White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/147767/

Version: Published Version

Article:
Offor, K., Vaci, L. and Mihaylova, L. (2019) Traffic estimation for large urban road network with high missing data ratio. Sensors, 19 (12). 2813. ISSN: 1424-8220

https://doi.org/10.3390/s19122813

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

sensors

MDPI

Article

Traffic Estimation for Large Urban Road Network with High Missing Data Ratio

Kennedy John Offor [1,*], Lubos Vaci [2] and Lyudmila S. Mihaylova [1]

[1] Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK; l.s.mihaylova@sheffield.ac.uk
[2] Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK; l.vaci@sheffield.ac.uk
* Correspondence: kjoffor1@sheffield.ac.uk

Received: 29 May 2019; Accepted: 20 June 2019; Published: 24 June 2019

Abstract: Intelligent transportation systems require the knowledge of current and forecasted traffic states for effective control of road networks. The actual traffic state has to be estimated as the existing sensors does not capture the needed state. Sensor measurements often contain missing or incomplete data as a result of communication issues, faulty sensors or cost leading to incomplete monitoring of the entire road network. This missing data poses challenges to traffic estimation approaches. In this work, a robust spatio-temporal traffic imputation approach capable of withstanding high missing data rate is presented. A particle based approach with Kriging interpolation is proposed. The performance of the particle based Kriging interpolation for different missing data ratios was investigated for a large road network comprising 1000 segments. Results indicate that the effect of missing data in a large road network can be mitigated by the Kriging interpolation within the particle filter framework.

Keywords: particle filtering; road traffic; state estimation; Bayesian inference; Kriging; missing data imputation

1. Introduction

Intelligent Transportation Systems (ITS) require accurate knowledge of traffic states for effective traffic monitoring and control. Traffic states are usually estimated from noisy sensor measurements [1] using various approaches, which can typically be subdivided into model based approaches, data driven approaches or streaming-data-driven [2]. An overview of the different modelling methodologies is given in Refs. [3,4]. These modelling methodologies include microscopic, macroscopic and mesoscopic approaches. Microscopic traffic models [3–6] describe the motion of each individual vehicle with a high level of detail. Hao et al. [7,8] recently proposed a model predictive control (MPC) termed urban cell transmission model (UCTM) for optimal switching of traffic lights at intersections based on the predicted traffic.

In macroscopic models [9,10], traffic state is represented by aggregating behaviour of the traffic, usually in terms of the average speed and the average density over a given period of time. Mesoscopic models [3] employ some features of microscopic and macroscopic approaches by utilising varying levels/degrees of detail to model traffic behaviour. This is achieved by modelling some locations with aggregated measurements as in macroscopic and the remaining locations are modelled down to the details of individual vehicles as is done in the case of microscopic.

Macroscopic models are sufficient to produce acceptable estimation accuracy when compared to the computational overhead of the microscopic models. Hence, they are the preferred choice for most practical purposes such as traffic control/management, road pricing and changes in infrastructure.

Most traffic estimation approaches are model based [2], while the new trend is to develop data driven approaches [11].

Data driven methods rely on historical data or streaming/real-time data. Within the last decade, there has been growing interest in applying Kriging to traffic state prediction; for directional traffic volume using global position system (GPS) data [12], annual average traffic count interpolation using origin-destination data [13], estimating annual average daily traffic [14,15], traffic volume prediction [16,17], and traffic volume imputation [18].

Kriging is one of the most flexible data driven methods originally employed in geo-statistics or spatio-temporal analysis. Kriging exploits the spatial dependence (either covariance or variogram) using a weighted sum of observed data points to interpolate the values at locations of interest. It was initially used for copper mining by Krige [19] and was developed further by Matheron [20]. Since then, it has been applied in other fields such as spatial analysis and computer experiments. In the recent times, the method has been applied in traffic prediction [12].

One of the major challenges faced in traffic prediction is the issue of missing or sparse data. Traffic measurements are generally captured with different types of sensors and transmitted through communication infrastructure for processing and utilisation. These infrastructures are subject to failure and malfunction occasionally leading to incomplete/missing data, sometimes more than 40% [21]. The problem of sparse data is caused by the high cost of installing and managing traffic measurement devices making them impractical to cover all locations needed for effective observation of the full road network.

To address these challenges, researchers resorted to various methods and approaches such as missing data imputation [22], compressive sensing and historical averages [23], and Kriging interpolation [16]. In Ref. [24], a particle filter (PF) with the stochastic compositional model (SCM) is developed to estimate traffic states in motorways. Boundary measurements (inflow and outflow) were used to estimate the traffic state within the segments. The study reported that missing boundary measurements affected the estimated accuracy. A solution to the problem of missing data in particle filter measurement update was proposed in our previous work [25]. Here Kriging methods were used to interpolate the missing data which is subsequently used for the computation of the PF likelihood for traffic state estimation.

This work extends the approach of [25] for a larger road network of several kilometres with 1000 segments under the influence of missing data and/or sensor failure. The task of training such large segments would be resource intensive, hence the reduced measurement space, proposed in Ref. [26], was used to select the most influential segments in the road network. The drawback of Ref. [26] is that the most influential segments are selected based on only available measurements without considering any missing values. This would mean that if some interconnected segments have missing data, they would not be used leading to information loss. To address this, we propose a new approach, that combines the benefits of both approaches ([25,26]) in a single framework. First, the most probable segments are selected using the column based matrix decomposition (CBMD). Second, when there are missing measurements in the segments, Kriging is used to estimate the measurements before the measurement update step (see Section 5.3).

The present work differs from Refs. [12,16] in the following ways. Whereas [12,16] employed Kriging for predicting vehicular speed, our approach used Kriging to estimate missing vehicular flow and speed measurements which is then applied to the computation of measurement update step for particle filter estimation. In addition, our approach combines speed and vehicular flow estimation unlike the former which predicted only speed. The approaches used in Refs. [12,16] are similar, the only difference being that the former proposes the use of alternate distance metric called Approximate Road Distance Network (ARDN) to replace the Euclidean distance metric.

The contribution of this work is twofold: (i) A robust spatio-temporal traffic imputation that can withstand higher missing data proportion via information exchange between correlated segments. (ii) the use of Kriging and particle filter to address the effects of high missing/sparse data. A Column

based matrix decomposition is used to select most influential segments in large road network and then imputing any missing measurements in the selected segments using Kriging.

The rest of the paper is organised as follows. Section 2 discusses some related work followed by the presentation of traffic flow and measurement model used in this work in Section 3. A background theory of random sets, covariance and variograms, which is the building block of Kriging is presented in Section 4. Recursive Bayesian estimation and particle filters (PF) are presented in Section 5. The proposed method was discussed in Section 5.3. Results and discussion are presented in Section 6.2 with conclusions being drawn in Section 7.

2. Related Work

A review of three different missing data imputation methods was presented by Ref. [27]. These include interpolation, prediction and statistical learning. The interpolation method imputes missing measurement at the particular location by averaging all historical measurements at that location at similar times of day. Prediction methods use a deterministic mathematical description to model the relationship between historical and future data. The statistical methods on the other hand treats the traffic as a random variable and tries to capture stochastic nature of the traffic pattern into the imputation algorithm.

In Refs. [28–30], a multi-resolution approximation using linear combinations of basis functions was proposed to address computational complexity of large datasets. The novelty in Ref. [28] lies in the use of multiple basis functions computed at lower resolutions closer to observation locations and then combining them to capture the different covariance functions with varying properties. The solution is achieved by dividing the spatial domain recursively into small regions and smaller sub-regions until the fine-scale dependencies are captured.

