Particle Filter Approach to Nonlinear System Identification under Missing Observations with a Real Application

R.Bhushan Gopaluni [*]

[*] *Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, Canada V6T 1Z3 (Tel: 604 827 5668; e-mail: gopaluni@chml.ubc.ca).*

Thomas B. Schön [**]

[**] *Division of Automatic Control, Linköping University, SE-581 83 Linköping, Sweden, e-mail: schon@isy.liu.se*

Adrian G. Wills [***]

[***] *School of Electrical Engineering and Computer Science, University of New Castle, Australia*

Abstract: This article reviews authors' recently developed algorithm for identification of nonlinear state-space models under missing observations and extends it to the case of unknown model structure. In order to estimate the parameters in a state-space model, one needs to know the model structure and have an estimate of states. If the model structure is unknown, an approximation of it is obtained using radial basis functions centered around a *maximum a posteriori* estimate of the state trajectory. A particle filter approximation of smoothed states is then used in conjunction with expectation maximization algorithm for estimating the parameters. The proposed approach is illustrated through a real application.

1. INTRODUCTION

Nonlinear models are commonly used to describe the behavior of many processes. Process variables, typically, can be divided into latent variables (that are not measured) and measured variables. A combination of latent and measured variables can be elegantly used to represent the dynamic behavior of a nonlinear process in the following form,

$$x_{t+1} = f(x_t, u_t, \theta) + w_t$$
$$y_t = g(x_t, u_t, \theta) + v_t \tag{1}$$

where $x_t \in \mathbf{R}^n$ is the n-dimensional state vector, $u_t \in \mathbf{R}^s$ is the s-dimensional input vector, $y_t \in \mathbf{R}^m$ is the m-dimensional output or measurement vector, and w_t, v_t are independent and identically distributed Gaussian noise sequences of appropriate dimension and variances Q and R respectively, $\theta \in \mathbf{R}^p$ is a p-dimensional parameter vector and $f(.)$, $g(.)$ are some nonlinear functions that describe the dynamics of the process. The nonlinear functions $f(.)$ and $g(.)$ are usually obtained using physical laws such as energy and mass balance expressions for the process. However, often, due to the complexity of chemical processes, it is difficult to develop accurate and reliable nonlinear functions. This article reviews the authors' previously developed nonlinear identification algorithm for known functions $f(.)$ and $g(.)$, and extends it to an algorithm for approximation and estimation of $f(.)$ and $g(.)$ using a combination of radial basis functions and Expectation Maximization (EM) algorithm.

The complexity of the parameter estimation problem considered in this article arises due to unknown nonlinearities, and presence of unmeasured latent variables. If the latent variables are measured, then the model parameters can be estimated using a straightforward nonlinear least squares method (Ljung [1999]). If the process dynamic functions are linear, then any sub-space approach can be used (Van Overschee and Moor [1996]). On the other hand, if the process dynamic functions are known but nonlinear and latent variables are not measured, then approximate maximum likelihood approaches such as the one based on local linearization in (Goodwin and Agüero [2005]), and those based on particle filter approximation in (Gopaluni [2008], Schön et al. [2006]) can be used.

The algorithm presented in this article extends the one in authors' previous work on parameter estimation for known model structure (Gopaluni [2008]). The central idea is to find the parameter vector, θ, that maximizes the likelihood function of the observations, y_t. Due to the presence of

latent variables, x_t, and missing observations, it is difficult to develop this likelihood function. On the other hand, due to Markov property of latent variables, it is rather straightforward to develop a joint likelihood function of the latent and measured variables. Hence, expectation maximization, a maximum likelihood approach, that iteratively maximizes the likelihood of the observations by maximizing the joint likelihood function in each iteration is used. EM algorithm is implemented by iteratively finding the expected value of the joint likelihood function in the first step and maximizing it in the second step (Dempster et al. [1977]).

This approach using EM algorithm for parameter estimation poses two problems. A structure of the process model (or in other words, the functions $f(.)$ and $g(.)$), and the distribution of noise sequence is needed to develop the joint likelihood function. Moreover, since the process is nonlinear, the distributions of latent variables, x_t, and measurements are likely to be non-Gaussian even if Gaussian noise is assumed. As a result, the expected value of the joint likelihood function required in EM algorithm can not be analytically calculated. In this article, the authors' approach for nonlinear system identification for the case of known model structure is explained, and then extended to an approach that uses radial basis functions to approximate the process dynamics. The expected value of the joint likelihood function is then approximated using particle filters. The proposed approach is also extended to handle missing observations.

