where $q_{1,j}$, $j = 1, \cdots, k_1$, is the unit eigenvectors of a super matrix M_1. A sample version of the super matrix M_1 is constructed as $\hat{M}_1 = \sum_{h=1}^{h_0} \sum_{\ell=1}^{J} \sum_{\ell'=1}^{J} \hat{\Omega}_{x,\ell\ell'}(h)\, \hat{\Omega}_{x,\ell\ell'}^T(h)$, where $\hat{\Omega}_{x,\ell\ell'}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} \zeta_{t,\ell} \zeta_{t+h,\ell'}^T$, $\zeta_{t,\ell}$ is the ℓ^{th} column vector of $\xi_{\cdot,t,\cdot}$, h is an integer variate, and h_0 is a pre-specified positive integer. Then, the sample version of $\hat{Q}_1 = [\hat{q}_{1,1}, \hat{q}_{1,2}, \cdots, \hat{q}_{1,k_1}]$, where $\hat{q}_{1,1}, \cdots, \hat{q}_{1,k_1}$ are the eigenvectors of $\hat{M}_1$ corresponding to its k_1 largest eigenvalues. The same procedure is applied to the transpose of $\xi_{\cdot,t,\cdot}$ to construct $\hat{M}_2$ to estimate the back-loading matrix $\hat{Q}_2$. For the purpose of prediction, the number of row factors k_1 and column factors k_2 can be estimated through the cross-validation method. $e_{\cdot,\cdot,\cdot}$ is a $N \times J \times T$ tensor where each slice $e_{\cdot,t,\cdot}$ is a $N \times J$ matrix. Overall, Eq.(3.2) is a tensor-based factor model where the latent common factor component contains convoluted spatiotemporal correlations and the co-movement pattern of network-wide traffic flows. The idiosyncratic component contains location-specific traffic features. More specifically, the front loading Q_1 reflects the (row) spatial dependence, i.e. the network-level traffic flow dependence between locations. The back loading Q_2 reflects the (column) dependence among the FPC scores of the traffic flow curves, i.e. the network-level traffic flow dependence in varying magnitude.

Eq.(3.3) and Eq.(3.4) follow autoregressive processes that reflect the dynamics and temporal dependence of co-movement of traffic flows and local traffic features. $vec(F_t)$ is the vectorized version of F_t from Eq. (3.2). The coefficients $A_1, \cdots, A_p$ are $r \times r$ matrices, where $r = k_1 k_2$, and $u_t = (u_{1t}, \cdots, u_{rt})'$ are white noise, p is the lag order. The coefficients $B_{i,1}, \cdots, B_{i,q_i}$ are $J \times J$ matrices, $\varepsilon_{i,t} = (\varepsilon_{1,t}, \cdots, \varepsilon_{N,t})'$ is white noise, q_i is the lag order associated with location $i = 1, \cdots, N$. The coefficients $A_1, \cdots, A_p$ and $B_{i,1}, \cdots, B_{i,q_i}$ of two autoregressive models can be estimated by the conventional Yule–Walker method and the innovations algorithm. Both p and q_i can be chosen based on BIC criterion (Brockwell and Davis, 2016).

3.1.2. *The modeling mechanism and data structure.* Fig. 4 shows the working mechanism flow chart of the proposed methodology. This approach sequentially concatenates a functional principal component analysis method, a matrix-variate factor model, and two autoregressive models for networked functional prediction.

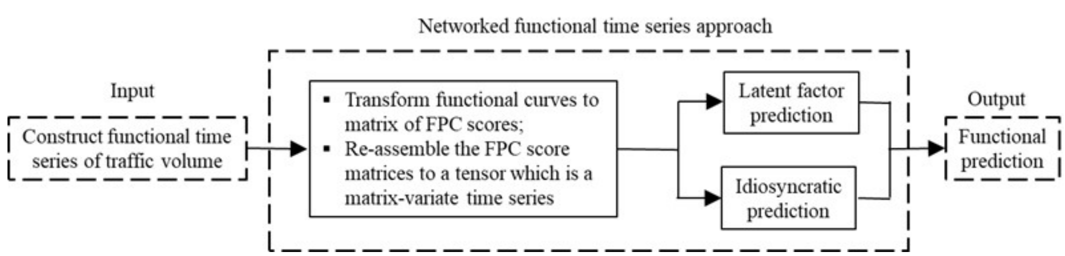


Fig. 4. Flow chart of the proposed methodology

According to the flow chart in Fig. 4, a length of T functional time series curves at a location is transformed to a $T \times J$ matrix of FPC scores via FPCA method. The matrices of FPC scores from all locations of interest are re-assembled to form a $N \times J \times T$ tensor which is a matrix-variate time series. Fig. 5 shows the structure of the re-assembled FPC score tensor. This matrix-variate time series is then decomposed into a latent common factor component and an idiosyncratic component. The prediction is carried out separately for the latent common factor series and the idiosyncratic series. Two predictions are combined and then inversely projected back to the functional space to obtain the functional prediction.

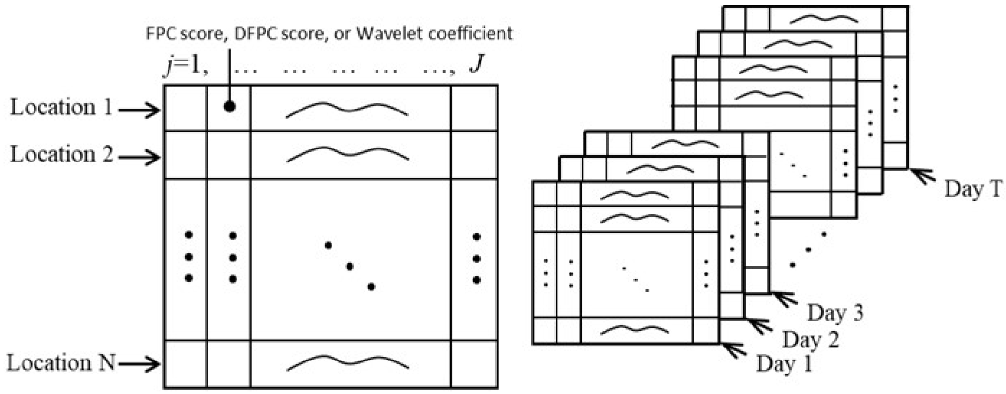


Fig. 5. Data structure of a matrix-variate time series

It is noteworthy that the latent factor component not only agglomerates the networked co-movement pattern and the spatiotemporal correlations of traffic flows, but also further reduces the dimensions of the FPC score tensor in both column J and row N directions. The idiosyncratic component reflects traffic flow characteristics at an individual location. Therefore, the combined prediction reflects both network-level and local traffic dynamic features.