Nychka et al. [28] used radial basis functions (RBF) and special type of Gaussian Markov random field (GMRF) called spatial autoregressive (SAR) model to model the spatial correlation among the coefficients while Katzfuss et al. [30] automatically determines the appropriate basis/covariance function. Whereas in Ref. [29], it was assumed that the sub-domains are independent, Ref. [30] assumed depended sub-domains and performed full-scale approximation. As the computations are done locally in parallel, it is possible to fuse multisensor data sources, in which case, the different covariance functions are used for each data source. While both mentioned that the approach could be extended to non-stationary functions, it was not implemented nor was there a derivation for such. The different methods on missing traffic data imputation considered small road network in the range of a few kilometres.

3. Model Formulation

Consider a general discrete time state space system of a form,

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{v}_k \quad \text{or} \quad p(\mathbf{x}_{k+1}|\mathbf{x}_k), \tag{1}$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{w}_k \quad \text{or} \quad p(\mathbf{z}_k|\mathbf{x}_k), \tag{2}$$

where $\mathbf{f}_k$ is a nonlinear function of the target state vector $\mathbf{x}_k$, $\mathbf{h}_k$ represents a nonlinear relationship between sensor output $\mathbf{z}_k$ and target state vector $\mathbf{x}_k$ affected by a measurement noise $\mathbf{w}_k$. In addition, $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ is the probability density function of the new state $\mathbf{x}_{k+1}$ given the previous state $\mathbf{x}_k$, and $p(\mathbf{z}_k|\mathbf{x}_k)$ is the likelihood function of the measurement $\mathbf{z}_k$ given the state $\mathbf{x}_k$.

3.1. Stochastic Compositional Traffic Flow Model

In this work, a stochastic compositional model (SCM) [31] is considered for modelling of a motorway/freeway vehicle traffic evolution expressed as $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ in Equation (1). This extended cell-transition model incorporates traffic speed and uses forward and backward waves to describe the complex relationship between traffic behaviour especially for a large road network. SCM utilises

sending and receiving functions which models the stochastic nature of traffic state evolution. The vehicles that are able to leave a cell are represented by receiving functions while those that are allowed to enter a cell are determined by the receiving functions. The receiving functions are usually less than or equal to the sending functions to obey the law of conservation of vehicles.

In the SCM, the road network is divided into a given number of cells, n, also called segments. Each segment has a length L_i and number of lanes l_i as shown in Figure 1, where $i = (1, \ldots, n)$. At any given time period k, a certain number of vehicles $Q_{i,k}$ crosses the boundary between two segments i and $i+1$. The number of vehicles in a given cell i within the same time period k is represented by $N_{i,k}$ with their average speed given by $v_{i,k}$.

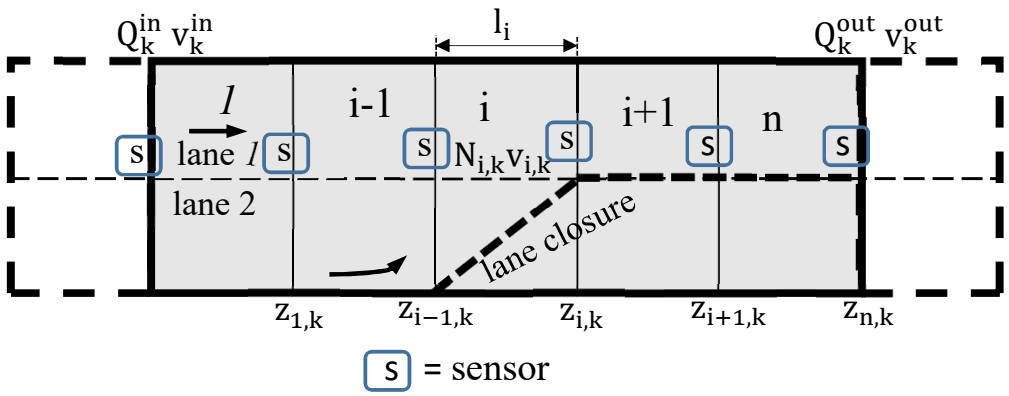

Figure 1. Stochastic compositional model (SCM) road network showing segments and measurement points [31].

The overall state vector at time k is given by $\mathbf{x}_k = [\mathbf{x}_{1,k}^T, \mathbf{x}_{2,k}^T, \ldots, \mathbf{x}_{n,k}^T]^T$ where $\mathbf{x}_{i,k} = [N_{i,k}, v_{i,k}]^T$ is the local state vector at segment i. Equations (3)–(5) models traffic state evolution within the cells.

$$\mathbf{x}_{1,k+1} = f_1(Q_k^{in}, v_k^{in}, \mathbf{x}_{1,k}, \mathbf{x}_{2,k}, \boldsymbol{\eta}_{1,k}) \tag{3}$$

$$\mathbf{x}_{i,k+1} = f_i(\mathbf{x}_{i-1,k}, \mathbf{x}_{i,k}, \mathbf{x}_{i+1,k}, \boldsymbol{\eta}_{i,k}) \tag{4}$$

$$\mathbf{x}_{n,k+1} = f_n(\mathbf{x}_{n-1,k}, \mathbf{x}_{n,k}, Q_k^{out}, v_k^{out}, \boldsymbol{\eta}_{n,k}) \tag{5}$$

where Q_k^{in}, called inflow represents the number of vehicles entering the first cell with average speed v_k^{in}, Q_k^{out}, called outflow represents the number of vehicles leaving the last segment with average speed v_k^{out}. The inflow and outflow vehicles with their average speeds are called the boundary conditions and are supplied to the model at the beginning.

3.2. Measurement Model

Consider the road network shown in Figure 1 divided into different segments with n boundaries. The traffic state at a boundary $j \in \mathscr{J} = 1, 2, \ldots, n$ is sampled at discrete time steps t_s, $s = 1, 2, \ldots$, to give $\mathbf{z}_{j,s} = (Q_{j,s}, v_{j,s})^T$. The measurements at all the boundaries are collected into a matrix given by $\mathbf{Z}_s = (\mathbf{z}_{1,s}^T, \mathbf{z}_{2,s}^T, \ldots, \mathbf{z}_{n,s}^T)^T$. The relationship between the sampling interval Δt_s and the state update time step Δt_k Equation (1) is such that sampling interval is split into q state update time steps. That is, $\Delta t_s = q\Delta t_k$. The measurement model and noise is represented by Equation (6) with error $\boldsymbol{\xi}_s = (\xi_{Q_{j,s}}, \xi_{v_{j,s}})^T$. With the assumption of Gaussian noise, $\mathbf{z}_{j,s}$ can be expressed as:

$$\mathbf{z}_{j,s} = \begin{pmatrix} Q_{j,s} \\ \\ v_{j,s} \end{pmatrix} + \boldsymbol{\xi}_s. \tag{6}$$

4. Random Sets, Covariances and Variograms

4.1. Random Set

Let $\mathbf{z}(\mathbf{r})$ be a non-stationary set of m measurements $\mathbf{z}(\mathbf{r}) = [z(\mathbf{r}_1), z(\mathbf{r}_2), \ldots, z(\mathbf{r}_n)]^T$, in our case average vehicle speeds $\mathbf{v}(\mathbf{r})$ or vehicle counts $\mathbf{N}(\mathbf{r})$, observed in segments $i = 1, \ldots, n$ at locations $\mathbf{r}_i = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n)$, where $\mathbf{r}_i = (r_i^x, r_i^z)^T$. Note that location of each sensor is uniquely defined by the segment topology (Figure 1). Then, according to Equation (7),

$$\mathbf{z}(\mathbf{r}) \sim \mathcal{N}\big(\boldsymbol{\mu}(\mathbf{r}), \mathbf{K}(\mathbf{r})\big) \tag{7}$$

the random set of variables $\mathbf{z}(\mathbf{r})$ can be approximated by a Gaussian process which is uniquely defined by the drift, i.e., the mean field $\boldsymbol{\mu}(\mathbf{r}) = E[\mathbf{z}(\mathbf{r})]$, and the corresponding covariance $\mathbf{K}(\mathbf{r}^T, \mathbf{r}) = \big(\mathbf{z}(\mathbf{r}) - \boldsymbol{\mu}(\mathbf{r})\big)^T \big(\mathbf{z}(\mathbf{r}) - \boldsymbol{\mu}(\mathbf{r})\big)$. Individual samples $z(\mathbf{r}_i)$ of the set Equation (7) can be also expressed as

