# Identification of Nonlinear State-Space Models: The Case of Unknown Model Structure

**R.Bhushan Gopaluni** *

* *Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4, Email: gopaluni@chml.ubc.ca.*

**Abstract:** This article presents an algorithm for identification of nonlinear state-space models when the "true" model structure of a process is unknown. In order to estimate the parameters in a state-space model, one needs to know the model structure and have an estimate of states. An approximation of the model structure 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 an example.

*Keywords:* Nonlinear Systems, Maximum Likelihood Parameter Estimation, Expectation Maximization, Particle Filters.

## 1. INTRODUCTION

Nonlinear models are commonly used to describe the behavior of many chemical 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,

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

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 typically 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 provides an algorithm for approximation and estimation of $f(.)$ and $g(.)$ using a combination of radial basis functions and Expectation Maximization (EM) algorithm Shumway and Stoffer (2000).

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 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 the one based on particle filter approximation in Gopaluni (2008) can be used.

The algorithm presented in this article extends the one in author's previous work on parameter estimation for known model structure Gopaluni (2008). The central idea is to find the parameter vector, $\theta$, that maximizes the likelihood function of the observations, $y_t$. Due to the presence of latent variables, $x_t$, 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 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 required in EM algorithm. Moreover, since the process is nonlinear, the distribution of latent variables, $x_t$, and measurements is not 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, an approach that uses radial basis functions to approximate the process dynamics and particle filters to approximate the expected value of the joint likelihood function is presented.

The rest of this paper is divided into following sections: The problem is mathematically defined in section 2, a summary of expectation maximization algorithm is presented in section 3, the proposed algorithm is presented in section 4, and an example is presented in section 5.

## 2. PROBLEM DEFINITION AND NOTATION

As explained in the previous section, it is assumed that the process dynamics are unknown and therefore 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. Hence, the model in (1) is approximated using radial basis functions as follows:

$$x_{t+1} = \sum_{i=1}^{I_x} h_i \rho_i(x_t, u_t, c_i, \Sigma_x) + Ax_t + Bu_t + w_t$$
$$y_t = \sum_{i=1}^{I_y} g_i \gamma_i(x_t, u_t, d_i, \Sigma_y) + Cx_t + Du_t + v_t$$

where $\rho_i(.,.)$ and $\gamma_i(.,.)$ are the radial basis functions centered around $c_i$ and $d_i$ with variance $\Sigma_x$ and $\Sigma_y$ respectively, $A$, $B$, $C$, $D$ are appropriate matrices that are used to capture any linear dynamics in the model. $w_t$, and $v_t$ are identically and independently distributed Gaussian noise sequences with zero mean and covariances $Q$ and $R$ respectively. $h_i$ 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, linear terms involving the matrices $A$, $B$, $C$, and $D$ are added. In this article, radial Guassian basis functions of the following form are used:

$$\rho_i(x_t, u_t, c_i, \Sigma_x) = e^{\left[-(\bar{x}_t - c_i)^T \Sigma_x^{-1} (\bar{x}_t - c_i)\right]}$$
$$\gamma_i(x_t, u_t, d_i, \Sigma_y) = e^{\left[-(\bar{x}_t - d_i)^T \Sigma_y^{-1} (\bar{x}_t - d_i)\right]}$$

where $\bar{x}_t$ is the concatenated vector of states and inputs. The input-output data from the nonlinear model in (1) are denoted by $\{y_{1:T}, u_{1:T}\}$, where $y_{1:T}$ are the observations from time, $t = 1$, to $t = T$, and $u_{1:T}$ are corresponding inputs during that time period. The parameter vector $\theta$ includes all the constant parameters in the above model that describe the process behavior, and is defined as $\theta = (\theta_l, \theta_{nl})$, where $\theta_l$ consists of all "linear" parameters, $h_i$, $g_i$, $Q$ and $R$, and $\theta_{nl}$ consists of all "nonlinear" parameters, $c_i$, $d_i$, $\Sigma_x$, $\Sigma_y$.

