This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Fundamental Limits of Missing Traffic Data Estimation in Urban Networks

Shangbo Wang, *Member, IEEE*, and Guoqiang Mao, *Fellow, IEEE*

***Abstract*— Traffic data estimation plays an important role because traffic data often suffers from missing data problems, caused by a variety of reasons, i.e., temporary deployment of sensors, sensor malfunction and communication failure. Existing research on missing data estimation has mostly focused on using data-driven or model-driven models to estimate the missing data, and there is a lack of study on the achievable estimation accuracy and the conditions to achieve accurate missing data estimation. In this paper, we investigate the fundamental limits of missing traffic data estimation accuracy in urban networks using the spatial-temporal random effects model. We derive the squared flow error bound (SFEB) for the cases of the Fisher matrix being a singular and non-singular matrix, respectively. We show that the sufficient and necessary condition of the existence of an unbiased estimator is that the number of missing points is less than or equal to the rank of the Fisher matrix. For the case that no unbiased estimator can be found, we derive an inequality for the SPEB and show that the SFEB is readily determined by the covariance matrix of the unknown (missing) parameter vector, flow correlation between the unknown and the available data, and the sensor locations. Furthermore, we develop an optimal spatial-temporal Kriging estimator which is efficient in both cases where the causal relationship among available data points exists or does not exist. Our theoretical findings can be used to develop a sensor location optimization strategy to minimize the SFEB.**

***Index Terms*— Cramer-Rao lower bound (CRLB), squared flow error bound (SFEB), fisher matrix, spatial-temporal kriging.**

I. Introduction

TRAFFIC data plays an important role in the development and implementation of ITS (Intelligent Transportation Systems). For example, both the ATIS (Advance Traveler Information System) which acquires, analyzes and presents information to assist travelers navigating from the source to the destination, and the ATMS (Advance Traffic Management System) which integrates various technology to improve the road traffic flow and road safety, rely heavily on highly accurate traffic data to provide users with up-to-date traffic information and guidance and for real-time traffic control [1]. Furthermore, accurate and relevant real-time traffic data can lead to a lot of improvements in many areas such as dynamic traffic control [2]–[7], improvement of incident management [8], [9] and traffic congestion detection and reduction [10].

Manuscript received May 5, 2018; revised October 4, 2018; accepted February 21, 2019. This work was supported by the Australia Research Council (ARC). The Associate Editor for this paper was S. E. Li. *(Corresponding author: Shangbo Wang.)*

The authors are with the School of Computing and Communications, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: shangbo.wang@uts.edu.au; guoqiang.mao@uts.edu.au).

Digital Object Identifier 10.1109/TITS.2019.2903524

Missing data problem has greatly hindered the collection and subsequent analysis, estimation and prediction of traffic flow data. Traffic data may become missing due to temporary deployment of sensors, detector malfunction or lossy communication systems. Specifically, due to high deployment costs, permanent traffic sensors may be installed on a subset of roads only [11] and some other roads may only be equipped with temporary sensors, which provide traffic data within limited time periods. Furthermore, failures, caused by detector malfunction and lossy communication systems, may also result in incomplete traffic data [12], [13]. It was reported in [14] that at hundreds of detection points within PeMS traffic flow database, more than 5% of data are missing. Qu et al. reported that about 10% of daily traffic flow is usually missing in Beijing [13]. In another study [15], almost a quarter of data from San Antonio, Texas, were found to be missing. The missing data has severe impact on many ITS applications, most of which rely on reliable, accurate and complete data [16]–[18]. For instance, traffic flow prediction relies on the complete historical data and the prediction performance will reduce sharply with incomplete data.

To tackle the missing data problem, a number of imputation methods have been proposed in the recent decade. Existing imputation methods can generally be classified into the following four categories: historical neighboring imputation methods, interpolation based methods, prediction based methods and statistical learning based methods [14].

The historical neighboring methods are naive methods which fills a missing data point with a known data point collected on the same site at the same time but from a neighboring day [19], [20]. The interpolation based methods estimate the missing data by arithmetic or weighted average of the available data, which is temporally, spatially, or both temporally and spatially close to the missing data. Some well known interpolation based methods include multiple imputation scheme [21], correlative kNN (k Nearest Neighbor) scheme [22], sectional kNN scheme [23] and LLS (Local Least Squares) scheme [24]. Prediction based methods directly apply the traffic prediction methods to fill in the missing data, including Autoregressive Integrated Moving Average (ARIMA) [25], [26], Seasonal ARIMA (SARIMA) [27], Space-Time ARIMA (ST-ARIMA) [28], [29] and Feed-Forward Neural Network (FFNN) [18], [30]. The main limitation of the prediction based methods is that data points measured after the missing data cannot be utilized to improve the imputation performance. Accurate imputation becomes difficult if there is no sufficient

1524-9050 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

measured data points before the missing points or a consecutive sequence of data points are lost. The most frequently used statistical learning methods are Probabilistic Principal Component Analysis (PPCA) [13], Kernel Probabilistic Principal Component Analysis (KPPCA) [14] and tensor completion techniques [31]–[34]. The PPCA and KPPCA can be considered as an orthogonal projection of data onto the principal subspace, in a way that maximizes the variance of the projected data [13]. Tensor completion techniques incorporate the measured traffic data into a tensor and then estimate the missing data using Tucker decomposition based imputation method (TDI).

To the authors' knowledge, existing research efforts mostly deal with developing certain data-driven or model-driven methods for missing data estimation. There is a lack of study on the achievable estimation accuracy and the conditions to achieve accurate estimation results. In this paper, we characterize estimation accuracy in terms of a performance measure called the Squared Flow Error Bound (SFEB) and focus on the study of the fundamental limits of missing data estimation accuracy by using Fisher matrix to derive SFEB. The Fisher matrix is an important tool for evaluating the accuracy of the parameter estimation technique. The inverse of the Fisher matrix yields the Cramer-Rao Lower Bound (CRLB) which provides asymptotically a lower bound for the covariance matrix of unbiased estimators [35]. We investigate the fundamental limits of missing data estimation accuracy under different scenarios: the Fisher matrix is a full-rank matrix, which is the sufficient condition to find an efficient estimator or a singular matrix, with which an efficient estimator can be found under certain conditions. We also investigate the impact of relationship between the number of missing points and rank of the Fisher matrix on the fundamental limits. We show the performance optimality of the spatial-temporal Kriging estimator proposed for missing data estimation.

Furthermore, almost all proposed historical neighboring imputation methods, interpolation based methods and prediction based methods only dealt with either single missing data imputation or the case that the number of missing data is not large. The imputing performances of the proposed methods greatly depend on the surrounding observed data of the missing points. Thus, their performances suffer when the missing ratio goes high. Although the proposed PPCA/KPPCA based imputation methods utilize all observed data to fit missing points and theoretically can cope with multiple detector cases, the aforementioned limitations remain.

In a presence of a large data set with a high missing ratio, some available data may have weak spatio-temporal correlation with the missing data. PPCA/KPPCA based approaches have dominant overfitting problem. The adopted interpolation based approaches, i.e., original *k*NN or correlative *k*NN, are not able to work well because a large portion of missing data leads to difficult selection of nearest neighbors. Intuitively, the missing data physically close to the observed data has a relatively high correlation and thus can be estimated with relatively high accuracy, then the imputing results can be considered as additional observed data in the next iteration for estimating missing data further away. Motivated by the intuition and the aforementioned shortcomings of existing imputation techniques, in this paper, we propose an iterative multiple-point spatial-temporal Kriging technique, which can greatly reduce the computational complexity when the size of the observed parameter vector becomes large and is proved to be optimal in estimation accuracy under certain conditions.

The following is a detailed summary of our contributions:

- Considering an arbitrary traffic network, we derive the Squared Flow Error Bound (SFEB) for the case of the Fisher matrix being a full-rank or singular matrix, respectively;
- We show the sufficient and necessary condition for the existence of an unbiased estimator is that the number of missing points is less than or equal to the rank of the Fisher matrix;
- In the case that no unbiased estimator can be found, we show an inequality for the SFEB; and
- the SFEB is readily determined by the covariance matrix of the unknown (missing) parameter vector, flow correlation between the unknown data and the available data, and the sensor locations.
- We further show that the spatial-temporal Kriging estimator is optimal in both cases where a causal relationship among available data points exists or does not exist;

The rest of the paper is outlined as follows: Section II reviews the related work. Section III proposes the system model and derives the SFEB for missing data estimation. Section IV proposes the optimal and iterative spatial-temporal Kriging estimator, and prove the optimality of the iterative spatial-temporal Kriging estimator. Section V validates our theoretical findings using real traffic data. Section VI concludes the paper.

II. Related Work

A number of missing data imputation methods have been investigated in the recent decade. Beretta et al. assessed the performance of the nearest neighbor algorithm, quantifying the effect imputation yields on the data structure and on inferential and predictive statistics [19]. Tak et al. proposed a sectional *k*-NN method, which imputes missing data based on road sections sharing the same traffic property [23]. Cai *et al.* [22] introduced the correlative *k*-NN model which was superior to the original *k*-NN model because it replaces the physical distance by both the physical distance and the correlation coefficient between the historical traffic data of the two roads. Tan *et al.* [31], explored the ability of tensor based method for multi-loop detector's missing data imputation, which completes the missing data by tensor decomposition. Qu *et al.* [13] proposed the PPCA based method which integrated MLE (Maximum Likelihood Estimation) into traditional PCA approach. Li *et al.* [14] compared PPCA method and KPPCA (Kernel PPCA) method, which assumes a nonlinear relationship between observed samples and latent variables.

The aforementioned literature only focuses on developing a certain method to improve estimation accuracy. To authors' knowledge, there is still no literature dealing with the fundamental limits for traffic data estimation and systematical

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

comparison of different methods. CRLB, as a well-known criterion, has been widely used in various applications for performance evaluation. For instance, Shen et al. derived the squared position error bound (SPEB) via equivalent Fisher information (EFI) for a general framework and cooperative wideband localization systems [36], [37]. Yu and Dutkiewicz [38] derived CRLB for mobile tracking in non-line-of-sight (NLoS) environment when measurements of distance, heading angle, and velocity are employed to estimate the mobile position.

III. System Model and Error Bound on Missing Data Estimation

A. Road Map

In this section, we will firstly propose the urban traffic network model and the spatial-temporal random effects (STRE) model. Based on the proposed urban traffic network model, we will derive the SFEB of parameter vector estimation and show the relation among SFEB, the co-variance matrix of the real flow vector and the co-variance matrix of the observed flow vector. Furthermore, we will present the relation between rank of the Fisher Information Matrix (FIM) and the number of missing points, and give the sufficient and necessary condition of existence of unbiased estimator.

B. Network Model

Consider a two-dimensional traffic network with N_l roads, which are composed of N_m roads with detectors and N_u roads without detectors, $N_l = N_m + N_u$. At a time instant t, the traffic flow measured by the well-functioned detectors is represented as $q_{1,\mathrm{w},t}, q_{2,\mathrm{w},t} \cdots q_{M,\mathrm{w},t}$, while those flow over the rest $N_m - M$ roads with malfunctioned detectors and N_u unmeasured roads are represented as $q_{1,\mathrm{m,t}}, q_{2,\mathrm{m},t} \cdots q_{N_m-M,\mathrm{m},t}$ and $q_{1,\mathrm{u},t}, q_{2,\mathrm{u},t} \cdots q_{N_u,\mathrm{u},t}$, respectively. The letters "w", "m" and "u" denote well-functioned, malfunctioned and unmeasured roads, respectively. The data structure can be more clearly viewed by formulating the traffic flow on each road over consecutive time instants into a spatial-temporal matrix (see (3), shown at the top of the next page). Our task is to estimate the missing traffic data on the malfunctioned roads and unmeasured roads at each missing time instant. The set of missing data on all malfunctioned roads and unmeasured roads at all time instants is denoted by $\mathcal{N}_{\mathrm{D}} = \{1, 2, \cdots N_{\mathrm{D}}\} \triangleq \mathcal{N}_{\mathrm{M}} \cup \mathcal{N}_{\mathrm{U}}$, where N_{D} is the total number of missing points on the malfunctioned roads and unmeasured roads, $\mathcal{N}_{\mathrm{M}}$ and $\mathcal{N}_{\mathrm{U}}$ denote the set of missing points on the malfunctioned roads and the unmeasured roads, respectively. In this paper, we model the traffic process in urban networks with the spatial-temporal random effects (STRE) model by which the observations can be represented as a sum of a deterministic function, small-scale variation, fine-scale variation caused by the nugget effect in geostatistics, and measurement error [1], [39], that is

$$Z(\boldsymbol{s};t) = \mu_t(\mathbf{s}) + \boldsymbol{S_t}(\boldsymbol{s})' \boldsymbol{\eta_t} + \xi(\boldsymbol{s};t) + \varepsilon(\boldsymbol{s};t) \tag{1}$$

$$\boldsymbol{\eta_{t+1}} = \boldsymbol{H_{t+1}} \boldsymbol{\eta_t} + \boldsymbol{\zeta_{t+1}} \tag{2}$$

In a traffic network, $Z(\boldsymbol{s};t)$ represents the traffic flow on a finite number of road segments $\boldsymbol{s} = \{s_1 \cdots s_N\}$ at the time instant of t, $\mu_t(\boldsymbol{s})$ represents the average flow on $\boldsymbol{s}$ at the t-th time instant, $\boldsymbol{S_t}(\boldsymbol{s})' \boldsymbol{\eta_t}$ can be interpreted as the flow fluctuation caused by the flow variation from L selected neighboring roads, $\xi(\boldsymbol{s};t)$ represents the variation on $\boldsymbol{s}$ at time instant t caused by nugget effect and flow generation or dissipation within some specific sections, $\varepsilon(\boldsymbol{s};t)$ is the measurement error on $\boldsymbol{s}$ at time instant t, $\xi(\boldsymbol{s};t)$ and $\varepsilon(\boldsymbol{s};t)$ can be modeled as independent white Gaussian process with mean zero and variances σ_ξ^2 and σ_ε^2, respectively, $\boldsymbol{H_{t+1}}$ is a first-order auto-regressive matrix and $\boldsymbol{\zeta_{t+1}}$ is a innovation vector. In the next subsection, we will derive the *squared flow error bound* (SFEB) based on (1) and (2).