3.2. *Functional Neural Network.*

Neural Networks are widely and extensively used for big data regression, classification and prediction. Fig. 6 shows the architecture of the functional neural network used for networked functional traffic flow prediction in this research. It consists of one input layer and one output layer with multiple hidden layers in between. $x_{i,t}(s)$ is a function to describe traffic profile at location i and day t. To train the neural network, $x_{i,t-1}(s), \cdots, x_{i,t-p}(s)$ in the input layer is pairwised with corresponding $x_{i,t}(s)$ in the output layer, where $i = 1, \cdots, N$ and $t = p + 1, \cdots, T$. There are totally $(T - p)$ pairwised input-output traffic curves for the neural network training. $\upsilon_i^{(u)}$ is the activated output of an individual neuron in the hidden layers where u is the index of hidden layers and i is the index of neurons in each hidden layer. $n_1, n_2, \cdots, n_u$ denote the number of neurons in each hidden layer respectively. $b^{(u)}$ is a vector of bias in the hidden layers.

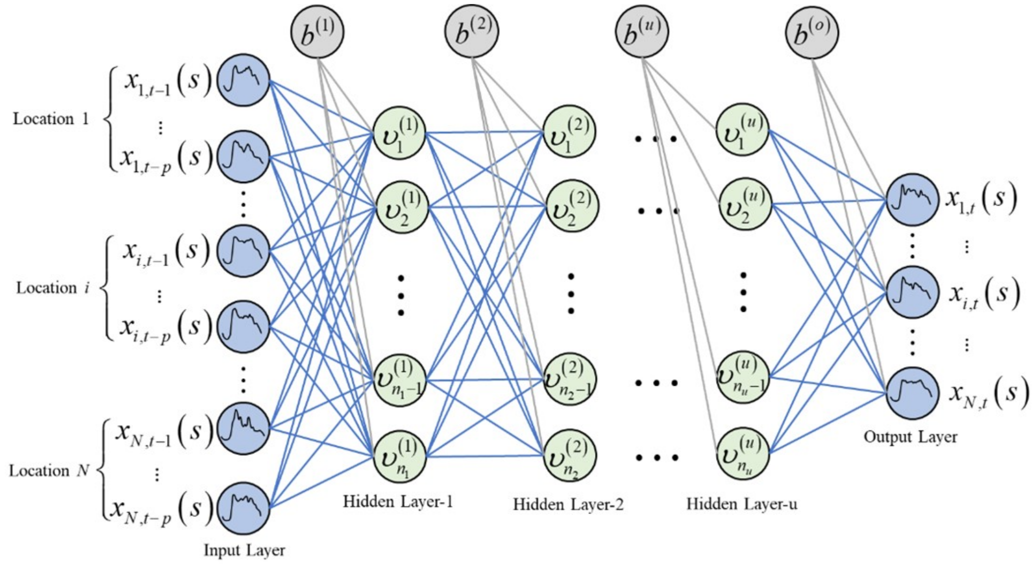


Fig. 6. Schematic architecture of the functional neural network

In the model training process, this structure of neural network captures the spatiotemporal dependence of traffic flows between current day and previous days and all locations of interest. This convoluted dependence is stored in the weight matrix and then used to predict traffic flow profile for near future days given traffic flow profiles on the adjacent past p days.

In the functional neural network, the mathematical form of a generalized neuron i in the first hidden layer can be expressed as Eq. (3.5). For generality, the functional neural network can accept scalar values as part of its input. Thus, Eq. (3.5) includes both functional input and scalar input.

$$\upsilon_i^{(1)} = g\left(\sum_{k=1}^{K} \int_{\mathrm{T}} \beta_{i,k}(s)\, x_k(s)\, dt + \sum_{d=1}^{D} \omega_{i,d}^{(1)} z_d + b_i^{(1)}\right) \tag{3.5}$$

where $g(\cdot)$ is a link function, which is called activation function in the context of neural network theory. K denotes the total number of neurons in the input layer that take functional data. $x_k(s)$ is a function from the k^{th} neuron in the input layer. $\beta_{i,k}(s)$ is the weight function between neuron i in the first hidden layer and neuron k in the input layer. Both $x_k(s)$ and $\beta_{i,k}(s)$ can be expressed by basis expansion, i.e. $x_k(s) = \sum_{h=1}^{H} C_{kh}\phi_{kh}(s)$ and $\beta_{i,k}(s) = \sum_{m=1}^{M} C_{ikm}\phi_{ikm}(s)$ where $\phi(s)$ is basis function, C is coefficient, H and M denote the number of basis functions. D denotes the total number of neurons in the input layer that take scalar value. z_d is the scalar input from the d^{th} neuron in the input layer. $\omega_{i,d}$ is the weight between neuron i and d. b_i is the bias term for neuron i.

In this research, because all traffic flow profiles are transformed to functional curves, the functional neural network does not take scalar input. Given an input function $x_k(s)$, the output of neuron i in the first hidden layer is $\upsilon_i^{(1)}$which is a real number. The neuron is called a functional neuron and $\beta_{i,k}(s)$ is called functional weight. It is worth mentioning that the functional weight can show the cumulative effect of the predictor onto the response variable over time, whereas a scalar weight in a non-functional neural network only reflects pointwise and static effect. As the output of a generalized functional neuron is a scalar value, we need such neurons only in the first hidden layer of the neural network to take functional input. Neurons after the first hidden layer are regular neurons that take scalar output from the previous layer. Functional neural network has a solid theoretical basis. Cybenko (1989) provided theoretical proof that a MLP neural network with functional inputs and a single hidden layer can uniformly approximate any continuous function of n real variables given only mild conditions on the activation function. Hence, MLP maintains its universal approximation property through a functional representation.

4. Prediction and comparative results.

Four ground-truth datasets of traffic volume (i.e. two datasets for weekday and two for weekend) are used to assess the effectiveness and efficiency of the proposed approaches, i.e. NIFTS and FNN. The raw data include two years of traffic volumes that were collected from the inductive loop detectors at 51 locations on Highway 401 in the City of Toronto. The data are aggregated to 288 data points per day at 5-minute intervals and transformed to a functional curve using a conventional smoothing technique of basis expansion with roughness penalty. The smoothing procedure including determination of basis functions, regularity and smoothing parameters may refer to Craven and Wahba (1978); Ramsay and Silverman (2005). Each functional curve represents an intraday traffic flow profile. At each location, the data are classified into two functional time series, i.e. weekday and weekend respectively, due to obvious difference in pattern. The functional traffic time series are divided into training set and test set. The training set is used for model estimation and the test set is used to validate model capacity for prediction.