Many process variables in the chemical industry are measured irregularly due to physical and economical constraints. However, often times, a discrete model valid at a higher sample rate than most measurements is desired. Such a model can be estimated from irregular measurements by treating all unavailable measurements as missing observations. For linear systems, EM algorithm has been extensively used for parameter estimation under missing observations (Shumway and Stoffer [2000], Isaksson [1993]), and also applied in practice (Raghavan et al. [2006]). Other approaches for linear systems based on lifting techniques (Li et al. [2003]) and continuous time identification (Wang and Gawthrop [2001]) have also been reported. While the importance of estimating nonlinear processes under missing observations has long been recognized (Gudi et al. [1995], Tatiraju et al. [1999]), to the best knowledge of the authors, no work has been reported for nonlinear stochastic processes. The published work on parameter estimation for nonlinear systems treats only states as missing data. In this paper the EM algorithm is adapted to also handle nonlinear processes with missing observations.

2. EXPECTATION MAXIMIZATION ALGORITHM

Expectation Maximization is an elegant optimization algorithm that constructs a complete likelihood function including the hidden states and missing observations, and maximizes the likelihood function of observed data through iterations. A brief description of the EM algorithm is presented in this section to facilitate the development of the proposed algorithm in later sections.

For the state-space model described in this article, let $p(y_{1:T}|\theta)$ [1] denote the joint likelihood function of the observed data. The maximum likelihood estimate of the parameter vector is obtained by maximizing this observed data likelihood function. For certain classes of state-space models, such as linear systems, it is possible to derive an explicit expression for this joint density. However, for the model considered in this paper, it is difficult to develop such an expression. Instead, using the Markov property of the model it is straightforward to develop an expression for the complete (including states and observations) likelihood function, $p(x_{1:T}, y_{1:T}|\theta)$. In light of this feature of the Markovian state-space model, the joint probability density function of the states and observations is iteratively maximized to obtain a maximizing θ for $p(y_{1:T}|\theta)$. This maximization approach is called EM algorithm and can be summarized in four steps: (1). Choose an initial guess of the parameter vector $\theta_0 \in \Omega$. (2). Estimate the states given the parameter vector and the observations and evaluate $Q(\theta_i, \theta) = \int \log[p(x_{1:T}, y_{1:T}|y_{1:T}, \theta)]p(x_{1:T}|y_{1:T}, \theta_i)dx_{1:T}$. (3). Maximize $Q(\theta_i, \theta)$ with respect to θ. Call the maximizing value θ_{i+1}. (4). Repeat the above two steps until the change in parameter vector is within a specified tolerance level. The second step in the above algorithm is called E-step and the third step is called M-step. The likelihood function increases monotonically through these iterations. Due to the nonlinear nature of the dynamics it is not possible to analytically evaluate the Q-function. In the next section, an approximation of the Q-function and an approach to maximize it are presented.

3. THE Q FUNCTION

In this section an approximation of the Q function that can handle missing observations, is developed using the Markovian property of the state-space model.

3.1 Full Data Case

In the rest of this article, it is assumed that the inputs are known and all the density functions of the form $p(.|.,., u_{1:T})$ are denoted by $p(.|.,.)$ without explicitly showing the input dependence. Consider the case where all the observations $\{y_1, \cdots, y_T\}$ and the inputs $\{u_1, \cdots, u_T\}$ are available. Then, using the Markov property of the state space model, the joint density function of states and outputs can be written as

$$p(x_{1:T}, y_{1:T}|y_{1:T}, \theta) =$$
$$= p(x_1|y_{1:T}, \theta) \prod_{t=2}^{T} p(x_t|x_{t-1}, \theta) \prod_{t=1}^{T} p(y_t|x_t, \theta)$$

Performing the integrations in the expression for Q, the following form of Q function can be obtained

---

[1] $y_{1:T}$ denotes the set $\{y_1, \cdots, y_T\}$.

