
Proceedings Paper:
Offor, K., Hawes, M. and Mihaylova, L.S. (2018) Short term traffic flow prediction with particle methods in the presence of sparse data. In: 2018 21st International Conference on Information Fusion (FUSION). 2018 21st International Conference on Information Fusion (FUSION), 10-13 Jul 2018, Cambridge, UK. IEEE, UK, pp. 1185-1192. ISBN: 978-0-9964527-6-2.

https://doi.org/10.23919/ICIF.2018.8455496

Short Term Traffic Flow Prediction with Particle Methods in the Presence of Sparse Data

Kennedy J. Offor
Dept of Automatic Control and Systems Engineering
The University of Sheffield
Sheffield, United Kingdom
Email: kjoffor1@sheffield.ac.uk

Matthew Hawes
Dept of Automatic Control and Systems Engineering
The University of Sheffield
Sheffield, United Kingdom
Email: m.hawes@sheffield.ac.uk

Lyudmila Mihaylova
Dept of Automatic Control and Systems Engineering
The University of Sheffield
Sheffield, United Kingdom
Email: l.s.mihaylova@sheffield.ac.uk

***Abstract*—Traffic prediction approaches face challenges when presented with sparse or missing data. This can be caused by numerous factors such as: i) sensors not being operational; ii) communication issues; iii) cost prohibiting full monitoring of a road network. This present work adds to existing body of knowledge by proposing a particle based framework for dealing with these challenges. An expression of the likelihood function is derived for the case when the missing value is calculated based on Kriging interpolation. With the Kriging interpolation, the missing values of the measurements are predicted, which are subsequently used in the computation of likelihood terms in the particle filter algorithm. The results show 23% to 36.34% improvement in RMSE values for the synthetic data used.**

I. Introduction

Traffic state estimation and forecasting is an essential part of Intelligent Transportation System (ITS) for effective traffic monitoring and control. Most traffic estimation approaches are model based [1], while the new trend is to develop data driven approaches [2], [3]. Traffic modelling methods are used to understand the evolution of traffic and estimate the traffic state [4]–[7].

An overview of different models is given in [7]–[9]. These include microscopic, macroscopic and mesoscopic models. Microscopic traffic models [7], [9]–[11], describe the motion of each individual vehicle with a high level of detail. Macroscopic models [12], [13] represent the aggregated behaviour of the traffic, usually in terms of the average speed and the average density. Mesoscopic models [14] uses varying levels/degrees of detail to model traffic behaviour. Some areas are modelled with aggregated measurements as in macroscopic models and other areas the detail goes down to individual vehicles as in microscopic models. Due to its computational efficiency for most practical purposes such as traffic management, road pricing and changes in infrastructure, the macroscopic model is sufficient to produce acceptable estimation.

The cell transmission model (CTM) [13] models traffic flow using macroscopic details by dividing the road into contiguous segments called cells. An extension of CTM, the stochastic compositional model (SCM) for traffic flow [5] uses probability distributions known as sending and receiving functions, which control the number of vehicles that could leave from one cell to the next, to model the stochastic nature of traffic state evolution.

The SCM was employed with the particle filter (PF) for estimating traffic state in motorways in [15]. Measurements at the boundaries were used to estimate the traffic state within the segments. It was reported that estimation accuracy is affected at boundaries without measurements.

A major challenge in traffic prediction is the problem of missing or sparse data. Communication infrastructure upon which traffic measurements are transmitted for processing and utilization often experience failure leading to missing data, which could be more than 40% in some cases [16]. The cost of installing and managing traffic sensing devices is high making it impractical to cover all locations needed for effective observation of the full road network resulting in sparse data. Various methods and approaches have been applied by researchers to address these problems such as missing data imputation [17], compressive sensing and historical averages [18], Kriging interpolation [19].

In [20], a review of three different missing data imputation methods, interpolation, prediction and statistical learning is presented. The interpolation method uses the historical average of measurements from a given sensor at similar times of day (e.g. all weekdays at 9am) to help cope with missing data. Prediction methods use a deterministic mathematical description to model the relationship between historical and future data. The statistical methods differ from the other two by modelling the stochastic nature of the traffic pattern into the imputation algorithm.

This work adds to the existing body of knowledge by proposing the use of Kriging and particle filtering to address the challenge of sparse traffic data. It uses Kriging to compute missing values at unobserved locations, which are subsequently used in the computation of likelihood terms in the particle filter algorithm. This approach combines the benefits of Kriging, which is a powerful geospatial method and a particle filter, which can capture the scholastic variations in traffic flow.

The rest of the paper is organised as follows. Section II discusses the traffic flow and measurement model used in this work. Traffic state interpolation and prediction using Kriging

and particle filters (PF) are presented in sections III and IV, respectively. Results and discussion are presented in Section V with conclusions being drawn in Section VI

II. Traffic Evolution and Measurement Model

A. Traffic Flow Model

Consider a system with the following state equation

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \boldsymbol{\omega}_k \tag{1}$$

and observation equation

$$\mathbf{z}_k = g(\mathbf{x}_k) + \boldsymbol{\xi}_k. \tag{2}$$

Here $f(.)$ and $g(.)$ are the state and observation model functions, $\mathbf{x}_k$ is state vector, $\mathbf{z}_k$ is the measurement vector, $\boldsymbol{\omega}_k$ and $\boldsymbol{\xi}_k$ are the state and observation errors, respectively and k is discrete time index.

The stochastic compositional model (SCM) for traffic flow [5], which is an adaptation of the cell-transmission model [13], uses sending and receiving functions to model the stochastic nature of traffic state evolution. The sending functions represent the vehicles that are able to leave a cell while the receiving functions determine the vehicles that are allowed to enter a cell. Figure 1 sows how the road is divided into n segments, also called cells, with length L_i and l_i lanes. The number of vehicles crossing the boundary between segments i and $i+1$ at time k is represented by $Q_{i,k}$. $N_{i,k}$ represents the number of vehicles in segment i with average speed given by $v_{i,k}$.

