Parameter Estimation in Continuous-Time Dynamic Models in the Presence of Unmeasured States and Non-Stationary Disturbances

M. Saeed Varziri, Kim B. McAuley[*] and P. James McLellan

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

Saeed.varziri@chee.queensu.ca, kim.mcauley@chee.queensu.ca, james.mclellan@chee.queensu.ca

Mathematical models that describe chemical engineering processes are not exact. Therefore, it is important to develop parameter estimation algorithms that account for possible model uncertainties. In this article, as a follow-up to earlier work by Poyton et al.[1] and Varziri et al.[2], we investigate the performance of an Approximate Maximum Likelihood Estimation (AMLE) algorithm for parameter estimation in nonlinear dynamic models with model uncertainties and stochastic disturbances. We examine the applicability of AMLE to cases in which some of the states are unmeasured and we demonstrate that AMLE can be employed in models with non-stationary process disturbances. Theoretical confidence interval expressions are obtained and are compared to empirical box plots from Monte Carlo simulations. Use of the methodology is illustrated using a continuous-stirred-tank-reactor model.

[*] To whom correspondence should be addressed. E-mail: kim.mcauley@chee.queensu.ca. Phone number: +1 (613) 533 6637. Fax number: +1 (613) 533 6637

Keywords: Maximum likelihood, Principal differential analysis, Parameter estimation, Dynamic models, Non-Stationary Stochastic disturbances, State estimation, B-splines

1 Introduction

Parameter estimation in nonlinear dynamic models is generally treated as a traditional nonlinear least-squares (TNLS) minimization problem where a weighted sum of the squared deviations of the model responses from the observed data is minimized, subject to the ordinary differential equations (ODEs) that describe the dynamic model.[3-6] If the ODE model has an analytical solution then, by substituting this solution into the objective function, the parameter-estimation problem can be transformed into an unconstrained nonlinear minimization problem. Unfortunately this is rarely the case, because it is usually not possible to find analytical solutions for nonlinear ODE models that represent dynamic chemical processes. In such cases, TNLS methods require that the model ODEs, possibly along with sensitivity ODEs[7], be solved numerically and repeatedly for each parameter perturbation. Repeated numerical solution of the ODEs makes the traditional methods computationally expensive and also prone to numerical stability problems.[8]

The aforementioned problems have motivated new parameter-estimation algorithms that do not require repeated numerical solution of the differential equation models. One method to avoid repeated integration is to fit an empirical curve (e.g., smoothing splines) to the observed data.[1, 9-12] Once the empirical curve is obtained, it can be differentiated and substituted into the model differential equations, thereby transforming the ODEs into algebraic equations. However, these methods often result in biased parameter estimates since there is no guarantee that the empirical curve is consistent with solution of the ODEs.[13] Biegler and coworkers[8, 14] used collocation discretization to approximate the solution of the ODEs, assuming that the solution of the ODEs can be expressed as a linear combination of some basis functions (e.g., polynomials). The constant coefficients of the (polynomial) basis functions are obtained by requiring the response trajectories to satisfy the ODEs. Basis-function representation of the response

trajectories transforms the ODEs into algebraic equations, and hence the ODE constraints in the nonlinear parameter estimation problem become algebraic constraints. The main expense of collocation-based methods is that they usually require the solution of a large-scale nonlinear minimization problem to simultaneously select the basis-function coefficients and determine the optimal parameter estimates. Fortunately, advanced optimization algorithms are capable of carrying out these kinds of problems quickly and efficiently.[15]

Traditional nonlinear least-squares and collocation-based methods enforce the model equations (in ODE and algebraic form, respectively) as hard constraints, implying that the structure of the model is perfect. More often than not, however, mathematical models describing chemical processes are only approximately true, and discrepancies between the model and physical reality should not be neglected. When it is not appropriate to assume that the model structure is perfect, implementing the model equations as hard constraints in the parameter-estimation problem results in biased or inconsistent parameter estimates.

Poyton et al.[1] showed that iterative Principal Differential Analysis (iPDA) is an effective method for parameter estimation in ODE models. In iPDA, basis functions (i.e., B-splines) are used to discretize the model ODEs, transforming them into algebraic equations. Unlike the previous collocation-based methods, Poyton's technique treats the model equations as soft constraints, allowing for possible model imperfections. The iPDA objective function was originally minimized in an iterative way, with each iteration consisting of two steps. Optimal B-spline coefficients are obtained in the first step, given the most recent parameter estimates; in the second step, fundamental model parameters are estimated using the B-spline coefficients obtained from the first step. The iterations continue until the parameter estimates converge.

Recently, Ramsay et al.[16] proposed another spline-based method that also accounts for possible model uncertainties. In this profile-based method, the dimensionality of the parameter estimation problem is reduced because the spline coefficients are treated as functions of the fundamental model parameters. Ramsay's method uses a two-level optimization scheme wherein the outer (primary) optimizer

determines optimal values of fundamental model parameters to minimize the sum of squared prediction errors. The inner (secondary) optimizer selects spline coefficients to minimize a weighted sum of the squared prediction errors and a model-based penalty using the parameter estimates from the outer optimization loop. In the present article we focus on Poyton's iPDA objective function since it has a direct relationship to the likelihood function of stochastic differential equation models that are appropriate for describing chemical processes.

We have shown that minimizing the iPDA objective function is equivalent to maximizing the log-likelihood of the conditional joint density function of the states and measurements, given the fundamental model parameters, when model mismatch results from additive stochastic white-noise disturbance inputs[2]. We have also shown that fundamental model parameters and B-spline coefficients can be minimized simultaneously. Since we no longer solve the optimization problem using the two-step iterative method, and because a likelihood criterion is approximated using B-splines, we call the technique Approximate Maximum Likelihood Estimation (AMLE) throughout the remainder of this article.

The contributions in the current article are as follows. We demonstrate that AMLE can be readily used in parameter estimation in cases in which some of the states are not observed. We also show that AMLE can estimate the unmeasured states along with fundamental model parameters. This capability naturally leads to the application of AMLE to parameter estimation in dynamic models driven by non-stationary process disturbances, because non-stationary disturbances can be considered to be unmeasured states.[17] Another important contribution is the development of theoretical confidence-interval expressions for parameter estimates and B-spline coefficients. The application of the AMLE technique is examined using a multi-input multi-output (MIMO) nonlinear Continuous Stirred Tank Reactor (CSTR) model. In the CSTR case study, theoretical confidence intervals are compared to confidence intervals from Monte Carlo simulations to confirm the theoretical results.

The manuscript is organized as follows. First, we briefly review the AMLE algorithm using a multivariate nonlinear dynamic model. Detailed information about the algorithm and its mathematical

basis can be found in Poyton et al.[1] and Varziri et al.[2] We then show how AMLE can be used for cases in which unmeasured states or non-stationary disturbances are present and we obtain theoretical confidence-interval expressions for the fundamental model parameters. Next, we use a nonlinear CSTR example to study the effectiveness of AMLE in several different scenarios: i) when all states are measured, ii) when temperature is measured, but the concentration is not, and iii) when a non-stationary disturbance enters the material-balance differential equation and both states are measured. Finally, we highlight some of the remaining challenges that need to be addressed to make AMLE more applicable for parameter estimation in dynamic chemical process models.

1.1 Review of the AMLE algorithm

Varziri et al. [2] derived the AMLE objective function using a Bayesian argument for a multi-input multi-output first-order nonlinear dynamic system. Here, for simplicity, we use a first-order model with two inputs and two states to review the AMLE algorithm:

$$\begin{cases} \dot{x}_1(t) = f_1(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + \eta_1(t) \\ \dot{x}_2(t) = f_2(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + \eta_2(t) \\ \\ x_1(0) = x_{10} \\ x_2(0) = x_{20} \\ \\ y_1(t_{m1j}) = x_1(t_{m1j}) + \varepsilon_1(t_{m1j}) \\ y_2(t_{m2j}) = x_2(t_{m2j}) + \varepsilon_2(t_{m2j}) \end{cases} \tag{1}$$

x_1 and x_2 are state variables, u_1 and u_2 are input variables, and y_1 and y_2 are output variables. f_1 and f_2 are suitably well-behaved nonlinear functions (Lipschitz continuous) and $\boldsymbol{\theta}$ is the vector of fundamental model parameters. $\eta_1(t)$ and $\eta_2(t)$ are independent continuous zero-mean stationary white-noise processes with intensities Q_{p1} and Q_{p2} respectively. ε_1 and ε_2 are zero-mean independent Normal random variables with variances σ_{m1}^2 and σ_{m2}^2 respectively. t_{m1j} and t_{m2j} are the time points at which outputs y_1 and y_2 are measured. We assume that there are N_1 and N_2 measurement times for y_1 and y_2, respectively.

In AMLE the system states trajectories are approximated using linear combinations of B-spline basis functions.[18, 19]

$$x_{\sim k}(t) = \sum_{i=1}^{c_k} \beta_{ki}\phi_{ki} \qquad k = 1,2 \tag{2}$$

where β_{ki}, $i = 1 \cdots c_k$ and $\phi_{ki}(t)$ $i = 1 \cdots c_k$ are B-spline coefficients and B-spline basis functions, respectively, for the kth state. Equation (2) can be written in matrix form:

$$x_{\sim k}(t) = \varphi_k^T(t)\boldsymbol{\beta}_k \qquad k = 1,2 \tag{3}$$

where $\varphi_k(t)$ is a vector containing the c_k basis functions and $\boldsymbol{\beta}_k$ is vector of c_k spline coefficients. Note that the "~" subscript is used to imply an empirical curve that can be easily differentiated:

$$\dot{x}_{\sim k}(t) = \frac{d}{dt}\left(\sum_{i=1}^{c_k} \beta_{ki}\phi_{ki}(t)\right) = \sum_{i=1}^{c_k} \beta_{ki}\dot{\phi}_{ki}(t) = \dot{\varphi}_k^T\boldsymbol{\beta}_k \qquad k = 1,2 \tag{4}$$

In AMLE, the spline coefficients $\boldsymbol{\beta}_k$ and the vector of fundamental model parameters $\boldsymbol{\theta}$ are obtained so that the following objective function is minimized:

$$\begin{aligned}
&\frac{1}{\sigma_{m1}^2}\sum_{j=1}^{N_1}(y_1(t_{m1j}) - x_{\sim 1}(t_{m1j}))^2 + \\
&\frac{1}{\sigma_{m2}^2}\sum_{j=1}^{N_2}(y_2(t_{m2j}) - x_{\sim 2}(t_{m2j}))^2 + \\
&\frac{1}{Q_{p1}}\int\left(\dot{x}_{\sim 1}(t) - f_1(x_{\sim 1}(t), x_{\sim 2}(t), u_1(t), u_2(t), \boldsymbol{\theta})\right)^2 dt + \\
&\frac{1}{Q_{p2}}\int\left(\dot{x}_{\sim 2}(t) - f_2(x_{\sim 1}(t), x_{\sim 2}(t), u_1(t), u_2(t), \boldsymbol{\theta})\right)^2 dt
\end{aligned} \tag{5}$$

We will refer to the first two terms $\sum(y_k(t_{mj}) - x_{\sim k}(t_{mj}))^2$ in the objective function as SSE (the sum of squared prediction errors), while PEN (the model-based penalty) will be used to refer to the third and fourth terms $\int\left(\dot{x}_{\sim k}(t) - f_k(x_{\sim 1}(t), x_{\sim 2}(t), u_1(t), u_2(t), \boldsymbol{\theta})\right)^2 dt$. The SSE and PEN terms in (5) are weighted by the reciprocals of the measurement noise variances and process disturbance intensities, respectively. In its original form, AMLE (which was called iPDA by Poyton et al.[1]) left the values of these weighting coefficients unspecified as tuning parameters that could be adjusted by trial and error, and the tuning

requirement was therefore listed as a disadvantage of the methodology. Recently, we used a Bayesian argument to show that, in the dynamic model described in (1), the optimal weighting coefficients are the reciprocals of the measurement noise variances for the SSE terms and reciprocals of the process disturbance intensities for the PEN terms and hence the objective function in (5) was derived.[2] However, we acknowledge that selecting the optimal weighting factors requires knowledge of measurement variances and process disturbance intensities. In real-world applications, the true values of these parameters are not known and estimates must be obtained. Measurement noise variances can be estimated from replicate measurements; however, estimating the process disturbance intensities is a difficult task requiring expert knowledge of the dynamic system and the corresponding mathematical model. Therefore, efficient algorithms for estimating the unknown variances and intensities need to be developed, which is a subject of our ongoing research.

Another shortcoming highlighted in our previous publications on AMLE was the lack of a systematic method to assess the variability of the parameter estimates (and the B-spline coefficients). We address this problem in Section 3 of the current article where we present the confidence interval results. We briefly review the major advantages of AMLE:

- AMLE provides an easy-to-implement method for estimating fundamental model parameters and system states in dynamic models in which stochastic process disturbances are present.

- The AMLE objective considers model uncertainties and measurement noise at the same time, and the trade off between poor measurements and model imperfections is addressed by the optimal choice of weighting coefficients.

- AMLE uses B-spline basis functions to discretize the differential equations in the dynamic model, thereby transforming them into algebraic equations. Hence, AMLE circumvents the repeated numerical integration used by traditional methods.

- The discretization approach used in AMLE removes the requirement of deriving and integrating the sensitivity differential equations used by conventional methods.

- AMLE inherits and combines the advantages of Kalman filters for state and parameter estimation in stochastic dynamic models and the benefits of the collocation-based methods for state and parameter estimation in deterministic dynamic models.

In addition to these advantages, we will demonstrate in the next section that AMLE can readily be applied to problems with unmeasured states and non-stationary disturbances, which are important in chemical processes.

