A Trade-off between Accuracy and Complexity: Short-term Traffic Flow Prediction with Spatio-temporal Correlations

Peibo Duan[1], Guoqiang Mao[2], Changsheng,Zhang[3], and Jun Kang[4]

***Abstract*— Considering spatio-temporal correlation between traffic in different roads has benefit for building an accurate spatio-temporal model for traffic prediction. However, it implies high computational complexity for model building in the context of a complicated network topology, e.g., urban network. Hence, this paper develops a method for capturing and quantifying the intricate spatio-temporal correlations. The contributions of this paper are as follows. First, we offer a physically intuitive approach to capture the spatio-temporal correlation between traffic in different roads, which is related to the road network topology, time-varying speed, and time-varying trip distribution. With this approach, only the parameters, namely time-varying lags, in our STARIMA (Space-Time Autoregressive Integrated Moving Average) based model should be adjusted in different time periods of the day. It guarantees the prediction accuracy and makes the predictor readily amendable to suit changing road and traffic conditions. Second, a metric named traffic transition probability calculated based on trip distribution, as well as a threshold ε are applied for selecting the most spatio-temporally correlated neighbors of a target road. Thus, the complexity of model building will be reduced dramatically. Trace-driven experiments are conducted from two aspects. First, our proposed prediction method has superior accuracy compared with ARIMA and the back propagation neural network model (BPNN) based method, but has much reduced computational complexity. Second, the results show that the prediction accuracy is not always proportional to the increase in the number of spatial neighbors considered for a target road. The trade-off between accuracy and complexity depends on the configuration of ε.**

I. INTRODUCTION

Traffic flow prediction, as an integral component in many Intelligent Transportation Systems (ITS) applications, has the potential to help individual travelers to make a better travel decision, alleviate traffic congestion and reduce carbon emissions. Generally speaking, traffic flow prediction can be broadly classified into short-term traffic prediction (ranging from 5 minute to 30 minutes) and long-term traffic prediction (ranging from 30 minutes to several days or months). In particular, short-term traffic prediction provides the much-needed information to drivers for real-time route optimization and travel time prediction, and to traffic authorities for road traffic management and traffic signal control. Thus, it has spurred great interest from traffic engineers and researchers in this area [1], [2].

Recently, developing models that consider spatio-temporal correlation between traffic data in analysis became in the center of interest [3], [2], [4]. Unfortunately, existing methods still have certain challenges which are discussed as follows. First, the spatio-temporal correlation between traffic at different observation points is not stationary but varies with time of the day [3]. To this end, multiple prediction models corresponding to different times of the day have been constructed to suit time-varying spatio-temporal traffic correlations [5], [6]. Obviously, the complexity will be increased due to the re-estimation of large amount of model parameters. Third, 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 traffic conditions. Lastly, in a complicated road network, e.g., urban network, 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 paper, we develop an effective method to capture and quantify the intricate spatio-temporal correlation between the traffic in different roads. To evaluate the effectiveness of this method, we further design a spatio-temporal model based on STARIMA (Space-Time Autoregressive Integrated Moving Average). Specifically, the contributions of the paper are briefly summarized as follows:

- A physically intuitive approach to traffic prediction is developed that captures the time-varying spatio-temporal correlation between traffic in different roads, which is related to the road network topology, time-varying speed, and time-varying trip distribution. With this approach, only the parameter, namely time-varying lag , in our STARIMA based model should be adjusted in different time periods of the day. It guarantees the prediction accuracy and makes the predictor readily amendable to suit changing road and traffic conditions
- A metric named traffic transition probability is calculated based on trip distribution and applied for selecting the most spatio-temporally correlated neighbors of a target road with the aid of a threshold ε. Thus, the complexity of model building will be reduced dramatically.
- Experiments using real traffic traces are conducted,

[1]Peibo Duan is with School of Computing and Communications, University of Technology, Sydney, Australia `Peibo.Duan@student.uts.edu.au`