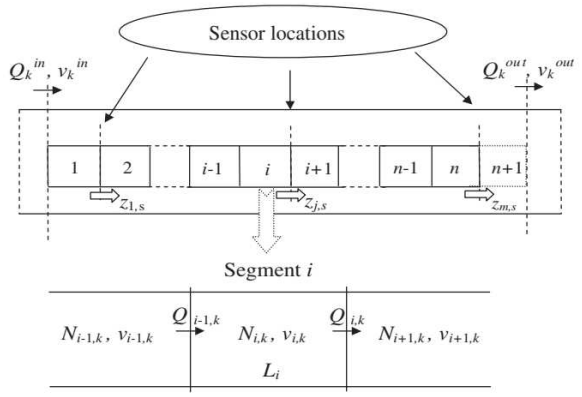

Fig. 1. SCM road network showing segments and measurement points [5].

The overall state vector at time t_k is given by $\mathbf{x}_k = [\mathbf{x}_{1,k}^T, \mathbf{x}_{2,k}^T, ..., \mathbf{x}_{n,k}^T]^T$ where $\mathbf{x}_{i,k} = [N_{i,k}, v_{i,k}]^T$ is the local state vector at segment i.

The evolution of traffic state within the segments is modelled with equations (3) to (5). The boundary conditions are the number of vehicles entering the first segment (inflow) Q_k^{in}, with average speed v_k^{in} and the number of vehicles leaving the last segment (outflow) Q_k^{out}, with corresponding average speed v_k^{out} within the time interval $\Delta t_k = t_{k+1} - t_k$. These are not estimated but supplied to the model by traffic sensors as boundary conditions. The reader is referred to [5] for a detailed algorithm.

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

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

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

B. Measurement Model

For a road segment with n boundaries, the traffic state at a boundary $j \in \mathscr{J} = 1, 2, ..., n$ is sampled at discrete time steps t_s, $s = 1, 2, ...$, to give $\mathbf{z}_{j,s} = (Q_{j,s}, v_{j,s})^T$. The matrix of measurements taken at all of the n boundaries is given by $\mathbf{Z}_s = [\mathbf{z}_{1,s}^T, \mathbf{z}_{2,s}^T, ..., \mathbf{z}_{n,s}^T]^T$. The sampling interval Δt_s is usually split into q state update time steps Δt_k. That is, $\Delta t_s = q\Delta t_k$.

With the assumption of Gaussian noise, the measurement $\mathbf{z}_{j,s}$ can be expressed as:

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

Here, $Q_{j,s}$ is the number of vehicles crossing segment j within time step s with average speed $v_{j,s}$ and $\boldsymbol{\xi}_s = [\xi_{Q_{j,s}}, \xi_{v_{j,s}}]^T$ is the error.

III. Spatial Traffic Data Estimation Using Kriging

The Kriging algorithm [21] is a point based estimation method which relies on exploiting the spatial correlation, of the data points. The Kriging algorithm attempts to interpolate the values at an unobserved location using statistics of the spatial variation between pairs of observed locations in a given region.

A. Variogram

The measurement (speed and count) $\mathbf{z}(\mathbf{y})$ observed at two-dimensional locations $\mathbf{y}$, are modelled according to equation (7) into two parts called the *drift* $\boldsymbol{\mu}$ and the *residuals* (or error) $\boldsymbol{\epsilon}(\mathbf{y})$.

$$\mathbf{z}(\mathbf{y}) = \boldsymbol{\mu} + \boldsymbol{\epsilon}(\mathbf{y}) \tag{7}$$

The *drift* is the average value of measurements and it is assumed to be constant in a given region of interest. The *residual*, $\boldsymbol{\epsilon}(\mathbf{y}) = \mathbf{z}(\mathbf{y}) - \boldsymbol{\mu}$, is a zero-mean-valued random quantity with covariance $Cov(h)$ given by (9), which models the correlation of measurements at different locations based on their separation distance called *lag* h.

Given measurements $\mathbf{z}(\mathbf{y}_i)$ and $\mathbf{z}(\mathbf{y}_j)$ at locations $\mathbf{y}_i$ and $\mathbf{y}_j$, respectively, the *lag* is given by $h = ||\mathbf{y}_i - \mathbf{y}_j||_2$, where $||.||_2$ is the l_2 norm. For a straight stretch of motorway this is equivalent to the path length through the road network. For convenience, $\mathbf{z}(\mathbf{y}_i)$ is denoted as $\mathbf{z_i}$ and $\mathbf{z}(\mathbf{y}_j)$ as $\mathbf{z}_{i+h}$. The following is then used to calculate the covariance:

$$Cov(h) = E[\{\mathbf{z}(\mathbf{y}_i) - \boldsymbol{\mu}\}^T \{\mathbf{z}(\mathbf{y}_j) - \boldsymbol{\mu}\}]. \quad (8)$$

which is equivalent to:

$$Cov(h) = E[\boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_{i+h}] \quad (9)$$

The covariance is a function of the lag distance, h, between pairs of locations and it is used to compute the empirical semi-variogram of the random process. The empirical semi-variogram, $\hat{\gamma}(h)$, can be computed using:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [\boldsymbol{\epsilon}_i - \boldsymbol{\epsilon}_{i+h}]^T [\boldsymbol{\epsilon}_i - \boldsymbol{\epsilon}_{i+h}] \quad (10)$$

where $N(h)$ is the number of pairs of observations with *lag*, h.

B. Nugget, Sill and Range

If a graph of the lag and variogram of the measurements are plotted, then Figure 2 is generated. Three points on the graph are useful in the computation of the model variogram for finding the Kriging weight. The variogram increases with the lag between points up to a point where it flattens out and remains constant. All pairs of location within the range are correlated while locations father than the range are not. The distance at which this flattening starts is called the *range* and the value of the variogram at that point is called the *sill*. All points at a location less than the range are correlated and all points with separation more than the range are not correlated. At lag distance of zero $\gamma(0)$ is expected to be zero. This is not so in practise but exhibits what is known as *nugget* effect. This is attributed to either error in the measurements or when the spatial variation is less than the sampling rate.

