Multiple Imputations Particle Filters: Convergence and Performance Analyses for Nonlinear State Estimation With Missing Data

Xiao-Ping Zhang, *Senior Member, IEEE*, Ahmed Shaharyar Khwaja, *Member, IEEE*, Ji-An Luo, Alon Shalev Housfater, and Alagan Anpalagan, *Senior Member, IEEE*

***Abstract*—In this paper, we present a multiple imputations particle filter (MIPF) to deal with non-linear state estimation when part of the observations are missing. The MIPF uses randomly drawn values called imputations to provide a replacement for the missing data and then uses the particle filter to estimate non-linear state with the data. Unlike the existing techniques, we do not assume a linear system and also take into account the time-varying transition matrix when accounting for missing data. We present the convergence analysis of the MIPF and show that it is almost surely convergent. We present examples with a non-linear time-varying model, which demonstrate that the MIPF can effectively deal with missing data in nonlinear problems. Comparison with existing techniques further validates the improvement offered by the proposed MIPF.**

***Index Terms*—Particle filter, missing data, imputations, non-linear state estimation, multiple imputations particle filter.**

I. Introduction

NON-LINEAR state estimation deals with estimation of input states, provided a measurement sequence from a sensor with the unknown input sequence. The observation and/or state transition matrices represent how the input and output sequences evolve with time are non-linear. Theoretically, the optimal non-linear filter for state estimation may be generated by the use of Bayesian techniques, whose main idea is to estimate the probability density function of the nonlinear state on the basis of the given observations.

Analytical approximations of the optimal nonlinear method include the extended Kalman filter (EKF) [1], the sigma point Kalman filter (SPKF) [2], etc. EKF is based on first order expansion of the non-linearities and approximates system state distribution and noise densities by Gaussian random variables [3]. It is known to be effective for systems that are linear or near linear but has poor robustness [4] and fails to give good estimates if the system is not approximated well by a localized linearization. SPKF also estimates state distribution by Gaussian random variables, but is based on deterministic sampling approach and uses a set of deterministically chosen weighted sample points; it is able to achieve accuracy obtained by second or third order Taylor Series expansion [3]. Its computational complexity is of the same order as the EKF and its implementation is easier compared to the EKF. Particle filter (PF) [5]–[7] solves the estimation problem numerically and is useful for non-linear non-Gaussian state-space models that can not be solved using analytic means.

In addition, data from multiple sensors may be used in non-linear state estimation problems so as to complement the data of one sensor by the data of the other sensor. This allows to extract the maximum amount of information about the environment that is being sensed, e.g., radar measurements from individual sensors are integrated in the fusion center to obtain a global surveillance picture. EKF and PF have been used for data fusion, a couple of relevant references are [8] and [9].

Missing data arises in many practical applications involving single or multiple sensors due to, e.g., hardware power limitations, unreliable channel, sensor failure, low signal-to-noise ratio, data loss, expensive acquisition equipment, limited storage space for data, etc. The above mentioned standard approaches for state estimation can not be used in such a situation without modifications. Some references where this issue has been addressed are described in the following.

One approach is to use a PF based on compressive sensing [14] to deal with missing data. However, this approach requires measurement matrices that are either physically not realizable or in case they are physically realizable, result in performance degradation. It also requires that the states be sparse in a particular dictionary or transform, which may not always be possible and leads to performance degradation when such conditions are not met. This approach is also susceptible to mismatch errors [15].

Another possible approach is to use Expectation Maximization (EM) algorithm [16], [17] that provides an iterative way to minimize the likelihood function of incomplete data. This algorithm is applied to resolve missing data in [18]. It ignores the state dynamics.

Manuscript received January 27, 2015; revised May 31, 2015; accepted July 23, 2015. Date of publication August 06, 2015; date of current version November 17, 2015. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN239031. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Xiaopeng Yang.

X.-P. Zhang, A.S. Khwaja, A. Housfater, and A. Anpalagan are with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail: xzhang@ee.ryerson.ca).

J.-A. Luo was with (CASPAL), Ryerson University, Toronto, ON M5B 2K3, Canada. He is now with the Institute of Information and Control, Hangzhou Dianzi University, Hangzhou 310018, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2015.2465360

A few other references deal with missing data problem in the context of data fusion. In [10], optimal fusion of measurements obtained from two sensors that suffer from missing data is described. In [11]–[13], missing data problem is described to solve asynchronous fusion problem. This problem arises in case where the scanning process of different sensors may not be synchronized and have different measurement rates. Thus data fusion can be seen as handling missing data in case where one data set is available at a higher rate and the other one at a lower rate. In [11], an EKF is formulated with a time-varying transition matrix for carrying out asynchronous data fusion for a radar network; it does not deal explicitly with missing data. In [12], Kalman filter based fusion of asynchronous data fusion is carried out for linear systems in the absence of data. In [13], asynchronous multi-sensor fusion for non-linear systems in case of missing data is described using SPKF. In [12] and [13], when there are no missing data, fusion is carried out normally. In the presence of missing data, the main idea of these references is to use the previous state and error covariance estimates.

Another approach for non-linear state estimation with missing data is to use a multiple imputations particle filter (MIPF) technique that consists of replacing missing data with multiple imputations, i.e., randomly drawn values, to form multiple complete data sets. Multiple imputation method is based on the Bayesian framework that allows to simulate the posterior distribution of missing values by imputing each data with several values according to one or more estimation models. This technique was used in [19] and [20] to deal with multi-rate data arising in case of asynchronous fusion. However, these references did not explicitly deal with missing data problem and did not analyze the convergence of the solution obtained by MIPF, which is necessary to show that it is able to estimate the states correctly from missing data. In [21], we presented some preliminary convergence and experimental results relating to MIPF.

In this paper, we propose MIPF for non-linear state estimation problem in case of missing data and present convergence analysis of MIPF. We extend results in [22] and show that MIPF is almost surely convergent. We further present an example using the state and system equations for the non-stationary growth model from [6] that shows the usefulness of MIPF in dealing with missing data in nonlinear problems. We also compare MIPF with various existing techniques and show its superior performance compared to these algorithms.

Our earlier work [19], did not present any convergence analysis. In our later work [21], we presented a preliminary convergence analysis. In addition, we did not compare the performance of our method with other existing methods. Results were compared only to a PF with complete data. In this paper, besides presenting a detailed convergence analysis, we compare the performance of MIPF with EKF, SPKF and EM algorithm followed by application of PF. Unlike our work in [21] where only 10% and 20% missing data were considered, we show the dependency of performance of MIPF on a range of missing samples varying till 50%. We also show the dependency of the performance of MIPF on the number of imputations.

This paper is organized as follows: Section II formulates the problem and Section III reviews particle filters and multiple imputations. MIPF algorithm is presented in Section IV, whereas Section V carries out convergence analysis of the algorithm. Section VI applies the algorithm for non-linear state estimation and compares performance with respect to several existing algorithms. Finally, Section VII concludes the paper.

NOTATION

φ_t	state transition equation at discrete time t
ψ_i	measurement transformation of observer i
ϕ	imputation proposal function
$\mathcal{U}_t^j$	imputed data set j at time t
K	number of observers
M	total number of imputations
t	discrete time index, $t \in N$
N	total number of particles
$R_{t,k}$	indicates if data from observer k at time t is missing
R_t	missing data indicator vector
$U_{t,k}$	noisy observation of the state by observer k at time t
U_t	noisy observations of the state by all observers at time t
w_t^i	weighting coefficient of particle i at time t
$w_t^{j,i}$	weighting coefficient of proposal function approximation particle i at time t using imputation j
X_t	hidden state of the system at time t
X_t^i	ith particle at time t
$X_t^{j,i}$	ith particle at time t using imputation j
Y_t	all available observations at time t
$Y_{0:t}$	all available observations from time 0 to time t
Z_t	all missing observations at time t
Z_t^j	jth imputation at time t
$\pi(X_t \mid X_{0:t}^i, Y_{0:t})$	importance function

II. PROBLEM FORMULATION

Consider a time-varying stochastic system with X_t denoting the state at time instance t. X_t behaves according to a non-homogenous Markov chain with transition probabilities described by the following equation:

$$X_t = \varphi_t(X_{t-1}, \mathcal{W}_t), \tag{1}$$

$\mathcal{W}_t$ is a random evolution noise, assumed to be independent identically distributed (iid) stochastic process and φ_t is the non-homogenous evolution transformation. Let the system be observed by K sensors where the measurements are modeled as:

$$U_t = \begin{pmatrix} \psi_1(X_t, \mathcal{V}_{t,1}) \\ \vdots \\ \psi_K(X_t, \mathcal{V}_{t,K}) \end{pmatrix} = \begin{pmatrix} U_{t,1} \\ \vdots \\ U_{t,K} \end{pmatrix} \tag{2}$$