2 Unmeasured states and non-stationary disturbances

Handling unmeasured states in AMLE is straightforward. There is no SSE term for the unmeasured state in the objective function, because the number of measurements in the corresponding summation is zero. For instance, if x_1 is not measured, the new objective function is:

$$\begin{aligned} &\frac{1}{\sigma_{m2}^2}\sum_{j=1}^{N_2}(y_2(t_{m2j})-x_{\sim 2}(t_{m2j}))^2 + \\ &\frac{1}{Q_{p1}}\int\left(\dot{x}_{\sim 1}(t)-f_1(x_{\sim 1}(t),x_{\sim 2}(t),u_1(t),u_2(t),\boldsymbol{\theta})\right)^2 dt + \\ &\frac{1}{Q_{p2}}\int\left(\dot{x}_{\sim 2}(t)-f_2(x_{\sim 1}(t),x_{\sim 2}(t),u_1(t),u_2(t),\boldsymbol{\theta})\right)^2 dt \end{aligned} \tag{6}$$

When a state is measured, there are two sources of information (i.e., the measured data and the corresponding differential equation) that can be used to estimate the states and eventually the model parameters. However, with an unmeasured state we are only left with the differential equation and hence, the B-spline curve should be flexible enough to follow the solution of the differential equation quite closely. More flexibility can be achieved by placing more knots within the time frame over which the integrals in (6) are evaluated. Parameter estimation in a nonlinear MIMO CSTR with an unmeasured state is studied in Section 4.2 as an example.

So far, we have considered dynamic models with stationary Gaussian process disturbances. Dynamic models with non-stationary disturbances can be handled by considering these disturbances as unmeasured states. To accommodate such problems, we consider the model in (7) below, which is a slightly modified version of the model in (1). We have added a non-stationary disturbance ($d_1(t)$) to the

right-hand-side of the first differential equation. Please note that it is possible to add a non-stationary disturbance to any of the differential equations in the model.

In (7), $d_1(t)$ is a non-stationary disturbance driven by $\eta_{d1}(t)$ which is a continuous zero-mean stationary white-noise process with intensity $Q_{pd1.}$ $f_{d1}(.)$ is a function defining the non-stationary noise model and $\boldsymbol{\theta}_{d1}$ is the vector of noise model parameters.[20] If $\boldsymbol{\theta}_{d1}$ is not known (which generally is the case) it can be estimated along with the vector of fundamental model parameters $\boldsymbol{\theta}$.

$$\begin{cases} \dot{x}_1(t) = f_1(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + d_1(t) + \eta_1(t) \\ \dot{x}_2(t) = f_2(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + \eta_2(t) \\ \dot{d}_1(t) = f_{d1}(d_1(t), \boldsymbol{\theta}_{d1}) + \eta_{d1}(t) \\ \\ x_1(0) = x_{10} \\ x_2(0) = x_{20} \\ d_1(0) = 0 \\ \\ y_1(t_{m1j}) = x_1(t_{m1j}) + \varepsilon_1(t_{m1j}) \\ y_2(t_{m2j}) = x_2(t_{m2j}) + \varepsilon_2(t_{m2j}) \end{cases} \tag{7}$$

By treating $d_1(t)$ like any other unmeasured state, assuming that x_1 and x_2 are measured, the AMLE objective function can be written as:

$$\begin{aligned} &\frac{1}{\sigma_{m1}^2}\sum_{j=1}^{N_1}(y_1(t_{m1j}) - x_{\sim 1}(t_{m1j}))^2 + \\ &\frac{1}{\sigma_{m2}^2}\sum_{j=1}^{N_2}(y_2(t_{m2j}) - x_{\sim 2}(t_{m2j}))^2 + \\ &\frac{1}{Q_{p1}}\int\left(\dot{x}_{\sim 1}(t) - f_1(x_{\sim 1}(t), x_{\sim 2}(t), d_{\sim 1}(t), u_1(t), u_2(t), \boldsymbol{\theta})\right)^2 dt + \\ &\frac{1}{Q_{p2}}\int\left(\dot{x}_{\sim 2}(t) - f_2(x_{\sim 1}(t), x_{\sim 2}(t), u_1(t), u_2(t), \boldsymbol{\theta})\right)^2 dt + \\ &\frac{1}{Q_{d1}}\int\left(\dot{d}_{\sim 1}(t) - f_{d1}(d_{\sim 1}(t), \boldsymbol{\theta}_{d1})\right)^2 dt \end{aligned} \tag{8}$$

where, $d_{\sim 1}(t) = \varphi_{d1}^T(t)\boldsymbol{\beta}_{d1}$. Note that (8) is minimized with respect to $\boldsymbol{\theta}$, $\boldsymbol{\theta}_{d1}$, $\boldsymbol{\beta}$, and $\boldsymbol{\beta}_{d1}$.

As an example, parameter estimation in a nonlinear MIMO CSTR with a non-stationary disturbance is studied in Section 4.3.

3 Theoretical confidence intervals

Varziri et al.[2] considered the model in (1) and showed that minimizing the AMLE objective function is equivalent to maximizing the likelihood of the conditional joint density function of states and measured output values given the model parameter values, $p(\mathbf{x},\mathbf{y}_m \mid \boldsymbol{\theta})$, where $\mathbf{x}$ is the vector of states and $\mathbf{y}_m$ is the vector of measured outputs (this likelihood function is recommended by Maybeck [21] as the most appropriate for combined state and parameter estimation). Therefore, in AMLE we maximize the log-likelihood of $p(\mathbf{x},\mathbf{y}_m \mid \boldsymbol{\theta})$, denoted by $L(\mathbf{x},\boldsymbol{\theta},\mathbf{y}_m)$:

$$L(\mathbf{x},\boldsymbol{\theta},\mathbf{y}_m) = \ln\big(p(\mathbf{x},\mathbf{y}_m \mid \boldsymbol{\theta})\big) \tag{9}$$

$$\hat{\boldsymbol{\theta}},\hat{\mathbf{x}} = \arg\min_{\boldsymbol{\theta},\mathbf{x}} \{-L(\mathbf{x},\boldsymbol{\theta},\mathbf{y}_m)\} \tag{10}$$

Analogous objective functions have been used in nonlinear filtering problems.[22-24] Under general regularity conditions, parameter estimates obtained from the minimization problem in (10) are consistent, asymptotically unbiased, asymptotically Normally-distributed and asymptotically efficient.[21, 25, 26] To obtain confidence intervals for the model parameters (and B-spline coefficients), we exploit the properties of the maximum likelihood estimator.

Assuming that x can be represented using B-splines as in (3), the parameter estimation problem reduces to:

$$\min_{\boldsymbol{\theta},\boldsymbol{\beta}} \{-L(\boldsymbol{\beta},\boldsymbol{\theta},\mathbf{y}_m)\} \tag{11}$$

By stacking the fundamental model parameters and spline coefficients in a vector $\boldsymbol{\tau} = [\boldsymbol{\theta}^T, \boldsymbol{\beta}^T]^T$ the problem becomes:

$$\min_{\boldsymbol{\tau}} \{-L(\boldsymbol{\tau},\mathbf{y}_m)\} \tag{12}$$

We assume that L satisfies the required regularity conditions (e.g., Kay [26], Appendix 7B, page 212). Then, from the asymptotic properties of maximum likelihood estimators (e.g. Kay [26], Theorem 7.3, page 183) we have:

$$\hat{\boldsymbol{\tau}} \sim N(\boldsymbol{\tau}, \mathbf{I}^{-1}(\boldsymbol{\tau})) \tag{13}$$

where $\mathbf{I}(\boldsymbol{\tau})$ is the Fisher information matrix[3] evaluated at the true values of $\boldsymbol{\tau}$:

$$\mathbf{I}(\boldsymbol{\tau}) = -E\left\{\frac{\partial^2 L(\boldsymbol{\tau}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}^2}\right\} = E\left\{\frac{\partial L(\boldsymbol{\tau}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}}^T \frac{\partial L(\boldsymbol{\tau}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}}\right\} \tag{14}$$

In the examples of Section 4, we have approximated $\mathbf{I}(\boldsymbol{\tau})$ by equation (A. 5) in Appendix A evaluated at $\hat{\boldsymbol{\tau}}$.

Approximate 100(1-α)% confidence intervals for the model parameters and spline coefficients can be obtained as follows:

$$\boldsymbol{\tau} = \hat{\boldsymbol{\tau}} \pm z_{\alpha/2} \times \mathrm{diag}(\mathbf{I}^{-1}(\hat{\boldsymbol{\tau}})) \tag{15}$$

The confidence intervals obtained from (15) are approximate because:

1- The model is nonlinear with respect to the parameters and spline coefficients;

2- The maximum likelihood estimates are asymptotically Normal;

3- The state trajectories are approximated by B-splines;

4- The true parameter values are not known; therefore, the approximate Fisher information matrix is evaluated at the estimated parameter values rather than true parameter values, and the observed measurements rather than taking expected values over the joint distribution of the observations.

Note also that we have used the standard Normal random deviate in equation (15), rather than a value from student's t distribution because we are assuming that the noise variance is known.

4 Case study

In this section we examine our results using a MIMO nonlinear CSTR model[27]. In Section 4.1 we estimate 4 parameters in the CSTR model, assuming that both temperature and concentration can be measured. In Section 4.2, we assume that only temperature can be measured and concentration is unmeasured. In Section 4.3, we consider the CSTR where both states are measured and non-stationary disturbances are present in the system.

We use the simulations to compare parameter estimation results obtained using the proposed AMLE algorithm and TNLS. In all case studies, the Simulink™ toolbox of MATLAB™ was used to solve the nonlinear dynamic models (using the ode45 solver) and to generate noisy measurements. TNLS parameter estimates were obtained using the lsqnonlin optimizer (default solver option), which employs a subspace trust region method and is based on the interior-reflective Newton method. Sensitivity equations were solved along with the model differential equations.

The AMLE objective function, in all examples, was optimized simultaneously over the fundamental model parameters and B-spline coefficients using the IP-OPT solver[15] via AMPL™. Initially, we tried to use lsqnonlin of MATLAB™ for minimizing the AMLE objective function; although we had satisfactory convergence, the computation time was prohibitively high due to the large dimension of combined vector of model parameters and B-spline coefficients. Hence, we switched to IP-OPT, which converged very quickly so that the computation time was not an issue. We note again that AMLE, and collocation-based methods in general, circumvent potential problems associated with numerical integration at the expense of solving a large nonlinear programming problem; hence, fast and efficient nonlinear programming solvers such as IP-OPT[15] are essential for the successful implementation of these algorithms, especially for larger problems.

4.1 Nonlinear MIMO CSTR with measured temperature and concentration

We consider the following model that represents a MIMO nonlinear CSTR. The model equations consist of material and energy balances[27] with additional stochastic disturbance terms:

$$\frac{dC_A(t)}{dt} = \frac{F(t)}{V}\left(C_{A0}(t) - C_A(t)\right) - gC_A(t) + \eta_1(t)$$
$$\frac{dT(t)}{dt} = \frac{F(t)}{V}\left(T_0(t) - T(t)\right) + \beta_1\left(T(t) - T_{cin}(t)\right) - \beta_2 gC_A(t) + \eta_2(t)$$

$$C_A(0) = 1.569\ (\text{kmol m}^{-3}) \qquad (16)$$
$$T(0) = 341.37\ (\text{K})$$

$$y_1(t_i) = C_A(t_i) + \varepsilon_1(t_i)$$
$$y_2(t_j) = C_A(t_j) + \varepsilon_2(t_j)$$

,

$$g = k_{ref}\exp\left(-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right),\ \beta_1 = -\frac{aF_c^{b+1}(t)}{V\rho C_p\left(F_c(t) + \frac{aF_c^b(t)}{2\rho_c C_{pc}}\right)}\ \beta_2 = \frac{\left(-\Delta H_{rxn}\right)}{\rho C_p}$$

where $E\left\{\eta_1(t_i)\eta_1(t_j)\right\} = \sigma_{p1}^2\delta(t_i - t_j)$, $E\left\{\eta_2(t_i)\eta_2(t_j)\right\} = \sigma_{p2}^2\delta(t_i - t_j)$ ($\delta(.)$ is the Dirac delta function), $\varepsilon_1(t_{mj})$ $j = 1..N_1$ and $\varepsilon_2(t_{mj})$ $j = 1..N_2$ are white-noise sequences with variances σ_{m1}^2 and σ_{m2}^2, respectively. We also assume that η_1, η_2, ε_1, and ε_2 are independent.

This stochastic differential equation model is nonlinear in the states (C_A, T) and parameters and does not have an analytical solution.

