A Unified Spatio-Temporal Model for Short-Term Traffic Flow Prediction

Peibo Duan, *Member, IEEE*, Guoqiang Mao, *Fellow, IEEE*, Weifa Liang, *Senior Member, IEEE*, and Degan Zhang, *Member, IEEE*

***Abstract*— This paper proposes a unified spatio-temporal model for short-term road traffic prediction. The contributions of this paper are as follows. First, we develop a physically intuitive approach to traffic prediction that captures the time-varying spatio-temporal correlation between traffic at different measurement points. The spatio-temporal correlation is affected by the road network topology, time-varying speed, and time-varying trip distribution. Distinctly different from previous black-box approaches to road traffic modeling and prediction, parameters of the proposed approach have physically intuitive meanings which make them readily amendable to suit changing road and traffic conditions. Second, unlike some existing techniques that capture the variation of spatio-temporal correlation by a complete re-design and calibration of the model, the proposed approach uses a unified model that incorporates the physical factors potentially affecting the variation of spatio-temporal correlation into a series of parameters. These parameters are relatively easy to control and adjust when road and traffic conditions change, thereby greatly reducing the computational complexity. Experiments using two sets of real traffic traces demonstrate that the proposed approach has superior accuracy compared with the widely used space–time autoregressive integrated moving average (STARIMA) and the back propagation neural network approaches, and is only marginally inferior to that obtained by constructing multiple STARIMA models for different times of the day, however, with a much reduced computational and implementation complexity.**

***Index Terms*— Spatio-temporal correlation, time-varying lag, trip distribution, digraph, unified.**

I. Introduction

ACCURATE short-term traffic flow prediction can benefit both road users and traffic management authorities. On one hand, road users can use traffic prediction to make better travel decisions, choose a faster route to reach the destination, and reduce fuel costs. On the other hand, traffic management authorities can utilize traffic prediction to improve traffic operation efficiency and apply more effective traffic control strategies to alleviate traffic congestion and improve the efficiency of road networks [1]–[6].

Existing work for short-term traffic prediction suffers from a number of shortcomings. First, the accuracy of a prediction model heavily depends on the traffic flow data which is spatially and temporally correlated [7]. It is challenging for the prediction model to take full account of the intricate spatio-temporal correlation. Second, the spatio-temporal correlation between traffic at different observation points is not stationary but varies with time of the day [8]. To this end, multiple prediction models corresponding to different times of the day have been constructed to suit time-varying spatio-temporal traffic correlations [9], [10]. Third, many approaches adopt a black-box approach to traffic prediction, e.g., principal component analysis based techniques, neural network-based techniques. The parameters of the developed traffic prediction models lack physically intuitive explanations. As a consequence, it becomes very difficult, if possible, for traffic operators to adjust the model parameters to suit changing road topology and traffic conditions. Lastly, in two-dimensional road networks, e.g., urban road networks, the estimation of time-varying spatio-temporal correlation, which forms the basis of traffic prediction, becomes more intricate since the spatio-temporal correlation is also strongly affected by the trip distribution and road topology.

In lieu of the aforementioned challenges, in this study, we design a unified spatio-temporal model based on STARIMA (Space-Time Autoregressive Integrated Moving Average) which captures the intricate spatio-temporal correlation structure between road traffic and hence can potentially deliver more accurate traffic flow prediction. Furthermore, parameters of the developed predictor have physically intuitive meanings, which make the model readily amendable to suit changing road topology and traffic conditions. Specifically, the following contributions are made in this paper:

- A physically intuitive approach to traffic prediction is developed that captures the time-varying spatio-temporal correlation between traffic at different measurement points. Distinctly different from previous black-box approaches to road traffic modeling and prediction, parameters of the proposed approach have physically intuitive

Manuscript received November 12, 2017; revised April 16, 2018 and July 20, 2018; accepted September 20, 2018. The Associate Editor for this paper was B. F. Ciuffo. *(Corresponding author: Degan Zhang.)*

P. Duan is with the School of Computing and Communication, The University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: peibo.duan@student.uts.edu.au).

G. Mao is with the School of Computing and Communication, The University of Technology Sydney, Ultimo, NSW 2007, Australia, and also with Data61, CSIRO, Eveleigh, NSW 2015, Australia (e-mail: g.mao@ieee.org).

W. Liang is with the Research School of Computer Science, Australian National University, Canberra, ACT 0200, Australia (e-mail: wliang@cs.anu.edu.au).

D. Zhang is with the School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China (e-mail: gandegande@126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2018.2873137

meanings which make them readily amendable to suit changing road and traffic conditions.
- Unlike some existing techniques which capture the variation of spatio-temporal correlation by a complete re-design and calibration of the model, the proposed approach uses a unified model which explicitly incorporates the impact of those physical factors affecting the variation of spatio-temporal correlation into the model parameters.
- Experiments using real traffic traces are conducted, which demonstrate that the proposed approach has superior accuracy compared with the STARIMA and the back propagation neural network model (BPNN, back propagation neural network) based approaches, and is only marginally inferior to that obtained by constructing multiple STARIMA models for different time period of the day, however with a much reduced computational complexity.

The rest of this paper is organized as follows. In Section II, existing research closely related to our work is reviewed. In Section III, the unified spatio-temporal model is developed based on a digraph model of the road network. The strategy and algorithm for estimating parameters of the proposed prediction model are presented in Section IV. After that, we evaluate the performance of the proposed methods in Section V. Finally, we draw the conclusion in Section VI.

II. Related Work

Depending on the traffic information employed for prediction, traffic prediction models can also be classified into: (i) temporal models which predict future traffic at a particular location of interest using historical (temporal) traffic data at the same location [11]–[13], (ii) spatio-temporal models which explore both historical traffic information and traffic information of spatially close measurement points for prediction [5], [14], [15].

Temporal models have been extensively applied in the past two decades. Particularly, time series based methods such as the ARIMA (Autoregressive Integrated Moving Average) model and its variants have attracted significant attention [2], [13], [16]. Van Der Voort *et al*. [16] proposed a Kohonen ARIMA (KARIMA) model, which applies Kohonen self-organizing map technique to classify the input data into a set of clusters, and then establishes an individually tuned ARIMA model for each cluster. Williams *et al*. [13] developed a seasonal ARIMA (SARIMA) model, which tries to identify seasonal patterns in the traffic to capture the cyclical variation of traffic states, such as peak and off-peak hours in each work day. In another work, Abadi *et al*. [2] used the SARIMA model to obtain accurate short-term prediction with limited input data. To capture the stochastic and nonlinear characteristics of historical traffic data in a temporal model, techniques from areas such as machine learning, economics, and stochastic analysis are also employed by researchers for traffic flow prediction. Some examples include Artificial Neural Network (ANN) [15], [17], Bayesian Network (BN) [18] and Support Vector Regression (SVR) [19]. However, spatial traffic correlation that can potentially be explored to improve the prediction accuracy was not considered in the aforementioned research.

To overcome the above shortcomings, spatio-temporal models have emerged as an efficient way to improve the prediction accuracy. Williams [20] developed a multivariate ARIMA model, denoted by ARIMAX (ARIMA with **ex**ogenous variables), which uses exogenous variables to capture the influence of upstream flows on downstream flows. An extension based on the ARIMAX model was developed by Stathopoulos and Karlaftis [21] by setting up various ARIMAX models for different time periods of the day. Xia *et al*. [3] proposed a spatio-temporal weighted KNN model, named STW-KNN, which predicts the traffic flow of a road by finding the most correlated flow from historical records at K adjacent up/downstream roads. The novelty of their research lies in the adoption of a state vector to describe the traffic conditions and a suitable distance metric to determine the proximity and correlation of traffic flows at different roads. In [22], Sun *et al*. modeled the road network as a Bayesian network where a road is represented as a node and the causal relation between two adjacent roads is represented as an edge. The joint probability distribution between the nodes with known data and the ones to be predicted was described by a Gaussian mixture model (GMM) where the parameters are estimated using the competitive expectation maximization algorithm. Bayesian network is also applied in Horvitz *et al*.'s work [23] which modeled traffic flow in the road, as well as the factors (e.g., incident, major events, weather) potentially affecting the variation of traffic flow as the nodes in the Bayesian network. To find the causal relation between nodes, a heuristic search together with a Bayesian scoring criterion to guide the search was performed over the models. Lv *et al*. [4] considered the traffic data as variables in the space-time cube. The generic traffic flow features embedded in these input variables are learned by a stacked auto-encoder model, a kind of neural networks. The model is trained in a greedy layerwise fashion and then used for forecasting. Deep learning was also used in [24] where Polson and Sokolov applied l_1-regularization technique to identify the spatio-temporal patterns. The experimental results showed that the preidctor was able to provide precise short-term traffic flow predictions even in the case that traffic flow regime changed drastically. Mitrovic *et al*. [25] used a singular value decomposition (SVD) based technique to construct a relationship matrix with which the traffic data of a few selected roads is able to map to that of the whole network. The traffic flows of the selected roads are then predicted by the SVR models and extrapolated to the whole network using the aforementioned relationship matrix.