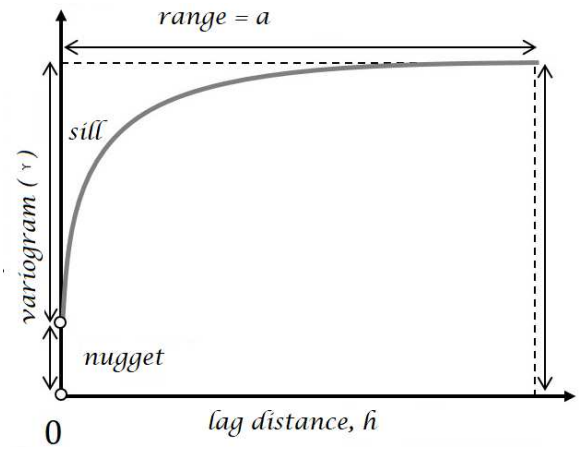

Fig. 2. Plot of empirical variogram γ and lag h

The computed empirical semi-variogram from the given number of observed locations is used to determine the parameters (*nugget*, *range* and *sill* in the chosen variogram model. To obtain values at unobserved locations, the values of these parameters are then used to calculate the co-variogram matrix between all the contributing observed locations and the unobserved locations required to find the Kriging weights. The variogram models to choose from include exponential, spherical, Gaussian, linear or power model [22]. The exponential model was used in this work because it gives the best fit for the dataset. This is given by,

$$\gamma(h) = c_0 + c\{1 - e^{-\frac{h}{a}}\} \quad (11)$$

where c_0 is the nugget, and $c = sill - c_0$ is the maximum of the correlated variance and α is the range.

C. Ordinary Kriging

After computing the semi-variograms, the values at an unobserved location, $(\mathbf{y}_u)$ can be estimated as a weighted sum of the measurements at the observed locations. This is given by:

$$\hat{\mathbf{z}}_u = \sum_{i=1}^{m} w_i \mathbf{z}_i = \mathbf{w}^T \mathbf{Z}. \quad (12)$$

where $\hat{\mathbf{z}}_u$ is the estimated value at unknown location $\mathbf{y}_u$ and $\mathbf{w} = [w_1, w_2, ..., w_m]^T$ are the Kriging weights at the m observed locations. To ensure an optimal solution the unbiased estimation constraint $E[\mathbf{z}_u - \hat{\mathbf{z}}_u] = 0$ is applied. The Kriging weights can be computed by minimizing the estimator error variance $Var[.]$ with the constraint $\sum_i w_i = 1$,

$$\min_{w_i \in \Re} Var[\mathbf{z}_u - \hat{\mathbf{z}}_u]. \quad (13)$$

Let the variogram and co-variogram of all the m observed locations to be used in the Kriging interpolation be expressed in matrix form as:

$$\mathbf{A} = \begin{bmatrix} \gamma_{1,1} & \cdots & \gamma_{1,m} \\ \vdots & \cdots & \vdots \\ \gamma_{m,1} & \cdots & \gamma_{m,m} \end{bmatrix} \quad (14)$$

and the co-variogram of the point to be estimated and the m contributory observed locations represented as:

$$\mathbf{b} = \begin{bmatrix} \gamma_{u,1} \\ \vdots \\ \gamma_{u,m} \end{bmatrix} \quad (15)$$

With this notation, the mean squared error (MSE) can be expressed as,

$$\begin{aligned} MSE &= E[(\mathbf{z}_u - \hat{\mathbf{z}}_u)^T (\mathbf{z}_u - \hat{\mathbf{z}}_u)] \\ &= Var[\mathbf{z}_u - \hat{\mathbf{z}}_u] \\ &= Var[\mathbf{z}_u] + Var[\hat{\mathbf{z}}_u] - 2Cov[\mathbf{z}_u \hat{\mathbf{z}}_u] \\ &= \gamma_u + \mathbf{w}^T \mathbf{A} \mathbf{w} - 2\mathbf{w}^T \mathbf{b}. \end{aligned} \quad (16)$$

Equation (16) is minimized by introducing a Lagrange multiplier, -2λ with the constraint $\mathbf{w}^T \mathbf{1} = 1$, where $\mathbf{1}$ is a vector of ones.

$$MSE = \gamma(\mathbf{y}_u) + \mathbf{w}^T \mathbf{A} \mathbf{w} - 2\mathbf{w}^T \mathbf{b} + 2\lambda(\mathbf{w}^T \mathbf{1} - 1). \quad (17)$$

By partial differentiation wrt $\mathbf{w}$ we get:

$$\frac{\partial MSE}{\partial \mathbf{w}} = 2\mathbf{A}\mathbf{w} - 2\mathbf{b} + 2\lambda \mathbf{1} = 0. \quad (18)$$

This implies that:

$$\mathbf{A}\mathbf{w} = \mathbf{b} - \lambda\mathbf{1}. \tag{19}$$

The Lagrange multiplier, λ is computed by solving equation (19) by direct substitution of values,

$$\begin{aligned} \lambda\mathbf{1} &= \mathbf{b} - \mathbf{A}\mathbf{w} \\ \lambda\mathbf{A}^{-1}\mathbf{1} &= \mathbf{A}^{-1}\mathbf{b} - \mathbf{w} \\ \lambda\mathbf{1}^T\mathbf{A}^{-1}\mathbf{1} &= \mathbf{1}^T\mathbf{A}^{-1}\mathbf{b} - \underbrace{\mathbf{1}^T\mathbf{w}}_{1} \\ \lambda &= \tfrac{\mathbf{1}^T\mathbf{A}^{-1}\mathbf{b}-1}{\mathbf{1}^T\mathbf{A}^{-1}\mathbf{1}} \end{aligned} \tag{20}$$

and with that the weights can be computed using:

$$\mathbf{w} = \mathbf{A}^{-1}(\mathbf{b} - \lambda\mathbf{1}). \tag{21}$$

Algorithm 1 gives a summary of this Kriging interpolation procedure.