[2]Guoqiang Mao is with the School of Computing and Communication, The University of Technology Sydney, Australia. He is also with Data61, CSIRO, Australia `Guoqiang.Mao@uts.edu.au`

[3]Changsheng Zhang is with School of Computer Application, Northeastern University, Shenyang, China `zhangchangsheng@ise.neu.edu.cn`

[4]Jun Kang is with School of Information Engineering, Chang'an University, Xi'an, China. Jun Kang is the corresponding author `junkang@chd.edu.cn`

which demonstrate that the proposed approach has superior accuracy compared with the ARIMA and the back propagation neural network model (BPNN, back propagation neural network) based approaches, but much reduced computational complexity. In addition, the results show that the prediction accuracy is not always proportional to the increase in the number of spatial neighbors considered for a target road. The trade-off between accuracy and complexity depends on the configuration of ε.

The rest of the paper is organized as follows. In Section II, existing research closely related to our work is reviewed. We illustrate the methodologies in Section III. The performance of our proposed methods are evaluated in Section IV. Finally, Section V draws the conclusion.

II. Related Work

Temporal models for traffic flow prediction have been extensively applied in the past two decades. Particularly, a kind of typical representatives are ARIMA-based models [5], [7] which integrate knowledge of the underlying traffic process in the form of traffic models. Anthony and Karlaftis [7] developed a multivariate ARIMA model, denoted as ARIMAX where the parameters are re-estimated in different time periods of the day (e.g. peak hours and off-peak hours). Similarly, Gurcan and Anton [5] built different ARIMA models in different traffic states (Major Accident, Minor Incident, Instability, Normal Driving) for the same purpose.

Spatio-temporal models are another category of traffic prediction techniques that has been widely used in recent years. The commonly used methods are on the basis of the STARIMA model which describes the spatial relations among locations by means of a spatial weight matrix. In order to improve the predictive performance of STARIMA model, a lot of work utilized some techniques which could capture the impact of traffic parameters (i.e. travel speed) on the spatial weight matrix [8], [9], [10]. Min and Wynter [8] proposed a new way to calculate the spatial order between two locations by dividing their physical distance by average speed in different time periods. As a result, a dynamic spatial weight matrix would be re-evaluated in different time periods. Similarly, Tao et al. [9] incorporated the concepts of dynamic spatial weights and dynamic spatial neighborhood by the dynamic spatial weight matrix in order to better capture the spatial heterogeneity and temporal nonstationarity in network data.

Due to the nature that more information (spatial and temporal information) is required in the STARIMA-based model, a large number of parameters and coefficients should be estimated. Accordingly, the high computational complexity makes such category of models incapable of providing an accurate forecast in a timely manner when the network is large and road structure is complex. In practice, we only need to predict the traffic flow in a road of interest with the traffic in its most spatio-temporally correlated neighbors. However, little research has provided solutions. A distinguished work is done by Athanasios et al. [2] who developed a more efficient way to evaluate the spatial weight matrix when the road structure was complex. More precisely, they used a graph matrix to manage spatial dependence between roads. With this graph matrix, the most spatio-temporally correlated neighbors of a road were determined by the score calculated according to the Pearson product-moment correlation-coefficient-based metric. However, this method can be regarded as "gray-box" approach since there is no theoretical analysis on the variation of spatio-temporal correlation. To overcome this shortcoming, our previous work [3] proposed a convenient technique to adjust the lags of the STARIMA model dynamically to suit different traffic states. This was validated using measured traffic data on a one-dimensional road segment.

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, traffic flow prediction or traffic flow demand estimation and so forth. To extend the trip distribution matrix to the digraph model, we propose the concepts of turning rate and traffic transition probability which are capable of accurately capturing the traffic distribution among roads with road intersections.