Another major class of spatio-temporal models is the STARIMA based methods. In the STARIMA, a spatial weight matrix W is introduced that comprises two components: a spatial adjacency structure and a spatial weighting structure [7], [14]. As for the spatial adjacency, it reflects first-order spatial relations between all observations where two directly adjacent observations are termed as first-order spatial neighbors. For the spatial weight, it is the element of W that expresses the spatial correlation between two

first-order neighbors. The parameters in the STARIMA model are (p_λ, d, q_m) where p and q are time lags for the STAR and the STMA models respectively, d is the degree of differencing, λ and m are the spatial orders for the STAR and the STMA models respectively. The improvements in the performance of a STARIMA model are primarily shown in the aspect of capturing the temporal variation of spatio-temporal correlation. A common method is to re-estimate the parameters of the STARIMA model in each traffic state of the day to better capture the traffic similarity in the same state. For example, Min and Wynter [7] redefined a spatial order as an ordering with respect to the Euclidean distance traveled by vehicles within a unit time interval. As travel speed varies temporally, the spatial weight matrix is re-evaluated in different time periods of the day. Similarly, Cheng *et al*. [14] transformed the static spatial weight matrix into a dynamic one by defining the spatial weight as a function of the time-varying speed between two neighboring locations. Unfortunately, with the rapid variation of traffic conditions, this causes a large increase in the number of estimated parameters and an explosive growth of computational time. To improve the efficiency of estimating the parameters in multiple STARIMA models corresponding to different times of the day, Salamanis *et al*. [26] only employed a prescribed number of spatially correlated neighbors of a road of interest. They analyzed the degree of the spatio-temporal correlation between the traffic from different measurement points using a Pearson product–moment correlation-coefficient-based metric, which is based on the cross correlation function. Our previous work [8] proposed a convenient technique to adjust the lags of the STARIMA model dynamically to suit different traffic states, which was validated using measured traffic data on a highway.

As mentioned in Section I, to apply the approach developed in [8] to an intricate two-dimensional road network, a number of challenges need to be conquered, including the explicit consideration of the road topology and trip distribution in traffic prediction. As for the road topology, most studies use graph-theoretic techniques to transform a road network into a mathematical model convenient for subsequent analysis. Kelly [27] modeled the road network by an incidence matrix. Each column in the matrix corresponds to a road and each row corresponds to a measurement point in a road. The column for a road comprises entries of 0s or 1s with 1 indicating a particular measurement point is on a particular road and 0 otherwise. The 1s in a row suggest which roads pass through that measurement point. However, the dimension of an incidence matrix quickly explodes for even a moderate number of roads and measurement points. To overcome the scalability problem, Salamanis *et al*. [26] represented a road network by an adjacency matrix where each column and each row represented a road. If two roads are adjacent, the corresponding entry in the adjacency matrix is 1; otherwise, the entry becomes 0. It is worth noting that all aforementioned methods modeled the road network as an undirected graph, that is, prediction must be executed before specifying a particular traffic direction. Unlike existing graph-theoretic techniques, we employ a digraph to model the road network which can better capture directionality of road traffic flows.

In the literature, the spatial pattern of traffic between origins and destinations is usually expressed by a trip distribution matrix based on the undirected graph model of traffic network and widely used in the traffic state estimation [28], traffic flow prediction [2] or traffic flow demand estimation [29] and so forth. To extend the trip distribution matrix to the digraph model, we propose the concepts of turning rate and traffic transition probability (TTP) which are capable of accurately capturing the traffic distribution among roads with road intersections. To estimate turning rate or TTP, we apply the gravity model based method where not only traffic data, but also the spatial separation between two locations is considered [30]. As the gravity model merely requires the traffic information at the origin and the destination, the adverse impact of missing traffic measurements on some roads along the paths between the origin and the destination can be omitted. Indeed, in the real life, it is economically prohibitive to deploy traffic detectors across the whole road network.

III. Unified Spatio-Temporal Model

In this paper, we introduce a unified spatio-temporal model, which is based on the STARIMA model. For completeness, we briefly introduce the STARIMA(p_λ, d, q_m) model as follows:

$$(\boldsymbol{I} - \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \boldsymbol{\phi}_{kl} \boldsymbol{W}_l L^k)(1-L)^d Y(t) = (\boldsymbol{I} - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \boldsymbol{\theta}_{kl} \boldsymbol{W}_l L^k)\varepsilon_t. \quad (1)$$

In (1), $Y(t)$ is a $\mathcal{N} \times 1$ vector including the traffic flow collected from $\mathcal{N}$ observation points at t where t is considered to be a discrete variable, representing a time interval with integer index $t = 1, 2, \ldots$; L is the lag operator with $Ly_i(t) = y_i(t-1), i \in \mathcal{N}$; $\boldsymbol{\phi}_{kl}$ and $\boldsymbol{\theta}_{kl}$ are the coefficients; $\boldsymbol{W}$ is the spatial weight matrix, and ε_t is white noise.

The establishment of the STARIMA model, especially the parameters $\{p, q, \lambda, m\}$, are closely related to the correlation of traffic data at the corresponding spatial and temporal lags. Therefore, when the correlation structure of the underlying traffic process changes, the model for traffic prediction also needs to change adaptively for more accurate prediction. In the literature, this has been done by setting up multiple models for different time periods of the day. The establishment of multiple models incurs much greater complexity and computational costs. Furthermore, the condition triggering the transition between these multiple models is also not always clear as the road traffic process does not necessarily repeat itself following exact temporal cycles, e.g., the occurrence of traffic peaks may easily shift by several minutes or tens of minutes from day to day. To handle above challenges, we propose a unified STARIMA model where the impact of the time-varying spatio-temporal correlation is taken into account by adjusting the values of temporal lags in the model. In the proposed approach, the time-varying components are captured by a series of parameters where each parameter has physically intuitive meanings and can be directly related to the road topology, trip distribution, travel speed and distance.

To better illustrate the establishment process of the unified spatio-temporal model, we model a road network as a digraph. We partition the road network into a set of *road segments*. Each road segment is a piece of road bounded by two road intersections and there is no intersection within a road segment. A very long road segment may be further partitioned into multiple smaller road segments. We call a particular travel direction of a road segment a *link*. Depending on whether the road is one-way or two-way, a road segment may be represented by one or two links [27]. Without losing generality, we further assume that there is at most one measurement point within a link. If there are multiple measurement points within a long road, this can be readily handled by dividing the long road into multiple road segments where each segment contains up to one measurement point only. Drawing from graph theory, an arrangement of links can be modeled as a digraph $D = (V, E)$ with a set of V of vertices and a set E of arcs (or directed edges). The vertex set $V = \{V_1, V_2, \ldots, V_{\mathcal{N}}\}$ and $V_i \in V$ represents the i-th link or a particular point, e.g., a measurement point, if it exists, within the i-th link. There is an arc $e_{i,j} \triangleq (V_i, V_j)$, $e_{i,j} \in E$, going from V_i to V_j if there is traffic traveling *directly* from V_i to V_j. Based on the digraph model, a route from link i to link j is defined as a path from V_i to V_j, including a finite sequence of arcs connecting a sequence of vertices that are all distinct from one another. Moreover, the number of arcs is denoted by l, which is the *path length*. Since a vertex $V_i \in V$ has both incoming and outgoing arcs, the neighbors of V_i are classified into two categories. The first category is a set of vertices that are the links located upstream of link i. We denote it by V_i^{1-}. Correspondingly, the second category is a set of vertices including the neighbors of V_i that are the links located downstream of link i. We denote it by V_i^{1+}.

In the following, we first explore the spatio-temporal correlation between $V_i \in V$ and $V_j \in V_i^{1+}$. To begin with, we introduce the concept termed "turning rate" $\pi_{i,j}$ to represent the ratio of traffic at V_i and traveling to V_j. Then, we approximately estimate the incoming traffic at V_j from V_i by:

$$y_{i,j}(t) = \pi_{i,j}\, y_i(t - \tau_{i,j}), \quad \tau_{i,j} \in \mathbb{Z}^+, \tag{2}$$

where $y_i(t - \tau_{i,j})$ represents the traffic at V_i at $t - \tau_{i,j}$. $\tau_{i,j}$ is the time-varying lag corresponding to the time required to travel from V_i to V_j because at that time lag, the (approximate same) set of vehicles $y_{i,j}(t)$ have reached V_j. Note that, the turning rate $\pi_{i,j}$ varies over the time of the day. In this paper, we assume that $\pi_{i,j}$ remains constant during a given time period of the day, e.g., peak or off-peak hours. The estimation of $\pi_{i,j}$ and $\tau_{i,j}$ will be discussed in next section. Utilizing the lag operator L, Equation (2) can be rewritten as

$$y_{i,j}(t) = \pi_{i,j} L^{\tau_{i,j}} y_i(t), \quad \tau_{i,j} \in \mathbb{Z}^+. \tag{3}$$