After computing the weights, estimated values at unknown location are given by (12) and their variance computed as:

$$\begin{aligned} \sigma^2_{e_u} &= Var[\mathbf{z}_u - \hat{\mathbf{z}}_u] \\ &= \gamma_u + \mathbf{w}^T\mathbf{A}\mathbf{w} - 2\mathbf{w}^T\mathbf{b} \\ &= \gamma_u + \mathbf{w}^T(\mathbf{b} - \lambda\mathbf{1}) - 2\mathbf{w}^T\mathbf{b} \\ &= \gamma_u - \mathbf{w}^T\mathbf{b} - \lambda. \end{aligned} \tag{22}$$

Equation (22) provides useful information about how confident we are with the estimation accuracy. This information is used to compute the weighting factor to improve computation of particle predictor likelihood.

Algorithm 1 Kriging Algorithm for Spatial Interpolation [23]

1) Determine location of all sensors (measurement/input points)
2) Compute the distance (lag, h) between all measurement locations
   For $u = 1$ to U, U number of unknown locations to be computed
   Do the following
   a) Determine the measurement locations that will contribute to the interpolation at each unknown location y_u.
   b) Compute the distance between all measurement locations in the above step.
   c) Compute the empirical semivariogram of all the contributory measurement location pairs above.
   d) Fit the exponential semi-variogram model to obtain the nugget, sill and range.
   e) Compute the distances of point y_u to all the measurement locations identified in step (a).
   f) Compute the semivariogram of the distances above.
   g) Compute the vector $\mathbf{w}$ containing the weight factors of the point u using (21).
   h) Compute the estimated value of this point u using (12).

IV. Traffic Estimation via Bayesian Inference and Particle Filtering

A. Bayesian Estimation of Traffic State

In Bayesian estimation the posterior probability density function (PDF) $p(\mathbf{x}_k|\mathbf{Z}^k)$ of the traffic state $\mathbf{x_k}$ at time t_k is evaluated, given a set of measurements $\mathbf{Z}^k = \{\mathbf{z}_{1:k}\}$, collected up to time t_k using Bayes' rule as:

$$p(\mathbf{x}_k|\mathbf{Z}^k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Z}^{k-1})}{p(\mathbf{z}_k|\mathbf{Z}^{k-1})}. \tag{23}$$

The likelihood, $p(\mathbf{z}_k|\mathbf{x}_k)$ is defined by the observation model (2), and $p(\mathbf{z}_k|\mathbf{Z}^{k-1})$ is a normalizing constant. The prior or state prediction $p(\mathbf{x}_k|\mathbf{Z}^{k-1})$ is updated recursively using the Chapman-Kolmogorov equation given by [24]:

$$p(\mathbf{x}_k|\mathbf{Z}^{k-1}) = \int_{\mathbb{R}^{n_x}} p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{Z}^{k-1})d\mathbf{x}_{k-1}. \tag{24}$$

When the system model (1) is linear, equations (23) and (24) are analytically tractable and the Kalman filter [25] is used to obtain optimal solutions under certain constraints. When the system is highly non-linear, the recursive solution becomes expensive to compute and numerical approximations methods such as the extended Kalman filter [26], [27] and particle filter [15], [24], [28] are often employed to obtain acceptable solutions.

B. Particle Filtering for Traffic State Estimation

The particle filter estimates the traffic state by taking a sufficient number of random samples from the PDF with assigned weights. When a new measurement becomes available, it is used to compute what is known as the likelihood and a normalized form of the weights computed. The new state of the system is then updated with the computed weights. Degeneracy can be avoided by re-sampling, i.e removing particles with low weights and replicating those with high weights [24].

1) Improved Likelihood Computation: The likelihood function term $p(\mathbf{z}_k|\mathbf{x}_k)$, is computed when a new measurement arrives. 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,

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

where $\hat{\mathbf{z}}_s^{krig}$ represents the value estimated by Kriging (when measurement is not available), $\circ$ is the Hadamard product, $\boldsymbol{\beta}_s = [\beta_{s,1}, \beta_{s,2}, ..., \beta_{s,n}]^T$ is a weighting factor introduced to vary the level of confidence placed on the Kriged values. The value of β is 1 if we fully trust the Kriging result and less than 1 otherwise.

The modified particle filter procedure is shown in Algorithm 2.

Algorithm 2 PF Algorithm for Prediction with Improved Likelihood [15]

1) Initialization
   At $k = 0$; define all boundary conditions: number of samples, weight of samples as below,
   For $i = 1, ...N_{pf}$, N_{pf} number of particles;
   - generate N_{pf} samples $\{\mathbf{x}_0^{(i)}\}$ from the initial distribution $p(\mathbf{x}_0)$
   - initialize the particle weights $w_0^{(i)} = \frac{1}{N_{pf}}$.