$$z(\mathbf{r}_i) = \mu(\mathbf{r}_i) + \delta(\mathbf{r}_i) \tag{8}$$

with drift $\mu(\mathbf{r}_i)$ and residual $\delta(\mathbf{r}_i)$. Based on the degree to which the moments, in our case mean and variance of the random set $\mathbf{z}(\mathbf{r})$ are dependent (or independent) on a spatial relationship between points $\mathbf{r}_i = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n)$, the 1st or the 2nd-order of stationarity can be recognised. While, the 1st-order stationary random set assumes a constant mean $E[\mathbf{z}(\mathbf{r})] = \boldsymbol{\mu}(\mathbf{r}) = \mu$, the 2nd-order stationary random field assumes a linear drift between the increments $E[\mathbf{z}(\mathbf{r})] = \boldsymbol{\mu}(\mathbf{r}) = \sum_{n=0}^{N} \alpha_n \mathbf{f}(\mathbf{r})$. In the rest of this sequel, we assume that the random set $\mathbf{z}(\mathbf{r})$ is *intrinsic* and stationary on the 2nd-order.

4.2. Covariance

By assuming that random variables $\mathbf{z}(\mathbf{r})$ are 2nd-order stationary and isotropic, the covariance function $\mathbf{K}(\mathbf{r}_i, \mathbf{r}_j)$ in Equation (7) reads as follows:

$$\begin{aligned} \mathbf{K}(\mathbf{r}_i, \mathbf{r}_j) &= E\big[(\mathbf{z}(\mathbf{r}_i) - \mu)^T (\mathbf{z}(\mathbf{r}_j) - \mu)\big], \\ \mathbf{K}(\mathbf{r}_i, \mathbf{r}_j) &= E\big[(\mathbf{z}(\mathbf{r}_i) - \mu)^T (\mathbf{z}(\mathbf{r}_i + \mathbf{h}) - \mu)\big]. \end{aligned} \tag{9}$$

The role of covariance function $\mathbf{K}(\mathbf{r}_i, \mathbf{r}_j)$ is to model the correlation between measurements $z(\mathbf{r}_i)$ and $z(\mathbf{r}_j)$ observed at locations $\mathbf{r}_i$ and $\mathbf{r}_j$ based on their separation distance called a *lag* h. Since the correlation between two random variables solely depends on their spatial distance, and not at all on their location, the *lag* can be conveniently expressed as an Euclidean l_2 norm, defined as $h = ||\mathbf{r}_j - \mathbf{r}_i||_2$. Above statements can be summarised into the following equalities, i.e., the isotropy assumptions:

$$\begin{aligned} \mathbf{h} = \mathbf{r}_j - \mathbf{r}_i \implies ||\mathbf{h}||_2 = h \implies \\ \mathbf{K}(\mathbf{r}_i, \mathbf{r}_j) = \mathbf{K}(\mathbf{r}_i, \mathbf{r}_i + \mathbf{h}) = \mathbf{K}(h). \end{aligned} \tag{10}$$

For a straight stretch of motorway (Figure 1) this is equivalent to the path length through the road network.

The process of covariance function $\mathbf{K}(\mathbf{r}_i, \mathbf{r}_j)$ modelling requires to find a covariance curve that has the best fit to the empirical data Equation (9), possibly being a subject to constraints. The covariance models to choose from include exponential, spherical, Gaussian, linear or power model [14]. In this work, the best fit for the dataset was achieved by the exponential model depicted in Figure 2a and given by

$$\mathbf{K}(\mathbf{r}_i, \mathbf{r}_j) = \begin{cases} c - a & if |\mathbf{r}_j - \mathbf{r}_i| = 0, \\ (c - a) e^{-\frac{3|\mathbf{r}_j - \mathbf{r}_i|}{b}} \} & if |\mathbf{r}_j - \mathbf{r}_i| > 0, \end{cases} \tag{11}$$

where a is the nugget, b is the range and c is still.

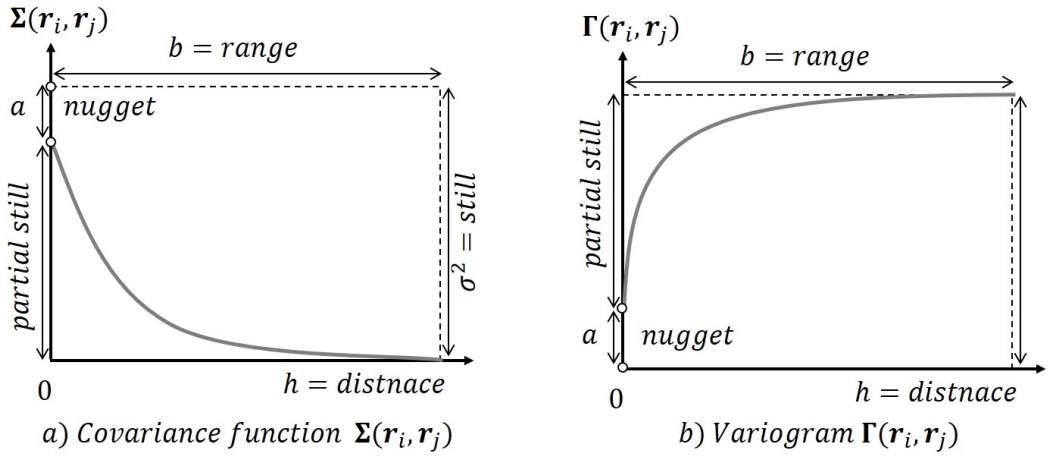

Figure 2. Covariance and variogram models.

As can be observed in Figure 2a, observe that the covariance function decreases with distance, so it can be thought of as a similarity function.

5. Recursive Bayesian Estimation

5.1. Bayesian Estimation

Consider a general discrete time state space system represented as in Equations (1) and (2). The goal of Bayesian estimation is to infer the state variable $\mathbf{x}_k$ as defined in Section 3.1 with the available sensor measurements $\mathbf{z}_{1:k}$. By using the Bayesian framework, this estimation problem relates to the recursive evaluation of the probability density function (PDF) $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ in two consecutive steps, the prediction and the measurement update of the state vectors.

$$p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) \xrightarrow[\text{Update}]{\text{Prediction}} p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \tag{12}$$

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \xrightarrow[\text{Update}]{\text{Measurement}} p(\mathbf{x}_k|\mathbf{z}_{1:k}) \tag{13}$$

The prediction state density $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ of state $\mathbf{x}_k$ is calculated from the prior PDF $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ by using Chapman-Kolmogorov equation

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1}. \tag{14}$$

Equation (14) follows the 1st order Markov property which assumes that $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ only depends on state $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$ at time k and $k-1$ respectively. The measurement update $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ is computed from the prior distribution of Equation (14) and measurements $\mathbf{z}_{1:k}$ by a Bayesian formula which results in

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{15}$$

The 1st order Markov property for Equation (15) implies that $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ only depends on measurement $\mathbf{z}_k$ at time k.

