

# PARAMETER AND STATE ESTIMATION IN NONLINEAR STOCHASTIC CONTINUOUS-TIME DYNAMIC MODELS WITH UNKNOWN DISTURBANCE INTENSITY

M. S. Varziri, K. B. McAuley\* and P. J. McLellan

Department of Chemical Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6

Approximate Maximum Likelihood Estimation (AMLE) is an algorithm for estimating the states and parameters of models described by stochastic differential equations (SDEs). In previous work (Varziri et al., *Ind. Eng. Chem. Res.*, **47**(2), 380–393, (2008); Varziri et al., *Comp. Chem. Eng.*, in press), AMLE was developed for SDE systems in which process-disturbance intensities and measurement-noise variances were assumed to be known. In the current article, a new formulation of the AMLE objective function is proposed for the case in which measurement-noise variance is available but the process-disturbance intensity is not known a priori. The revised formulation provides estimates of the model parameters and disturbance intensities, as demonstrated using a nonlinear CSTR simulation study. Parameter confidence intervals are computed using theoretical linearization-based expressions. The proposed method compares favourably with a Kalman-filter-based maximum likelihood method. The resulting parameter estimates and information about model mismatch will be useful to chemical engineers who use fundamental models for process monitoring and control.

L'estimation des vraisemblances maximums approximatives (AMLE) est un algorithme destiné à l'estimation des états et des paramètres de modèles décrits par les équations différentielles stochastiques (SDE). Dans un précédent travail (Varziri et al., 2008a, 2008b), l'AMLE a été mis au point pour des systèmes SDE dans lesquels les intensités de perturbations et les variances de bruits de mesure sont supposées connues. On propose dans cet article une nouvelle formulation de la fonction objectif de l'AMLE pour le cas où la variance de bruit de mesure est disponible mais où l'intensité des perturbations de procédé n'est pas connue a priori. La formulation révisée fournit des estimations des paramètres de modèle et des intensités de perturbations, comme le démontre une étude de simulation en CSTR non linéaire. Les intervalles de confiance des paramètres sont calculés par ordinateur au moyen d'expressions basées sur la linéarisation théorique. La méthode proposée se compare favorablement à une méthode de vraisemblance maximum reposant sur le filtre de Kalman. Les estimations de paramètres qui en résultent et l'information sur la discordance de modèle seront utiles aux ingénieurs en génie chimique qui utilisent des modèles fondamentaux pour la surveillance et le contrôle des procédés.

**Keywords:** maximum likelihood, parameter estimation, dynamic models, stochastic disturbances, B-splines, principal differential analysis

## INTRODUCTION

Developing mechanistic mathematical models is of great importance in many disciplines in engineering and science (Biegler and Grossman, 2004). In fundamental dynamic models of chemical processes, which are based on material and energy balances, it is important for the modeller to obtain appropriate values of kinetic and transport parameters using experimental data. The problem of parameter estimation can be formulated using the following continuous-time stochastic dynamic model. To keep the notation simple, a Single-Input Single-Output (SISO) model with a known initial condition is used; extension to Multi-Input Multi-Output (MIMO) systems with unknown initial conditions is straightforward (Varziri et al.,

2008a, in press).

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t), \theta) + \eta(t) \\ x(t_0) &= x_0 \\ y(t_{mj}) &= x(t_{mj}) + \varepsilon(t_{mj})\end{aligned}\quad (1)$$

$x \in \mathfrak{R}$  is the state variable,  $u \in \mathfrak{R}$  is the input variable and  $y \in \mathfrak{R}$  is the output variable.  $\theta \in \mathfrak{R}^p$  is the vector of model parameters and

\* Author to whom correspondence may be addressed.  
E-mail address: mcauleyk@chee.queensu.ca  
Can. J. Chem. Eng. 86:828–837, 2008  
© 2008 Canadian Society for Chemical Engineering  
DOI 10.1002/cjce.20100

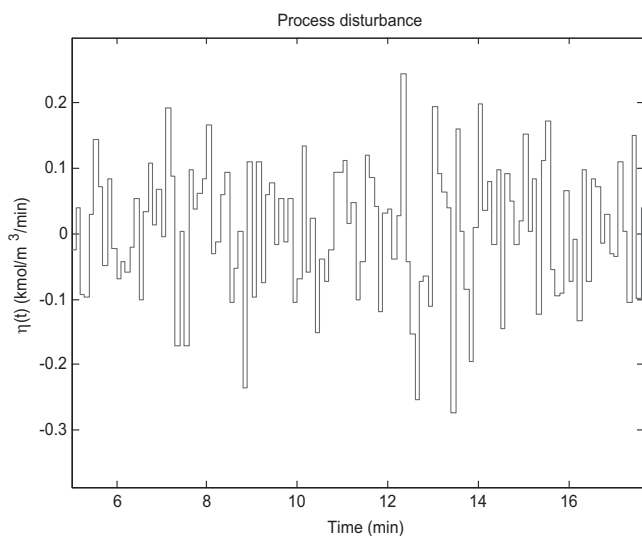


Figure 1. Process disturbance.

$f : \mathfrak{R} \times \mathfrak{R} \times \mathfrak{R}^p \rightarrow \mathfrak{R}$  is a nonlinear function of the state variables, the input variables and the model parameters. We assume that  $f$  satisfies some regularity conditions (Kloeden and Platen, 1992) so that Equation (1) has a unique solution.  $\varepsilon$  is a discrete-time zero-mean uncorrelated Normal random variable with variance  $\sigma_m^2$ .  $\eta(t)$  is a continuous zero-mean stationary white-noise process with covariance matrix  $E\{\eta(t)\eta(t + \tau)\} = Q\delta(\tau)$ , where  $Q$  is the corresponding process disturbance intensity and  $\delta(\cdot)$  is the Dirac delta function. The random noise trajectory  $\eta(t)$  is a series of random steps with a switching time of  $\Delta t$ , where  $\Delta t \rightarrow 0$ . For the corresponding discrete-time white-noise process (Maybeck, 1979):

$$E\{\eta(j_1 \Delta t)\eta(j_2 \Delta t)\} = \begin{cases} \frac{Q}{\Delta t} & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases} \quad (2)$$

where  $j_1$  and  $j_2$  are integers. We also assume that the process disturbance  $\eta(t)$  and the measurement noise  $\varepsilon(t)$  are not correlated. In Figure 1, a discrete random white noise sequence is shown, where the sampling interval for the step functions is  $\Delta t = 0.1$  min., and the variance of the white noise sequence is  $Q/\Delta t = 4 \times 10^{-2}$  (kmol/m<sup>3</sup>/min)<sup>2</sup>. As  $\Delta t \rightarrow 0$ , and the disturbance intensity  $Q$  remains fixed, the variance of the random steps becomes infinite and we obtain the continuous white noise  $\eta(t)$  in Equation (1).

A continuous-time model is used because chemical processes that we are concerned with have a continuous-time nature, and mathematical models for these processes are derived based on continuous-time material, energy, and momentum balances. Using discrete-time models to represent these processes involves inherent approximations. Also, the continuous-time representation is more convenient if some of the data that are used for parameter estimation are irregularly sampled (Kristensen et al., 2004; Varziri et al., 2008b, in press).

Parameter values are generally selected so that the model predictions are as close as possible (usually in the sense of sum of squared errors (SSE)) to the measured responses of the process. Deviations between the model predictions and the measured responses arise from two sources: (i) measurement errors and (ii) process disturbances or model-plant mismatch. Measurement errors, which arise due to sensor inaccuracy and fluctuations, only influence the measured outputs at the current time. Process disturbances and model-plant mismatch, however, can influence

the future behaviour of the process, and hence future measured responses. If discrepancies caused by process disturbances and model mismatch are very small, it is appropriate to model lumped dynamic systems using ordinary differential equations (ODEs). On the other hand, when unknown disturbances and mismatch cause significant discrepancies, it is appropriate to use stochastic differential equation (SDE) models that explicitly account for unmeasured process disturbances in the states (Jazwinski, 1970; Maybeck, 1979).