The expectation maximization algorithm plays a central role in the method developed in this article, and hence a summary of EM algorithm is presented below. It is an elegant optimization algorithm that constructs a complete likelihood function including the latent states and 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 $p(y_{1:T}|\theta)$ denote the 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 due to the presence of latent states. Instead, using the Markov property of the states 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 states, the joint probability density function of the states and observations is iteratively maximized to obtain a maximizing $\theta$ 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, say $\theta_0$.
(2) Estimate the states given the parameter vector and the observations and evaluate

$$Q(\theta_i, \theta) = \int \log[p(x_{1:T}, y_{1:T}|y_{1:T}, \theta)] p(x_{1:T}|y_{1:T}, \theta_i) dx_{1:T} \quad (2)$$

where $p(x_{1:T}|y_{1:T})$ is the joint conditional density function of the states given the observations, and $\theta_i$ is an estimate of the parameter vector from a previous iteration.
(3) Maximize $Q(\theta_i, \theta)$ with respect to $\theta$. Call the maximizing value $\theta_{i+1}$.
(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, $p(y_{1:T}|\theta)$, increases monotonically through these iterations. Due to the nonlinear nature of the dynamics it is not possible to analytically evaluate the $Q$-function in (2). In the next section, an approximation of the $Q$-function and an approach to maximize it are presented.

## 3. MAIN ALGORITHM

### *3.1 Approximation of Q function*

A number of approximations of EM algorithm, involving different approximations of Q function have been proposed in the literature. They either involve approximating the nonlinearities Goodwin and Agüero (2005) or approximating expected value of the joint likelihood function using particle filters and other simulation based approaches Schön et al. (2006); Gopaluni (2008). Methods involving approximation of nonlinearities will fail if the nonlinearities are prominent and on the other hand, methods involving approximation of expected value are usually computationally intensive. In this article, the $Q$ function is approximated using a combination of particle filters and smoothers. The complete details of this approximation and its extension to handle missing data are presented in the author's work in Gopaluni (2008). For continuity a summary of this approach is presented in this section.

The $Q$ function, using Markov property of states, can be expanded to

$$\begin{aligned}
&Q(\theta_i, \theta) \\
&= \int \log[p(x_1|y_{1:T}, \theta)] p(x_1|y_{1:T}, \theta_i) dx_1 \\
&\quad + \sum_{t=2}^{T} \int \log[p(x_t|x_{t-1}, \theta)] p(x_{t-1:t}|y_{1:T}, \theta_i) dx_{t-1:t} \\
&\quad + \sum_{t=1}^{T} \int \log[p(y_t|x_t, \theta)] p(x_t|y_{1:T}, \theta_i) dx_t. \qquad (3)
\end{aligned}$$

From the above expression, it is easy to notice that in order to obtain an approximation of the $Q$ function, the following density functions are needed:

(1) $p(x_1|y_{1:T}, \theta_i)$,
(2) $p(x_{t-1:t}|y_{1:T}, \theta_i)$, and
(3) $p(x_t|y_{1:T}, \theta_i)$.

It is possible to obtain the following particle approximations of the above density functions (please see Gopaluni (2008) for details)

$$\begin{aligned}
p(x_1|y_{1:T}, \theta_i) &= \sum_{i=1}^{N} w_{1|1}^{(i)} \delta(x_1 - x_1^{(i)}) \\
p(x_t|y_{1:T}, \theta_i) &= \sum_{i=1}^{N} w_{t|T}^{(i)} \delta(x_t - x_t^{(i)}) \\
p(x_{t-1}, x_t|y_{1:T}, \theta_i) &= \sum_{i=1}^{N} w_{t-1,t}^{(i)} \delta(x_{t-1} - x_{t-1}^{(i)}) \delta(x_t - x_t^{(i)})
\end{aligned}$$

where $w_{1|1}^{(i)}$, $w_{t|T}^{(i)}$ and $w_{t-1,t}^{(i)}$ are appropriate weights calculated using Bayes rule and importance sampling Klaas et al. (2006), $\delta(.)$ is the Kronecker delta function, and $x_t^{(i)}$ are $N$ samples of states obtained through simulations. Using the above approximations of the density functions in the $Q$ function, one can write the following expression,

$$\begin{aligned}
Q(\theta_i, \theta) \approx &\sum_{i=1}^{N} w_{1|1}^{(i)} \log[p(x_1^{(i)}|y_{1:T}, \theta)] + \sum_{t=2}^{T} \sum_{i=1}^{N} w_{t-1,t}^{(i)} \\
&\log[p(x_t^{(i)}|x_{t-1}^{(i)}, y_{1:T}, \theta)] + \sum_{t=1}^{T} \sum_{i=1}^{N} w_{t|T}^{(i)} \log[p(y_t|x_t^{(i)}, \theta)]
\end{aligned} \qquad (4)$$