4.1. *Model implementation and specification.*

4.1.1. *Network-integrated functional time series approach.* To implement the network-integrated functional time series approach, every functional time series from a location is transformed to a matrix of FPC scores. The FPC scores in each column correspond to a principal component. The FPC scores in each row correspond to a traffic curve. The number of rows is equal to the number of days, i.e. the length of the functional time series. The number of columns is equal to the number of eigenfunctions in Hilbert space that are remained to approximate the original data. All matrices of FPC scores from all locations of interest are re-assembled to form a tensor which is a matrix-variate time series as shown in Fig. 5. In addition, two additional transformation methods, i.e. dynamic FPCA and wavelet method, are introduced to the line of NIFTS approach as alternatives to FPCA to assess the role of transformation in prediction. Three transformation methods have different mathematical basis. The theoretical details of the dynamic FPCA may refer to Hörmann, Kidziński and Hallin (2015) and the wavelet method may refer to Mallat (1989); Percival and Walden (2000). Three transformation methods use different ways to extract features of the functional data. The dynamic FPCA transforms the functional curves to dynamic FPC scores. The wavelet method extracts the features of the functional curves to wavelet coefficients. They replace the FPC scores in Eq. (3.1). Through comparison, we can identify which transformation method is more suitable for functional traffic prediction.

The number of eigenfunctions or the number of wavelet coefficients is determined by the percentage of variance explained (PVE). To minimize loss of information in the data, save for a parsimony number of eigenfunctions or wavelet coefficients, we let three transformations satisfy 99% PVE for four ground-truth datasets. Table 1 shows the number of eigenfunctions and wavelet coefficients reserved respectively in the transformation. According to Eq. (3.2), the specifications of the front-loading and the back-loading of the matrix factor model are determined by the cross-validation method. The optimal choice of row factors k_1 reserved for the front-loading and column factors k_2 for the back-loading are shown in Table 1 conditional on the transformation method in Eq. (3.1). The lag order of vector autoregressive models in Eq. (3.3) and (3.4) is determined by BIC criterion (Brockwell and Davis, 2016).

Table 1 Specification of Integrated Functional Time Series Approach

Transformation Method	The number of FPCs or wavelet coefficients reserved by 99% PVE	Matrix Factor Model		The number of FPCs or wavelet coefficients reserved by 99% PVE	Matrix Factor Model	
		Row Factors k_1	Col Factors k_2		Row Factors k_1	Col Factors k_2
	Highway 401 Westbound Dataset			Highway 401 Eastbound Dataset		
FPCA	18	2	1	20	2	3
Dynamic FPCA	21	3	2	22	2	3
Wavelet	25	1	2	29	3	1

4.1.2. *Functional neural network approach.* The functional neural network for traffic prediction is set up as the architecture in Fig. 6. Neurons in the input layer take in functional data and neurons in the output layer give out predicted functional curves. Before it can be used for prediction, the neural network has to be trained with training data, called supervised learning. The training process is to find weight matrices that can map the relationship or dependency between the response and predictors. Hence, the training data are pairwised between the input layer and output layer for supervised learning. It is assumed that the traffic flow profile in the current day is correlated to those in the adjacent past p days. Hence, the functional traffic curves from the past p days from all locations of interest are loaded to the input layer and the pairwised functional traffic curves in the current day from all locations are loaded to the output layer. There are $(T - p)$ pairs of data that can be used to train the neural network given length of T functional time series, where p is lags. The network-level spatial correlations of traffic flows are convoluted in the functional weights and numerical weight matrices in the learning process.

The hyperparameters of the functional neural network such as the number of hidden layers, neurons in each hidden layer, epochs, validation split, learning rate, number of bases, activation choice, etc. are optimized with k-fold cross validation method. Details about how to determine the hyperparameters of functional neural network may refer to Thind, Multani and Cao (2020).

Table 2 Specification of Functional Neural Network

Model "sequential"	Highway 401 WB Dataset		Highway 401 EB Dataset	
Layer (type)	Output Shape	Param #	Output Shape	Param #
dense (Dense)	(None, 176)	215248	(None, 184)	216384
dense_1 (Dense)	(None, 128)	22656	(None, 128)	23680
dense_2 (Dense)	(None, 144)	18576	(None, 144)	18576
dense_3 (Dense)	(None, 754)	109330	(None, 725)	105125
Total params	365,810		363,765	
Trainable params	365,810		363,765	
Non-trainable params	0		0	
Final epoch (plot to see history):				
loss	0.01371		0.01194	
mean_squared_error	0.01371		0.01194	
val_loss	0.02163		0.02839	
val_mean_squared_error	0.02163		0.02839	

The training and validation of the functional neural network in this research employ the Tensorflow package in R (i.e. Keras) as the backend engine. The specification of the trained functional neural network is shown in Table 2. Fig. 7(a) and (b) show the plots of the network training and validation curves and their convergence with two ground-truth datasets. Once the supervised learning is completed, the neural network can be used for prediction. Given the most recent p days of traffic profiles as input, the trained functional neural network can output a traffic profile for the next day.

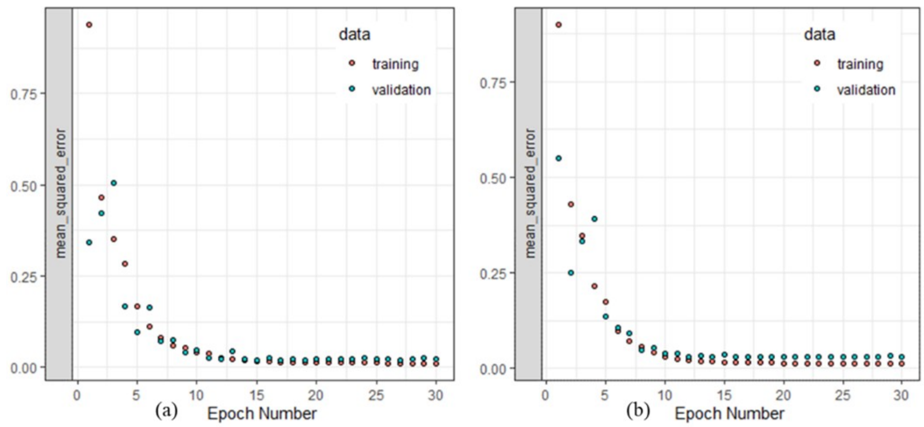


Fig. 7 Training and validation of Functional Neural Network with Highway 401 datasets of (a) WB and (b) EB