Parameter estimation in ODEs is usually treated as a constrained nonlinear optimization problem that requires iterative numerical solution of the ODEs (Bard, 1974; Bates and Watts, 1988; Seber and Wild, 1989; Ogunnaike and Ray, 1994). Robust and efficient methods such as multiple shooting (Bock, 1981, 1983) and collocation-based techniques (Biegler, 1984) have also been developed. Unfortunately, these methods, which neglect process disturbances, can result in biased parameter estimates if disturbances are significant (Voss et al., 2004).

Several methods are available for parameter estimation in SDEs (Nielsen et al., 2000). Maximum Likelihood (ML) methods have been used for parameter estimation in stochastic dynamic models (Maybeck, 1982). In the classical ML method, the parameters are selected so that the conditional probability density function of the measured outputs, given the model parameters, is maximized (e.g., Shumway and Stoffer, 2000; Kristensen et al., 2004).

$$p(y_{m1}, \dots, y_{mn} | \theta) = \prod_{i=1}^n p(y_{i-1}, \dots, y_1, \theta) \quad (3)$$

where  $y_{mj} = y(t_{mj})$ . The conditional density functions used in the classical ML method are generally difficult to obtain and require the successive solution of a partial differential equation (the forward Kolmogorov equation (Jazwinski, 1970; Maybeck, 1982)). The reason is that the probability density of the observations depends on the probability density of the underlying system states, which undergo a nonlinear mapping. One way to avoid this difficulty is to use extended-Kalman-filter-related methods in which the nonlinear model is successively linearized so that the time evolution of the state covariance matrix can be described by a set of ODEs (e.g., Jazwinski, 1970). Extended Kalman filters combine numerical integration of the model differential equations with a linearization-based solution for the state covariance matrix (Maybeck, 1982). If the model is highly nonlinear, linearization-based methods may perform poorly. In such cases, the state covariance matrix can be estimated based on deterministic sampling techniques (Julier and Uhlmann, 2004) or ensemble averaging and Monte Carlo-related algorithms (Evensen, 2003; Andrieu et al., 2004). These methods, however, are computationally intensive.

Classical ML estimates can be obtained using the Expectation Maximization (EM) method (Roweis and Ghahramani, 2001; Goodwin and Aguero, 2005). EM is an elegant method; however, it requires calculating the expected value of the log-likelihood of the joint density of the measured observations and unobserved states, conditioned on the parameter values, which is also computationally intensive and requires calculation of integrals.

ML methods are not confined to the classical case; other density functions such as the conditional joint density function of the states and measurements, given the model parameters, can also be used (Maybeck, 1982). In our previous work (Varziri et al., 2008a, in press), we developed an Approximate Maximum Likelihood Estimation (AMLE) method that minimizes the

conditional joint density function of the states and measurements, given the parameters, while assuming a piece-wise polynomial discretization scheme for the state trajectories of the dynamic model. The minimization criterion in this algorithm is sometimes referred to as the Joint MAP-ML criterion since it leads to Maximum A Posteriori (MAP) state estimates (Yeredor, 2000). Unlike Kalman-filter-based or classical ML methods, AMLE does not require the time-varying state covariance matrix, due to the convenient form of its objective function (shown in (11)) that arises from the discretization of the state trajectories.

Although AMLE does not require the time-varying state covariance matrix, the intensity of the model disturbance and the variance of the measurement noise are required to form the objective function. True values or reasonable estimates of some or all of these parameters are generally not known a priori by the modeller, and hence these parameters need to be either adjusted by trial and error (manual Kalman filter tuning) or estimated along with model parameters. For linear dynamic systems, cross validation has been used to simultaneously estimate the noise parameters and model parameters (Gong et al., 1998); however, for nonlinear problems, cross validation can be computationally very intensive (Solo, 1996). ML methods have been widely used for estimating noise variances; however, the estimation problem is generally a difficult and often ill-conditioned problem (Solari, 1969; Warnes and Ripley, 1987; Dee, 1995; Dee and Da Silva, 1999), so that simplifying assumptions about the structure of the noise covariance matrix need to be made to make the estimation problem tractable. In engineering problems, it is often reasonable to assume that the covariance matrix of the measurement noise is reasonably well known because this information is either provided by the manufacturer of the measurement device or can be determined from replicate measurements. Disturbance intensity information is typically poorly known.

In this article, we present a modified formulation of the AMLE objective function following an approach developed by Heald and Stark (2000), for the case in which measurement-noise variance is available but the process-disturbance intensity is not known. We then apply the proposed algorithm to estimate the states and parameters of a nonlinear Continuous Stirred Tank Reactor (CSTR) in a simulation study, assuming that the measurement noise variances are known, but the process disturbance intensities are unknown. We also compare the proposed method to an extended Kalman filter-based maximum likelihood method (Kristensen et al., 2004) in this case study. The paper is organized as follows. In AMLE Fitting Criterion Section, the AMLE algorithm is briefly reviewed. Estimation of process disturbance intensities is discussed in Estimation of Disturbance Intensity Section. In Simulation Case Study: Nonlinear CSTR Section, the nonlinear CSTR case study is presented, followed by the summary and conclusions in Summary and Conclusions Section.

## AMLE FITTING CRITERION

In the following paragraphs we briefly review the AMLE algorithm.

At the discrete time  $t_i$ , where  $t_i = t_{i-1} + \Delta t$  and the sampling interval  $\Delta t$  is small, Equation (1) can be written using the following Euler approximation.

$$\begin{aligned} x(t_{i-1} + \Delta t) &= x(t_i) \\ &= x(t_{i-1}) + f(x(t_{i-1}), u(t_{i-1}), \theta) \Delta t + \eta(t_{i-1}) \Delta t \end{aligned} \quad (4)$$

Note that the variance of random noise term  $\eta(t_{i-1}) \Delta t$  (increment of a Wiener process with constant diffusion  $Q$ ) is  $Q \Delta t$ . We refer to this variance as  $\sigma_p^2$  because it is the variance of the stochastic step disturbances entering the process. Consider  $x(t_i)$  at  $q + 1$  uniformly spaced time points,  $t_i, i = 0 \dots q$  so that  $q \Delta t = T$ , where  $T = t_q - t_0$  is the overall time span for the model predictions. For brevity, we define  $x_i = x(t_i)$  and  $\mathbf{x}$  as a vector containing  $x_i$ s for  $i = 0 \dots q$ . Please note that the set of times at which the measurements are available is within the  $[t_0, t_q]$  interval and is denoted by  $t_{mj} (j = 1 \dots n)$ . The measurement times  $t_{mj}$  do not need to be uniformly spaced. The vector of outputs at observation times  $y(t_{mj}) (j = 1 \dots n)$  and its corresponding state vector of true values  $x(t_{mj}) (j = 1 \dots n)$  and measurement noise vector  $\varepsilon(t_{mj}) (j = 1 \dots n)$  are denoted by  $\mathbf{y}_m, \mathbf{x}_m$ , and  $\varepsilon_m$  respectively.

Note that we have used different discretization schemes for the measurements and state trajectories so that we can treat the state trajectories as continuous-time functions by letting  $\Delta t$  (the sampling interval for state discretization) go to zero as shown in (9). As explained in Introduction Section, the chemical processes that we are concerned with are continuous-time in nature and their corresponding mathematical models are derived based on first principle material and energy balances in continuous-time. The measurements however, are generally taken at discrete times. Problems such as existence of unmeasured states, states measured at irregularly sampled times, and different states measured at different points in time are commonly faced by chemical engineers when estimating parameters in nonlinear dynamic processes (Varziri et al., 2008b, in press). The formulation proposed in this article allows for addressing these problems by using continuous-time state trajectories and discrete-time measurements.

We shall now consider several probability density functions whose log-likelihood could be maximized to obtain optimal parameter estimates:

$$p(\mathbf{y}_m | \theta)$$

$$p(\mathbf{x}, \theta | \mathbf{y}_m)$$

$$p(\mathbf{x} | \mathbf{y}_m, \theta)$$

$$p(\mathbf{x}, \mathbf{y}_m | \theta)$$

The first density function, which is the same density function as shown in Equation (3), is used in classical ML estimation (Seber and Wild, 1989). As noted in the introduction, forming such density functions for SDE models can be very difficult if the model is nonlinear.

The second density function is the posterior joint distribution of  $\mathbf{x}$  and  $\theta$  that leads to optimal state and parameter estimates in a Bayesian framework (Seber and Wild, 1989). Using this density function, however, requires a reasonable prior distribution for the parameters, which can be hard to obtain. The modeller might also opt to use a non-informative prior. In this case, minimizing density function 2 is equivalent to minimizing density function 4.