C_A is the concentration of the reactant A, T is the reactor temperature, V is the volume and T_{ref} = 350 K is a reference temperature. The true values of the parameters to be estimated are: E/R = 8330.1 K, k_{ref} = 0.461 min^{-1}, a =1.678E6, $b = 0.5$. The initial parameter guesses were set at 50% of the true parameter values. 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_{cin}, and the coolant flow rate F_c. Values for the various other known constants[27] are as follows: $V = 1.0$ m^3, $C_p = 1$ cal g^{-1}K^{-1}, ρ = 1E6 g m^{-3}, $C_{pc} = 1$ cal g^{-1}K^{-1}, ρ_c = 1E6 g m^{-3}, and $-\Delta H_{rxn}$ = 130E6 cal kmol^{-1}. The initial steady-state operating point is: C_{As} = 1.569 kmol m^{-3} and T_s = 341.37 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 1. [28]

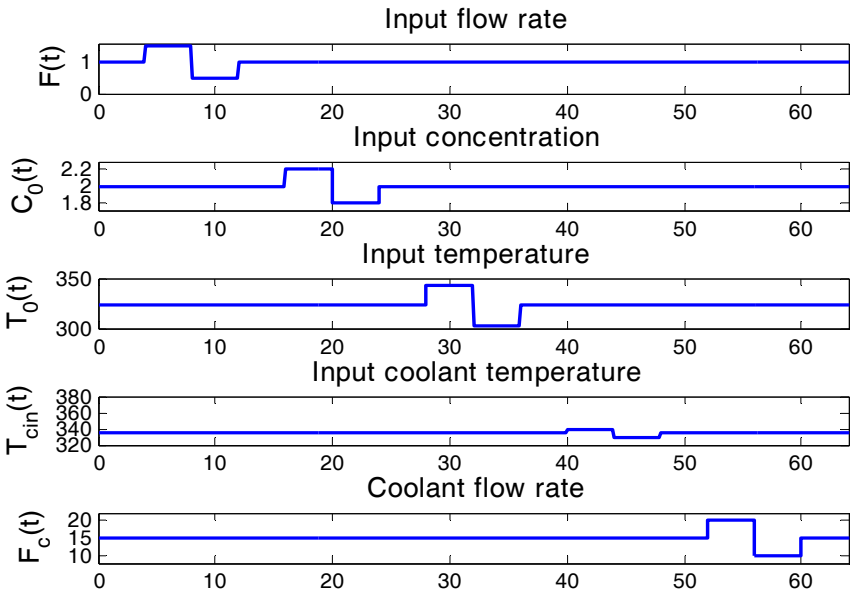

Figure 1 Input scheme for MIMO nonlinear CSTR

Each input consists of a step up, followed by a step down, and then a step back to the steady-state point.

We assume that concentration and temperature are measured and initial values are known Temperature is measured once every 0.3 minutes while concentration is measured once per minute. The duration of the simulated experiment is 64 minutes, so that there are 213 temperature measurements and 64 concentration measurements. The noise variance for the concentration and temperature measurements are $\sigma_{m1}^2 = 4\times10^{-4}\ (\text{kmol/m}^3)^2$ and $\sigma_{m2}^2 = 6.4\times10^{-1}\ \text{K}^2$ respectively. The corresponding process noise intensities for the stochastic disturbances are $Q_{p1} = 4\times10^{-3}\ (\text{kmol/m}^3)^2/\text{min}$ and $Q_{p2} = 4\,\text{K}^2/\text{min}$. From equation (5) , the AMLE objective function is:

$$
\begin{aligned}
&\frac{1}{\sigma_{m1}^2}\sum_{j=1}^{64}\left(y_1(t_{m1j})-C_{A\sim}(t_{m1j})\right)^2+\\
&\frac{1}{\sigma_{m2}^2}\sum_{j=1}^{213}\left(y_2(t_{m2j})-T_{\sim}(t_{m2j})\right)^2+\\
&\frac{1}{Q_{p1}}\int_{t=0}^{64}\left(\frac{dC_{A\sim}(t)}{dt}-\frac{F(t)}{V}\left(C_{A0}(t)-C_{A\sim}(t)\right)+gC_{A\sim}(t)\right)^2dt+\\
&\frac{1}{Q_{p2}}\int_{t=0}^{64}\left(\frac{dT_{\sim}(t)}{dt}-\frac{F(t)}{V}\left(T_0(t)-T_{\sim}(t)\right)-\beta_1\left(T_{\sim}(t)-T_{cin}(t)\right)+\beta_2 gC_{A\sim}(t)\right)^2dt
\end{aligned}
\tag{17}
$$

For the temperature trajectory, B-spline knots were placed at the observation times (one knot at every 0.3 minutes) and for the concentration trajectory, we placed one knot at every 0.2 minutes. For both the concentration and temperature trajectories, 5 Gaussian Quadrature points were used between every two knots to numerically calculate the integrals in (17).

We repeated this parameter estimation problem using 100 different sets of noisy observations to study the sampling properties of the parameter estimates. The Monte Carlo box plots for AMLE and TNLS parameter estimates are shown in Figures 2 to 5. The AMLE and TNLS predicted responses are with the true responses in Figures 6 to 9. On average, the AMLE parameter estimates are better than than the TNLS parameter estimates (more precise and less biased) and the AMLE response trajectories are closer to the true trajectories than are the TNLS trajectories.

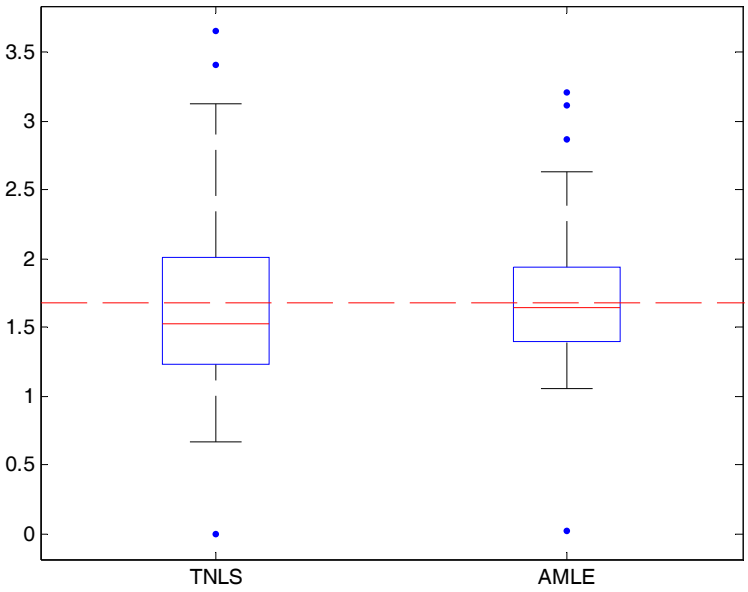

Figure 2 Box plots for a using TNLS and AMLE

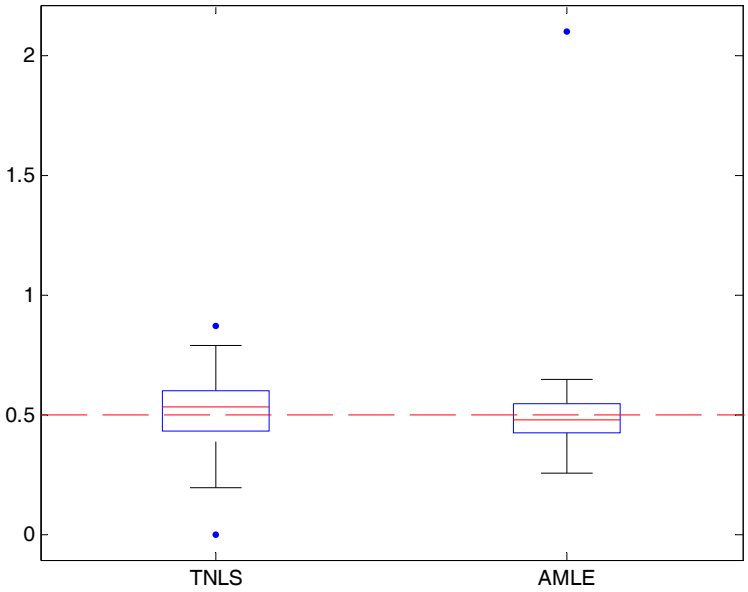

Figure 3 Box plots for b using TNLS and AMLE

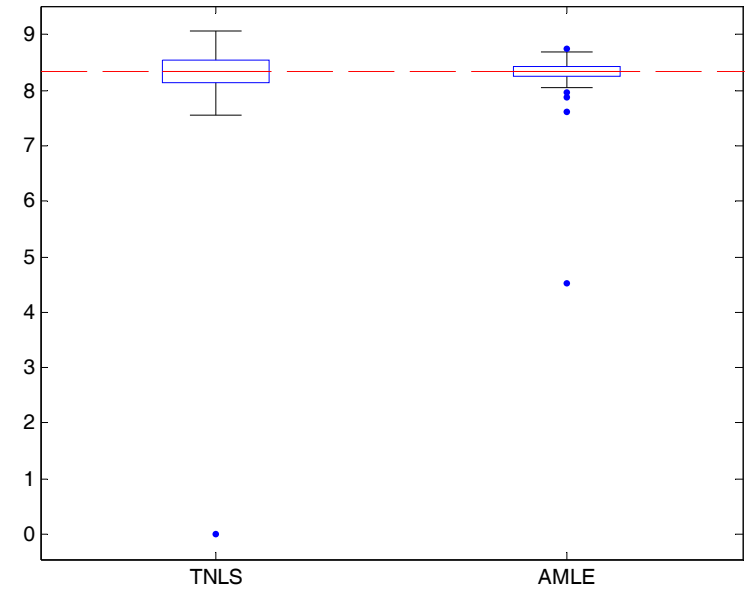

Figure 4 Box plots for $\frac{E}{R}$ using TNLS and AMLE

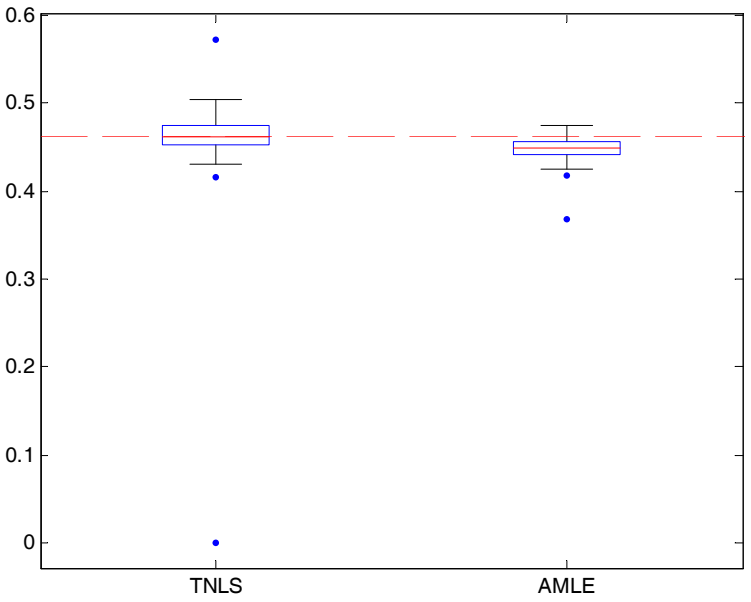

Figure 5 Box plots for k_{ref} using TNLS and AMLE

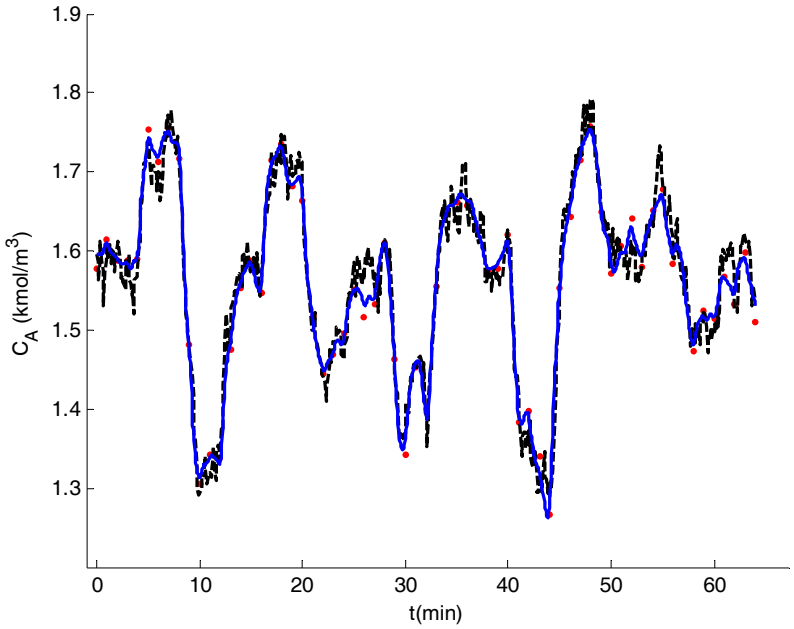

Figure 6 Observed, true, and predicted concentration response for AMLE

(• simulated data, ---- response with true parameters and true stochastic noise, — AMLE response)

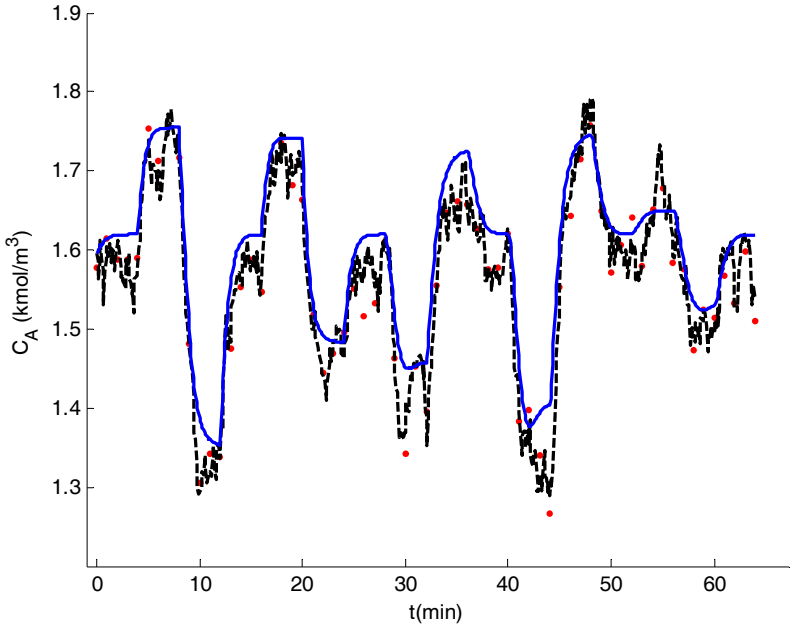

Figure 7 Observed, true, and predicted concentration response for TNLS

(• simulated data, ---- response with true parameters and true stochastic noise, — TNLS response)