III. Methodologies

A. Road Network Model

We decompose a road network into a series of one-dimensional road segments. Each road segment is defined as a route between two intersections. As a road segment may have two ways, we use the concept of "**link**" defined in [11] to represent a particular way in a road segment. Hence, a road segment consists of one or two links. Without loss of generality, we assume only one detector station is equipped in a link. If there are more than one detectors, we can further divide the link into smaller links where each small link satisfies the aforementioned assumption. Based on the above definitions, we model a road network with $\mathcal{N}$ links as a digraph, denoted by $D = (V, E)$. Specifically, the vertex set $V = \{V_1, V_2, ..., V_{\mathcal{N}}\}$ and $V_i \in V$ represents the i-th link or a particular point, e.g., a measurement point in 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 digraph model, a route in the road network is defined as a path including finite sequence of arcs connecting a sequence of vertices which are all distinct from one another.

B. Spatio-temporal Correlation Quantification

Consider two detector stations A and B where A is located downstream of B and the distance between A and B is S. Suppose the vehicles travel from B to A with a stable average speed v, then approximately time $t = S/v$ is needed for the vehicles to arrive at 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.

Based on the digraph of a road network, we use $P_{i,j}^l$ to denote the set of paths originated from vertex i with length l. Given the k-th path $P_{i,j}^l(k) \in P_{i,j}^l$, the time-varying lag upon $P_{i,j}^l(k)$ is $\tau_{P_{i,j}^l(k)}$, which is calculated in the following way:

$$\tau_{P_{i,j}^l(k)} = \sum_{(V_{n_s}, V_{n_{s+1}}) \in P_{i,j}^l(k)} \tau_{n_s, n_{s+1}}. \tag{1}$$

To obtain the time-varying lags upon all the paths between two measurements, we can use a search algorithm like breadth-first search (BFS).

Obviously, it implies high computational complexity for considering the spatio-temporal correlation between the traffic in any roads and any paths in a spatio-temporal model. To solve this problem, we quantify the spatio-temporal correlation based on the trip distribution in the road network. Specifically, we capture the trip distribution by introducing the concept of **turning rate** between two adjacent vertices V_i and $V_j, i, j \in \mathcal{N}$, which is denoted by $\pi_{i,j}$ and calculated by

$$\pi_{i,j} = \frac{\sum y_{i,j}(t)}{\sum y_i(t - \tau_{i,j})}. \tag{2}$$

In equation (2), $y_{i,j}(t)$ is the traffic entering into V_j from V_i at t; $y_i(t - \tau_{i,j})$ is the total traffic in V_i at $t - \tau_{i,j}$ where $\tau_{i,j}$ is the time-varying lag between V_i and V_j; $\sum$ represents the sum of traffic in a given time period. Clearly, $\pi_{i,j}$ is 0 in the case that there is no arc between V_i and V_j, or there is no traffic entering into V_j from V_i. Otherwise, $\pi_{i,j}$ is non-zero. Based on the turning rate, we calculate the probability of traffic entering V_j from V_i through $P_{i,j}^l(k)$ by

$$\pi_{P_{i,j}^l(k)}(k) = \prod_{(V_{n_s}, V_{n_{s+1}}) \in P_{i,j}^l(k)} \pi_{n_s, n_{s+1}}. \tag{3}$$

Thus, we propose a metric named **traffic transition probability**, which is defined as the total traffic entering V_j from V_i through all the paths in $P_{i,j}^l$, is estimated by

$$\pi_{i,j}^l = \sum_{k=1}^{|P_{i,j}^l|} \pi_{P_{i,j}^l(k)}(k). \tag{4}$$

Based on the road network model, we use a $\mathcal{N} \times \mathcal{N}$ **turning rate matrix**, denoted by A_D, to represent the turning rate between V_i and $V_j, \forall i, j \in \mathcal{N}$. With the turning rate matrix A_D, we define A_D^l as