U_t denotes the noisy observation of the state X_t where $\mathcal{V}_{t,k}$ is an *i.i.d* noise process and ψ_k is the measurement transformation for sensor k. We denote the k-entry of the K-dimensional vector U_t as $U_{t,k}$. At each time instance t, we consider the possibility that some sensor observations may be missing. In order to handle these missing observations, we introduce a random indicator variable $R_{t,k}$, which corresponds to the observation $U_{t,k}$; this variable indicates if observation U_t^k is available or not.

$$R_{t,k} = \begin{cases} 1, & \text{observation is available from sensor } k \text{ at time } t \\ 0, & \text{observation is missing from sensor } k \text{ at time } t \end{cases}$$

Next, we define the missing information set Z_t as the collection of observations U_t^k at time instance t for all observers $k = 1, \ldots, K$ such that $R_{t,k} = 0$. Similarly, the available information set Y_t is the collection of U_t^k for all $k = 1, \ldots, K$ such that $R_{t,k} = 1$. It is assumed that the missing data mechanism is independent of the missing observations given the available observations:

$$P(R_{t,k}|Z_t, Y_t) = P(R_{t,k}|Y_t) \text{ for all } k. \tag{3}$$

This standard statistical assumption is known as missing at random (MAR) [23]. Our objective is to obtain the posteriori probability density function of the state given all past and present observations. We write this density function as $p(X_t|Y_{0:t})$ where $Y_{0:t}$ denotes all observations from the initial time instance to time instance t.

III. Particle Filters and Multiple Imputations

A. Particle Filters

Consider K noisy observations of the state X_t as introduced in Section II. We assume that there is no missing information and all K observations are available for processing at each time instance t. We wish to obtain the probability density function of the state X_t given all past and current observations $Y_{0:t}$, i.e., $p(X_t|Y_{0:t})$. We can express this density using known quantities by the following set of equations:

$$p(X_t|Y_{0:t-1}) = \int p(X_t|X_{t-1})p(X_{t-1}|Y_{0:t-1})dX_{t-1} \tag{4}$$

$$P(X_t|Y_{0:t}) = \frac{p(Y_t|X_t)p(X_t|Y_{0:t-1})}{P(Y_t|Y_{0:t-1})} \tag{5}$$

As noted in [24], (4) does not have a closed form solution in the general case. The particle filter can be viewed as a technique that performs stochastic integration on (4). The particle filter uses a set of N randomly drawn states, also known as particles, X_t^i and their corresponding weights w_t^i where i indexes the particles at each time instance t. These particles, $\{w_t^i, X_t^i\}_{i=1}^N$, discretize the density $p(X_t|Y_{0:t})$ as follows:

$$p(X_t|Y_{0:t}) \approx \sum_{i=1}^{N} w_t^i \delta(X_t - X_t^i), \tag{6}$$

where δ denotes a dirac function. The particles are obtained in a recursive manner. At each time instance t, the particles from time instance $t-1$ are used in sampling from an importance function $\pi(X_t|X_{0:t}^i, Y_{0:t})$:

$$X_t^i \sim \pi(X_t|X_{0:t}^i, Y_{0:t}) \text{ for } 1 \le i \le N,$$

The particle weighting coefficient w_t^i is obtained by the following calculation:

$$w_t^i = w_{t-1}^i \frac{p(Y_t|X_t^i)p(X_t^i|X_{t-1}^i)}{\pi(X_t^i|X_{0:t}^i, Y_{0:t})}.$$

Note that in an application with missing data, the standard particle filtering algorithm does not incorporate the response information $R_{t,k}$ when performing the procedure and is not guaranteed to be stable. Thus, in the presence of missing data, the particle filter can diverge or its performance can be severely degraded.

B. Handling Missing Data With Multiple Imputations

Missing data can introduce bias into the statistical estimation process and the existence and severity of the missing data bias depends on the missing data mechanism. Unfortunately, the structure of the missing data mechanism is almost never available for analysis. However, under the condition of MAR (3), it can be shown that one does not need to know the structure of the missing data mechanism. Consider K sensors that are all observing the same hidden state X, we let Y denote the set of all available observations from the sensors and Z denote the set of all missing observations. Moreover, let $R = [R_1, \ldots, R_K]^T$ be the K-dimensional indicator vector for the response of the sensors. Note that this problem is not time dependent and there is no state space model. We wish to compute the probability density $p(X|Y, R)$, which can be written as:

$$p(X|Y, R) = \int p(X|Y, Z, R)p(Z|Y, R)dZ, \tag{7}$$

Using the condition of MAR as defined in (3), it can be shown [23] that (7) reduces to:

$$p(X|Y, R) = \int p(X|Y, Z)p(Z|Y)dZ.$$

Thus, we do not need to know the statistical structure of the missing information mechanism. We can approximately compute the density $p(X|Y, R)$ by the Monte Carlo approximation:

$$p(X|Y) = \lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} p(X|Y, Z_j), \tag{8}$$

where $Z_j \sim p(Z|Y)$ are the multiple imputations and M is the number of imputations.

Note that multiple imputations do not use the past observations and the state transition equation in estimating the density $p(X|Y)$. This is significant since many real world problems are well modeled by a Markov structure, which is dependant on the past values to determine the present ones. Thus, in such applications, we expect that the performance of the multiple imputations will be degraded.

IV. Multiple Imputations Particle Filter

We present a new algorithm that resolves the mentioned deficiencies in the particle filtering and multiple imputations algorithms. This algorithm uses both the state and observation dynamics while accounting for the missing data.

A. Approximation of the Imputing Function

The multiple imputations method draws imputations from the missing data probability density $p(Z_t|Y_{0:t})$ as shown by (8). However, in the context of a state space model with missing information, this density is unknown. In similar applications with unknown missing data probability density, a common solution is to draw the imputations using Markov Chain Monte Carlo (MCMC) methods [17]. These MCMC methods are iterative in nature and may not be applicable in some problem domains such as real time systems. In this section, we present a new approach to perform the imputation process by utilizing particle approximation techniques. The imputing probability density $p(Z_t|Y_{0:t})$ can be written as:

$$p(Z_t|Y_{0:t}) = \int p(Z_t|X_t)p(X_t|Y_{0:t})dX_t. \tag{9}$$

Note that the filtering density $p(X_t|Y_{0:t})$ appears inside the integral; as we do not know this density, we are unable to sample directly from $p(Z_t|Y_{0:t})$. However, (9) suggests the following approximation:

First, find a discrete density $\widetilde{p}(X_t|Y_{0:t})$ that approximates the true filtering density reasonably well. Then using the relationship given in (9), one can obtain the discrete density $\widetilde{p}(Z_t|Y_{0:t})$, which will approximate the desired density $p(Z_t|Y_{0:t})$. We write the approximate filtering density $\widetilde{p}(X_t|Y_{0:t})$ using a particle approximation as follows:

$$\widetilde{p}(X_t|Y_{0:t}) = \sum_{i=1}^{N} \widetilde{w}_t^i \delta(X_t - \widetilde{X}_{t,i}), \tag{10}$$

where the particle set $\{\widetilde{w}_t^i, \widetilde{X}_t^i\}_{i=1}^N$ is obtained by performing the particle filtering with no regard for missing data, as in Section III-A. Substituting this approximation into (9) and assuming $p(Z_t|\widetilde{X}_{t,i}) = p(Y_t|\widetilde{X}_{t,i})$, i.e., the missing data's pdf is the same as that with observable data, we obtain the proposal function $\phi(Z_t|Y_{0:t})$:

$$p(Z_t|Y_{0:t}) \approx \phi(Z_t|Y_{0:t}) = \sum_{i=1}^{N} \widetilde{w}_t^i p(Z_t|\widetilde{X}_{t,i}).$$

B. MIPF Algorithm

First, we draw the random observations or imputations, from a proposal function ϕ:

$$Z_t^j \sim \phi(Z_t|Y_{0:t}) \text{ for } j = 1, \ldots, M,$$

Note that we can write the filtering probability density $p(X_t|Y_{0:t})$ as follows [25]:

$$p(X_t|Y_{0:t}) = \int p(X_t|U_{0:t-1}, Y_t)p(Z_t|Y_{0:t})dZ_t. \tag{11}$$

By forming the imputed data sets $\mathcal{U}_t^j = \{Z_t^j, Y_t\}$ and taking a Monte Carlo approximation, we can write (11) as follows:

$$p(X_t|Y_{0:t}) \approx \frac{1}{M}\sum_{j=1}^{M} p\left(X_t|U_{0:t-1}, \mathcal{U}_t^j\right). \tag{12}$$