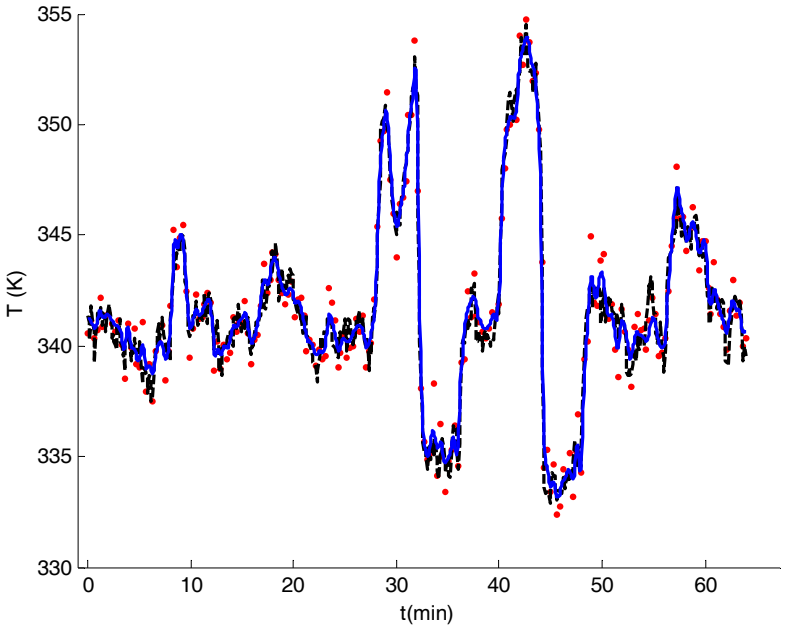

Figure 8 Observed, true, and predicted temperature response for AMLE

(• simulated data, ---- response with true parameters and true stochastic noise, — AMLE response)

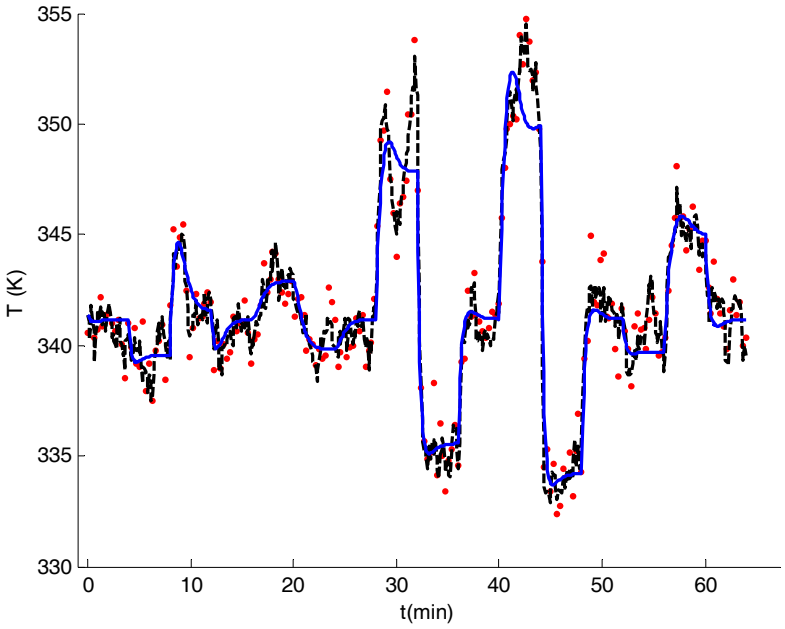

Figure 9 Observed, true, and predicted temperature response for TNLS

(• simulated data, ---- response with true parameters and true stochastic noise, — TNLS response)

Table 2 presents 95% confidence intervals (derived in Section 3) for this case study. Please note that the confidence intervals shown correspond to one particular data set chosen randomly from the 100 data sets used to generate the box plots, and the approximate Fisher Information matrix is evaluated using these parameter estimates.

Table 1. 95% Confidence Intervals for AMLE parameter estimates

Parameter Estimates	Lower Bound	Upper Bound
a	1.4501	2.6322
b	0.3307	0.5400
E/R	8.3152	8.7592
k_{ref}	0.4575	0.4866

By inspecting the Monte Carlo box plots and the scatter plot matrix given in Appendix B, and comparing them to the theoretical results in Table 2, we see that the empirical and theoretical confidence intervals are consistent.

4.2 Nonlinear MIMO CSTR with unmeasured concentration

In this example we consider the same CSTR model as in the previous section. The only difference is that concentration is unmeasured.

From equation (6) , the appropriate AMLE objective function is:

$$\begin{aligned}
&\frac{1}{\sigma_{m2}^2}\sum_{j=1}^{213}\left(y_2(t_{m2j})-T_{\sim}(t_{m2j})\right)^2+\\
&\frac{1}{Q_{p1}}\int_{t=0}^{64}\left(\frac{dC_{A\sim}(t)}{dt}-\frac{F(t)}{V}\left(C_{A0}(t)-C_{A\sim}(t)\right)+gC_{A\sim}(t)\right)^2dt+\\
&\frac{1}{Q_{p2}}\int_{t=0}^{64}\left(\frac{dT_{\sim}(t)}{dt}-\frac{F(t)}{V}\left(T_0(t)-T_{\sim}(t)\right)-\beta_1\left(T_{\sim}(t)-T_{cin}(t)\right)+\beta_2 gC_{A\sim}(t)\right)^2dt
\end{aligned} \tag{18}$$

because N_1=0. We used 100 sets of noisy observations to study the sampling properties of the parameter estimates. The Monte Carlo box plots are shown in Figures 10 to 13. The AMLE and TNLS predicted responses are compared against the true responses in Figures 14 to 17.

From Figures 10 to 17 we observe that, on average, AMLE parameter estimates are better than those of TNLS, and the AMLE response trajectories are closer to the true trajectories than are the TNLS trajectories. We also observe that, since concentration is not measured, both TNLS and AMLE parameter estimates are worse than those in the previous section when both states were measured.

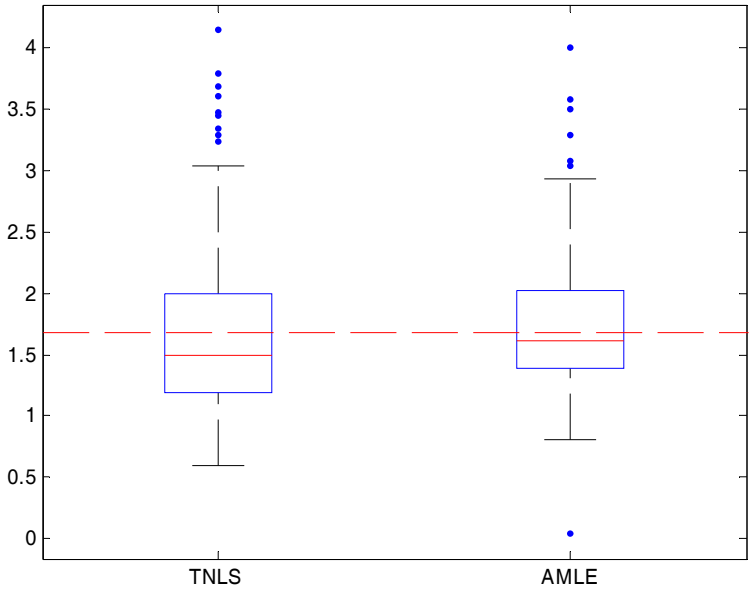

Figure 10 Box plots for a using TNLS and AMLE (unmeasured concentration)

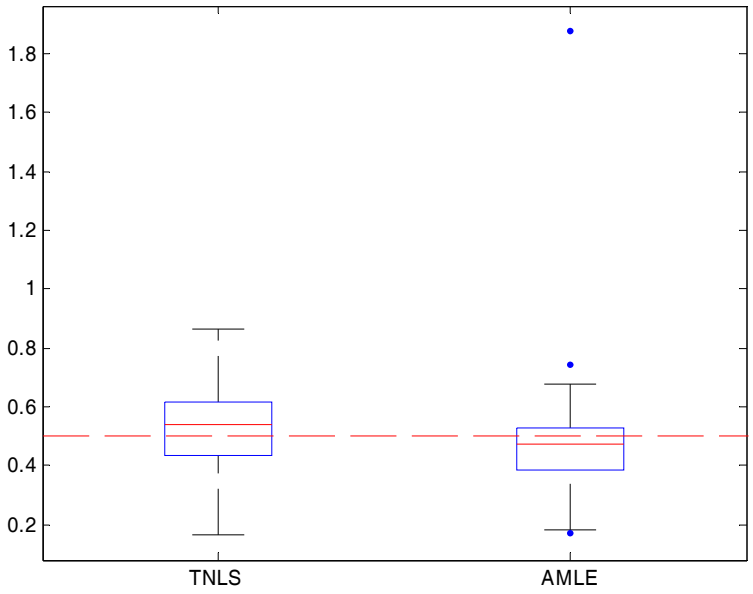

Figure 11 Box plots for b using TNLS and AMLE (unmeasured concentration)

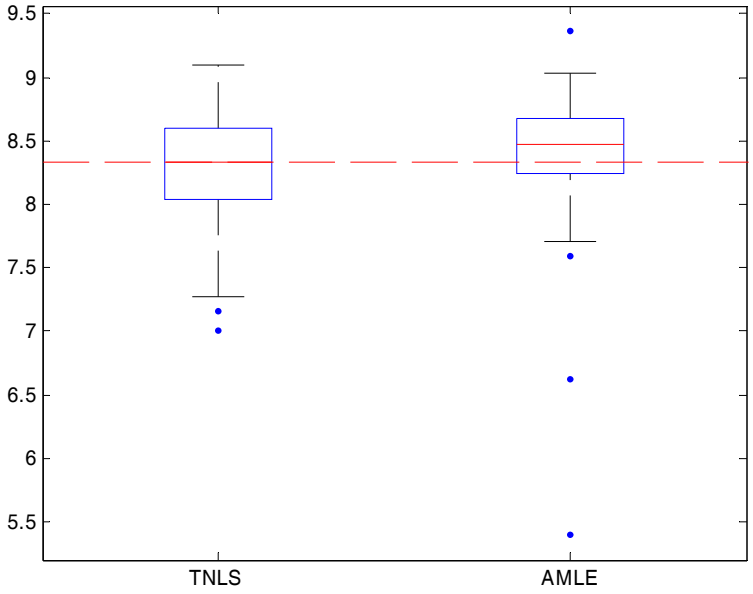

Figure 12 Box plots for $\frac{E}{R}$ using TNLS and AMLE (unmeasured concentration)

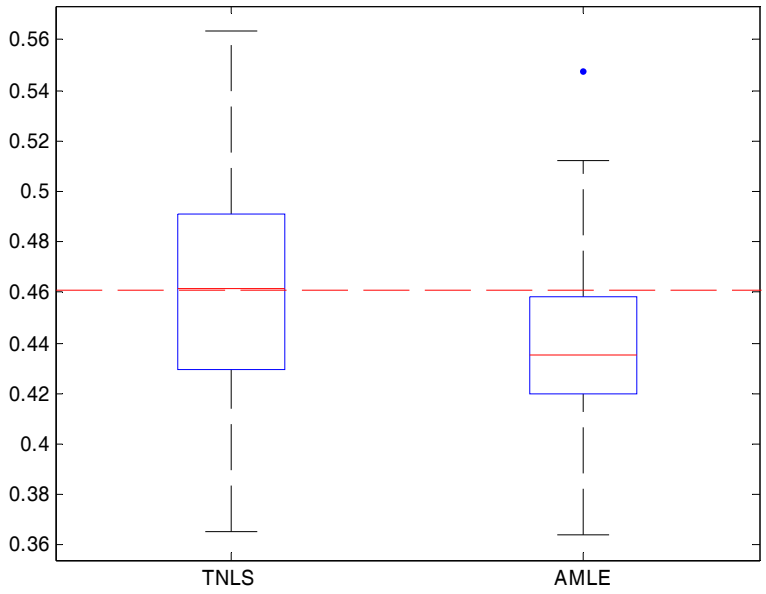

Figure 13 Box plots for k_{ref} using TNLS and AMLE (unmeasured concentration)

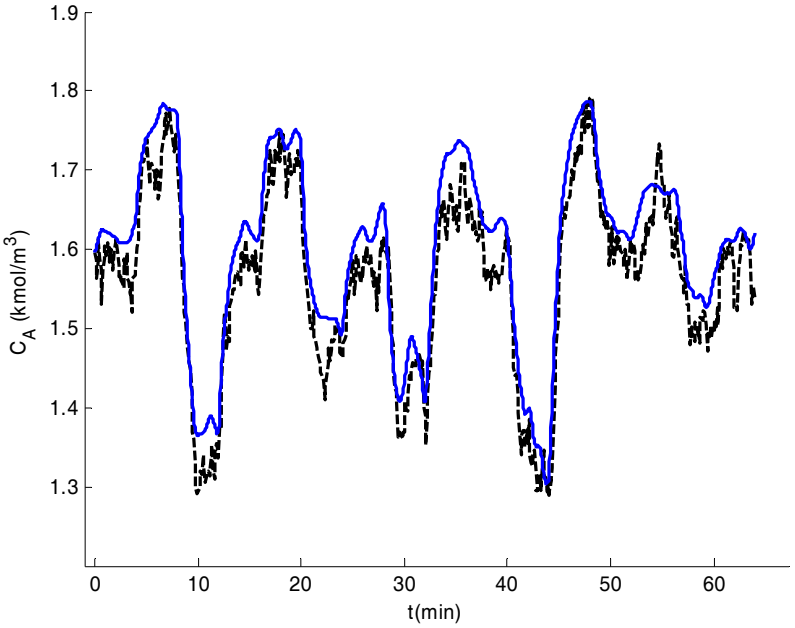

Figure 14 Observed, true, and predicted concentration response for AMLE (unmeasured concentration)