C. Error Bound on Missing Data Estimation

Based on (1), the traffic flow on each road segment can be expressed by a weighted sum of flows on its neighboring roads and independent random variations. Note that the neighboring roads not only include adjacent road segments, but also include the so-called l-th order neighbors, where l represents the spatial order of neighbors. For example, the first-order neighbors are those links that directly incident to the target road segments, while the second-order neighbors are indirectly connected to target road segments, via the first-order neighbors [40]. Yang it et al. [41] underlined that positive correlations exist among traffic collected at hundreds of sensors distributed on the entire road network sparsely, not just the neighborhood surrounding the target road segments.

Let us define an unknown parameter vector $\boldsymbol{\theta_1}$ that includes all missing datum on the $N_m - M$ malfunctioned roads, and an unknown parameter vector $\boldsymbol{\theta_2}$ that includes all unmeasured data points on N_u unmeasured roads, and a known parameter vector $\boldsymbol{\theta_3}$ that includes all measured data points on $N_m - M$ malfunctioned roads and M measured roads, i.e.,

$$\begin{aligned}
\boldsymbol{\theta}_1 &= \left[q_{1,\mathrm{m}} \quad q_{2,\mathrm{m}} \quad \cdots \quad q_{K_{\mathrm{m}},\mathrm{m}}\right]^{\mathrm{T}} \\
\boldsymbol{\theta_2} &= \left[q_{1,\mathrm{u},1} \quad \cdots \quad q_{N_{\mathrm{u}},\mathrm{u},1} \quad \cdots \quad q_{1,\mathrm{u},\mathrm{T}} \quad \cdots \quad q_{N_{\mathrm{u}},\mathrm{u},\mathrm{T}}\right]^{\mathrm{T}} \\
\boldsymbol{\theta_3} &= \left[q_{1,w,1} \quad \cdots \quad q_{M,w,1} \quad \cdots \quad q_{1,w,T} \quad \cdots \quad q_{M,w,T}\right. \\
&\qquad \left. q_{1,\mathrm{m,av}} \quad q_{2,\mathrm{m,av}} \quad \cdots \quad q_{K_{\mathrm{av}},\mathrm{m,av}}\right]^{\mathrm{T}}
\end{aligned}$$

where K_{m} and K_{av} are the number of missing data points caused by temporary failure of detectors and available data points on the malfunctioned roads, respectively. Let us define R_1 and R_2 as the data missing ratio and the failure rate of detectors, respectively, then the relation between R_1 and R_2, and K_{m}, K_{av} can be expressed by

$$\begin{aligned}
R_1 &= \frac{N_m + (N_m - M)(1 - R_2)}{N_m + N_u} \\
K_{\mathrm{m}} &= (N_m - M) R_2 T \\
K_{\mathrm{av}} &= (N_m - M)(1 - R_2) T
\end{aligned} \tag{4}$$

Then the following theorem can be obtained:

Theorem 1: An unbiased estimator of the parameter vector $\boldsymbol{\theta} = \left[\boldsymbol{\theta_1}^T \ \boldsymbol{\theta_2}^T \ \boldsymbol{\theta_3}^T\right]^T$ *with finite variance exists iff (if only if)* $\boldsymbol{A}^T \boldsymbol{R_\epsilon^{-1}} \boldsymbol{A} + \boldsymbol{R_\theta^\dagger}$ *is a full-rank matrix, where* $\boldsymbol{A}$ *is the scaling matrix and* $\boldsymbol{R_\epsilon}$ *is the error co-variance matrix given by (8) and (9), respectively,* $\boldsymbol{R_\theta}$ *is the co-variance matrix of the real flow vector,* $\boldsymbol{R_\theta^\dagger}$ *is the Moore-Penrose pseudo-inverse of* $\boldsymbol{R_\theta}$.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

$$\boldsymbol{G}=\left[\begin{array}{ccc|ccc|ccc} q_{1,\mathrm{w},1} & \cdots & q_{M,\mathrm{w},1} & q_{1,\mathrm{m},1} & \cdots & q_{N_m-M,\mathrm{m},1} & q_{1,\mathrm{u},1} & \cdots & q_{N_\mathrm{u},\mathrm{u},1} \\ q_{1,\mathrm{w},2} & \cdots & q_{M,\mathrm{w},2} & q_{1,\mathrm{m},2} & \cdots & q_{N_m-M,\mathrm{m},2} & q_{1,\mathrm{u},2} & \cdots & q_{N_\mathrm{u},\mathrm{u},2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ q_{1,\mathrm{w},t} & \cdots & q_{M,\mathrm{w},t} & q_{1,\mathrm{m},t} & \cdots & q_{N_m-M,\mathrm{m},t} & q_{1,\mathrm{u},t} & \cdots & q_{N_\mathrm{u},\mathrm{u},t} \\ q_{1,\mathrm{w},t+1} & \cdots & q_{M,\mathrm{w},t+1} & q_{1,\mathrm{m},t+1} & \cdots & q_{N_m-M,\mathrm{m},t+1} & q_{1,\mathrm{u},t+1} & \cdots & q_{N_\mathrm{u},\mathrm{u},t+1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ q_{1,\mathrm{w},T} & \cdots & q_{M,\mathrm{w},T} & q_{1,\mathrm{m},T} & \cdots & q_{N_m-M,\mathrm{m},T} & q_{1,\mathrm{u},T} & \cdots & q_{N_\mathrm{u},\mathrm{u},T} \end{array}\right] \tag{3}$$

Based on the given condition, the mean squared error (MSE) matrix of $\hat{\boldsymbol{\theta}}$ by any estimators satisfies the following inequality

$$\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^H\right\}\succeq \boldsymbol{R_\theta}-\boldsymbol{R_\theta}\boldsymbol{A}^T\boldsymbol{R}_{\boldsymbol{q_o}}^{-1}\boldsymbol{A}\boldsymbol{R_\theta}$$

where $\boldsymbol{M_1}\succeq\boldsymbol{M_2}$ means $\boldsymbol{M_1}-\boldsymbol{M_2}$ is positive semi-definite, $\boldsymbol{R}_{\boldsymbol{q_o}}$ is the co-variance matrix of the observed flow vector.

Proof: For simplicity, we abbreviate the n-th road segment and the t-th time instant as the (n,t)-th point. Recall that equations (1) and (2) give us the information that in a traffic network, traffic flow on each (n,t)-th point can be represented as the summation of a mean value, variation and error term. The variation is caused by the variation on its neighboring time instant and road segments, and evolves with time. By Kriging technique [42], the observed traffic flow on the (n,t)-th point can be expressed by

$$\begin{aligned} q_{n,t} &= q_{n,t,\mathrm{r}}+\xi(n;t)+\varepsilon(n;t) \\ &= \bar{q}_{n,t,\mathrm{r}}+\boldsymbol{r}_{\boldsymbol{n,t}}^{\mathrm{T}}\boldsymbol{G}_{\boldsymbol{n,t}}^{-1}\left(\boldsymbol{\theta}_{\boldsymbol{n,t},\mathrm{r}}-\bar{\boldsymbol{\theta}}_{\boldsymbol{n,t},\mathrm{r}}\right)+\zeta(n;t) \\ &\quad+\xi(n;t)+\varepsilon(n;t) \end{aligned} \tag{5}$$

where $q_{n,t,\mathrm{r}}$ and $\bar{q}_{n,t,\mathrm{r}}$ are the (n,t)-th instantaneous flow and average flow, respectively, $\boldsymbol{r_{n,t}}$ is a $[(N_m+N_u)T-1]\times 1$ column vector containing the empirical cross correlation between the (n,t)-th data point and the rest spatial-temporal data points $\left(n',t'\right)=(1\cdots N_m+N_u,1\cdots T),\left(n',t'\right)\neq(n,t)$, $\boldsymbol{G_{n,t}}$ is a $[(N_m+N_u)T-1]\times[(N_m+N_u)T-1]$ empirical co-variance matrix consisting of co-variance between each pair of spatial-temporal data points in the rest spatial-temporal data set, $\boldsymbol{\theta}_{\boldsymbol{n,t},\mathrm{r}}$ is a $[(N_m+N_u)T-1]\times 1$ column vector containing instantaneous flow at all unmeasured data points and measured data points, $\bar{\boldsymbol{\theta}}_{\boldsymbol{n,t},\mathrm{r}}$ is the $[(N_m+N_u)T-1]\times 1$ average real flow vector, $\zeta(n;t)$ is the flow variation introduced by measurement error on rest spatial-temporal data points and flow generation or dissipation within some specific sections, $\xi(n;t)$ is the flow variation caused by nugget effect and $\varepsilon(n;t)$ is the measurement error for the (n,t)-th observation. For simplicity, we define a new $[MT+K_{\mathrm{av}}-1]\times 1$ column vector $\boldsymbol{\theta}_{\boldsymbol{3}}^{(\boldsymbol{n,t})}$ obtained by removing the (n,t)-th element from $\boldsymbol{\theta_3}$. Then $\boldsymbol{r_{n,t}}$, $\boldsymbol{G_{n,t}}$ and $\boldsymbol{\theta_{n,t}}$ can be expressed by

$$\begin{aligned} \boldsymbol{\theta_{n,t}} &= \left[\boldsymbol{\theta}_{\boldsymbol{1}}^{\mathrm{T}}\quad\boldsymbol{\theta}_{\boldsymbol{2}}^{\mathrm{T}}\quad\boldsymbol{\theta}_{\boldsymbol{3}}^{(\boldsymbol{n,t})\mathrm{T}}\right]^{\mathrm{T}} \\ \boldsymbol{r_{n,t}} &= \mathbb{E}\left[\left(q_{n,t}-\bar{q}_{n,t}\right)\left(\boldsymbol{\theta_{n,t}}-\bar{\boldsymbol{\theta}}_{\boldsymbol{n,t}}\right)\right] \\ &= \mathbb{E}\left[\left(\tilde{q}_{n,t}+\varepsilon_{n,t}\right)\begin{pmatrix}\tilde{\boldsymbol{\theta}}_{\boldsymbol{1}}+\boldsymbol{\xi_1}+\boldsymbol{\varepsilon_1}\\ \tilde{\boldsymbol{\theta}}_{\boldsymbol{2}}+\boldsymbol{\xi_2}+\boldsymbol{\varepsilon_2}\\ \tilde{\boldsymbol{\theta}}_{\boldsymbol{3}}^{(\boldsymbol{n,t})}+\boldsymbol{\xi_3}+\boldsymbol{\varepsilon}_{\boldsymbol{3}}^{(\boldsymbol{n,t})}\end{pmatrix}\right] \\ &= \left[\boldsymbol{r}_{\boldsymbol{n,t,\theta_1}}^{\mathrm{T}}\quad\boldsymbol{r}_{\boldsymbol{n,t,\theta_2}}^{\mathrm{T}}\quad\boldsymbol{r}_{\boldsymbol{n,t,\theta_3^{(n,t)}}}^{\mathrm{T}}\right]^{\mathbf{T}} \\ \boldsymbol{G_{n,t}} &= \mathbb{E}\left[\left(\boldsymbol{\theta_{n,t}}-\bar{\boldsymbol{\theta}}_{\boldsymbol{n,t}}\right)\left(\boldsymbol{\theta_{n,t}}-\bar{\boldsymbol{\theta}}_{\boldsymbol{n,t}}\right)^{\mathrm{T}}\right] \\ &= \mathbb{E}\left[\begin{array}{c}\begin{pmatrix}\tilde{\boldsymbol{\theta}}_{\boldsymbol{1}}+\boldsymbol{\xi_1}+\boldsymbol{\varepsilon_1}\\ \tilde{\boldsymbol{\theta}}_{\boldsymbol{2}}+\boldsymbol{\xi_2}+\boldsymbol{\varepsilon_2}\\ \tilde{\boldsymbol{\theta}}_{\boldsymbol{3}}^{(\boldsymbol{n,t})}+\boldsymbol{\xi_3}+\boldsymbol{\varepsilon}_{\boldsymbol{3}}^{(\boldsymbol{n,t})}\end{pmatrix}\\ \times\begin{pmatrix}c\tilde{\boldsymbol{\theta}}_{\boldsymbol{1}}+\boldsymbol{\xi_1}+\boldsymbol{\varepsilon_1}\\ \tilde{\boldsymbol{\theta}}_{\boldsymbol{2}}+\boldsymbol{\xi_2}+\boldsymbol{\varepsilon_2}\\ \tilde{\boldsymbol{\theta}}_{\boldsymbol{3}}^{(\boldsymbol{n,t})}+\boldsymbol{\xi_3}+\boldsymbol{\varepsilon}_{\boldsymbol{3}}^{(\boldsymbol{n,t})}\end{pmatrix}\end{array}\right] \\ &= \boldsymbol{R}_{\boldsymbol{\theta_{n,t,r}}}+\left(\sigma_\xi^2+\sigma_\varepsilon^2\right)\boldsymbol{I} \end{aligned}$$

where $\boldsymbol{r_{n,t,\theta}}$ is the flow correlation between the (n,t)-th data point and $\boldsymbol{\theta}$, $\boldsymbol{R}_{\boldsymbol{\theta_k\theta_l}}$ is the co-variance matrix between $\boldsymbol{\theta_k}$ and $\boldsymbol{\theta_l}$, σ_ξ^2 is variance of nugget effect and flow generation or dissipation, σ_ε^2 is variance of the measurement error, $\tilde{q}_{n,t}$ is the flow variation at the (n,t)-th data point, $\tilde{\boldsymbol{\theta}}_{\boldsymbol{1}}$,$\tilde{\boldsymbol{\theta}}_{\boldsymbol{2}}$ and $\tilde{\boldsymbol{\theta}}_{\boldsymbol{3}}^{(\boldsymbol{n,t})}$ are the flow variation vectors for $\boldsymbol{\theta_1}$,$\boldsymbol{\theta_2}$ and $\boldsymbol{\theta}_{\boldsymbol{3}}^{(\boldsymbol{n,t})}$, respectively.