$$A_D^l = \underbrace{A_D \times A_D \times ... \times A_D}_{l}. \tag{5}$$

It is easy to find that the $(i,j)^{th}$ entry of A_D^l is the traffic transportation probability between V_i and V_j along all the paths with the length l. Clearly, the spatio-temporal correlation between two links is weak if the traffic transition probability between these two links is small (0 means no correlation). Further, we introduce a **threshold** ε that is set manually. We say there is a weak correlation between two adjacent vertices V_i and V_j if the $(i,j)^{th}$ entry in A_D^l is non-zero and less than ε. For ease of description, we denote A_D^l after filtering by ε as $A_D^{l/\varepsilon}$. With the aid of $A_D^{l/\varepsilon}$, we can obtain the maximal spatial order between any pair of vertices in D, denoted as $\lambda_{i,j}, i, j \in \mathcal{N}$.

C. Spatio-temporal Model

Based on the above methods, we present a spatio-temporal model based on STARIMA(p_λ, d, q_m) which is defined as follows:

$$\begin{aligned}(I - &\sum_{k=1}^{p}\sum_{l=0}^{\lambda_k} \phi_{kl} W_l L^k)(1-L)^d Y(t) \\ =(I - &\sum_{k=1}^{q}\sum_{l=0}^{m_k} \theta_{kl} W_l L^k)\epsilon_t.\end{aligned} \tag{6}$$

In this paper, $Y(t) = \{y_1(t), y_2(t), ..., y_N(t)\}$ is a $\mathcal{N} \times 1$ vector including the traffic flow from $\mathcal{N}$ links at$t$, L is the lag operator: $y_i(t-1) = Ly_i(t), i \in \mathcal{N}$, $\phi_{kl}$and θ_{kl} are coefficients, W_l is the spatial weight matrix, and ϵ_t is white noise. There are three steps to set up a STARIMA model including 1) Model Identification, 2) Parameter Estimation and 3) Diagnostic Checking [12].

We provide the derivation of our STARIMA based model with consideration of time-varying lags as follows. Given a vertex $V_j \in V$, the relationship between the traffic of V_j and the traffic from its upstream links $V_i \in V$ can be represented as a function of time-varying lags along all the paths from V_i to V_j:

$$y_j(t) = \boldsymbol{\phi_{i,j}} \sum_{i \in \mathcal{N}} \sum_{P_k \in P_{i,j}^l} \pi_{i,j}^l(k) y_i(t - \tau_{i,j}^{P_k}(t)), \tag{7}$$

where $\boldsymbol{\phi_{i,j}}$ is the coefficient with the same definition in the original STARIMA model. With the lag operator L defined in the STARIMA model, we can rewrite $y_i(t - \tau_{i,j}^{P_k}(t))$ by $y_i(t - \tau_{i,j}^{P_k}(t)) = L^{\tau_{i,j}^{P_k}(t)} y_i(t)$. Given a $l \in \lambda_{i,j}$, we define two $\mathcal{N} \times \mathcal{N}$ matrices, respectively denoted as Π_l and Ξ_l, namely the traffic transition probability matrix and the time-varying lag matrix. For $\forall i, j \in \mathcal{N}$, $\Pi_l(i,j)$ is a $\mathcal{N}$ dimension vector in which each element is $\pi_{P_{i,j}^l(k)}(k), k \in |P_{i,j}^l|$, and $\Xi_l(i,j)$ is also a $\mathcal{N}$ dimension vector in which each element is $L^{\tau_{i,j}^{P_k}(t)}, P_k \in P_{i,j}^l$. In this way, formulation (7) can be rewritten as

$$y_j(t) = \phi_{i,j} \sum_{i \in \mathcal{N}} \Pi_l(i,j) \Xi_l(i,j) y_i(t). \tag{8}$$