To compare the third and the fourth density functions, we note that, from Bayes' rule:

$$p(\mathbf{x} | \mathbf{y}_m, \theta) \times p(\mathbf{y}_m | \theta) = p(\mathbf{x}, \mathbf{y}_m | \theta) \quad (5)$$

Using the fourth density function is preferable because  $p(\mathbf{x}, \mathbf{y}_m | \theta)$  contains information about  $p(\mathbf{y}_m | \theta)$ , which, in turn, depends on  $\theta$  (Maybeck, 1982).

Using Equations (1) and (4), the density function  $p(\mathbf{x}, \mathbf{y}_m | \boldsymbol{\theta})$  can be written as (Varziri et al., 2008a, in press):

$$p(\mathbf{x}, \mathbf{y}_m | \boldsymbol{\theta}) = \frac{1}{2\pi^{n/2}\sigma_m^n} \exp\left(-\frac{\text{SSE}}{2\sigma_m^2}\right) \times \frac{1}{2\pi^{q/2}\sigma_p^q} \exp\left(-\frac{\text{PEN}}{2\sigma_p^2}\right) \quad (6)$$

where

$$\text{SSE} = (\mathbf{y}_m - \mathbf{x}_m)^T (\mathbf{y}_m - \mathbf{x}_m)$$

$$\text{PEN} = \sum_{i=1}^q (x_i - x_{i-1} - f(x_{i-1}, u_{i-1}, \boldsymbol{\theta}) \Delta t)^2$$

$$\sigma_p^2 = Q \Delta t \quad (7)$$

Please note that minimizing (6) with respect to  $\sigma_m^2$  leads to  $\hat{\sigma}_m^2 = \text{SSE}/n$ . As explained in Estimation of Disturbance Intensity Section, this estimate overlooks the uncertainty in the estimated states.

If  $Q$  and  $\sigma_m^2$  are known, (6) can be simplified to

$$p(\mathbf{x}, \mathbf{y}_m | \boldsymbol{\theta}) = C_1 \times \exp\left(-\frac{\text{SSE}}{2\sigma_m^2}\right) \times \exp\left(-\frac{\text{PEN}}{2\sigma_p^2}\right) \quad (8)$$

where  $C_1$  depends on  $\sigma_m^2$  and  $\sigma_p^2$ , but not on the model parameters or the states. We will let  $\Delta t$  to shrink to zero to obtain the continuous-time equivalent of (8) (shown in Equation (11)) that is used in our algorithm (Varziri et al., 2008a, in press). Note that when  $\Delta t \rightarrow 0$ , then  $C_1 \rightarrow \infty$ . To avoid this problem, the probability density functional (Jazwinski, 1970) is defined as:

$$\lim_{\Delta t \rightarrow 0} \frac{p(\mathbf{x}, \mathbf{y}_m | \boldsymbol{\theta})}{C_1} = \lim_{\Delta t \rightarrow 0} \exp\left(-\frac{\text{SSE}}{2\sigma_m^2}\right) \times \exp\left(-\frac{\text{PEN}}{2\sigma_p^2}\right) \quad (9)$$

From (7) it can be noted that SSE does not depend on  $\Delta t$ . We can also see that:

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \exp\left(-\frac{\text{PEN}}{2\sigma_p^2}\right) &= \exp\left(\lim_{\Delta t \rightarrow 0} \left(-\frac{\text{PEN}}{2\sigma_p^2}\right)\right) \\ &= \exp\left(\lim_{\Delta t \rightarrow 0} \frac{\sum_{i=1}^q (x_i - x_{i-1} - f(x_{i-1}, u_{i-1}, \boldsymbol{\theta}) \Delta t)^2}{2Q\Delta t}\right) \\ &= \exp\left(\lim_{\Delta t \rightarrow 0} \frac{\sum_{i=1}^q \left(\frac{x_i - x_{i-1}}{\Delta t} - f(x_{i-1}, u_{i-1}, \boldsymbol{\theta})\right)^2 \Delta t}{2Q}\right) \\ &= \exp\left(\frac{1}{2Q} \int_{t_0}^{t_q} (\dot{x}(t) - f(x(t), u(t), \boldsymbol{\theta}))^2 dt\right) \end{aligned} \quad (10)$$

Note that optimal state and parameter estimates are obtained by substituting (10) in (9) and minimizing the negative of the natural logarithm of the probability density functional as follows:

$$\begin{aligned} -\ln\left(\lim_{\Delta t \rightarrow 0} \frac{p(\mathbf{x}, \mathbf{y}_m | \boldsymbol{\theta})}{C_1}\right) &= \frac{(\mathbf{y}_m - \mathbf{x}_m)^T (\mathbf{y}_m - \mathbf{x}_m)}{2\sigma_m^2} \\ &+ \frac{1}{2Q} \int_{t_0}^{t_q} (\dot{x}(t) - f(x(t), u(t), \boldsymbol{\theta}))^2 dt \end{aligned} \quad (11)$$

Since  $x(t)$  is an unknown function, minimizing (11) over  $x(t)$  and  $\boldsymbol{\theta}$  is an infinite-dimensional problem (a calculus of variations problem) which is generally hard to solve.

To turn the problem into a finite-dimensional problem, the state variable,  $x(t)$ , is assumed to be sufficiently accurately approximated by a basis function expansion. B-splines provide convenient basis functions due to their compact support and other favourable properties (Ramsay and Silverman, 2005; Poyton et al., 2006; Ramsay et al., 2007; Varziri et al., 2008a, in press):

$$x(t) \approx x_{\sim}(t) = \sum_{i=1}^c \beta_i \phi_i \quad (12)$$

where  $\beta_i$ ,  $i = 1 \dots c$  are B-spline coefficients and  $\phi_i(t)$   $i = 1 \dots c$  are B-spline basis functions (de Boor, 2001). The knot sequences for the B-spline basis functions should be selected in a way that they provide enough flexibility for the B-spline curves to follow the possibly sharp features of the state trajectories. However, caution should be practiced since too fine of a knot sequence could significantly slow down the parameter-estimation algorithm. Our experience shows that if the sampling interval is not too large, placing one knot at each observation point and one or two knots in between every two observation points leads to sound results. Once the knot sequence is rich enough to capture the important features of the state trajectories, not much is gained by further refining the grid. Please refer to Poyton et al. (2006) and Varziri et al. (2008b, in press), for further discussions and examples on knot sequence strategies. Note that Equation (12) can be written in matrix form:

$$x_{\sim}(t) = \boldsymbol{\Phi}^T(t) \boldsymbol{\beta} \quad (13)$$

where  $\boldsymbol{\Phi}(t)$  is a vector containing the  $c$  basis functions and  $\boldsymbol{\beta}$  is vector of  $c$  spline coefficients. The B-spline expansion,  $x_{\sim}(t)$ , can easily be differentiated:

$$\dot{x}_{\sim}(t) = \frac{d}{dt} \left( \sum_{i=1}^c \beta_i \phi_i(t) \right) = \sum_{i=1}^c \beta_i \dot{\phi}_i(t) = \dot{\boldsymbol{\Phi}}^T \boldsymbol{\beta} \quad (14)$$

By substituting (12) and (14) into (11) we have the following finite-dimensional optimization problem:

$$\begin{aligned} \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\theta}, \boldsymbol{\beta}} \left\{ \frac{(\mathbf{y}_m - \mathbf{x}_{\sim m})^T (\mathbf{y}_m - \mathbf{x}_{\sim m})}{2\sigma_m^2} \right. \\ \left. + \frac{1}{2Q} \int_{t_0}^{t_q} (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u(t), \boldsymbol{\theta}))^2 dt \right\} \end{aligned} \quad (15)$$

Minimizing (15) provides point estimates for the model parameters and the spline coefficients. The spline coefficients can then be used to determine the estimated state trajectory  $x_{\sim}(t)$ . In order to obtain approximate confidence intervals for the model parameters, the inverse of the Fisher information matrix can be used as an approximation to the covariance matrix of the combined vector of states and parameters (Varziri et al., 2008). The AMLE criterion in nonlinear problems generally produces inconsistent and biased parameter estimates but it may outperform ML in the overall mean squared error especially when a small number of measured data is used (Yeredor, 2000) and



is much easier to formulate. Equation (11) was derived based on the assumption that  $\sigma_m^2$  and  $Q$  (the noise parameters) are known. In the next section, we consider the more usual case where  $Q$  is unknown and must be estimated along with the model parameters,  $\theta$ .