5.2. Particle Filter

Arguably, the most popular algorithm for nonlinear recursive estimation is the particle filter (PF), extensively evaluated in Ref. [32]. PF represents any arbitrary probability density function $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ by samples or particles x_k^l, where $l = 1, \ldots, N_p$ is the number of particles, i.e.,

$$x_k^l \approx p(\mathbf{x}_k|\mathbf{z}_{1:k}), \tag{16}$$

The particles are used to form an approximative distribution as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \hat{p}(\mathbf{x}_k|\mathbf{z}_{1:k}) = \sum_{i=1}^{N} w_{k|k}^l \delta(\mathbf{x}_k - x_k^l), \tag{17}$$

where $\hat{p}(\mathbf{x}_k|\mathbf{z}_{1:k})$ is an approximated distribution, $\delta(\mathbf{x}_k - x_k^l)$ is a the Dirac delta function and $w_{k|k}^l$ the weights of the particles satisfying $\sum_{l=1}^{N_p} w_{k|k} = 1$. The time update of the Bayesian recursion Equation (1) is in case of PF evaluated as

$$\begin{aligned} &p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \\ &\approx \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) \sum_{l=1}^{N} w_{k-1|k-1}^l \delta(\mathbf{x}_{k-1} - x_{k-1}^l) \mathbf{d}\mathbf{x}_{k-1}, \\ &\approx \sum_{l=1}^{N} w_{k-1|k-1}^l \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) \delta(\mathbf{x}_{k-1} - x_{k-1}^l) \mathbf{d}\mathbf{x}_{k-1}, \\ &\approx \sum_{l=1}^{N} w_{k-1|k-1}^l p(\mathbf{x}_k|\mathbf{x}_{k-1}). \end{aligned} \tag{18}$$

The particles x_{k-1}^l in above Equation (18) are sampled from proposal distribution $\pi(\mathbf{x}_k|x_{k-1}^l)$, i.e., $x_{k-1}^l \approx \pi(\mathbf{x}_k|x_{k-1}^l)$. Proposal distribution is very often defined by the state transition PDF, that is $\pi(\mathbf{x}_k|x_{k-1}^l) = p(\mathbf{x}_k|x_{k-1}^i)$. In this case, the weights updates result to

$$\begin{aligned} w_{k|k-1}^l &= \frac{p(\mathbf{x}_k|x_{k-1}^l)}{\pi(\mathbf{x}_k|x_{k-1}^l, z_k)} w_{k-1|k-1}^l, \\ &= \frac{p(\mathbf{x}_k|x_{k-1}^l)}{p(\mathbf{x}_k|x_{k-1}^l)} w_{k-1|k-1}^l = w_{k-1|k-1}^l. \end{aligned} \tag{19}$$

The measurement update $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ (2) is computed by a Equation (15), which can be in terms of the particles x_k^l represented as

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{z}_{1:k}) &\propto p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_{1:k-1}), \\ &\approx \sum_{l=1}^{N_p} w_{k|k-1}^l p(\mathbf{z}_k|\mathbf{x}_k) \delta(\mathbf{x}_k - x_k^l). \end{aligned} \tag{20}$$

Similarly, the particle filter weights are updated as

$$w_{k|k}^l = \frac{w_{k|k-1}^l p(\mathbf{z}_k|x_k^l)}{\sum_{l'=1}^{N_p} w_{k|k-1}^{l'} p(\mathbf{z}_k|x_k^{l'})}. \tag{21}$$

The denominator in Equations (15) and (21) is only a normalising factor independent of $\mathbf{x}_k$ thus can be safety omitted if the distribution is numerically normed as shown by Equations (20) and (21).

$$\begin{aligned} w_{k|k}^{l} &\propto w_{k|k-1}^{l} p(\mathbf{z}_k|x_k^l), \\ &\approx w_{k|k-1}^{l} p(\mathbf{z}_k|x_k^l). \end{aligned} \tag{22}$$

The MC recursion tempt to degrade over time as all relative weights would tend to zero except for one that tends to one. Therefore, when particle depletion ratio reaches 0.5 a Sampling Importance Resampling (SIR) or Sampling Importance Sampling (SIS) techniques are applied in the recursion.

5.3. Missing Measurement Interpolation and Improved Likelihood Computation

For a large road network with many segments, using all the measurements in the particle filter measurement update step becomes computationally intensive. Column based matrix decomposition approach similar (as earlier stated in Section 1) to the work of [26] is employed to select the most probable segments that would give acceptable accuracy.

The idea is to select m most influential segments from all available segments n using CBMD and then estimating missing measurements (if any) of the most influential segments using Kriging for improved particle likelihood computation. Let $\mathbf{Z}_n \in \mathbb{R}^{k\times n}$ represent a set of all segments or measurement locations in a given time period, where the rows, k is number of time instances at which the measurements were taken and columns n number of road segments.

The goal of CBMD is to approximate the measuremtns $\mathbf{Z}_n$ with a subset of measurements $\mathbf{Z}_m \in \mathbb{R}^{k\times m}$ where $m < n$ is a subset of the measurements using singular value decomposition (SVD) as Ref. [33]:

$$\mathbf{Z}_m = \mathbf{Z}_n \mathbf{\Phi} \tag{23}$$

where $\mathbf{\Phi} \in \mathbb{R}^{n\times m}$ is the transformation matrix that expresses every column of all measurement $\mathbf{Z}_n$ in terms of the basis in $\mathbf{Z}_m$. Having computed the SVD, the right singular matrix is used to assign a probability P_{z_i} to each selected location according to:

$$P_{z_i} = \frac{1}{r}\sum_{j=1}^{r} v_{i,j}^2, i = 1, ..., n \tag{24}$$

where $v_{i,j}$ is the ith element of the jth right singular vector and r is the rank of the matrix. From the probabilities computed, m locations with the highest probabilities are chosen as the reduced measurement to approximate the entire network which is used in computing the particle filter likelihood during the measurement update step.

The likelihood function term $p(\mathbf{z}_k|\mathbf{x}_k)$ in Equation (21), is computed when a new measurement arrives. The performance of the PF degrades substantially when there is a missing measurement. For the multivariate Gaussian distribution, the PDF is given by:

$$p(\mathbf{z}_k|\mathbf{x}_k) = \frac{1}{\sqrt{2\pi}|\mathbf{R}|} e^{-0.5\boldsymbol{\nu}\mathbf{R}^{-1}\boldsymbol{\nu}^T}, \tag{25}$$

where $\mathbf{R}$ is the covariance matrix of the measurement data, $|\mathbf{R}| \equiv det(\mathbf{R})$ is the determinant of $\mathbf{R}$ and $\boldsymbol{\nu}$ is the difference between the PF predicted value ($\bar{\mathbf{z}}_s$) and measurement ($\mathbf{z}_s$), given by:

$$\boldsymbol{\nu} = \mathbf{z}_s - \bar{\mathbf{z}}_s. \tag{26}$$

The measurement matrix $\mathbf{z}_s$ can be expressed as,

$$\mathbf{z}_s = \begin{cases} \mathbf{z}_s^{meas} & \text{measurement from sensor;} \\ \hat{\mathbf{z}}_s^{krig} & \text{estimated by Kriging.} \end{cases} \tag{27}$$

where $\hat{\mathbf{z}}_s^{krig}$ represents the value estimated by Kriging (when measurement is not available). Note that the sampling time index t_s is split into q update time indices t_k as mentioned in Section 3.2. The measurement update state of the PF is performed only when $t_k \equiv t_s$. The modified particle method is presented in Algorithm 1.

Algorithm 1 Particle Filter for Traffic State Estimation with Kriging Estimated Measurements [24]

1. Road network approximation
   Use compressed sensing to select m most significant locations out of the n segments to be used for the measurement update step as defined in Section 5.3.
2. Initialisation
   At $k = 0$; define all boundary conditions: number of samples, weight of samples as below,
   For $l = 1, \ldots N_p$, N_p number of particles;
   - generate N_p samples $\{\mathbf{x}_0^{(l)}\}$ from the initial distribution $p(\mathbf{x}_0)$
   - initialise the particle weights $w_0^{(l)} = \frac{1}{N_p}$.