We use $Y(t)$ to represent a $\mathcal{N} \times 1$ vector in which each element is the traffic flow data observed at each link at time t, on the basis of formulation (8), we can get the following equation

$$Y(t) = \boldsymbol{\phi} \Pi_l^T \Xi_l Y(t), \tag{9}$$

where $\boldsymbol{\phi}$ is the vector of $\phi_{i,j}, i, j \in \mathcal{N}$; Π_l^T is the transportation of Π_l. We use $\lambda = max\{\lambda_{i,j}, i, j \in \mathcal{N}\}$ as the spatial order parameter in the STARIMA model, Thus, the format of our STARIMA based model is presented as

$$
\begin{aligned}
&(I - \sum_{l=0}^{\lambda} \phi_l \Pi_l^T \Xi_l L^d) Y(t) = \\
&(I - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \theta_{kl} W_l L^d) \epsilon_t.
\end{aligned} \tag{10}
$$

The main difference between the parameters estimation of our STARIMA model and original one lies in that additional time-varying lags should be first considered in the phase of model identification. In each time period, e.g, peak hour or off-peak hour, the time-varying lags and λ can be estimated using Algorithm 1.

Algorithm 1 The estimation of λ and time-varying lags between measurements (links) in road network

1 **Input:** Ω, A_D, $\mathcal{N}$, ε, $\lambda = +\infty$
2 **for** $\forall i, j \in \mathcal{N}$
3 **if** $A_D(i,j) \neq 0$
4 $\quad \tau_{i,j} \leftarrow \frac{S_{i,j}}{v_{i,j} t_{lag}}$
5 **endif**
6 **endfor**
7 **while** $l \leq \lambda$
8 $\quad A_D^{l/\varepsilon} \leftarrow A_D^l$
9 $\quad$ **for** $\forall i, j \in \mathcal{N}$
10 $\quad\quad$ **if** $A_D^l(i,j) \neq 0$ and $A_D^l(i,j) \geq \varepsilon$
11 $\quad\quad\quad P_{i,j}^l$: using BFS algorithm
12 $\quad\quad$ **for** $\forall P_{i,j}^l(k) \in P_{i,j}^l$
13 $\quad\quad\quad \tau_{P_{i,j}^l(k)} = \sum_{(V_{n_s}, V_{n_{s+1}}) \in P_{i,j}^l(k)} \tau_{n_s, n_{s+1}}$
14 $\quad\quad$ **endfor**
15 $\quad\quad$ **elseif** $A_D^l(i,j) \neq 0$ and $A_D^l(i,j) < \varepsilon$
16 $\quad\quad\quad A_D^{l/\varepsilon}(i,j) \leftarrow 0$
17 $\quad\quad$ **endif**
18 $\quad$ **endfor**
19 $\quad$ **if** all non-zero $A_D^l(i,j) < \varepsilon$
20 $\quad\quad \lambda \leftarrow l$
21 $\quad$ **else**
22 $\quad\quad l \leftarrow l + 1$
23 $\quad$ **endif**
24 **endwhile**

In Algorithm 1, the inputs include $\{\Omega, A_D, \mathcal{N}, \varepsilon, \lambda\}$. More precisely, Ω is time period clusters which is obtained using some classification algorithms such as ISODATA [3]. As the classification algorithm is not the key point in this paper, thus, we mainly divide different time periods in one day into off-peak or on-peak hour. In other words, $\Omega = \{\Omega_1 : Peak, \Omega_2 : Off-peak\}$. A_D can be estimated based on the traffic data in two adjacent links during a time period. $\mathcal{N}$ is the number of links in the road network and ε is an empirical value set manually. To find the maximal spatial order between measurements, the initial value of λ is set by $\lambda = +\infty$. The time-varying lag between each pair of adjacent links is calculated in line III-C. It is to be observed that $v_{i,j}$ should be the average speed computed as the length of the segment between two detectors located on two links at time t divided by the total time required to travel the segment, in other words, could be regarded as space mean speed (SMS). However, the speed collected by detectors, mostly refer to the time mean speed (TMS) [13], [6]. To infer the SMS from TMS, we use a more common model which better reflects the relationship between TMS and SMS in [14]. From line 9 to 18, we search all the paths $P_{i,j}^l$ according the the traffic correlation between V_i and V_j. We get $A_D^{l/\varepsilon}$ by comparing $A_D^l(i,j)$ and ε in line 15 and 16. Line 19 to 23 are used for determining λ. The computational complexity of Algorithm 1 mainly consists of two parts. One is the implementation of BFS algorithm. The other one is the determination of time-varying lags. However, as the topology of a road network is fixed, thus, the BFS algorithm is only proceeded once.