## ESTIMATION OF DISTURBANCE INTENSITY

Heald and Stark (2000) developed an iterative algorithm that leads to approximate classical ML estimates for measurement noise variance and process disturbance intensity in a discrete dynamic system. As mentioned in the introduction, using the classical ML criterion is not straightforward because the density function to be maximized is:  $p(y_m | \theta, \sigma_m, \sigma_p) = \int p(y_m, x | \theta, \sigma_m, \sigma_p) dx$  and this integration is generally very difficult or not tractable at all. Heald and Stark (2000) used Laplace's method (described by Mackay, 2004) to approximate this integral. Their development leads to the following measurement noise estimator:

$$\hat{\sigma}_m^2 = \frac{\text{SSE}}{n - \gamma_m} \quad (16)$$

where  $\gamma_m = (1/\hat{\sigma}_m^2) \text{Trace}(\mathbf{A}^{-1})$  and  $\mathbf{A}$  is the Hessian matrix of  $-\ln p(x | y_m, \theta, \sigma_m, \sigma_p)$  with respect to  $x$  evaluated at  $\arg \min_x \{-\ln p(x | y_m, \theta, \sigma_m, \sigma_p)\}$ . Substituting  $\gamma_m$  in (16) and solving for  $\hat{\sigma}_m^2$  we have:

$$\hat{\sigma}_m^2 = \frac{\text{SSE}}{n} + \frac{\text{Trace}(\mathbf{A}^{-1})}{n} \quad (17)$$

If we try to obtain an estimate of  $\sigma_m^2$  by maximizing (6) (or by minimizing  $-\ln p(x, y_m | \theta)$ ) we will get  $\hat{\sigma}_m^2 = \text{SSE}/n$  which lacks the second term in (17); this would have been a good estimate if the true state trajectory, rather than the estimated state trajectory,  $x_{\sim}$ , was used to calculate SSE. The term,  $\text{Trace}(\mathbf{A}^{-1})/n$  can therefore be thought of as a compensation for deviations of  $x_{\sim}$  from  $x$ ; this makes more sense if we consider  $\mathbf{A}^{-1}$  as an approximation to the inverse of the Fisher information matrix which is, in turn, an approximation for the covariance matrix of  $x_{\sim}$ . Having Equation (17) at our disposal and assuming that  $\sigma_m^2$  is known, we can pick an initial estimate of  $Q$  and solve the minimization problem in (15) to get  $\hat{\theta}, \hat{\beta}$ , and therefore  $x_{\sim}$ . The estimates obtained can then be used to evaluate  $\hat{\sigma}_m^2$  using (17). Next, a new estimate of  $Q$  is obtained to minimize the loss function  $(\hat{\sigma}_m^2/\sigma_m^2 - 1)^2$ . This two-step optimization scheme can be summarized as follows:

Outer optimization problem:

$$\hat{Q} = \arg \min_Q \left( \frac{(y_m - x_{\sim}(Q))^T (y_m - x_{\sim}(Q))}{n\sigma_m^2} + \frac{\text{Trace}(\mathbf{A}^{-1})}{n\sigma_m^2} \Big|_{\hat{\theta}, \hat{\beta}} - 1 \right)^2 \quad (18)$$

where  $\hat{\theta}, \hat{\beta}$  are obtained as the solution of the inner optimization problem

$$\hat{\theta}, \hat{\beta} = \arg \min_{\theta, \beta} \left\{ \frac{(y_m - x_{\sim})^T (y_m - x_{\sim})}{2\sigma_m^2} + \frac{1}{2\hat{Q}} \int_{t_0}^{t_q} (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u(t), \theta))^2 dt \right\} \quad (19)$$

The overall minimization problem presented in Equations (18) and (19) consists of outer and inner optimizations. The outer optimization is to minimize the discrepancy between the estimated and the known measurement variance, while the inner minimizes the AMLE criterion in Equation (15) using an estimate of  $Q$  obtained from the outer optimization. The results of this overall optimization problem provide the modeller with  $\hat{\theta}, \hat{\beta}$  and  $\hat{Q}$ .  $\hat{\theta}$  is the desired estimate for the fundamental model parameters. The estimated spline coefficients  $\hat{\beta}$ , can be used to obtain estimated state trajectories  $x_{\sim}$  and  $\hat{Q}$  provides information about the magnitude of the model uncertainty and process disturbance.

## SIMULATION CASE STUDY: NONLINEAR CSTR

We consider the same two-state CSTR example as in our previous work (Varziri et al., 2008). The model equations consist of material and energy balances (Marlin, 2000) with additional stochastic disturbance terms:

$$\begin{aligned} \frac{dC_A(t)}{dt} &= \frac{F(t)}{V} (C_{A0}(t) - C_A(t)) - gC_A(t) + \eta_1(t) \\ \frac{dT(t)}{dt} &= \frac{F(t)}{V} (T_0(t) - T(t)) + \beta_1 (T(t) - T_{\text{cin}}(t)) \\ &\quad + \beta_2 gC_A(t) + \eta_2(t) \\ C_A(0) &= 1.569 \text{ (kmol m}^{-3}\text{)} \\ T(0) &= 341.37 \text{ (K)} \\ y_1(t_i) &= C_A(t_i) + \varepsilon_1(t_i) \\ y_2(t_j) &= C_A(t_j) + \varepsilon_2(t_j) \\ g &= k_{\text{ref}} \exp \left( -\frac{E}{R} \left( \frac{1}{T} - \frac{1}{T_{\text{ref}}} \right) \right), \\ \beta_1 &= -\frac{aF_c^{b+1}(t)}{V\rho C_p (F_c(t) + \frac{aF_c^b(t)}{2\rho C_{pc}})}, \\ \beta_2 &= \frac{(-\Delta H_{\text{rxn}})}{\rho C_p} \end{aligned} \quad (20)$$

where  $E \{ \eta_1(t_i) \eta_1(t_j) \} = Q_{p1} \delta(t_i - t_j)$ ,  $E \{ \eta_2(t_i) \eta_2(t_j) \} = Q_{p2} \delta(t_i - t_j)$  ( $\delta(\cdot)$  is the Dirac delta function),  $\varepsilon_1(t_{mj})$   $j = 1 \dots N_1$  and  $\varepsilon_2(t_{mj})$   $j = 1 \dots N_2$  are white-noise sequences with variances  $\sigma_{m1}^2$  and  $\sigma_{m2}^2$  respectively. We also assume that  $\eta_1, \eta_2, \varepsilon_1$ , and  $\varepsilon_2$  are independent.  $C_A$  is the concentration of the reactant A,  $T$  is the reactor temperature,  $V$  is the volume and  $T_{\text{ref}} = 350$  K is a reference temperature. This stochastic differential equation model is nonlinear in the states ( $C_A$  and  $T$ ) and in the parameters, and does not have an analytical solution. The true values of the parameters to be estimated are:  $E/R = 8330.1$  K,  $k_{\text{ref}} = 0.461$   $\text{min}^{-1}$ ,  $a = 1.678\text{E}6$ ,  $b = 0.5$ . To examine the robustness of the proposed algorithm to the initial parameter values, the initial guess for each parameter was randomly drawn from a Normal distribution with a mean of 50% of the true value of the corresponding parameter and a variance of roughly 15% of the true parameter value.

Parameters  $a$  and  $b$  account for the effect of the coolant flow rate,  $F_c$ , on the heat transfer coefficient. This nonlinear system has five inputs: the reactant flow rate  $F$ , the inlet reactant concentration  $C_{A0}$ , the inlet temperature  $T_0$ , the coolant inlet temperature  $T_{\text{cin}}$ , and the coolant flow rate  $F_c$ . Values for the various other known

constants are as follows:  $V = 1.0 \text{ m}^3$ ,  $C_p = 1 \text{ cal g}^{-1} \text{ K}^{-1}$ ,  $\rho = 1\text{E}6 \text{ g m}^{-3}$ ,  $C_{pc} = 1 \text{ cal g}^{-1} \text{ K}^{-1}$ ,  $\rho_c = 1\text{E}6 \text{ g m}^{-3}$ , and  $-\Delta H_{rxn} = 130\text{E} + 6 \text{ cal kmol}^{-1}$ . The initial steady-state operating point is:  $C_{As} = 1.569 \text{ kmol m}^{-3}$  and  $T_s = 341.37 \text{ K}$ .