4.1.3. *Other approaches in comparative study.* In addition, FVCAR, SparseVAR and traditional AR models are also included in the comparative study. A classical functional varying coefficient regression model has functional response, covariates, coefficients as well as error term. For more details of the classical functional regression model, interested readers can refer to Chiou, Müller and Wang (2003); Yao, Müller and Wang (2005b); Ramsay and Silverman (2005); Zhu, Li and Kong (2012). A functional varying coefficient model with network structure was recently proposed by Zhu, Cai and Ma (2022). In this research, to fit the autoregressive structure of the functional traffic time series, the classical functional regression model is adapted to $y_{t,i}(s) = \beta_{0,i}(s) + \sum_{n=1}^{p} \beta_{n,i}(s)\, y_{t-n,i}(s) + Z_{t,i}(s)$, where $y_{t,i}(s)$ is the functional traffic curve on day t and at location i, s denotes intraday time, $\beta_{0,i}(s)$ and $\beta_{n,i}(s)$ are functional coefficients, n denotes the number of lags in day ($n = 1, 2, \ldots, p$), and $Z_{t,i}(s)$ is a white noise function with zero mean and covariance function $G_{t,i}(s,u) = Cov\{Z_{t,i}(s), Z_{t,i}(u)\}$. This model is called functional varying coefficient autoregressive model (FVCAR). All coefficient functions can be forced to be smooth via the use of roughness penalties. The FVCAR model is applied to every single location without considering spatial correlation between traffic flows from different locations. It is intended via comparison to assess the impact on the accuracy of prediction by the spatial correlations. At each location, it is assumed that the traffic flow curve in the current day is correlated to the adjacent past p days. Hence, traffic curves from the past p days are regressed on the curve in the current day for model estimation. The functional coefficients reflect the continuous change of the correlations between model input and output over time. The lag p takes the same value as the previous two models (i.e. NIFTS and FNN) for the comparative study. The rolling-ahead prediction is performed for five consecutive days.

A sparseVAR model is applied to every single location without taking network-level spatiotemporal correlations of traffic flows into account. The sparseVAR takes transformed FPC scores as its input to perform prediction. This approach imposes regularity on coefficients to reduce the complexity of the regression model. Interested readers can refer to Basu and Michailidis (2015); Efron et al. (2004); Friedman, Hastie and Tibshirani (2010); Tibshirani (1996) for more details of sparseVAR. The sparseVAR model specification follows BIC criterion (Brockwell and Davis, 2016).

A traditional pointwise autoregressive (AR) model is also included in the comparative study. As a benchmark method for time series prediction, it is intended via comparison to see the difference between pointwise interval-based prediction and functional curve prediction. The AR model is applied to every single location with no capacity to consider network-level correlations. The traffic time series with a 24-hour interval as a lag is used to estimate an AR model, then the model is used to perform traffic prediction for five consecutive days.

4.1.4. *Measurement of prediction accuracy.* The accuracy of prediction is measured with six indices, thus assessing the predictive capacity of each approach. The six indices that measure prediction errors include mean square prediction error (MSPE), median of MSPE (mMSPE), mean absolute percentage prediction error (MAPPE), median of MAPPE (mMAPPE), mean error of prediction (MEP) (Aneiros-Pérez and Vieu, 2006), and coefficient of variation (C.O.V).

The mean square prediction error is a primary indicator of prediction accuracy which contains combined magnitude of variance and bias due to approach. It is calculated by $MSPE = 1/n \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ where y_i denotes sample observation, $\hat{y}_i$ denotes predicted value from the model, and n is the sample size. The mean absolute percentage prediction error is to measure the average difference between observations and predictions in a percentage using an equation of $MAPPE = 1/n \sum_{i=1}^{n} |y_i - \hat{y}_i|/y_i \times 100\%$. The median of MSPE as well as the median of MAPPE indicate the distribution of the corresponding MSPE and MAPPE. The mean error of prediction is the ratio of MSPE over sample variance of the observed response which is calculated by$MEP = \left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \big/ n\right) \big/ Var(y)$. MEP can be seen as a rescaling of the MSPE. The coefficient of variation is defined as the ratio of the standard deviation to the mean, i.e. $c.o.v = \sigma/\mu$. It shows the extent of variability relative to the mean of the corresponding variable which is often expressed as a percentage. It is a standardized measure of dispersion of predictive error distribution.

4.2. *Comparative results.*

The comparative outcomes of five approaches with two additional transformation methods and four ground-truth traffic datasets is summarized in Table 3 and 4. The predictive capacity of each approach is reflected by the six indices, where each measurement is an average of predictions for five consecutive days at all locations along eastbound and westbound of Highway 401. The implication of the outcomes can be revealed and interpreted from three perspectives: (1) comparison between three functional approaches; (2) comparison between three transformation methods; (3) comparison between network-level prediction and location-specific prediction. Additionally, the prediction outcome with a pointwise interval-based AR method is also included in the comparison.

4.2.1. *Three lines of functional approaches.* There are three lines of functional approaches for traffic prediction, i.e. NIFTS, FNN and FVCAR. The result shows that the NIFTS approach outperforms both FNN and FVCAR. FVCAR does better than FNN. The statistical methods outperform the neural network-based method in this comparative study. The advantages of the NIFTS approach lie in its delicate structure that seamlessly concatenates the FPCA transformation, a matrix-variate factor model, and a vector autoregressive model. This model structure treats the networked traffic flows as one tensor and decomposes it to capture both network-level co-movement and location-specific traffic flow features. The spatiotemporal cross-correlations convolute in the model, thus significantly improving accuracy of prediction. In contrast, the neural network-based method and FVCAR have no such clear structure to account for both the network-level co-movement and location-specific traffic features. In addition, it is a consensus that the class of neural network models requires huge datasets for its training process in order to have excellent performance in prediction. On the contrary, a statistical method can be satisfied with relatively small datasets. The compromised performance of FNN in prediction is likely due to the short length of functional traffic time series. Relatively speaking, statistical models are not so sensitive as neural networks to the size of dataset.