   End for
2) Start the iteration for $k = 1, 2, ...$
   a) Prediction stage
      For $i = 1, ..., N_{pf}$,
      sample $\mathbf{x}_k^{(i)} \sim p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)})$ according to SCM model equations
      End for
   b) Use measurements to compute likelihoods and update the weights
      This step is performed when the sampling time t_s equals the iteration count t_k
      i. Estimate missing measurements with Kriging using Algorithm 1
      ii. Compute the likelihoods
      Use model (6) to compute the likelihood, $p(\mathbf{z}_s|\mathbf{x}_s^{(i)})$ of the particles
      iii. Update the weights of the particles using the likelihood $p(\mathbf{z}_s|\mathbf{x}_s^{(i)})$ calculated from model (6)
      For $i = 1, ..., N_{pf}$
      $\omega_s^{(i)} = \omega_{s-1}^{(i)} p(\mathbf{z}_s|\mathbf{x}_s^{(i)})$
      End For
      iv. Normalize the weights: $\hat{\omega}_s^{(i)} = \frac{\omega_s^{(i)}}{\sum_{i=1}^{N_{pf}} \omega_s^{(i)}}$.
   c) Update the predicted states (Output): $\hat{\mathbf{x}}_s = \sum_{i=1}^{N} \hat{\omega}_s^{(i)} \mathbf{x}_s^{(i)}$
   d) Re-sample the weights (Selection) only when $t_k = t_s$

V. Performance Evaluation

A. Investigation with Synthetic Data

Here, the performance of the particle filter with Kriging for traffic state estimation is evaluated using synthetic data from a 4km stretch of motorway over a period of three hours. This is split into eight segments, each with a length of 0.5km and three lanes. For more details on the process for obtaining the synthesized data see [15].

The modelling consists of periods of normal flow and congestion which was modelled by random changes (increase and decrease) in the inflow between time interval of $(1.12 \leq t < 1.17) hours$ and $(1.70 \leq t < 1.82) h$ and outflow speed (decrease) between $(2.40 \leq t \leq 2.65) hours$.

To test the effect of assigning different values between 0 and 1 to $\boldsymbol{\beta}_s$, the simulation was repeated three times each with 200 independent Monte Carlo runs. First, by using all the measurements available at the segment boundaries. Second, removing measurements at two locations (boundary segments) and interpolating them using Kriging with equal weights assigned to the interpolated measurement and actual measurement in the likelihood computation. Lastly, by assigning a weight of 0.2 to the Kriging interpolated values.

In order to test the prediction accuracy for different levels of sparsity, three statistical measures namely the root mean squared error (RMSE), the absolute percentage error (APE) and the mean absolute error (MAE) were computed. Note, $\mathbf{z}_i$ is the ground truth or actual measurement, $\hat{\mathbf{z}}_i$ is the estimated value and m_r is number of independent Monte Carlo runs.

$$RMSE = \sqrt{\frac{1}{m_r} \sum_{i=1}^{m_r} [\mathbf{z}_i - \hat{\mathbf{z}}_i]^2}. \tag{28}$$

$$APE = \sum_{i=1}^{m_r} \frac{|[\mathbf{z}_i - \hat{\mathbf{z}}_i]|}{\mathbf{z}_i} * 100. \tag{29}$$

$$MAE = \frac{1}{m_r} \sum_{i=1}^{m_r} |[\mathbf{z}_i - \hat{\mathbf{z}}_i]|. \tag{30}$$

The RMSE and APE for 200 independent Monte Carlo runs were plotted in Figures 3 and 4, respectively. The results show that there is an improvement in the estimation accuracy when the Kriging interpolated values were assigned weights. The results for the two Kriged examples reach the same accuracy for location 4. When the middle sensor is removed there is still information up and down flow from the missing sensor. As a result more information can be applied to the interpolation process allowing a more accurate estimate. Instead when there are no sensor down flow for example there is less information therefore the interpolation is less accurate.

The results in Figures 5 and 6 show that estimation accuracy is better when all the measurements are used in computing the likelihood, followed by that computed with kriging interpolated measurements. The accuracy of the estimation performed without using any measurement is the least as can be seen in the figures. The results show improvement of 23% to 36.34% at different segments in RMSE values for the synthetic data used.

Figures 7 and 8 shows the plot of velocity and flow at segments boundaries 2 to 6. Segments 1 and 8 were not included as they were the inflow and outflow segments. It is evident that the estimation when all measurements (second plot from bottom) are used provides the best accuracy, followed by that where Kriging was used to estimate the missing measurements

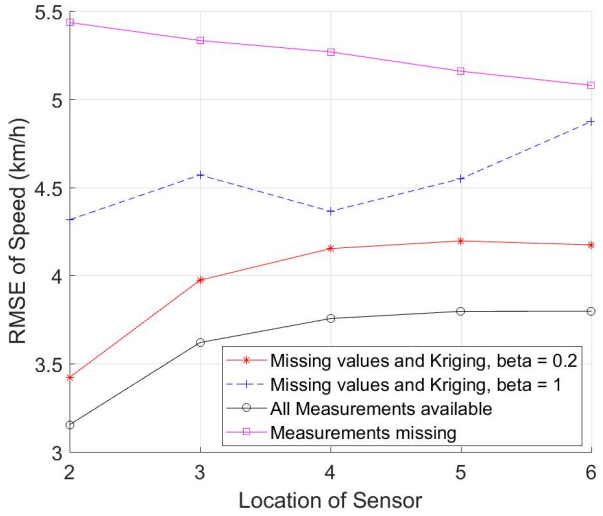

Fig. 3. The root mean squared error (RMSE) of speed over locations

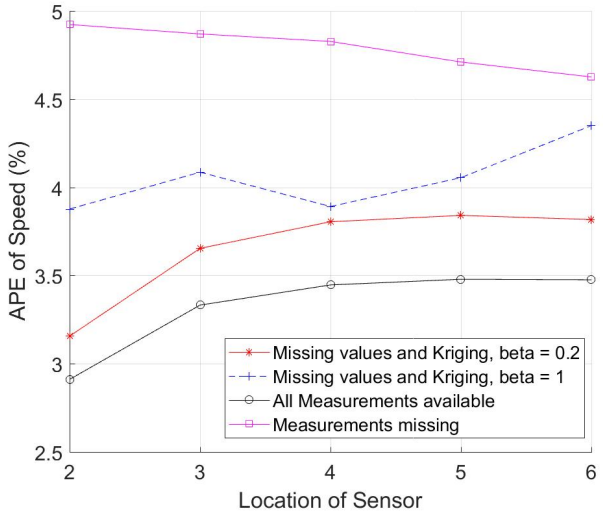

Fig. 4. The absolute percentage error (APE) of speed over locations

(third plot from bottom). The least accurate result was obtained when the missing values were not included in the likelihood computation (first plot from the top). This is apparent during congestion between time interval $(2.40 \leq t \leq 2.65)$ hours.