In this example, there is no temperature controller, and perturbations are introduced into each of the five inputs using the input scheme shown in Figure 2 (Poyton, 2005).

The back-to-back step changes in the five system inputs generated enough excitation so that the model parameters could successfully be estimated. However, it should be noted that the focus of this article is not experimental design for parameter estimation purposes. Better designs exist that lead to less biased and more precise parameter estimates.

We assume that concentration and temperature are measured; however, we treat the initial concentration and temperature conditions as unknowns that are estimated along with other state values. Note that known initial conditions can be forced as constraints in the AMLE minimization problem (Varziri et al., 2008a, in press). Temperature is measured once every 0.3 min while concentration is measured once per minute. The duration of the simulated experiment is 64 min, so that there are 213 temperature measurements and 64 concentration measurements. The noise variance for the concentration and temperature measurements are  $\sigma_{m1}^2 = 4 \times 10^{-4} (\text{kmol/m}^3)^2$  and  $\sigma_{m2}^2 = 6.4 \times 10^{-1} \text{ K}^2$ , respectively. The corresponding process noise intensities for the stochastic disturbances used in the simulations are  $Q_{p1} = 4 \times 10^{-3} (\text{kmol/m}^3)^2/\text{min}$  and  $Q_{p2} = 4 \text{ K}^2/\text{min}$ .

The model in Equation (20) was simulated using the SIMULINK™ toolbox in MATLAB™. To approximate the continuous-time Gaussian noise we used the “band limited white noise block” (which generates discrete-time normally distributed random numbers) with a sampling time (0.01 min) that was considerably smaller than the system time constants (approximately 5min).

We assume that measurement noise variances are known but the process disturbance intensities are unknown. From Equation (18) the appropriate outer objective function is  $(J_{1C} + J_{1T})$  shown below in Equation (21), which is minimized with respect to  $Q_{p1}$  and  $Q_{p2}$ , and the inner objective function is  $(J_{2C} + J_{2T})$  which is minimized with respect to model parameters and states, given estimates for  $Q_{p1}$  and  $Q_{p2}$ :

$$\begin{aligned}
 J_{1C} &= \frac{1}{64\sigma_{m1}^2} \left( \sum_{j=1}^{64} (y_1(t_{m1j}) - C_{A\sim}(t_{m1j}))^2 + \text{Trace} \left( \left( \frac{\partial^2 J_{2C}}{\partial C_{A\sim}^2} \right)^{-1} \right) \right) - 1 \\
 J_{2C} &= \frac{1}{2\sigma_{m1}^2} \sum_{j=1}^{64} (y_1(t_{m1j}) - C_{A\sim}(t_{m1j}))^2 \\
 &\quad + \frac{1}{2Q_{p1}} \int \left( \dot{C}_{A\sim}(t) - \left( \frac{F(t)}{V} (C_{A0}(t) - C_{A\sim}(t)) + gC_{A\sim}(t) \right) \right)^2 dt \\
 J_{1T} &= \frac{1}{213\sigma_{m2}^2} \left( \sum_{j=1}^{213} (y_2(t_{m2j}) - T_{\sim}(t_{m2j}))^2 + \text{Trace} \left( \left( \frac{\partial^2 J_{2T}}{\partial T_{\sim}^2} \right)^{-1} \right) \right) - 1 \\
 J_{2T} &= \frac{1}{2\sigma_{m2}^2} \sum_{j=1}^{213} (y_2(t_{m2j}) - T_{\sim}(t_{m2j}))^2 + \frac{1}{2Q_{p2}} \\
 &\quad \times \int \left( \dot{T}_{\sim}(t) - \left( \frac{F(t)}{V} (T_0(t) - T_{\sim}(t)) - \beta_1 (T_{\sim}(t) - T_{cin}(t)) - \beta_2 gC_{A\sim}(t) \right) \right)^2 dt
 \end{aligned} \tag{21}$$

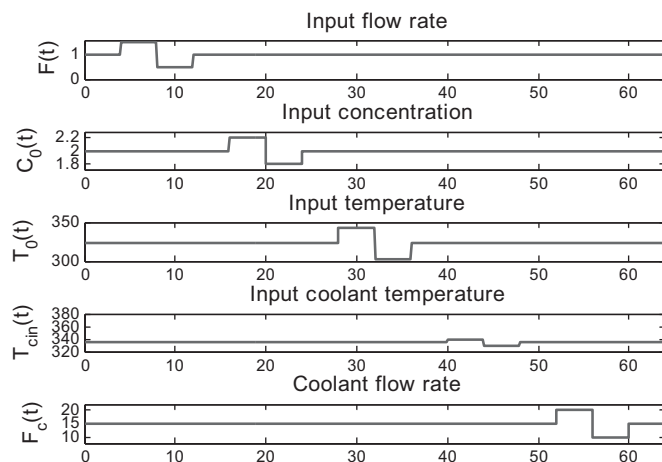


Figure 2. Input scheme for MIMO nonlinear CSTR.

For the temperature and concentration trajectories, 200 equally spaced B-spline knots were used.

The minimization was performed using the IPOPT nonlinear programming solver (Wächter and Biegler, 2006) with the model implemented using AMPL™ and the lsqnonlin routine in MATLAB™. Both of the inner and outer optimization problems can indeed be solved using MATLAB™. Based on our experience however, the computation time for solving the inner problem in MATLAB™ was prohibitively long. Thus, IPOPT and AMPL™ were used to solve the larger, inner optimization problem. Although we do not have formal convergence results for the general form of the proposed inner and outer optimization problems, we did not face any convergence issues in this case study.

In order to investigate the sampling properties of the parameter estimates, empirical sampling distributions were formed by repeating this parameter estimation problem using 500 different sets of parameter initial guesses and simulated noisy observations. The Monte Carlo histograms and scatter plot matrix for the parameter estimates are shown in Figures 2–4, respectively. The histograms in Figure 3 zoom in on the main parts of the distributions for the estimated disturbance intensities. Corresponding boxplots that show all of the 500 estimates are provided in the Appendix. Note that the distributions of the disturbance intensity estimates are broad and somewhat asymmetric.

The scatter plot matrix in Figure 5 indicates a strong nonlinear co-dependency between estimates for parameters  $a$  and  $b$ . There is noticeable linear co-dependency between estimates for  $a$  and  $k_{ref}$  and modest nonlinear co-dependency between estimates for  $a$  and  $E/R$ . Note also the strong nonlinear co-dependency between the estimate for  $a$  and  $Q_{p1}$ , the estimated disturbance intensity for the material balance. There is modest correlation between the estimates of  $k_{ref}$  and  $E/R$ . Apart from the strong co-dependency with the estimate for  $a$ , estimates for second heat transfer coefficient parameter  $b$  show relatively little co-dependency with other parameter estimates. Overall, the model parameter estimates are good.

Approximate  $100(1 - \alpha)\%$  confidence intervals for the model parameters and spline coefficients can be obtained as follows (e.g., Varziri et al., 2008):

$$\theta = \hat{\theta} \pm z_{\alpha/2} \times \sqrt{\text{diag}(\mathbf{I}^{-1}(\hat{\theta}))} \tag{22}$$

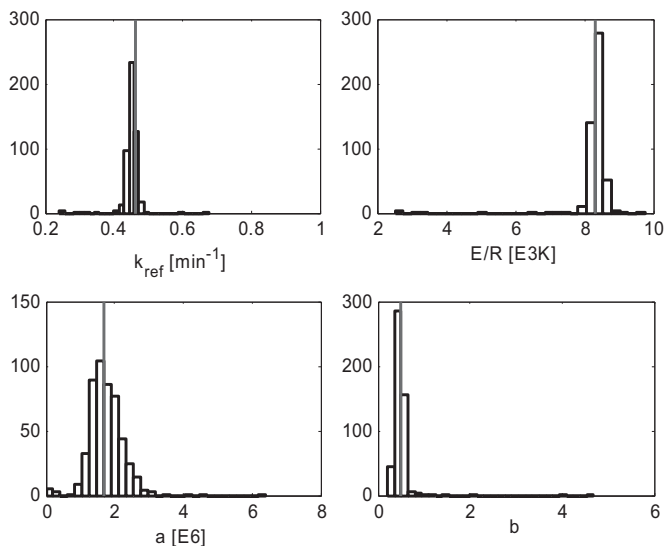


Figure 3. Histograms for AMLE parameter estimates with unknown disturbance intensity.