   End for
3. Start the iteration for $k = 1, 2, \ldots$
   (a) Prediction stage
   For $l = 1, ..., N_p$,
   sample $\mathbf{x}_k^{(l)} \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}^{(l)})$ according to SCM model equations
   End for
   (b) Measurement Update:
   This step is performed when the sampling time t_s equals the iteration count t_k as defined in Section 3.2
   i. Estimate missing measurements in the m most significant locations with Kriging using Equations (28) and (33)
   ii. Compute the likelihoods
   Based on Equation (6) compute the likelihood, $p(\mathbf{z}_s | \mathbf{x}_s^{(l)})$ of the particles using Equations (25)–(27)
   iii. Update the weights of the particles using the likelihood $p(\mathbf{z}_s | \mathbf{x}_s^{(i)})$ calculated from Equation (6)
   For $l = 1, ..., N_p$
   $\omega_s^{(l)} = \omega_{s-1}^{(l)} p(\mathbf{z}_s | \mathbf{x}_s^{(l)})$
   End For
   iv. Normalise the weights: $\hat{\omega}_s^{(l)} = \frac{\omega_s^{(l)}}{\sum_{l=1}^{N_p} \omega_s^{(l)}}$.
   (c) Update the predicted states (Output): $\hat{\mathbf{x}}_s = \sum_{l=1}^{N_p} \hat{\omega}_s^{(l)} \mathbf{x}_s^{(l)}$
   (d) Re-sample the weights (Selection) only when $t_k = t_s$

5.4. Missing Data Estimation via Kriging Models

The underlying idea of Kriging is to obtain the response $z(\mathbf{r}_u)$, interpreted as a random variable positioned at the location $\mathbf{r}_u$, by interpolating random variables $\mathbf{z}(\mathbf{r}) = [z(\mathbf{r}_1), z(\mathbf{r}_2), \ldots, z(\mathbf{r}_n)]^T$ from Equation (7), i.e., observations $z(\mathbf{r}_i)$, $i = 1, \ldots, n$ at locations $\mathbf{r}_i$. The Kriging predictor incorporates the covariance structure among the observation points $z(\mathbf{r}_i)$ into the weights $\mathbf{w}(\mathbf{r}_u) = [w_1(\mathbf{r}_u), w_2(\mathbf{r}_u), \ldots, w_n(\mathbf{r}_u)]^T$ for predicting $\hat{z}(\mathbf{r}_u)$ as a linear combination:

$$\hat{z}(\mathbf{r}_u) = \sum_{i=1}^{n} w_i(\mathbf{r}_u) z(\mathbf{r}_i) = \mathbf{w}^T(\mathbf{r}_u) \mathbf{z}(\mathbf{r}). \tag{28}$$

In order to assess the accuracy of the Kriging prediction $\hat{z}(\mathbf{r}_u)$ w.r.t the real (true) value $z(\mathbf{r}_u)$ an error $\epsilon(\mathbf{r}_u)$ is declared

$$\epsilon(\mathbf{r}_u) = z(\mathbf{r}_u) - \hat{z}(\mathbf{r}_u). \tag{29}$$

The following criteria, evaluated in terms of mean $E\left[\epsilon(\mathbf{r}_u)\right]$ and variance $\mathrm{Var}\left[\epsilon(\mathbf{r}_u)\right] = \sigma_\epsilon^2(\mathbf{r}_u)$ of the prediction error $\epsilon(\mathbf{r}_u)$ Equation (29), apply to any type of Kriging interpolation:

- Lack of bias, implies that $E\left[\epsilon(\mathbf{r}_u)\right] = 0$ and therefore $E\left[\hat{z}(\mathbf{r}_u)\right] = E\left[z(\mathbf{r}_u)\right]$. Thus, the Kriging interpolation is said to be the globally unbiased Equation (30).

$$E[\mathbf{z}(\mathbf{r_u}) - \hat{\mathbf{z}}(\mathbf{r_u})] = 0 \tag{30}$$

- Minimum variance, implies that the mean of the squared deviations $\mathrm{Var}[\epsilon(\mathbf{r}_u)] = \sigma_\epsilon^2(\mathbf{r}_u)$ must be the minimal Equation (31),

$$\min_{\mathbf{1}^T\mathbf{w}=1} \mathrm{Var}[z(\mathbf{r}_u) - \hat{z}(\mathbf{r}_u)] \quad \text{or} \quad \min_{\mathbf{1}^T\mathbf{w}=1} \sigma_\epsilon^2(\mathbf{r}_u), \tag{31}$$

subject to

$$\sum_{i=1}^{m} w_i = \mathbf{1}^T\mathbf{w} = 1 \tag{32}$$

Kriging weights $\mathbf{w}(\mathbf{r}_u)$ are chosen such that the mean squared prediction error $\sigma_\epsilon^2(\mathbf{r}_u)$ Equation (31), also known as Kriging variance or Kriging error is minimised as $\min_{\mathbf{1}^T\mathbf{w}=1} \sigma_\epsilon^2(\mathbf{r}_u)$ over all $\mathbf{z}(\mathbf{r_u})$ Equation (31) subject to the unbiased conditions of Equations (30) and (32) by a Lagrange multiplier 2λ to give,

$$\mathbf{w}(\mathbf{r}_u) = \mathbf{K}(\mathbf{r}_i, \mathbf{r}_j)^{-1}\mathbf{K}(\mathbf{r}_i, \mathbf{r}_u) - \mu\mathbf{1}. \tag{33}$$

where matrix $\mathbf{K}(\mathbf{r}_i, \mathbf{r}_j)$ is the covariance between the individual samples and column vector $\mathbf{k}(\mathbf{r}_i, \mathbf{r}_u)$ is a covariance function between samples and the point to interpolate.

6. Performance Evaluation

A road network with 1000 segments was simulated using SUMO software [34] to validate the proposed method. The segments are spaced 0.5 km apart and measurements (number of vehicles crossing each segment boundary with their average speed) taken every second. Traffic signs were installed at some locations to model the effect of congestion.

This is an extension of the previous work [25] where a smaller number segments was considered. The aggregate traffic flow and speed were sampled every 60 s and the results collected over a period of 10,800 s (3 h). Two types of vehicles, bus and passenger car was defined with the parameters as in Table 1.

Table 1. SUMO simulation parameters.

	Car	**Bus**
Max speed	25 m/s	20 m/s
Acceleration	1.0 m/s^2	0.8 m/s^2
Deceleration	4.5 m/s^2	4.5 m/s^2
Sigma (driver perfection)	0.5	0.5
Length	5 m	10 m
Minimum Separation	2.5 m	3 m

The vehicles were added randomly into the network through the inflow boundary every one second and they travel through the network until they get to the last boundary when they leave the network. As a vehicle crosses each induction loop, it is counted with its speed. The average speed of

the vehicles arriving at an induction loop over a period is recorded as the average speed. The entire statistics, flow, occupancy, and speed are collected in an output file for further processing.

6.1. Simulation Design

To simulate different scenarios such as congestion and free flow, (i) the number of lanes were decreased from 3 to 2, and (ii) the rate of vehicle injection into the network is varied at different time periods. Figures 3 and 4 show the spatio-temporal evolution of the traffic and their corresponding average speed, respectively, for a 100 segment section. The average speed of the vehicles varies around 100 km/h when the flow was around 2000 veh/h. Between time interval [1.5 h, 1.7 h], the flow was increased slightly to cause congestion, this resulted to a decrease in the average speed as can be seen in the first spike from Figure 4.