(---- response with true parameters and true stochastic noise, — AMLE response)

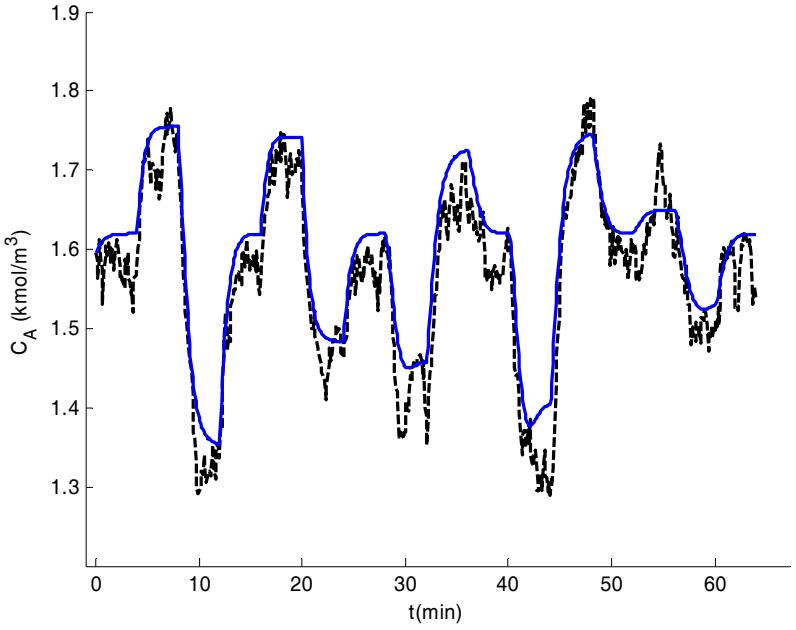

Figure 15 Observed, true, and predicted concentration response for TNLS (unmeasured concentration)

(---- response with true parameters and true stochastic noise, — TNLS response)

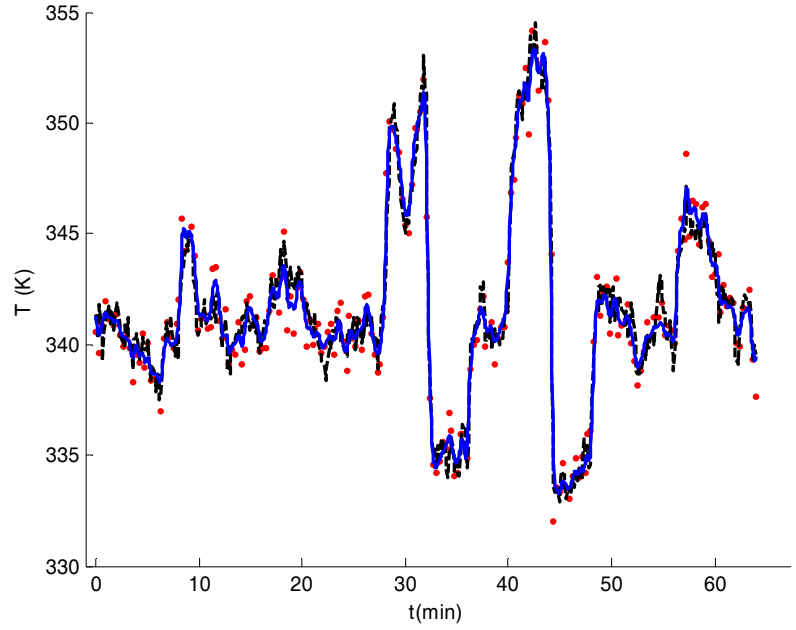

Figure 16 Observed, true, and predicted temperature response for AMLE (unmeasured concentration)

(• simulated data, ---- response with true parameters and true stochastic noise, — AMLE response)

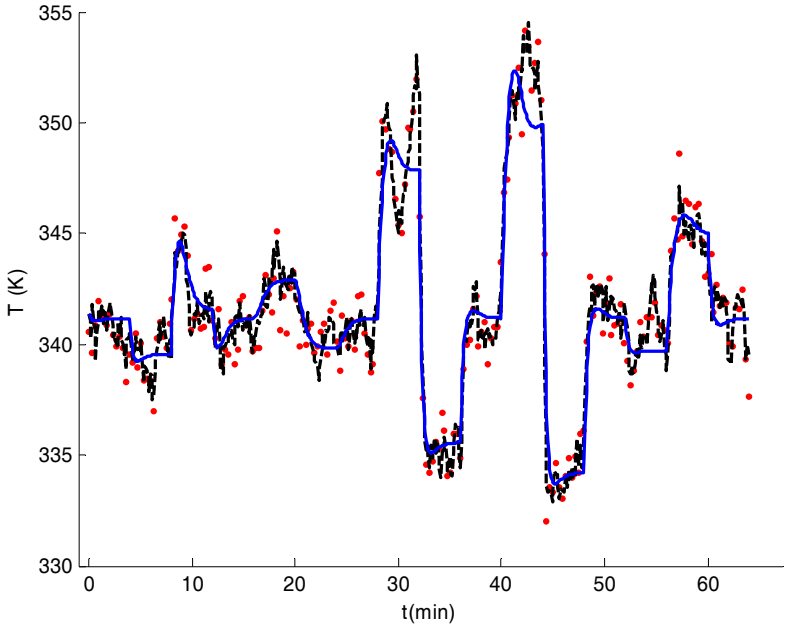

Figure 17 Observed, true, and predicted temperature response for TNLS (unmeasured concentration)

(• simulated data, ---- response with true parameters and true stochastic noise, — TNLS response)

Theoretical 95% confidence intervals for AMLE parameter estimates are presented in Table 2.

Table 2. 95% Confidence Intervals for AMLE parameter estimates (unmeasured concentration)

Parameter Estimates	Lower Bound	Upper Bound
a	1.0750	2.4559
b	0.3062	0.5765
E/R	8.0477	8.9000
k_{ref}	0.3932	0.4577

Note that removing the concentration measurements resulted in slightly wider confidence intervals. However, since the temperature measurements are more frequent and more precise than concentration measurements, removing the concentration measurements did not worsen the confidence intervals as

dramatically as removing the temperature measurements does (not shown). A scatter plot matrix of the parameter estimates for this case study is presented in Appendix B.

4.3 Nonlinear MIMO CSTR with non-stationary disturbance

In the third case study we consider the same CSTR model as in previous sections but with an additional non-stationary disturbance affecting the concentration differential equation. This non-stationary disturbance could be used to account for a meandering input that is not included in the fundamental part of the material balance equation. We assume that both concentration and temperature are measured and we estimate the vector of fundamental model parameters along with the non-stationary disturbance:

$$\begin{aligned}
&\frac{dC_A(t)}{dt} = \frac{F(t)}{V}\left(C_{A0}(t) - C_A(t)\right) - gC_A(t) + d_1(\mathrm{t}) + \eta_1(t) \\
&\frac{dT(t)}{dt} = \frac{F(t)}{V}\left(T_0(t) - T(t)\right) + \beta_1\left(T(t) - T_{cin}(t)\right) - \beta_2 gC_A(t) + \eta_2(t) \\
&\frac{dd_1(t)}{dt} = \eta_3(t)
\end{aligned} \tag{19}$$

$$\begin{aligned}
&C_A(0) = 1.569\,(\mathrm{kmol\,m^{-3}}) \\
&T(0) = 341.37\,(\mathrm{K}) \\
&d_1(0) = 0(\mathrm{kmol\,m^{-3}\,min^{-1}})
\end{aligned}$$

$$\begin{aligned}
&y_1(t_i) = C_A(t_i) + \varepsilon_1(t_i) \\
&y_2(t_j) = C_A(t_j) + \varepsilon_2(t_j)
\end{aligned}$$

$$g = k_{ref}\exp\left(-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right),\ \beta_1 = -\frac{aF_c^{b+1}(t)}{V\rho C_p\left(F_c(t) + \frac{aF_c^b(t)}{2\rho_c C_{pc}}\right)}\ \beta_2 = \frac{\left(-\Delta H_{rxn}\right)}{\rho C_p}$$

$E\{\eta_3(t)\eta_3(t-\tau)\} = Q_{p3}\delta(\tau)$ where $Q_{p3} = 6\times10^{-2}$ $(\mathrm{kmol/m^3\,/\,min})^2$ and everything else remains the same as in Section 4.1.

From equation (8), the AMLE objective function is:

$$\frac{1}{\sigma_{m1}^2}\sum_{j=1}^{64}\left(y_1(t_{m1j})-C_{A\sim}(t_{m1j})\right)^2+$$
$$\frac{1}{\sigma_{m2}^2}\sum_{j=1}^{213}\left(y_2(t_{m2j})-T_{\sim}(t_{m2j})\right)^2+$$
$$\frac{1}{Q_{p1}}\int_{t=0}^{64}\left(\frac{dC_{A\sim}(t)}{dt}-\frac{F(t)}{V}\left(C_{A0}(t)-C_{A\sim}(t)\right)+gC_{A\sim}(t)-d_{1\sim}(t)\right)^2dt+ \qquad (20)$$
$$\frac{1}{Q_{p2}}\int_{t=0}^{64}\left(\frac{dT_{\sim}(t)}{dt}-\frac{F(t)}{V}\left(T_0(t)-T_{\sim}(t)\right)-\beta_1\left(T_{\sim}(t)-T_{cin}(t)\right)+\beta_2 gC_{A\sim}(t)\right)^2dt+$$
$$\frac{1}{Q_{p3}}\int_{t=0}^{64}\left(\dot{d}_{1\sim}(t)\right)^2dt$$

The input scheme and the knot placements for the temperature and concentration trajectories are the same as the previous examples. For the non-stationary disturbance trajectory, $d_{1\sim}$, knots were placed at 0.3 minute intervals.

Figures 18 to 21 show the Monte Carlo box plots for the parameter estimates. Estimated concentration and temperature trajectories are presented (for one data set) in Figures 22 to 25.

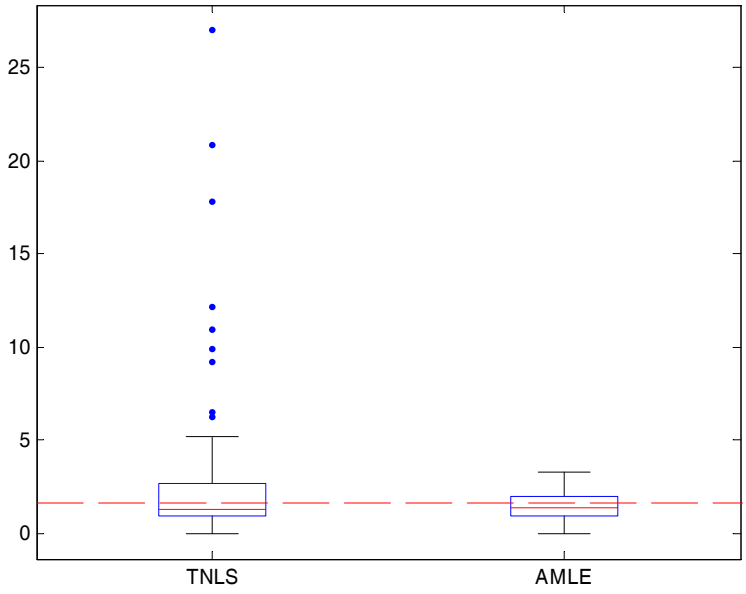

Figure 18 Box plots for a using TNLS and AMLE (with non-stationary disturbance)

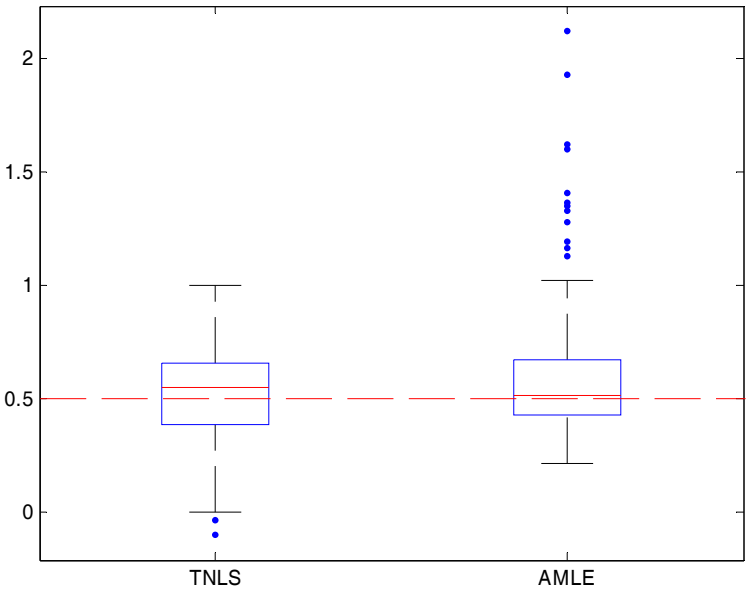

Figure 19 Box plots for b using TNLS and AMLE (with non-stationary disturbance)

