Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

A Bayesian Method for Estimating Parameters in Stochastic Differential Equations

Hadiseh Karimi*. Kimberley B. McAuley**

Department of Chemical Engineering, Queen's University, Kingston, K7L3N6
**Canada (e-mail: hadiseh.karimi@queensu.ca).*
***Canada (Tel: 613-533-6000 ext.77562; e-mail: kim.mcauley@queensu.ca).*

Abstract: A Bayesian algorithm is developed for estimating parameters in nonlinear stochastic differential equation (SDE) models. The proposed algorithm uses prior information about parameters and builds on the approximate expectation maximization (AEM) algorithm (Karimi and McAuley, 2014a). A nonlinear continuous stirred tank reactor (CSTR) model is used to compare the effectiveness of the Bayesian algorithm to that of the AEM algorithm. For the CSTR example studied, the proposed method provides more accurate parameter estimates, especially for small data sets.

Keywords: Parameter estimation, maximum likelihood, stochastic modeling, stochastic systems, nonlinear equations, uncertain dynamic systems, differential equations, noise levels.

1. INTRODUCTION

Fundamental dynamic models are derived using material, energy and momentum balances. Stochastic terms are sometimes introduced on the right-hand sides of the resulting differential equations to account for disturbances and model mismatch (Jones et al., 1989). The resulting equations are called stochastic differential equations (SDEs). In this paper, we consider Multi-Input Multi-Output (MIMO) nonlinear SDE models of the form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\eta}(t) \ , \tag{1.a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \, , \tag{1.b}$$

$$\mathbf{y}(t_{m\,r,j}) = \mathbf{g}(\mathbf{x}(t_{m\,r,j}), \mathbf{u}(t_{m\,r,j}), \boldsymbol{\theta}) + \boldsymbol{\varepsilon}(t_{m\,r,j}) \, , \tag{1.c}$$

where $\mathbf{x} \in R^X$ is the vector of state variables, t is time, $\mathbf{f} : R^X \times R^U \times R^P \rightarrow R^X$ is a vector of nonlinear functions, $\mathbf{u} \in R^U$ is the vector of input variables and $\boldsymbol{\theta} \in R^P$ is the vector of unknown model parameters. $\boldsymbol{\eta}(t) \in R^X$ is a continuous zero-mean stationary Gaussian white-noise process with covariance matrix $E\{\boldsymbol{\eta}(t_1)\boldsymbol{\eta}(t_2)\}=\mathbf{Q}\ \delta(t_2-t_1)$, where $\mathbf{Q}$ is the corresponding diagonal power spectral density function :

$$\mathbf{Q} = \begin{bmatrix} Q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_X \end{bmatrix}. \tag{2}$$

The matrix of power spectral density function sometimes referred to as the process disturbance intensity matrix (Varziri et al., 2008). $\delta(.)$ is the Dirac delta function and $\mathbf{x_0} \in R^X$ is a vector of initial conditions for the state variables. Some of these initial conditions may be known to the modeler and others may be unknown values that require estimation along with the model parameters. $\mathbf{y} \in R^Y$ is the vector of measured output variables. The times at which measurements are available for the rth response (r=1…Y) are denoted by $t_{m\,r,j}$ (j = 1…N_r) where N_r is the number of measurements for the rth response. $\mathbf{g} \in R^Y$ is a vector of nonlinear mappings and $\boldsymbol{\varepsilon} \in R^Y$ is a vector of zero-mean random variables. If the measurement errors are independent, the corresponding covariance matrix has the following form:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_Y^2 \end{bmatrix}. \tag{3}$$

Consider the vector $\mathbf{Y_m}$ that contains the stacked measured value $\mathbf{Y_m} = [y_1(t_{m\,1,1}) \ldots y_1(t_{m\,1,N_1}) \ldots y_Y(t_{m\,Y,1}) \ldots y_Y(t_{m\,Y,N_Y})]^{\mathrm{T}}$ and $\mathbf{X_m} = [x_1(t_{m\,1,1}) \ldots x_1(t_{m\,1,N_1}) \ldots x_Y(t_{m\,Y,1}) \ldots x_Y(t_{m\,Y,N_Y})]^{\mathrm{T}}$ which contains the stacked values of the state variables at the measurement times. $\mathbf{U_m}$ and $\boldsymbol{\varepsilon_m}$ are corresponding vectors for the input variables and random errors, respectively so that:

$$\mathbf{Y_m} = \mathbf{G}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon_m} \, . \tag{4}$$

where $\mathbf{G}$ is $\mathbf{G}=[\mathbf{g}, \ldots, \mathbf{g}]^{\mathrm{T}}{}_{1\times NY}$. The index m for a variable indicates that the values of that variable are taken at measurement times. The existence of a solution of an SDE is ensured when globally Lipschitz, linear growth and boundedness conditions are satisfied (Liptser and Bishwal, 2000). Since $\boldsymbol{\eta}(t)$ does not have a simple mathematical interpretation, SDEs are often written in the differential form (Liptser and Shiryaev, 2000):

$$\mathrm{d}\,\mathbf{x} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})dt + \mathbf{Q}d\mathbf{W} \, , \tag{1.d}$$

whrere $\mathbf{W}(t)$ is a Wiener process. Modelers often have knowledge about the accuracy of their measurements (*i.e.*, the diagonal elements of $\mathbf{\Sigma}$), but do not have knowledge about the values of the diagonal elements of $\mathbf{Q}$, (i.e., $\mathbf{Q_d}=[Q_1,\ldots,Q_X]^T$). Let $\zeta = [\boldsymbol{\theta}^T, \mathbf{x}_{0u}^T, \mathbf{Q_d}^T]^T$ be the vector of unknown parameters in the SDE model where $\mathbf{x}_{0u}$ is a vector of the unknown initial conditions. SDE models are used for simulation, design and optimization of chemical processes and for model predictive control (McLean and McAuley, 2012). Therefore, accurate and reliable parameter estimation techniques for SDE models are beneficial for chemical engineers. Maximum likelihood estimation (MLE) methods are commonly used to estimate parameters in SDE models because of their asymptotic efficiency and consistency (Casella and Berger, 1990).

A challenge in estimating parameters in SDE models for chemical engineering systems is that experiments and measurements are often limited due to cost or difficulties in measuring certain variables. Performing further experiments is expensive and may not be feasible (McLean and McAuley, 2012). As a result, the number of data values for parameter estimation may be limited and some of the states are often not measured. While MLE methods provide satisfactory solutions for parameter estimation in many SDE models, they provide noticeably biased parameter estimates when only limited data are available for parameter estimation (Casella and Berger, 1990; Ninness and Henriksen, 2010).

In chemical engineering applications, prior information about some of the parameters is often known to the modeler (e.g., reasonable initial guesses and physically realistic ranges for parameter values). Box and Draper (1964) introduced the use of Bayesian methods for estimating parameters in chemical engineering models so that prior knowledge about parameter values could be accounted for. One benefit of Bayesian parameter estimation methods is that they can provide improved parameter estimates, especially when available datasets are small (Robert and Casella, 1999). In general, in Bayesian methods, the probability density function of the parameters given the measured data $p(\zeta \mid \mathbf{Y_m})$ is maximized to estimate the unknown parameters. This joint probability function is referred to as the posterior density function (Jang and Gopaluni, 2011). The posterior density function can be obtained from:

$$p(\zeta \mid \mathbf{Y_m}) = \frac{p(\mathbf{Y_m} \mid \zeta)\, p(\zeta)}{p(\mathbf{Y_m})} . \qquad (5)$$

The numerator on the right-hand side is the product of the probability density function of the measurements given parameters $p(\mathbf{Y_m} \mid \zeta)$ and the prior distribution of the parameters $p(\zeta)$, which contains knowledge about the possible values of ζ. The likelihood function of the parameters given the measurements is defined as

$$L(\zeta \mid \mathbf{Y_m}) = p(\mathbf{Y_m} \mid \zeta) . \qquad (6)$$

The denominator in (5), which ensures that the posterior integrates to unity, does not depend on the parameter values. The prior probability $p(\zeta)$ is important when there is limited data available to provide reliable estimates for some of the model parameters. When a large quantity of informative data is available, the posterior probability will be dominated by the likelihood function In nonlinear models with unmeasured states, evaluation of the posterior density function is a major challenge requiring calculation of complicated integrals of probability density functions (Jang and Gopaluni, 2011; Ljung, 1999). Computationally intensive Markov Chain Monte Carlo (MCMC) algorithms (also referred to as particle filtering), which require very few assumptions about the posterior density function, have been used to compute these integrals (Coleman and Block, 2006; Jang and Gopaluni, 2011; Robert and Casella, 1999). MCMC methods are used to approximate the posterior densities in SDE models (Geweke and Tanizaki, 2001; Jang and Gopaluni, 2011; Ninness and Henriksen, 2010), in mixed models (Gelman, 2006) and in ordinary differential equation models (Coleman and Block, 2006). MCMC methods are particularly computationally demanding when the number of states and parameters is large (Gopaluni, 2010). Benefits and drawbacks of MCMC methods are summarized by Chen et al. (2004).