Next, the algorithm performs particle filtering on each data set U_t^j to obtain an approximation as in (6):

$$p(X_t|U_{0:t-1}, \mathcal{U}_t^j) \approx \sum_{i=1}^{N} w_t^{j,i}\delta\left(X_t - X_t^{j,i}\right), \tag{13}$$

where $X_t^{j,i}$ is the ith particle for the jth imputation at time instance t and $w_t^{j,i}$ is its weight. Finally, the algorithm combines the multiple particle filtering results by substituting (13) into (12) to obtain an approximation of the desired density as follows:

$$p(X_t|Y_{0:t}) \approx \frac{1}{M}\sum_{j=1}^{M}\sum_{i=1}^{N} w_t^{j,i}\delta\left(X_t - X_t^{j,i}\right).$$

V. MIPF Convergence Analysis

Given the algorithm described in the previous section, an obvious and critical question to ask is how the approximate distribution $p(X_t|Y_{0:t})$ obtained by the multiple imputations particle filter is related to the true density. We begin by reformulating the state space model and the multiple imputations particle filter. We then follow by analyzing its convergence. This analysis extends results by Crisan *et al.* [22].

A. Probability Space Formulation

Let $(\Omega, \mathcal{F}, P)$ be a probability space where $\mathcal{F} = \mathcal{B}(R^{n_x})$ is the Borel set of R^{n_x}, the Borel set is the standard set of all possible probability events on R^{n_x}. On this probability space we define a vector-valued stochastic process $X = \{X_t, t \in N\}$ where n_x is the dimension of the state space of X. The process X is Markov with initial distribution $X_0 \sim \mu$ and probability transition kernel $K(x_t|x_{t-1})$:

$$p(X_t \in A|X_{t-1} = x_{t-1}) = \int_A K(x_t|x_{t-1})dx, A \in \mathcal{B}(R^{n_x})$$

The process X can be viewed as hidden state that we wish to obtain, for example in radar application this would be the true object position. Moreover, we define a stochastic process $W = \{W_t, t \in N\}$ where $W_t = \{W_t^1, \ldots, W_t^k\}$ and W_t^i is a n_w-dimensional vector for $1 \leq i \leq k$. The process W is conditionally independent of X:

$$p(W_t \in B|X_t = x_t) = \int_B g(w_t|x_t)dw_t, B \in \mathcal{B}(R^{n_w \times k}) \tag{14}$$

We may think of the process W as k noisy observations of the hidden markov process X, in our application these observations would represent all observations from multiple sensors. We let the density of W conditional on X have the following factorization

$$g(w_t|x_t) = \prod_{i=1}^{k} g_i\left(w_t^i|x_t\right) \tag{15}$$

We can think of this factorization as the statement that given an array of k sensors, each sensor is independent of the other. Thus, combining the statements (14) and (15) we have

$$P(W_t^i \in C|X_t = x_t) = \int_C g(w_t^i|x_t)dw_t^i, C \in \mathcal{B}(R^{n_w})$$

Consider the non-response vector-valued stochastic process $R = \{R_t, t \in N\}$ where R is a n_w-dimensional vector. We let

$r_t^i \in (0,1)$ be indicator variables. We introduce the following sets:

$$z_t = \left\{w_t^i | r_t^i = 0 \text{ for } 0 \le i \le n_w\right\}$$
$$y_t = \left\{w_t^i | r_t^i = 1 \text{ for } 0 \le i \le n_w\right\}$$

with a probability density

$$h(z_t | r_{0:t}, y_{0:t}) = p(Z_t \in dz_t | R_{0:t} = r_{0:t}, Y_{0:t} = y_{0:t})$$

The density h represents our knowledge of the non-response mechanism, later we will show that we can use that knowledge to improve our knowledge of the underlying state X.

B. Multiple Imputations Particle Filter

Consider the following probability distributions of interest:

$$\eta_{i|k:j} \triangleq p(X_i \in dx_i | Y_k = y_k, \ldots, Y_j = y_j)$$
$$\psi_t \triangleq p(X_t \in dx_t | Y_{0:t} = y_{0:t}, R_{0:t} = r_{0:t})$$

The probability density ψ_t can be thought of as the posterior probability density which combines the data from observations and non-response while the family of probability densities $\eta_{i|k:j}$ describe the probability distribution of the state given only a set of observations. For notational convenience, we write it as $\eta_{t|t} = \eta_{t|t:t}$. We are interested in obtaining the density ψ_t so that we can compute estimates such as MAP and MMSE. The distributions ψ_t and $\eta_{t|t}$ are related by the following expression:

$$\psi_t = \int \eta_{t|t} h(z_t | y_{0:t}) dz_t \tag{16}$$

Standard Bayesian filtering theory gives us the following equation:

$$\eta_{t|t} = \frac{g(w_t|x_t)\eta_{t|t-1}}{\int g(w_t|x_t)\eta_{t|t-1}}$$

Substituting this expression into (16), we have

$$\psi_t = \int h(z_t | w_{0:t-1}, y_t) \left(\frac{g(w_t|x_t)\eta_{t|t-1}}{\int g(w_t|x_t)\eta_{t|t-1}} \right) dz_t \tag{17}$$

This equation cannot be generally solved except for very specific models such as linear Gaussian. Thus we resort to approximation strategies. Consider a set of values or particles distributed approximately according to $\eta_{t-1|t-1}$, we sample $x_t^i \sim K(x_t | x_{t-1}^i)$ as the standard bootstrap procedure. Thus, the particles are distributed approximately according to $\eta_{t|t-1}$ and we have the following empirical distributions:

$$\eta_{t|t-1}^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_t^i}$$
$$\eta_{t|t}^N = \sum_{i=1}^{N} w_t^i \delta_{x_t^i}$$

where

$$w_t^i = \frac{g(w_t | x_t^i)}{\sum_{i=1}^N g(w_t | x_t^i)}$$

Now we wish to incorporate the additional knowledge of the non-response. This can be done by substituting the empirical distribution $\eta_{t|t}^N$ into (16) in place of the true distribution $\eta_{t|t}$. Having done so, we get the Monte Carlo approximation

$$\psi_t^N = \int h(z_t | w_{0:t-1}, y_t) \eta_{t|t}^N dz_t$$

The integral above cannot be evaluated explicitly in general, therefore, we again resort to an approximation scheme. For simplicity, we apply a naive Monte Carlo procedure to approximate the integral, let $z_t^j \sim h(z_t | w_{0:t-1}, y_t)$ for $0 \le j \le M$, then we can write

$$\psi_t^{N,M} = \eta_{t|t}^N \frac{1}{M} \sum_{j=1}^{M} \delta_{z_t^j}$$

Therefore, we have an approximation $\psi_t^{N,M}$ to the desired probability density ψ_t in terms of two particle sets $\{w_y^i, x_y^i\}_{i=1}^N$ and $\{z_t^j\}_{j=1}^M$.

C. Almost Sure Convergence of the MIPF

The integral expression (17) can be thought of a sequence of three transformations whose overall result is taking a probability density ψ_{t-1} to the next one in time ψ_t. We can write this sequence of maps as $\eta_{t|t-1} \to \eta_{t|t} \to \psi_t$ where $\eta_{t|t-1}$ and $\eta_{t|t}$ are intermediate distributions due to the particle filtering. We wish to prove the convergence of this sequence of mappings. We begin with an abstract argument and later show how it is related to the algorithm at hand. Consider a metric space (E, d) and let a_t, b_t, d_t be sequences of continuous functions $a_t, b_t, d_t : E \to E$ indexed by $t \in N$. In addition, let

$$l_t \triangleq d_t \circ a_t \circ b_t$$

where $(a \circ b)(x) = a(b(x))$. Now, we introduce two, not necessarily continuous, perturbation functions $f^M, c^N : E \to E$ in the following way: We assume that as N and M increase, c^N and f^M converge to the identity function respectively. We perturb l_t using these two functions in the following way

$$l_t^{N,M} = f^M \circ d_t \circ c^N \circ a_t \circ c^N \circ b_t$$

Let e_M and e_N be a sequence of elements in the metric space E indexed by M and N, respectively and let $e \in E$ denote a single element of E. We require that f^M, c^N satisfy the following conditions for all such sequences e_M, e_N:

$$\lim_{N\to\infty} e_N = e \Rightarrow \lim_{N\to\infty} c^N(e_N) = e \tag{18}$$
$$\lim_{M\to\infty} e_M = e \Rightarrow \lim_{M\to\infty} f^M(e_M) = e \tag{19}$$