Note that the co-variance with the data vector on the malfunctioned road segments can be determined by historical data because of the temporary failure of detectors. The co-variance with the data vector on the unmeasured road segments can be determined by historical data measured by temporarily installed detectors on those road segments. In the case that some unmeasured road segments have no temporary detectors installed previously, $\boldsymbol{\theta_2}$ can be considered as a latent parameter vector, and $\boldsymbol{\theta_{n,t}}$ becomes $\left[\boldsymbol{\theta}_{\boldsymbol{1}}^{\mathrm{T}}\ \boldsymbol{\theta}_{\boldsymbol{3}}^{(\boldsymbol{n,t})\mathrm{T}}\right]^{\mathrm{T}}$.

The variance of the flow variation $\zeta(n;t)$ can be determined by

$$\begin{aligned} var\left(\zeta(n;t)\right) &= var\left(q_{n,t,\mathrm{r}}\right)-\boldsymbol{r}_{\boldsymbol{n,t}}^{\mathrm{T}}\boldsymbol{G}_{\boldsymbol{n,t}}^{-1}\boldsymbol{R}_{\boldsymbol{\theta_{n,t,r}}}\boldsymbol{G}_{\boldsymbol{n,t}}^{-\mathrm{T}}\boldsymbol{r_{n,t}} \\ &= var\left(q_{n,t,\mathrm{r}}\right)-\boldsymbol{r}_{\boldsymbol{n,t}}^{\mathrm{T}}\boldsymbol{U}_{\boldsymbol{\theta_{n,t,r}}}\boldsymbol{\Lambda}_{\boldsymbol{\theta_{n,t,r}}}\boldsymbol{U}_{\boldsymbol{\theta_{n,t,r}}}^{\mathrm{H}}\boldsymbol{r_{n,t}} \end{aligned}$$

where $\boldsymbol{R}_{\boldsymbol{\theta_{n,t,r}}}$ is the co-variance matrix of the real instantaneous flow vector $\boldsymbol{\theta_{n,t}}$, $\boldsymbol{\Lambda}_{\boldsymbol{\theta_{n,t,r}}}$contains non-zero Eigenvalues of $\frac{\lambda_k}{\lambda_k+\sigma_\xi^2}$, $k=1\cdots K_{n,t}$ in its diagonal, λ_k is the non-zero Eigenvalues of $\boldsymbol{R}_{\boldsymbol{\theta_{n,t,r}}}$, $\boldsymbol{U}_{\boldsymbol{\theta_{n,t,r}}}$ is the Eigenspace of $\boldsymbol{R}_{\boldsymbol{\theta_{n,t,r}}}$ dedicated to λ_k, $k=1\cdots K_{n,t}$. Recall that $\xi(n;t)$ and $\varepsilon(n;t)$ are independent white Gaussian process in space and time with mean zero and variance σ_ξ^2 and σ_ε^2, then (5) can be rewritten as

$$\begin{aligned} q_{n,t} &= \bar{q}_{n,t,\mathrm{r}}+\boldsymbol{r}_{\boldsymbol{n,t}}^{\mathrm{T}}\boldsymbol{G}_{\boldsymbol{n,t}}^{-1}\left(\boldsymbol{\theta}_{\boldsymbol{n,t},\mathrm{r}}-\bar{\boldsymbol{\theta}}_{\boldsymbol{n,t},\mathrm{r}}\right)+\epsilon_{n,t} \\ &= \boldsymbol{r}_{\boldsymbol{n,t}}^{\mathrm{T}}\boldsymbol{G}_{\boldsymbol{n,t}}^{-1}\boldsymbol{\theta}_{\boldsymbol{n,t},\mathrm{r}}+\bar{q}_{n,t,\mathrm{r}}-\boldsymbol{r}_{\boldsymbol{n,t}}^{\mathrm{T}}\boldsymbol{G}_{\boldsymbol{n,t}}^{-1}\bar{\boldsymbol{\theta}}_{\boldsymbol{n,t},\mathrm{r}}+\epsilon_{n,t} \end{aligned} \tag{6}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

where $\epsilon_{n,t}$ is an i.i.d zero-mean Gaussian process with variance of $var\left(q_{n,\mathrm{r}}\right)-\boldsymbol{r}_{n,t}^{\mathrm{T}}\boldsymbol{U}_{\theta_{n,t,r}}\boldsymbol{\Lambda}_{\theta_{n,t,r}}\boldsymbol{U}_{\theta_{n,t,r}}^{\mathrm{H}}\boldsymbol{r}_{n,t}+\sigma_{\xi}^{2}+\sigma_{\varepsilon}^{2}$. Formulating $MT+K_{\mathrm{av}}$ observations into a system of linear equations (SLE), which can be expressed by

$$\boldsymbol{q_o}=\boldsymbol{A}\boldsymbol{\theta}+\boldsymbol{u}+\boldsymbol{\epsilon} \tag{7}$$

where $\boldsymbol{q_o}$ is a $[MT+K_{\mathrm{av}}]\times 1$ observation flow vector, $\boldsymbol{\theta}$ is a $(N_m+N_u)\,T\times 1$ instantaneous real traffic flow vector containing flow on each road segment at T time instants, $\boldsymbol{A}$ is a $[MT+K_{\mathrm{av}}]\times(N_m+N_u)\,T$ scaling matrix obtained by $\boldsymbol{r}_{n,t}^{\mathrm{T}}\boldsymbol{G}_{n,t}^{-1}$, $n=1\cdots N_m+N_u$, $t=1\cdots T$, $\boldsymbol{u}$ is a $[MT+K_{\mathrm{av}}]\times 1$ bias vector and $\boldsymbol{\epsilon}$ is the measurement error vector, that is

$$\begin{aligned}
&\boldsymbol{A}=\left[\boldsymbol{A_i}\quad \boldsymbol{A_j}\quad \boldsymbol{A_{\ell 1}}\right]\\
&\boldsymbol{A_i}=\begin{bmatrix}\underbrace{\left[\boldsymbol{G}_{\boldsymbol{1}}^{-\mathrm{T}}\right]_{1}^{K_{\mathrm{m}}}\boldsymbol{r_1}\quad\cdots\quad\left[\boldsymbol{G}_{\boldsymbol{MT}}^{-\mathrm{T}}\right]_{1}^{K_{\mathrm{m}}}\boldsymbol{r_{MT}}}_{MT}\\ \underbrace{\begin{matrix}\left[\boldsymbol{G}_{\boldsymbol{MT+1}}^{-\mathrm{T}}\right]_{1}^{K_{\mathrm{m}}} & \cdots & \left[\boldsymbol{G}_{\boldsymbol{MT+K_{\mathrm{av}}}}^{-\mathrm{T}}\right]_{1}^{K_{\mathrm{m}}}\\ \times\boldsymbol{r_{MT+1}} & & \times\boldsymbol{r_{MT+K_{\mathrm{av}}}}\end{matrix}}_{K_{\mathrm{av}}}\end{bmatrix}^{\mathrm{T}}\\
&\boldsymbol{A_j}=\begin{bmatrix}\underbrace{\begin{matrix}\left[\boldsymbol{G}_{\boldsymbol{1}}^{-\mathrm{T}}\right]_{K_{\mathrm{m}}+1}^{K_{\mathrm{m}}+N_{\mathrm{u}}T} & \cdots & \left[\boldsymbol{G}_{\boldsymbol{MT}}^{-\mathrm{T}}\right]_{K_{\mathrm{m}}+1}^{K_{\mathrm{m}}+N_{\mathrm{u}}T}\\ \times\boldsymbol{r_1} & & \times\boldsymbol{r_{MT}}\end{matrix}}_{MT}\\ \underbrace{\begin{matrix}\left[\boldsymbol{G}_{\boldsymbol{MT+1}}^{-\mathrm{T}}\right]_{K_{\mathrm{m}}+1}^{K_{\mathrm{m}}+N_{\mathrm{u}}T} & \cdots & \left[\boldsymbol{G}_{\boldsymbol{MT+K_{\mathrm{av}}}}^{-\mathrm{T}}\right]_{K_{\mathrm{m}}+1}^{K_{\mathrm{m}}+N_{\mathrm{u}}T}\\ \times\boldsymbol{r_{MT+1}} & & \times\boldsymbol{r_{MT+K_{\mathrm{av}}}}\end{matrix}}_{K_{\mathrm{av}}}\end{bmatrix}^{\mathrm{T}}\\
&\left[\boldsymbol{A_{\ell 1}}\right]_{k=1}^{MT+K_{\mathrm{av}}}\\
&\quad=\begin{bmatrix}\left[\boldsymbol{G}_{\boldsymbol{k}}^{-\mathrm{T}}\right]_{K_{\mathrm{m}}+N_uT+1}^{K_{\mathrm{m}}+N_uT+k-1}\boldsymbol{r_k}\\ \boldsymbol{0}\\ \left[\boldsymbol{G}_{\boldsymbol{k}}^{-\mathrm{T}}\right]_{K_{\mathrm{m}}+N_uT+k}^{K_{\mathrm{m}}+N_uT+MT+K_{\mathrm{av}}-1}\boldsymbol{r_k}\end{bmatrix}^{\mathrm{T}}
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\boldsymbol{\theta}=\left[\boldsymbol{\theta}_{\boldsymbol{1},\mathrm{r}}^{\mathrm{T}}\quad\boldsymbol{\theta}_{\boldsymbol{2},\mathrm{r}}^{\mathrm{T}}\quad\boldsymbol{\theta}_{\boldsymbol{3},\mathrm{r}}^{\mathrm{T}}\right]^{\mathrm{T}},\quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A}+\boldsymbol{R}_{\boldsymbol{\theta}}^{\dagger}\boldsymbol{u}=\bar{\boldsymbol{q}}_{\mathrm{r}}-\boldsymbol{A}\bar{\boldsymbol{\theta}}\\
&\boldsymbol{\epsilon}=\left[\boldsymbol{\epsilon_1}\quad\cdots\quad\boldsymbol{\epsilon}_{MT+K_{\mathrm{av}}}\right]^{\mathrm{T}},\quad \bar{\boldsymbol{q}}_{\mathrm{r}}=\left[\bar{q}_{1,\mathrm{r}}\quad\cdots\quad\bar{q}_{MT+K_{\mathrm{av}},\mathrm{r}}\right]^{\mathrm{T}}
\end{aligned} \tag{9}$$

where $[\boldsymbol{G}]_{k1}^{k2}$ denotes the sub-matrix that extracts $[k1\sim k2]$-th rows of $\boldsymbol{G}$, $\bar{\boldsymbol{\theta}}$ is the real average traffic flow vector, $\boldsymbol{R_\epsilon}$ is $[MT+K_{\mathrm{av}}]\times[MT+K_{\mathrm{av}}]$ diagonal matrix with the n-th element on its diagonal $var\left(q_{n,t,\mathrm{r}}\right)-\boldsymbol{r}_{n,t}^{\mathrm{T}}\boldsymbol{U}_{\theta_{n,t,r}}\boldsymbol{\Lambda}_{\theta_{n,t,r}}\boldsymbol{U}_{\theta_{n,t,r}}^{\mathrm{H}}\boldsymbol{r}_{n,t}+\sigma_{\xi}^{2}+\sigma_{\varepsilon}^{2}$.

Let $\hat{\boldsymbol{\theta}}$ denote an estimate of the vector $\boldsymbol{\theta}$ based on observation $\boldsymbol{q_o}$. Then, the mean squared error (MSE) matrix of $\hat{\boldsymbol{\theta}}$ satisfies the information inequality [43]

$$\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{H}}\right\}\succeq \mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}^{-1} \tag{10}$$

$$\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\|\,\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\,\|^{2}\right\}\geq \mathrm{tr}\left(\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}^{-1}\right) \tag{11}$$

where $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$ is the Fisher Information Matrix (FIM) for the parameter vector $\boldsymbol{\theta}$, $\boldsymbol{A}\succeq\boldsymbol{B}$ denotes that the matrix $\boldsymbol{A}-\boldsymbol{B}$ is a positive semi-definite matrix, $\|\,\boldsymbol{.}\,\|$ is the Euclidean norm of its argument, $\mathrm{tr}\,\{\boldsymbol{.}\}$is the trace operation.

Note that the parameter vector $\boldsymbol{\theta}$ is a random parameter and $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$ is a Bayesian information matrix that does not depend on any particular $\boldsymbol{\theta}$, but requires an average over all possible $\boldsymbol{\theta}$.

