Predicting traffic speed in urban transportation subnetworks for multiple horizons

Justin Dauwels [1], Aamer Aslam [1], Muhammad Tayyab Asif [1], Xinyue Zhao [1], Nikola Mitrovic [1], Andrzej Cichocki [2] and Patrick Jaillet [3,4]

[1] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.
[2] Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama, Japan.
[3] Laboratory for Information and Decision Systems, MIT, Cambridge, MA.
[4] Center for Future Urban Mobility, Singapore-MIT Alliance for Research and Technology, Singapore.

***Abstract*—Traffic forecasting is increasingly taking on an important role in many intelligent transportation systems (ITS) applications. However, prediction is typically performed for individual road segments and prediction horizons. In this study, we focus on the problem of collective prediction for multiple road segments and prediction-horizons. To this end, we develop various matrix and tensor based models by applying partial least squares (PLS), higher order partial least squares (HO-PLS) and N-way partial least squares (N-PLS). These models can simultaneously forecast traffic conditions for multiple road segments and prediction-horizons. Moreover, they can also perform the task of feature selection efficiently. We analyze the performance of these models by performing multi-horizon prediction for an urban subnetwork in Singapore.**

I. Introduction

Advancements in sensor technologies have helped intelligent transportation systems (ITS) to improve the efficiency of existing transportation infrastructure [1]. Traffic prediction is important for many ITS applications such as route guidance, urban traffic control and management, and sustainable mobility [1]–[3]. Consequently, traffic prediction has garnered considerable attention in the field of transportation studies [4]–[13]. However, these studies, mostly focus on the development of separate prediction models for each road segment and prediction horizon. Traffic parameters such as speed, flow and travel time tend to be influenced by traffic conditions on the neighboring roads [14], [15]. These relationships are sometimes incorporated in forecasting models to improve the prediction performance [4]. However, it would be highly inefficient to develop individual prediction models in the presence of such spatial and temporal dependencies.

Some studies [14], [16], [17] have addressed the problem of unified models, albeit in a limited manner. The predictors proposed in [16] and [17] are capable of performing multi-horizon prediction using a single model. However, these techniques still require separate models for individual links. Lippi et. al. proposed Markov logic networks (MLN) to perform simultaneous prediction of traffic states on multiple roads and for multiple prediction horizons [14]. They defined the traffic state of each road as either congested or normal, based on a certain threshold value. To this end, they considered the threshold value of 50 mph to classify the traffic conditions on the roads [14]. The classification based approach is not suitable for many ITS applications such as route guidance, which requires continuous values of traffic parameters. Consequently, we consider traffic prediction as a regression problem rather than a classification problem.

We propose unified models for multi-horizon prediction of road networks. To this end, we develop a common low-dimensional representation for various temporal states (past, current and future) of the road network. Future traffic conditions tend to be dependent upon current and past states of the road network [4]. We employ various partial least squares (PLS) based methods to extract these temporal and spatial dependencies. PLS based methods have found applications in many areas such as chemometrics [18], bioinformatics [19], [20] and neuroscience [21]. In this study, we apply these techniques for multi-horizon speed prediction of a subnetwork. We develop unified models for network-wide traffic prediction by considering partial least squares (PLS) and its tensor variants, i.e., higher order partial least squares (HO-PLS) and N-way partial least squares (N-PLS). For analysis, we consider a generic road network which comprises of 266 road segments. This network contains expressway segments as well as arterial roads.

The rest of the paper is organized as follows. In section II, we discuss different unified prediction models based on PLS, N-PLS and HO-PLS. In section III, we explain the data set for this study. In section IV, we analyze the performance of different prediction models for a generic road network. Finally, in section V, we summarize our contributions and suggest topics for future work.

II. Traffic prediction models

In this section, we discuss different unified models for collective traffic prediction. We represent the road network, composed of p road segments, by a directed graph $G = (N,E)$. The road segment s_j is defined as an edge on the graph such that $\{s_j \in E\}_{j=1}^{p}$. We define the weight of each edge $z(s_j,t_i)$ as the average speed on that road segment s_j during the time interval $(t_i - \Delta t, t_i)$, where Δt = 5 minutes.

To perform prediction, traditional forecasting models try to find the relationship f_k between k-step ahead traffic state $z(s_j, t_i + k\Delta t)$ of the road and certain current/past traffic states of that road s_j and the neighboring roads. Let us represent the road s_j and the neighboring roads by a set $\Theta(s_j) = \{\theta_u\}_{u=1}^{q}$. The goal of the predictor would be to find the suitable relationship function for k-step ahead prediction:

$$z(s_j, t_i + k\Delta t) = f_k\big(z(\theta_1, t_i)...z(\theta_q, t_i - m_{\theta_q}\Delta t)\big) + \varepsilon, \quad (1)$$

where ε represents the residual error. We refer to this forecasting model as the individual model. The choice of relevant spatial $\Theta(s_j)$ and temporal m_{θ_q} features greatly influences the performance of a predictor [4]. Various methods have been proposed to extract relevant temporal features [17], [22], [23]. However, these methods are computationally intensive. Moreover, optimal selection of spatial features remains an open problem even for simple cases such as expressway networks [4].