We need the following lemma from [22]:

Lemma 1: Let a_t, b_t, k_t and c^N be as defined above. Then if c^N satisfies condition in (18), we have

$$\lim_{N\to\infty} k_t^N = k_t$$

Moreover, k_t^N satisfies

$$\lim_{N\to\infty} e_N = e \Rightarrow \lim_{N\to\infty} k_t^N(e_N) = k_t(e)$$

then we can prove the following lemma:

Lemma 2: Let d_t, a_t, b_t and f^M, c^N be as defined above, then if conditions given in (18)–(19) are satisfied we have

$$\lim_{N\to\infty} e_N = e \Rightarrow \lim_{M\to\infty} f_M(\lim_{N\to\infty}(d_t \circ k_t^N)(e_N)) = l_t(e)$$

Let $l_t^{N,M} = f^M \circ d_t \circ k_t^N$ where $k_t^N = c^N \circ a_t \circ c^N \circ b_t$ and $k_t = a_t \circ b_t$. Then by Lemma 2, we know that

$$\lim_{N\to\infty} e_N = e \Rightarrow \lim_{N\to\infty} k_t^N(e_N) = k_t(e)$$

since d_t is continuous,

$$\lim_{N\to\infty} k_t^N(e_N) = k_t(e) \Rightarrow \lim_{N\to\infty}(d_t \circ k_t^N)(e_N) = (d_t \circ k_t)(e).$$

Now set $\widehat{e} = (d_t \circ k_t)(e)$ and $\widehat{e}_M = \lim_{N\to\infty}(d_t \circ k_t^N)(e_N)$ for all M then clearly, $\widehat{e}, \widehat{e}_M \in E$ and $\lim_{M\to\infty}\widehat{e}_M = \widehat{e}$. Using the condition in (19), we have

$$\begin{aligned}\lim_{M\to\infty}\widehat{e}_M = \widehat{e} &\Rightarrow \lim_{M\to\infty} f_M(\widehat{e}_M) = \widehat{e}\\ &\Rightarrow \lim_{M\to\infty} f_M(\lim_{N\to\infty}(d_t \circ k_t^N)(e_N))\\ &= (d_t \circ k_t)(e) = l_t(e)\end{aligned}$$

Now we specialize this abstract discussion to the domain of multiple imputation particle filters. Consider the following definition of convergence: Let H be a Hilbert space with inner product $\langle,\rangle$. A sequence $\{x_t\}, x_t \in H$ is weakly convergent to $x_0 \in H$ if for every element $f \in H$ we have

$$\lim_{t\to\infty}\langle x_t, f\rangle = \langle x_0, f\rangle$$

Let $E = \mathcal{P}^{n_x}$ be the space of probability measures with the standard inner product, then E is a Hilbert space. We endow this space with the topology of weak convergence as defined above. Let ν be an arbitrary probability measure and φ is any continuous bounded function, then we define

$$\begin{aligned}(a_t(\nu), \varphi) &= \left(\int_{R^{n_x}} g(y_t|x_t)\nu(dx_t)\right)^{-1}\int_{R^{n_x}}\varphi(x_t)g(y_t|x_t)\\ &\quad \nu(dx_t)\\ (b_t(\nu), \varphi) &= \int_{R^{n_x}}\varphi(x_t)K(dx_t|x_{t-1})\nu(dx_{t-1})\end{aligned}$$

Assuming K is feller, i.e., $x \to K(x,\cdot)$ is continuous and the function g is continuous, bounded and strictly positive, then a_t, b_t can be shown to be continuous. Also, we define the function $d_t : \mathcal{P}^{n_x} \to \mathcal{P}^{n_x}$

$$\begin{aligned}d_t(\nu) &\triangleq \int_{R^{n_z}}\nu(dx_t|z_t, y_t)H(dz_t)\\ (d_t(\nu), \varphi) &= \int_{R^{n_x}}\int_{R^{n_z}}\varphi(x_t)\nu(dx_t|z_t, y_t)H(dz_t)\\ &= (H, \nu\varphi)\end{aligned}$$

We assume d_t to be a continuous operator, this is a reasonable assumption since we can interpret that requirement as the fact that adding or removing observers will influence the quality of observation in a continuous manner (where the continuity is in the function space sense). Moreover, we define the perturbation functions c^N and f^M as:

$$c^N = \frac{1}{N}\sum_{i=1}^{N}\delta_{V_i}$$

$$f^M = \frac{1}{M}\sum_{j=1}^{M}\delta_{W_j}$$

where V_i are $i.i.d.$ random variables with common distribution ν_c and W_j are $i.i.d.$ random variables with common distribution ν_f. We have the following lemma:

Lemma 3: If c^N and f^M are as defined above, then they satisfy conditions (18)–(19) almost surely. If we consider the empirical measure $\psi_t^{N,M}$, it is easy to see that

$$\psi_t^{N,M} = f^M \circ d_t \circ c^N \circ a_t \circ c^N \circ b_t(\psi_{t-1}^{N,M}) = k_t^{N,M}(\psi_{t-1}^{N,M})$$

Thus, we have the following theorem:

Theorem 1: Assuming the transition kernel K is feller and g is bounded, continuous and strictly positive, then almost surely $\psi^{N,M}$ is convergent to ψ_t as follows:

$$\lim_{N\to\infty}\lim_{M\to\infty}\psi_t^{N,M} = \psi_t$$

where the convergence is in the weak sense. Using the definitions above we have

$$\eta_{t|t}^N = (c^N \circ a_t \circ c^N \circ b_t)(\eta_{t-1|t-1}^N) = k_t^N(\eta_{t-1|t-1}^N)$$

and if $\lim_{N\to\infty}\eta_{t-1|t-1}^N = \eta_{t-1|t-1}$ then $\lim_{N\to\infty}\eta_{t|t}^N = \eta_{t|t}$. Also, we see that

$$\psi_{t|t}^{N,M} = (f^M \circ d_t \circ k_t^N)(\eta_{t-1|t-1}^N)$$

then by Lemmas 2 and 3 we have,

$$\lim_{M\to\infty}(f^M\lim_{N\to\infty}(d_t \circ k_t^N)(\eta_{t-1|t-1}^N)) = d_t \circ a_t \circ b_t = \psi_t$$

Subsequently, we can write

$$\lim_{N\to\infty}\lim_{M\to\infty}\psi_t^{N,M} = \psi_t.$$

VI. Performance Analysis

In this section, we analyze the performance of MIPF through simulation results. We consider the state and observation equations given in the non-linear model used in [6] and utilize them to generate data. The model in [6] is chosen as it is well known and has been used in other existing literature such as [26]. We remove a varying number of samples from these data randomly and apply MIPF to estimate the non-linear state. We further apply techniques in [12] and [13], as well as the EM algorithm to get non-linear state estimation with missing data and compare their performance to MIPF. We make a comparison of performance by comparing the results obtained by all these methods. In addition, we compare the results with those obtained using a PF applied to the complete data. We present figures with one set of simulation results as well as give results with root mean square averaged over several iterations.