Another observation from the figures is that the estimation accuracy is consistent under free flow conditions and begins to get worse as congestion sets in. During the period when the network is congested from time interval $(2.40 \leq t \leq 2.65)$ hours the estimate without the full measurements used in the likelihood computation could not capture the decrease in speed. Incorporating the Kriged values improves the accuracy a little while the estimate with full measurements is closest to the true value.

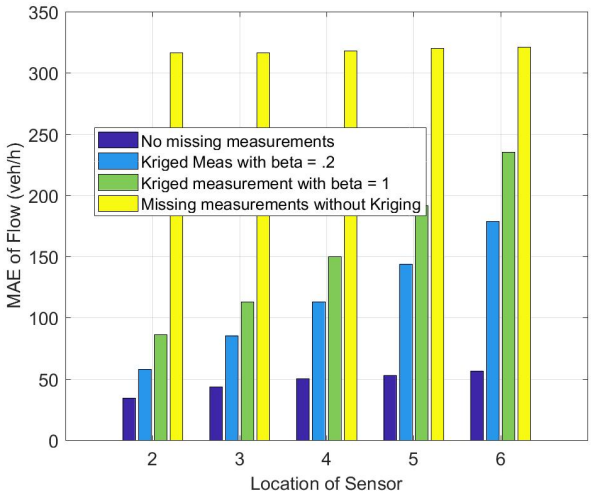

Fig. 5. The mean absolute error (MAE) of flow for locations

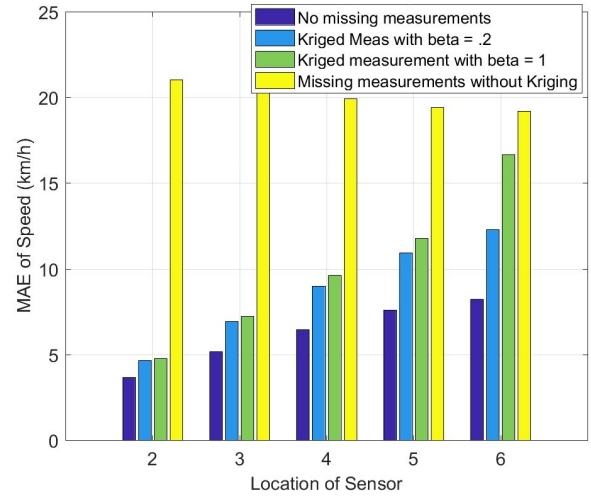

Fig. 6. The mean absolute error (MAE) of speed for locations

B. Investigation with Real Data

The modified algorithm was further tested with real data from the E-17 motorway in Belgium [15], which is usually congested. The test data consist of a day measurement recorded by sensors installed at locations CLOF to CLO9 as shown in Figure 9. Measurements at location CLOE to CLOB were removed and then interpolated using Kriging with the following parameters, free flow speed v_{free} = 120 km/h, minimum speed v_{min} = 7.4 km/h, critical density ρ_{crit} = 20.89 veh/km/lane, jam density ρ_{jam} = 180 veh/km and a $\beta = 0.5$. Figures 10 show the plot of the estimated values and the ground truth. The estimates follow the pattern of measured states at most of the points as seen from the plot. The speed flow and flow-density diagrams are plotted in Figures 11 and 12. The shape of the figures resemble the fundamental diagram of traffic flow confirming the validity of the approach.