IV. Experimental Validation

A. Experimental Setup

In this section, we use the measurement data collected from I-205 NB Portland-area freeway to establish the validity and accuracy of the proposed traffic flow predictor [1] The freeway in Fig. 1 covers 8.23km including a major road and on- and a series of on and off-ramps. As there are no detector stations at off-ramps, we roughly infer the amount of traffic leaving the freeway through an off-ramp from the traffic collected at the existing detector stations. For instance, consider the off-ramp between station 1046 and 1047, we can infer the traffic y_{off} by $y_{off} = y_{1046} - (y_{1047} - y_{5047})$. 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 ($t_{lag} = 20s$). 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 [15], [16]. In addition, all the experiments are implemented in 64-bit Windows operating system with 16G memory.

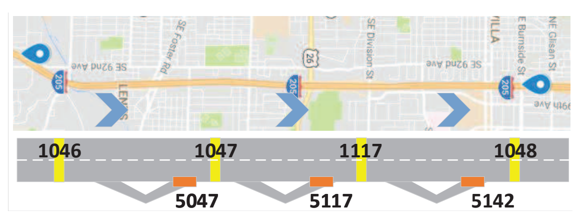

Fig. 1. The map and topology of I-205 NB freeway

We compare our proposed method (denoted as STARIMA*) with other two approaches, respectively the ARIMA(p,d,q) based model (denoted as ARIMA*) in which the parameters and coefficients would be re-evaluated in different time periods, and the BPNN model. ARIMA* is a linear predictive models, while BPNN is a non-linear predictive model. We use a $4 \times 20 \times 1$ BPNN model

[1]The data can be downloaded from: http://portal.its.pdx.edu.

including a hidden layer and an output layer to predict the traffic flow at each measurement point. The configuration of parameters of BPNN is based on the counterparts used in the experiment in [17]. Specifically, 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 activation function. Besides, we set the momentum coefficient to be 0.7, and the learning rate to be 0.3. An 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. 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, the performance measure MSE can be expressed as $\frac{1}{N}\sum_{n=1}^{N}(\hat{y}_n - y_n)^2$, and MAPE is $\frac{1}{N}\sum_{n=1}^{N}|\frac{\hat{y}_n - y_n}{y_n}|$.

B. Experimental Results

Based on the traffic flow collected at the stations, we intuitively divided a day into five time periods (Ω) and estimate the time-varying lag between two adjacent measurements which are shown in in Table I. It verifies that the time-varying lag has a close relation with the distance of two measurements, as well as different travel speeds during different time periods of the day. Then, we calculate the time-varying lags between any two measurements Algorithm 1. For instance, the time-varying lag between station 1046 and 1048 is 11 $(5+3+3)$ within 6:00am-9:00am, while 8 $(4+2+2)$ within 9:00am-16:00pm.

TABLE I

THE TIME VARYING LAG BETWEEN MEASUREMENTS IN THE MAJOR ROAD DURING DIFFERENT TIME PERIODS OF THE DAY

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

The non-zero elements in the turning rate matrix A_D are presented in Table II. To save space, the values of the turning rates estimated by both methods are rounded off to the two decimal places. With A_D, we can calculate traffic transition probability between two reachable measurements. For instance, the traffic transition probability between detector stations 1046 and 1048 is 0.43 $(0.78 \times 0.69 \times 0.8)$.