$$Q(\theta_i, \theta) = \int \log[p(x_1|y_{1:T}, \theta)]p(x_1|y_{1:T}, \theta_i)dx_1$$

$$+ \sum_{t=2}^{T} \int \log[p(x_t|x_{t-1}, \theta)]p(x_{t-1:t}|y_{1:T}, \theta_i)dx_{t-1:t}$$

$$+ \sum_{t=1}^{T} \int \log[p(y_t|x_t, \theta)]p(x_t|y_{1:T}, \theta_i)dx_t. \qquad (2)$$

From the above expression, notice that approximations of the density functions $p(x_1|y_{1:T}, \theta_i)$, $p(x_{t-1:t}|y_{1:T}, \theta_i)$, $p(x_t|y_{1:T}, \theta_i)$ would allow one to approximate the Q function.

3.2 Missing Data in Output

Suppose that only a portion of the output measurements at time instants $\{t_1, \cdots, t_\gamma\}$ are available and that they are not available at time instants $\{s_1, \cdots, s_\beta\}$. In other words only $\{y_{t_1}, \cdots, y_{t_\gamma}\}$ out of $\{y_1, \cdots, y_T\}$ are available. For notational simplicity, it is also assumed that y_1 and y_T are available. Then the Q function can be written as

$$Q(\theta_i, \theta) = \int \log[p(x_1|y_{t_1:t_\gamma}, \theta)]p(x_1|y_{t_1:t_\gamma}, \theta_i)dx_1$$

$$+ \sum_{t=2}^{T} \int \log[p(x_t|x_{t-1}, \theta)]p(x_{t-1:t}|y_{t_1:t_\gamma}, \theta_i)dx_{t-1:t}$$

$$+ \sum_{t=t_1}^{t_\gamma} \int \log[p(y_t|x_t, \theta)]p(x_t|y_{t_1:t_\gamma}, \theta_i)dx_t$$

$$+ \sum_{t=s_1}^{s_\beta} \int \log[p(y_t|x_t, \theta)]p(x_t, y_t|y_{t_1:t_\gamma}, \theta_i)dx_t dy_t \qquad (3)$$

In order to approximate the Q functions, approximations of the following density functions are required:
Full data case

(1) $p(x_t|y_{1:T}, \theta)$
(2) $p(x_{t-1}, x_t|y_{1:T}, \theta)$

Missing data case

(1) $p(x_t|y_{t_1:t_\gamma}, \theta)$
(2) $p(x_{t-1}, x_t|y_{t_1:t_\gamma}, \theta)$
(3) $p(x_t, y_t|y_{t_1:t_\gamma}, \theta)$ for $t \notin \{t_1, \cdots, t_\gamma\}$

Notice that the maximum dimensionality of the above density functions is $max(2n, n+m)$, and hence the accuracy of these density functions does not deteriorate with increase in the size of available measurements as is the case with the method suggested in Andrieu et al. [2004].

4. IDENTIFICATION WITH KNOWN MODEL STRUCTURE

The following particle approximations of the density functions identified above can be obtained using Bayes' rule and importance sampling.

$$p(x_1|y_{t_1:t_\gamma}, \theta) = \sum_{i=1}^{N} \bar{w}_{1|1}^{(i)}\delta(x_1 - x_1^{(i)})$$

$$p(x_{t-1}, x_t|y_{t_1:t_\gamma}, \theta) = \sum_{i=1}^{N} \bar{w}_{t,t-1}^{(i)}\delta(x_{t-1} - x_{t-1}^{(i)})\delta(x_t - x_t^{(i)})$$

$$p(x_t|y_{t_1:t_\gamma}, \theta) = \sum_{i=1}^{N} \bar{w}_{t|T}^{(i)}\delta(x_t - x_t^{(i)})$$

$$p(x_t, y_t|y_{t_1:t_\gamma}, \theta) = \sum_{i=1}^{N} \bar{w}_{t|x}^{(i)}\delta(x_t - x_t^{(i)})\delta(y_t - y_t^{(i)})$$