In this article a computationally efficient algorithm is proposed for estimating parameters and states in nonlinear SDE models when the modeler has some prior knowledge about some of the parameters. This algorithm is developed using a Bayesian approach. Recently, we developed three approximate MLE algorithms for estimating parameters in nonlinear SDE models (Karimi and McAuley, 2013; Karimi and McAuley, 2014a; Karimi and McAuley, 2014b). These MLE-based methods, which do not require prior knowledge about parameters, are computationally efficient, but can provide poor parameter estimates when data sets are small. Here, we develop an approximate Bayesian expectation maximization (ABEM) algorithm that builds on our previous approximate expectation maximization (AEM) method (Karimi and McAuley, 2014a). The inclusion of prior information about parameters in the resulting objective function leads to improved parameter estimates, especially when data are sparse or noisy. First, an analytical expression for the posterior density function is derived and used to develop a suitable objective function for parameter estimation. The proposed algorithm is then tested using a CSTR model and results are compared with those from the AEM method. It is shown that the proposed ABEM method provides more accurate parameter estimates for the example studied.

2. DEVELOPMENT OF THE APPROXIMATE BAYESIAN EXPECTATION MAXIMIZATION ALGORITHM

In Bayesian approaches, the posterior density function $p(\zeta \mid \mathbf{Y}_m)$ is maximized to obtain the parameter estimates. Maximizing $p(\zeta \mid \mathbf{Y}_m)$ is equal to minimizing -ln $p(\zeta \mid \mathbf{Y}_m)$. When developing the AEM methodology, we showed that (Karimi and McAuley, 2014a):

$$
\begin{aligned}
-2\ln\, p(\mathbf{Y}_m \mid \boldsymbol{\zeta}) \approx \text{constant} \;& + [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \\
& \times [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})] + q \ln[\det(\mathbf{Q})] \\
& + \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} \\
& \times [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]\, \mathrm{d}t, \qquad (7)
\end{aligned}
$$

where $\mathbf{x}_{\sim}$ is the vector of state estimates obtained using B-spline basis functions:

$$\mathbf{x}_{\sim}(t) = \mathbf{B}\boldsymbol{\Phi}(t). \qquad (8)$$

$\boldsymbol{\Phi}(t)$ is a matrix of spline functions:

$$\boldsymbol{\Phi}(t) = \begin{bmatrix} \boldsymbol{\varphi}_1^T(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}_2^T(t) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\varphi}_X^T(t) \end{bmatrix}, \qquad (9)$$

and $\mathbf{B} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_X]^{\mathrm{T}}$, (10)

where $\boldsymbol{\beta}_s$ is the vector containing c_s B-spline coefficients for the sth state:

$$\boldsymbol{\beta}_s = [\beta_{s,1}, \dots, \beta_{s,c_s}]^{\mathrm{T}} \text{ for } \quad s=1,\dots,X. \qquad (11)$$

Since $p(\mathbf{Y}_m)$ is data-dependent and parameter-independent, (5) can be written as:

$$p(\boldsymbol{\zeta} \mid \mathbf{Y}_m) \propto p(\mathbf{Y}_m \mid \boldsymbol{\zeta}) p(\boldsymbol{\zeta}). \qquad (12)$$

Taking the logarithm of both sides gives:

$$\ln\, p(\boldsymbol{\zeta} \mid \mathbf{Y}_m) = \text{constant} + \ln\, p(\mathbf{Y}_m \mid \boldsymbol{\zeta}) + \ln\, p(\boldsymbol{\zeta}). \qquad (13)$$

Assuming that the probability density functions of **Q,** $\mathbf{x}_{0u}$ and $\boldsymbol{\theta}$ are independent of each other gives:

$$p(\boldsymbol{\zeta}) = p(\boldsymbol{\theta}) p(\mathbf{x}_{0u}) p(\mathbf{Q}). \qquad (14)$$