Based on (2), we obtain the traffic at V_j:

$$y_j(t) = \sum_{V_i \in V_j^{1-}} y_{i,j}(t). \tag{4}$$

Unfortunately, not every vertex in V_j^{1-} has measurement data available since in real applications many links may not be equipped with traffic detectors. Denoting the subset of vertices with measurement data in V_j^{1-} by $\widehat{V_j^{1-}}$, whereas the subset of vertices without measurement data by $\widetilde{V_j^{1-}}$. In this case, $y_j(t)$ can be expressed as the sum of the traffic coming from $\widehat{V_j^{1-}}$ and $\widetilde{V_j^{1-}}$:

$$y_j(t) = \sum_{V_{i_1} \in \widehat{V_j^{1-}}} y_{i_1,j}(t) + \sum_{V_{j_1} \in \widetilde{V_j^{1-}}} y_{j_1,j}(t). \tag{5}$$

In (5), the traffic from $\widehat{V_j^{1-}}$ can be calculated directly. As for the traffic from $\widetilde{V_j^{1-}}$, we should estimate it by considering the traffic upstream from the adjacent neighbors of V_{j_1}. Moreover, if there is still no measurement traffic upstream from the adjacent neighbors of V_{j_1}, we have to further consider the traffic upstream from the neighbors that are far away from V_{j_1}. For the sake of simplicity, we term $\widehat{V_j^{1-}}$ as the first *in-level-available* vertices of V_j. Second *in-level-available* vertices of V_j, denoted by $\widehat{V_j^{2-}}$, and so on. Correspondingly, $\widetilde{V_j^{l-}}, l \geqslant 1$ are termed as the l-th *in-level-unavailable* vertices of V_j. To find $\widehat{V_j^{l-}}$ and $\widetilde{V_j^{l-}}$ in the general case, a BFS (breadth first search) based algorithm is designed and applied. We will present such algorithm in next section.

We use $P_{i_l,j}^l, V_{i_l} \in \widehat{V_j^{l-}}$, to denote a set of paths where each path $P_z \in P_{i_l,j}^l$ starts from V_{i_l} and ends at V_j via $l-1$ vertices respectively belong to $\widehat{V_j^{1-}}, \widehat{V_j^{2-}}, \ldots, \widehat{V_j^{(l-1)-}}$. We use $y_{i_l,j}^l(t)$ to denote the traffic traveling from V_{i_l} to V_j along $\forall P_z \in P_{i_l,j}^l$. Besides, $y_{i_l,j}^l(t)$ is estimated by

$$\begin{aligned} y_{i_l,j}^l(t) &= \sum_{p_z \in P_{i_l,j}^l} \pi_{i_l,j}^{P_z} y_{i_l}(t - \tau_{i_l,j}^{P_z}) \\ &= \sum_{p_z \in P_{i_l,j}^l} \pi_{i_l,j}^{P_z} L^{\tau_{i_l,j}^{P_z}} y_{i_l}(t). \end{aligned} \tag{6}$$

Particularly, $\pi_{i_l,j}^{P_z}$ and $\tau_{i_l,j}^{P_z}$ are respectively the turning rate and time-varying lag between V_{i_l} and V_j upon path $P_z \in P_{i_l,j}^l$. Both $\pi_{i_l,j}^{P_z}$ and $\tau_{i_l,j}^{P_z}$ are estimated on the basis of $\pi_{i,j}$ and $\tau_{i,j}$. Suppose there is a λ_j satisfying $\widetilde{V_j^{\lambda_j-}} = \varnothing$. With (6), we can calculate $y_j(t)$ by

$$y_j(t) = \sum_{l=1}^{\lambda_j} \sum_{V_{i_l} \in \widehat{V_j^{l-}}} y_{i_l,j}^l(t). \tag{7}$$

Up to this point, we draw a clear and physically intuitive picture of the spatio-temporal correlation between any two links. To better illustrate above process, we give an artificial instance in Fig. 1a where the gray nodes are the vertices with measured data, whereas the red nodes are the vertices without measured data. In this instance, $\widehat{V_1^{1-}} = \{V_2\}$ and $\widetilde{V_1^{1-}} = \{V_3\}$. As there is no traffic measured at V_3, we should consider $\widehat{V_1^{2-}} = \{V_5, V_6\}$. Since $\widetilde{V_1^{2-}} = \varnothing$, we get $\lambda_1 = 2$. Finally, we calculate $y_1(t) = y_{2,1}^1(t) + y_{5,1}^2(t) + y_{6,1}^2(t)$.

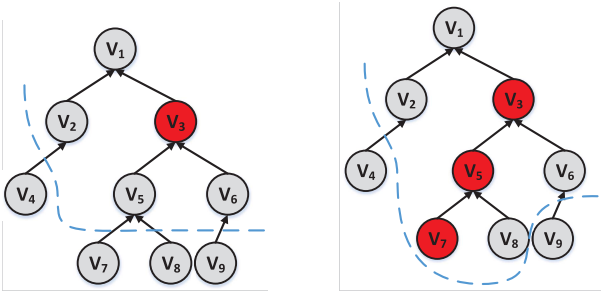


Fig. 1. Traffic flow prediction for a vertex (link) in an artificial road network with consideration of the situations that there is (not) enough traffic data.

In (7), a big challenge we face is that in some situations, there is no such value of λ_j which satisfies $\widetilde{V_j^{\lambda_j-}} = \varnothing$. In other words, there are no enough detectors to provide sufficient data to calculate $y_j(t)$. Consider Fig.1b where the digraph structure of the road network topology is the same as the one in Fig.1a. However, not only V_3 but also V_5 and V_7 do not have measured data. In order to estimate the traffic at V_3, the BFS algorithm will be executed until the leaf node V_7 is achieved. As the traffic at V_7 can not be inferred from its child nodes, it is impossible to accurately estimate the traffic at V_5 and V_3. Furthermore, the traffic at V_1 can not be calculated via (7). To tackle this problem, we assume that a BFS algorithm terminates when there is $l = \lambda_j$ satisfying each node in $\widetilde{V_j^{\lambda_j-}}$ has no child. In this way, $y_j(t)$ consists of two parts. The first part is the traffic from measured links while the second part is the traffic from unmeasured links. Thus (7) can be expressed as follows:

$$y_j(t) = \sum_{l=1}^{\lambda_j} \sum_{V_{i_l} \in \widehat{V_j^{l-}}} y_{i_l,j}^{l}(t) + \sum_{V_i \in \widetilde{V_j^{\lambda_j-}}} y_{i,j}^{\lambda_j}(t) \tag{8}$$

For simplicity, we use $\widehat{y_j(t)}$ and $\widetilde{y_j(t)}$ to represent the first and second part in (8) respectively. Based on (2), (3) and (8), $\widehat{y_j(t)}$ can be estimated by

$$\widehat{y_j(t)} = \sum_{l=1}^{\lambda_j} \sum_{V_{i_l} \in \widehat{V_j^{l-}}} \sum_{p_z \in P_{i_l,j}^{l}} \pi_{i_l,j} L^{\tau_{i_l,j}} y_{i_l}(t) \tag{9}$$

Assuming that there are $\hat{\mathcal{N}} \leqslant \mathcal{N}$ links in the road network with measured data. We then define two $\hat{\mathcal{N}} \times 1$ vectors $Y(t) = \{y_j(t) | j \in \hat{\mathcal{N}}\}'$ and $\widehat{Y(t)} = \{\widehat{y_j(t)} | j \in \hat{\mathcal{N}}\}'$. Then, (9) can be expressed as

$$\widehat{Y(t)} = \sum_{l=1}^{\lambda} \hat{\boldsymbol{\phi}}_l Y(t), \tag{10}$$

In (10), $\hat{\boldsymbol{\phi}}_l$ is a $\hat{\mathcal{N}} \times \hat{\mathcal{N}}$ matrix where the $(i, j)^{th}$ entry is $\sum_{p_z \in P_{i,j}^{l}} \pi_{i,j} L^{\tau_{i,j}}$ if $V_i \in \widehat{V_j^{l-}}$, Otherwise, the entry is equal to 0. Beside, we define λ as the maximal value of λ_j, $j \in \hat{\mathcal{N}}$, mathematically, denoted as

$$\lambda = \max_{\forall j \in \hat{\mathcal{N}}} \lambda_j. \tag{11}$$

With estimated results of π, τ and λ (using the methods in the next section), $\widetilde{y_j(t)}$ can be calculated by $y_j(t) - \widehat{y_j(t)}$. We define a $\hat{\mathcal{N}} \times 1$ vector $\widetilde{Y(t)} = \{\widetilde{y_j(t)} | j \in \hat{\mathcal{N}}\}$. Then we construct a STARIMA model for $\widetilde{Y(t)}$, formulated as follows:

$$\widetilde{Y(t)} = \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \widetilde{\boldsymbol{\phi}_{kl}} \boldsymbol{W}_l L^k \widetilde{Y(t)} + \varepsilon_t - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \widetilde{\boldsymbol{\theta}_{kl}} \boldsymbol{W}_l L^k \varepsilon_t. \tag{12}$$

Unlike original STARIMA model where l refers to the spatial order between two vertices, in (12), l is the path length. The $(i, j)^{th}$ entry of W_l is 1 if $V_i \in \widehat{V_j^{l-}}$. Otherwise the entry is 0. Eq. (12) can also denoted as

$$\begin{aligned}(\boldsymbol{I} - \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \widetilde{\boldsymbol{\phi}_{kl}} \boldsymbol{W}_l L^k)(1-L)^d \widetilde{Y(t)} \\ = (\boldsymbol{I} - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \widetilde{\boldsymbol{\theta}_{kl}} \boldsymbol{W}_l L^k)\varepsilon_t.\end{aligned} \tag{13}$$