We then predict the traffic flow of the day at 4 stations with $\varepsilon = 0$, that is, we consider all the links which are spatio-temporally correlated with the link of interest. From Table III, we can see that the best predicted results are obtained by our proposed model. Comparing with the forecasting results obtained by STARIMA* and ARIMA*, it further validates that the accuracy of prediction is improved if we consider both spatial and temporal correlation between traffic data. In the worst case there is $\sim 16\%$ (at station 1117) gap between the MAPE of our method and ARIMA*, as well as BPNN.

TABLE II

THE ESTIMATED TURNING RATES BETWEEN ADJACENT LINKS

From	1046		1047		1117	
To	1047	Off	1117	Off	1048	Off
0:00am-6:00am	0.78	0.22	0.69	0.31	0.80	0.20
6:00am-9:00am	0.79	0.21	0.54	0.46	0.79	0.21
9:00am-16:00pm	0.90	0.10	0.55	0.45	0.84	0.16
16:00pm-18:00pm	0.76	0.24	0.47	0.53	0.77	0.23
18:00pm-24:00pm	0.88	0.12	0.66	0.34	0.85	0.15

TABLE III

THE MAPE/MSE OF ONE-DAY TRAFFIC FLOW PREDICTION USING STARIMA* ($\varepsilon = 0$), ARIMA* AND BPNN

St.	STARIMA*	ARIMA*	BPNN
1046	19.29%/184.91	21.08%/152.90	38.43%/394.25
1047	12.57%/103.54	19.98%/124.04	44.26%/405.67
1117	35.95%/413.51	51.47%/687.21	51.28%/481.24
1048	15.72%/116.64	18.81%/142.97	35.11%/375.22

TABLE IV

THE MAPE/MSE OF ONE-DAY TRAFFIC FLOW PREDICTION USING STARIMA* WITH DIFFERENT CONFIGURATIONS OF ε

St.	$\varepsilon = 0$	$\varepsilon = 0.3$	$\varepsilon = 0.4$	$\varepsilon = 0.6$
1046	19.29%/184.91	19.29%/184.91	19.29%/184.91	19.29%/184.96
1047	12.57%/103.54	12.57%/103.54	12.57%/103.54	15.78%/123.41
1117	35.95%/413.51	38.70%/481.35	40.66%/517.12	41.22%/578.62
1048	15.72%/116.64	15.38%/138.33	16.92%/136.43	18.52%/154.79

In order to make an intuitive grasp of the impact of different configurations of εs on the prediction performance by our proposed method, we proceed an experiment by setting $\varepsilon = 0.3, 0.4, 0.6$. The corresponding results are listed in Table IV. Comparing with Table III, we can see that the results at station 1046 and 1047 are the same as the ones when $\varepsilon = 0$ because the transportation probability between each of these three stations and the correlated roads are above 0.6. Therefore, the formulations of STARIMA* are the same, thereby the predicted results are the same. The case is different at other stations because the traffic transition probability will be affected by ε. For instance, the traffic transition probability between detector stations 1046 and 1148 is 0.43. Hence, station 1046 is out of the consideration when we predict the traffic at station 1148. On the other hand, the accuracy of the prediction at target location does not always decrease when few correlated neighbors of the target road are considered. For instance, the result at station 1048 at $\varepsilon = 0.3$ are better than that at $\varepsilon = 0$. This can be explained by the overfitting that may exist when we consider too many correlated neighbors of the target road. Moreover, the difference between MAPE of the results estimated at

$\varepsilon = 0.3$ and $\varepsilon = 0.4$ ($\varepsilon = 0.4$ and $\varepsilon = 0.6$) in different stations is less than 3%. However, the efficiency of the model building will be improved due to the fact that few number of parameters are to estimate.