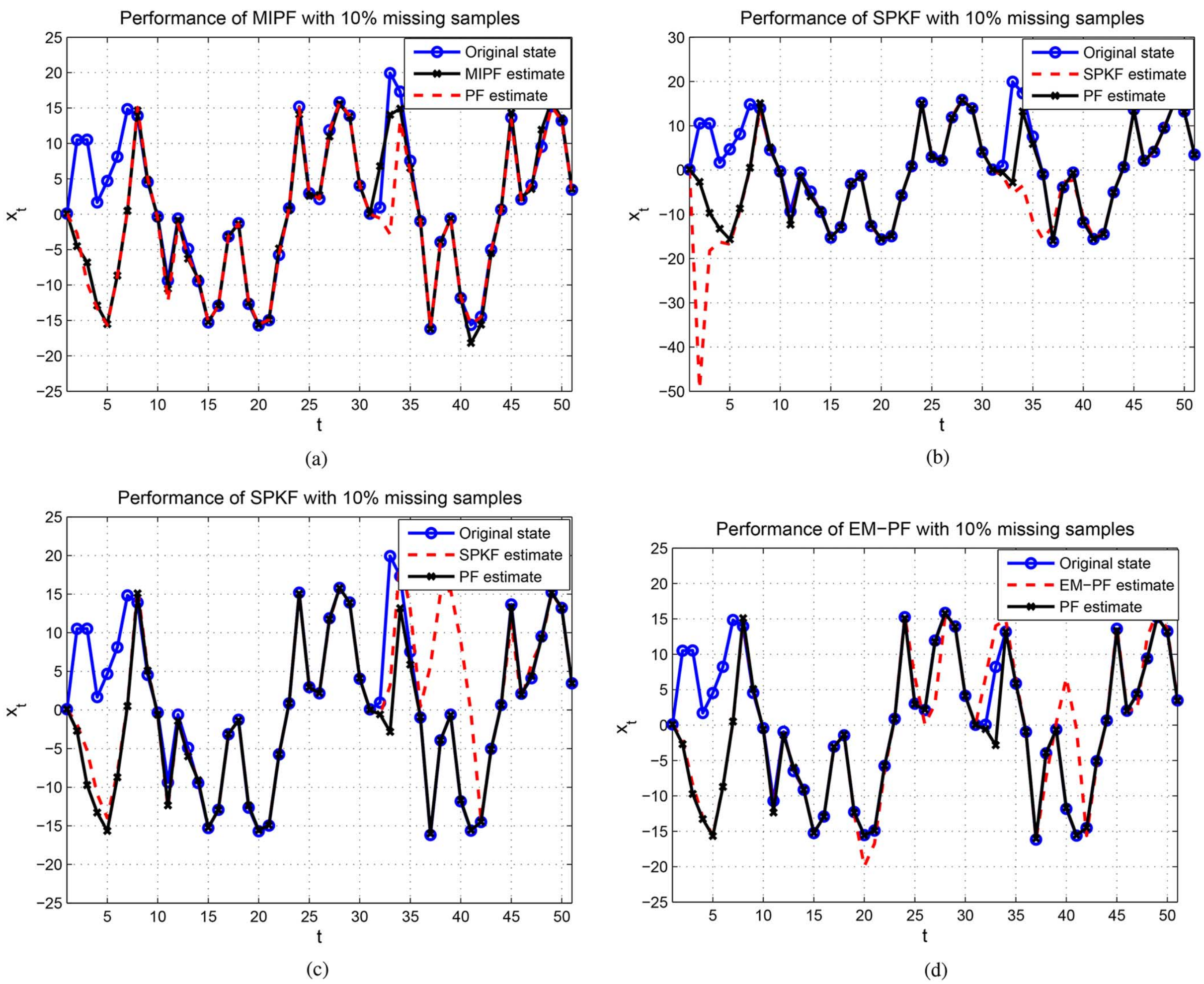

Fig. 1. Simulation results with 10% missing data. The results with PF are without any missing data and are used as a benchmark. SPKF shows divergent behavior, the results shown here with SPKF are the best results. (a) Results with MIPF. (b) Results with EKF. (c) Results with SPKF. (d) Results with EM-PF.

A. Non-Linear Model

We use the following non-linear model that represents the nonstationary growth model at tth instance [6]:

$$x_t = 0.5x_{t-1} + 25x_{t-1}/(1 + x_{t-1}^2) + 8\cos(1.2(t-1)) + w_t \tag{20}$$

$$y_t = x_t^2/20 + v_t \tag{21}$$

The state evolves according to (20), whereas the observations are made according to (21). The process and measurement noise are considered to be Gaussian with mean 0 and variance of 10 and 1, respectively, i.e., $w_t \sim \mathcal{N}(0, 10)$ and $v_t \sim \mathcal{N}(0, 1)$.

B. Results and Explanation

We simulate data for a total number of $T = 50$ iterations by varying t from 1 to 50 in the non-linear model. We first apply PF to complete data for non-linear state estimation and consider 30 particles, i.e., $N = 30$. Subsequently, we remove data randomly and apply MIPF to the data with missing samples. Note that in general, multiple imputations do not depend on the type of missing data mechanism and imputed data may be valid under any mechanism. The simulation examples in this section mainly present missing completely at random (MCAR) case. This is done for the ease of data simulation and in addition as MCAR implies MAR [28], we can use the results of using MIPF with MCAR data to examine the performance of the proposed approach. The condition of the missing data mechanism being independent of the missing observations still holds in the case of MCAR. Furthermore, it is known that multiple imputations can be used for MCAR [29] data and using MIPF for MCAR data can improve the estimation performance as we impute missing values based on the system model.

Please note that we use consider MCAR case as it is easier to simulate and MIPF can be applied to MCAR data. Using PF only on available observations and ignoring the missing observations may give unbiased results for MCAR case, as it is known that an estimator that uses only the available data can give an unbiased estimate for such a case. However, it is also known that by ignoring the missing data, especially if the missing data percentage compared to the available data is high, statistical power is reduced as the estimation is carried out on part of the data [30]. In addition, a PF that only uses the available data will not estimate the states for the instants where the data are not available. On the other hand, MIPF generates the missing data and tries to compensate for the uncertainty of missing observations through multiple imputations and averaging of the results. MIPF also estimates the states for those instants where the observations are absent.

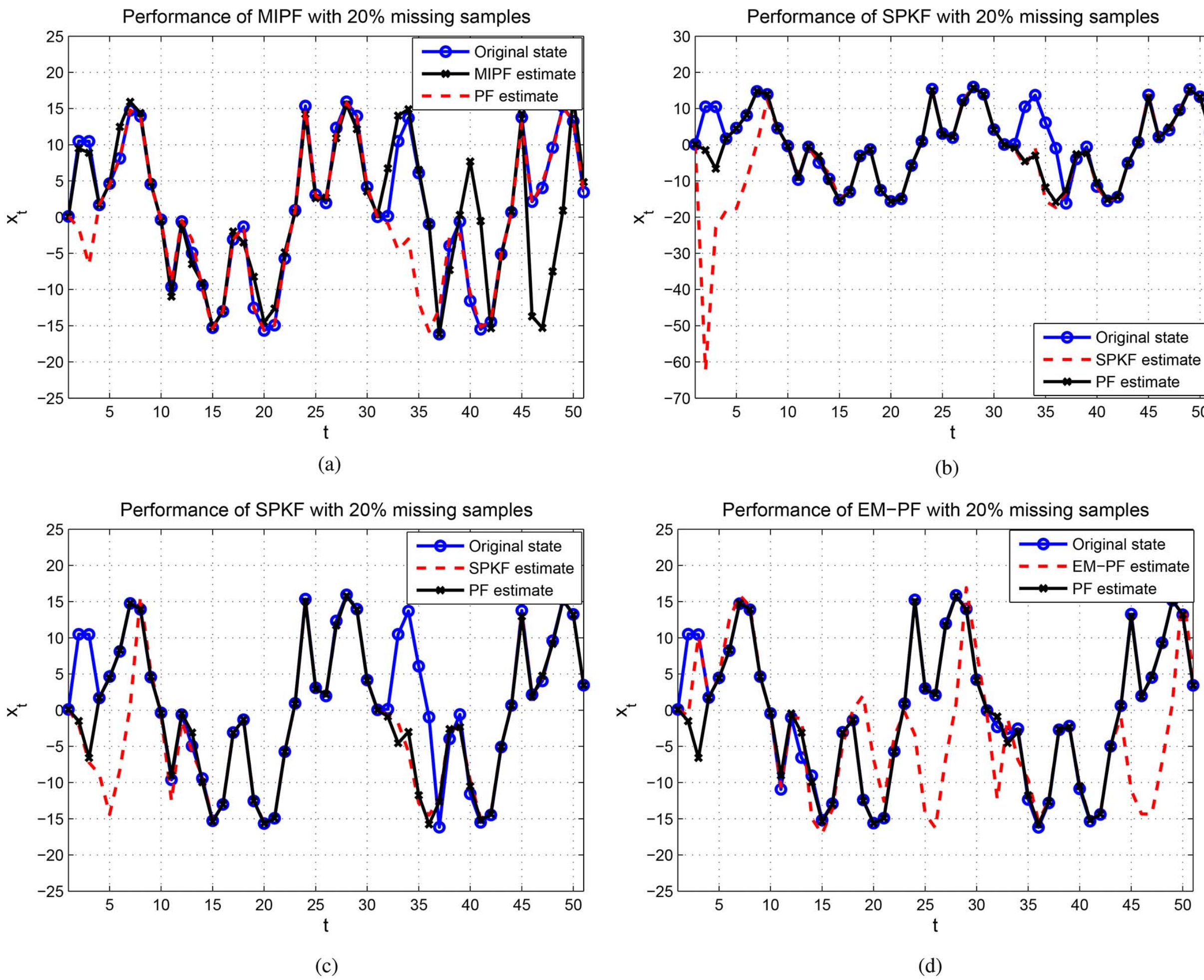

Fig. 2. Simulation results with 20% missing data. The results with PF are without any missing data and are used as a benchmark. SPKF shows divergent behavior, the results shown here with SPKF are the best results. (a) Results with MIPF. (b) Results with EKF. (c) Results with SPKF. (d) Results with EM-PF.