where $x_t^{(i)}$ represent state particles, $y_t^{(i)}$ represent particles of missing observations, $\bar{w}_{1|1}^{(i)}$, $\bar{w}_{t,t-1}^{(i)}$, $\bar{w}_{t|T}^{(i)}$, $\bar{w}_{t|x}^{(i)}$ are weights proportional to the corresponding density functions, and $\delta(.)$ is a dirac-delta function. These weights depend on the knowledge of the functions $f(.)$ and $g(.)$ or their respective approximations. Complete derivations of the above expressions for particle approximations of various density functions are presented in Gopaluni [2008]. By combining the equations for the filter, smoother and the joint density functions, one can approximate the Q function. Once an approximation of the Q function is available, it is possible to maximize it with respect to the parameter vector and obtain the next iterate of the EM algorithm. By substituting the particle approximations of the density functions in (3), approximate Q function can be written as

$$Q(\theta', \theta) \approx \sum_{i=1}^{N} \bar{w}_{1|1}^{(i)} \log[p(x_1^{(i)}|y_{t_1:t_\gamma}, \theta)] + \sum_{t=2}^{T}\sum_{i=1}^{N} \bar{w}_{t,t-1}^{(i)}$$

$$\log[p(x_t^{(i)}|x_{t-1}^{(i)}, y_{t_1:t_\gamma}, \theta)] + \sum_{t=t_1}^{t_\alpha}\sum_{i=1}^{N} \bar{w}_{t|T}^{(i)} \log[p(y_t|x_t^{(i)}, \theta)]$$

$$+ \sum_{t=s_1}^{s_\beta}\sum_{i=1}^{N} \bar{w}_{t|x}^{(i)} \log[p(y_t^{(i)}|x_t^{(i)}, \theta)]$$

$$(4)$$

Then the EM algorithm can be summarized as

(0). **Initialization**: Initialize the parameter vector to θ_0. Set $i = 0$. (1).**Expectation**: Evaluate the approximate Q function according to (4) using $\theta' = \theta_i$. (2). **Maximization**: Maximize the Q function with respect to θ and call the maximizing parameter, θ_{i+1}. Maximization can be performed using any standard optimization algorithm. Then set $\theta = \theta_{i+1}$. (3). **Iterate**: Repeat steps 1 and 2 until the change in parameter vector is within a specified tolerance level.

The maximization step is performed using standard nonlinear programming algorithms.

5. IDENTIFICATION WITH UNKNOWN MODEL STRUCTURE

As mentioned in the previous section, the weights in the particle approximations of various density functions depend on knowing the structure of state and measurement dynamic functions, namely $f(.)$ and $g(.)$. Hence, if the process dynamics are unknown, then an approximation

of the dynamics is needed to apply EM algorithm. It is well-known that any function can be approximated to an arbitrary degree of accuracy using basis functions such as radial basis functions. Therefore, the model in (1) is approximated using radial basis functions as follows:

$$x_{t+1} = \sum_{i=1}^{I_x} h_i \rho_i(x_t, u_t, c_i, \sigma_x) + A x_t + B u_t + w_t$$

$$y_t = \sum_{i=1}^{I_y} g_i \gamma_i(x_t, u_t, d_i, \sigma_y) + C x_t + D u_t + v_t$$

where $\rho_i(.,.)$ and $\gamma_i(.,.)$ are the radial basis functions centered around c_i and d_i with radii σ_x^2 and σ_y^2 [2] respectively, A, B, C, D are appropriate matrices that are used to capture any linear dynamics in the model. h_i and g_i are constant vectors of appropriate dimensions. I_x and I_y are the number of basis functions used in the state and observation equations. Theoretically, even linear dynamics in the process can be approximated if sufficiently large number of radial basis functions are used. In order to reduce the total number of parameters, and capture linear dynamics directly, linear terms involving the matrices A, B, C, and D are added. In this article, radial Gaussian basis functions of the following form are used:

$$\rho_i(x_t, u_t, c_i, \sigma_x) = e^{\left[-(\bar{x}_t - c_i)^T \sigma_x^{-2} (\bar{x}_t - c_i)\right]}$$

$$\gamma_i(x_t, u_t, d_i, \sigma_y) = e^{\left(-\left[\bar{x}_t - d_i\right]^T \sigma_y^{-2} (\bar{x}_t - d_i)\right)}$$

where $\bar{x}_t = [x_t \quad u_t]^T$ [3] is the concatenated vector of states and inputs. The parameter vector θ includes all the constant parameters in the above model that describe the process behavior, and is defined as $\theta = (\theta_l, \theta_{nl})$, where θ_l consists of all "linear" parameters, h_i, g_i, Q and R, and θ_{nl} consists of all "nonlinear" parameters, c_i, d_i, σ_x, σ_y.