In this study, we focus on an alternate approach to tackle the problem of traffic prediction. Instead of developing models for individual links and prediction horizons, we develop a unified model for the entire network and different horizons. For this purpose, we represent the current traffic conditions at time t_i on the road network by a row vector $\mathbf{a}_c(t_i) \in \mathbb{R}^{1\times p}$ where $\mathbf{a}_c(t_i) = [z(s_1, t_i) \ ... \ z(s_p, t_i)]$. We represent a certain past traffic state of the network by $\mathbf{a}_l(t_i) \in \mathbb{R}^{1\times p}$ where $\{\mathbf{a}_l(t_i) = [z(s_1, t_i - l\Delta t)...z(s_p, t_i - l\Delta t)]\}_{l=1}^{m_l}$ and m_l represents the maximum lag value. We combine these current and past states of the network to obtain the feature vector $\mathbf{x}(t_i)$ at time t_i, such that $\mathbf{x}(t_i) = [\mathbf{a}_c(t_i)\, \mathbf{a}_1(t_i)... \mathbf{a}_{m_l}(t_i)]$. Let us now consider the future traffic states of the road network. We represent the k-step ahead traffic state of the network by the row vector $\mathbf{g}_k(t_i) \in \mathbb{R}^{1\times p}$, where $\{\mathbf{g}_k(t_i) = [z(s_1, t_i + k\Delta t) \ ... \ z(s_p, t_i + k\Delta t)]\}_{k=1}^{m_k}$ and m_k represents the maximum prediction horizon. Network-wide traffic state for multiple future horizons $\{k\}_{k=1}^{m_k}$ will then be $\mathbf{y}(t_i) = [\mathbf{g}_1(t_i) \ ... \ \mathbf{g}_{m_k}(t_i)]$. The unified prediction model at time t_i can then be represented as:

$$\mathbf{y}(t_i) = \mathbf{x}(t_i)\mathbf{B} + \mathbf{e}, \quad (2)$$

where $\mathbf{B} \in \mathbb{R}^{h_1\times h_2}$ can be considered as the relationship matrix with dimensions $h_1 = p(1 + m_l)$ and $h_2 = pm_k$ and the row vector $\mathbf{e} \in \mathbb{R}^{1\times h_2}$ contains the errors. We refer to this model as the unified model. It may seem that both (1) and (2) represent similar prediction models. In fact, the individual model might seem more attractive as it incorporates a richer set of relationship functions. However, for individual model, we need to develop separate predictors for each link and prediction horizon. Moreover, we also need to extract appropriate set of spatial and temporal features for each predictor. Due to the inherent difficulty in extraction of optimal spatial feature, most individual models only utilize temporal features [4].

In the unified model, we learn the network-wide relation between the input features and future network states. Traffic states of different roads in a network tend to influence each other. These relationships can be helpful in the development of low-dimensional representations even for generic networks [24]–[26]. In the unified model, we extend the notion of low-dimensional models to include the state of the network during different time instances (past, current and future). We develop a unified low-dimensional model by linking the future and current/past traffic states of the road network with the help of relationship matrix $\mathbf{B}$. To this end, we apply PLS to obtain the these low-dimensional patterns and calculate the relationship matrix $\mathbf{B}$ in (2). For HO-PLS and N-PLS, we consider the tensor variant of (2) to perform traffic prediction. We utilize these models to perform network-wide prediction for multiple horizons. As a result, we do not have to extract suitable features for each link and prediction horizon. We now briefly explain different matrix and tensor based unified models for traffic prediction.

A. Partial least squares

To perform prediction at time t_i, we need to learn the relationship matrix $\mathbf{B}$, between the future traffic conditions $\mathbf{y}(t_i)$ and the input features for the network $\mathbf{x}(t_i)$. We learn this relationship by considering the historical data. To this end, we create a feature matrix $\mathbf{X} \in \mathbb{R}^{m_d\times h_1}$ and the corresponding prediction matrix $\mathbf{Y} \in \mathbb{R}^{m_d\times h_2}$, where m_d is the size of training data, $h_1 = p(1 + m_l)$ and $h_2 = pm_k$. We will perform prediction up to 30 minutes, hence we set $m_k = 6$. Following the procedure adopted by other related studies, we set the maximum lag $m_l = 6$ [16]. The feature matrix is obtained by stacking the feature vectors $\{\mathbf{x}(t_i)\}_{i=1}^{m_d}$ such that $\mathbf{X}^{\mathrm{T}} = [\mathbf{x}(t_1)^{\mathrm{T}} \ ... \ \mathbf{x}(t_{m_d})^{\mathrm{T}}]$. Similarly, the prediction matrix is obtained by stacking the corresponding traffic states for multiple future horizons, such that $\mathbf{Y}^{\mathrm{T}} = [\mathbf{y}(t_1)^{\mathrm{T}} \ ... \ \mathbf{y}(t_{m_d})^{\mathrm{T}}]$. These matrices can be represented by the following low-dimensional model [20]:

$$\begin{aligned} \mathbf{X} &= \mathbf{X}_S\mathbf{X}_L^{\mathrm{T}} + \mathbf{E}_x, \\ \mathbf{Y} &= \mathbf{Y}_S\mathbf{Y}_L^{\mathrm{T}} + \mathbf{E}_y, \end{aligned} \quad (3)$$

where the matrices $\mathbf{X}_S \in \mathbb{R}^{m_d\times r}$ and $\mathbf{X}_L \in \mathbb{R}^{r\times h_1}$ contain the score vectors and loading vectors of the feature matrix $\mathbf{X}$, respectively. Similarly, the matrices $\mathbf{Y}_S \in \mathbb{R}^{m_d\times r}$ and $\mathbf{Y}_L \in \mathbb{R}^{r\times h_2}$ contain the the score vectors and loading vectors of the prediction matrix $\mathbf{Y}$, respectively. The matrices $\mathbf{E}_x$ and $\mathbf{E}_y$ represent the error terms. The factor r represents the number of latent components used to obtain the low-dimensional representation, where $r < \min\big(p(1 + m_l), pm_k\big)$.

PLS decomposes the feature and prediction matrices with the constraint that the resultant components maximize the covariance between these two matrices i.e. $\mathbf{X}$ and $\mathbf{Y}$ [20], [27]. The commonly used methods to perform PLS regression are statistically inspired modification of the PLS (SIMPLS) and nonlinear iterative PLS (NIPALS). We chose SIMPLS as it is faster than NIPLS [28]. By applying SIMPLS, the relationship matrix will be $\mathbf{B}_{\mathrm{pls}} = \mathbf{R}\mathbf{Y}_L^{\mathrm{T}}$, where $\mathbf{R}$ is the so called weight matrix for $\mathbf{X}$ [28].

B. N-way partial least squares

The feature matrix $\mathbf{X}$ contains the information about the current state $\mathbf{a}_c(t_i)$ as well as certain past states $\{\mathbf{a}_l(t_i)\}_{l=1}^{m_l}$ of the network. Let us represent the feature matrix $\mathbf{X}$ in terms of

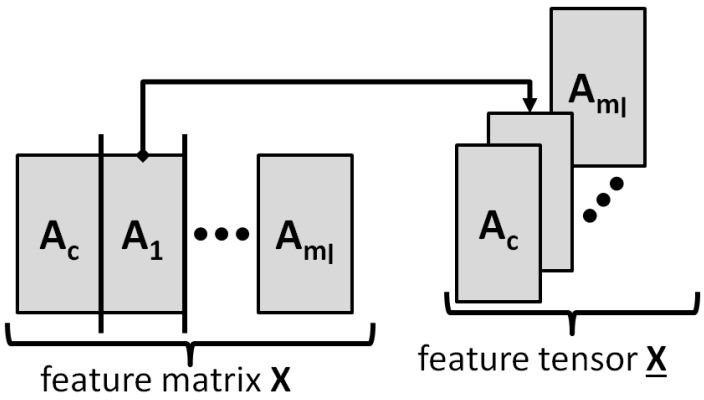


Fig. 1: Formation of 3-way feature tensor $\underline{\mathbf{X}}$ from the feature matrix $\mathbf{X}$.

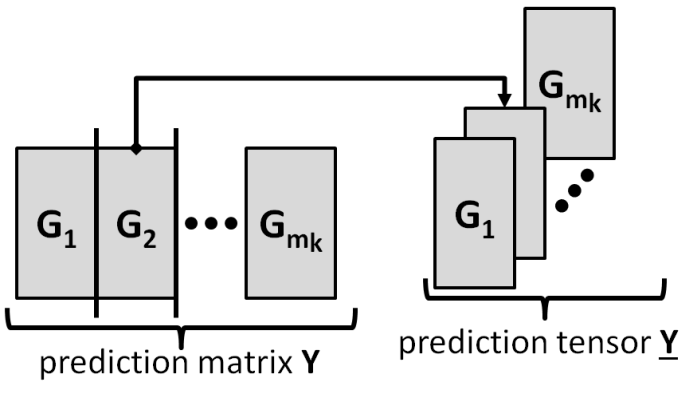


Fig. 2: Formation of 3-way prediction tensor $\underline{\mathbf{Y}}$ from the prediction matrix $\mathbf{Y}$.