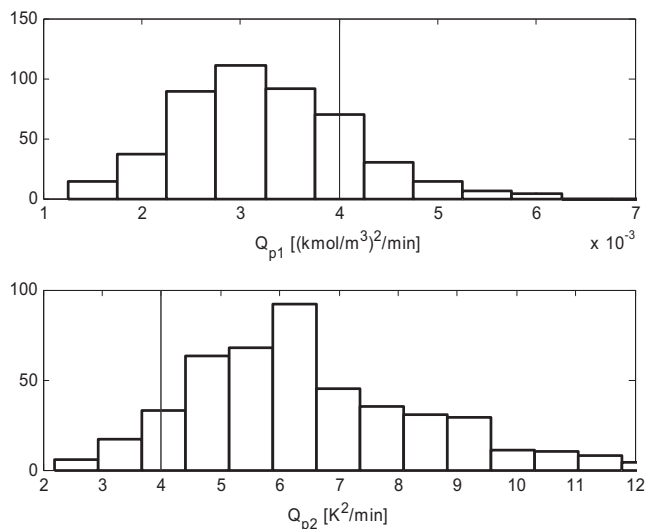


Figure 4. Histograms for AMLE process intensity estimates.

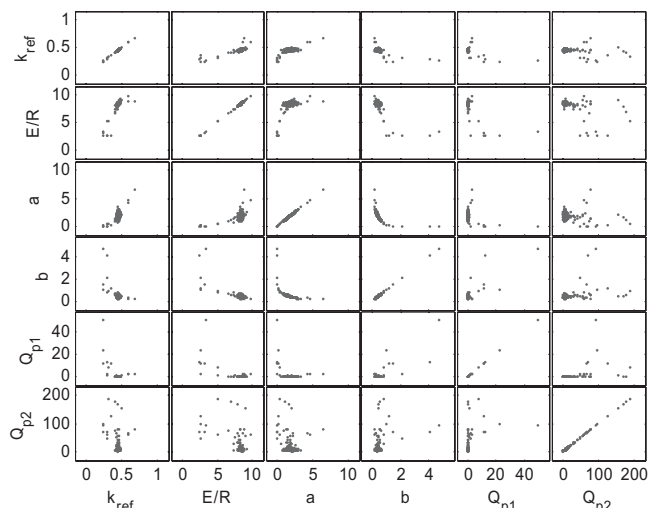


Figure 5. Scatter plot matrix for AMLE estimates.

Parameter	True value	Estimate	SD	Lower bound	Upper bound
$a$	1.678	2.0189	0.2823	1.4656	2.5722
$b$	0.5	0.4298	0.0508	0.3302	0.5294
$E/R$	8.3301	8.4492	0.0989	8.2553	8.6431
$k_{ref}$	0.4610	0.4619	0.0062	0.4498	0.4740
$C_{A0}$	1.5965	1.5769	0.0139	1.5497	1.6041
$T_0$	341.3754	340.59	0.5361	339.5379	341.6396

where  $I$  is the Fisher information matrix for the inner objective function ( $J_{2C} + J_{2T}$ ) in Equation (21). Since obtaining the Fisher information matrix involves calculating an expectation, we have used an approximation. The approximate confidence intervals are calculated based on the estimated process noise intensity and parameters, and therefore do not take into account the uncertainty that is introduced due to the variance in the estimated intensity and model parameters. Also, inaccuracies due to the Laplace approximation and also due to nonlinearity of the equations are reflected in the approximate confidence intervals. However, as shown below, these intervals are quite consistent with the empirical sampling distributions.

The parameter estimation results for one of the 500 simulated data sets are listed in Table 1. The estimated noise intensities for this run are  $\hat{Q}_{p1} = 2.39 \times 10^{-3} \text{ (kmol/m}^3\text{)}^2\text{/min}$  and  $\hat{Q}_{p2} = 3.33 \text{ (K}^2\text{/min)}$ . The approximate theoretical confidence intervals calculated using these intensities along with the estimated parameters agree with the Monte Carlo results. The corresponding AMLE estimated trajectories are presented in Figure 6. The estimated trajectories follow the true trajectories closely.

For comparison purposes, we used the same simulated data and estimated the parameters using a classical ML-based algorithm developed by Kristensen et al. (2004). This algorithm, which is based on the extended Kalman filter, estimates parameters in nonlinear stochastic differential equations models by maximizing the classical likelihood density function,  $p(y_m | \theta)$ . The estimation

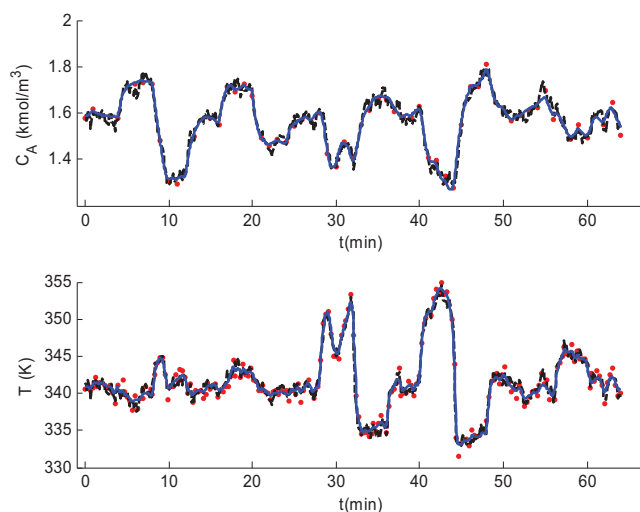


Figure 6. Estimated B-Spline trajectories of  $C_A$  and  $T$ , for one of the 500 considered samples using AMLE (—, B-spline fit; ●, measured data; --- response with true parameters and true stochastic disturbances).

**Table 2.** 95% confidence intervals for classical ML model parameter estimates [obtained using CSTM algorithm of Kristensen et al. (2003)]

Parameter	True value	Estimate	SD	Lower bound	Upper bound
$a$	1.678	2.2324	0.5612	1.1324	3.3324
$b$	0.5	0.4131	0.0916	0.2336	0.5926
$E/R$	8.3301	8.1721	0.1951	7.7897	8.5545
$k_{ref}$	0.4610	0.4715	0.0112	0.4495	0.4935
$C_{A0}$	1.5965	1.5770	0.0323	1.5137	1.6403
$T_0$	341.3754	340.55	0.7926	338.9965	342.1035

was carried out using the CSTM software developed by Kristensen et al. (2003). This software is publicly available and easy to use. The estimated noise intensities obtained using the CSTM package in this case are  $\hat{Q}_{p1} = 3.00 \times 10^{-3}$  (kmol/m<sup>3</sup>)<sup>2</sup>/min and  $\hat{Q}_{p2} = 2.63$  (K<sup>2</sup>/min). The rest of the results are presented in Table 2. Like the AMLE method proposed in this article, the CSTM software can accommodate unknown initial conditions for state variables, irregularly sampled data and unknown disturbance intensities.

The parameter estimation results are quite comparable for the two methods. The AMLE algorithm is easier to set up and converges faster than the classical ML-based method of Kristensen et al. for this case study, presumably because the proposed method does not require recursive solution of Riccati equations to obtain the Kalman gain and the estimated state covariance matrix. Parameter estimates for the complete set of 500 simulated data sets were not computed using the CSTM software because of the long run times.

## SUMMARY AND CONCLUSIONS

In this paper, an algorithm is proposed to estimate parameters, states, and process noise intensities in nonlinear continuous-time stochastic dynamic systems. The algorithm, which is an extension of the AMLE algorithm previously proposed by Varziri et al. (2008a, in press), is a two-level nonlinear minimization problem. The outer objective function is minimized with respect to the process noise intensity  $Q$  so that an estimate of the measurement noise variance is close to a known true value. The inner objective function is minimized over the model parameters  $\theta$  and states  $x$  by maximizing  $p(x, y_m | \theta)$  given the disturbance intensity  $Q$ . Setting up the proposed algorithm is much easier than classical likelihood methods in which  $p(y_m | \theta)$  is maximized. Using the likelihood corresponding to  $p(x, y_m | \theta)$ , leads to an objective function that is easy to derive and compute.