Taking the logarithm of both sides of (14) and substituting into (13) gives:

$$\begin{aligned}
\ln\, p(\boldsymbol{\zeta} \mid \mathbf{Y}_m) = \text{constant} \;& + \ln\, p(\mathbf{Y}_m \mid \boldsymbol{\zeta}) + \ln\, p(\boldsymbol{\theta}) \\
& + \ln\, p(\mathbf{x}_{0u}) + \ln\, p(\mathbf{Q}).
\end{aligned} \qquad (15)$$

If the modeler assigns a Gaussian distribution to the ith parameter θ_i (with mean $\bar{\theta}_i$ and variance $\sigma_{\theta,i}^2$) to account for prior knowledge:

$$p(\theta_i) = \frac{1}{\sqrt{2\pi}\sigma_{\theta,i}} \exp\left(-\frac{(\theta_i - \bar{\theta}_i)^2}{2\sigma_{\theta,i}^2}\right). \qquad (16)$$

Let $\bar{\boldsymbol{\theta}} = [\bar{\theta}_1, \dots, \bar{\theta}_P]^{\mathrm{T}}$ denote the vector of assigned prior parameter means and $\boldsymbol{\sigma}_{\theta}^2 = [\sigma_{\theta,1}^2, \dots, \sigma_{\theta,P}^2]^{\mathrm{T}}$ denote the vector of assigned prior parameter variances. If the assigned values of the model parameters are independent:

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{P} \frac{1}{\sqrt{2\pi}\sigma_{\theta,i}} \exp\left(-\frac{(\theta_i - \bar{\theta}_i)^2}{2\sigma_{\theta,i}^2}\right). \qquad (17)$$

The modeler may also assign a Gaussian distribution for the sth unmeasured initial state value $x_{0u,s}$ (with mean $\bar{x}_s$ and variance $\sigma_{x,s}^2$) to account for prior knowledge:

$$p(x_{0u,s}) = \frac{1}{\sqrt{2\pi}\sigma_{x,s}} \exp\left(-\frac{(x_{0u,s} - \bar{x}_s)^2}{2\sigma_{x,s}^2}\right). \qquad (18)$$

Let $\bar{\mathbf{x}}_{0u} = [\bar{x}_1, \dots, \bar{x}_{N_u}]^{\mathrm{T}}$ denote the vector of assigned means for the initial values where N_u is the number of unknown initial state values and $\boldsymbol{\sigma}_{0u}^2 = [\sigma_{x,1}^2, \dots, \sigma_{x,N_u}^2]^{\mathrm{T}}$ denote the vector of variances assigned for these initial states. If the assigned initial state values are independent:

$$p(\mathbf{x}_{0u}) = \prod_{s=1}^{N_u} \frac{1}{\sqrt{2\pi}\sigma_{x,s}} \exp\left(-\frac{(x_{0u,s} - \bar{x}_s)^2}{2\sigma_{x,s}^2}\right). \qquad (19)$$

Since the modeler would usually have no prior information about **Q**, $p(\mathbf{Q})$ in (15) can be assumed to be a uniform distribution (between 0 and ∞) indicating that all positive values are equally likely:

$$p(\mathbf{Q}) \propto \begin{cases} 1 & \text{if } \mathbf{Q} > 0 \\ 0 & \text{otherwise} \end{cases}. \qquad (20)$$

Taking the natural logarithms of both sides of (17), (19) and (20) and substituting these expressions and $\ln\, p(\mathbf{Y}_m \mid \boldsymbol{\zeta})$ from (7) into (15) gives:

$$
\begin{aligned}
-2\ln\, p(\boldsymbol{\zeta} \mid \mathbf{Y}_m) \approx \text{constant} \;& + [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \\
& \times [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})] + q \ln[\det(\mathbf{Q})] \\
& + \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} \qquad (21) \\
& \times [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]\, \mathrm{d}t + 2\ln\left(\sqrt{2\pi}\sigma_{\theta,i}\right)^P \\
& + \sum_{i=1}^{P} \frac{(\theta_i - \bar{\theta}_i)^2}{\sigma_{\theta,i}^2} + 2\ln\left(\sqrt{2\pi}\sigma_{x,s}\right)^{N_u} + \sum_{s=1}^{N_u} \frac{(x_{0u,s} - \bar{x}_s)^2}{\sigma_{x,s}^2}.
\end{aligned}
$$

Because the B-spline expressions in (8) can be readily integrated with respect to time, the integrand in the fifth term on the right-hand side of (21) becomes an algebraic expression. As a result, there is no need for numerical solution of differential equations when estimating **B** and $\boldsymbol{\theta}$.