these different temporal states such that $\mathbf{X} = [\mathbf{A}_c\ \mathbf{A}_1 ... \mathbf{A}_{m_l}]$, where the current matrix is $\mathbf{A}_c^{\mathrm{T}} = [\mathbf{a}_c(t_1)^{\mathrm{T}} ... \mathbf{a}_c(t_{m_d})^{\mathrm{T}}]$ and the different lag matrices are $\{\mathbf{A}_l^{\mathrm{T}} = [\mathbf{a}_l(t_1)^{\mathrm{T}}\ ...\ \mathbf{a}_l(t_{m_d})^{\mathrm{T}}]\}_{l=1}^{m_l}$. Naturally, certain degree of correlation exists between these temporal states. Therefore, a tensor representation might be more suitable to extract correlations between different temporal features. To this end, we modify the feature matrix $\mathbf{X} \in \mathbb{R}^{m_d \times h_1}$ to obtain a 3-way feature tensor $\underline{\mathbf{X}} \in \mathbb{R}^{m_d \times p \times h_t}$, where $h_t = 1 + m_l$ or $h_1 = ph_t$. The feature tensor $\underline{\mathbf{X}}$ is obtained by stacking together these temporal matrices $\{\mathbf{A}_c, \mathbf{A}_1, ...\mathbf{A}_{m_l}\}$ to form a 3-way array (see Fig. 1). Traditional PLS methods cannot handle multi-way arrays. Therefore, we consider N-way PLS (N-PLS) to perform prediction with tensor based network states. To this end, we apply a commonly used algorithm for N-PLS modeling, which is called trilinear-PLS2 (tri-PLS2) [21], [29], [30]. In tri-PLS2, the independent variables are represented in the form of 3-way array, where as the dependent variables are represented by a matrix [29]. In this study, we apply tri-PLS2 to obtain the common patterns between the feature tensor $\underline{\mathbf{X}}$ and the prediction matrix $\mathbf{Y}$. These patterns are then used to learn the regression coefficients [31]. During the rest of the paper, we will refer to this particular implementation (tri-PLS2) as N-PLS. The low-dimensional model for N-PLS can be written as:

$$\begin{aligned} \underline{\mathbf{X}} &= \sum_{i=1}^{r} \lambda_i\, \mathbf{v}_i^{(1)} \circ \mathbf{v}_i^{(2)} \circ \mathbf{v}_i^{(3)} + \underline{\mathbf{E}}, \\ \mathbf{Y} &= \mathbf{Y}_S \mathbf{Y}_L^{\mathrm{T}} + \mathbf{E}_y, \end{aligned} \tag{4}$$

where $\mathbf{v}_i^{(n)}$ are the column-i vectors of n-mode factor matrix $\mathbf{V}^{(n)}$, λ_i are scalers [27], [32] and the tensor $\underline{\mathbf{E}}$ contains the residuals. The matrices $\mathbf{Y}_S = [\mathbf{y}_{S1} ... \mathbf{y}_{Sr}]$ and $\mathbf{Y}_L = [\mathbf{y}_{L1} ... \mathbf{y}_{Lr}]$ contain the score vectors and loading vectors of $\mathbf{Y}$, respectively. The symbol $\circ$ denotes the vector outer product. Similar to PLS, N-PLS will decompose the feature tensor $\underline{\mathbf{X}}$ and the prediction matrix $\mathbf{Y}$ by finding a suitable common latent subspace [27]. To this end, the algorithm performs successive rank-one decompositions such that covariance between the scores vectors $\mathbf{v}_i^{(1)}$ and $\mathbf{y}_{Si}$ is maximized. For tensor decomposition, N-PLS uses canonical decomposition (CP)/ parallel factor analysis (PARAFAC). Tucker model is sometimes considered superior to PARAFAC model for tensor decomposition [27], [32]. In the next section, we discuss a variant of PLS which employs Tucker decomposition to find the common subspace between the independent and the response variables.

C. Higher-order partial least squares

Similar to N-PLS, in higher-order partial least squares (HO-PLS) we try to find the common latent space between the current/past network states and the future traffic conditions. However, instead of applying PARAFAC decomposition, we consider Tucker decomposition to obtain the low-dimensional models for tensors. We create the feature tensor $\underline{\mathbf{X}} \in \mathbb{R}^{m_d \times p \times h_t}$ by stacking together the temporal feature matrices $\{\mathbf{A}_c, \mathbf{A}_1 ... \mathbf{A}_{m_l}\}$ (see Fig. 1). The prediction matrix $\mathbf{Y}$ is composed of multiple future states of the road network. Therefore, we can represent the prediction matrix as $\mathbf{Y} = [\mathbf{G}_1 ... \mathbf{G}_{m_k}]$, where the matrices $\{\mathbf{G}_k^{\mathrm{T}} = [\mathbf{g}_k(t_1)^{\mathrm{T}} ... \mathbf{g}_k(t_{m_d})^{\mathrm{T}}]\}_{k=1}^{m_k}$ contain the k-step ahead states of the road network during different time instances $\{t_i\}_{i=1}^{m_d}$. We now stack these matrices $\{\mathbf{G}_i\}_{i=1}^{m_k}$ to obtain the feature tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{m_d \times p \times m_k}$ (see Fig. 2). In higher order PLS (HO-PLS), we represent the feature tensor $\underline{\mathbf{X}}$ as the sum of rank-$(1, L_2, L_3)$ tucker blocks and the prediction tensor $\underline{\mathbf{Y}}$ as the sum of rank-$(1, M_2, M_3)$ tucker blocks [27]. This can be expressed as:

$$\begin{aligned} \underline{\mathbf{X}} &= \sum_{i=1}^{r} \underline{\mathbf{C}}_i \times_1 \mathbf{t}_i \times_2 \mathbf{V}_i^{(2)} \times_1 \mathbf{V}_i^{(3)} + \underline{\mathbf{E}}_x, \\ \underline{\mathbf{Y}} &= \sum_{i=1}^{r} \underline{\mathbf{D}}_i \times_1 \mathbf{t}_i \times_2 \mathbf{W}_i^{(2)} \times_1 \mathbf{W}_i^{(3)} + \underline{\mathbf{E}}_y, \end{aligned} \tag{5}$$

where $\mathbf{t}_i \in \mathbb{R}^{m_d \times 1}$ is the i^{th} latent vector, $\{\mathbf{V}_i^{(2)}, \in \mathbb{R}^{p \times L_2}, \mathbf{V}_i^{(3)} \in \mathbb{R}^{h_t \times L_3}, \mathbf{W}_i^{(2)} \in \mathbb{R}^{p \times M_2}, \mathbf{W}_i^{(3)} \in \mathbb{R}^{m_k \times M_3}\}_{i=1}^{r}$ are the loading matrices for their respective modes and $\{\underline{\mathbf{C}}_i \in \mathbb{R}^{1 \times L_2 \times L_3}\}_{i=1}^{r}$, $\{\underline{\mathbf{D}}_i \in \mathbb{R}^{1 \times M_2 \times M_3}\}_{i=1}^{r}$ are the core tensors. The tensors $\underline{\mathbf{E}}_x$ and $\underline{\mathbf{E}}_y$ represent the residuals. In (5), $\times_n$ represents n-mode product between a tensor and a matrix [32].

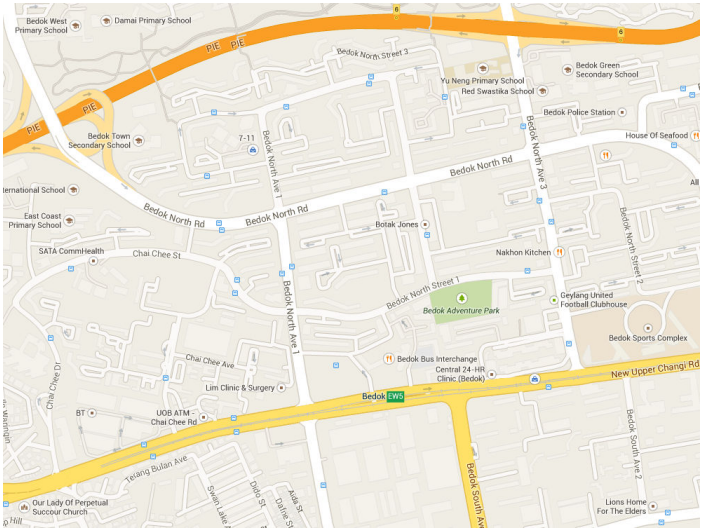


Fig. 3: Test network for performance analysis.

The block-tucker decomposition model employed in (5) is slightly different than the standard tucker decomposition. Nonetheless, the block-tucker decomposition can also be represented in the standard tucker decomposition form as follows:

$$\begin{aligned} \underline{\mathbf{X}} &= \underline{\overline{\mathbf{C}}} \times_1 \mathbf{T} \times_2 \overline{\mathbf{V}}^{(2)} \times_3 \overline{\mathbf{V}}^{(3)} + \underline{\mathbf{E}}_x, \\ \underline{\mathbf{Y}} &= \underline{\overline{\mathbf{D}}} \times_1 \mathbf{T} \times_2 \overline{\mathbf{W}}^{(2)} \times_3 \overline{\mathbf{W}}^{(3)} + \underline{\mathbf{E}}_y, \end{aligned} \tag{6}$$

where $\mathbf{T} = [\mathbf{t}_1 ... \mathbf{t}_r]$, and the n-mode loading matrices are $\{\overline{\mathbf{V}}^{(n)} = [\mathbf{V}_1^{(n)} ... \mathbf{V}_r^{(n)}]\}_{n\in\{2,3\}}$ and $\{\overline{\mathbf{W}}^{(n)} = [\mathbf{W}_1^{(n)} ... \mathbf{W}_r^{(n)}]\}_{n\in\{2,3\}}$. The core tensors $\underline{\overline{\mathbf{C}}} \in \mathbb{R}^{r\times rL_2\times rL_3}$ and $\underline{\overline{\mathbf{D}}} \in \mathbb{R}^{r\times rM_2\times rM_3}$ have block diagonal structure such that $\underline{\overline{\mathbf{C}}} = \text{blockdiag}(\underline{\mathbf{C}}_1, ... \underline{\mathbf{C}}_r)$ and $\underline{\overline{\mathbf{D}}} = \text{blockdiag}(\underline{\mathbf{D}}_1, ... \underline{\mathbf{D}}_r)$ [27]. Similar to N-PLS and PLS, we will train the HO-PLS model using historical data. We then employ the trained HO-PLS model to perform network wide prediction for multiple future horizons.