According to (10), $\widetilde{Y(t)} = Y(t) - \widehat{Y(t)} = (\boldsymbol{I} - \sum_{l=1}^{\lambda} \hat{\boldsymbol{\phi}}_l) Y(t)$. Substituting it into (13), we have the unified spatio-temporal model in the following way:

$$\begin{aligned}(\boldsymbol{I} - \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \widetilde{\boldsymbol{\phi}_{kl}} \boldsymbol{W}_l L^k)(1-L)^d (\boldsymbol{I} - \sum_{l=1}^{\lambda} \hat{\boldsymbol{\phi}}_l) Y(t) \\ = (\boldsymbol{I} - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \widetilde{\boldsymbol{\theta}_{kl}} \boldsymbol{W}_l L^k)\varepsilon_t.\end{aligned} \tag{14}$$

For simplicity, we define $\boldsymbol{\phi}_{\pi,\tau,\lambda_1} = \boldsymbol{I} - \sum_{l=1}^{\lambda} \hat{\boldsymbol{\phi}}_l$, $\boldsymbol{\phi}_{p,\lambda_2} = \boldsymbol{I} - \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \widetilde{\boldsymbol{\phi}_{kl}} \boldsymbol{W}_l L^k$, $\nabla^d = (1-L)^d$, and $\boldsymbol{\theta}_{q,m} = (\boldsymbol{I} - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \widetilde{\boldsymbol{\theta}_{kl}} \boldsymbol{W}_l L^k)$. Finally, (14) can be rewritten as

$$\boldsymbol{\phi}_{\pi,\tau,\lambda_1} \boldsymbol{\phi}_{p,\lambda_2} \nabla^d Y(t) = \boldsymbol{\theta}_{q,m} \varepsilon_t \tag{15}$$

In our model, we put the physical factors potentially affecting such spatio-temporal correlation in the component $\boldsymbol{\phi}_{\pi,\tau,\lambda_1}$, which is independent of $\boldsymbol{\phi}_{p,\lambda_2}$ and $\boldsymbol{\theta}_{q,m}$. Besides, π reflects the trip distribution between adjacent links, and τ reflects the travel time delay between links in terms of the travel speed and route length; λ_1 reflects the number of spatially correlated links surrounding a link of interest. In this case, the accuracy of traffic flow prediction greatly relies on the estimation of $\boldsymbol{\phi}_{\pi,\tau,\lambda_1}$, relies to a lesser extent on $\boldsymbol{\phi}_{p,\lambda_2}$ and $\boldsymbol{\theta}_{q,m}$. The term "unified" in our proposed model is mainly manifested in the following aspects: 1) a day is divided into different time periods (e.g. peak and off peak hours) where traffic state in each time period can be regarded as static. The prediction model (15) in different time periods is identified by only adjusting $\boldsymbol{\phi}_{\pi,\tau,\lambda_1}$, which is estimated using the historical traffic data from the same time period of different days. 2) $\boldsymbol{\phi}_{p,\lambda_2}$ and $\boldsymbol{\theta}_{q,m}$ are required to be estimated once only based on traffic data and $\boldsymbol{\phi}_{\pi,\tau,\lambda_1}$ in any time period of the day. SACF (spatial autocorrelation function) and SPACF (spatial partial ACF) are applied to estimate $\boldsymbol{\phi}_{p,\lambda_2}$ and $\boldsymbol{\theta}_{q,m}$.

Consequently, the challenging problem in model identification is the determination of λ_1, τ, and π, which will be further discussed in the next section.

IV. Methodology for Parameter Estimation

In this section, we first propose the kernel strategies to estimate τ and π. After that, a BFS based algorithm is proposed to estimate λ_1 as well as τ and π. Finally, the computational complexity of model construction is analyzed.

A. Time-Varying Lags τ

Consider two detector stations A and B with distance S where the vehicles keep a stable average speed v, then approximately $t = S/v$ is needed for vehicles to travel from B to A. In other words, the traffic flow collected at station A is strongly correlated with that at B t time ago. Thus the temporal lag with the maximum correlation should be $\tau = [t/t_{lag}]$ where t_{lag} is the length of one temporal lag. As v is time-varying, τ will change over the time. Therefore, we name τ as time-varying lag.

In (6), $\tau_{i,j}^{P_z}$ can be abbreviated as $\tau_{i,j}$ if the length of P_z is $l = 1$. $\tau_{i,j}$ is the time-varying lag between two adjacent links and estimated by

$$\tau_{i,j} = \frac{S_{i,j}}{v_{i,j} t_{lag}}. \tag{16}$$

where $S_{i,j}$ is the distance between V_i and V_j along the road, $v_{i,j}$ is the average traffic speed from link i to link j. Here the situation that $\tau_{i,j}$ may not be an integer is ignored for simplicity.

As a matter of fact, $v_{i,j}$ is the space mean speed (SMS). However, the speed collected by detectors is mostly the time mean speed (TMS) [10], [31]. To derive the SMS from the TMS, a commonly used technique, which is also adopted in this paper, is via the equation $v_{tms} = v_{sms} + \sigma^2/v_{sms}$ where v_{tms} and v_{sms} are the corresponding TMS and SMS respectively and $\sigma^2 = E((v_{ins} - v_{sms})^2)$ with v_{ins} being the instantaneous vehicle speed and $E(v_{ins}) = v_{tms}$. Han *et al.* [31] assumed a quadratic relationship between $E(v_{ins}^2)$ and $E(v_{ins})$ by $E[v_{ins}^2] = aE(v_{ins})^2 + bE(v_{ins}) + c$ where the parameters $\{a, b, c\}$ were estimated using 9304 traffic samples as $\{a, b, c\} = \{1.22, -15.21, 207.95\}$.

In the case that the length l of P_z is $l > 1$, the physical significance of $\pi_{i,j}^{P_z}$ within a sampling time interval t, denoted as $\tau_{i,j}^{P_z,t}$ can be interpreted as the sum of the delay caused by the travel time from link i to link j upon the path P_z. We use Ω to denote a set of time period clusters where the label of a cluster represents a specific time period of the day. We divide the successive time intervals of a day $T = \{1, 2, 3, \ldots\}$ into different clusters where the successive time intervals in a cluster compose a time period $T_m^n \in \Omega_n \subseteq \Omega$. As the classification algorithm is not the focus of this research, we use the ISODATA algorithm given in [8], or roughly make a division according to the observation of the traffic flow data variation. After that, we have $\pi_{i,j}^{P_z}$ within a specific time period of day by

$$\tau_{i,j}^{P_z} = \left[\frac{\sum_{t \in T_m^n} \tau_{i,j}^{P_z,t}}{|T_m^n|}\right], \tag{17}$$

where $\lceil x \rceil$ is the smallest integer that is greater than or equal to x. Further note that, the absence and breakdown of traffic detectors causes data missing, e.g., traffic speed and traffic flow. Thus, the aforementioned way to estimate $\tau_{i,j}^{P_z,t}$ is not available in this situation. Consider traffic flow or traffic speed data are not observable in a link $V_{miss} \in V$, we use a KNN based method [32] to estimate the TMS v_{miss} at V_{miss} by $v_{miss} = \sum_{k=1}^{K} v_k/K$, where $v_k, k \in K$ is the TMS at the k-th nearest links of V_{miss} ordered with respect to the Euclidean distance.

B. Turning Rate Estimation

In (2), $\pi_{i,j}$ is a special case of $\pi_{i,j}^{P_z}$ (in (6)) in the case that the length of P_z is $l = 1$. Indeed, the physical significance of $\pi_{i,j}^{P_z}$ is the ratio of the traffic attached to V_j with the traffic produced in V_i and traveling in P_z. Due to the fact that the turning rates at different intersections along a path are i.i.d., a simple way to estimate $\pi_{i,j}^{P_z}, l > 1$ is the accumulation of the turning rate between any two adjacent links in the path P_z from link i to link j. However, such estimation method is an intuitive, but not a general approach since a prior knowledge of the turning rate between any two adjacent links are needed. As the estimation of turning rate between two adjacent links V_i and V_j is closely correlated with the traffic at these two vertices, it is hard to infer $\pi_{i,j}$ once there is data missing in any link of V_i and V_j.

To overcome the aforementioned problem, we come up with a method motivated by the gravity model that is widely used for estimating the trip distribution between two zones. More precisely, the principle of gravity model states that the number of trips between two traffic zones is directly proportional to the number of trip attractions generated by the destination zone and inversely proportional to a function of travel time between the two zones [33]. Based on the gravity model, we estimate $\pi_{i,j}^{P_z}$ by the following three-steps procedure which only requires the traffic at both ends of a path P_z, rather than the traffic from each link in the path.