Initially, we will give the FIM for deterministic parameter vector $\boldsymbol{\theta}$, then the PDF of parameters is incorporated to derive the FIM for random parameter vector. Let us define $f\left(\boldsymbol{q_o}|\boldsymbol{\theta}\right)$ as the likelihood ratio of the observation vector $\boldsymbol{q_o}$ conditioned on $\boldsymbol{\theta}$, then the FIM for a deterministic parameter vector $\boldsymbol{\theta}$ is given by [43]

$$\mathrm{J}_{\boldsymbol{\theta},\mathrm{d}}\triangleq\mathbb{E}_{\boldsymbol{q_o}}\left\{\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f\left(\boldsymbol{q_o}|\boldsymbol{\theta}\right)\right]\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f\left(\boldsymbol{q_o}|\boldsymbol{\theta}\right)\right]^{\mathrm{T}}\right\} \tag{12}$$

From (7) to (9), the likelihood function can be expressed by

$$\begin{aligned}
f\left(\boldsymbol{q_o}|\boldsymbol{\theta}\right)&=\frac{\exp\left\{-\frac{1}{2}\left(\boldsymbol{q_o}-\boldsymbol{A}\boldsymbol{\theta}-\boldsymbol{u}\right)^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{q_o}-\boldsymbol{A}\boldsymbol{\theta}-\boldsymbol{u}\right)\right\}}{(2\pi)^{\frac{(MT+K_{\mathrm{av}})}{2}}\det^{\frac{1}{2}}\left(\boldsymbol{R_\epsilon}\right)}\\
&\propto\exp\left\{-\frac{1}{2}\left(\boldsymbol{q}_{\boldsymbol{o}}^{\mathrm{T}}-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}-\boldsymbol{u}^{\mathrm{T}}\right)\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{q_o}-\boldsymbol{A}\boldsymbol{\theta}-\boldsymbol{u}\right)\right\}
\end{aligned} \tag{13}$$

Substituting (13) in (5), we have the FIM $\mathrm{J}_{\boldsymbol{\theta},\mathrm{d}}$ as

$$\mathrm{J}_{\boldsymbol{\theta},\mathrm{d}}=-\mathbb{E}_{\boldsymbol{q_o}}\left\{-\frac{\partial^{2}\left\{\begin{matrix}\left(\boldsymbol{q}_{\boldsymbol{o}}^{\mathrm{T}}-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}-\boldsymbol{u}^{\mathrm{T}}\right)\\ \times\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\left(\boldsymbol{q_o}-\boldsymbol{A}\boldsymbol{\theta}-\boldsymbol{u}\right)\end{matrix}\right\}}{2\partial\boldsymbol{\theta}^{2}}\right\} \tag{14}$$

$$\begin{aligned}
&=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A}\\
&=\begin{bmatrix}\boldsymbol{A}_{\boldsymbol{i}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A_i} & \boldsymbol{A}_{\boldsymbol{i}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A_j} & \boldsymbol{A}_{\boldsymbol{i}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A_{\ell 1}}\\ \boldsymbol{A}_{\boldsymbol{j}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A_i} & \boldsymbol{A}_{\boldsymbol{j}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A_j} & \boldsymbol{A}_{\boldsymbol{j}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A_{\ell 1}}\\ \boldsymbol{A}_{\boldsymbol{\ell 1}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A_i} & \boldsymbol{A}_{\boldsymbol{\ell 1}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A_j} & \boldsymbol{A}_{\boldsymbol{\ell 1}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A_{\ell 1}}\end{bmatrix}
\end{aligned} \tag{15}$$

Then, a lower bound of co-variance matrix of $\hat{\boldsymbol{\theta}}$ can be straightly obtained by (10) and (11) when $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A}$ is a full-rank matrix, that is $MT+K_{\mathrm{av}}=(N_m+N_u)\,T$. Unfortunately, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{A}$ is a singular matrix because $\boldsymbol{A}$ is a fat matrix. Thus, an unbiased estimation of the entire parameter vector is impossible to be found because some components in the parameter vector may have infinite variance [44]. In such scenario, (10) should be rewritten as

$$\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{H}}\right\}\succeq \mathrm{J}_{\boldsymbol{\theta},\mathrm{d}}^{\dagger}$$

where $\mathrm{J}_{\boldsymbol{\theta},\mathrm{d}}^{\dagger}$ is the Moore-Penrose pseudoinverse of $\mathrm{J}_{\boldsymbol{\theta},\mathrm{d}}$. There are three approaches coping with the case of the FIM being a singular matrix, that is, incorporating the PDF of parameter vector to covert the original deterministic parameter estimation problem into a Bayesian estimation problem; posing deterministic constrains on the parameter vector to reduce the parameter dimension; using a biased estimator [44]. In this paper, we focus on incorporating the a priori knowledge of the parameter vector.

Let us define $\boldsymbol{f}\left(\boldsymbol{\theta}\right)$ as the PDF of parameter vector $\boldsymbol{\theta}$, then the joint PDF of observation and parameter vector can be

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

expressed by

$$f(\boldsymbol{q_o}, \boldsymbol{\theta}) = f(\boldsymbol{q_o}|\boldsymbol{\theta}) \boldsymbol{f}(\boldsymbol{\theta})$$

where $f(\boldsymbol{q_o}|\boldsymbol{\theta})$ is given by (13), and thus the FIM becomes

$$\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}} = \mathbb{E}_{\boldsymbol{\theta}}\left(\mathrm{J}_{\boldsymbol{\theta},\mathrm{d}} + \mathrm{J}_{\boldsymbol{\theta}}\right)$$

where $\mathrm{J}_{\boldsymbol{\theta}}$ is the FIM for the a priori knowledge of the parameter vector and can be expressed by

$$\begin{aligned}\mathrm{J}_{\boldsymbol{\theta}} &= \mathbb{E}_{\boldsymbol{\theta}}\left\{\left[\frac{\partial}{\partial \boldsymbol{\theta}}\ln \boldsymbol{f}(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial \boldsymbol{\theta}}\ln \boldsymbol{f}(\boldsymbol{\theta})\right]^{\mathrm{T}}\right\}\\ &= -\mathbb{E}_{\boldsymbol{\theta}}\left\{\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\ln \boldsymbol{f}(\boldsymbol{\theta})\right\}\end{aligned}$$

Let us define $\boldsymbol{R_\theta}$ as the co-variance matrix of the real flow vector, then the PDF of $\boldsymbol{\theta}$ can be expressed by

$$\boldsymbol{f}(\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)\boldsymbol{R}_{\boldsymbol{\theta}}^{\dagger}\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)^{\mathrm{T}}\right\}$$

where $\bar{\boldsymbol{\theta}}$ has the same definition as (9), $\boldsymbol{R}_{\boldsymbol{\theta}}^{\dagger}$ is the Moore-Penrose pseudoinverse of $\boldsymbol{R_\theta}$. Note that $\boldsymbol{R_\theta}$ has a possibility of being a singular matrix because of the spatial-temporal correlation between traffic at different road segments and time instants. The rank of $\boldsymbol{R_\theta}$ intuitively represents the number of "sources" in the traffic network. Hence, the FIM for the a priori knowledge of the parameter vector can be obtained by

$$\mathrm{J}_{\boldsymbol{\theta}} = \boldsymbol{R}_{\boldsymbol{\theta}}^{\dagger}$$

and the FIM for the joint PDF of observation and parameter vector can be expressed by

$$\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-\mathbf{1}}\boldsymbol{A} + \boldsymbol{R}_{\boldsymbol{\theta}}^{\dagger} \tag{16}$$

where $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-\mathbf{1}}\boldsymbol{A}$ and $\boldsymbol{R}_{\boldsymbol{\theta}}^{\dagger}$ are of the rank of $MT + K_{\mathrm{av}}$ and r, respectively. The rank of $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$ has the following relation

$$\begin{aligned}\mathrm{R}\left(\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}\right) &\leq \mathrm{R}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-\mathbf{1}}\boldsymbol{A}\right) + \mathrm{R}\left(\boldsymbol{R}_{\boldsymbol{\theta}}^{\dagger}\right)\\ &= MT + K_{\mathrm{av}} + r\\ &= MT + (N_m - M)(1 - R_2)T + r\end{aligned}$$

Thus, the necessary condition that an unbiased estimator of the entire parameter vector with finite variance exists is that $MT + (N_m - M)(1 - R_2)T + r \geq (N_m + N_u)T$. In the following context, we will derive the lower error bound assuming that $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$ is a full-rank matrix and a singular matrix, respectively. The former means that there exists an unbiased estimator with finite variance for the entire parameter vector $\boldsymbol{\theta}$ and the latter means that no unbiased estimator can be found and thus parameter transformation or biased estimator should be applied. Determining the error bound requires inverting the FIM $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$ in (16), which can be rewritten as

$$\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-\mathbf{1}}\boldsymbol{A} + \boldsymbol{R}_{\boldsymbol{\theta}}^{\dagger}$$

With the Woodbury matrix identity, the inverse of $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$ can be expressed by

$$\begin{aligned}\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}^{-\mathbf{1}} &= \left(\boldsymbol{R}_{\boldsymbol{\theta}}^{\dagger} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\epsilon}}^{-\mathbf{1}}\boldsymbol{A}\right)^{-1}\\ &= \boldsymbol{R_\theta} - \boldsymbol{R_\theta}\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{R_\epsilon} + \boldsymbol{A}\boldsymbol{R_\theta}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{R_\theta}\\ &= \boldsymbol{R_\theta} - \boldsymbol{R_\theta}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{q_o}}^{-\mathbf{1}}\boldsymbol{A}\boldsymbol{R_\theta}\end{aligned} \tag{17}$$

From (11), the mean squared error can be obtained from the trace of $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}^{-\mathbf{1}}$, that is

$$\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\|\,\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\,\|^2\right\} \geq \mathrm{tr}\left(\boldsymbol{R_\theta}\right) - \mathrm{tr}\left(\boldsymbol{R_\theta}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{q_o}}^{-\mathbf{1}}\boldsymbol{A}\boldsymbol{R_\theta}\right) \tag{18}$$

Note that $\boldsymbol{A}$ in (11) is given by (8), which is determined by the flow correlation between available data and the rest parameter vector (partial available data and entire unknown data) $\boldsymbol{r_{n,t}}$, and the co-variance matrix of the rest parameter vector $\boldsymbol{G_{n,t}}$. In the case of appropriate selection of sensor locations so that there is no causal relationship between each pair of available data points, $\boldsymbol{A}$ can be directly expressed by a product of cross correlation matrix between the available data and the unknown parameter vector, and inverse matrix of co-variance matrix of the unknown parameter vector. Then, (18) can be rewritten as

$$\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\|\,\hat{\boldsymbol{\theta}}_{\boldsymbol{u}} - \boldsymbol{\theta}_{\boldsymbol{u}}\,\|^2\right\} \geq \mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{\theta_u}}\right) - \mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}\boldsymbol{R}_{\boldsymbol{q_o}}^{-\mathbf{1}}\boldsymbol{R}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}^{\mathrm{H}}\right) \tag{19}$$

where $\boldsymbol{R}_{\boldsymbol{\theta_u}}$ and $\boldsymbol{R}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}$ are the co-variance matrix of the unknown parameter vector and cross correlation matrix between the available data and the unknown parameter vector, respectively. Then Theorem 1 is proved. □

When $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$ is a singular matrix, the following corollary can be obtained:

Corollary 1: In the case that $J_{\boldsymbol{\theta},r}$ is a singular matrix, an unbiased estimator of the unknown parameter vector $\boldsymbol{\theta_u}$ exists iff (if only if) the number of missing points is less than or equal to the rank of $J_{\boldsymbol{\theta},r}$ i.e. $(N_m + N_u)T - K_{av} - MT \leq r_J$, where r_J is the rank of $J_{\boldsymbol{\theta},r}$. Based on the given condition, the variance of $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}$ satisfies the following inequality

$$\boldsymbol{var_\theta}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\right) \succeq \begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\boldsymbol{U_1}\boldsymbol{\Lambda}_I^{-\mathbf{1}}\boldsymbol{U}_{\mathbf{1}}^H\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{\Lambda}_I$ and $\boldsymbol{U_1}$ are the non-zero eigenvalues and the corresponding eigenspace, respectively. When $(N_m + N_u)T - K_{av} - MT > r_J$, no unbiased estimator can be found and the variance of $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}$ satisfies the following inequality

$$\begin{aligned}\boldsymbol{H_u}\boldsymbol{var_\theta}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\right)\boldsymbol{H}_{\boldsymbol{u}}^H \succeq\ & \boldsymbol{H}\boldsymbol{U_1}\boldsymbol{\Lambda}_I^{-\mathbf{1}}\boldsymbol{U}_{\mathbf{1}}^H\boldsymbol{H}^{\boldsymbol{H}}\\ & + \left(\sigma_{\xi}^2 + \sigma_{\varepsilon}^2\right)\left(\boldsymbol{H_a}\boldsymbol{H}_{\boldsymbol{a}}^H - 2\boldsymbol{H}\boldsymbol{A}^{\dagger}\boldsymbol{A}^{\dagger H}\boldsymbol{H}^{\boldsymbol{H}}\right)\end{aligned}$$

where $\boldsymbol{H}$ is an arbitrary scaling matrix whose row space is orthogonal to the eigenspace dedicated to the zero eigenvalues of $J_{\boldsymbol{\theta},r}$, $\boldsymbol{H_u}$ and $\boldsymbol{H_a}$ are the sub-matrices of $\boldsymbol{H}$ corresponding to $\boldsymbol{\theta_u}$ and $\boldsymbol{\theta_a}$, respectively.

Proof: From [45], the general form of CRLB can be expressed by

$$\begin{aligned}\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\begin{matrix}\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{E}\left(\hat{\boldsymbol{\alpha}}\right)\right)\times\\ \left(\hat{\boldsymbol{\alpha}} - \boldsymbol{E}\left(\hat{\boldsymbol{\alpha}}\right)\right)^{\mathrm{H}}\end{matrix}\right\} &\succeq \frac{\partial \boldsymbol{E}\left(\hat{\boldsymbol{\alpha}}\right)}{\partial \boldsymbol{\theta}}\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}^{\dagger}\left(\frac{\partial \boldsymbol{E}\left(\hat{\boldsymbol{\alpha}}\right)}{\partial \boldsymbol{\theta}}\right)^{\mathrm{H}}\\ &= \frac{\partial \boldsymbol{E}\left(\hat{\boldsymbol{\alpha}}\right)}{\partial \boldsymbol{\theta}}\left(\boldsymbol{R_\theta} - \boldsymbol{R_\theta}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{q_o}}^{-\mathbf{1}}\boldsymbol{A}\boldsymbol{R_\theta}\right)\\ &\quad\times\left(\frac{\partial \boldsymbol{E}\left(\hat{\boldsymbol{\alpha}}\right)}{\partial \boldsymbol{\theta}}\right)^{\mathrm{H}}\end{aligned} \tag{20}$$

where $\boldsymbol{\alpha} = \boldsymbol{h}(\boldsymbol{\theta})$ and $\frac{\partial \boldsymbol{E}(\hat{\boldsymbol{\alpha}})}{\partial \boldsymbol{\theta}}$ is a $(N_m + N_u)T \times (N_m + N_u)T$ matrix. To guarantee the right side of (20) has no unbounded eigenvalue on its diagonal, the row space of $\frac{\partial \boldsymbol{E}(\hat{\boldsymbol{\alpha}})}{\partial \boldsymbol{\theta}}$ must be orthogonal to the eigenspace dedicated to the zero eigenvalues of $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$. Assume that $\boldsymbol{h}$ is a linear function and $\boldsymbol{\alpha}$ can be expressed by $\boldsymbol{\alpha} = \boldsymbol{H\theta}$. By unbiased estimation of $\boldsymbol{\alpha}$, (20) can be rewritten as

$$\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\left[\hat{\boldsymbol{\alpha}} - \boldsymbol{E}(\hat{\boldsymbol{\alpha}})\right]\left[\hat{\boldsymbol{\alpha}} - \boldsymbol{E}(\hat{\boldsymbol{\alpha}})\right]^{\mathrm{H}}\right\} \succeq \boldsymbol{H}\mathrm{J}^{\dagger}_{\boldsymbol{\theta},\mathrm{r}}\boldsymbol{H}^{\mathrm{H}} \tag{21}$$