In this section, we have discussed various unified models to perform network wide prediction for multiple horizons. For comparison, we will consider support vector regression (SVR), which is commonly used for traffic forecasting [5], [8], [10], [11], [13], [33], [34]. We will train individual SVR models for each link and prediction horizon. To avoid the computation cost associated with spatial feature selection, we will only consider temporal features for SVR [4]. In the next section, we explain the data set used for performance analysis.

III. Data Set

In this section, we explain the data set used in this study. The test network consists of $p = 266$ road segments from the Bedok area in southeastern part of Singapore. The network contains sections of pan island expressway (PIE) as well as arterial roads in the vicinity of Bedok mass rapid transit (MRT) station (see Fig. 3). The speed data was obtained courtesy of Land Transport Authority (LTA) of Singapore. For analysis, we use speed data from the months of August and September, 2011. The data represents the average speed on each road segment with the averaging interval of 5 minutes. We perform training on the speed data of August and analyze the performance of the proposed algorithms by using the data from September. We calculate percentage root mean square distortion (PRD) to compare the performance of different prediction methods. For matrices, we define PRD as:

TABLE I: Number of latent factors (r) used for each prediction algorithm.

Algorithm	PLS	N-PLS	HO-PLS
Latent Factors (r)	50	75	63

TABLE II: Time taken (in sec) to perform multi-horizon prediction at one time instance. The network consists of 266 road segments and prediction was performed for six horizons (5 min, ... 30 min). The time for SVR was computed by running the SVR predictors in series.

Algorithm	PLS	N-PLS	HO-PLS	SVR
Time (sec)	0.1	0.4	0.1	7.9

$$\text{PRD}(\%) = \frac{\| \mathbf{Y} - \hat{\mathbf{Y}} \|_{\text{F}}}{\| \mathbf{Y} \|_{\text{F}}} \times 100\%, \tag{7}$$

where $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ are the actual and predicted matrices, respectively. The Frobenius norm of a matrix $\mathbf{Y} \in \mathbb{R}^{I_1\times I_2}$ is defined as:

$$\| \mathbf{Y} \|_{\text{F}} = \sqrt{\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2} y_{i_1 i_2}^2}. \tag{8}$$

For tensors, we define PRD as:

$$\text{PRD}(\%) = \frac{\| \underline{\mathbf{Y}} - \underline{\hat{\mathbf{Y}}} \|_{\text{F}}}{\| \underline{\mathbf{Y}} \|_{\text{F}}} \times 100\%, \tag{9}$$

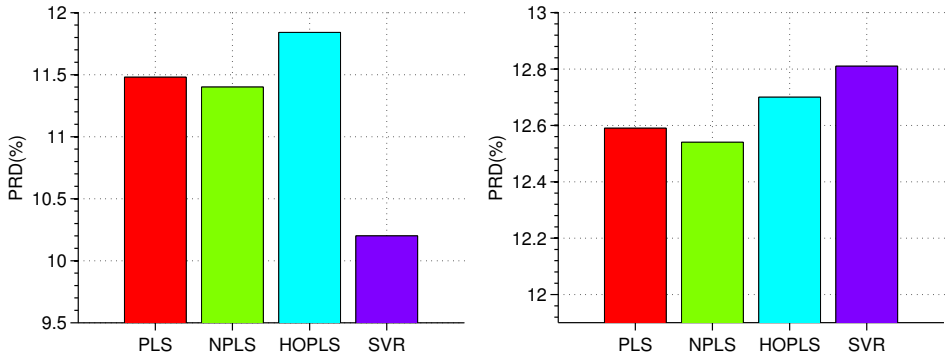
where $\underline{\mathbf{Y}}$ and $\underline{\hat{\mathbf{Y}}}$ are the actual and predicted tensors, respectively. The Frobenius norm of a tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1\times I_2\times I_3}$ is defined as:

$$\| \underline{\mathbf{Y}} \|_{\text{F}} = \sqrt{\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\sum_{i_3=1}^{I_3} y_{i_1 i_2 i_3}^2}. \tag{10}$$

In the next section, we compare the performance of the proposed prediction algorithms.