- Divide a path P_z into a sequence of concatenate sub-path by $P_z = \cup_s P_{z_s}$ where the links without measured data are distributed into each sub-path P_{z_s}, whereas the links at the both ends of P_{z_s} have measured data;
- Apply a modified gravity model to calculate the turning rate upon sub-path P_{z_s};
- $\pi_{i,j}^{P_z} = \prod_s \pi_{i,j}^{P_{z_s}}$.

Suppose a vertex $V_i \in V$, as well as $\widehat{V_i^{l-}}$ where each $V_j \in \widehat{V_i^{l-}}$ has measured data and there is a path P_z with length l from V_i to V_j. We use $\widehat{P_{i,j}^{l}}$ to denote the collection of paths from V_i to $\forall V_j \in \widehat{V_i^{l-}}$. The gravity model based method is formulated as follows:

$$y_{i,j}^{P_z} = y_i \left[\frac{y_j \mathcal{C}_{i,j}^{P_z} \mathcal{B}_{i,j}^{P_z}}{\sum_{P_z \in \widehat{P_{i,j}^{l}}} y_j \mathcal{C}_{i,j}^{P_z} \mathcal{B}_{i,j}^{P_z}}\right]. \tag{18}$$

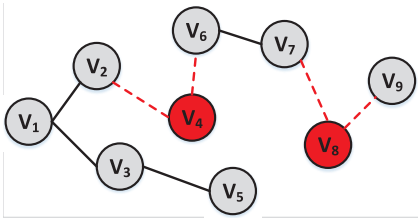


Fig. 2. Consider the path P_z from V_9 to V_1 where V_4 and V_8 have no traffic data. $P_z = P_{z_1} \cup P_{z_2}$ where P_{z_1} is the path from V_9 to V_6, whereas P_{z_2} is the path from V_6 to V_1. We first get the turning rate upon P_{z_1} is $\pi_{9,6}^{P_{z_1}}$, and we also get the the turning rate of P_{z_2} is $\pi_{6,1}^{P_{z_2}}$. Then we have $\pi_{9,1}^{P_z} = \pi_{9,6}^{P_{z_1}} \times \pi_{6,1}^{P_{z_2}}$.

Particularly, when $l = 1$, the turning rate between two adjacent links is calculated. We use t_{P_z} to denote the travel time of vehicles traveling along the path P_z and is calculated by $t_{P_z} = \tau_{i,j}^{P_z} \times t_{lag}$ based on (16). Thus the inverse function of travel time t_{P_z}, $\mathcal{C}_{i,j}^{P_z}$ in (18), can be obtained from the calibration process [33]. $\mathcal{B}_{i,j}$ is socioeconomic adjustment factor for the interchange between vertices V_i and V_j, and in this paper, $\mathcal{B}_{i,j} = 1$. Within a time period $T_m^n \in \Omega_n \in \Omega$, y_i, y_j and $y_{i,j}^{P_k}$ are defined in the following way:

$$\begin{aligned} y_i &= \sum_{t \in T_m^n} y_i(t), y_j = \sum_{t \in T_m^n} y_j(t) \\ y_{i,j}^{P_z} &= \sum_{t \in T_m^n} y_{i,j}^{P_z}(t). \end{aligned} \tag{19}$$

The objective of the gravity model is to estimate $y_{i,j}^{P_z}$ so that $\pi_{i,j}^{p_z}$ can be further estimated by $\pi_{i,j}^{p_z} = \frac{y_{i,j}^{P_z}}{y_i}$. We use an iterative procedure [33] to estimate y_j until convergence is reached:

$$y_{j,w} = \frac{y_j}{\sum_{i \in \mathcal{N}} \sum y_{i,j}^{P_z}} y_{j,w-1}. \tag{20}$$

In (20), w is the iteration number. Finally, we have $\pi_{i,j}^{p_z}$. To better understand the above estimation process, we give an example in Figure 2 where no detectors are configured at V_4 and V_8, causing the turning rates upon the dash and red lines can not be estimated directly.

C. Spatial Order λ_1 and Parameters Estimation Algorithm

To identify λ_1, as well as τ and π, a BFS based algorithm is designed in (1). In order to improve the efficiency, the estimation of τ, π and $\lambda_j, j \in \hat{\mathcal{N}}$ is executed in each vertex concurrently (line 2 to 24). With $\lambda_j, \forall j \in \hat{\mathcal{N}}$, λ_1 is calculated in a centralized way (line 25).

With the determination of τ, π, and λ_1, the uniform STARIMA model is set up according to the following three steps [34]:

- **Model Identification:** using STACF (space-time autocorrelation function) and STPACF (space-time partial autocorrelation function) to determine the maximum lags ($\{p, \lambda, q, m\}$) in the uniform STARIMA model.
- **Parameter Estimation:** estimating the model parameters ($\boldsymbol{\phi}_{p,\lambda_2}$ and $\boldsymbol{\theta}_{q,m}$) by non-linear optimization techniques;
- **Diagnostic Checking:** there are two phases in this process. In the first phase, the residuals will be examined in order to make the model adequately represents the data. In the second phase, it analyzes the statistical significance of the estimated parameters in order to avoid constructing a unduly complex (e.g., overfitting) model.

Algorithm 1 The Estimation of τ, π, and λ_1

1: **τ, π and λ_j for each link** $j \in \hat{\mathcal{N}}$
2: **Initialization:**
3: $P \leftarrow \varnothing$, $\lambda_j = 0$, $Q \leftarrow \varnothing$, $visited = 0$
4: $Q \leftarrow V_j$,$visited[j] = 1$
5: **while** $Q \neq \varnothing$ **do**
6: $V_{temp} \leftarrow$the head in the Q
7: V_i : there is an arc from V_i toV_{temp}
8: **while** $V_i \neq \varnothing$ **do**
9: **if** $visited[i] = 0$ **then**
10: **if** there is traffic data at V_i **then**
11: $\overbrace{V_{temp}^{1+}} \leftarrow V_i$
12: $P \leftarrow P_z$ from V_i to V_j with length l
13: **if** $\lambda_j < l$ **then**
14: $\lambda_j = l$
15: **end if**
16: Estimate $\tau_{i,j}^{P_z}$ and $\pi_{i,j}^{P_z}$
17: **else if** there is no traffic data at V_i **then**
18: $visited[i] = 1$
19: $Q \leftarrow V_i, \overbrace{V_{temp}^{1+}} \leftarrow V_i$
20: **end if**
21: V_i : the next vertex that there is an arc from V_i to V_{temp}
22: **end if**
23: **end while**
24: **end while**
25: **Calculate λ_1 using** (11)

The parameters τ, π and λ_1 in the modified STARIMA model can be regarded as "hyper parameters" like those in deep learning model, e.g., the number of hidden layers. The difference is that these hyper parameters have physical meanings. Besides, with Algorihtm 1, these hyper parameters can be easily estimated (the algorithm complexity is analyzed in the following paragraph). Comparing with most studies where a lot of parameters have to estimate due to the fact that multiple models, particularly non-parametric models built in different time periods of the day, the distinguished advantage of our method is that we only need to adjust τ, π and λ_1 in different time periods of the day, rather than re-estimating all the parameters repeatedly. Thus, the accuracy complexity trade-off can be guaranteed.

We now analyze the computational complexity of parameters estimation in the STARIMA model. According to literature [35], Dave suggested that the computational complexity of identifying parameter p using ACF (autocorrelation function, resp. parameter q using PACF, partial autocorrelation function) for the ARIMA model is $O(N_S N_l)$ where N_s is the

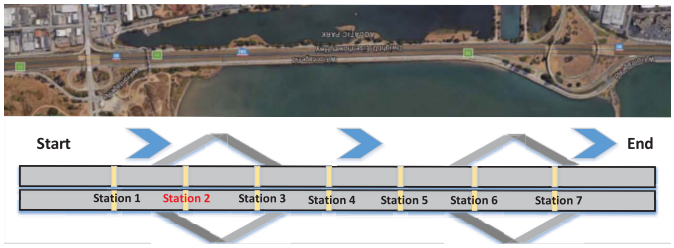


Fig. 3. The map and the topology of considered segment in I-80 freeway.

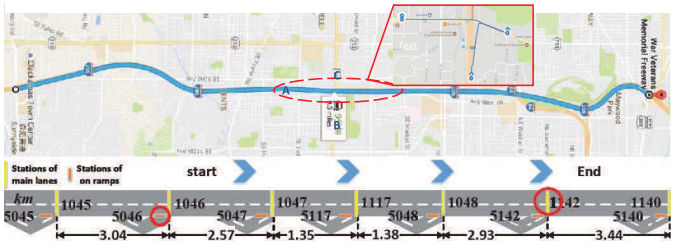


Fig. 4. The map and topology of I-205 NB freeway.

number of samples from an observation and N_l is the number of time lags [35]. Unlike the ARIMA model, the parameters p and λ_ks in the STARIMA model are identified by STACF (resp. STPACF for q and m_ks). Thus, the computational complexity of calculating STACF (STPACF) between two links is $O((\mathcal{N}-1)N_l N_s)$ where $\mathcal{N}-1$ is the maximal spatial lag between two links. Consider any pair of links and the number of time periods n in a day, we have the computational complexity of parameters estimation in the STARIMA model is $O(nN_l N_s \mathcal{N}^3)$.