In the above approximation, since the noise sequences are assumed to be Gaussian, the density functions, $p(x_t^{(i)}|x_{t-1}^{(i)}, y_{1:T}, \theta)$, and $p(y_t|x_t^{(i)}, \theta)$ can be written in terms of Gaussian density functions and hence,

$$\begin{aligned}
&\log[p(x_t^{(i)}|x_{t-1}^{(i)}, y_{1:T}, \theta)] = \\
&-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\det(Q)) - \frac{1}{2}(x_t^{(i)} - \hat{x}_{t-1}^{(i)})^T Q^{-1} \\
&(x_t^{(i)} - \hat{x}_{t-1}^{(i)})
\end{aligned}$$

$$\begin{aligned}
&\log[p(y_t|x_t^{(i)}, \theta)] = \\
&-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\det(R)) - \frac{1}{2}(y_t^{(i)} - \hat{\hat{x}}_{t-1}^{(i)})^T R^{-1} \\
&(y_t^{(i)} - \hat{\hat{x}}_{t-1}^{(i)})
\end{aligned}$$

where $\hat{x}_t^{(i)} = \sum_{i=1}^{I_x} h_i \rho_i(x_t^{(i)}, u_t, c_i, \Sigma_x) + Ax_t^{(i)} + Bu_t$, and $\hat{\hat{x}}_t^{(i)} = \sum_{i=1}^{I_y} g_i \gamma_i(x_t^{(i)}, u_t, d_i, \Sigma_y) + Cx_t^{(i)} + Du_t$.

### *3.2 Maximization of Q function*

The maximization of $Q$ function is performed in two steps using separable least squares. It is easy to notice that the parameters in $\theta_l$ enter the model linearly, while those in $\theta_{nl}$ enter the model nonlinearly. Hence, a two step procedure (called separable least squares) where the linear parameters are estimated in the first step using linear least squares, and the nonlinear parameters are estimated in the second step through nonlinear least squares. The procedure is explained below.

*Step 1* Starting with an initial guess for the nonlinear parameter vector, $\theta_{nl}$, the $Q$ function is maximized with respect to $\theta_l$. This maximization can be achieved through linear least squares. Before providing the maximizing value of the linear parameter vector, define the following matrices,

$$\begin{aligned}
\Omega_x &= [h_1 \; h_2 \; \cdots h_{I_x} \; A \; B] \\
s_t &= [I_1 \rho_1(x_t, u_t, c_1, \Sigma_x) \; I_1 \rho_2(x_t, u_t, c_2, \Sigma_x) \cdots \\
&\qquad I_1 \rho_{I_x}(x_t, u_t, c_{I_x}, \Sigma_x) \; x_t \; u_t]
\end{aligned}$$

where $I_1$ is a vector of ones of appropriate dimensions. Noting that the $Q$ function is quadratic in $\Omega_x$, through straightforward calculations, it can be shown that

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

where $< . >_{xx}$ is used to denote integration with respect to the density function $p(x_{t-1:t}|y_{1:T}, \theta)$. This integration can be approximated using the particle approximation of $p(x_{t-1:t}|y_{1:T}, \theta)$. 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,

$$\begin{aligned}
\Omega_y &= \left[g_1 \; g_2 \; \cdots g_{I_y} \; C \; D\right] \\
r_t &= [I_1 \gamma_1(x_t, u_t, d_1, \Sigma_y) \; I_1 \gamma_2(x_t, u_t, d_2, \Sigma_y) \cdots \\
&\qquad I_1 \gamma_{I_y}(x_t, u_t, d_{I_x}, \Sigma_y) \; x_t \; u_t]
\end{aligned}$$

and noticing that the $Q$ function is quadratic in $\Omega_y$, it can be shown that,

$$\Omega_y = \left[\sum_{t=1}^{T} < y_t r_t^T >_x\right] \left[\sum_{t=1}^{T} < r_t r_t^T >_{yx}\right]^{-1} \qquad (6)$$

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,

$$R = \frac{1}{T}\sum_{t=1}^{T} \left\langle (y_t - \Omega_y r_t) y_t^T \right\rangle_x .$$

The parameters in the matrices $\Omega_x$, $\Omega_y$, $Q$, and $R$ constitute the linear parameter vector, $\theta_l$.