Now that an approximate functional form of the state and measurement dynamic functions is available, the approximate Q function in (4) can be evaluated by estimating the appropriate weights. Unlike the case with known model structure, the maximization of Q function is performed in two steps using separable least squares. It is easy to notice that the parameters in θ_l enter the model linearly, while those in θ_{nl} enter the model nonlinearly. Hence, a two step procedure where the linear parameters are estimated in the first step using linear least squares, and the nonlinear parameters are estimated in the second step, using the previous estimate of linear parameters, through nonlinear least squares is proposed. The procedure is described below.

Step 1 Starting with an initial guess for the nonlinear parameter vector, θ_{nl}, the Q function is maximized with respect to θ_l. This maximization can be achieved through linear least squares. Define the following matrices,

---

[2] Please note that, for the sake of keeping the notation simple, it is assumed that all the basis functions in the state and measurement equations have identical radii. However, the development presented in this work is valid even if they are not identical.

[3] Please note that we have used T to denote length of data and also to denote transpose of a matrix. Usually the context makes this difference obvious.

$$\Omega_x = [h_1 \ h_2 \ \cdots h_{I_x} \ A \ B]$$
$$s_t = [I_1 \rho_1(x_t, u_t, c_1, \sigma_x) \ I_1 \rho_2(x_t, u_t, c_2, \sigma_x) \cdots$$
$$I_1 \rho_{I_x}(x_t, u_t, c_{I_x}, \sigma_x) \ x_t \ u_t]$$

where I_1 is a vector of ones of appropriate dimensions. Noticing that the Q function is convex and quadratic in Ω_x, through straightforward calculations, it can be shown that

$$\Omega_x = \left[\sum_{t=1}^{T} <x_t s_t^T>_{xx}\right] \left[\sum_{t=1}^{T} <s_t s_t^T>_{xx}\right]^{-1} \quad (5)$$

where $< . >_{xx}$ is used to denote integration with respect to the density function $p(x_{t-1}, x_t | y_{t_1:t_\gamma}, \theta)$. The integrations can be performed using particle approximation of the corresponding density function. The state co-variance can be shown to be

$$Q = \frac{1}{T} \sum_{t=1}^{T} \left\langle (x_{t+1} - \Omega_x s_t) x_{t+1}^T \right\rangle_{xx}.$$

Similarly, defining the matrices,

$$\Omega_y = \left[g_1 \ g_2 \ \cdots g_{I_y} \ C \ D\right]$$
$$r_t = [I_1 \gamma_1(x_t, u_t, d_1, \sigma_y) \ I_1 \gamma_2(x_t, u_t, d_2, \sigma_y) \cdots$$
$$I_1 \gamma_{I_y}(x_t, u_t, d_{I_x}, \sigma_y) \ x_t \ u_t]$$

and noticing that the Q function is convex and quadratic in Ω_y, it can be shown that,

$$\Omega_y = \left[\sum_{t=t_1}^{t_\gamma} <y_t r_t^T>_x + \sum_{t=s_1}^{s_\beta} <y_t r_t^T>_{yx}\right]$$
$$\left[\sum_{t=t_1}^{t_\gamma} <r_t r_t^T>_x + \sum_{t=s_1}^{s_\beta} <r_t r_t^T>_{yx}\right]^{-1} \quad (6)$$

where $< . >_x$ denotes integration with respect to the density function $p(x_t | y_{1:T}, \theta)$ and $< . >_{yx}$ denotes integration with respect to the density function $p(y_t | x_t, \theta)$. The measurement noise co-variance can be shown to be,

$$R = \frac{1}{T} \left[\sum_{t=t_1}^{t_\gamma} \left\langle (y_t - \Omega_y r_t) y_t^T \right\rangle_x + \sum_{t=s_1}^{s_\beta} \left\langle (y_t - \Omega_y r_t) y_t^T \right\rangle_{yx}\right]. \quad (7)$$

All the integrations are performed using corresponding particle approximations. The parameters in the matrices Ω_x, Ω_y, Q, and R constitute the linear parameter vector, θ_l.