To generate confidence intervals for the estimated parameters, we have used an approximation of the inverse of the Fisher information matrix as the asymptotic covariance matrix for the estimated parameters. Even though parameter estimates that are obtained by minimizing the proposed objective function will, in general, be biased for nonlinear models, there are examples that show parameter estimates that are obtained by maximizing  $p(x, y_m | \theta)$  can be better in the sense of mean-square-error, than classical ML estimates, especially when a small amount of measured data is used (Yeredor, 2000). The proposed AMLE parameter estimates are much easier to compute than the corresponding classical ML parameter estimates because it does

not require recursive solution of Riccati equations to obtain the state covariance matrix.

We have used a simple nonlinear CSTR with two states in a simulation study to examine the effectiveness of the proposed algorithm and the sampling properties of the parameter and intensity estimates. The measurement noise variances are assumed to be known, but the process noise variances are unknown. Initial state conditions were also assumed to be unknown. Four model parameters, along with two process noise intensities and two state trajectories, were estimated. Monte Carlo simulations showed that bias in the model parameter estimates is negligible. The process disturbance intensity estimates though, were slightly biased. This bias may result from Laplace's approximation, which is used in the computation of the disturbance intensity estimates. Overall, AMLE did a good job in jointly estimating the model parameters, state trajectories and process noise intensities. Theoretical confidence intervals were obtained for the model parameter estimates. These approximate confidence intervals are in agreement with the Monte Carlo results. We compared our parameter and disturbance intensity estimates with those from a classical ML-based method, using the same simulated data. In this example, the computation time required for the AMLE algorithm was shorter than that of the classical ML-based algorithm. We believe that AMLE is a potentially appealing parameter estimation algorithm that should be further studied and tested for more complicated parameter estimation problems. Some of the beneficial features of the proposed AMLE method are as follows: simplicity of implementation, recognizing and taking into account both process disturbance (model approximations) and measurement noise, handling unknown disturbance intensities and nonstationary disturbances, efficiently handling unknown initial state conditions using empirical spline functions, handling irregularly-sampled and missing state observations, producing good parameter estimates and approximate confidence intervals.

Industrial-scale problems are more complicated than the simple example presented here. Further studies are underway to investigate the performance of AMLE in practical problems with larger numbers of inputs, outputs, and parameters, and disturbances.

## END NOTES

Please note that the SDE in (1) can be more rigorously written as  $dx(t) = f(x(t), u(t), \theta)dt + dw(t)$  where  $dw(t)$  is the increment of a Wiener process or a Brownian motion with constant diffusion  $Q$  (Maybeck, 1979).

## ACKNOWLEDGEMENTS

Financial support from Cybernetica, E. I. du Pont de Nemours and Company, Hatch, Matrikon, SAS, MITACS (Mathematics of Information Technology and Complex Systems), the Natural Sciences and Engineering Research Council of Canada (NSERC), the Government of Ontario and Queen's University is gratefully acknowledged.

## NOMENCLATURE

$a$  CSTR model parameter relating heat-transfer coefficient to coolant flow rate



$b$	CSTR model exponent relating heat-transfer coefficient to coolant flow rate
$c_i$	number of spline coefficients for state $i$
$C_A$	concentration of reactant A ( $\text{kmol m}^{-3}$ )
$C_{A0}$	feed concentration of reactant A ( $\text{kmol m}^{-3}$ )
$C_{As}$	concentration of reactant A at steady state ( $\text{kmol m}^{-3}$ )
$C_p$	reactant heat capacity ( $\text{cal g}^{-1} \text{K}^{-1}$ )
$C_{pc}$	coolant heat capacity ( $\text{cal g}^{-1} \text{K}^{-1}$ )
$E\{\cdot\}$	expected value
$E/R$	activation energy over the ideal gas constant (K)
$F$	reactant volumetric flow rate ( $\text{m}^3 \text{min}^{-1}$ )
$F_c$	coolant volumetric flow rate ( $\text{m}^3 \text{min}^{-1}$ )
$f$	nonlinear function
$\mathbf{I}$	Fisher information matrix
$k_{\text{ref}}$	kinetic rate constant at temperature $T_{\text{ref}}$ ( $\text{min}^{-1}$ )
$l$	likelihood function
$N_i$	number of observations of state $i$
$p(\cdot)$	probability density function
$t_j$	$j$ th measurement time (min)
$T$	temperature of reactor contents (K)
$T_0$	reactant feed temperature (K)
$T_{\text{cin}}$	inlet temperature of coolant (K)
$T_s$	temperature of reactant at steady state value (K)
$T_{\text{ref}}$	reference temperature (K)
$u_i$	input to the differential equation for state $i$
$V$	volume of the reactor ( $\text{m}^3$ )
$x, \mathbf{x}$	state variables
$\tilde{x}$	B-spline approximation of the state
$y$	noisy output measurements
$\mathbf{y}_m$	stacked vector of measured outputs
$z_{\alpha/2}$	normal random deviate corresponding to an upper tail area of $\alpha/2$

## Greek Symbols

$\alpha$	significance level for confidence intervals
$\beta_i$	$i$ th B-spline coefficient
$\beta$	vector of B-spline coefficients
$\delta(\cdot)$	dirac delta function
$\Delta H_{\text{rxn}}$	enthalpy of reaction ( $\text{cal g}^{-1} \text{K}^{-1}$ )
$\varepsilon_i$	normally distributed measurement noise for state $i$
$\eta_i$	White Gaussian process disturbance for differential equation of the $i$ th state
$\theta$	vector of model parameters
$\rho$	density of reactor contents ( $\text{g m}^{-3}$ )
$\rho_c$	coolant density ( $\text{g m}^{-3}$ )
$\sigma_{mi}^2$	Measurement noise variance for the $i$ th state
$\sigma_{mi}^2$	process noise intensity of stochastic differential equation for state $i$
$\varphi_i$	$i$ th B-spline basis function
$\phi$	matrix containing all $\varphi_i$ s

## Abbreviations

AMLE	approximate maximum likelihood estimation
CSTR	continuous stirred tank reactor
MAP	maximum a posteriori
MIMO	multi-input multi-output
ML	maximum likelihood
ODE	ordinary differential equation
PEN	model-based penalty
SISO	single-input single-output
SSE	sum of squared errors

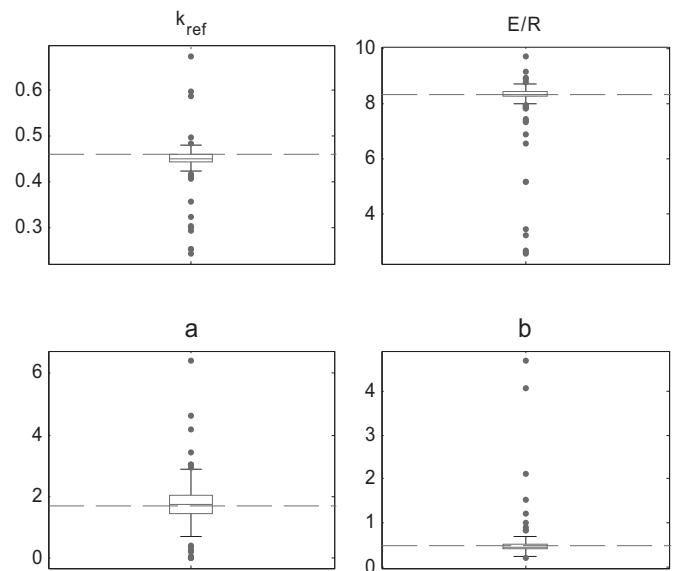


Figure 7. Boxplots for AMLE parameter estimates.