For MIPF, we consider 30 particles and 50 imputations, i.e., $N = 30$ and $M = 50$. One set of results of the non-linear state estimation using MIPF is shown in Figs. 1–3, with 10%, 20% and 50% missing samples in Figs. 1(a), 2(a) and 3(a), respectively. These figures show the true state, the estimated state by PF with complete data and the estimated state using MIPF with missing data. We compare the results of applying PF to the complete data and MIPF to missing data to verify if the MIPF can estimate the non-linear states even if data are missing. It can be seen that the performance of MIPF with 10% and 20% missing data is similar to that of the PF, which shows the effectiveness of MIPF in dealing with missing data. With 50% missing data, the difference between the results using PF and MIPF is more, which shows that with 50% missing data, the performance of MIPF degrades.

Next we compare the results of MIPF with existing techniques in [12] and [13]. The techniques in [12] and [13] use modifications of EKF and SPKF to estimate non-linear state with missing data. In addition, we compare with using the EM algorithm to estimate missing data and then apply PF to the data for non-linear state estimation. When using the EM algorithm, we first generate all the data, then remove varying number of samples from the data and subsequently estimate the mean and covariance of the data using the remaining data. The missing values are then generated using the estimated mean and covariance and then PF is applied to the data for non-linear state estimation. Results with 10%, 20% and 50% missing samples for [12], [13] and the EM algorithm are shown in Figs. 1(b), 2(b), 3(b), 1(c), 2(c), 3(c), 1(d), 2(d) and 3(d) respectively. It can be observed by comparing with Figs. 1(a), 2(a) and 3(a) that MIPF in general performs better than the other techniques. The performance of the EM algorithm is the second best, followed by SPKF and EKF. It can be seen that EKF and SPKF show some large differences with the true state and estimated state using PF, which indicates that these methods are sometimes not able to carry out estimation when data are missing. Compared to these techniques, MIPF does not show any large differences.

To further analyze and compare the performance of MIPF, we calculate root mean square error (RMSE) between the estimated and actual states as follows:

$$RMSE = \frac{\sum_{t=1}^{T}(x_t - \hat{x}_t)^2}{T} \tag{22}$$

We average the RMSE values obtained from a total number of 500 simulation experiments. We plot the results that show a comparison with other techniques in Figs. 4 and 5. We first discuss the results using PF with complete data, and using MIPF, EKF and SPKF with missing data shown in Fig. 4. It can be

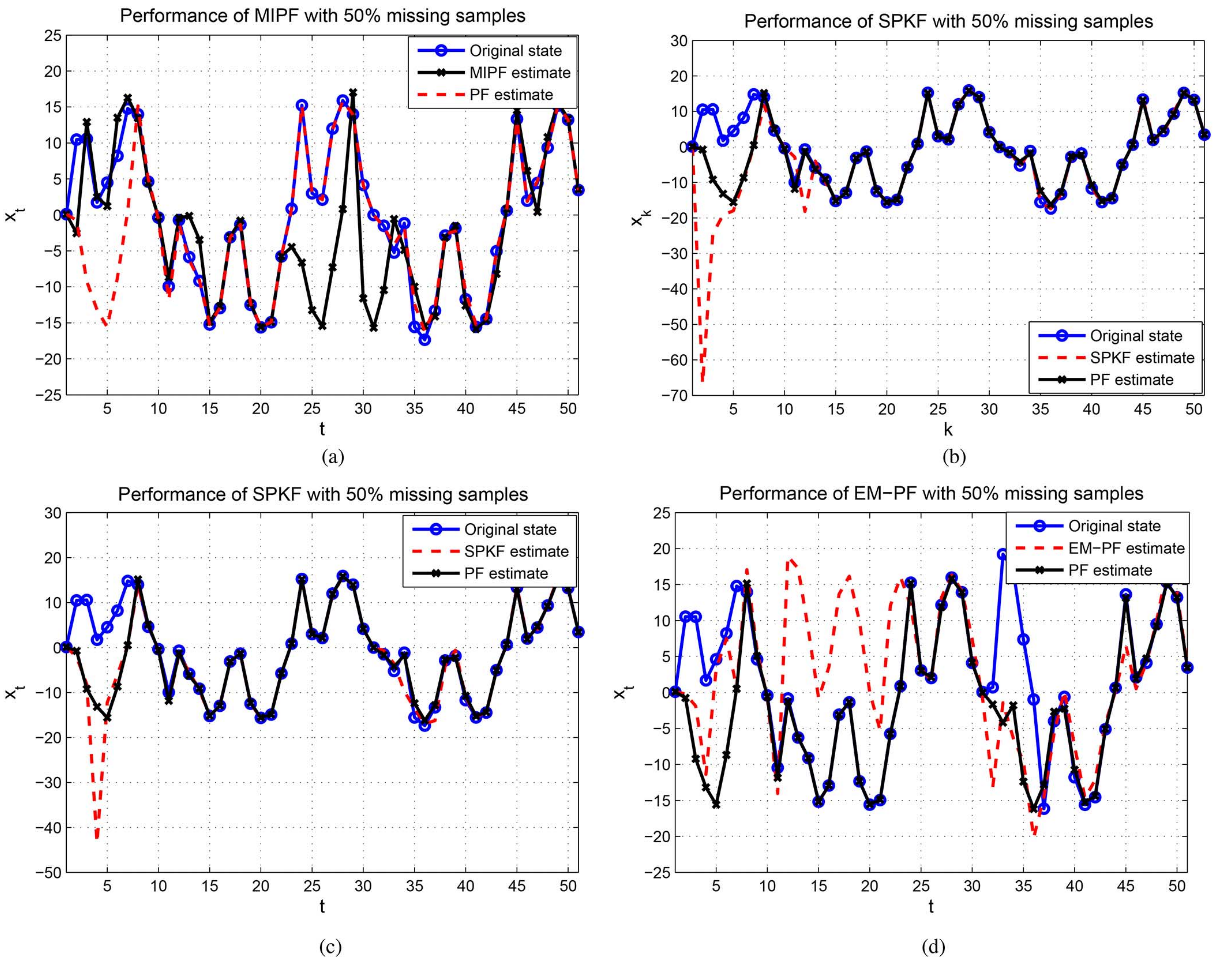

Fig. 3. Simulation results with 50% missing data. The results with PF are without any missing data and are used as a benchmark. SPKF shows divergent behavior, the results shown here with SPKF are the best results. (a) Results with MIPF. (b) Results with EKF. (c) Results with SPKF. (d) Results with EM-PF.

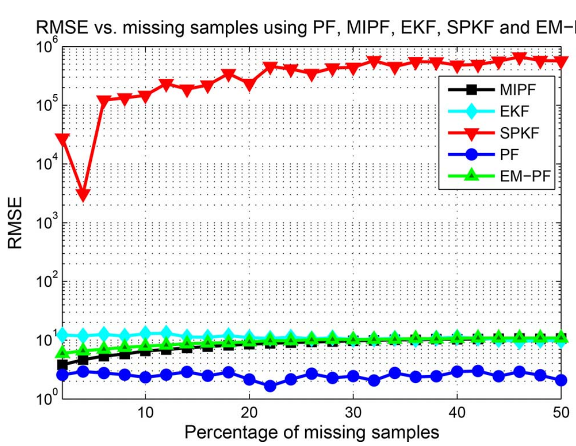

Fig. 4. Comparison of RMSE between PF, MIPF, EKF, SPKF and EM-PF. The results with PF are without any missing data and are used as a benchmark. SPKF shows divergent behavior.