The only differences between the ABEM and AEM objective function derived in our past work are the four last terms in (21), which are related to prior information about the parameters and initial states.

Estimates of the parameters and unknown initial states for SDE model (1) can be determined by minimizing objective function $p(\boldsymbol{\zeta} \mid \mathbf{Y}_m)$ in (21) with respect to the model

parameters, initial states and disturbance parameters in ζ and the B-splines parameters **B**:

$$\hat{\boldsymbol{\zeta}}, \hat{\mathbf{B}} = \arg \min_{\boldsymbol{\zeta}, \mathbf{B}} \; -2 \ln \, p(\boldsymbol{\zeta} \mid \mathbf{Y}_m) \,. \tag{22}$$

3. ILLUSTRATIVE EXAMPLE: NONLINEAR TWO-STATE CSTR MODEL

In this section, a two-state CSTR model from Marlin (1995) with additive stochastic disturbances is used to illustrate the use of ABEM:

$$\frac{\mathrm{d}\, C_A(t)}{\mathrm{d}\, t} = \frac{F(t)}{V}(C_{A0}(t) - C_A(t)) - k_r C_A(t) + \eta_C(t) \,, \tag{23a}$$

$$\frac{\mathrm{d}\, T(t)}{\mathrm{d}\, t} = \frac{F(t)}{V}(T_0(t) - T(t)) + U_A(T(t) - T_{in}(t)) + \gamma\, k_r C_A(t) + \eta_T(t), \tag{23b}$$

$$y_C(t_{m\,1,j}) = C_A(t_{m\,1,j}) + \varepsilon_C(t_{m\,1,j}), \tag{23c}$$

$$y_T(t_{m\,1,j}) = T(t_{m\,1,j}) + \varepsilon_T(t_{m\,1,j}) \,. \tag{23d}$$

The parameter estimation results obtained from the proposed ABEM approach are then compared to results obtained using AEM. Notice that stochastic disturbance terms ($\eta_C(t)$ and $\eta_T(t)$) appear on the right-hand sides of material balance (23a) and energy balance (23b), respectively, where C_A is the concentration of reactant A and T is temperature. The rate constant for the reaction is:

$$k_r = k_{ref} \; \exp(-E/R(1/T - 1/T_{ref})) \,. \tag{24}$$

U_A is a heat transfer coefficient that depends on the coolant flowrate F_c:

$$U_A = aF_c^{\,b+1}(V\rho c_p(F_c + aF_c^{\,b}(2\rho_c c_{pc})^{-1}))^{-1}, \tag{25}$$

and γ is:

$$\gamma = (-\Delta H_{rxn})(\rho c_p)^{-1} \,, \tag{26}$$

The true initial values for the states are $C_A(0)$= 1.569 kmol·m^{-3} and $T(0)$=341.37 K. For illustration, it is assumed that the initial concentration $C_A(0)$ is perfectly known, but the initial temperature $T(0)$ is measured at 343.51 K with a variance of S_T^2=5.0 K^2. Parameters that will be estimated using ABEM and AEM are model parameters $\boldsymbol{\theta}_{\mathrm{CSTR}} = [k_{ref}, E/R, a, b]^{\mathrm{T}}$, the initial temperature T(0), disturbance intensities Q_C and Q_T and the B-spline coefficients **B**. The inputs for this CSTR model are the feed flow rate F, the inlet concentration C_{A0}, the inlet temperature T_0, the coolant inlet temperature T_{in} and the flow rate of coolant to the cooling coil Fc. Known values of the model constants and known measurement variances σ_C^2 and σ_T^2 are provided in Table 1 (Marlin, 1995).

Table 1. Model constants (Marlin, 1995)

Model	Value	Units
c_p	4186.8	J·kg^{-1}·K^{-1}
c_{pc}	4186.8	J·kg^{-1}·K^{-1}
T_{ref}	350	K
V	1	m^3
ρ	1000	kg·m^{-3}
ΔH_{rxn}	-544.154×10^3	J·kmol^{-1}
σ_C^2	4×10^{-4}	kmol^2·m^{-6}
σ_T^2	0.64	K^2

These values were used along with the true values of the parameters (at the top of Table 2) and the step input trajectories in Fig. 1 to generate 200 sets of simulated data. In each simulation, C_A is measured n_C times and T is measured n_T times. The "ode45 solver" in MATLAB™ was used to generate the simulated data.