4.2.2. *Three transformation methods.* Two additional transformation methods (i.e. dynamic FPCA, or wavelet) are employed within the line of the NIFTS approach to replace FPCA. After transformation, functional time series are converted to either dynamic FPC scores or wavelet coefficients, then processed by the models in Eq. (3.2), (3.3) and (3.4). The difference in prediction reflect the impact of transformation on the accuracy of prediction. The result in Table 3 and 4 indicates that the FPCA transformation excels the dynamic FPCA and wavelet methods. The NIFTS approach with FPCA transformation shows a stable and robust predictive capacity for ground-truth datasets in both weekdays and weekends. The dynamic FPC transformation exhibits very close capacity to FPCA. The difference in predictive accuracy due to three transformation methods are likely caused by their respective theoretical basis and working mechanisms. The wavelet coefficients seem to be less well received by the matrix factor model for network-level prediction. It is likely because the wavelet method is inclined to capture local features of functional data with sharp fluctuations.

4.2.3. *Network-level vs. location-specific prediction.* The NIFTS and FNN approaches are capable of network-level prediction because both can process network-level functional time series all at once. They both take spatiotemporal correlations of traffic flows into account in prediction. On the other hand, the FVCAR approach is suitable for location-specific prediction. Among statistical methods, Table 3 and 4 show that the NIFTS approach produces much smaller errors in prediction than the FVCAR, sparseVAR, as well as AR models. The values of six indices exhibit consistent pattern for both weekday and weekend datasets. The differences reflect the impact and significance of spatiotemporal correlations on the accuracy of prediction between network-level prediction and location-specific prediction.

4.2.4. *Weekday vs. weekend.* The traffic flow pattern during weekends is usually different from the ones during weekdays. The traffic flow exhibits two peak periods respectively in the morning and afternoon during weekdays, but one peak period in the middle of the day during weekends. All approaches are applied to the traffic data collected from both weekdays and weekend. The comparative results in Table 3 and 4 show consistent pattern for the predictive capacity of all approaches. These datasets with different traffic patterns and features may cross-validate the efficiency and effectiveness of the proposed functional approaches.

Table 3 Error of prediction with two ground-truth datasets (weekday)

Methodology (Indicator / Dataset)		MSPE	mMSPE	MAPPE %	mMAPPE %	MEP	C.O.V	MSPE	mMSPE	MAPPE %	mMAPPE %	MEP	C.O.V
		Highway 401 Westbound Dataset						Highway 401 Eastbound Dataset					
Network-Integrated Functional Time Series Approach (NIFTS)	FPCA	607.27	535.45	7.39	6.92	0.269	0.089	901.63	757.29	8.95	8.23	0.395	0.102
	Dynamic FPCA	747.40	639.87	8.10	7.85	0.321	0.097	891.78	752.49	9.00	8.51	0.392	0.103
	Wavelet	724.17	635.52	10.24	10.11	0.313	0.095	1043.30	887.23	12.72	12.51	0.465	0.111
Functional Varying Coefficient Autoregressive model (FVCAR)		724.77	653.78	8.06	7.86	0.318	0.095	1002.69	806.92	9.06	8.94	0.428	0.106
Functional Neural Network (FNN)		1367.52	1163.65	9.98	9.52	0.576	0.126	1635.44	1149.01	12.17	11.29	0.668	0.134
Sparse Vector Autoregressive model (SparseVAR)		728.78	643.65	7.85	7.27	0.315	0.095	986.44	864.02	9.29	9.11	0.431	0.108
Autoregressive Model (AR)		1446.96	1411.66	13.96	14.03	0.650	0.141	1872.71	1664.54	14.61	14.90	0.835	0.153

Table 4 Error of prediction with two ground-truth datasets (weekend)

Methodology \ Indicator / Dataset		MSPE	mMSPE	MAPPE %	mMAPPE %	MEP	C.O.V	MSPE	mMSPE	MAPPE %	mMAPPE %	MEP	C.O.V
		Highway 401 Westbound Dataset						Highway 401 Eastbound Dataset					
Network-Integrated Functional Time Series Approach (NIFTS)	FPCA	567.46	485.32	7.10	6.78	0.401	0.092	538.54	440.60	7.14	6.86	0.346	0.090
	Dynamic FPCA	594.22	505.06	7.31	6.89	0.410	0.094	557.05	505.16	7.59	7.24	0.359	0.092
	Wavelet	833.39	714.24	9.75	9.18	0.599	0.114	807.26	704.88	10.18	9.95	0.511	0.110
Functional Varying Coefficient Autoregressive model (FVCAR)		668.03	600.36	7.66	7.33	0.479	0.102	579.00	557.62	7.27	7.01	0.365	0.094
Functional Neural Network (FNN)		1800.25	1657.03	12.95	12.61	1.199	0.164	1957.89	1694.43	13.26	12.99	1.183	0.171
Sparse Vector Autoregressive model (SparseVAR)		735.39	597.15	7.92	7.52	0.494	0.104	621.61	541.69	7.68	7.54	0.417	0.099
Autoregressive Model (AR)		1474.15	1495.81	13.26	13.44	1.051	0.158	1457.41	1613.82	13.44	13.68	0.980	0.156

4.2.5. *Visualization.* Fig. 8, 9, 10 and 11 show visualizations of observed traffic curves and predicted traffic curves with five approaches and two additional transformation methods in NIFTS for a weekday and weekend at multiple locations. Due to limited space, only 6 out of 51 locations are selected for presentation in a figure. The observed traffic profiles vary significantly from one location to another and exhibit location-specific features even though variation is partly due to stochasticity. The observed traffic profiles are plotted in black, whereas the predicted curves are plotted in different colors. The curves in red are predicted by the NIFTS approach with FPC transformation that more closely mimic the observed traffic profile at each location than other predicted curves. The visualization further verifies the numerical outcomes in Table 3 and 4.

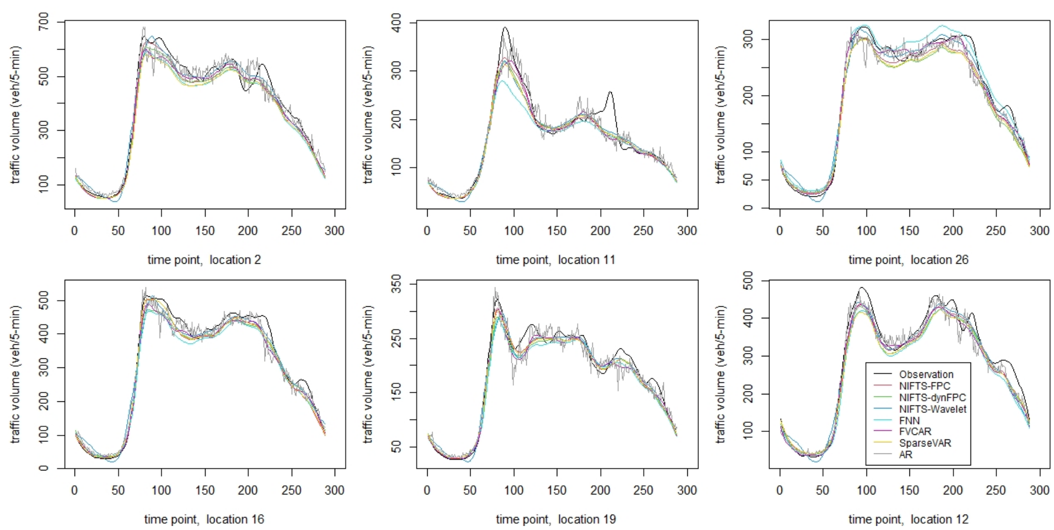


Fig. 8. Predictions at selected locations on Highway 401 WB (weekday)