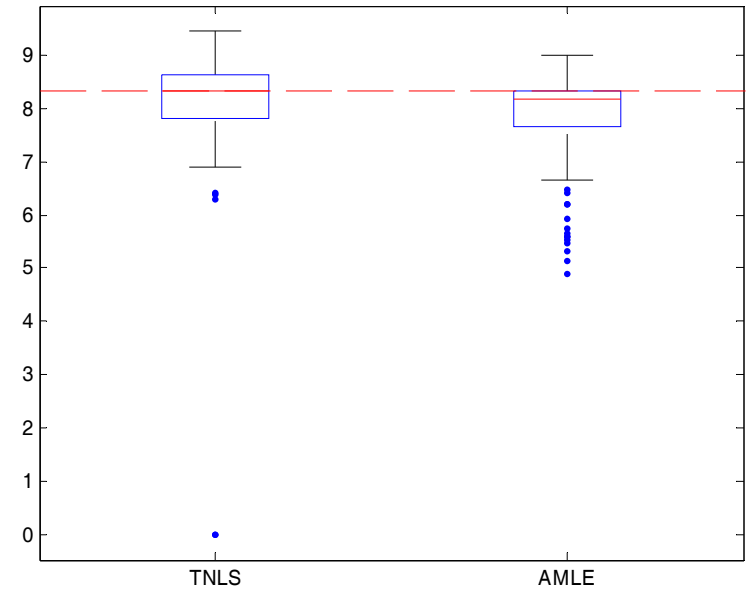

Figure 20 Box plots for $\frac{E}{R}$ using TNLS and AMLE (with non-stationary disturbance)

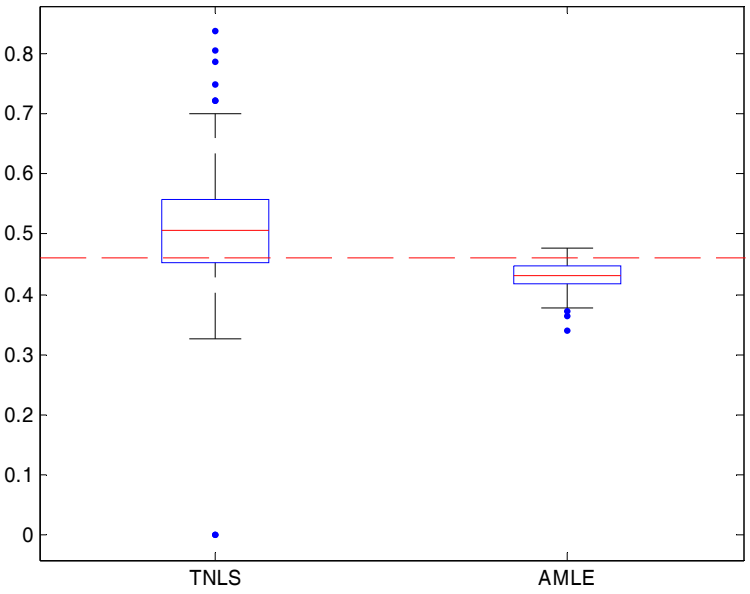

Figure 21 Box plots for k_{ref} using TNLS and AMLE (with non-stationary disturbance)

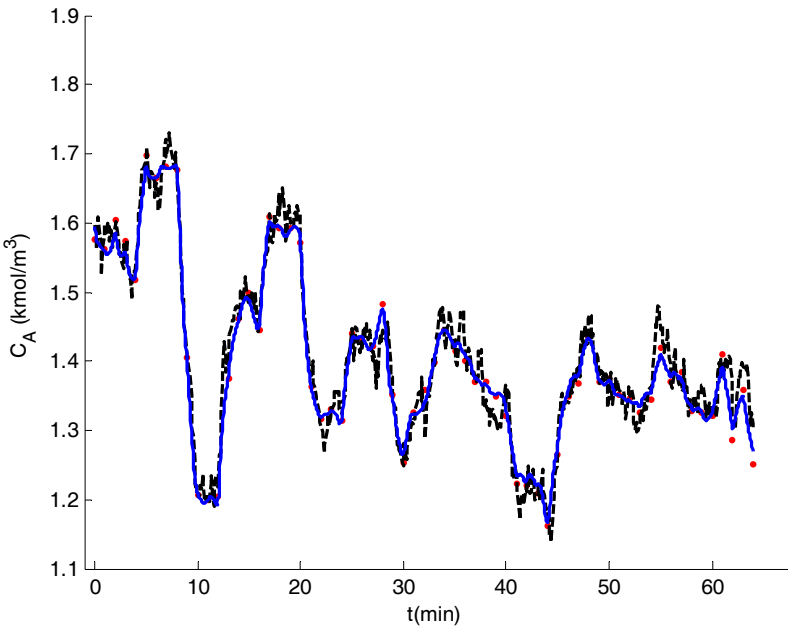

Figure 22 Observed, true, and predicted concentration response for AMLE, non-stationary example

(• simulated data, ----- response with true parameters and true stochastic noise, — AMLE response)

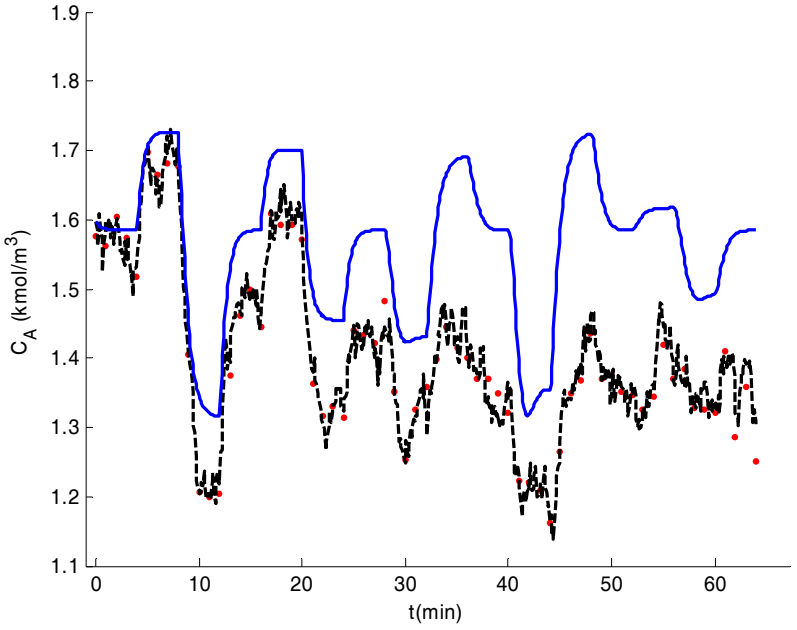

Figure 23 Measured, true, and predicted concentration responses using TNLS, non-stationary example

(• simulated data, ----- response with true parameters and true stochastic noise, — TNLS response)

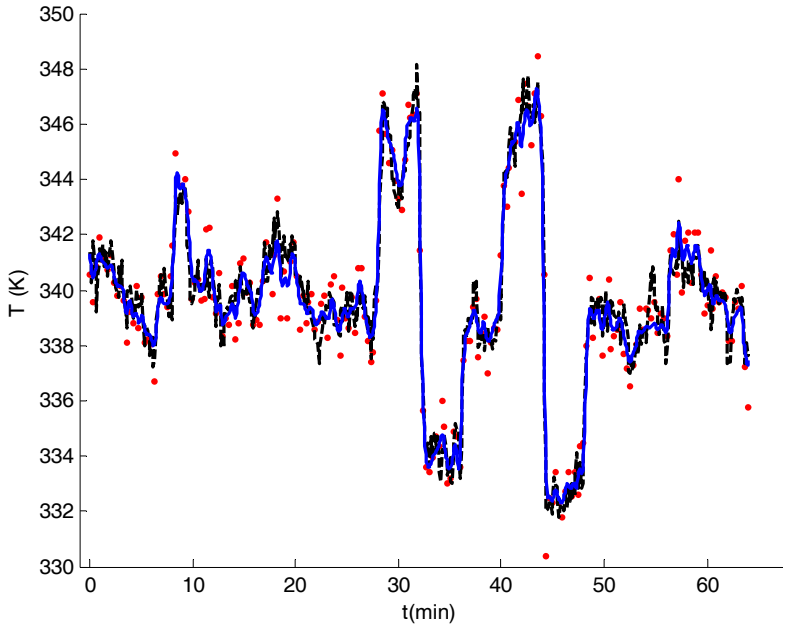

Figure 24 Observed, true, and predicted temperature response for AMLE, non-stationary example

(• simulated data, ----- response with true parameters and true stochastic noise, — AMLE response)

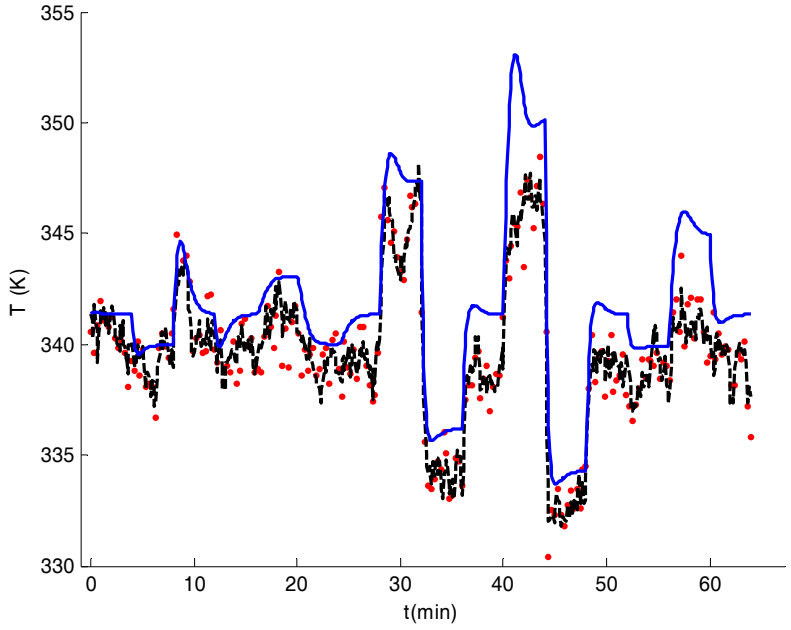

Figure 25 Observed, true, and predicted temperature response for TNLS, non-stationary example

(• simulated data, ----- response with true parameters and true stochastic noise, — TNLS response)

As expected, the advantage of AMLE over TNLS is more pronounced when there are non-stationary disturbance inputs, which are not accounted for using TNLS. The AMLE parameter estimates are less biased and more precise than TNLS parameter estimates. AMLE does a significantly better job at estimating (smoothing) the concentration and temperature trajectories and captures the sharp and meandering features in the response that appear due to the stationary and non-stationary disturbance terms in the model. The estimated disturbance in one of the data sets is illustrated in Figure 26.

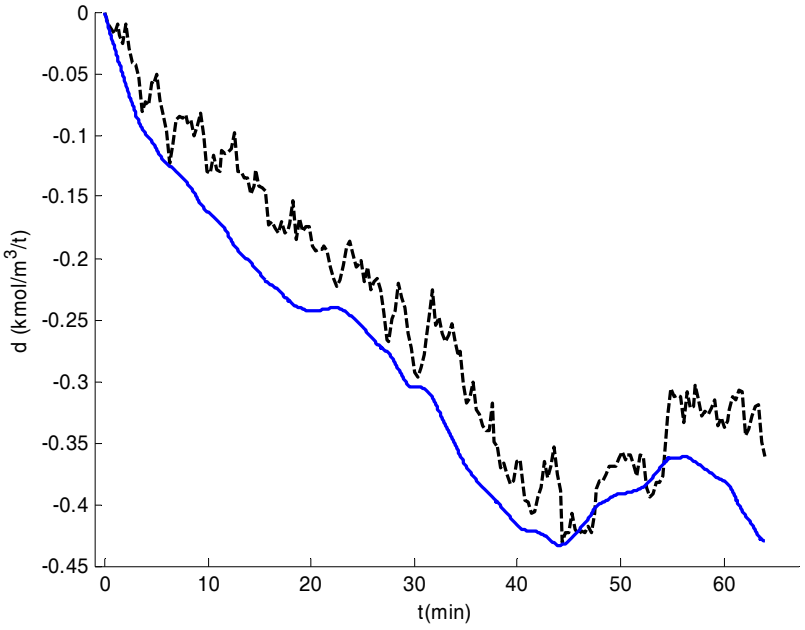

Figure 26 True, and estimated non-stationary disturbance using AMLE,

(----- non-stationary stochastic disturbance, — AMLE estimate)

Theoretical confidence intervals for the AMLE parameter estimates for the second example are presented in Table 3. As expected, these confidence intervals are wider than when there was no meandering disturbance (Table 1).

Table 3. 95% Confidence Intervals for AMLE parameter estimates (with non-stationary disturbance)

Parameter Estimates	Lower Bound	Upper Bound
a	0.6019	2.0615
b	0.3337	0.7259
E/R	7.8400	8.8890
k_{ref}	0.3924	0.4549

A scatter plot matrix for the parameter estimates is presented in Appendix B.

5 Summary and Conclusions

We show that AMLE can be readily used for parameter and state estimation in nonlinear dynamic models in which stochastic process disturbances and measurement noise are present and some states are not observed. We also show that the AMLE can be modified to accommodate unmeasured states by simply removing the corresponding sum-of-squared-error terms from the objective function. AMLE can handle parameter estimation in nonlinear dynamic systems in which non-stationary disturbances are present by treating these disturbances as unmeasured states. We develop theoretical confidence interval expressions for AMLE parameter estimates. The inference method is based on the asymptotic properties of maximum-likelihood estimators.

We test our results using a MIMO nonlinear CSTR model. In the first scenario, both states are measured. In the second, the temperature is measured, but the concentration is unmeasured. We estimate the concentration and temperature trajectories, along with four fundamental-model parameters, using both AMLE and TNLS. The parameter estimation results from AMLE are, on average, more precise and less biased than the TNLS results because AMLE is able to properly account for the process disturbances and measurement noise.

We also consider an additive non-stationary input disturbance (Brownian motion) in the material balance equation. Again AMLE and TNLS methods are compared and, in the case of AMLE, the non-stationary disturbance is estimated. The AMLE parameter estimates are more precise and less biased than TNLS parameter estimates. AMLE obtains significantly better estimates of the state trajectories. Confidence intervals for AMLE parameter estimates are obtained and compared with empirical box plots generated by Monte Carlo simulations. Confidence intervals and box plots are consistent.