Let $\boldsymbol{U_2}$ be the eigenspace dedicated to zero eigenvalues of $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$, then a sufficient condition that $\hat{\boldsymbol{\alpha}}$ is a variance-limited unbiased estimator is $\boldsymbol{HU_2} = \boldsymbol{0}$. By utilizing the eigenvector/eigenvalue representation of $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$

$$\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}} = \begin{bmatrix}\boldsymbol{U_1} & \boldsymbol{U_2}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{U_1^{\mathrm{H}}}\\ \boldsymbol{U_2^{\mathrm{H}}}\end{bmatrix}$$

the right side of (20) can be rewritten as $\boldsymbol{HU_1\Lambda_1^{-1}U_1^{\mathrm{H}}H^{\mathrm{H}}}$. By an unbiased estimator $\hat{\boldsymbol{\alpha}}$, we have

$$\boldsymbol{var_\theta}\left(\hat{\boldsymbol{\alpha}}\right) \succeq \boldsymbol{HU_1\Lambda_1^{-1}U_1^{\mathrm{H}}H^{\mathrm{H}}}$$

and

$$\begin{bmatrix}\boldsymbol{H_u} & \boldsymbol{H_a}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\\ \hat{\boldsymbol{\theta}}_{\boldsymbol{a}}\end{bmatrix} = \hat{\boldsymbol{\alpha}} \tag{22}$$

where $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{a}}$ are the $(N_m + N_u)T - K_{av} - MT \times 1$ and $K_{av} + MT \times 1$ vectors representing the estimated unknown and available parameters, respectively. Let us define the rank of $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$ as r_J. In the case that $(N_m + N_u)T - K_{av} - MT \le r_J$, the unbiased estimation of $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}$ can be obtained by

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{u}} = \left(\boldsymbol{H_u^{\mathrm{H}}H_u}\right)^{-1}\boldsymbol{H_u^{\mathrm{H}}}\left[\hat{\boldsymbol{\alpha}} - \boldsymbol{H_a}\left(\boldsymbol{\theta_a} + \boldsymbol{\xi} + \boldsymbol{\varepsilon}\right)\right]$$

then

$$\begin{aligned}\boldsymbol{var_\theta}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\right) &\succeq \begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\boldsymbol{U_1\Lambda_1^{-1}U_1^{\mathrm{H}}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}\\ &= \left(\boldsymbol{H_u^{\mathrm{H}}H_u}\right)^{-1}\boldsymbol{H_u^{\mathrm{H}}HU_1\Lambda_1^{-1}U_1^{\mathrm{H}}H^{\mathrm{H}}H_u}\left(\boldsymbol{H_u^{\mathrm{H}}H_u}\right)^{-1}\end{aligned} \tag{23}$$

In the case that $(N_m + N_u)T - K_{av} - MT > r_J$, (23) does not hold because $\boldsymbol{H_u}$ is no longer full-rank. The equation (23) can be transformed into

$$\boldsymbol{H_u}\hat{\boldsymbol{\theta}}_{\boldsymbol{u}} = \hat{\boldsymbol{\alpha}} - \boldsymbol{H_a}\left(\boldsymbol{\theta_a} + \boldsymbol{\xi} + \boldsymbol{\varepsilon}\right)$$

then

$$\begin{aligned}\boldsymbol{H_u var_\theta}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\right)\boldsymbol{H_u^{\mathrm{H}}} \succeq{}& \boldsymbol{HU_1\Lambda_1^{-1}U_1^{\mathrm{H}}H^{\mathrm{H}}} + \left(\sigma_\xi^2 + \sigma_\varepsilon^2\right)\boldsymbol{H_aH_a^{\mathrm{H}}}\\ &- 2\mathrm{cov}\left[\hat{\boldsymbol{\alpha}}, \boldsymbol{H_a}\left(\boldsymbol{\xi} + \boldsymbol{\varepsilon}\right)\right]\end{aligned} \tag{24}$$

Note that $\mathrm{cov}(\hat{\boldsymbol{\alpha}}, \boldsymbol{H_a}(\boldsymbol{\xi} + \boldsymbol{\varepsilon}))$ depends on $\boldsymbol{H_a}$ and $\hat{\boldsymbol{\alpha}}$, i.e., the applied estimator. It is well-known that a maximum likelihood (ML) estimator is asymptotically efficient [45]. By introducing $\boldsymbol{\alpha} = \boldsymbol{H\theta}$ to (7), it can be transformed to

$$\boldsymbol{q_o} = \boldsymbol{AH^{\dagger}\alpha} + \boldsymbol{u} + \boldsymbol{\epsilon}$$

then

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_{\mathrm{ml}} &= \left(\boldsymbol{AH^{\dagger}}\right)^{\dagger}\left(\boldsymbol{q_o} - \boldsymbol{u}\right)\\ &= \boldsymbol{\alpha} + \boldsymbol{HA^{\dagger}\epsilon}\end{aligned}$$

Then (24) can be rewritten as

$$\begin{aligned}\boldsymbol{H_u var_\theta}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\right)\boldsymbol{H_u^{\mathrm{H}}} \succeq{}& \boldsymbol{HU_1\Lambda_1^{-1}U_1^{\mathrm{H}}H^{\mathrm{H}}} + \left(\sigma_\xi^2 + \sigma_\varepsilon^2\right)\Big(\boldsymbol{H_aH_a^{\mathrm{H}}}\\ &- 2\boldsymbol{HA^{\dagger}A^{\dagger\mathrm{H}}H^{\mathrm{H}}}\Big)\end{aligned} \tag{25}$$

Then Corollary 1 is proved. □

Note that Theorem 1 and Corollary 1 indicate that the minimum mean squared error can be readily determined by covariance of the unknown parameter vector, flow correlation between the unknown parameter vector and the available data, and the sensor locations. Thus Theorem 1 and Corollary 1 can also be used to optimize the sensor locations in a large traffic network when the number of available sensors is limited.

IV. Optimal and Iterated Spatial-Temporal Kriging and Performance Analysis

In this section, we firstly propose an optimal spatial-temporal Kriging estimator and conduct its error analysis based on the data structure in (3). For simplicity of expression, we formulate the unknown parameters $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ into a new parameter vector $\boldsymbol{\theta_u}$, and the observed parameters into $\boldsymbol{q_o}$. It is often the case that the mean flow value is not constant over space and time in a traffic network. Thus, we utilize a simple spatial-temporal Kriging estimator proposed in [42], which can be expressed by

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\boldsymbol{u}} &= \mathbb{E}\left(\boldsymbol{\theta_u}|\boldsymbol{q_o}\right)\\ &= \bar{\boldsymbol{\theta}}_{\boldsymbol{u}} + \boldsymbol{R}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}\boldsymbol{R}_{\mathrm{q_o}}^{-1}\left(\boldsymbol{q_o} - \bar{\boldsymbol{\theta}}_{\boldsymbol{o}}\right) &\quad (26)\\ &= \bar{\boldsymbol{\theta}}_{\boldsymbol{u}} + \boldsymbol{R}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}\begin{bmatrix}\boldsymbol{R}_{\boldsymbol{\theta_o}} +\\ \left(\boldsymbol{\sigma_\xi^2} + \boldsymbol{\sigma_\varepsilon^2}\right)\boldsymbol{I}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{\theta_o} - \bar{\boldsymbol{\theta}}_{\boldsymbol{o}} +\\ \boldsymbol{\xi} + \boldsymbol{\varepsilon}\end{bmatrix} &\quad (27)\end{aligned}$$

where $\bar{\boldsymbol{\theta}}_{\boldsymbol{u}}$ and $\bar{\boldsymbol{\theta}}_{\boldsymbol{o}}$ are the $(N_m + N_u)T - K_{av} - MT \times 1$ and $K_{av} + MT \times 1$ average flow vectors for the unknown parameters and the observed parameters, respectively, $\boldsymbol{R}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}$ and $\boldsymbol{R}_{\mathrm{q_o}}$ are the co-variance matrix between $\boldsymbol{\theta_u}$ and $\boldsymbol{q_o}$, and auto correlation matrix of $\boldsymbol{q_o}$, respectively.

Remark 1: In this paper, it is assumed that the number of samples is sufficiently large so that the perturbations can be neglected. In practice, $\bar{\boldsymbol{\theta}}_{\boldsymbol{u}}$, $\boldsymbol{R}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}$, $\boldsymbol{R}_{\mathrm{q_o}}$ and $\bar{\boldsymbol{\theta}}_{\boldsymbol{o}}$ are perturbed by small drifts due to the possibility of insufficient statistics of the empirical data.

Equation (26) shows that spatial-temporal Kriging estimator is biased towards $\bar{\boldsymbol{\theta}}_{\boldsymbol{u}}$, rather than $\boldsymbol{\theta_u}$ if the dimension of $\boldsymbol{q_o}$ is not sufficiently large. The mean squared error (MSE) of (26) is given by

$$\begin{aligned}mse\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\right) &= \mathrm{tr}\left\{\mathbb{E}\left[\left(\boldsymbol{\theta_u} - \hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\right)\left(\boldsymbol{\theta_u} - \hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\right)^{\mathrm{H}}\right]\right\}\\ &= \mathrm{tr}\left\{\boldsymbol{R}_{\theta_{\mathrm{u}}} - \boldsymbol{R}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}\begin{bmatrix}\boldsymbol{R}_{\boldsymbol{\theta_o}} +\\ \left(\boldsymbol{\sigma_\xi^2} + \boldsymbol{\sigma_\varepsilon^2}\right)\boldsymbol{I}\end{bmatrix}^{-1}\boldsymbol{R}^{\mathrm{H}}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}\right\}\\ &= \mathrm{tr}\left(\boldsymbol{R}_{\theta_{\mathrm{u}}} - \boldsymbol{R}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}\boldsymbol{R}_{\mathrm{q_o}}^{-1}\boldsymbol{R}^{\mathrm{H}}_{\boldsymbol{\theta_u},\boldsymbol{q_o}}\right)\end{aligned} \tag{28}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

where $\boldsymbol{R}_{\theta_u}$ is the auto correlation matrix of $\boldsymbol{\theta_u}$. Since $\boldsymbol{R}_{\theta_u,q_o}\left[\boldsymbol{R}_{\theta_o}+\left(\sigma_\xi^2+\sigma_\varepsilon^2\right)\boldsymbol{I}\right]^{-1}\boldsymbol{R}_{\theta_u,q_o}^{\mathrm{H}}$ is a semi-positive definite matrix, then $mse\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}\right)\leq \operatorname{tr}\left(\boldsymbol{R}_{\theta_u}\right)$.

Remark 2: Note that in (27) and (28), Moore-Penrose pseudoinverse should be applied if $\boldsymbol{R}_{\theta_o}+\left(\sigma_\xi^2+\sigma_\varepsilon^2\right)\boldsymbol{I}$ is a singular matrix.

Corollary 2: Spatial-temporal estimator proposed by (27) is efficient for both cases where causal relationship among available data points exists or does not exist.

Proof: In the case that there is no causal relationship between each pair of available data points, it is straightforward to show that (27) is efficient because (28) achieves the CRLB proposed by (19). In the case that the causal relationship among available points exists and $\mathrm{J}_{\boldsymbol{\theta},\mathrm{r}}$ is not a singular matrix, the lower bound of MSE of $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}$ can be expressed by

$$\begin{aligned}\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}&\left\{\|\,\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}-\boldsymbol{\theta}_{\boldsymbol{u}}\,\|^2\right\}\\ &\geq \operatorname{tr}\left\{\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\left(\boldsymbol{R}_{\boldsymbol{\theta}}-\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{A}^{\mathbf{H}}\boldsymbol{R}_{\boldsymbol{q_o}}^{-1}\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{\theta}}\right)\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}\right\}\\ &=\operatorname{tr}\left\{\boldsymbol{R}_{\theta_u}-\begin{bmatrix}\boldsymbol{I}\ \boldsymbol{0}\end{bmatrix}\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{A}^{\mathbf{H}}\boldsymbol{R}_{\boldsymbol{q_o}}^{-1}\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}\right\}\end{aligned}\tag{29}$$

Compared with the MSE in (28), we have

$$\begin{aligned}&\operatorname{tr}\begin{pmatrix}\boldsymbol{R}_{\theta_u}-\boldsymbol{R}_{\theta_u,q_o}\boldsymbol{R}_{q_o}^{-1}\boldsymbol{R}_{\theta_u,q_o}^{\mathrm{H}}-\boldsymbol{R}_{\theta_u}\\ +\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{A}^{\mathbf{H}}\boldsymbol{R}_{\boldsymbol{q_o}}^{-1}\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}\end{pmatrix}\\ &=\operatorname{tr}\begin{pmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{A}^{\mathbf{H}}\boldsymbol{R}_{\boldsymbol{q_o}}^{-1}\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}\\ -\boldsymbol{R}_{\theta_u,q_o}\boldsymbol{R}_{q_o}^{-1}\boldsymbol{R}_{\theta_u,q_o}^{\mathrm{H}}\end{pmatrix}\\ &=\operatorname{tr}\left[\begin{pmatrix}\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{A}^{\mathbf{H}}\\ -\boldsymbol{R}_{\theta_u,q_o}^{\mathrm{H}}\boldsymbol{R}_{\theta_u,q_o}\end{pmatrix}\boldsymbol{R}_{q_o}^{-1}\right]\end{aligned}\tag{30}$$

Equation (30) indicates that the spatial-temporal estimator can be shown to be efficient if we can prove $\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{A}^{\mathbf{H}}=\boldsymbol{R}_{\theta_u,q_o}^{\mathrm{H}}\boldsymbol{R}_{\theta_u,q_o}$. Recall that $\boldsymbol{A}$ is a $MT+K_{\mathrm{av}}\times(N_m+N_u)\,T$ scaling matrix, each row of which is actually a product of a flow correlation vector, an inverse matrix of a cofactor matrix of $\boldsymbol{R}_{\boldsymbol{\theta}}$ and an insertion matrix. Thus, the row vector of $\boldsymbol{A}$ can be rewritten as

$$\boldsymbol{A}_{\boldsymbol{l}}=\boldsymbol{r}_{\boldsymbol{l}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\theta},\boldsymbol{l},\boldsymbol{l}}^{*-1}\boldsymbol{B}_{\boldsymbol{l}}\tag{31}$$

where $\boldsymbol{A}_{\boldsymbol{l}}$ stands for the l-th row of $\boldsymbol{A}$, $\boldsymbol{R}_{\boldsymbol{\theta},\boldsymbol{l},\boldsymbol{l}}^{*}$ is the (l,l)-th co-factor matrix of $\boldsymbol{R}_{\boldsymbol{\theta}}$ and $\boldsymbol{B}_{\boldsymbol{l}}$ is the l-th insertion matrix inserting the $(l+(N_m+N_u)\,T-MT+K_{\mathrm{av}})$-th column of $(N_m+N_u)\,T\times(N_m+N_u)\,T$ identity matrix with zero vector. By substituting (31) into (30), it is not hard to see that $\boldsymbol{r}_{\boldsymbol{l}}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{\theta},\boldsymbol{l},\boldsymbol{l}}^{*-1}\boldsymbol{B}_{\boldsymbol{l}}\boldsymbol{R}_{\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}$ is the flow correlation vector between the l-th available point and the unknown parameter vector, and thus $\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}\triangleq\boldsymbol{R}_{\theta_u,q_o}^{\mathrm{H}}$ and Corollary 2 is proved. □

Corollary 2 reveals that (27) is an optimal estimator. Equation (28) shows that the optimal spatial-temporal estimator requires an inverse operation of the auto correlation matrix of the observed parameter vector, which leads to high computational effort with a large size of the observed parameter vector. To tackle the problem, we propose a sub-optimal iterative multiple-points spatial-temporal Kriging technique, which iteratively imputes the missing points via the most appropriate available data. At each iteration, the most appropriate available data being used for imputation is selected such that the flow correlation with the missing points is maximum while the covariance matrix of the available data is not ill-conditioned. Note that the available data in the i-th iteration may be the estimated value in the $i-1$-th iteration.