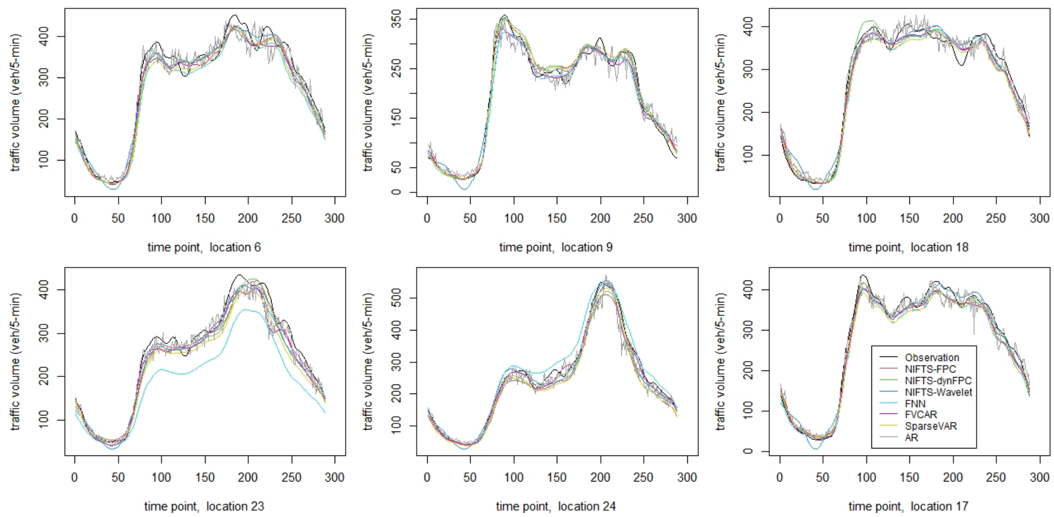


Fig. 9. Predictions at selected locations on Highway 401 EB (weekday)

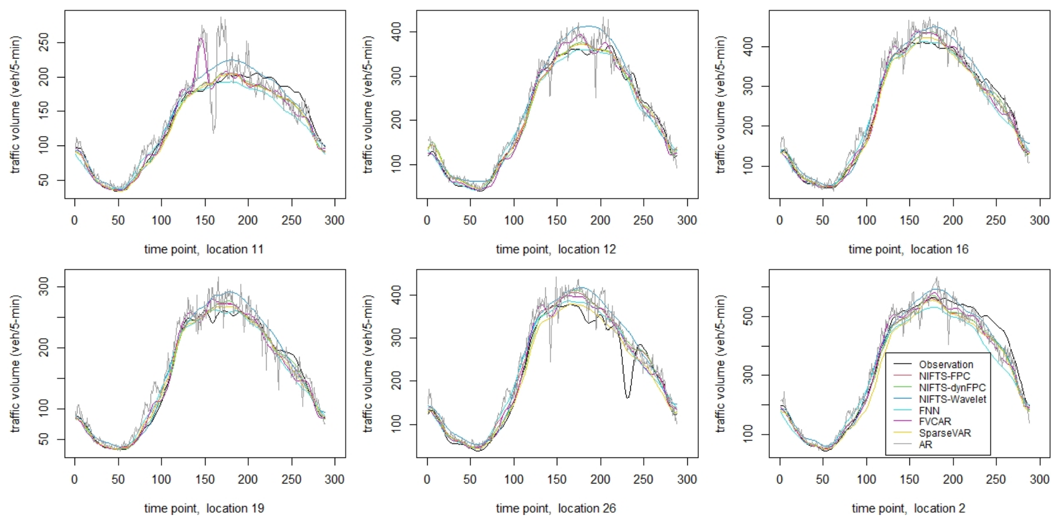


Fig. 10. Predictions at selected locations on Highway 401 WB (weekend)

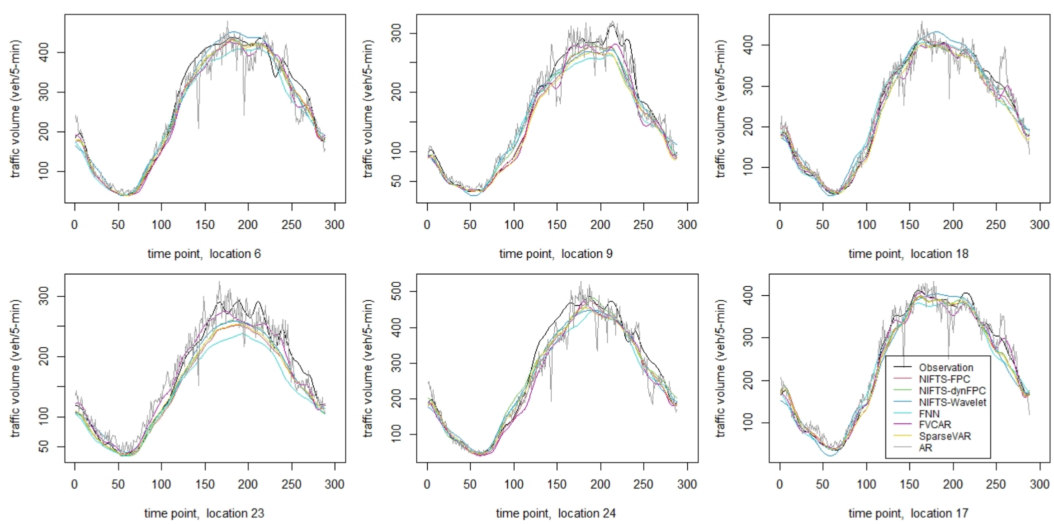


Fig. 11. Predictions at selected locations on Highway 401 EB (weekend)

In summary, from the empirical study based on four ground-truth datasets, the results indicate that (1) the NIFTS approach outperforms other approaches for network-level prediction; (2) the network-level prediction prevails the local-specific prediction on average; (3) the FPCA transformation method excels the dynamic FPCA and wavelet methods in line with the NIFTS approach. Overall, the functional approaches are able to incorporate network-level co-movement, spatiotemporal correlations of traffic flows, and local traffic features in the prediction process. It is noteworthy that the NIFTS approach produces the smallest error in network-level prediction with the FPCA transformation method, but the FNN approach provides a promising alternative for network-level prediction other than a statistical approach.