*Step 2* : In step one, it is assumed that the centers and widths of the radial basis functions are known. However, in practice, it is difficult estimate them. In this step an approach to estimate centers and radii is presented. The 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 is obtained using a modified Viterbi algorithm as described in Godsill et al. (2001). For the sake of completeness, Viterbi algorithm is described below [1] .

**Viterbi Algorithm**

1. *Initialization*: For $1 \leq i \leq N$, $\delta_1(i) = \log(p(x_1^{(i)}) + \log(p(y_1|x_1^{(i)}))$.
2. *Recursion*: For $2 \leq t \leq T$, and $1 \leq j \leq N$,
$$\delta_t(j) = \log(p(y_t|x_t^{(j)})) + \max_i \left[\delta_{t-1}(i) + \log(p(x_t^{(j)}|x_{t-1}^{(i)}))\right]$$
$$\psi_t(j) = \arg\max_i \left[\delta_{t-1}(i) + \log(p(x_t^{(j)}|x_{t-1}^{(i)}))\right]$$
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 $x_{MAP}(t) = x_t(i_t)$.

An estimate of $\theta_{nl}$ is now obtained from the data $\{x_{MAP}(1:T), y_{1:T}, u_{1:T}\}$ by fixing $\theta_l$ to its estimated value from step 1 and using nonlinear least-squares. Step 1 and Step 2 are iterated until changes in $\theta_l$ and $\theta_{nl}$ between iterations are less than a specified tolerance level.

### *3.3 Proposed Algorithm*

The complete proposed identification algorithm is summarized below:

0. **Initialization**: Initialize the parameter vector to $\theta_0$.
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 $\theta_l$ 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.

[1] for notational clarity, the parameter dependence is not shown in the density functions below

## 4. ILLUSTRATIVE EXAMPLES

The proposed approach is tried on data collected from a real continuous stirred tank reactor. The governing

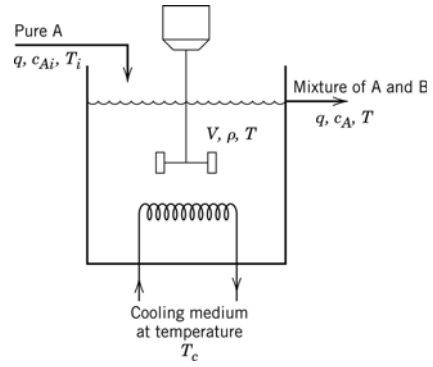


Fig. 1. Continuous Stirred Tank Reactor - Picture taken from Seborg et al. (2004).

equations of this popular CSTR, shown in figure 1, are given below (Morningred et al. (1992); Chen (2004))

$$\frac{dC_A}{dt} = \frac{q}{V}(C_{Ai} - C_A) - k_0 C_A e^{-E_A/T}$$
$$\frac{dT}{dt} = \frac{q}{V}(T_i - T) - \frac{\Delta H}{\rho C_p} k_0 C_A e^{-E_A/T} - \frac{\rho_c C_{pc}}{\rho C_p V} q_c (1 - e^{-\frac{hA}{q_c \rho_c C_{pc}}})(T - T_c)$$

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, $\Delta H$ is the reaction heat, $\rho$ and $\rho_c$ 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,

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

where the state vector is $x_t = [x_t(1) \quad x_t(2)] = [C_A(t) \quad T(t)]$, $\theta_1 = k_0$, $\theta_2 = (k_0 \Delta H)/(\rho C_p)$, $\theta_3 = hA$, $u_t = q_c$, $g(x_t, u_t, \theta) = x_t$ and $\Delta t$ is the discretization sample time. $C_{Ai}$ and $q_c$ are input variables. Real data [2] collected from this reactor is shown in figures 2 and 3. Our goal is to fit a state-space model to this data assuming that the energy and mass balance expressions provided above are unknown.