Let us define k_{av} and k_{im} as the constant number of the selected available data and the missing points being imputed at each iteration, respectively. Let e^{χ} be MSE of the iterated spatial-temporal Kriging when a specific imputation order $\chi\in\Phi$ is used. The cardinality of imputation order set Φ is determined by the number of total missing points $K_m+N_{\mathrm{u}}T$ and the missing points being imputed at each iteration k_{im}, where

$$|\Phi|=\prod_{i=1}^{\lfloor (K_m+N_{\mathrm{u}}T)/k_{\mathrm{im}}\rfloor}\binom{K_m+N_{\mathrm{u}}T-ik_{\mathrm{im}}}{k_{\mathrm{im}}}$$

Then e^{χ} can be expressed by

$$\begin{aligned}e^{\chi}&=\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\sum_{i=1}^{I}\|\,\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{\boldsymbol{i},\boldsymbol{\chi}}-\boldsymbol{\theta}_{\boldsymbol{u}}^{\boldsymbol{i},\boldsymbol{\chi}}\,\|^2\right\}\\ &=\sum_{i=1}^{L}\mathbb{E}_{\boldsymbol{q_o},\boldsymbol{\theta}}\left\{\|\,\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{\boldsymbol{i},\boldsymbol{\chi}}-\boldsymbol{\theta}_{\boldsymbol{u}}^{\boldsymbol{i},\boldsymbol{\chi}}\,\|^2\right\}&(32)\\ &=\sum_{i=1}^{L}e^{i,\chi}&(33)\end{aligned}$$

where L is the number of iterations, $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{\boldsymbol{i},\boldsymbol{\chi}}$ and $\boldsymbol{\theta}_{\boldsymbol{u}}^{\boldsymbol{i},\boldsymbol{\chi}}$ are the estimated and real unknown parameter vector of the i-th iteration for the order χ. In what follows, we give the closed form of e^{χ} and investigate the relation between e^{χ} and χ, k_{av} and k_{im}.

By (28), at the i-th iteration for a specific order χ, the $e^{i,\chi}$ can be expressed by

$$\begin{aligned}e^{i,\chi}&=\operatorname{tr}\left(\boldsymbol{R}_{\theta_u}^{\boldsymbol{i},\boldsymbol{\chi}}\right)-\operatorname{tr}\left(\boldsymbol{R}_{\theta_u,q_{av}}^{\boldsymbol{i},\boldsymbol{\chi}}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi}\right)\\ \boldsymbol{q}_{av}^{\boldsymbol{i},\boldsymbol{\chi}}&=\underset{q_{av}\in\left\{\boldsymbol{q}_{av}^{i-1,\chi}\cup\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i-1,\chi}\right\}}{\operatorname{argmax}}\left[\operatorname{tr}\left(\boldsymbol{R}_{\theta_u,q_{av}}^{\boldsymbol{i},\boldsymbol{\chi}}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi}\right)\right]\end{aligned}\tag{34}$$

where $\boldsymbol{R}_{\theta_u}^{\boldsymbol{i},\boldsymbol{\chi}}$, $\boldsymbol{R}_{q_{av}}^{\boldsymbol{i},\boldsymbol{\chi}}$ and $\boldsymbol{R}_{\theta_u,q_{av}}^{\boldsymbol{i},\boldsymbol{\chi}}$ are the $k_{\mathrm{im}}\times k_{\mathrm{im}}$, $k_{\mathrm{av}}\times k_{\mathrm{av}}$ and $k_{\mathrm{im}}\times k_{\mathrm{av}}$ auto correlation matrix of the selected unknown parameter vector, available parameter vector and covariance matrix between the selected unknown parameter vector and the available parameter vector, respectively. Note that $\boldsymbol{q}_{av}^{\boldsymbol{i},\boldsymbol{\chi}}$ in (34) is the most appropriate available data at the i-th iteration. Substituting (34) into (33), e^{χ} can be rewritten by

$$e^{\chi}=\operatorname{tr}\left(\boldsymbol{R}_{\theta_u}\right)-\sum_{i=1}^{I}\operatorname{tr}\left(\boldsymbol{R}_{\theta_u,q_{av}}^{\boldsymbol{i},\boldsymbol{\chi}}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi}\right)\tag{35}$$

Remark 3: Note that at the i-th iteration, where $i > 1$, $e^{i,\chi}$ may be slightly different from that given by (34), because there is possibility that some elements of $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i-1,\chi}$ are selected to be the available data. In such case, $\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i,\chi}$ should be rewritten by

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i,\chi} &= \bar{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i,\chi} + \boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\left(\boldsymbol{q}_{av}^{i,\chi} - \bar{\boldsymbol{\theta}}_{av}^{i,\chi}\right) \\ &= \bar{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i,\chi} + \boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\left(\boldsymbol{\theta}_{av}^{i,\chi} - \bar{\boldsymbol{\theta}}_{av}^{i,\chi} + \boldsymbol{\epsilon}^{i,\chi}\right) \end{aligned} \quad (36)$$

and

$$\begin{aligned} e^{i,\chi} &= \mathrm{tr}\left\{\mathbb{E}\left[\left(\boldsymbol{\theta}_{\boldsymbol{u}}^{i,\chi} - \hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i,\chi}\right)\left(\boldsymbol{\theta}_{\boldsymbol{u}}^{i,\chi} - \hat{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i,\chi}\right)^{\mathrm{H}}\right]\right\} \\ &= \mathrm{tr}\left\{\begin{array}{c}\mathbb{E}\left[\begin{array}{c}\boldsymbol{\theta}_{\boldsymbol{u}}^{i,\chi} - \bar{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i,\chi} - \\ \boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\left(\boldsymbol{\theta}_{av}^{i,\chi} - \bar{\boldsymbol{\theta}}_{av}^{i,\chi} + \boldsymbol{\epsilon}^{i,\chi}\right)\end{array}\right] \\ \times\left[\begin{array}{c}\boldsymbol{\theta}_{\boldsymbol{u}}^{i,\chi} - \bar{\boldsymbol{\theta}}_{\boldsymbol{u}}^{i,\chi} - \\ \boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\left(\boldsymbol{\theta}_{av}^{i,\chi} - \bar{\boldsymbol{\theta}}_{av}^{i,\chi} + \boldsymbol{\epsilon}^{i,\chi}\right)\end{array}\right]^{\mathrm{H}}\end{array}\right\} \\ &= \mathrm{tr}\left[\begin{array}{c}\boldsymbol{R}_{\theta_u}^{i,\chi} - 2\boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi} - \\ \boldsymbol{R}_{\theta_u,\epsilon}^{i,\chi}\boldsymbol{R}_{q_{\mathrm{av}}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi} - \\ \boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{\mathrm{av}}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,\epsilon}^{\mathrm{H},i,\chi} + \\ \boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{\mathrm{av}}}^{-1,i,\chi}\left(\begin{array}{c}\boldsymbol{R}_{\theta_{av}}^{i,\chi} + \\ \boldsymbol{\Lambda}_{\epsilon}^{i,\chi}\end{array}\right)\boldsymbol{R}_{q_{\mathrm{av}}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi}\end{array}\right] \end{aligned} \quad (37)$$

where $\boldsymbol{\epsilon}^{i,\chi}$ is the error term of the available data at the i-th iteration for the order χ, $\boldsymbol{\Lambda}_{\epsilon}^{i,\chi}$ is the co-variance matrix of $\boldsymbol{\epsilon}^{i,\chi}$, $\boldsymbol{R}_{\theta_u,\epsilon}^{i,\chi}$ is the co-variance matrix between $\boldsymbol{\theta}_{\boldsymbol{u}}^{i,\chi}$ and $\boldsymbol{\epsilon}^{i,\chi}$.

Remark 3 shows that equation (37) reveals the same result as (34) when $\boldsymbol{\Lambda}_{\epsilon}^{i,\chi} = \left(\sigma_{\xi}^2 + \sigma_{\varepsilon}^2\right)\boldsymbol{I}$, i.e. all selected available data is observed data, rather than containing the available data estimated from the $i-1$-th iteration, where $\boldsymbol{I}$ is the identity matrix. For $i = 1$, it is clear that $\boldsymbol{\epsilon}^{i,\chi}$ only consists of the measurement error $\xi + \varepsilon$ from the observed data. But for $i > 1$, $\boldsymbol{\epsilon}^{i,\chi}$ may contain the estimation error delivered in the $i-1$-th iteration due to the available data selection in the current iteration. The estimation error can be traced back to the observed data and can be classified into two types: measurement error propagation and error caused by insufficient statistics. The measurement error of the observed data can propagate with the evolution if the estimated value at the previous iteration is used as the available data at the current iteration. Error caused by insufficient statistics is attributable to that for some missing points at specific iterations, the selected available data may not cover all available information relating to the missing points. For example, in a specific iteration, there are $k'_{\mathrm{av}} + u$ observed data having spatial-temporal correlation with the missing points. However, the k_{av} selected available data can only be traced back to k'_{av} related observed data. The rest u observed data is not utilized and thus leads to additional estimation error.

From (37), e^{χ} can be expressed by

$$\begin{aligned} e^{\chi} &= \mathrm{tr}\left(\boldsymbol{R}_{\theta_u}\right) - \sum_{i=1}^{I}\mathrm{tr}\left(\boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi}\right) \\ &\quad - 2\sum_{i=1}^{I}\mathrm{tr}\left(\boldsymbol{R}_{\theta_u,\epsilon}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi}\right) \\ &\quad + \sum_{i=1}^{I}\mathrm{tr}\left\{\boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\left[\begin{array}{c}\boldsymbol{\Lambda}_{\epsilon}^{i,\chi} - \\ \left(\sigma_{\xi}^2 + \sigma_{\varepsilon}^2\right)\boldsymbol{I}\end{array}\right] \times\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi}\right\} \end{aligned} \quad (38)$$

The following lemma can be easily obtained:

Lemma 1: Iterated spatial-temporal kriging estimator can achieve the CRLB proposed by (18) iff (if only if) the selected available data contain all spatial-temporal information relating to the missing points at each iteration. In such case, the performance of iterated spatial-temporal kriging estimator is irrelevant to a specific imputation order χ.

Proof: When the selected available data contain all spatial-temporal information relating to the missing points at each iteration, $\boldsymbol{R}_{\theta_u,\epsilon}^{i,\chi}$ becomes zero, $\sum_{i=1}^{I}\mathrm{tr}\left(\boldsymbol{R}_{\theta_u,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_{av}}^{\mathrm{H},i,\chi}\right)$ achieves $\mathrm{tr}\left(\boldsymbol{R}_{\theta_u,q_o}\boldsymbol{R}_{q_o}^{-1}\boldsymbol{R}_{\theta_u,q_o}^{\mathrm{H}}\right)$, and $\boldsymbol{\Lambda}_{\epsilon}^{i,\chi}$ only contains the measurement error propagated from the previous iterations. If the measurement error comes totally from the observed data, it is clear that $\boldsymbol{\Lambda}_{\epsilon}^{i,\chi} = \left(\sigma_{\xi}^2 + \sigma_{\varepsilon}^2\right)\boldsymbol{I}$ for each iteration and thus the last term in (38) disappears. For the case that the measurement error does not come directly from the observed data, but fully or partially from the data estimated in the previous iteration, $\boldsymbol{\Lambda}_{\epsilon}^{i,\chi}$ can be expressed by

$$\boldsymbol{\Lambda}_{\epsilon}^{i,\chi} = \mathbb{E}\left(\left[\begin{array}{c}\boldsymbol{\xi}_1 + \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\xi}_{k_o} + \boldsymbol{\varepsilon}_{k_o} \\ \epsilon_1^{i-1,\chi} \\ \vdots \\ \epsilon_{k_{av}^i}^{i-1,\chi}\end{array}\right]\left[\begin{array}{c}\boldsymbol{\xi}_1 + \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\xi}_{k_o} + \boldsymbol{\varepsilon}_{k_o} \\ \epsilon_1^{i-1,\chi} \\ \vdots \\ \epsilon_{k_{av}^i}^{i-1,\chi}\end{array}\right]^{\mathbf{H}}\right) \quad (39)$$

where k_{av}^i is the number of available data estimated in the previous iteration, the rest $k_{av} - k_{av}^i$ available data are from the observed data, when $k_{av}^i = k_{av}$, the available data are fully from the estimated data. The equation (39) can be further derived as