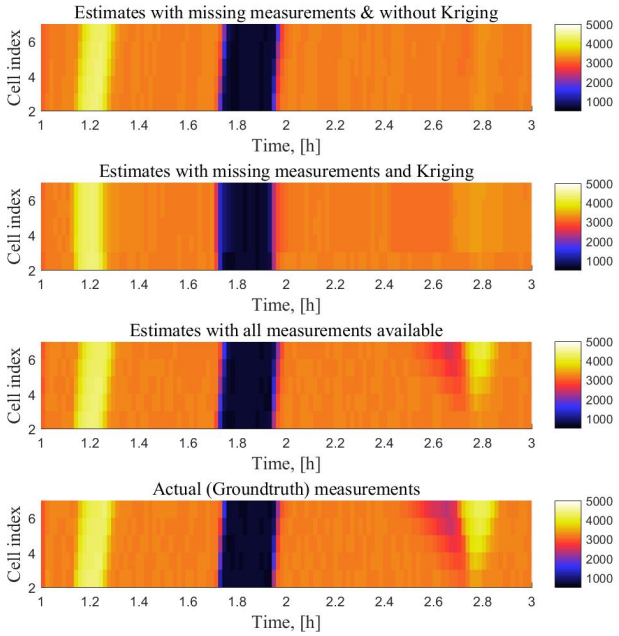

Fig. 7. The flow surf of segments 2 to 6

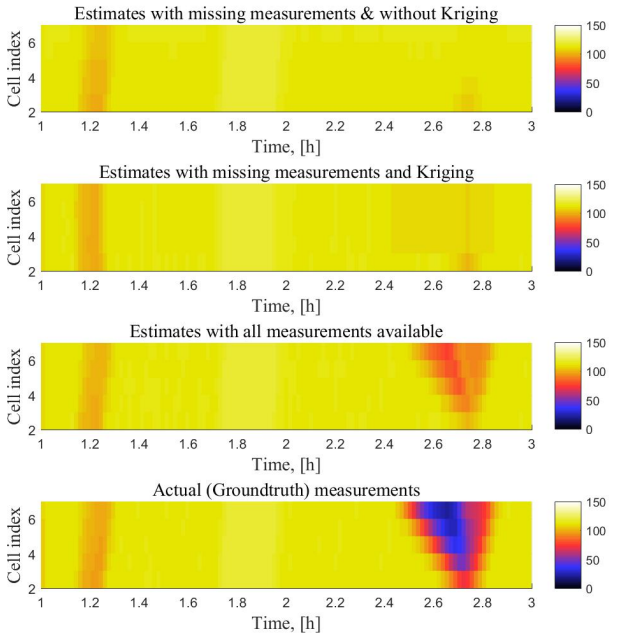

Fig. 8. Velocity surf of segments 2 to 6

VI. Conclusion

This paper presented a novel approach to tackle the problem of missing and sparse data in traffic estimation. This approach entails interpolating the missing values using Kriging with a level of confidence assigned to the kriged values by computing their interpolation error variance. This level of confidence is then used to compute the weight to be assigned during computation of innovation terms used in PF. This was tested using simulated and real data by assigning fixed test-values to the weighting factor. From the results presented benefit of lowering the weighting of interpolated values as compared to actual measurements has offered an improvement.

In a future work, the algorithm will be validated further by empirically computing the weighting factor β of the kriging estimate of the missing measurements with real data from a larger road network. In addition, the use of different methods in calculating the kriging variance would be investigated.

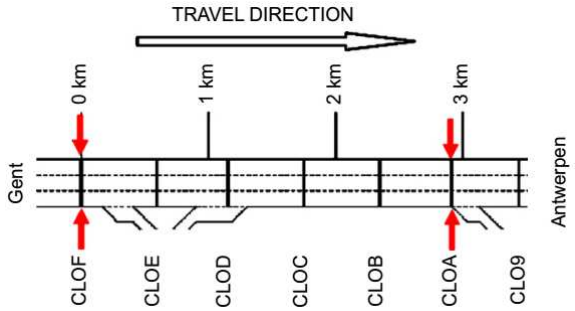

Fig. 9. Schematic diagram of the E17 freeway between Ghent and Antwerp Kruibeke, Belgium [15]

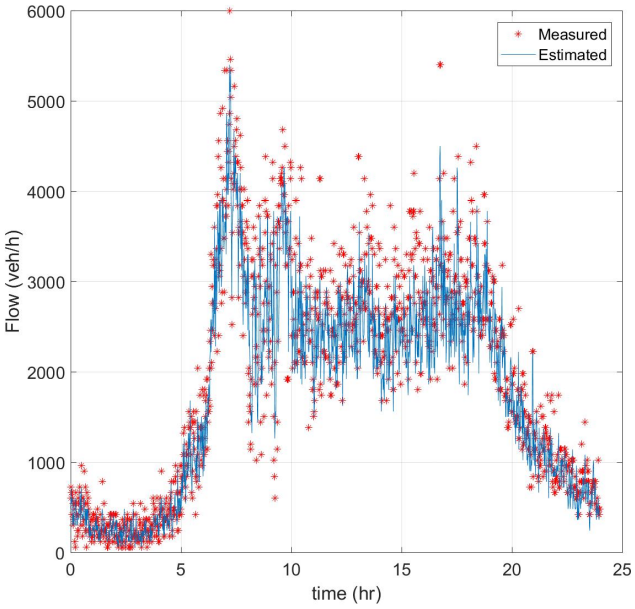

Fig. 10. Predicted states (solid line) and actual measured states (*) at CLOC

Acknowledgement

We appreciate the support of SETA project funded from the European Unions Horizon 2020 research and innovation programme under grant agreement No. 688082 and the Tertiary Education Trust Fund (TETFund, Nigeria). We also thank the Vlaams Verkeerscentrum Antwerpen, Antwerp, Belgium, for providing the real data used in this study.

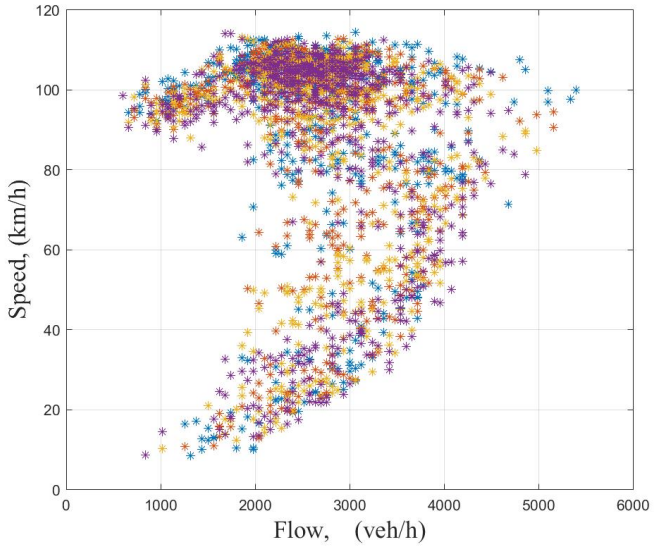

Fig. 11. Speed-low diagram for the PF with Kriging estimated measurements at CLOE, CLOD, CLOC and CLOB

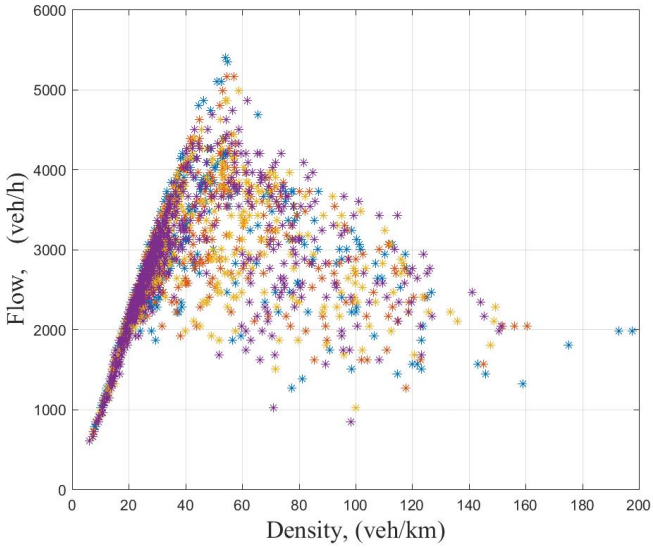

Fig. 12. Flow-density diagram for the PF with Kriging estimated measurements at CLOE, CLOD, CLOC and CLOB

REFERENCES

[1] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura, "Traffic state estimation on highway: A comprehensive survey," *Annual Reviews in Control*, vol. 43, 2017, pp. 128–151, 2017.