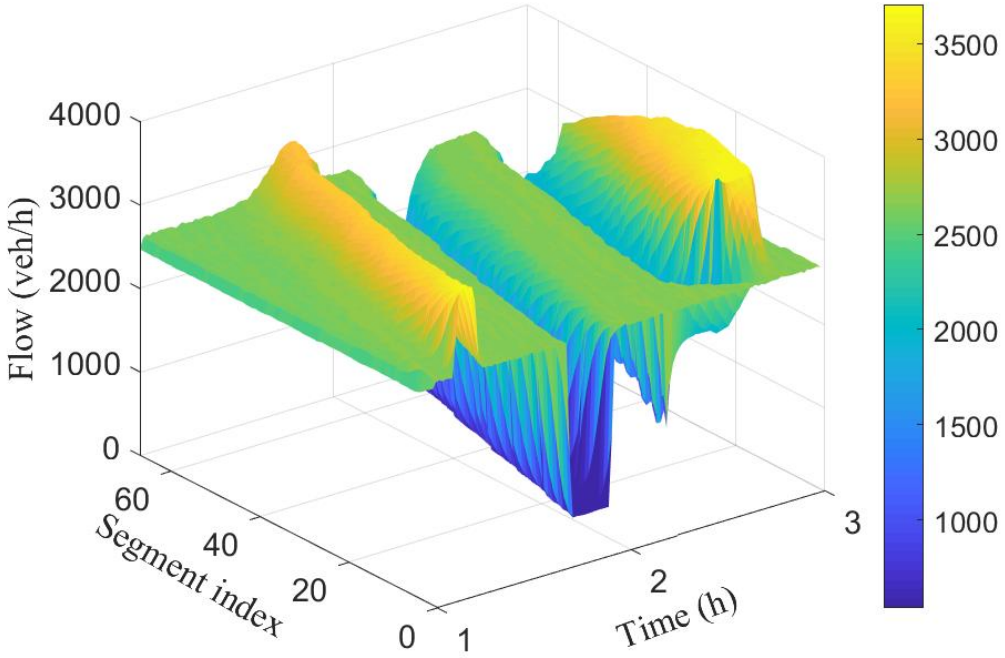

Figure 3. Spatio-temporal evolution of traffic flow for the 100 segment.

Observe that the effect was felt more close to the inflow boundary. The vehicles' speed increases marginally as they move into the network. Between time interval [1.6 h, 1.9 h] the flow was decreased leading to increase in the average speed. Finally, the number of lanes in segments 10 to 14 were reduced from 3 to 2 between time interval [2.4 h, 2.4 h] while maintaining vehicle injection rate. This results to substantive decrease in speed as can be seen (cone-shaped) in Figure 4. As the vehicles leave the segments with closed lanes, there is increase in their speed again.

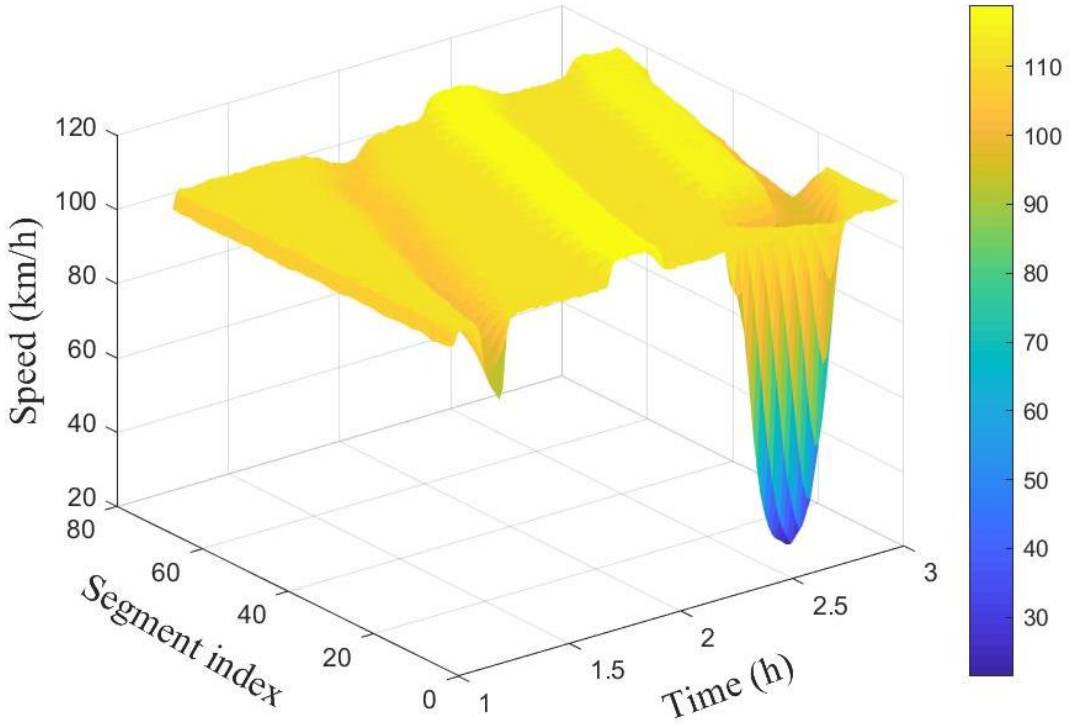

Figure 4. Spatio-temporal evolution of traffic speed for the 100 segment.

6.2. Results and Discussion

In order to test the prediction accuracy for different levels of sparsity, a statistical measure namely the root mean squared error (RMSE), was computed. Note, z_i is the ground truth or actual measurement, $\hat{z}_i$ is the estimated value and m_r is number of independent Monte Carlo runs. The measurements at some boundaries were randomly removed each time and then estimated using the proposed method. Different missing data rates (from 10, 20, ..., 70%) were investigated by randomly removing measurements at some locations using leave one out cross validation. This was repeated for 100 Monte Carlo runs and the average value used.

$$RMSE = \sqrt{\frac{1}{m_r}\sum_{i=1}^{m_r}[z_i - \hat{z}_i]^2}. \tag{34}$$

Figures 5 and 6 show the average estimation error for the different missing data ratios and the number of segments. It would be observed that the prediction error increases with the missing data rate. This is expected as less data is available for the computation. This effect could be reduced further by incorporating a mechanism known as multi resolution approach [30].

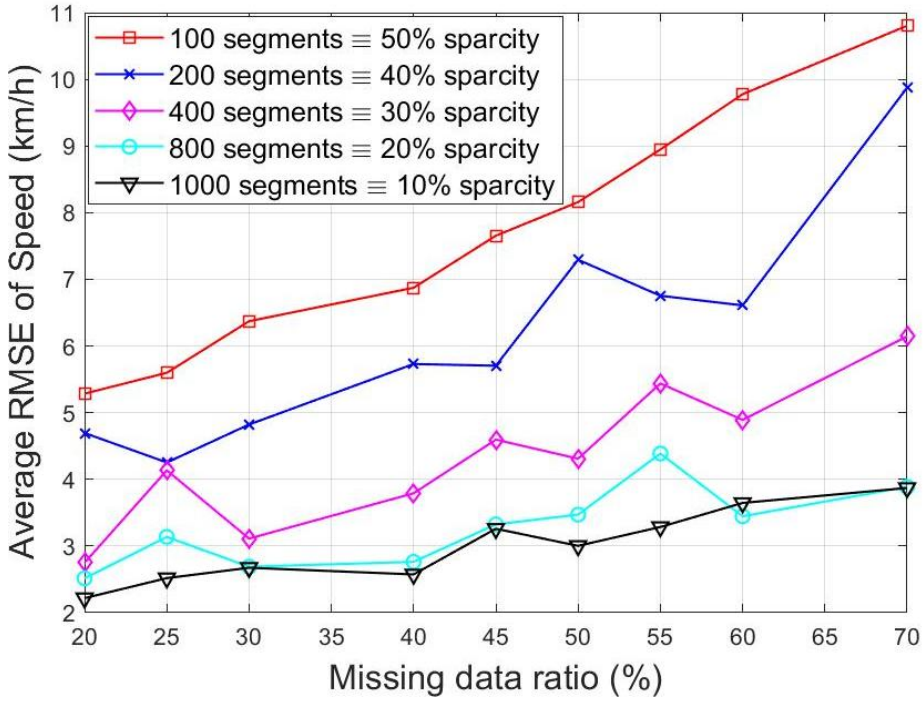

Figure 5. RMSE of speed at different missing data ratios.