Step 2 : In step one, it is assumed that the centers and widths of the radial basis functions are known. However, in practice, they are usually unknown. In this step, centers and radii are estimated. The basic idea is to obtain a *maximum a posteriori* (MAP) estimate of the state trajectory and fix centers and radii that provide the best possible predictions of MAP state estimate and the observations. The MAP estimate of the state under missing observations is obtained using a modified version of Viterbi algorithm described in Godsill et al. [2001]. The modified version ac-

counts for missing observations. The algorithm is described below for the sake of completeness [4].

Viterbi Algorithm

1. *Initialization*: For $1 \leq i \leq N$, $\delta_1(i) = \log(p(x_1^{(i)})) + \log(p(y_1|x_1))$.
2. *Recursion*: For $2 \leq t \leq T$, and $1 \leq j \leq N$,

$$\delta_t(j) = \begin{cases} \log(p(y_t|x_t^{(j)})) + \max_i [\delta_{t-1}(i) + \\ \quad \log(p(x_t^{(j)}|x_{t-1}^{(i)}))] \text{ if } y_t \text{ is measured} \\ \max_i \left[\log(p(y_t^{(i)}|x_t^{(j)})) + \delta_{t-1}(i) + \\ \quad \log(p(x_t^{(j)}|x_{t-1}^{(i)})) \right] \text{ if } y_t \text{ is missing} \end{cases}$$

$$\psi_t(j) = \arg\max_i \left[\delta_{t-1}(i) + \log(p(x_t^{(j)}|x_{t-1}^{(i)})) \right]$$

3. *Termination*: $i_T = \arg\max_i \delta_T(i)$ and $x_{MAP}(T) = x_T^{(i_T)}$.
4. *Backtracking*: For $t = T-1, T-2, \cdots, 1$, $i_t = \psi_{t+1}(i_{t+1})$, and $x_{MAP}(t) = x_t(i_t)$.

Now the centers and radii of radial basis functions are estimated using standard nonlinear least squares on the MAP trajectory of states, MAP estimate of missing observations, and the available input-output data.

5.1 The Algorithm

The complete proposed identification algorithm is summarized below:

0. **Initialization**: Initialize the parameter vector to θ_0.
1. **Expectation**: Approximate the expected value of the complete log-likelihood function (E-step) using particle filters.
2. **Maximum a Posteriori Estimate**: Obtain a maximum *a posteriori* estimate of the state trajectory using Viterbi algorithm. Using this MAP estimate of the state trajectory, fix the centers and variances of the radial basis functions. In other words, estimate $(\theta_{nl})_{i+1}$, where i denotes the number of EM algorithm iterations performed so far.
3. **Maximization**: Maximize the Q function with respect to θ_l and call the maximizing parameter, $(\theta_l)_{i+1}$. Then set $\theta_{i+1} = [(\theta_l)_{i+1} \ (\theta_{nl})_{i+1}]$.
4. **Iterate**: Repeat steps 1, 2, and 3 until the change in parameter vector is within a specified tolerance level.

6. EXAMPLE

The proposed approach is applied on data collected from a real continuous stirred tank reactor (CSTR). The governing equations of this popular CSTR, shown in figure 1, are given below (Morningred et al. [1992], Chen [2004])

$$\frac{dC_A}{dt} = \frac{q}{V}(C_{Ai} - C_A) - k_0 C_A e^{-E_A/T}$$

$$\frac{dT}{dt} = \frac{q}{V}(T_i - T) - \frac{\Delta H}{\rho C_p} k_0 C_A e^{-E_A/T} - \frac{\rho_c C_{pc}}{\rho C_p V} q_c$$