[2] C. Bachmann, B. Abdulhai, M. J. Roorda, and B. Moshiri, "A comparative assessment of multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling," *Transportation Research Part C*, vol. 26, no. 1, pp. 33–48, 2013.

[3] H.-p. Lu, Z.-y. Sun, and W.-c. Qu, "Big Data-Driven Based Real-Time Traffic Flow State Identification and Prediction," *Discrete Dynamics in Nature and Society*, vol. 2015, pp. 1–11, 2015.

[4] M. S. Ahmed and A. R. Cook, "Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques," *Transportation Research Record*, vol. 722, pp. 1–9, 1979.

[5] R. Boel and L. Mihaylova, "A compositional stochastic model for real time freeway traffic simulation," *Transportation Research Part B: Methodological*, vol. 40, no. 4, pp. 319–334, 2006.

[6] B. D. Greenshields, "The photographic method of studying traffic behavior," in *Proc. 13th Annual Meeting of the Highway Research Board*, (Washington, D.C.,), pp. 382–399, Highway Research Board, 1934.

[7] S. P. Hoogendoorn and P. H. L. Bovy, "Generic gas-kinetic traffic systems modeling with applications to vehicular traffic flow," *Transportation Research Part B*, vol. 35, no. 4, pp. 317 – 336, 2001.

[8] C. F. Daganzo, *Traffic flow theory, in Carlos F. Daganzo (ed.) Fundamentals of Transportation and Traffic Operations*. Bingley, UK: Emerald Group Publishing, 1997.

[9] M. Papageorgiou, M. Ben-Akiva, J. Bottom, P. H. L. Bovy, S. P. Hoogendoorn, N. B. Hounsell, A. Kotsialos, and M. Mcdonald, *Chapter 11, ITS and Traffic Management in: Handbooks in Operations Research and Management Science.*, vol. 14. Amsterdam: Elsevier, 2007.

[10] L. A. Pipes, "An Operational Analysis of Traffic Dynamics," *Journal of Applied Physics*, vol. 24, no. 3, pp. 274–281, 1953.

[11] P. G. Gipps, "A Behavioural Car-Following Model for Computer Simulation," *Transp Res Part B Methodol*, vol. 15, no. 2, pp. 105–111, 1981.

[12] A. Messmer and M. Papageorgiou, "METANET: A macroscopic simulation program for motorway networks," *Traffic Engineering and Control*, vol. 31, no. 8-9, pp. 466 – 470, 1990.

[13] C. F. Daganzo, "The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory," *Transpn. Res.-B*, vol. 28, no. 4, pp. 269–287, 1994.

[14] S. P. Hoogendoorn and P. H. L. Bovy, "State-of-the-art of vehicular traffic flow modelling," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 215, no. 4, pp. 283–303, 2001.

[15] L. Mihaylova, R. Boel, and A. Hegyi, "Freeway traffic estimation within particle filtering framework," *Automatica*, vol. 43, no. 2, pp. 290–300, 2007.

[16] Y. Zhang and Y. Zhang, "A Comparative Study of Three Multivariate Short-Term Freeway Traffic Flow Forecasting Methods With Missing Data," *Journal of Intelligent Transportation Systems*, vol. 20, no. 3, pp. 205–218, 2016.

[17] A. Ladino, A. Y. Kibangou, H. Fourati, and C. Canudas de Wit, "Travel time forecasting from clustered time series via optimal fusion strategy," in *Proc. European Control Conference 2016*, (Aalborg, Denmark), 2016.

[18] M. Hawes, H. M. Amer, and L. Mihaylova, "Traffic State Estimation via a Particle Filter with Compressive Sensing and Historical Traffic Data," in *Proc. 2016 19th International Conference on Information Fusion (FUSION)*, pp. 735–742, 2016.

[19] H. Zou, Y. Yue, Q. Li, and A. G. Yeh, "An improved distance metric for the interpolation of link-based traffic data using kriging: a case study of a large-scale urban road network," *International Journal of Geographical Information Science*, vol. 26, no. 4, pp. 667–689, 2012.

[20] Y. Li, Z. Li, and L. Li, "Missing traffic data: comparison of imputation methods," *IET Intelligent Transport Systems*, vol. 8, pp. 51–57, feb 2014.

[21] D. G. Krige, "A statistical approach to some basicmine problems on the Witwatersrand," *ournal of the Chemical, Metallurgical and Mining Society of South Africa*, vol. 52, no. 6, pp. 119–139, 1951.

[22] X. Wang and K. M. Kockelman, "Forecasting Network Data: Spatial Interpolation of Traffic Counts Using Texas Data," *Transportation Research Record*, vol. 2105, pp. 100–108, 2009.

[23] M. Oliver and R. Webster, "A tutorial guide to geostatistics: Computing and modelling variograms and kriging," *CATENA*, vol. 113, pp. 56–69, feb 2014.

[24] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[25] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems 1," *Transactions of the ASME - Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[26] S. J. Julier and J. K. Uhlmann, "A New Extension of the Kalman Filter to Nonlinear Systems," in *Proc. The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls*, 1997.

[27] S. Julier, J. Ulhmann, and H. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, 2000.

[28] Z. Chen, "Bayesian Filtering : From Kalman Filters to Particle Filters , and Beyond," Tech. Rep., McMaster University, Canada, 2003.