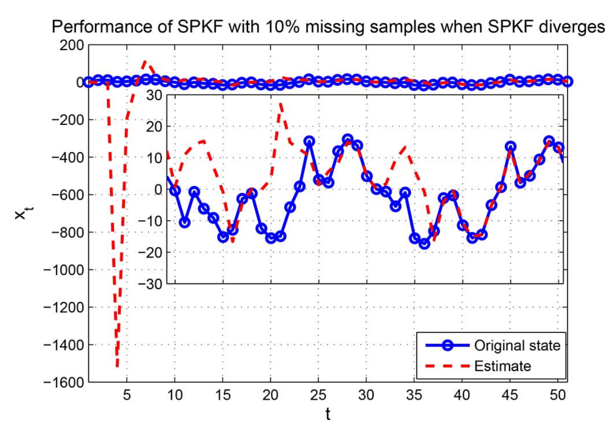

Fig. 5. Simulation results with 10% missing data showing divergence of SPKF.

seen that RMSE for MIPF is lower than EKF till 30% missing data and after that the performance of both filters becomes very similar, the maximum RMSE is approximately 10. It can be further seen that MIPF performs better than EM-PF. The RMSE increases for both MIPF and EM-PF with increasing number of missing samples. However, the rate of increase of RMSE decreases with increasing number of missing samples and reduces to 0 around 40% missing samples. EM-PF shows reasonable performance; however, it requires the use of full data for parameter estimation.

The error for SPKF is very high. In fact, we encountered two issues with using EKF and SPKF: We found that EKF and SPKF would sometimes diverge where there were missing data, the latter diverging more often than the former. SPKF also suffers from the problem that sometimes the Cholesky decomposition of the covariance matrix will not result in a semi-definite positive matrix. Therefore, we only considered those cases,

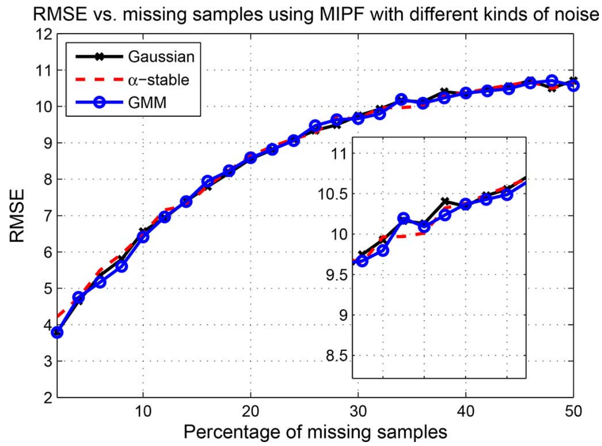

Fig. 6. Comparison of RMSE using MIPF with different kinds of noise. The parameters of α-stable noise are: 1.5, 0, 0.5, 0 and the GMM consists of two distributions with variance 1 and 2, and equal weights, i.e., 0.5.

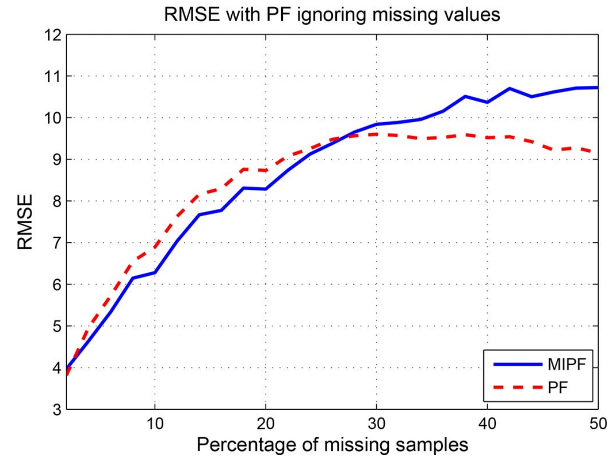

Fig. 7. Comparison of RMSE between MIPF and PF that ignores missing values.

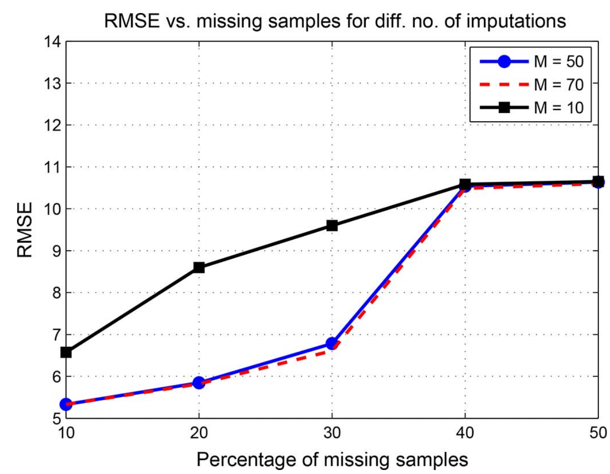

Fig. 8. RMSE with different imputations.

where the result of the decomposition was a semi-definite matrix. When the matrix is not semi-definite, we would re-run the simulation. An example showing the divergence of SPKF is shown in Fig. 5, where the divergence of SPKF can clearly be seen from very high differences between the estimated and the true states. From Figs. 1–3, we can also observe that when the SPKF does not diverge, its performance is better than EKF.

Note that we consider Gaussian measurement noise in the simulation section as we are using an already existing model. However, our method does not depend on Gaussian noise. RMSE with other commonly used alpha stable and Gaussian mixture model (GMM) distributions [27] as measurement noise are shown in Fig. 6 with varying percentage of missing samples. As it can be seen, the performance of MIPF with different kinds of measurement noise is similar and its performance is not dependant on the type of measurement noise.

We show simulation results with a PF ignoring the missing data, i.e., whenever the data are missing, the PF only updates the particles according to the observation equation and does not update the weights. Moreover, the PF does not estimate the states for the time instants where the observations are not available. The performance in terms of RMSE for both methods can be seen in Fig. 7. Note that as the PF does not estimate the states for the time instants when the observations are missing, therefore, for the PF, RMSE is measured only for the states that can be obtained from the observed data. Unlike the case with complete data, RMSE for PF is more than that of MIPF and also increases for PF with increasing percentage of missing samples. This further reinforces the utility for MIPF: the imputed data values not only replace the missing values in the observation, but also help in better state estimation. This can be seen by the performance of the PF, where only the available observations were used and the RMSE is more than that obtained from the MIPF. After 30% missing samples, the PF has lower RMSE than MIPF as the performance of MIPF degrades similar to Fig. 4.

We vary the number of imputations and plot the RMSE values in Fig. 8. We average the RMSE over 500 simulation instances. We vary the number of imputations as $M = 10, M = 50$ and $M = 70$. It can be seen that the RMSE for $M = 10$ is much higher than RMSE for the other two cases. Furthermore, it can be noticed that $M = 50$ and $M = 70$ have similar performance. The performance of $M = 10$ after 30% missing samples is almost the same like the other two cases. The reason is that as the number of missing samples increases, the performance of MIPF degrades and the estimates have large differences. Once the performance of MIPF degrades beyond a certain point, the increasing number of imputations has very little effect on performance improvement.

VII. Conclusion

In this paper, we applied the multiple imputations particle filter (MIPF) to non-linear state estimation problem with missing data. We showed that it can be used to deal with missing data by using imputations to provide a replacement for the missing data. We carried out a convergence analysis and showed that the MIPF is almost surely convergent. Simulation examples compared the performance of the MIPF with that of a particle filter applied to the complete data and showed that the MIPF performed well in the absence of data. We also compared the MIPF with several existing techniques such as variants of extended Kalman filter and sigma point Kalman filter, and expectation maximization algorithm that further underlined the superior performance of the MIPF for dealing with missing data.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their suggestions to help improve the quality of this paper.

REFERENCES

[1] M. Di, E. Joo, and L. Beng, "A comprehensive study of Kalman filter and extended Kalman filter for target tracking in wireless sensor networks," in *Proc. IEEE SMC*, 2008, pp. 2792–2797.

[2] A. Paul and E. Wan, "RSSI-based indoor localization and tracking using sigma-point Kalman smoothers," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 5, pp. 860–873, Oct. 2009.

[3] R. Merwe, E. Wan, S. Julier, A. Bogdanov, G. Harvey, and J. Hunt, "Sigma-point Kalman filters for nonlinear estimation and sensor fusion: Applications to integrated navigation," in *Proc. AIAA Guidance Nav. Control Conf.*, 2004.

[4] H. Tanizaki, *Nonlinear Filters: Estimation and Applications*. New York, NY, USA: Springer, 1996.

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

[6] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *Proc. IEE-F*, 1993, vol. 140, no. 2, pp. 107–113.

[7] P. Djuric, J. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. Bugallo, and J. Miguez, "Particle filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, Sep. 2003.

[8] H. Zhao and Z. Wang, "Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended Kalman filter for data fusion," *IEEE Sens. J.*, vol. 12, no. 5, pp. 943–953, May 2012.

[9] P. Pérez, J. Vermaak, and A. Blake, "Data fusion for visual tracking with particle," *Proc. IEEE*, vol. 92, no. 3, pp. 495–513, Mar. 2004.

[10] S. Mohamed and S. Nahavandi, "Optimal multisensor data fusion for linear systems with missing measurements," in *Proc. IEEE SOSE*, 2008, pp. 1–4.