[2] Please note that this data is available at http://homes.esat.kuleuven.be/ smc/daisy/

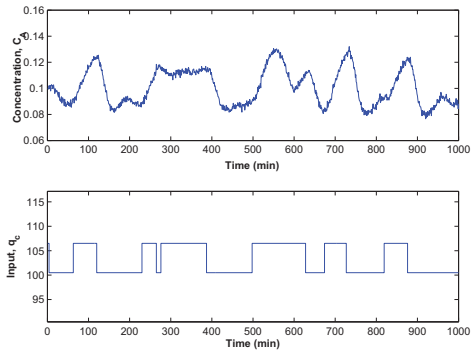


Fig. 2. The concentration, $C_A$, measurements.

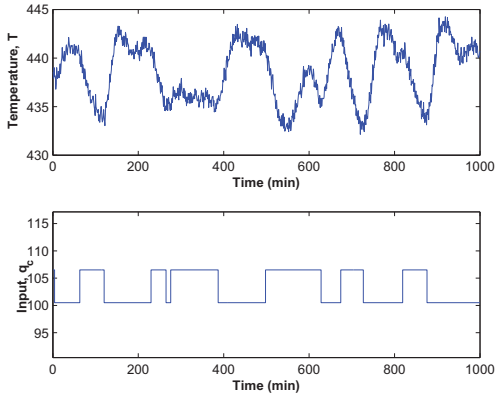


Fig. 3. The temperature, T, measurements.

The proposed algorithm is applied on this data, with a single radial basis function to describe the nonlinearities in the state and observation equations *i.e.,* with $I_x = 1$, $I_y = 1$. The accuracy of the model can definitely be increased by increasing the number of radial basis functions used. The predictions of concentration from this model for different prediction horizons are shown in figure 4. The %-fit, at these prediction horizons, calculated with the estimated model is comparable to that of input-output Hammerstein-Weiner (HW) models built 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 - both in states and observations. In this article, missing observations are not considered. However, as shown in Gopaluni (2008), it is possible to derive particle approximations of density functions, required to approximate the $Q$ function, even if there are some missing observations. Hence, extension of this approach to handle missing observations is rather straightforward.

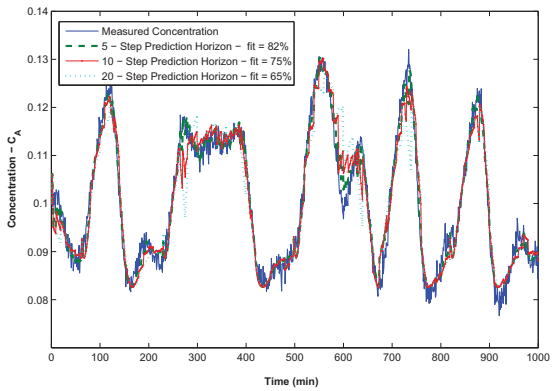


Fig. 4. True and predicted concentration profiles.

## 5. CONCLUSIONS

An approach to identify stochastic nonlinear systems using a combination of expectation maximization algorithm and particle filters is presented. In the proposed approach it is assumed that the model structure is unknown, and is approximated using radial basis functions. The expectation step in the algorithm is approximated using particle approximations of certain density functions. The maximization step is performed by separable least squares, where linear parameters are estimated using linear least squares, and the nonlinear parameters are estimated by using nonlinear least squares on a sequence of maximum *a posteriori* states and observations. The developed algorithm is applied to a real continuous stirred tank reactor. The proposed approach is easily extendable to handle missing observations.

## REFERENCES

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

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

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

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

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

Klaas, M., Briers, M., Nando, D., and Doucet, A. (2006). Fast particle smoothing: If I had a million particles. *International Conference on Machine Learning*.

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

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

Schön, T., Wills, A., and Ninness, B. (2006). Maximum likelihood nonlinear system estimation. In *Proceedings*

*of IFAC Symposium on System Identification*, 1003–1008.

Seborg, D., Edgar, T., and Mellichamp, D. (2004). *Process Dynamics and Control*. John Wiley, New Jersey.

Shumway, R. and Stoffer, D. (2000). *Time Series Analysis and Its Applications*. Springer.

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