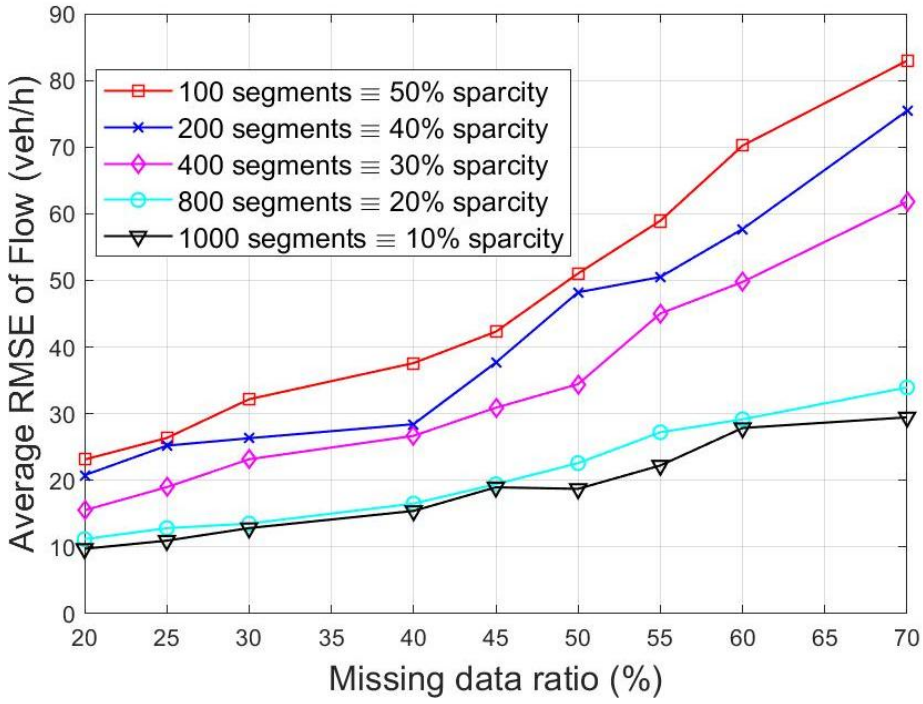

Figure 6. RMSE of flow at different missing data ratios.

The plots also show that the higher the number of segments used, the better the accuracy. This could be attributed to the fact that there is better information exchange within the network and hence, on average, more segments with available data are used for the higher dimensional segments scenario. This is in agreement with [35] where estimation accuracy in the presence of sparse sensor data was improved by exchanging particle weights between segments. For instance, a 10 segment network with 70% missing data ratio will result to using only 3 data points to estimate the remaining

7 missing locations. The chances of these 3 locations correlating with the other 7 is lower compared to when 30 out of 100 data points are available.

Figures 7 and 8 show the spatio-temporal evolution of the traffic flow and the corresponding average speed for the 100 segment scenario.

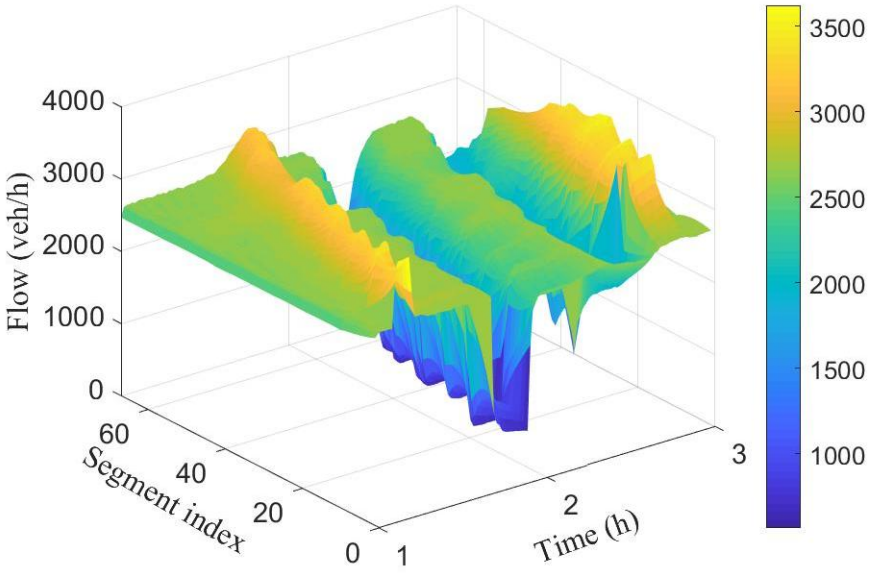

Figure 7. Estimated flow for the 100 segment with 30% missing data.

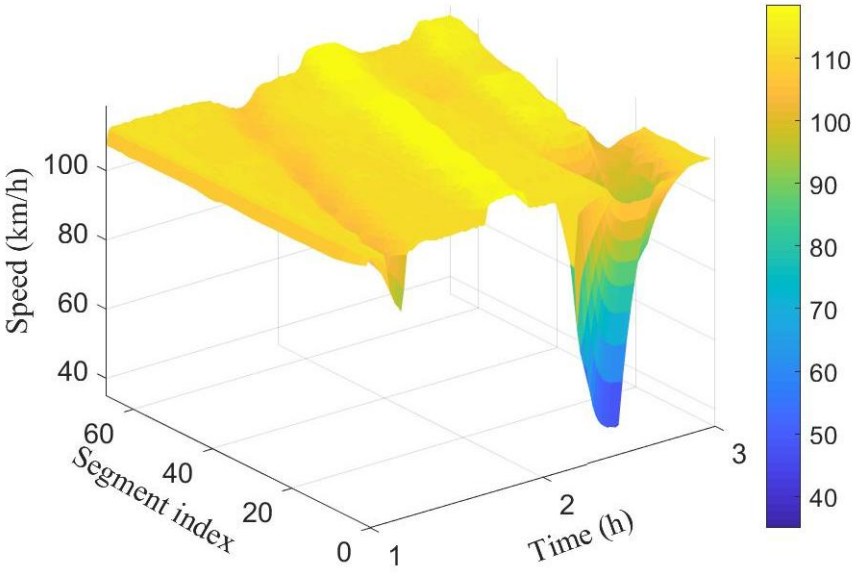

Figure 8. Estimated speed for the 100 segment with 30% missing data.

The number of vehicles crossing each boundary in space and their associated average speed is represented by the colour bar.Compared to the ground truth shown in Figures 3 and 4, it is evident that estimated number of vehicles crossing segment boundaries and the associated speeds have been estimated with a good accuracy. Observe also that the estimated flow and speed captured the periods where there is drop in number of vehicles and decrease in speed.

7. Conclusions

This paper presented a traffic estimation for a large road network with different missing data ratios. The computational overhead of the large network was addressed by using a method called

reduced measurement space proposed by [26] to select the most influential and information rich segments in the road network. These are subsequently used in the particle filter measurement update step. Missing data in the selected segments are imputed using Kriring. A 1000-segment road network was simulated using SUMO. Different missing data ratios ranging from 10% to 70% were tested for different sizes of road network ranging from 100 to 1000 segments.

The results indicate that considering a larger number of segments would reduce the overall estimation error even when the missing data ration is high. From the foregoing results and discussion, it is recommended that the best estimation accuracy would be obtained when the entire road network is considered at once. The effects of computational overhead could further be reduced by using a distributed approach with a central control unit.

Author Contributions: Ideas/conceptualisation, L.S.M., K.J.O.; methodology, L.S.M., K.J.O. and L.V.; software: particle filter/traffic models, L.S.M.; Kriging/sparsity, K.J.O.; implementation, K.J.O., L.S.M. and L.V.; validation, K.J.O., L.V.; formal analysis, K.J.O., L.S.M. and L.V.; investigation, K.J.O.; writing—original draft preparation, K.J.O.; writing—review and editing, L.V., L.S.M., K.J.O.; visualisation, K.J.O.