The ABEM objective function (J) for estimating parameters in the CSTR model is:

$$\begin{aligned} J = {} & \frac{1}{\sigma_C^2}\sum_{j=1}^{n_C}\left(y_C(t_{m\,1,j}) - C_{A\sim}(t_{m\,1,j})\right)^2 \\ & + \frac{1}{\sigma_T^2}\sum_{j=1}^{n_T}\left(y_T(t_{m\,2,j}) - T_{\sim}(t_{m\,2,j})\right)^2 + q_C \ln(Q_C) + q_T \ln(Q_T) \\ & + \frac{1}{Q_C}\int_{t_0}^{t_q}\left[\frac{\mathrm{d}\, C_{A\sim}(t)}{\mathrm{d}\, t} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r C_{A\sim}(t)\right]^2 \mathrm{d}\, t \\ & + \frac{1}{Q_{T,k}}\int_{t_0}^{t_q}\left[\frac{\mathrm{d}\, T_{\sim}(t)}{\mathrm{d}\, t} - \frac{F(t)}{V}(T_0(t) - T_{\sim}(t)) - U_A[T_{\sim}(t) - T_{in}(t)]\right. \\ & \left. - \gamma\, k_r C_{A\sim}(t)\right]^2 \mathrm{d}\, t + \frac{(k_{ref} - \bar{k}_{ref})^2}{\sigma_{k_{ref}}^2} + \frac{(E/R - \bar{E}/R)^2}{\sigma_{ER}^2} \\ & + \frac{(a - \bar{a})^2}{\sigma_a^2} + \frac{(b - \bar{b})^2}{\sigma_b^2} + \frac{(T_m(0) - T_{\sim}(0))^2}{S_T^2}, \end{aligned} \tag{27}$$

where $\bar{k}_{ref}$, $\sigma_{k_{ref}}^2$, $\bar{E}/R$, σ_{ER}^2, σ_a^2, $\bar{b}$ and σ_b^2 are means and variances of k_{ref}, E/R, a and b, respectively.

Each of the means was set at 70% of the true value for the corresponding parameter value. The corresponding standard deviations used in prior distributions specified in (16) were set at 60% of the initial guesses for the corresponding parameter. In the simulations, continuous white noise disturbances were approximated using discrete white noise sequences with a sampling interval of Δt=0.5 min so that in (27), $q_C = q_T$=128. The ABEM objective function and AEM objective function ((27) without the last 5 terms) were programed using AMPL™ and optimized using the IPOPT solver (Wächter and Biegler, 2006). Cubic B-splines were implemented with three equally-spaced knots per sampling interval, which was shown to be effective in our previous work (Karimi and McAuley, 2014b). Estimates for unknown parameters were obtained from 200 sets of simulated data

obtained using two different scenarios (100 replicate data sets each, obtained using different random noise, disturbance sequences and random initial guesses for the parameters chosen from a uniform distribution between 0% and 500% of their true values), as shown in Table 2. Medians and interquartile ranges (IQRs) from the 100 replicate estimates were used to quantify the effectiveness of the two parameter estimation approaches and are reported in Table 2.

In Scenario A, C_A and T were measured once every 0.5 min, so that 128 concentration measurements and 128 temperature measurements are available for parameter estimation.

Table 2. True parameter values, median values and IQRs for the estimates based on 100 Monte Carlo runs for two scenarios.

Parameter			k_{ref}	$(E/R)/10^3$	$a/10^6$	b	$T(0)$	Q_C	Q_T
Unit			min^{-1}	K			K	$kmol^2 \cdot m^{-6} \cdot min^{-1}$	$K^2 \cdot min^{-1}$
True			**0.461**	**8.3301**	**1.678**	**0.50**	**341.38**	**0.0010**	**4.0**
Scenario									
A	ABEM	Median	0.434	8.2171	1.486	0.49	343.30	0.010	5.1
		IQR	0.017	0.2058	0.394	0.10	1.89	0.004	1.4
	AEM	Median	0.435	8.2210	1.481	0.50	343.05	0.010	5.1
		IQR	0.018	0.2142	0.520	0.12	1.78	0.004	1.3
B	ABEM	Median	0.432	8.1753	2.081	0.44	343.17	0.001	10.1
		IQR	0.039	0.8863	0.440	0.08	1.83	0.000	2.3
	AEM	Median	0.417	8.1546	4.666	0.00	343.23	0.001	10.4
		IQR	0.075	1.2818	2.773	0.18	1.94	0.000	3.5