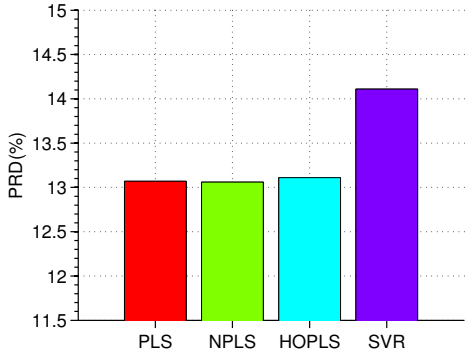
IV. Results and Discussion

In this section, we analyze the prediction performance of the proposed algorithms for a generic road network. For PLS, N-PLS and HO-PLS, we trained a unified prediction model for the entire network and various horizons. For comparison, we trained individual SVR predictors for each road segment and prediction horizon. As discussed earlier, we only used temporal features for SVR.

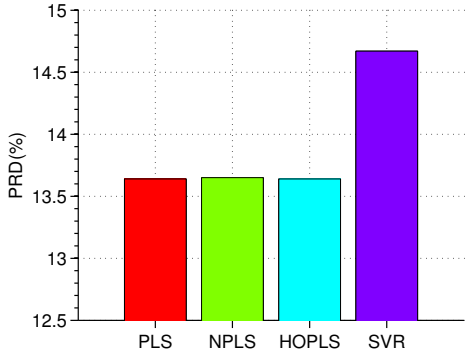


(a) Prediction Horizon: 5 minutes.

(b) Prediction Horizon: 10 minutes.



(c) Prediction Horizon: 15 minutes.



(d) Prediction Horizon: 30 minutes.

Fig. 4: Prediction performance of different algorithms in terms of PRD for various prediction horizons.

Fig. 4 summarizes the performance of the proposed algorithms for different prediction horizons. Table I shows the optimal number of latent factors used for each algorithm. PLS, N-PLS and HO-PLS perform slightly better than SVR for longer prediction horizons (see Fig. 4b, 4c and 4d). For the prediction horizon of 5 minutes, SVR has lower prediction error as compared to PLS based methods (see Fig. 4a). Amongst the PLS based methods, N-PLS achieves slightly higher accuracy as compared to PLS and HO-PLS.

Table II shows the time required for each method to perform multi-horizon prediction at one time instance for the entire network. The simulations were performed on a 24 core 2.64GHz CPU with 32 GB RAM. However, for standardized comparison, only one core was used. Moreover, no parallelization was done for SVR based predictors. As we trained individual SVR models for each link and prediction horizon, hence the processing time for SVR can be substantially improved by running multiple parallel threads. Interestingly, the processing times for PLS based methods are similar to those reported for models presented in [16], [35]. However, these studies developed customized models which are not readily available. Therefore, a direct comparison is not possible.

Let us now briefly discuss the problem of scalability of unified models for large road networks. The processing times seems to indicate that PLS, N-PLS and HO-PLS can also be used to perform real-time prediction for large-scale networks (see Table II). However, computational complexity of all three methods increases quadratically i.e. $O(p^2)$ with the network size p. Similarly the memory requirement also increase by $O(p^2)$. On the other hand, complexity of SVR (with temporal features) increases linearly i.e. $O(p)$ with the network size p. Hence, the unified models based on PLS, N-PLS and HO-PLS may not be feasible for large-scale networks.

In summary, we can state that the unified models achieve similar prediction accuracy as compared to the individual models (see Fig. 4). Moreover, for moderately sized networks, unified models based on PLS, N-PLS and HO-PLS can predict faster than traditional SVR based models (see Table II). However, the quadratic complexity in terms of both processing time and memory may limit their scalability for large networks. In such scenarios, network partition methods such as proposed in [35] might help to improve the scalability of PLS based prediction models.

V. Conclusions and Future Work

Traffic prediction is increasingly becoming an important part of many ITS applications such as route guidance and traffic management. In this study, we considered the problem of network-wide prediction for multiple horizons. To this end, we developed different forecasting models based on PLS, HO-PLS and N-PLS. We applied PLS and its tensor variants in the context of time series prediction. We employed

these techniques to obtain a common low-dimensional model between the current/past conditions of the network and the future network states. We then used the low-dimensional model to perform multi-horizon prediction for the entire subnetwork. We analyzed the performance of the proposed methods by performing speed prediction on a generic road network, which consisted of expressways as well as arterial roads.

In the future, we will incorporate the effect of abnormal traffic conditions arising due to accidents and road works in the unified models. Another aspect of this work involves the development of optimal network partition strategies to improve the scalability of these models for large road networks.

ACKNOWLEDGMENT

The research described in this project was funded in part by the Singapore National Research Foundation (NRF) through the Singapore MIT Alliance for Research and Technology (SMART) Center for Future Mobility (FM).

REFERENCES

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 4, pp. 1624–1639, Dec 2011.

[2] E. Schmitt and H. Jula, "Vehicle route guidance systems: Classification and comparison," in *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, Sept 2006, pp. 242–247.

[3] P. d'Orey and M. Ferreira, "Its for sustainable mobility: A survey on applications and impact assessment tools," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 2, pp. 477–493, April 2014.

[4] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where were going," *Transportation Research Part C: Emerging Technologies*, 2014.