[11] Y. Zhou, "A Kalman filter based registration approach for multiple asynchronous sensors," Tech. Rep., Ottawa, ON, Canada, , 2003 [Online]. Available: http://cradpdf.drdc.gc.ca/PDFS/unc17/p520870.pdf

[12] L. Yan, D. Zhou, M. Fu, and Y. Xia, "State estimation for asynchronous multirate multisensor dynamic systems with missing measurements," *IET Signal Process.*, vol. 4, no. 6, pp. 728–739, 2010.

[13] L. Yan, B. Xiao, Y. Xia, and M. Fu, "State estimation for asynchronous multirate multisensor nonlinear dynamic systems with missing measurements," *Int. J. Adapt. Control Signal Process.*, vol. 26, no. 6, pp. 516–529, 2012.

[14] E. Wang, J. Silva, and L. Carin, "Compressive particle filtering for target tracking," in *Proc. IEEE SSP*, 2009, pp. 233–236.

[15] Y. Chi, A. Pezeshki, L. Scharf, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[16] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[17] J. Schafer, *Analysis of Incomplete Multivariate Data*. Boca Raton, FL, USA: Chapman Hall/CRC, 1997.

[18] Y. Mustafa, V. Tolpekin, and A. Stein, "Application of the expectation maximization algorithm to estimate missing values in Gaussian Bayesian network modeling for forest growth," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1821–1831, May 2012.

[19] A. Housfater, X. Zhang, and Y. Zhou, "Nonlinear fusion of multiple sensors with missing data," in *Proc. ICASSP*, 2006, pp. 961–964.

[20] S. Imtiaz *et al.*, "Estimation of states of nonlinear systems using a particle filter," in *Proc. ICIT*, 2006.

[21] X.-P. Zhang, A. S. Khwaja, J.-A. Luo, A. S. Housfater, and A. Anpalagan, "Convergence analysis of multiple imputations particle filters for dealing with missing data in nonlinear problems," in *Proc. ISCAS*, 2014.

[22] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Trans. Signal. Process.*, vol. 50, no. 3, pp. 736–746, Mar. 2002.

[23] D. Rubin, *Multiple Imputation for Nonresponse in Suverys*. New York, NY, USA: Wiley, 1987.

[24] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statist. Comput.*, vol. 10, no. 2, pp. 197–208, 2000.

[25] J. Liu, A. Kong, and W. Wong, "Sequential imputations and Bayesian missing data problems," *J. Amer. Statist. Assoc.*, vol. 89, no. 425, pp. 278–288, 1994.

[26] F. Lindsten, "An efficient stochastic approximation EM algorithm using conditional particle filter," in *Proc. ICASSP*, 2013.

[27] H.-Q. Liu and H.-C. So, "Target tracking with line-of-sight identification in sensor networks under unknown measurement noises," *Progress Electromagn. Res.*, vol. 97, pp. 373–389, 2009.

[28] R. Mislevy and P.-K. Wu, "Missing responses and IRT ability estimation: Omits, choice, time limits, and adaptive testing," ETS Research Report Series, 1996. [Online]. Available: http://www.ets.org/Media/Research/pdf/RR-96-30.pdf

[29] D. Bouhlila and F. Sellaouti, "Multiple imputation using chained equations for missing data in TIMSS: A case study," Large-scale Assessments in Education 2013, 1:4 doi:10.1186/2196-0739-1-4.

[30] P. McKnight, K. McKnight, S. Sidani, and A. Figueredo, *Missing Data: A Gentle Introduction*. New York, NY, USA: Guilford, 2007.

Xiao-Ping Zhang received B.S. and Ph.D. degrees from Tsinghua University in 1992 and 1996, respectively, both in electronic engineering. He holds an M.B.A. in finance, economics and entrepreneurship with Honors from the University of Chicago Booth School of Business, Chicago, IL.

Since Fall 2000, he has been with the Department of Electrical and Computer Engineering, Ryerson University, where he is now Professor, Director of Communication and Signal Processing Applications Laboratory (CASPAL). He has served as Program Director of Graduate Studies. He is cross appointed to the Finance Department at the Ted Rogers School of Management at Ryerson University. His research interests include statistical signal processing, multimedia content analysis, sensor networks and electronic systems, computational intelligence, and applications in big data, finance, and marketing. He is a frequent consultant for biotech companies and investment firms. He is cofounder and CEO for EidoSearch, an Ontario based company offering a content-based search and analysis engine for financial data.

Dr. Zhang is a registered Professional Engineer in Ontario, Canada, and a member of Beta Gamma Sigma Honor Society. He is the general chair for MMSP'15. He is the publicity chair for ICME'06 and program chair for ICIC'05 and ICIC'10. He served as guest editor for Multimedia Tools and Applications, and the International Journal of Semantic Computing. He is a tutorial speaker in ACMMM2011, ISCAS2013, ICIP2013, and ICASSP2014. He is Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE SIGNAL PROCESSING LETTERS and for *Journal of Multimedia*.

Ahmed Shaharyar Khwaja received the B.Sc. degree in electronic engineering from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan, and the Ph.D. and M.Sc. degrees in signal processing and telecommunications from the University of Rennes 1, Rennes, France.

He is currently a Postdoctoral Research Fellow with the WINCORE Lab, Ryerson University, Toronto, ON, Canada. His research interests include compressed sensing, remote sensing, optimization problems in wireless communication systems and smart grid.

Ji-An Luo received the B.Sc. and M.Sc. degrees in electrical engineering from Hangzhou Dianzi University, Hangzhou, China, in 2005 and 2008 respectively. In 2013, he received the Ph.D. degree at the State Key Lab of Industrial Control Technology, Zhejiang University, Hangzhou, China.

From January to May 2010, he was with the Key Lab of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Science, Beijing. Since November 2010, he has been a visiting Ph.D. student in the Communications and Signal Processing Applications Lab (CASPAL), Ryerson University, Toronto. Currently, he is working at Hangzhou Dianzi University as a Lecturer. His research interests are array signal processing, radar signal processing, and statistical signal processing.

Alon Shalev Housfater, author photograph and biography unavailable at the time of publication.

Alagan Anpalagan received the B.A.Sc. M.A.Sc. and Ph.D. degrees in electrical engineering from the University of Toronto, Canada. He joined the ELCE Department at Ryerson University in 2001 and was promoted to Full Professor in 2010. He served the department as Graduate Program Director (2004–09) and the Interim Electrical Engineering Program Director (2009–10). Dr. Anpalagan's industrial experience includes working for three years at Bell Mobility, Nortel Networks and IBM Canada. He directs a research group working on radio resource management (RRM) and radio access and networking (RAN) areas within the WINCORE Lab. His current research interests include 5G wireless systems, energy harvesting and green communications technologies, cognitive radio resource management, wireless cross layer design and optimization, cooperative communication, M2M and sensor communication, small cell and heterogeneous networks.

Dr. Anpalagan served as Editor for the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS (2012–14), IEEE COMMUNICATIONS LETTERS (2010–13), *Springer Wireless Personal Communications* (2011–13), and *EURASIP Journal of Wireless Communications and Networking* (2004–2009). He also served as Guest Editor for three special issues: EURASIP SI in Radio Resource Management in 3G+ Systems and SI in Fairness in Radio Resource Management for Wireless Networks, and MONET SI on Green Cognitive and Cooperative Communication and Networking. He co-authored three edited books, *Design and Deployment of Small Cell Networks*, Cambridge University Press (2014), *Routing in Opportunistic Networks*, Springer (2013), *Handbook on Green Information and Communication Systems*, Academic Press (2012).

Dr. Anpalagan served as TPC Co-Chair, IEEE Globecom'15: SAC Green Communication and Computing, IEEE WPMC'12 Wireless Networks, IEEE PIMRC'11 Cognitive Radio and Spectrum Management, IEEE IWCMC'11 Workshop on Cooperative and Cognitive Networks and IEEE CCECE'04/08. He served as Vice Chair, IEEE SIG on Green and Sustainable Networking and Computing with Cognition and Cooperation (2015–), IEEE Canada Central Area Chair (2012–14), IEEE Toronto Section Chair (2006–07), ComSoc Toronto Chapter Chair (2004–05), and IEEE Canada Professional Activities Committee Chair (2009–11). Dr. Anpalagan also completed a course on Project Management for Scientist and Engineers at the University of Oxford CPD Center. He is a registered Professional Engineer in the province of Ontario, Canada, Senior Member of IEEE and Fellow of the Institution of Engineering and Technology (FIET).