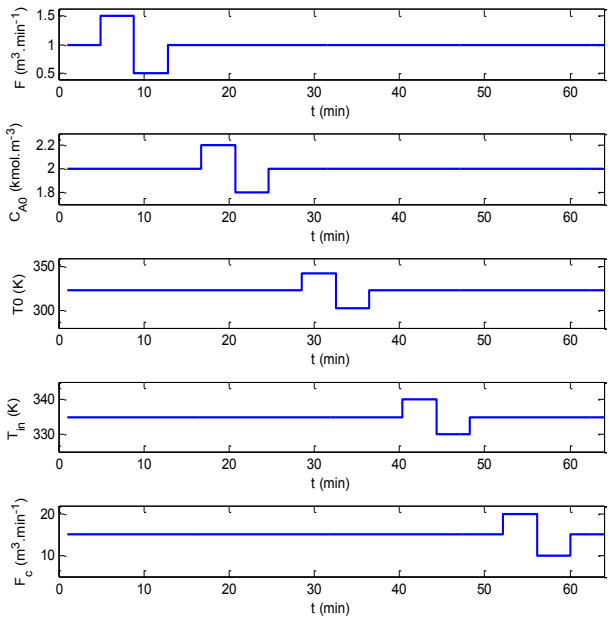

Fig. 1. Input trajectories for the CSTR (Varziri et al., 2008)

As expected, there is no significant difference between the estimates of parameters obtained from ABEM and AEM because there is a large dataset available for parameter estimation. Estimated trajectories ($C_{A\sim}$ and $T_{\sim}$) for one of the simulated data sets are shown in Fig. 2, along with the true state trajectories and the corresponding data values when ABEM is used for parameter estimation. As expected, the estimated state trajectories follow the true trajectories closely. In Scenario B, only 10 equally-spaced concentration measurements and 10 equally-spaced temperature measurements were available for parameter estimation. Estimates of model parameters obtained from ABEM are significantly less biased than those obtained from AEM. The estimates of b obtained in Scenario B are almost zero using AEM. The estimates of Q_C and Q_T obtained from ABEM are similar to those obtained from AEM. The reason for this similarity is that no prior information for Q_C and Q_T was specified. The estimates of Q_C and Q_T in both ABEM and AEM are biased because the data set is not informative enough to estimate them with good accuracy. Note that the widths of the IQRs obtained using AEM are larger than those obtained using ABEM in both Scenarios A and B. The results in Table 2 reveal that the ABEM parameter estimation algorithm was more effective than AEM for the CSTR example studied. Results in Table 2 suggest that ABEM has superior performance when there are small datasets and prior information is available for some parameters. In future, it will be important to test ABEM using larger-scale models.

4. CONCLUSIONS

A Bayesian method for estimating parameters and process disturbance intensities in nonlinear SDE models is proposed when the modeler has prior knowledge about some of the parameters. This approximate Bayesian expectation maximization (ABEM) method builds on the approximate expectation maximization (AEM) algorithm (Karimi and McAuley, 2014a) and uses prior information about parameters. The new ABEM method permits modelers to estimate model parameters and the magnitude of process disturbances in SDE models even when the size of the datasets are small. A two-state nonlinear CSTR model with stochastic disturbances and measurement noise was used to test the ABEM methodology. Parameter and disturbance intensity estimates were compared with those from the AEM method. The resulting ABEM parameter estimates are less biased and more precise than the corresponding estimates obtained using ABEM, especially in the case where the data set was small.