Application of AMLE relies on knowledge of measurement noise variances and stochastic process intensities. These constants, however, are usually unknown in real-world applications. Before AMLE can enjoy widespread use, a means for estimating the noise and process disturbance constants, along with the model parameters, needs to be developed. This is the subject of our ongoing research.

Acknowledgement

The authors thank MITACS, Cybernetica, DuPont Engineering Research and Development and SAS for financial support, and Drs. L. T. Biegler from Carnegie Mellon University and J. O. Ramsay from McGill University for technical advice.

Appendix A

Derivation of the Hessian for the AMLE objective function

In this appendix, we show how the exact or approximate Hessian matrix for the AMLE objective function can be derived by exploiting the least-squares structure of the objective function.

Rewriting equation (5) in the matrix-vector form where we replace the integrals by Gaussian Quadrature sums we have:

$$\begin{aligned}
&\frac{1}{\sigma_{m1}^2}(\mathbf{y}_1-\mathbf{x}_{\sim 1})^T(\mathbf{y}_1-\mathbf{x}_{\sim 1})+\\
&\frac{1}{\sigma_{m2}^2}(\mathbf{y}_2-\mathbf{x}_{\sim 2})^T(\mathbf{y}_2-\mathbf{x}_{\sim 2})+\\
&\frac{1}{Q_{p1}}\left(\dot{\mathbf{x}}_{\sim q1}-\mathbf{f}_1(\mathbf{x}_{\sim q1},\mathbf{x}_{\sim q2},\mathbf{u}_1,\mathbf{u}_2,\boldsymbol{\theta})\right)^T\mathbf{w}_1\left(\dot{\mathbf{x}}_{\sim q1}-\mathbf{f}_1(\mathbf{x}_{\sim q1},\mathbf{x}_{\sim q2},\mathbf{u}_1,\mathbf{u}_2,\boldsymbol{\theta})\right)+\\
&\frac{1}{Q_{p2}}\left(\dot{\mathbf{x}}_{\sim q2}-\mathbf{f}_2(\mathbf{x}_{\sim q1},\mathbf{x}_{\sim q2},\mathbf{u}_1,\mathbf{u}_2,\boldsymbol{\theta})\right)^T\mathbf{w}_2\left(\dot{\mathbf{x}}_{\sim q2}-\mathbf{f}_2(\mathbf{x}_{\sim q1},\mathbf{x}_{\sim q2},\mathbf{u}_1,\mathbf{u}_2,\boldsymbol{\theta})\right)
\end{aligned} \qquad \text{(A. 1)}$$

where $\mathbf{y}_i$ and $\mathbf{x}_i$, respectively, are vectors of outputs $y_i(t)$ and states $x_{\sim i}(t)$ evaluated at the observation points. $\mathbf{x}_{\sim qi}$, $\dot{\mathbf{x}}_{\sim qi}$, $\mathbf{f}_i$, and $\mathbf{u}_i$ are vectors containing $x_{\sim i}(t)$, $\dot{x}_{\sim i}(t)$, $f_i(t)$ and $u_i(t)$, respectively, evaluated at Gaussian Quadrature points. $\mathbf{w}_1$ and $\mathbf{w}_2$ are diagonal matrices containing the Gaussian Quadrature weights. Equation (A. 1) can further be simplified by introducing the following vectors and matrices:

$$\mathbf{x}_{\sim} = \begin{bmatrix} \mathbf{x}_{\sim 1} \\ \mathbf{x}_{\sim 2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varphi}_1 \boldsymbol{\beta}_1 \\ \boldsymbol{\varphi}_2 \boldsymbol{\beta}_2 \end{bmatrix}, \quad \mathbf{x}_{\sim q} = \begin{bmatrix} \mathbf{x}_{\sim q1} \\ \mathbf{x}_{\sim q2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varphi}_{q1} \boldsymbol{\beta}_1 \\ \boldsymbol{\varphi}_{q2} \boldsymbol{\beta}_2 \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \quad \mathbf{y}_m = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix},$$

$\boldsymbol{\varphi} = \begin{bmatrix} \boldsymbol{\varphi}_1 & \\ & \boldsymbol{\varphi}_2 \end{bmatrix}$ and $\boldsymbol{\varphi}_q = \begin{bmatrix} \boldsymbol{\varphi}_{q1} & \\ & \boldsymbol{\varphi}_{q2} \end{bmatrix}$. Note that $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_{q1}$ are concatenated vectors of $\varphi_1(t)$ evaluated at observation times and Gaussian Quadrature points, respectively. Analogous definitions apply to $\boldsymbol{\varphi}_2$ and $\boldsymbol{\varphi}_{q2}$.

We let $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ be $N_1 \times N_1$ and $N_2 \times N_2$ diagonal matrices with σ_{m1}^2 and σ_{m2}^2, respectively, on the diagonals. Then, the AMLE objective function in (A. 1) can be written in the following form:

$$L(\boldsymbol{\tau}, \mathbf{y}_m) = -\mathbf{g}(\boldsymbol{\tau}, \mathbf{y}_m)^T \mathbf{W} \mathbf{g}(\boldsymbol{\tau}, \mathbf{y}_m) \tag{A. 2}$$

where

$$\mathbf{g}(\boldsymbol{\tau}, \mathbf{y}_m) = \begin{bmatrix} \mathbf{y}_m - \boldsymbol{\varphi}\boldsymbol{\beta} \\ \\ \dot{\boldsymbol{\varphi}}_q \boldsymbol{\beta} - \mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\beta}) \end{bmatrix} \tag{A. 3}$$

and $\mathbf{W}$ is a diagonal matrix with:

$$diag(\mathbf{W}) = [diag(\boldsymbol{\Sigma}_1^{-1}), diag(\boldsymbol{\Sigma}_2^{-1}), diag(\frac{1}{Q_{p1}} \mathbf{w}_1), diag(\frac{1}{Q_{p2}} \mathbf{w}_2)].$$

Since, according to (A. 2), the objective function is in a least-squares form, the Hessian matrix can be written as:

$$\frac{\partial^2 L(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} = \frac{\partial \mathbf{g}(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}'}^T \left(\frac{\partial^2 L}{\partial \mathbf{g} \partial \mathbf{g}^T} \right) \frac{\partial \mathbf{g}(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}} + \sum_i \frac{\partial L}{\partial \mathbf{g}_i} \frac{\partial^2 \mathbf{g}_i}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} \tag{A. 4}$$

If the second-order-derivative terms in the summation on the right-hand side of the above expression can be neglected, then equation (A. 4) can be further simplified by the following approximation:

$$\frac{\partial^2 L(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} \cong -2 \frac{\partial \mathbf{g}(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}}^T \mathbf{W} \frac{\partial \mathbf{g}(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}} \tag{A. 5}$$

Note that $\frac{\partial \mathbf{g}}{\partial \boldsymbol{\tau}} = \left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}, \frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}}\right]$ and from(A. 3):

$$\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathbf{0} \\ -\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}} \end{bmatrix}, \quad \frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} -\boldsymbol{\varphi} \\ \dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}} \end{bmatrix} \tag{A. 6}$$

Hence:

$$\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\tau}}\right)^T \mathbf{W} \left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\tau}}\right) = \begin{bmatrix} \left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right)^T \mathbf{w}_{12} \left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right) & -\left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right)^T \mathbf{w}_{12} \left(\dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right) \\ -\left(\dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right)^T \mathbf{w}_{12} \left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right) & \boldsymbol{\varphi}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varphi} + \left(\dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right)^T \mathbf{w}_{12} \left(\dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right) \end{bmatrix} \tag{A. 7}$$

where $\mathbf{w}_{12} = \begin{bmatrix} \mathbf{w}_1 & \\ & \mathbf{w}_2 \end{bmatrix}$.

Note that, to obtain the inverse of the matrix in (A. 7), the following property can be used[3].

If all inverses exist then:

$$\begin{aligned} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} &= \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21} & -\mathbf{B}_{12}\mathbf{B}_{22}^{-1} \\ -\mathbf{B}_{22}^{-1}\mathbf{B}_{21} & \mathbf{B}_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{C}_{11}^{-1} & -\mathbf{C}_{11}^{-1}\mathbf{C}_{12} \\ -\mathbf{C}_{21}\mathbf{C}_{11}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12} \end{bmatrix} \end{aligned} \tag{A. 8}$$

where $\mathbf{B}_{22} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, $\mathbf{B}_{12} = \mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, $\mathbf{B}_{21} = \mathbf{A}_{21}\mathbf{A}_{11}^{-1}$, $\mathbf{C}_{11} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$, $\mathbf{C}_{12} = \mathbf{A}_{12}\mathbf{A}_{22}^{-1}$, $\mathbf{C}_{22} = \mathbf{A}_{22}^{-1}\mathbf{A}_{21}$.

Appendix B

Scatter plot matrices

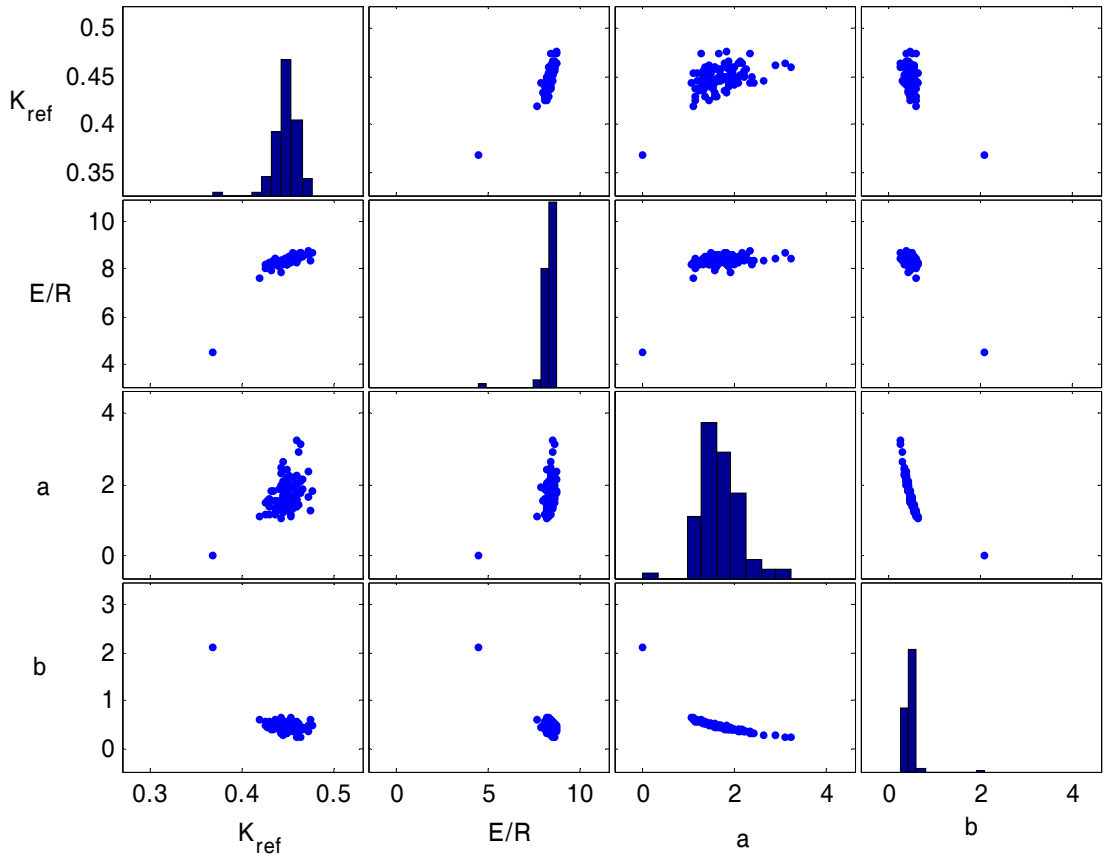

Figure B.1 Scatter plot matrix for 100 sets of parameter estimates obtained using AMLE with both states measured

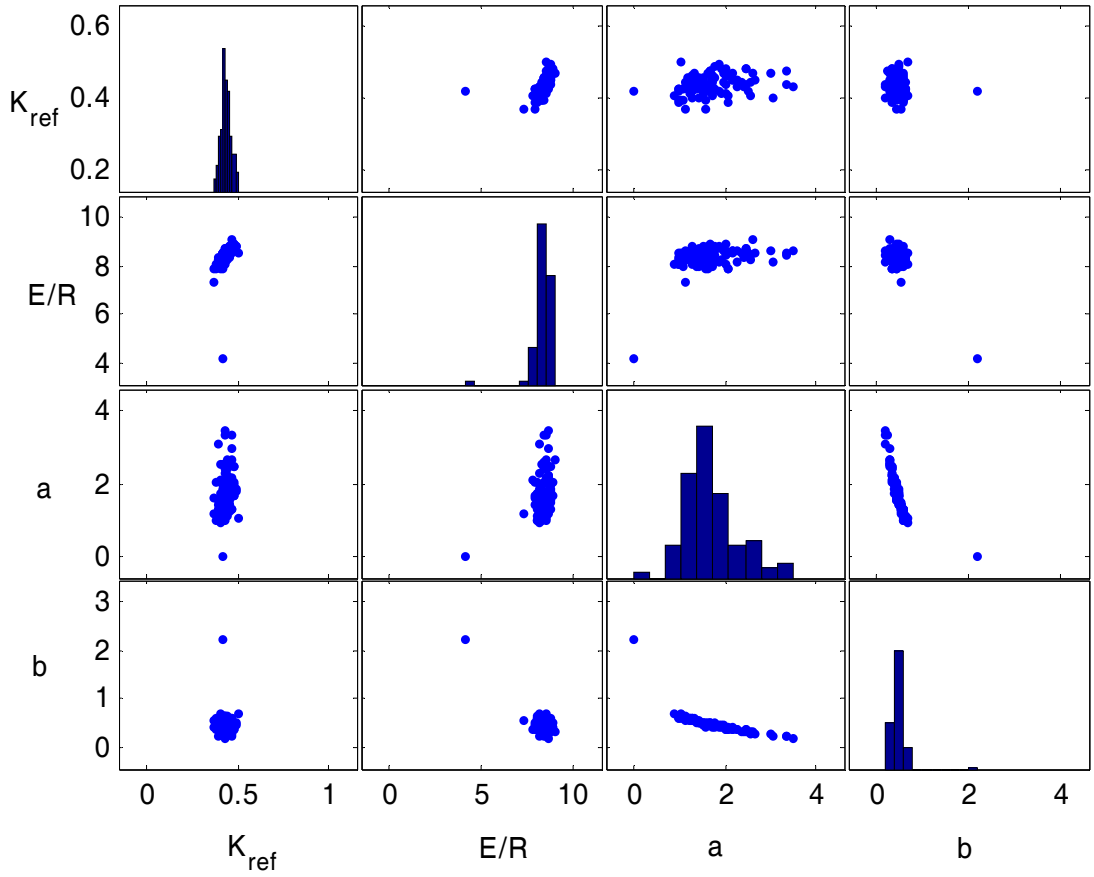

Figure B.2 Scatter plot matrix for 100 sets of parameter estimates obtained using MLE without concentration measurements