[4] for notational clarity, the parameter dependence is not shown in the density functions
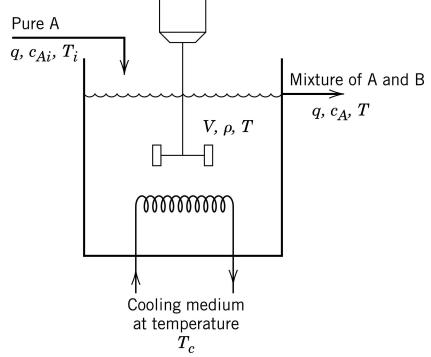
Fig. 1. Continuous Stirred Tank Reactor - Picture taken from Seborg et al. [2004].

$$(1 - e^{-\frac{hA}{q_c \rho_c C_{pc}}})(T - T_c)$$

where C_A is the concentration of the reactant in the reactor, T is the temperature in the reactor, q is the flow rate, V is the volume of the reactor, C_{Ai} and T_i are inflow concentration and temperature, $k_0 C_A e^{-E_A/T}$ is the reaction rate, ΔH is the reaction heat, ρ and ρ_c are the densities of the reactant and the cooling fluid respectively, C_p and C_{pc} are the corresponding specific heats, h and A are the effective heat transfer coefficient and area respectively, T_c and q_c are the temperature and flow rate of the cooling fluid. Finite difference discretization of the above continuous time differential equations results in the following model, where the state vector is $x_t = [x_t(1) \ x_t(2)] = [C_A(t) \ T(t)]$, and

$$f(x_t, u_t, \theta) = x_{t-1}$$
$$+ \Delta t \begin{bmatrix} \frac{q}{V}(C_{Ai} - x_{t-1}(1)) - \theta_1 x_{t-1}(1) e^{-E_A/x_{t-1}(2)} \\ \frac{q}{V}(T_i - x_{t-1}(2)) - \theta_2 x_{t-1}(1) e^{-E_A/x_{t-1}(2)} \end{bmatrix}$$
$$+ \begin{bmatrix} 0 \\ \frac{\rho_c C_{pc}}{\rho C_p V} u_{t-1} \left[1 - e^{-\theta_3 A/(u_{t-1}\rho_c C_{pc})} \right] (T_c - x_{t-1}(2)) \end{bmatrix}$$

where $\theta_1 = k_0$, $\theta_2 = (k_0 \Delta H)/(\rho C_p)$, $\theta_3 = hA$, $g(x_t, u_t, \theta) = x_t$ and Δt is the discretization sample time. C_{Ai} and q_c are input variables.

Using real measurements from an experiment conducted on this CSTR [5], different models are developed. The $f(.)$ and $g(.)$ functions in the model are approximated using radial basis functions with $I_x = 10$, $I_y = 10$. The data set consists of 1000 samples of concentration, C_A, and temperature, T, and coolant flow rate, q_c, measurements.

Full Data Case: The proposed algorithm is applied on this data, with randomly chosen initial guess for the paramter vector. The predictions of concentration from the estimated model for different prediction horizons are shown in figure 2. The %-fit, at these prediction horizons, calculated with the estimated model is comparable to that of input-output Hammerstein-Weiner (HW) models built

[5] Data obtained from "Database for the Identification of Systems" - http://homes.esat.kuleuven.be/smc/daisy/

Table 1. % fit of the models for different prediction horizons

% missing data	Prediction Horizon			
	1	5	10	20
0%	89%	82%	75%	65%
5%	88%	79%	73%	59%
10%	88%	79%	69%	58%
15%	87%	76%	71%	60%
20%	87%	76%	71%	58%

using Matlab system identification toolbox. However, it should be noted that while there is no realistic and fair way to compare the complexities of HW and state space models, an attempt is made to compare the "best" trial and error based HW model with the state-space model estimated using the proposed approach. The main advantage of the proposed method, over other nonlinear input-output identification methods, is in its ability to handle missing data.