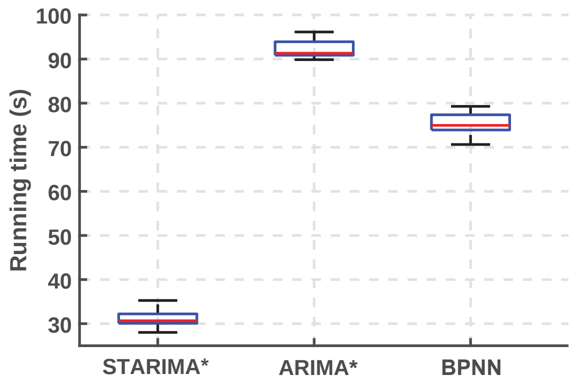

Fig. 2. The running time of STARIMA* ($\varepsilon = 0.6$), ARIMA* and BPNN

In Fig. 2, we present the running times including model building and forecasting for our method ($\varepsilon = 0.6$) and the counterparts. It is clear to see that less time is consumed for our proposed model. Based on the results in Table IV and Fig.2, it is sufficient to say that our proposed model is available for a more complicated road network.

V. Conclusions

In this paper, we develop a method for capturing and quantifying the intricate spatio-temporal correlations between traffic in different roads. More precisely, we capture the time-varying spatio-temporal correlation with the parameters, namely time-varying lags, which are related to the road network topology, time-varying speed, and time-varying trip distribution. Further, a metric named traffic transition probability calculated based on trip distribution, as well as a threshold ε are applied to select the most spatio-temporally correlated neighbors of a target road. With the aid of time-varying lag, the estimation of traffic transition probability and ε, we developed a STARIMA based model which removes the need to reconstruct different models for different time periods of the day for traffic prediction. Aside from a superior accuracy achieved based on our data set, the improved computational efficiency revealed through our work is that our proposed predictor potentially can be applied to a more complicated road network, e.g. the urban roads where there is a large number of road segments. However, the detector stations cannot cover every road in the urban network. Thus, it is part of our future work plan to extend our prediction technique with consideration of missing data.

Acknowledgment

I would especially like to thank Dr. Tom McBride, who has taught me more than I could ever give him credit for here.

References

[1] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transportation Research Part C: Emerging Technologies*, 2015.

[2] 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 Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1678–1687, 2016.

[3] P. Duan, G. Mao, C. Zhang, and S. Wang, "Starima-based traffic prediction with time-varying lags," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 1610–1615.

[4] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 1–17, 2017.

[5] G. Comert and A. Bezuglov, "An online change-point-based model for traffic parameter prediction," *IEEE Trans. Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1360–1369, 2013.

[6] J. Ahn, E. Ko, and E. Y. Kim, "Highway traffic flow prediction using support vector regression and bayesian classifier," in *2016 International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2016, pp. 239–244.

[7] A. Stathopoulos and M. G. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 2, pp. 121–135, 2003.

[8] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.

[9] 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," *Geographical Analysis*, vol. 46, no. 1, pp. 75–97, 2014.

[10] J. Ahn, E. Ko, and E. Y. Kim, "Highway traffic flow prediction using support vector regression and bayesian classifier," in *2016 International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2016, pp. 239–244.

[11] F. Kelly, "The mathematics of traffic in networks," *The Princeton companion to mathematics*, vol. 1, no. 1, pp. 862–870, 2008.

[12] 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.

[13] J. Han, J. W. Polak, J. Barria, and R. Krishnan, "On the estimation of space-mean-speed from inductive loop detector data," *Transportation planning and technology*, vol. 33, no. 1, pp. 91–104, 2010.

[14] A. D. May, *Traffic flow fundamentals*, 1990.

[15] S. Sun, C. Zhang, and G. Yu, "A bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124–132, 2006.

[16] L. Qu, L. Li, Y. Zhang, and J. Hu, "Ppca-based missing data imputation for traffic flow volume: a systematical approach," *IEEE Transactions on intelligent transportation systems*, vol. 10, no. 3, pp. 512–522, 2009.

[17] Y.-j. Kim, J.-s. Hong *et al.*, "Urban traffic flow prediction system using a multifactor pattern recognition model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2744–2755, 2015.