[5] J. Wang and Q. Shi, "Short-term traffic speed forecasting hybrid model based on chaos–wavelet analysis-support vector machine theory," *Transportation Research Part C: Emerging Technologies*, vol. 27, pp. 219–232, 2013.

[6] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining kohonen maps with ARIMA time series models to forecast traffic flow," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 5, pp. 307–318, 1996.

[7] B. L. Smith, B. M. Williams, and R. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.

[8] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 14, no. 2, pp. 871–882, June 2013.

[9] C. Quek, M. Pasquier, and B. Lim, "POP-TRAFFIC: a novel fuzzy neural approach to road traffic analysis and prediction," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 7, no. 2, pp. 133–146, June 2006.

[10] C.-H. Wu, J.-M. Ho, and D. Lee, "Travel-time prediction with support vector regression," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 5, no. 4, pp. 276–281, Dec 2004.

[11] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, 2009.

[12] C. Van Hinsbergen, J. Van Lint, and H. Van Zuylen, "Bayesian committee of neural networks to predict travel times with confidence intervals," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 5, pp. 498–509, 2009.

[13] G. Gopi, J. Dauwels, M. Asif, S. Ashwin, N. Mitrovic, U. Rasheed, and P. Jaillet, "Bayesian support vector regression for traffic speed prediction with error bars," in *Intelligent Transportation Systems - (ITSC), 2013 16th International IEEE Conference on*, Oct 2013, pp. 136–141.

[14] M. Lippi, M. Bertini, and P. Frasconi, "Collective traffic forecasting," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2010, pp. 259–273.

[15] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing, "Discovering spatio-temporal causal interactions in traffic data streams," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1010–1018.

[16] 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.

[17] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211–234, 2005.

[18] K. Hasegawa, M. Arakawa, and K. Funatsu, "Rational choice of bioactive conformations through use of conformation analysis and 3-way partial least squares modeling," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 2, pp. 253–261, 2000.

[19] D. V. Nguyen and D. M. Rocke, "Tumor classification by partial least squares using microarray gene expression data," *Bioinformatics*, vol. 18, no. 1, pp. 39–50, 2002.

[20] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *Subspace, Latent Structure and Feature Selection*. Springer, 2006, pp. 34–51.

[21] E. Martınez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/FMRI analysis by multiway partial least squares," *NeuroImage*, vol. 22, no. 3, pp. 1023–1034, 2004.

[22] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *Journal of transportation engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[23] X. Jiang and H. Adeli, "Dynamic wavelet neural network model for traffic flow forecasting," *Journal of Transportation Engineering*, vol. 131, no. 10, pp. 771–779, 2005.

[24] Z. Li, Y. Zhu, H. Zhu, and M. Li, "Compressive sensing approach to urban traffic sensing," in *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*. IEEE, 2011, pp. 889–898.

[25] M. T. Asif, N. Mitrovic, L. Garg, J. Dauwels, and P. Jaillet, "Low-dimensional models for missing data imputation in road networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3527–3531.

[26] A. Hofleitner, R. Herring, A. Bayen, Y. Han, F. Moutarde, and A. de La Fortelle, "Large scale estimation of arterial traffic and structural analysis of traffic patterns using probe vehicles," in *91st Transportation Research Board Annual Meeting*. TRB, 2012.

[27] Q. Zhao, C. F. Caiafa, D. P. Mandic, Z. C. Chao, Y. Nagasaka, N. Fujii, L. Zhang, and A. Cichocki, "Higher order partial least squares (HOPLS): A generalized multilinear regression method," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 7, pp. 1660–1673, 2013.

[28] S. de Jong, "SIMPLS: an alternative approach to partial least squares regression," *Chemometrics and intelligent laboratory systems*, vol. 18, no. 3, pp. 251–263, 1993.

[29] R. Bro, "Multiway calidration. multilinear PLS," *Journal of Chemometrics*, vol. 10, pp. 47–61, 1996.

[30] M. M. Sena and R. J. Poppi, "N-way PLS applied to simultaneous spectrophotometric determination of acetylsalicylic acid, paracetamol and caffeine," *Journal of pharmaceutical and biomedical analysis*, vol. 34, no. 1, pp. 27–34, 2004.

[31] A. K. Smilde, "Comments on multilinear pls," *Journal of Chemometrics*, vol. 11, no. 5, pp. 367–377, 1997.

[32] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[33] M. T. Asif, J. Dauwels, C. Y. Goh, A. Oran, E. Fathi, M. Xu, M. Dhanya, N. Mitrovic, and P. Jaillet, "Spatiotemporal patterns in large-scale traffic speed prediction," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 2, pp. 794–804, April 2014.

[34] W.-C. Hong, "Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm," *Neurocomputing*, vol. 74, no. 12, pp. 2096–2107, 2011.

[35] Y. Wang, J. van Schuppen, and J. Vrancken, "Prediction of traffic flow at the boundary of a motorway network," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 1, pp. 214–227, Feb 2014.