5. Discussion and Conclusion.

In this research, we propose functional approaches for network-level traffic prediction, thus setting a stage beneficial to both road operators and drivers, allowing more advanced operations and management of road network, traffic demand and real-time route guidance for connected vehicles in a connected traffic system. To inspect the effectiveness and efficiency of the proposed approaches, comparisons are carried out between three lines of functional approaches, three transformation methods, network-level vs. location-specific, and continuous vs. discrete prediction. The result shows that the NIFTS with FPCA transformation outperforms other approaches. Transformation methods have an impact on prediction. The network-level prediction outperforms the standalone location-specific prediction with consideration of spatiotemporal cross-correlations in prediction. As an alternative to the NIFTS, the FNN and FVCAR functional approaches also exhibit competitive prediction capacity.

The advantages of the NIFTS approach lie in its delicate structure that sequentially concatenates the FPCA transformation, a matrix-variate factor model, and a vector autoregressive model. This model structure treats the network traffic flows as one tensor, uses its decomposition capacity to capture both network-level co-movement and location-specific traffic flow features and convolutes spatiotemporal cross-correlations in the model, thus significantly improving accuracy of prediction. In addition, the NIFTS provides functional prediction simultaneously for all locations of interest across the traffic network and is able to provide a feasible, effective, and efficient way for large-scale network-level continuous-time traffic prediction.

Acknowledgments. We are very thankful to the Ministry of Transportation Ontario Canada and the ITS center at the University of Toronto for traffic data.

REFERENCES

Abadi, A., Rajabioun, T. and Ioannou, P. A. (2015). Traffic Flow Prediction for Road Transportation Networks With Limited Traffic Data. *IEEE Transactions on Intelligent Transportation Systems* **16** 653–662.

Abdulhai, B., Porwal, H. and Recker, W. (2002). Short-Term Traffic Flow Prediction Using Neuro-Genetic Algorithms. *Journal of Intelligent Transportation Systems* **7** 3–41.

Aneiros, G., Bongiorno, E. G., Cao, R., Vieu, P. et al. (2017). *Functional statistics and related fields*. Springer.

Aneiros-Pérez, G. and Vieu, P. (2006). Semi-functional partial linear regression. *Statistics & Probability Letters* **76** 1102–1110.

Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the Prediction of Stationary Functional Time Series. *Journal of the American Statistical Association* **110** 378–392.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Annals of Statistics* **43** 1535–1567.

Besse, P. C. and Cardot, H. (1996). Approximation spline de la prevision d'un processus fonctionnel autorégressif d'ordre 1. *Canadian Journal of Statistics* **24** 467–487.

Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In *Nonparametric functional estimation and related topics* 509–529. Springer.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. *Lecture Notes in Statistics*. Springer-Verlag, New York.

Brockwell, P. J. and Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer.

Cetin, M. and Comert, G. (2006). Short-Term Traffic Flow Prediction with Regime Switching Models. *Transportation Research Record: Journal of the Transportation Research Board* **1965** 23–31.

Chiou, J.-M. (2012). Dynamical functional prediction and classification, with application to traffic flow prediction. *The Annals of Applied Statistics* **6** 1588–1614.

Chiou, J.-M. and Li, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69** 679–699.

Chiou, J.-M., Müller, H.-G. and Wang, J.-L. (2003). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65** 405-423.

Chiou, J.-M. and Müller, H.-G. (2007). Diagnostics for functional regression via residual processes. *Computational Statistics & Data Analysis* **51** 4849–4863.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische Mathematik* **31** 377–403.

Crawford, F., Watling, D. P. and Connors, R. D. (2017). A statistical method for estimating predictable differences between daily traffic flow profiles. *Transportation Research Part B: Methodological* **95** 196–213.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* **2** 303–314.

Dai, X., Fu, R., Zhao, E., Zhang, Z., Lin, Y., Wang, F.-Y. and Li, L. (2019). DeepTrend 2.0: A light-weighted multi-scale traffic prediction model using detrending. *Transportation Research Part C: Emerging Technologies* **103** 142–157.

Dell'Acqua, P., Bellotti, F., Berta, R. and De Gloria, A. (2015). Time-Aware Multivariate Nearest Neighbor Regression Methods for Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems* **16** 3393–3402.

Deng, W., Lei, H. and 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.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics* **32** 407–499.

Ermagun, A. and Levinson, D. (2018). Spatiotemporal traffic forecasting: review and proposed directions. *Transport Reviews* **38** 786–814.

Feng, X., Ling, X., Zheng, H., Chen, Z. and Xu, Y. (2019). Adaptive Multi-Kernel SVM With Spatial–Temporal Correlation for Short-Term Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems* **20** 2001–2013.

Ferraty, F. (2011). Recent advances in functional data analysis and related topics.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Science & Business Media.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software* **33** 1–22.

Gu, Y., Lu, W., Xu, X., Qin, L., Shao, Z. and Zhang, H. (2020). An Improved Bayesian Combination Model for Short-Term Traffic Prediction With Deep Learning. *IEEE Transactions on Intelligent Transportation Systems* **21** 1332–1342.

Guardiola, I. G., Leon, T. and Mallor, F. (2014). A functional approach to monitor and recognize patterns of daily traffic profiles. *Transportation Research Part B: Methodological* **65** 119–136.

Guo, X. and Zhu, Q. (2009). A traffic flow forecasting model based on BP neural network. In *2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS)* **3** 311–314.

Hörmann, S. and Kidziński, Ł. (2015). A Note on Estimation in Hilbertian Linear Models. *Scandinavian Journal of Statistics* **42** 43–62.

Hörmann, S., Kidziński, Ł. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77** 319–348.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics* **38** 1845–1884.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. Springer-Verlag, New York.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics* **179** 66–82.

Jia, T. and Yan, P. (2020). Predicting Citywide Road Traffic Flow Using Deep Spatiotemporal Neural Networks. *IEEE Transactions on Intelligent Transportation Systems* 1–11.

Kamarianakis, Y. and Prastacos, P. (2005). Space–time modeling of traffic flow. *Computers & Geosciences* **31** 119–133.