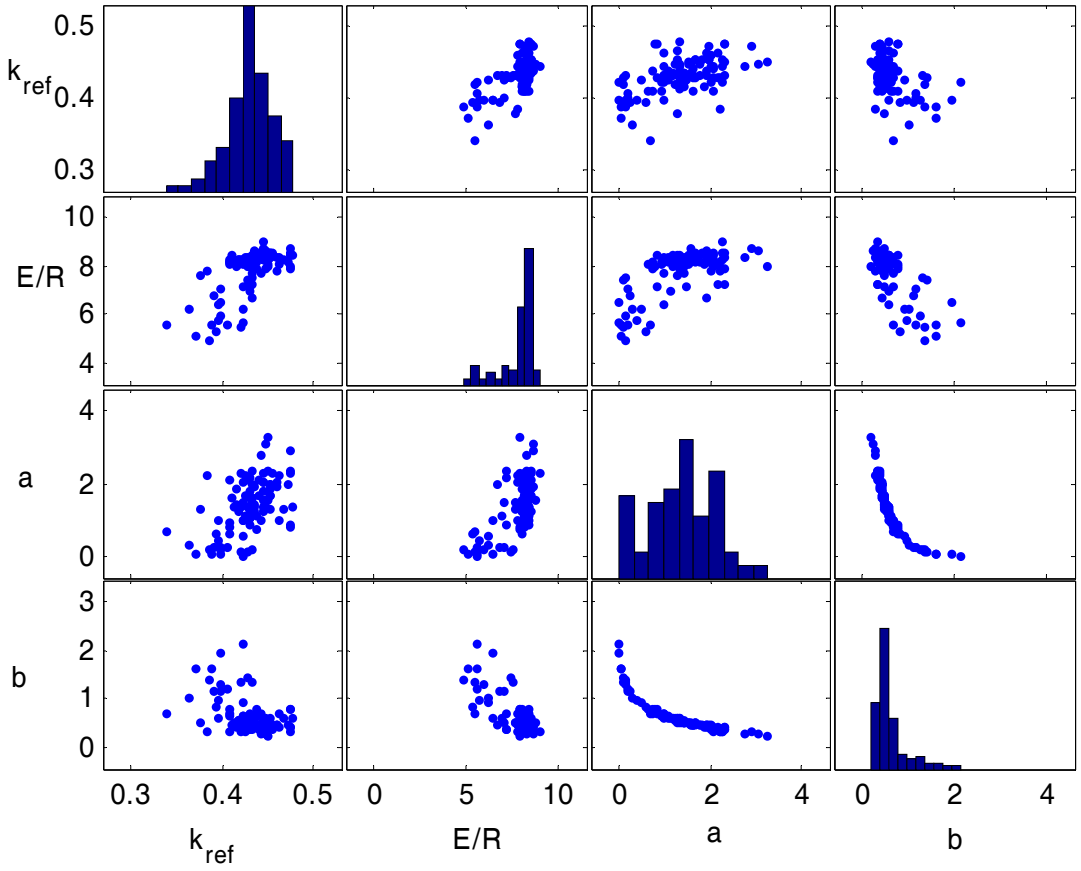

Figure B.3 Scatter plot matrix for 100 sets of parameter estimates obtained from AMLE with both temperature and concentration measured and a non-stationary disturbance in the material balance

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	kmol m^{-3}
C_{A0}	Feed concentration of reactant A	kmol m^{-3}

C_{As}	Concentration of reactant A at steady state	kmol m^{-3}
C_p	Reactant heat capacity	cal $g^{-1}K^{-1}$
C_{pc}	Coolant heat capacity	cal $g^{-1}K^{-1}$
d_1	Non-stationary disturbance term	kmol m^{-3} min^{-1}
$d_{1\sim}$	B-spline approximation of the non-stationary disturbance	kmol m^{-3} min^{-1}
$E\{.\}$	Expected value	
E/R	Activation energy over the ideal gas constant	K
F	Reactant volumetric flow rate	m^3 min^{-1}
F_c	Coolant volumetric flow rate	m^3 min^{-1}
f_i, $\mathbf{f}_i$	Nonlinear function on the right-hand side of the differential equation for state i	
f_d	Nonlinear function on the right-hand side of the non-stationary disturbance differential equation	
$\mathbf{g}$	Vector of combined sum-of-squared-errors and model-based penalties	
$\mathbf{I}$	Fisher information matrix	
k_{ref}	Kinetic rate constant at temperature T_{ref}	min^{-1}
L	Log-likelihood function	
N_i	Number of observations of state i	
$p(.)$	Probability density function	
Q_{pi}	Process noise intensity of stochastic differential equation for state i	
Q_{pd}	Process noise intensity of stochastic differential equation for non-stationary disturbance	
t_{mij}	j-th measurement time for the i-th state	min

T	Temperature of reactor contents	K
T_0	Reactant feed temperature	K
T_{cin}	Inlet temperature of coolant	K
T_s	Temperature of reactant at steady state value	K
T_{ref}	Reference temperature	K
u_i, $\boldsymbol{u}_i$	Input to the differential equation for state i	
V	Volume of the reactor	m^3
$\mathbf{w}_i$	Matrix of Gaussian quadrature weights for calculating the model-based penalty integrals of differential equation for state i	
$\mathbf{W}$	Overall weighting matrix to define the log-likelihood function in a least-squares form	
x, $\mathbf{X}$	State variables	
$x_{i\sim}$	B-spline approximation of the i-th state	
y, $\mathbf{Y}$	Noisy output measurements	
$\mathbf{y}_m$	Stacked vector of measured outputs	
$z_{\alpha/2}$	Normal random deviate corresponding to an upper tail area of $\alpha/2$	
α	Significance level for confidence intervals	
β_{ij}	j-th B-spline coefficient of the i-th state	
$\boldsymbol{\beta}_i$	Vector of B-spline coefficients of the i-th state	
$\boldsymbol{\beta}_d$	Vector of B-spline coefficients of the disturbance term	
$\delta(.)$	Dirac delta function	

ΔH_{rxn}	Enthalpy of reaction	cal g^{-1} K^{-1}
ε_i	Normally distributed measurement noise for state *i*	
η_i	White Gaussian process disturbance for differential equation of the *i*-th state	
η_d	White Gaussian process disturbance for the non-stationary disturbance differential equation	
$\boldsymbol{\theta}$	Vector of model parameters	
$\boldsymbol{\theta}_d$	Vector of disturbance model parameters	
ρ	Density of reactor contents	g m^{-3}
ρ_c	Coolant density	g m^{-3}
σ^2_{mi}	Measurement noise variance for the *i*-th state	
Σ_i	Measurement noise covariance matrix of the *i*-th state vector	
Σ	Measurement noise covariance matrix	
$\boldsymbol{\tau}$	Combined vector of model parameters and spline coefficients	
ϕ_{ij}	*j*-th B-spline basis function of the *i*-th state	
φ_i	Vector of B-spline basis functions for the *i*-th state	
φ_{d1}	Vector of B-spline basis functions for disturbance d_1	
$\boldsymbol{\varphi}_i$	Matrix of all φ_i s evaluated at the *i*-th state observation times	
$\boldsymbol{\varphi}$	Matrix containing all $\boldsymbol{\varphi}_i$ s	
$\boldsymbol{\varphi}_{qi}$	Matrix of all φ_i s evaluated at the quadrature points of the differential equation corresponding to the *i*-th state	

$\boldsymbol{\varphi}_q$ Matrix containing all $\boldsymbol{\varphi}_{qi}$ s

AMLE Approximate Maximum Likelihood Estimation

CSTR Continuous Stirred Tank Reactor

diag Diagonal elements of a matrix

iPDA iteratively-refined Principal Differential Analysis

MIMO Multi-Input Multi-Output

ODE Ordinary Differential Equation

PEN Model-based penalty

SSE Sum of Squared Errors

TNLS Traditional Nonlinear Least-Squares

Literature Cited

(1) A.A. Poyton; M.S. Varziri; K.B. McAuley; P. J. McLellan; J. O. Ramsay, Parameter estimation in continuous-time dynamic models using principal differential analysis, *Comp. Chem. Eng*. **2006**, 30, 698-708.

(2) Varziri M. S.; McAuley, K. B.; McLellan P. J., Selecting optimal weighting factors in iPDA for parameter estimation in continuous-time dynamic models, submitted to *Comp. Chem. Eng*. , **2007**.

(3) Seber G. A. F.; Wild C. J., *Nonlinear Regression*; John Wiley and Sons, Inc. 1989.

(4) Bates D. M; Watts D. G., *Nonlinear Regression Analysis and its Applications*; John Wiley and Sons, Inc., 1988.

(5) Bard Y., *Nonlinear Parameter Estimation*; Academic Press, Inc., New York, 1974.

(6) B. A. Ogunnaike; W. H. Ray, *Process Dynamics, Modeling and Control*; Oxford University Press: New York, 1994.

(7) Leis J. R.; Kramer M. A., ALGORITHM 658: ODESSA – an ordinary differential equation solver with explicit simultaneous sensitivity analysis. *ACM Trans. Math. Software,* **1988**, 14 , 61-67.

(8) Biegler L. T.; Grossman I. E., Retrospective on optimization. *Comp. Chem. Eng.* **2004**, 28, 1169-1191.

(9) Tang Y. P., On the estimation of rate constants for complex kinetic models, *Ind. Eng. Chem. Fund.* **1971**, 10, 321-322.

(10) Swartz J.; Bremermann H., Discussion of parameter estimation in biological modelling: algorithms for estimation and evaluation of estimates, *J. Math. Bio.* **1975**,1**,** 241-275.

(11) Varah J. M., A spline least squares method for numerical parameter estimation in differential equations. *SIAM J. Scientific Computing* **1982**, 3**,** 28-46.

(12) Vajda S.; Valko P., A direct-indirect procedure for estimation of kinetic parameters, *Comp. Chem. Eng.* **1986**, 10, 1, 49-58.

(13) Hosten L. H., A comparative study of short cut procedures for parameter estimation in differential equations, *Comp. Chem. Eng.* **1979**, 3, 117-126.

(14) Biegler L. T., Short note solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Comp. Chem. Eng.* **1984**, 8, 243-248.

(15) Wächter A. ; Biegler L. T., On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Math. Prog*. **2006**, 106(1), 25-57.

(16) Ramsay J. O.; Hooker G.; Campbell D.; Cao J., Parameter estimation for differential equations: a generalized smoothing approach, accepted to appear in the *J. Royal Stat. Soc.* **2007**.

(17) Gagnon L.; MacGregor J. F., State estimation for continuous emulsion polymerization, *Can. J. Chem. Eng*. **1991**, 69, 648-656.

(18) de Boor C., *A practical Guide to Splines*; Springer: New York, 2001.

(19) Ramsay J. O.; Silverman B. W., *Functional Data Analysis*, 2nd edition; Springer 2005.

(20) Maybeck P. S., *Stochastic Models, Estimation, and Control* Vol1; Academic Press, 1979.

(21) Maybeck P. S., *Stochastic Models, Estimation, and Control* Vol2; Academic Press, 1982.

(22) Mortensen R. E., Maximum-Likelihood recursive nonlinear filtering, *J. Opt. Theo. and Appl.* **1968**, 2, No.6, 386.

(23) Jazwinski A. H., *Stochastic Processes and Filtering*; Academic Press, 1970.

(24) Evensen G.; Dee D.; Schroter J., Parameter Estimation in Dynamical Models, In *Ocean Modeling and Parametrizations*; Chassignet E.P.; Verron J.; Kluwer Academic: The Nederlands, 1998, 373-398

(25) C.R. Rao, *Linear Statistical Inference and Applications*, 2nd edition; John Wiley & Sons, 1973.

(26) Kay S. M., *Fundamentals of Statistical Signal Processing*; Prentice Hall Signal Processing Series, 1993.

(27) Marlin T. E., *Process Control: Designing Processes and Control Systems for Dynamic Performance*, 2nd edition; McGraw-Hill, 2000.

(28) A.A. Poyton, *Application of principal differential analysis to parameter estimation in fundamental dynamic models*, M.Sc. Thesis, Queen's University, Kingston, Canada, 2005.