$$\boldsymbol{\Lambda}_{\epsilon}^{i,\chi} = \left[\begin{array}{cc}\left(\sigma_{\xi}^2 + \sigma_{\varepsilon}^2\right)\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_{\epsilon}^{i,\chi}\end{array}\right]$$

where $\boldsymbol{R}_{\epsilon}^{i,\chi}$ is not necessarily a diagonal matrix because of possibility of cross correlation among the estimation error at different iterations. To evaluate $\boldsymbol{\Lambda}_{\epsilon}^{i,\chi}$, we need to investigate how $\boldsymbol{\epsilon}^{i,\chi}$ propagates with i. Let us define $\epsilon_n^{i,\chi}$ as the estimation error at the n-th missing point at the i-th iteration for an arbitrary imputation order, then it can be expressed by

$$\epsilon_n^{i,\chi} = \boldsymbol{R}_{\theta_n,q_{av}}^{i,\chi}\boldsymbol{R}_{q_{av}}^{-1,i,\chi}\boldsymbol{\epsilon}^{i-1,\chi} = \sum_{k=1}^{k_{av}^i} w_k^i \epsilon_k^{i-1,\chi} \quad (40)$$

where $w_k^i, k = 1 \cdots k_{av}^i$ is the weight vector revealing correlation between $\boldsymbol{\epsilon}^{i-1,\chi}$ is the error vector estimated at the i-1-th iteration. The k-th error term at the $i-1$-th iteration $\epsilon_k^{i-1,\chi}$ can be further traced back to the observed data with the same procedure as (40). Thus, $\epsilon_n^{i,\chi}$ can be further expressed by

$$\epsilon_n^{i,\chi} = \sum_{k=1}^{k_o}\left(\prod_{\ell=1}^{i} w_{k,n}^{\ell}\right)\left(\boldsymbol{\xi}_k + \boldsymbol{\varepsilon}_k\right)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

where $\prod_{\ell=1}^{i} w_k^{\ell}$ is the correlation between the k-th observed data and the n-th missing point at the i-th iteration, and thus it is the (n, k)-th element of the matrix $\boldsymbol{R}_{\theta_n,q_o}^{i,\chi}\boldsymbol{R}_{q_o}^{-1,i,\chi}$. Thus, the last term in (38) can be straightforwardly transformed to

$$\sum_{i=1}^{I}\mathrm{tr}\left\{\begin{array}{c}\boldsymbol{R}_{\theta_u,q_o}^{i,\chi}\boldsymbol{R}_{q_o}^{-1,i,\chi}\begin{bmatrix}\left(\sigma_\xi^2+\sigma_\varepsilon^2\right)\boldsymbol{I}- \\ \left(\sigma_\xi^2+\sigma_\varepsilon^2\right)\boldsymbol{I}\end{bmatrix} \\ \times\boldsymbol{R}_{q_o}^{-1,i,\chi}\boldsymbol{R}_{\theta_u,q_o}^{\mathrm{H},i,\chi}\end{array}\right\}=0$$

Then Lemma 1 is proved. □

Remark 4: Because the performance is irrelevant to the imputation order when the selected available data cover all related information, Lemma 1 is specially useful when we only want to estimate partial unknown data. In such case, the benefits of iterated spatial-kriging technique becomes more dominant because it can greatly lower the computational complexity without performance degradation. In a traffic network, the available data that cover all information relating to the missing points is not hard to be determined because of road topology.

Lemma 1 also suggests that there is possibility to simplify a certain form of matrix multiplication by reducing matrix dimension. From Lemma 1, the following corollary can be easily derived:

Corollary 3: For any $K' \times N'$ matrix $\boldsymbol{A}$ and $N' \times N'$ non-singular Hermitian matrix $\boldsymbol{M}$, where $\boldsymbol{A}$ and $\boldsymbol{M}$ can be written in their sub-matrices forms

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_N\end{bmatrix} \tag{41}$$

$$\boldsymbol{M}=\begin{bmatrix}\boldsymbol{M}_{11} & \boldsymbol{M}_{12} & \cdots & \boldsymbol{M}_{1N} \\ \boldsymbol{M}_{12}^H & \boldsymbol{M}_{22} & \cdots & \boldsymbol{M}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{M}_{1N}^H & \boldsymbol{M}_{2N}^H & \cdots & \boldsymbol{M}_{NN}\end{bmatrix} \tag{42}$$

,$\boldsymbol{A}_n$ and $\boldsymbol{M}_{mn}^H, m \neq n$ are with the size of $K' \times N_n'$ and $N_n' \times N_m'$, respectively, if there exists matrices $\boldsymbol{\lambda}_n, n = 1\cdots N$ that fulfill

$$\boldsymbol{A}_m=\sum_{n=1}^{N}\boldsymbol{\lambda}_n\boldsymbol{M}_{nm}^H,\quad m=1\cdots N \tag{43}$$

then the following equation can be obtained

$$\boldsymbol{A}\boldsymbol{M}^{-1}\boldsymbol{A}^H=\sum_{m,n=1}^{N}\boldsymbol{\lambda}_m\boldsymbol{M}_{mn}\boldsymbol{\lambda}_n^H \tag{44}$$

Proof: Suppose that we use N' observed data to impute K' missing points. The N' observed data can be arbitrarily classified into N groups. The matrix $\boldsymbol{M}$ is the co-variance matrix of the N' observed data which can be expressed by the form of (42). Each sub-matrix $\boldsymbol{M}_{mn}$ can be considered as cross-correlation between the m-th and n-th group. We assume that the K' missing points is actually a linear combination of N groups of the observed data with the form:

$$\boldsymbol{\theta}_\mathrm{u}=\sum_{n=1}^{N}\boldsymbol{\lambda}_n\boldsymbol{\theta}_\mathrm{o}^{(n)}$$

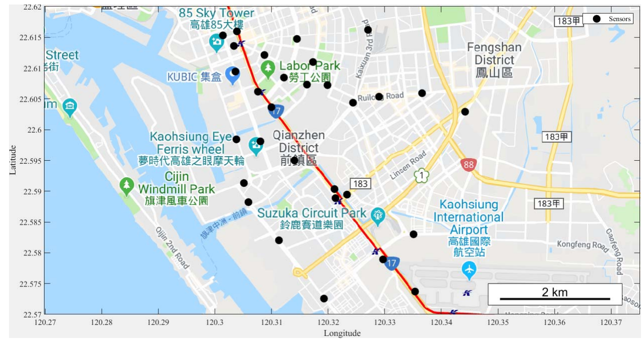

Fig. 1. Selected map for experiment: urban area of Kaohsiung, Taiwan.

where $\boldsymbol{\theta}_\mathrm{o}^{(n)}$ is the n-th group of the observed data and the definition of $\boldsymbol{\lambda}_n$ is the same as that is given in (43). From (28), the MSE of estimation error of the K' missing points can be determined by

$$mse\left(\boldsymbol{\theta}_\mathrm{u}|\boldsymbol{q}_o\right)=\boldsymbol{R}_{\theta_\mathrm{u}}-\boldsymbol{A}\boldsymbol{M}^{-1}\boldsymbol{A}^\mathrm{H} \tag{45}$$

where $\boldsymbol{A}$ is the cross-correlation matrix between $\boldsymbol{\theta}_\mathrm{u}$ and $\boldsymbol{\theta}_\mathrm{o}$, and can be expressed by (43). Suppose that there exists K' available data with the same correlation to the observed data and covers all information about the missing points, then the MSE of estimation error of the K' missing points based on the K' available data can be expressed by

$$mse\left(\boldsymbol{\theta}_\mathrm{u}|\boldsymbol{q}_{av}\right)=\boldsymbol{R}_{\theta_\mathrm{u}}-\sum_{m,n=1}^{N}\boldsymbol{\lambda}_m\boldsymbol{M}_{mn}\boldsymbol{\lambda}_n^\mathrm{H} \tag{46}$$

From lemma 1, we know that (45) must be identical to (46). Then Corollary 3 is proved. □

Remark 5: Note that Corollary 3 can also be verified by matrix operation. It is specially useful for the large dimension of $\boldsymbol{M}$, where inverse operation becomes cumbersome. Corollary 3 can also be used to support Lemma 1.

V. Experimental Results

In this section, we use the real traffic data provided by the transportation department of Kaohsiung, Taiwan to validate our theoretical findings.

The selected flow data was collected by 30 loop detectors on the road segments located in an urban area of Kaohsiung, Taiwan, over the period of 01/04/2013 to 30/04/2013 (Fig. 1). The black nodes in Fig. 1 represent the deployed sensors. The horizontal and vertical axis represent the longitude and latitude, respectively. Each sensor provides the flow data of 10-minute interval from 00:00 to 23:59 for each day. To validate the theoretical findings in our paper, the malfunctioned sensors and the unmeasured spatio-temporal points are intentionally generated with different missing ratios that ranges from 25% to 45% at every 5% increment. We consider the Missing Completely at Random manner (MCR), where the malfunctioned sensors and the unmeasured spatio-temporal points are independently and uniformly distributed over the spatio-temporal domain.

We calculated the SFEB (Squared Flow Error Bound) determined by the co-variance matrix of the real flow vector,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

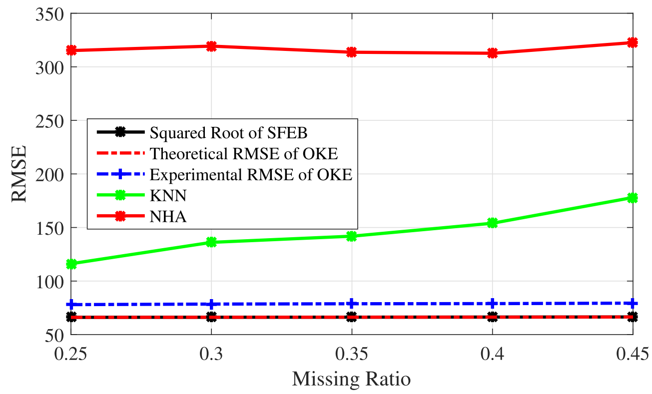

Fig. 2. Comparison of the squared root of SFEB, RMSE of the optimal Kriging estimator, KNN and NHA for SNR = 10dB.

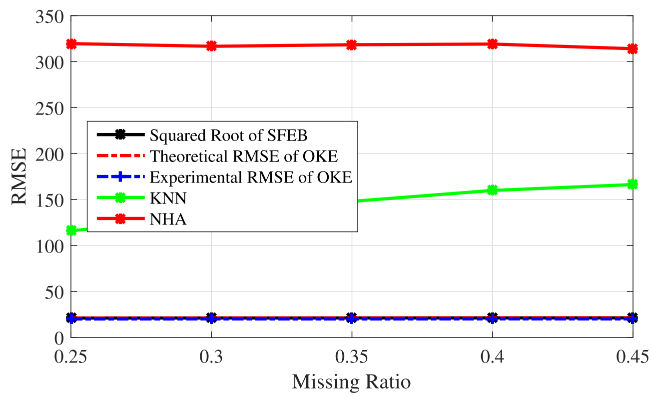

Fig. 3. Comparison of the squared root of SFEB, RMSE of the optimal Kriging estimator, KNN and NHA for SNR = 20dB.

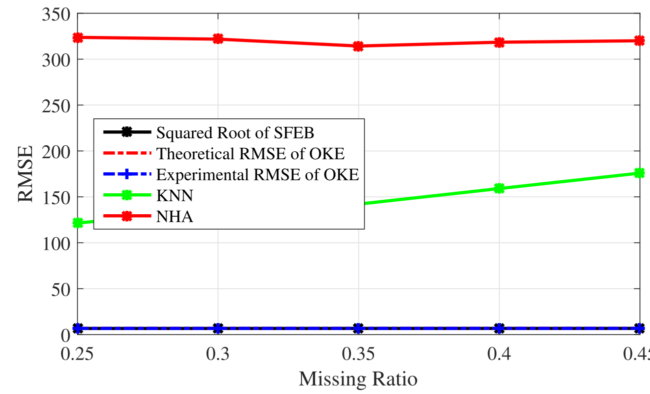

Fig. 4. Comparison of the squared root of SFEB, RMSE of the optimal Kriging estimator, KNN and NHA for SNR = 30dB.

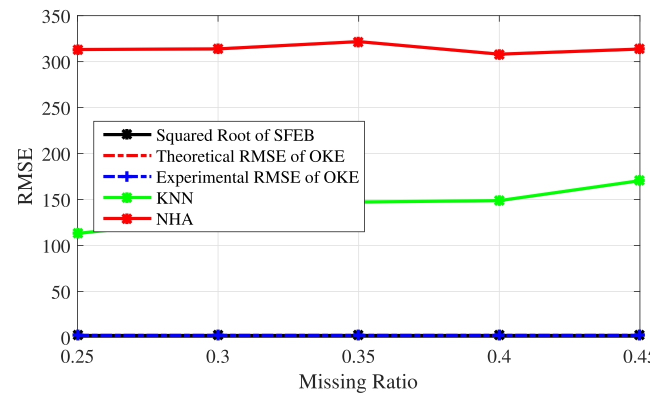

Fig. 5. Comparison of the squared root of SFEB, RMSE of the optimal Kriging estimator, KNN and NHA for SNR = 40dB.

scaling matrix and the co-variance matrix of the observed flow vector. We estimated the flow values on the malfunctioned sensors and the unmeasured spatio-temporal points by using the optimal spatial-temporal Kriging estimator, Nearest Historical Average (NHA) and K Nearest Neighbor (KNN) methods, and compared their performance to each other in terms of RMSE (Root Mean Squared Error). The optimal and iterated spatial-temporal Kriging estimators were proposed in our paper. NHA is the most commonly used method in the data imputation because it shows a stable performance regardless of the missing data size with an easy implementation [23]. NHA replaces the missing data by the arithmetic average or weighted average of the nearest historical data [46]. NHA does not incorporate the information from neighboring roads at the same day and is based on the assumption that traffic pattern at the same sensor at the same time instant is similar from day to day. KNN methods is to fill the missing data with the arithmetic or weighted average of data on K neighboring sensors. The K neighboring sensors are selected by searching for the data with close physical distances or equivalent distances [22].