Karlaftis, M. G. and Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies* **19** 387–399.

Klepsch, J., Klüppelberg, C. and Wei, T. (2017). Prediction of functional ARMA processes with an application to traffic data. *Econometrics and Statistics* **1** 128–149.

Laharotte, P.-A. (2016). Contributions à la prévision court-terme, multi-échelle et multi-variée, par apprentissage statistique du trafic routier, Ph.D. Thesis, Université de Lyon.

Lana, I., Del Ser, J., Velez, M. and Vlahogianni, E. I. (2018). Road Traffic Forecasting: Recent Advances and New Challenges. *IEEE Intelligent Transportation Systems Magazine* **10** 93–109.

Liu, L., Zhen, J., Li, G., Zhan, G., He, Z., Du, B. and Lin, L. (2020). Dynamic Spatial-Temporal Representation Learning for Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems* 1–15.

Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F.-Y. (2015). Traffic Flow Prediction With Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems* **16** 865–873.

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y. and Wang, Y. (2017). Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. *Sensors* **17** 818.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11** 674–693.

Min, W. and Wynter, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies* **19** 606–616.

Mitrovic, N., Asif, M. T., Dauwels, J. and Jaillet, P. (2015). Low-Dimensional Models for Compressed Sensing and Prediction of Large-Scale Traffic Data. *IEEE Transactions on Intelligent Transportation Systems* **16** 2949–2954.

Müller, H.-G. (2005). Functional Modelling and Classification of Longitudinal Data*. *Scandinavian Journal of Statistics* **32** 223–240.

Müller, H.-G. (2008). *Functional modeling of longitudinal data. In Longitudinal Data Analysis (G. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs, eds.) 223–251*. CRC Press, Boca Raton, FL.

Müller, H.-G. and Zhang, Y. (2005). Time-Varying Functional Regression for Predicting Remaining Lifetime Distributions from Longitudinal Trajectories. *Biometrics* **61** 1064–1075.

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

Nanthawichit, C., Nakatsuji, T. and Suzuki, H. (2003). Application of Probe-Vehicle Data for Real-Time Traffic-State Estimation and Short-Term Travel-Time Prediction on a Freeway. *Transportation Research Record: Journal of the Transportation Research Board* **1855** 49–59.

Oh, S.-d., Kim, Y.-j. and Hong, J.-s. (2015). Urban Traffic Flow Prediction System Using a Multifactor Pattern Recognition Model. *IEEE Transactions on Intelligent Transportation Systems* **16** 2744–2755.

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press.

Qiao, F., Yang, H. and Lam, W. H. K. (2001). Intelligent simulation and prediction of traffic flow dispersion. *Transportation Research Part B: Methodological* **35** 843–863.

Ramsay, J. O. and Dalzell, C. J. (1991). Some Tools for Functional Data Analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* **53** 539–572.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis, 2nd ed.*, 2nd ed. Springer-Verlag New York.

Rice, J. A. (2004). Functional and longitudinal data analysis: Perspectives on smoothing. *Statistica Sinica* **14** 631–647.

Rossi, F. and Conan-Guez, B. (2005). Functional multi-layer perceptron: a non-linear tool for functional data analysis. *Neural Networks* **18** 45–60.

Rossi, F., Delannay, N., Conan-Guez, B. and Verleysen, M. (2005). Representation of functional data in neural networks. *Neurocomputing* **64** 183–210.

Tan, H., Wu, Y., Shen, B., Jin, P. J. and Ran, B. (2016). Short-Term Traffic Prediction Based on Dynamic Tensor Completion. *IEEE Transactions on Intelligent Transportation Systems* **17** 2123–2133.

Thind, B., Multani, K. and Cao, J. (2020). Deep Learning with Functional Inputs. *arXiv:2006.09590 [cs, stat]*.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58** 267–288.

Vlahogianni, E. I., Karlaftis, M. G. and Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies* **43** 3–19.

Wagner-Muns, I. M., Guardiola, I. G., Samaranayke, V. A. and Kayani, W. I. (2018). A Functional Data Analysis Approach to Traffic Volume Forecasting. *IEEE Transactions on Intelligent Transportation Systems* **19** 878–888.

Wang, J., Chen, R. and He, Z. (2019). Traffic speed prediction for urban transportation network: A path based deep learning approach. *Transportation Research Part C: Emerging Technologies* **100** 372–385.

Wang, J., Deng, W. and Guo, Y. (2014). New Bayesian combination method for short-term traffic flow forecasting. *Transportation Research Part C: Emerging Technologies* **43** 79–94.

Wang, D., Liu, X. and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics* **208** 231–248.

Wang, Y. and Papageorgiou, M. (2005). Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. *Transportation Research Part B: Methodological* **39** 141–167.

Wang, Y., Coppola, P., Tzimitsi, A., Messmer, A., Papageorgiou, M. and Nuzzolo, A. (2011). Real-Time Freeway Network Traffic Surveillance: Large-Scale Field-Testing Results in Southern Italy. *IEEE Transactions on Intelligent Transportation Systems* **12** 548–562.

Zheng Weizhong, Lee Der-Horng and Shi Qixin (2006). Short-term freeway traffic flow prediction: bayesian combined neural network approach. *Journal of Transportation Engineering* **132** 114–121.

Billy M. Williams and Lester A. Hoel (2003). Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering* **129** 664–672.

Yao, F., Lei, E. and Wu, Y. (2015). Effective dimension reduction for sparse functional data. *Biometrika* **102** 421–437.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005a). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association* **100** 577–590.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33** 2873–2903.

Yu, H., Wu, Z., Wang, S., Wang, Y. and Ma, X. (2017). Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction in Transportation Networks. *Sensors* **17** 1501.

Zhang, Y., Zhang, Y. and Haghani, A. (2014). A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transportation Research Part C: Emerging Technologies* **43** 65–78.

Zhao, X., Marron, J. S. and Wells, M. T. (2004). The functional data analysis view of longitudinal data. *Statistica Sinica* **14** 789–808.

Zhu, X., Cai, Z. and Ma, Y. (2022). Network Functional Varying Coefficient Model. *Journal of the American Statistical Association* **117** 2074-2085.

Zhu, H., Li, R. and Kong, L. (2012). Multivariate varying coefficient model for functional responses. *Annals of statistics* **40** 2634.

Zhu, S., Zhao, Y., Zhang, Y., Li, Q., Wang, W. and Yang, S. (2020). Short-Term Traffic Flow Prediction With Wavelet and Multi-Dimensional Taylor Network Model. *IEEE Transactions on Intelligent Transportation Systems* 1–6.