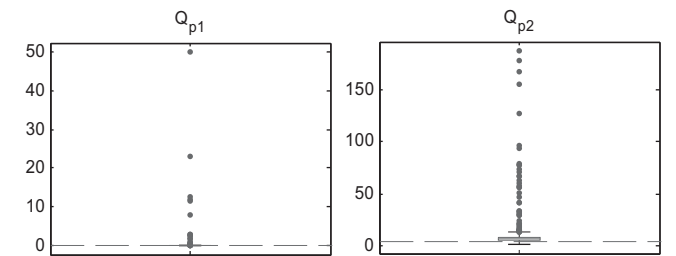


Figure 8. Boxplots for AMLE process intensity estimates.



Figure 9. Zoomed boxplots for AMLE process intensity estimates.

## APPENDIX

Figures 7–9 are given in Appendix.

## REFERENCES

- Andrieu, C., A. Doucet, S. S. Singh and V. B. Tadic, "Particle Methods for Change Detection, System Identification, and Control," Proc. IEEE 92(3), 423–438 (2004).
- Bard, Y., "Nonlinear Parameter Estimation," Academic Press, Inc., New York (1974).
- Bates, D. M. and D. G. Watts, "Nonlinear Regression Analysis and its Applications," John Wiley and Sons, Inc., New York (1988).
- Biegler, L. T., "Short Note Solution of Dynamic Optimization Problems by Successive Quadratic Programming and

- Orthogonal Collocation,” *Comput. Chem. Eng.* **8**, 243–248 (1984).
- Biegler, L. T. and I. E. Grossman, “Retrospective on Optimization,” *Comput. Chem. Eng.* **28**, 1169–1192 (2004).
- Bock, H. G., “Numerical Treatment of Inverse Problems in Chemical Reaction Kinetics,” in KH, Ebert, P, Deuflhard and W, Jäger, Eds., “Modelling of Chemical Reaction Systems,” Springer (1981), pp. 102–125 (Chapt. 8).
- Bock, H. G., “Recent Advances in Parameter Identification for Ordinary Differential Equations,” in P. Deuflhard and E. Hairer, Eds., “Progress in Scientific Computing,” Birkhäuser (1983), pp. 95–121.
- de Boor, C., “A Practical Guide to Splines,” Springer, New York (2001).
- Dee, D. P., “On-Line Estimation of Error Covariance Parameters for Atmospheric Data Assimilation,” *Mon. Weather Rev.* **123**(4), 1128–1145 (1995).
- Dee, D. P. and A. M. Da Silva, “Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part I: Methodology,” *Mon. Weather Rev.* **127**(8), 1822–1834 (1999).
- Evensen, G., “The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation,” *Ocean Dyn.* **53**, 343–367 (2003).
- Gong, J., G. Wahba, D. R. Johnson and J. Tribbia, “Adaptive Tuning of Numerical Weather Prediction Models: Simultaneous Estimation of Weighting, Smoothing, and Physical Parameters,” *Mon. Weather Rev.* **126**, 210–231 (1998).
- Goodwin, G. C. and J. C. Aguero, “Approximate EM Algorithms for Parameter and State Estimation in Nonlinear Stochastic Models,” *Proc. IEEE Conf. Dec. Cont.* 368–373 (2005).
- Heald, J. P. M. and J. Stark, “Estimation of Noise Levels for Models of Chaotic Dynamical Systems,” *Phys. Rev. Lett.* **84**, 2366–2369 (2000).
- Jazwinski, A. H., “Stochastic Processes and Filtering Theory,” Academic Press, New York (1970).
- Julier, S. J. and J. K. Uhlmann, “Unscented Filtering and Nonlinear Estimation,” *Proc. IEEE* **92**(3), 401–422 (2004).
- Kloeden, P. E. and E. Platen, “Numerical Solution of Stochastic Differential Equations,” Springer-Verlag (1992).
- Kristensen, N. R., H. Madsen and S. B. Jorgensen, “Continuous Time Stochastic Modelling-CSTM 2.3,” Technical University of Denmark, Lyngby, Denmark (2003).
- Kristensen, N. R., H. Madsen and S. B. Jorgensen, “Parameter Estimation in Stochastic Grey-Box Models,” *Automatica* **40**, 225–237 (2004).
- Mackay, D. J. C., “Information Theory, Inference, and Learning Algorithms,” version 7, Cambridge University Press (2004).
- Marlin, T. E., “Process Control: Designing Processes and Control Systems for Dynamic Performance,” 2nd ed., McGraw-Hill, New York (2000).
- Maybeck, P. S., “Stochastic Models, Estimation, and Control,” Vol. 1, Academic Press, New York (1979).
- Maybeck, P. S., “Stochastic Models, Estimation, and Control,” Vol. 2, Academic Press, New York (1982).
- Nielsen, J. N., H. Madsen and P. C. Young, “Parameter Estimation in Stochastic Differential Equations: An Overview,” *Ann. Rev. Cont.* **24**, 83–94 (2000).
- Ogunnaike, B. A. and W. H. Ray, “Process Dynamics, Modeling and Control,” Oxford University Press, New York (1994).
- Poyton, A., “Application of principal differential analysis to parameter estimation in fundamental dynamic models,” M.Sc. Thesis, Queen’s University, Kingston, Ontario, Canada (2005).
- Poyton, A. A., M. S. Varziri, K. B. McAuley, P. J. McLellan and J. O. Ramsay, “Parameter Estimation in Continuous-Time Dynamic Models Using Principal Differential Analysis,” *Comp. Chem. Eng.* **30**, 698–708 (2006).
- Ramsay, J. O. and B. W. Silverman, “Functional Data Analysis,” 2nd ed., Springer, New York (2005).
- Ramsay, J. O., G. Hooker, D. Campbell and J. Cao, “Parameter Estimation for Differential Equations: A Generalized Smoothing Approach,” *J.R. Stat. Soc. Series B (Stat. Methodol.)* **69**(5), 741–796 (2007).
- Roweis, S. and Z. Ghahramani, “Learning Nonlinear Dynamical Systems Using the Expectation-Maximization Algorithm,” in S. Haykin, Ed., “Kalman Filtering and Neural Networks,” John Wiley & Sons, New York (2001), pp. 175–220.
- Seber, G. A. F. and C. J. Wild, “Nonlinear Regression,” John Wiley and Sons, Inc., New York (1989).
- Shumway, R. H. and D. S. Stoffer, “Time Series Analysis and its Applications,” Springer, New York (2000).
- Solari, M. E., “The Maximum Likelihood Solution of the Problem of Estimating a Linear Functional Relationship,” *J. Roy. Statist. Soc. B* **31**, 372–375 (1969).
- Solo, V., “A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems,” In *Proc. IEEE ICIP96*, IEEE, IEEE press (1996), pp. 89–92.
- Varziri, M. S., K. B. McAuley and P. J. McLellan, “Parameter Estimation in Continuous-Time Dynamic Models in the Presence of Unmeasured States and Non-Stationary Disturbances,” *Ind. Eng. Chem. Res.* **47**(2), 380–393 (2008).
- Varziri, M. S., A. Poyton, K. B. McAuley, P. J. McLellan and J. O. Ramsay, “Selecting Optimal Weighting Factors in iPDA for Parameter Estimation in Continuous-Time Dynamic Models,” *Comp. Chem. Eng.* 2008a (in press, DOI: 10.1016/j.compchemeng.2008.04.005).
- Varziri, M. S., K. B. McAuley and P. J. McLellan, “Approximate Maximum Likelihood Parameter Estimation for Nonlinear Dynamic Models: Application to a Lab-Scale Nylon Reactor Model,” *Ind. Eng. Chem. Res.* 2008b (in press, DOI: 10.1021/ie800503v).
- Voss, H. U., J. Timmer and J. Kurths, “Nonlinear Dynamical System Identification From Uncertain and Indirect Measurements,” *Int. J. Bifur. Chaos.* **14**(6), 1905–1933 (2004).
- Wächter, A. and L. T. Biegler, “On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming,” *Math. Prog.* **106**(1), 25–57 (2006).
- Warnes, J. J. and B. D. Ripley, “Problems With Likelihood Estimation of Covariance Functions of Spatial Gaussian Processes,” *Biometrika* **74**(3), 640–642 (1987).
- Yeredor, A., “The Joint MAP-ML Criterion and its Relation to ML and Extended Least-Squares,” *IEEE Trans. Signal Process.* **48**(12), 3484–3492 (2000).

---

*Manuscript received February 1, 2008; revised manuscript received June 2, 2008; accepted for publication June 15, 2008.*