Missing Data Case: In order to test the ability of the proposed algorithm to estimate parameters in presence of missing observations, four different subsets of the data are created by randomly choosing a fraction of the measurements. These subsets are created with 5%, 10%, 15%, and 20% of the observations missing. Table 1 shows the percentage fit of the various models. It is well-known that the bias and variance errors of the estimated parameters increase with increase in the fraction of missing observations. This phenomenon can be observed in table 1 through the deterioration in the predictive ability of the models obtained from data sets with higher percentages of missing observations. The relatively poor performance of the models at large prediction horizons can be explained by the choice of input excitation used. The input-output data from the process is collected by exciting the input using a bi-level PRBS signal. It is believed that a multi-level PRBS excitation signal will improve the %-fit of the above models.
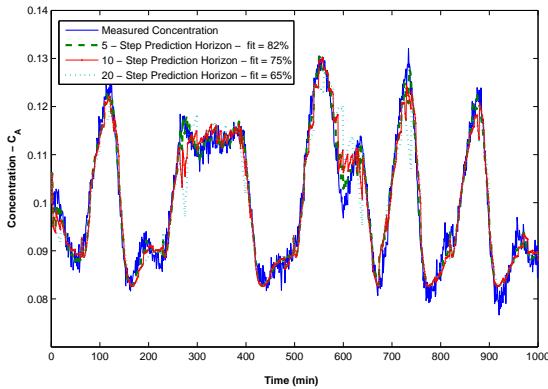
Fig. 2. True and predicted concentration profiles - Full data case.

7. CONCLUSIONS

An approach to identify stochastic nonlinear systems using a combination of expectation maximization algorithm and particle filters is presented. This approach is extended to handle missing observations and to the case of unknown model structure. The developed algorithm is applied to a real continuous stirred tank reactor.

REFERENCES

C. Andrieu, A. Doucet, S.S. Singh, and V.B. Tadic. Particle methods for change detection, system identification, and control. *Proceedings of the IEEE*, 92:423–438, 2004.

W.S. Chen. *Bayesian Estimation by Sequential Monte Carlo Sampling For Nonlinear Dynamical Systems*. Ph.d. thesis, Ohio State University, 2004.

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B*, 39:1–38, 1977.

S.J. Godsill, A. Doucet, and M. West. Maximum a posteriori sequence estimation using Monte Carlo simulations. *Ann. Inst. Stat. Math*, 53(1):82–96, 2001.

G.C. Goodwin and J.C. Agüero. Approximate EM algorithms for paramter and state estimation in nonlinear stochastic models. *Proceedings of IEEE Conference on Decision and Control*, pages 368–373, 2005.

R. B. Gopaluni. Identification of nonlinear processes with known structure under missing observations. In *Proceedings of IFAC World Congress*, Seoul, Korea, 2008.

R. D. Gudi, S. L. Shah, and M. R. Gray. Adaptive multirate state and parameter estimation strategies with application to a bioreactor. *AIChE Journal*, 41(11): 2451–2464, 1995.

A. J. Isaksson. Analysis of identified 2-D noncausal models. *IEEE Transactions on Information Theory*, 39: 525–534, 1993.

D. Li, S.L. Shah, T. Chen, and K. Qi. Application of dual-rate modeling to CCR octane quality inferential control. *IEEE Transactions on Control Systems Technology*, 11 (1):43–51, 2003.

L. Ljung. *System Identification: Theory for the user*. Prentice Hall, 1999.

J.D. Morningred, B. E. Paden, D. E. Seborg, and D. A. Mellichamp. An adaptive nonlinear predictive controller. *Chemical Engineering Science*, 47:755–762, 1992.

H. Raghavan, A. K. Tangirala, R. B. Gopaluni, and S. L. Shah. Identification of chemical processes with irregular output sampling. *Control Engineering Practice*, 14(5): 467–480, 2006.

T.B. Schön, A. Wills, and B. Ninness. Maximum likelihood nonlinear system estimation. In *Proceedings of IFAC Symposium on System Identification*, pages 1003–1008, 2006.

D.E. Seborg, T.F. Edgar, and D.A. Mellichamp. *Process Dynamics and Control*. John Wiley, New Jersey, 2004.

R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications*. Springer, 2000.

S. Tatiraju, M. Soroush, and R. Mutharasan. Multi-rate nonlinear state and parameter estimation in a bioreactor. *Biotechnology and Bioengineering*, 63(1):22–32, 1999.

P. Van Overschee and B. De Moor. *Subspace identification for linear systems*. Kluwer, 1996.

L. Wang and P. Gawthrop. On the estimation of continuous transfer functions. *International Journal of Control*, 74(9):889–904, 2001.