The computational complexity of parameters estimation for the unified spatio-temporal model mainly relies on the following two parts: 1) the identification of τ, π, and λ which relies on Algorithm 1 for paths searching, executed by the $\mathcal{N}$ vertices in parallel with computational complexity $O(\mathcal{N}^2)$ 2) the identification of parameters in STARIMA model that has the complexity is $O(N_l N_s \mathcal{N}^3)$. As a result, the total computational complexity is $O(n\mathcal{N}^2 + N_l N_s \mathcal{N}^3) = O((\frac{n}{\mathcal{N}} + N_l N_s)\mathcal{N}^2)$. Generally speaking, $\frac{n}{\mathcal{N}} + N_l N_s \ll nN_l N_s$ when a large amount of samples is considered at each observation.

V. Experimental Validation

A. Experimental Setup

In order to verify the performance of the proposed model, two datasets are used, thereafter referred as the dataset from one-dimensional freeway and the dataset from two-dimensional freeway incorporating on- and off-ramps (Fig.3 and 4).[1] The reason for using different datasets is the need for exploring the impact of different road network topology on the prediction accuracy of the unified spatio-temporal model. For example, with the aid of the dataset from the two-dimensional network, we can clearly present the estimation of turning rate with the methods provided in Section IV-B.

The dataset in the first group is sampled from six dual-loop detector stations deployed on a road segment of Interstate 80 (I-80) freeway in Emeryville, California, which are numbered by 1, 3, 4, 5, 6 and 7 (Fig.3). Furthermore, 10-days traffic data is recorded with sampling interval of 30 seconds ($t_{lag} = 30s$). We regard the mean traffic flow of every 3 data points as one data point. Thus, 960 (2880/3)/day $\times$9 data points are available for training model, and the data in the last day are used for prediction.

The dataset in the second group is collected from I-205 NB Portland-area freeway. The freeway in Fig. 4 covers 10.09 miles (16.24km) including a major road with on- and off-ramps. In addition, the freeway is equipped with 14 detector stations to record the traffic traveling from north to south. Particularly, we select the data within 10 working days (Monday to Friday) from Sept. 19, 2011 to Sept. 30, 2011 with sampling interval of 20 seconds ($t_{lag} = 20s$). The locations of the detectors are marked by yellow and orange lines in the figure. The yellow lines are the detectors installed at the major road, while the orange lines are the detectors installed at the entrance from the on-ramp to the major road. The station surrounded by the red circle means there is no available traffic data. We use the first 9-days data to train the model and the data in the last day to validate the prediction. Theoretically, there should be 4320 data at each station in one day. Unfortunately, there are some missing and dirty data inside. Hence, we use a commonly used way, named historical average, to replace the missing data by the average of the known values [22], [36].

We compare our proposed model (denoted as uSTARIMA) with other three approaches, respectively the STARIMA(p, q, λ, m) (denoted as STARIMA), multiple STARIMA based method (denoted as STARIMA*) in which the parameters and coefficients would be re-evaluated in different time periods of the day, and the BPNN method. The STARIMA and STARIMA* are both linear predictive method, while BPNN is a non-linear predictive method. We use a $4 \times 20 \times 1$ BPNN model including a hidden layer and an output layer to predict the traffic flow at each measurement point. There are 4 input nodes which respectively denote the traffic flow data collected from the same measurement at t, $t - 10min$, $t - 20min$ and $t - 30min$. There are 20 nodes in the hidden layer and one node in the output layer. The initial weights are randomly distributed inside a range $[-0.12, 0.12]$ and the thresholds have initial values of 0. We use the sigmoid function as the active function. Besides, we set the momentum coefficient to be 0.7, and the learning rate to be 0.3. A gradient descent optimization algorithm is used to adjust the weights and thresholds by calculating the gradient of the loss function iteratively until the sum of squared errors is no more than the learning error set by 0.01. We use R language running on 64-bit Windows system with 4 CPUs and 16G RAM. With the aid of starma [2] and neuralnet packages,[3] we develop our uniform

[1]The first set of data can be downloaded from: http://ngsim-community.org. The second set of data can be downloaded from: http://portal.its.pdx.edu.

[2]https://cran.r-project.org/web/packages/starma/starma.pdf

[3]https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf

TABLE I

THE TIME-VARYING LAGS BETWEEN STATIONS WITH SPATIAL ORDER $l \geqslant 2$

From	s_3	s_3	s_4
To	s_5	s_6	s_6
0:00am-6:00am	1	2	1
6:00am-9:00am	2	3	2
9:00am-16:00pm	1	2	1
16:00pm-18:00pm	2	3	2
18:00pm-24:00pm	1	2	1

TABLE II

THE MAPE/MSE OF ONE-DAY TRAFFIC FLOW PREDICTION USING USTARIMA, STARIMA*, STARIMA AND BPNN

St.	uSTARIMA	STARIMA*	STARIMA	BPNN
s_3	17.80%/206.33	14.86%/164.21	22.19%/245.37	31.52%/315.22
s_4	17.12%/191.25	15.84%/179.06	25.68%/259.24	24.75%/238.97
s_5	15.13%/178.59	14.92%/159.57	20.41%/209.73	30.21%/334.61
s_6	14.41%/136.27	12.65%/112.44	23.77%/142.57	22.45%/215.72

model as well as the other counterparts. Particularly, starma packages integrated three-stage iterative modeling procedure. In order to verify the prediction accuracy and the efficiency of the proposed scheme, the metrics of the mean square error (MSE), the mean absolute percentage error (MAPE) and running time are considered. More precisely, let $\hat{y}$ be the estimate of N-dimensional vector y, then MSE can be expressed as $MSE(\hat{y}, y) = 1/N \sum_{n=1}^{N} (\hat{y}_n - y_n)^2$, and MAPE is calculated by $MAPE(\hat{y}, y) = 1/N \sum_{n=1}^{N} |\frac{\hat{y}_n - y_n}{y_n}|$.

B. Experimental Results for One-Dimensional Freeway

According to the traffic data collected at stations 3 and 6 on I-80 freeway, we intuitively set $\Omega_1 \in \Omega$ (peak hour) by $\Omega_1 = \{T_1^1\}$ where T_1^1 covers the time period from 6:30am to 9:00am. Correspondingly, Ω_2 (off-peak hour) is the set of time periods outside the range of 6:30am-9:00am. Hereafter, we provide the MAPE/MSE of the traffic prediction at different time of the day in Table II.

From the experimental results, we can observe that the best performance is the one achieved by STARIMA*. Such phenomenon can be explained by the fact that the simple road topology structure enables the time-varying spatio-temporal correlation can be successfully captured by re-estimating all the parameters ($\{p, q, \lambda, m\}$ and $\{\phi_{p,\lambda}, \theta_{q,m}\}$) of STARIMA* in each time period of the day. In comparison, our proposed uSTARIMA model has a slight increase in prediction error ($\sim$ 3% of the measured value). The loss of accuracy is caused by that a nearly monotonous structure of uSTARIMA is set up in different time periods. In other words, there is no obvious difference between the parameters $\{\pi, \tau, \lambda_1\}$ of uSTARIMA in different time periods. For instance, the turning rate between any two adjacent links is a constant value "1" since there is no intersection in the study site. Further, as the road segment (between s_3 and s_6) is not long, we can find the maximal time-varying lag is 3 between s_3 and s_6 in peak hour from Table I. On the contrary, the minimal time-varying lag is 1 between s_3 and s_5 in off-peak hour. The time-varying lags between any adjacent stations are not presented in Table I since the values are smaller than 1, but

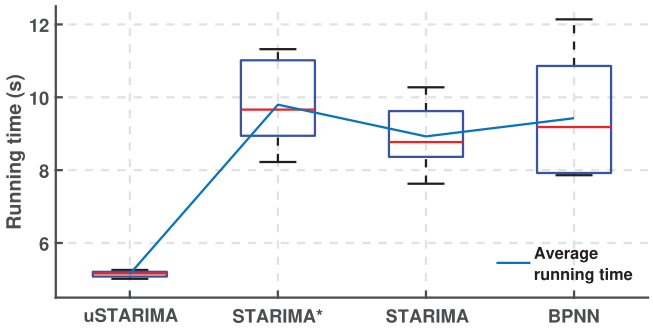


Fig. 5. The running time of STARIMA(Ξ), STARIMA*, ARIMA* and BPNN.

TABLE III

THE TIME VARYING LAG BETWEEN TWO NEIGHBORING LINKS ON THE MAJOR ROAD IN DIFFERENT TIME PERIODS OF THE DAY

From	1045	1046	1047	1117	1048
To	1046	1047	1117	1048	1140
0:00am-6:00am	5	4	2	2	10
6:00am-9:00am	6	5	3	3	12
9:00am-16:00pm	5	4	2	2	10
16:00pm-18:00pm	6	5	3	3	13
18:00pm-24:00pm	5	4	2	2	10

approximately equal to 1 according to formulation (17). Also, $\lambda_1 = 3$ based on the graph model of the study site. The slight change of $\{\pi, \tau, \lambda_1\}$ has no significant impact on the estimation of $\phi_{\pi,\tau,\lambda_1}$. Thus, the performance of uSTARIMA is a little worse than STARIMA*. In practice, the forecasting accuracy of uSTARIMA is sensitive to the fluctuation of above three parameters $\{\pi, \tau, \lambda_1\}$. This can be observed from the experimental results on the basis of the study site in the second group, which we will illustrate in the next subsection. In addition, comparing with the STARIMA and BPNN technique, unsurprisingly, the proposed technique achieves much better prediction accuracy. Particularly, the MAPE of the proposed technique is at least 5%, at most 15% better than that achieved by these two techniques.