Funding: The authors acknowledge the support from the SETA project funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 688082 and the Tertiary Education Trust Fund (TETFund, Nigeria). The APC was funded by Department of Automatic Control and Systems Engineering, University of Sheffield.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Ferrara, A.; Sacone, S.; Siri, S. State estimation in freeway traffic systems. In *Freeway Traffic Modelling and Control. Advances in Industrial Control*; Springer International Publishing: Cham, Switzerland, 2018; pp. 169–190.
2. Seo, T.; Bayen, A.M.; Kusakabe, T.; Asakura, Y. Traffic state estimation on highway: A comprehensive survey. *Annu. Rev. Control* **2017**, *43*, 128–151. [CrossRef]
3. Hoogendoorn, S.P.; Bovy, P.H.L. Generic gas-kinetic traffic systems modeling with applications to vehicular traffic flow. *Transp. Res. Part B* **2001**, *35*, 317–336. [CrossRef]
4. Papageorgiou, M.; Ben-Akiva, M.; Bottom, J.; Bovy, P.H.L.; Hoogendoorn, S.P.; Hounsell, N.B.; Kotsialos, A.; Mcdonald, M. ITS and traffic management. In *Handbooks in Operations Research and Management Science*; Elsevier: Amsterdam, The Netrerlands, 2007; Chapter 11, Volume 14.
5. Pipes, L.A. An operational analysis of traffic dynamics. *J. Appl. Phys.* **1953**, *24*, 274–281. [CrossRef]
6. Gipps, P.G. A behavioural car-following model for computer simulation. *Transp. Res. Part B Methodol.* **1981**, *15*, 105–111. [CrossRef]
7. Hao, Z.; Boel, R.; Li, Z. Model based urban traffic control, part I: Local model and local model predictive controllers. *Transp. Res. Part C Emerg. Technol.* **2018**, *97*, 61–81. [CrossRef]
8. Hao, Z.; Boel, R.; Li, Z. Model based urban traffic control, part II: Local model and local model predictive controllers. *Transp. Res. Part C Emerg. Technol.* **2018**, *97*, 23–44. [CrossRef]
9. Messmer, A.; Papageorgiou, M. METANET: A macroscopic simulation program for motorway networks. *Traffic Eng. Control* **1990**, *31*, 466–470.
10. Daganzo, C.F. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transp. Res. Part B* **1994**, *28*, 269–287. [CrossRef]
11. Lu, H.-P.; Sun, Z.-Y.; Qu, W.-C. Big data-driven based real-time traffic flow state identification and prediction. *Discret. Dyn. Nat. Soc.* **2015**, *2015*, 1–11. [CrossRef]
12. Braxmeier, H.; Schmidt, V.; Spodarev, E. Kriged road-traffic maps. In *Interfacing Geostatistics and GIS*; Pilz, J., Ed.; Springer: Heidelberg, Germany, 2009; Chapter 9, pp. 105–119.
13. Lowry, M. Spatial interpolation of traffic counts based on origin-destination centrality. *J. Transp. Geogr.* **2014**, *36*, 98–105. [CrossRef]
14. Wang, X.; Kockelman, K.M.; Murray, W.J. Forecasting network data: Spatial interpolation of traffic counts from Texas data. *Transp. Res. Rec.* **2009**, *2105*, 100–108. [CrossRef]

15. Gastaldi, M.; Gecchele, G.; Rossi, R. Estimation of annual average daily traffic from one-week traffic counts. A combined ANN-Fuzzy approach. *Transp. Res. Part C Emerg. Technol.* **2014**, *47*, 86–99. [CrossRef]
16. Zou, H.; Yue, Y.; Li, Q.; Yeh, O.G.A. An improved distance metric for the interpolation of link-based traffic data using kriging: A case study of a large-scale urban road network. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 667–689. [CrossRef]
17. Song, Y.; Wang, X.; Wright, G.; Thatcher, D.; Wu, P.; Felix, P. Traffic volume prediction with segment-based regression Kriging and its implementation in assessing the impact of heavy vehicles. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20* 234–243. [CrossRef]
18. Yang, H.; Yang, J.; Han, L.D.; Liu, X.; Pu, L.; Chin, S.-M.; Hwang, H.-L. A Kriging based spatiotemporal approach for traffic volume data imputation. *PLoS ONE* **2018**, *13*, 1–11. [CrossRef]
19. Krige, D.G. A statistical approach to some basic mine problems on the Witwatersrand. *J. Chem. Metall. Soc. S. Min. Afr.* **1951**, *52*, 119–139.
20. Matheron, G. Principles of geostatistics. *Econ. Geol.* **1963**, *58*, 1246–1266. [CrossRef]
21. Zhang, Y. A comparative study of three multivariate short-term freeway traffic flow forecasting methods with missing data. *J. Intell. Transp. Syst.* **2016**, *20*, 205–218. [CrossRef]
22. Ladino, A.; Kibangou, Y.; Fourati, H.; Canudas de Wit, C. Travel time forecasting from clustered time series via optimal fusion strategy. In Proceedings of the 2016 European Control Conference (ECC), Aalborg, Denmark, 29 June–1 July 2016.
23. Hawes, M.; Amer, H.M.; Mihaylova, L. Traffic state estimation via a particle filter with compressive sensing and historical traffic data. In Proceedings of the 2016 19th International Conference on Information Fusion (FUSION), Heidelberg, Germany, 5–8 July 2016; pp. 735–742.
24. Mihaylova, L.; Boel, R.; Hegyi, A. Freeway traffic estimation within particle filtering framework. *Automatica* **2007**, *43*, 290–300. [CrossRef]
25. Offor, K.J.; Hawes, M.; Mihaylova, L.S. Short term traffic flow prediction with particle methods in the presence of sparse data. In Proceedings of the 2018 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 10–13 July 2018; pp. 1205–1212.
26. Hawes, M.; Amer, H.M.; Mihaylova, L. Traffic state estimation via a particle filter over a reduced measurement space. In Proceedings of the 2017 20th International Conference on Information Fusion (Fusion), Xi'an, China, 10–13 July 2017.
27. Li, Y.; Li, Z.; Li, L. Missing traffic data: Comparison of imputation methods. *IET Intell. Transp. Syst.* **2014**, *8*, 51–57. [CrossRef]
28. Nychka, D.; Bandyopadhyay, S.; Hammerling, D.; Lindgren, F. A multiresolution Gaussian process model for the analysis of large spatial datasets. *J. Comput. Graph. Stat.* **2015**, *24*, 579–599. [CrossRef]
29. Katzfuss, M.; Hammerling, D. Parallel inference for massive distributed spatial data using low-rank models. *Stat. Comput.* **2017**, *27*, 363–375. [CrossRef]
30. Katzfuss, M. A multi-resolution approximation for massive spatial datasets. *J. Am. Stat. Assoc.* **2017**, *112*, 201–214. [CrossRef]
31. Boel, R.; Mihaylova, L. A compositional stochastic model for real-time freeway traffic simulation. *Transp. Res. Part B Methodol.* **2006**, *40*, 319–334. [CrossRef]
32. Gustafsson, F. Particle filter theory and practice with positioning applications. *IEEE Aerosp. Electron. Syst. Mag.* **2010**, *25*, 53–82. [CrossRef]
33. Mitrovic, N.; Asif, M.T.; Dauwels, J.; Jaillet, P. Low-dimensional models for compressed sensing and prediction of large-scale traffic data. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2949–2954. [CrossRef]
34. Behrisch, M.; Bieker, L.; Erdmann, J.; Krajzewicz, D. SUMO—Simulation of urban mobility—An overview. In Proceedings of the SIMUL 2011, The Third International Conference on Advances in System Simulation, Barcelona, Spain, 23–28 Oktober 2011; pp. 63–68.
35. Mihaylova, L.; Hegyi, A.; Gning, A.; Boel, R.K. Parallelized particle and Gaussian sum particle filters for large-scale freeway traffic systems. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 36–48. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).