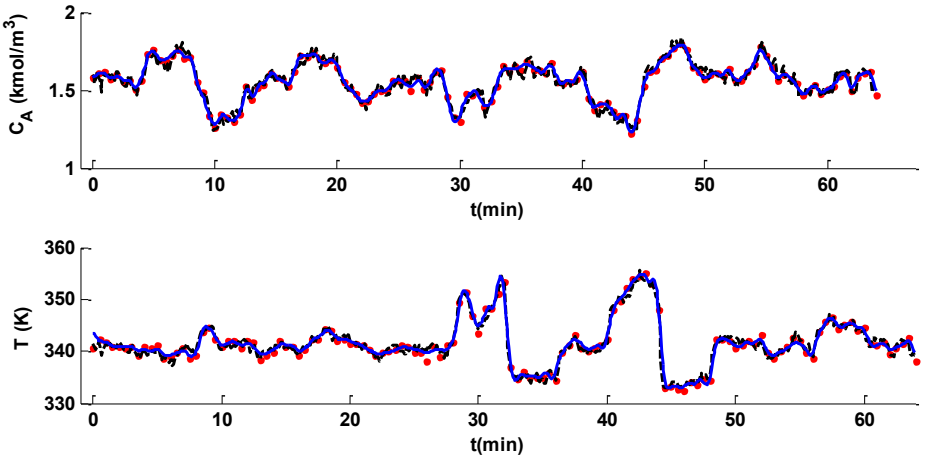

Fig. 2. Measured, true, and predicted concentration and temperature responses for the ABEM method for one dataset in scenario A (• simulated data, ----- response with true parameters and true stochastic noise, ___ABEM response).

REFERENCES

Casella, G., Berger, R. (1990). *Statistical Inference.* Brooks/Cole Publishing Company, Pacific Grove, California.

Box, G. E., and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, Series B, 26 (2), 211-252.

Chen, W., Bakshi, B. R., Goel, P. K., and Ungarala, S. (2004). Bayesian estimation via sequential Monte Carlo sampling: unconstrained nonlinear dynamic systems. *Industrial Engineering Chemistry & Research,* 43, 4012-4025.

Coleman, M. C., and Block, D. E. (2006). Bayesian parameter estimation with informative priors for nonlinear systems. *AICHE Journal*, 52, 651-667.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, 1,515-534.

Geweke, J., and Tanizaki, H. (2001). Bayesian estimation of state-space models using the Metropolis–Hastings algorithm within Gibbs sampling. *Computational Statistics & Data Analysis,* 3, 151-170.

Gopaluni, R. B. (2010). Nonlinear system identification under missing observations: The case of unknown model structure. *Journal of Process Control*, 20, 314-324.

Jang, S., and Gopaluni, R. (2011). Parameter estimation in nonlinear chemical and biological processes with unmeasured variables from small data sets. *Chemical Engineering Science*, 66, 2774-2787.

Jones, R., MacGregor, J., and Murphy, K. (1989). State estimation in wastewater engineering: Application to an anaerobic process. *Environmental Monitoring and Assessment,* 13, 271-282.

Karimi, H., and McAuley, K. B. (2014a). An approximate expectation maximisation algorithm for estimating parameters in nonlinear dynamic models with process disturbances. *The Canadian Journal of Chemical Engineering*, 92, 835-850.

Karimi, H., and McAuley, K. B. (2014b). A maximum-likelihood method for estimating parameters, stochastic disturbance intensities and measurement noise variances in nonlinear dynamic models with process disturbances. *Computers and Chemical Eng*ineering, 67, 178-198.

Karimi, H., and McAuley, K. B. (2013). An Approximate Expectation Maximization Algorithm for Estimating Parameters, Noise Variances, and Stochastic Disturbance Intensities in Nonlinear Dynamic Models. *Industrial Engineering Chemistry & Research,* 52, 18303-18323.

Liptser R.S., Shiryaev A.N. (2000). *Statistics of Random Processes: II. Applications.* Springer, New York.

Ljung, L. (1999). *System identification*, Prentice Hall, New Jersey.

Marlin, T. E. (1995). *Process control: designing processes and control systems for dynamic performance,* McGraw-Hill, New York.

McLean, K. A., and McAuley, K. B. (2012). Mathematical modelling of chemical processes-obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *The Canadian Journal of Chemical Engineering,* 90 (2), 351-366.

Ninness, B., and Henriksen, S. (2010). Bayesian system identification via Markov chain Monte Carlo techniques. *Automatica,* 46, 40-51.

Robert, C. P., and Casella, G. (1999). *Monte Carlo statistical methods*, Springer, New York.

Varziri, M., Poyton, A., McAuley, K., McLellan, P., and Ramsay, J. (2008). Selecting optimal weighting factors in iPDA for parameter estimation in continuous-time dynamic models. *Computers and Chemical Engineering*. 32, 3011-3022.

Wächter, A., and Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming,* 106, 25-57.