Fig. 5 shows the running time of each approach. From the figure, we know that the running time of uSTARIMA is much less than the other two methods, attributable to the unified model employed for traffic prediction during different time periods.

C. Experimental Results for Two-Dimensional Network

Based on the traffic data collected at 6 stations on the major road of I-205 NB freeway, we intuitively divide a day into three time periods. Specially, $\Omega_1 = \{T_1^1, T_2^1\}$ where T_1^1 covers the time period from 6:00am to 9:00am and T_2^1 covers the time period from 16:00pm to 18:00pm. Correspondingly, Ω_2 consists of the set of time periods outside of T_1^1 and T_2^1 . Since the major road in the freeway is long, we divide it into a set of links where each detector station is distributed in one link. In this paper, we mainly provide the traffic prediction at each detector station on the major road.

We list the time-varying lags between two neighboring links in the major road in Table III. It further verifies that the time-varying lag has a close relation with the distance of

TABLE IV
THE ESTIMATION OF TURNING RATES AT THE INTERSECTIONS OF MAJOR ROAD AND OFF-RAMPS (GRAVITY BASED METHODS/DATA-DRIVEN METHOD)

From	1045		1046		1047		1117		1048	
To	1046	Off	1047	Off	1117	Off	1048	Off	1140	Off
0:00am-6:00am	0.72/-	0.28/-	0.81/0.78	0.19/0.22	0.77/0.69	0.23/0.31	0.86/0.80	0.14/0.20	0.87/0.86	0.13/0.14
6:00am-9:00am	0.73/-	0.27/-	0.83/0.79	0.17/0.21	**0.48/0.54**	**0.52/0.46**	0.86/0.79	0.14/0.21	0.89/0.90	0.11/0.10
9:00am-16:00pm	0.71/-	0.29/-	0.88/0.90	0.12/0.10	**0.56/0.55**	**0.44/0.45**	0.88/0.84	0.12/0.16	0.88/0.79	0.12/0.21
16:00pm-18:00pm	0.73/-	0.27/-	0.89/0.76	0.11/0.24	**0.43/0.47**	**0.57/0.53**	0.79/0.77	0.21/0.23	0.84/0.73	0.16/0.27
18:00pm-24:00pm	0.71/-	0.29/-	0.89/0.88	0.11/0.12	0.79/0.66	0.21/0.34	0.88/0.85	0.12/0.15	0.91/0.86	0.09/0.14

two stations, as well as different travel speeds during different time periods of the day. Then, we calculate the time-varying lags between stations by means of Algorithm 1. For instance, the time-varying lag between station 1045 and 1117 is 14 $(6+5+3)$ within 6:00am-9:00am, while 11 $(5+4+2)$ within 9:00am-16:00pm.

As the vehicles coming from on-ramps will move into the major road, the turning rate at the intersection between on-ramps and major road is equal to 1. However, the vehicles at the intersection of off-ramps and major road have two alternatives. One is leaving the freeway through off ramps, and the other one is to keep traveling straightly on the major road. The results of turning rate estimation are presented in Table IV. Table IV presents the turning rates respectively estimated by gravity based method and data-driven method. We regard the data-driven based method as the "actual scenario" where the turning rate at an intersection is calculated by the ratio of "the traffic flow streaming into the off-ramps" to "traffic flow traveling from the major road". As there is no detector configured in the off-ramps, we cannot obtain the traffic flow streaming into the off-ramps directly. However, we can roughly estimate it using the traffic flow data collected at two adjacent detector stations as well as the stations configured in the on-ramp between these two stations. For instance, suppose we have time-spaced traffic flow data at station 1046, 1047 and 5047, respectively y_{1046}, y_{1047} and y_{5047}. Then the traffic flow streaming into the off-ramps between station 1046 and 1047 is calculated by "$y_{off} = y_{1046} - (y_{1047} - y_{5047})$". Given a time period T, we estimate turning rate by "$\frac{\sum_T y_{off}}{\sum_T y_{1046}}$". Note, there is no data at station 5046, thus, we can only use the data-driven method to estimate the turning rates at the other intersections. To save space, the values of the turning rates estimated by both methods are rounded off to the two decimal places. From Table IV, we can observe that the results obtained from gravity based methods are approximately the same as the ones calculated by data-driven method. As we have mentioned in Section IV-B, one advantage of gravity based model is that it can be used to estimate the turning rate even there is missing data in some roads such as the estimation turning rates between station 1045 and off-ramp at the second column (with '-'). Except for the turning rates labeled in red, all the other turning rates show that above 70% vehicles will keep traveling on the major road. As for the turning rates between station 1047 and the downstream off-ramp, we observe that the off-ramp is connected with a road named "SE Powell Blvd" across the segment between station 1047 and 1117 (circled by the dashed line in red). In [37], Stoll *et al*. indicated that Powell Blvd road was a major arterial road in the Portland metropolitan area and carried between 45,000 and 30,000 vehicles a day. The large traffic volume in Powell Blvd road implies that a lot of vehicles will leave the major road and move into Powell Blvd road (e.g. the vehicles traveling from A to B or from A to C). Therefore, in each time period of the daytime (from 6:00 am to 18:00 pm), the turning rate between station 1047 and 1117 are less than the ones estimated between any other pairs of stations. The discrepancies in the estimated turning rates further verify our idea that road trip distribution has a critical influence on the analysis of spatio-temporal correlation.

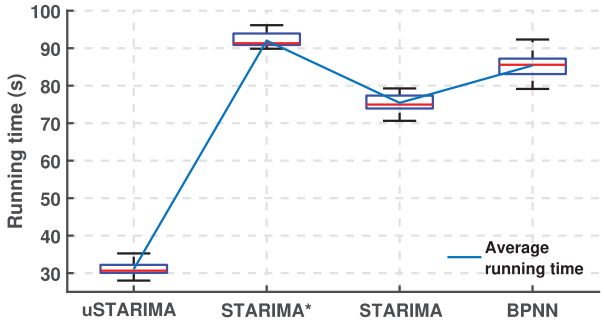


Fig. 6. The running time of uSTARIMA, STARIMA*, STARIMA and BPNN.

TABLE V
THE MAPE/MSE OF ONE-DAY TRAFFIC FLOW PREDICTION USING USTARIMA, STARIMA*, STARIMA AND BPNN

St.	uSTARIMA	STARIMA*	STARIMA	BPNN
1045	24.21%/165.57	17.08%/159.75	25.14%/202.69	41.14%/452.84
1046	19.29%/184.91	18.32%/175.76	23.72%/195.07	28.43%/394.25
1047	**12.57%/103.54**	12.78%/119.49	22.60%/154.25	34.26%/405.67
1117	35.95%/413.51	29.97%/297.06	46.81%/594.62	45.28%/481.24
1048	**15.72%/116.64**	16.54%/139.72	17.19%/130.84	35.11%/375.22
1140	19.03%/121.13	15.46%/108.54	24.11%/115.07	37.37%/326.33

With turning rate and time-varying lag estimated above, we predict the traffic flow at 6 stations (without station 1142). From Table V, we can see that the MAPE of our proposed model is at most $\sim$ 6% (at stations 1046 and 1117) lower than STARIMA*. Note that the forecasting results obtained from our proposed model have the best accuracy. This can be illustrated that the time-varying spatio-temporal correlation affected by the frequent variation of travel speed and trip distribution in the study site of I-205NB freeway can be better captured by the introduced parameters $\{\pi, \tau, \lambda_1\}$

in our method. For instance, given two stations 1047 and 1048, the gap between the time-varying lag in peak and off-peak hours can be 7 (30 in peak hour from 16:00 pm to 18:00 pm and 23 in off-peak hour from 9:00 am to 16:00 pm). Distinctly different from uSTARIMA, the determination of lags in STRAIMA* is by means of STACF and STPACF which depend on the assumption that we are comfortable making with respect to the constancy of the trend in the data. It is difficult to select accurate number of lags. Thus, the forecasting accuracy of STARIMA* will be reduced in some cases, e.g., station 1047 and 1048. Unsurprisingly, in the worst case there is $\sim$ 22% (at station 1117) gap between the MAPE of uSTARIMA and BPNN, and at worst $\sim$ 10% (at station 1047) gap between the MAPE of uSTARIMA and STARIMA.

In Fig. 6, we present the running time of different methods. It is clear to see that less time is consumed for our proposed model, which is consistent with the result in Fig.5. Based on the results in Table V and Fig.6, it is sufficient to say that our proposed model is also available for the two-dimensional road network.