The experimental results are shown in Fig. 2 to Fig. 5, which compare the squared root of SFEB, RMSE of optimal Kriging estimator, KNN and NHA for SNR (Signal to Noise Ratio) is equal to 10dB, 20dB, 30dB and 40dB, respectively. SNR in this paper is defined by the ratio between the variance of real traffic flow data to the variance of measurement noise. To validate the impact on SFEB by different measurement noise power, we intentionally create measurement noise with different SNR. As can be seen from Fig. 2 to Fig. 5, theoretical RMSE of the Optimal Kriging Estimator (OKE) given by Equation (28) coincides with the squared root of SFEB given by Theorem 1 for all missing ratio and all SNRs. The experimental RMSE of the OKE depicts a gap with the squared root of SFEB for SNR = 10dB while generally coincides with the squared root of SFEB for higher SNR values. The experimental RMSE of the OKE is obtained by Equation (27), which requires to determine the co-variance matrix between the unmeasured data and the observed data, and the auto-correlation matrix of the observed data. With larger measurement noise, the determined co-variance matrix and auto-correlation matrix are subject to perturbation due to insufficient statistics, and thus the gap is created. RMSE delivered by KNN and NHA is dominantly larger than the squared root of SFEB and RMSE of OKE for all missing ratios and all SNR. The results show that performance of any missing points estimators is lowered by SFEB, and the OKE is an efficient estimator, and thus Theorem 1 and Corollary 2 are validated. It can be seen that NHA indeed shows a general stable performance for different missing ratio and the performance of KNN degrades with an increasing missing ratio. SFEB increases with an increasing variance of the measurement noise.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

VI. Conclusions and Future Work

In this paper, we analyzed the CRLB of missing traffic data imputation and the sufficient and necessary condition for the existence of an unbiased estimator for both the case that the Fisher matrix is a non-singular and the case that the Fisher matrix is a singular matrix. We show the CRLB strongly relies on the three parameters, viz. covariance of the unknown parameter vector, flow correlation between the unknown parameter vector and the available data, and the sensor locations. We have shown the importance of relationship between the number of missing points and the rank of the Fisher matrix, i.e., an unbiased estimator of the unknown parameter vector can be found if only if the number of missing points is less than or equal to the rank of the Fisher matrix. We also derived an inequality on the variance of estimation error for the case that the above mentioned condition is not fulfilled.

We proved that the spatial-temporal estimator proposed in [42] is optimal for both cases that causal relationship between available data points exists or not. To overcome high computational complexity with a large size of the observed parameter vector, we proposed a sub-optimal iterative multiple-points spatial-temporal Kriging technique. We derived the variance of estimation error and showed the sufficient and necessary condition with which the iterative multiple-points spatial-temporal Kriging estimator can achieve the CRLB, i.e. the selected available data covers all information relating to the missing points at each iteration, in such case, imputation order becomes irrelevant to the performance. We further derived an useful corollary that can simplify a certain form of matrix multiplication by reducing matrix dimension. This corollary becomes more significant with a large matrix dimension where inverse operation becomes cumbersome.

The proposed results in this paper can be used as a benchmark to evaluate the performance of missing data estimators in future work. The performance of any missing data estimators is lower bounded by the proposed SFEB. Corollary 1 can be utilized to optimize the sensor locations and the number of deployed sensors in the network. An application example is to deploy the sensors on the locations so that FIM is a full-rank matrix. Lemma 1 can be utilized to reduce computational complexity with a large size of the observed parameter vector. As an example, the missing points can be estimated iteratively with optimal selection of available data according to Lemma 1 at each iteration, rather than using the optimal Kriging estimator.

References

[1] Y.-J. Wu, F. Chen, C.-T. Lu, and S. Yang, “Urban traffic flow prediction using a spatio temporal random effects model,” *J. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 282–293, 2016. doi: 10.1080/15472450.2015.1072050.

[2] H. S. Mahmassani, “Dynamic network traffic assignment and simulation methodology for advanced system management applications,” *Netw. Spatial Econ.*, vol. 1, nos. 3–4, pp. 267–292, Sep. 2001.

[3] E. Mitsakis, J. M. Salanova, and G. Giannopoulos, “Combined dynamic traffic assignment and urban traffic control models,” *Procedia-Social Behav. Sci.*, vol. 20, pp. 427–436, Jan. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042811014297

[4] S. Amini, I. Gerostathopoulos, and C. Prehofer, “Big data analytics architecture for real-time traffic control,” in *Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Jun. 2017, pp. 710–715.

[5] M. A. Kafi, Y. Challal, D. Djenouri, M. Doudou, A. Bouabdallah, and N. Badache, “A study of wireless sensor networks for urban traffic monitoring: Applications and architectures,” *Procedia Comput. Sci.*, vol. 19, pp. 617–626, Jan. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187705091300690X

[6] S. Akinboro, J. A. Adeyiga, A. Omotosho, and A. Akinwunmi, “Mobile road traffic management system using weighted sensors,” *Int. J. Interact. Mobile Technol.*, vol. 11, no. 5, pp. 147–160, Jul. 2017.

[7] A. Faghri and K. Hamad, “Application of GPS in traffic management systems,” *GPS Solutions*, vol. 5, no. 3, pp. 52–60, Jan. 2001.

[8] K. G. Zografos, K. N. Androutsopoulos, and G. M. Vasilakis, “A real-time decision support system for roadway network incident response logistics,” *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 1, pp. 1–18, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X00000437

[9] W. Wong and S. C. Wong, “Evaluation of the impact of traffic incidents using GPS data,” *Proc. Inst. Civil Eng.-Transp.*, vol. 169, no. 3, pp. 148–162, 2016.

[10] Z. Y. chuan, Z. X. qing, Z. liting, and C. Zhenting, “Traffic congestion detection based On GPS floating-car data,” *Procedia Eng.*, vol. 15, pp. 5541–5546, 2011, cEIS 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187770581102529X

[11] N. Caceres, L. M. Romero, F. G. Benitez, and J. M. del Castillo, “Traffic flow estimation models using cellular phone data,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1430–1441, Sep. 2012.

[12] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya, “Detecting errors and imputing missing data for single-loop surveillance systems,” *Transp. Res. Record, J. Transp. Res. Board*, vol. 1855, no. 1, pp. 160–167, 2003. doi: 10.3141/1855-20.

[13] 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.

[14] L. Li, Y. Li, and Z. Li, “Efficient missing data imputing for traffic flow by considering temporal and spatial dependence,” *Transp. Res. C, Emerg. Technol.*, vol. 34, pp. 108–120, Sep. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X13001095

[15] S. Turner, L. Albert, B. Gajewski, and W. Eisele, “Archived intelligent transportation system data quality: Preliminary analyses of san antonio transguide data,” *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1719, pp. 77–84, 2000. doi: 10.3141/1719-10.

[16] J. van Lint, S. Hoogendoorn, and H. J. van Zuylen, “Accurate freeway travel time prediction with state-space neural networks under missing data,” *Transp. Res. C, Emerg. Technol.*, vol. 13, pp. 347–369, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X0500029X

[17] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, “Short-term traffic forecasting: Overview of objectives and methods,” *Transp. Rev.*, vol. 24, no. 5, pp. 533–557, 2004. doi: 10.1080/0144164042000195072.

[18] M. Karlaftis and E. L. Vlahogianni, “Statistical methods versus neural networks in transportation research: Differences, similarities and some insights,” *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X10001610

[19] L. Beretta and A. Santaniello, “Nearest neighbor imputation algorithms: A critical evaluation,” *BMC Med. Inform. Decis. Making*, vol. 16, no. 3, p. 74, Jul. 2016. doi: 10.1186/s12911-016-0318-z.

[20] J. Chen and J. Shao, “Nearest neighbor imputation for survey data,” *J. Statist.*, vol. 16, no. 2, pp. 113–132, 2000.

[21] D. Ni, J. D. Leonard, A. Guin, and C. Feng, “Multiple imputation scheme for overcoming the missing values and variability issues in ITS data,” *J. Transp. Eng.*, vol. 131, no. 12, pp. 931–938, 2005. doi: 10.1061/(ASCE)0733-947X(2005)131:12(931).

[22] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, “A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting,” *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 21–34, Jan. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X15003812

[23] S. Tak, S. Woo, and H. Yeo, “Data-driven imputation method for traffic data in sectional units of road links,” *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1762–1771, Jun. 2016.

[24] G. Chang, Y. Zhang, and D. Yao, “Missing data imputation for traffic flow based on improved local least squares,” *Tsinghua Sci. Technol.*, vol. 17, no. 3, pp. 304–309, Jun. 2012.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[25] H. Dong, L. Jia, X. Sun, C. Li, and Y. Qin, "Road traffic flow prediction with a time-oriented ARIMA model," in *Proc. 15th Int. Joint Conf. INC, IMS IDC*, Aug. 2009, pp. 1649–1652.

[26] M. Zhong, P. Lingras, and S. Sharma, "Estimation of missing traffic counts using factor, genetic, neural, and regression techniques," *Transp. Res. C, Emerg. Technol.*, vol. 12, no. 2, pp. 139–166, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X04000191

[27] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, 2003. [Online]. Available: http://ascelibrary.org/doi/abs/10.1061/%28ASCE%290733-947X%282003%29129%3A6%28664%29

[28] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches," *Transp. Res. Record, J. Transp. Res.*, vol. 1857, pp. 74–84, Jan. 2003. doi: 10.3141/1857-09.

[29] Z. Yan, "Traj-arima: A spatial-time series model for network-constrained trajectory," in *Proc. 2nd Int. Workshop Comput. Transp. Sci. (IWCTS)*, 2010, pp. 11–16. doi: 10.1145/1899441.1899446.

[30] 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," *Transp. Res. C, Emerg. Technol.*, vol. 13, no. 3, pp. 211–234, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X05000276

[31] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transp. Res. C, Emerg. Technol.*, vol. 28, pp. 15–27, Mar. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X12001532

[32] M. T. Asif, N. Mitrovic, J. Dauwels, and P. Jaillet, "Matrix and tensor based methods for missing data estimation in large traffic networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 1816–1825, Jul. 2016.

[33] B. Ran, H. Tan, Y. Wu, and P. J. Jin, "Tensor based missing traffic data completion with spatial temporal correlation," *Phys. A, Stat. Mech. Appl.*, vol. 446, pp. 54–63, Mar. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437115008870

[34] L. Garg, J. Dauwels, A. Earnest, and K. P. Leong, "Tensor-based methods for handling missing data in quality-of-life questionnaires," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1571–1580, Sep. 2014.

[35] A. Klein and G. Mélard, "Computation of the Fisher information matrix for time series models," *J. Comput. Appl. Math.*, vol. 64, no. 1, pp. 57–68, 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0377042795000062

[36] Y. Shen and M. Z. Win, "Fundamental limits of wideband localization—Part II: Cooperative networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4956–4980, Oct. 2010.

[37] Y. Shen, H. Wymeersch, and M. Z. Win, "Fundamental limits of wideband localization—Part II: Cooperative networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4981–5000, Oct. 2010.

[38] K. Yu and E. Dutkiewicz, "CRLB derivation for mobile tracking in NLOS propagation environments," in *Proc. Int. Symp. Commun. Inf. Technol. (ISCIT)*, Oct. 2012, pp. 1148–1153.

[39] N. Cressie, T. Shi, and E. L. Kang, "Fixed rank filtering for spatio-temporal data," *J. Comput. Graph. Statist.*, vol. 19, no. 3, pp. 724–745, 2010. doi: 10.1198/jcgs.2010.09051.

[40] A. Ermagun, S. Chatterjee, and D. Levinson, "Using temporal detrending to observe the spatial correlation of traffic," *PLoS ONE*, vol. 12, no. 5, 2017, Art. no. e0176853. doi: 10.1371/journal.pone.0176853.

[41] S. Yang, S. Shi, X. Hu, and M. Wang, "Spatiotemporal context awareness for urban traffic modeling and prediction: Sparse representation based variable selection," *PLoS ONE*, vol. 10, no. 10, 2015, Art. no. e0141223. doi: 10.1371/journal.pone.0141223.

[42] N. Cressie and C. Wikle, *Statist.for Spatio-Temporal Data* (CourseSmart Series). Hoboken, NJ, USA: Wiley, 2011. [Online]. Available: https://books.google.com.au/books?id=-kOC6D0DiNYC

[43] H. Van Trees, K. Bell, and Z. Tian, *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*. Hoboken, NJ, USA: Wiley, 2013. [Online]. Available: https://books.google.com.au/books?id=dnvaxqHDkbQC

[44] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.

[45] S. M. Kay, *Fundamentals Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[46] J. Conklin, "Data imputation strategies for transportation management systems," Center Transp. Stud., Univ. Virginia, Charlottesville, VA, USA, Res. Rep. UVACTS-13-0-80, 2003. [Online]. Available: https://books.google.com.au/books?id=yTsJIAAACAAJ

Shangbo Wang (S'17–M'18) received the master's and Dr.Ing. degrees from the University of Duisburg-Essen, Germany, in 2007 and 2014, respectively. He was with Siemens AG, Munich, Germany, and Continental AG, Lindau, Germany, as a Research Engineer and a Development Engineer, respectively. In 2015, he joined the University of Technology Sydney as a Research Assistant. His research interests include statistical signal processing, intelligent transportation systems, digital signal processing, wireless localization, and mobile communication.

Guoqiang Mao (S'98–M'02–SM'07–F'18) was with the School of Electrical and Information Engineering, The University of Sydney. The Center is among the largest university research centers in Australia in wireless communications and networking. In 2014, he joined the University of Technology Sydney as a Professor of wireless networking and the Director of the Center for Real-time Information Networks. He has authored or co-authored about 200 papers in international conferences and journals, which have been cited more than 4500 times. His research interests include intelligent transport systems, applied graph theory and its applications in telecommunications, the Internet of Things, wireless sensor networks, wireless localization techniques, and network performance analysis. He was a recipient of the Top Editor Award for outstanding contributions to the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, in 2011, 2014, and 2015. He has served as a chair, a co-chair, and a TPC member for a large number of international conferences. He is a Co-Chair of the IEEE Intelligent Transport Systems Society Technical Committee on Communication Networks. He has been an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS since 2014 and has been the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY since 2010.