VI. Conclusions

In this paper, we developed a unified spatio-temporal model, which does not need a complete re-design and calibration of the prediction model for short-term traffic flow prediction during the day. In the model, the spatio-temporal traffic correlation is captured by the turning rate at the intersections, as well as the time-varying lag which is formulated as a function of the spatial separation and the travel speed between two measurement points. Fundamentally, a better performance is achieved because, instead of using a black-box approach to model the traffic correlations, the proposed method explicitly takes into account the road topology, trip distribution and travel speed and offers a physically intuitive approach to capturing the spatio-temporal correlation between traffic at different locations. In this sense, a deeper insight revealed through our work is that by incorporating the knowledge of the underlying road topology into traffic prediction, a better accuracy can be achieved. As the STARIMA model is unable to capture the non-linear trend of traffic data, thereby, it is part of our future work to explore other non-linear models with spatio-temporal correlations to solve traffic prediction problems.

References

[1] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 61–78, May 2016.

[2] A. Abadi, T. Rajabioun, and P. A. Ioannou, "Traffic flow prediction for road transportation networks with limited traffic data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 653–662, Apr. 2015.

[3] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang, "A distributed spatial–temporal weighted model on mapreduce for short-term traffic flow forecasting," *Neurocomputing*, vol. 179, pp. 246–263, Feb. 2016.

[4] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[5] P. Lopez-Garcia, E. Onieva, E. Osaba, A. D. Masegosa, and A. Perallos, "A hybrid method for short-term traffic congestion forecasting using genetic algorithms and cross entropy," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 557–569, Feb. 2016.

[6] E. Y. Kim, "MRF model based real-time traffic flow prediction with support vector regression," *Electron. Lett.*, vol. 53, no. 4, pp. 243–245, Feb. 2017.

[7] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C Emerg. Technol.*, vol. 19, no. 4, pp. 606–616, 2011.

[8] P. Duan, G. Mao, C. Zhang, and S. Wang, "STARIMA-based traffic prediction with time-varying lags," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 1610–1615.

[9] G. Comert and A. Bezuglov, "An online change-point-based model for traffic parameter prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1360–1369, Sep. 2013.

[10] J. Ahn, E. Ko, and E. Y. Kim, "Highway traffic flow prediction using support vector regression and Bayesian classifier," in *Proc. Int. Conf. Big Data Smart Comput. (BigComp)*, Jan. 2016, pp. 239–244.

[11] Y. Xie, Y. Zhang, and Z. Ye, "Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 22, no. 5, pp. 326–334, 2007.

[12] D. Billings and J.-S. Yang, "Application of the ARIMA models to urban roadway travel time prediction—A case study," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, vol. 3. Oct. 2006, pp. 2529–2534.

[13] B. Williams, P. Durvasula, and D. Brown, "Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transp. Res. Rec., J. Transp. Res. Board*, no. 1644, pp. 132–141, 1998.

[14] T. Cheng, J. Wang, J. Haworth, B. Heydecker, and A. Chow, "A dynamic spatial weight matrix and localized space–time autoregressive integrated moving average for network modeling," *Geograph. Anal.*, vol. 46, no. 1, pp. 75–97, 2014.

[15] S.-D. Oh, Y.-J. Kim, and J.-S. Hong, "Urban traffic flow prediction system using a multifactor pattern recognition model," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2744–2755, Oct. 2015.

[16] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, 1996.

[17] L. Song, "Improved intelligent method for traffic flow prediction based on artificial neural networks and ant colony optimization," *J. Converg. Inf. Technol.*, vol. 7, no. 8, pp. 272–280, 2012.

[18] J. Chen, K. H. Low, Y. Yao, and P. Jaillet, "Gaussian process decentralized data fusion and active sensing for spatiotemporal traffic modeling and prediction in mobility-on-demand systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 3, pp. 901–921, Jul. 2015.

[19] Y.-S. Jeong, Y.-J. Byon, M. M. Castro-Neto, and S. M. Easa, "Supervised weighting-Online learning algorithm for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1700–1707, Dec. 2013.

[20] B. Williams, "Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling," *J. Transp. Res. Board*, no. 1776, pp. 194–200, 2001.

[21] A. Stathopoulos and G. M. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transp. Res. C, Emerg. Technol.*, vol. 11, no. 2, pp. 121–135, 2003.

[22] S. Sun, C. Zhang, and G. Yu, "A Bayesian network approach to traffic flow forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 124–132, Mar. 2006.

[23] E. J. Horvitz, J. Apacible, R. Sarin, and L. Liao. (2012). "Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service." [Online]. Available: https://arxiv.org/abs/1207.1352

[24] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017.

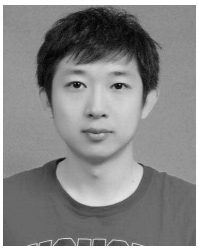
[25] N. Mitrovic, M. T. Asif, J. Dauwels, and P. Jaillet, "Low-dimensional models for compressed sensing and prediction of large-scale traffic data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2949–2954, Oct. 2015.

[26] A. Salamanis, D. D. Kehagias, C. K. Filelis-Papadopoulos, D. Tzovaras, and G. A. Gravvanis, "Managing spatial graph dependencies in large volumes of traffic data for travel-time prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1678–1687, Jun. 2016.

[27] F. Kelly, "The mathematics of traffic in networks," in *The Princeton Companion to Mathematics*, vol. 1. no. 1, T. Gowers, Ed. Princeton, NJ, USA: Princeton Univ. Press, 2008, pp. 862–870.

[28] M. Ben-Akiva, M. Bierlaire, D. Burton, H. N. Koutsopoulos, and R. Mishalani, "Network state estimation and prediction for real-time traffic management," *Netw. Spatial Econ.*, vol. 1, nos. 3–4, pp. 293–318, Sep. 2001.

[29] T. Djukic *et al.*, "Advanced traffic data for dynamic OD demand estimation: The state of the art and benchmark study," in *Proc. TRB 94th Annu. Meeting Compendium Papers*, 2015, pp. 1–16.
[30] Y. Zhang, M. Roughan, C. Lund, and D. Donoho, "An information-theoretic approach to traffic matrix estimation," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun.*, 2003, pp. 301–312.
[31] J. Han, J. W. Polak, J. Barria, and R. Krishnan, "On the estimation of space-mean-speed from inductive loop detector data," *Transp. Planning Technol.*, vol. 33, no. 1, pp. 91–104, 2010.
[32] U. Yildirim and Z. Çataltepe, "Short time traffic speed prediction using data from a number of different sensor locations," in *Proc. 23rd Int. Symp. Comput. Inf. Sci. (ISCIS)*, Oct. 2008, pp. 1–6.
[33] N. J. Garber and L. A. Hoel, *Traffic & Highway Engineering*. Boston, MA, USA: Cengage, 2014.
[34] P. E. Pfeifer and S. J. Deutrch, "A three-stage iterative procedure for space-time modeling phillip," *Technometrics*, vol. 22, no. 1, pp. 35–47, 1980.
[35] D. Hale, "An efficient method for computing local cross-correlations of multi-dimensional signals," in *CWP-544: Consortium Project on Seismic Inverse Methods for Complex Structures*, no. 656. CO, USA: Center for Wave Phenomena, 2006, pp. 253–260.
[36] L. Qu, L. Li, Y. Zhang, and J. Hu, "PPCA-based missing data imputation for traffic flow volume: A systematical approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 512–522, Sep. 2009.
[37] N. B. Stoll, T. Glick, and M. A. Figliozzi, "Using high-resolution bus GPS data to visualize and identify congestion hot spots in urban arterials," *Transp. Res. Rec., J. Transp. Res. Board*, no. 2539, pp. 20–29, 2016.



Peibo Duan (S'16–M'18) received the B.S. and M.S. degrees from Northeastern University, Shenyang, China, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree with the School of Computing and Communication, The University of Technology Sydney, under the supervision of Prof. G. Mao. His current research interests include intelligent transportation system and distributed constraint optimization problem.

Guoqiang Mao (S'98–M'02–SM'08–F'18) joined The University of Technology Sydney as a Professor with the Wireless Networking and Director of Center for Real-time Information Networks in 2014. He has published about 200 papers in international conferences and journals, which have been cited over 5000 times. His research interests include intelligent transport systems, applied graph theory and its applications in telecommunications, Internet of Things, wireless sensor networks, wireless localization techniques, and network performance analysis. He is a fellow of IET.

Weifa Liang (M'99–SM'01) received the B.Sc. degree from Wuhan University, China, in 1984, the M.E. degree from the University of Science and Technology of China in 1989, and the Ph.D. degree from the Australian National University in 1998, all in computer science. He is currently a Full Professor with the Research School of Computer Science, Australian National University. His research interests include design and analysis of energy efficient routing protocols for wireless ad hoc and sensor networks, cloud computing, software-defined networking, design and analysis of parallel and distributed algorithms, approximation algorithms, combinatorial optimization, and graph theory.

Degan Zhang (M'01) was born in 1970. He received the Ph.D. degree from Northeastern University, China. He is currently a Professor with the Tianjin Key Lab of Intelligent Computing and Novel Software Technology, Key Lab of Computer Vision and System, Ministry of Education, Tianjin University of Technology, Tianjin, China. His research interests include IOT, WSN, and IOV.