PARAMETER ESTIMATION IN NONLINEAR CONTINUOUS-TIME DYNAMIC MODELS WITH MODELLING ERRORS AND PROCESS DISTURBANCES

by

M. Saeed Varziri

A thesis submitted to the Department of Chemical Engineering

in conformity with the requirements for the degree of

Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

(June, 2008)

To my parents

Nadereh & Hassan

Abstract

Model-based control and process optimization technologies are becoming more commonly used by chemical engineers. These algorithms rely on fundamental or empirical models that are frequently described by systems of differential equations with unknown parameters. It is, therefore, very important for modellers of chemical engineering processes to have access to reliable and efficient tools for parameter estimation in dynamic models. The purpose of this thesis is to develop an efficient and easy-to-use parameter estimation algorithm that can address difficulties that frequently arise when estimating parameters in nonlinear continuous-time dynamic models of industrial processes.

The proposed algorithm has desirable numerical stability properties that stem from using piece-wise polynomial discretization schemes to transform the model differential equations into a set of algebraic equations. Consequently, parameters can be estimated by solving a nonlinear programming problem without requiring repeated numerical integration of the differential equations.

Possible modelling discrepancies and process disturbances are accounted for in the proposed algorithm, and estimates of the process disturbance intensities can be obtained along with estimates of model parameters and states. Theoretical approximate confidence interval expressions for the parameters are developed.

Through a practical two-phase nylon reactor example, as well as several simulation studies using stirred tank reactors, it is shown that the proposed parameter estimation algorithm can address difficulties such as: different types of measured responses with different levels of measurement noise, measurements taken at irregularly-spaced sampling times, unknown initial conditions for some state variables, unmeasured state variables, and unknown disturbances that enter the process and influence its future behaviour.

Co-Authorship

The material presented in Chapters 3, 4, 5 and 6 has been submitted either wholly or in part to refereed journals. References to those journals are indicated in the corresponding chapter introductions. Dr. Jim McLellan and Dr. Kim McAuley edited and suggested revisions for the manuscripts and this thesis, assisted with the organization of the material, and provided technical advice.

Acknowledgements

First and foremost, I would like to thank my Mother and Father (Nadereh and Hassan) for their unending love and unconditional support throughout my life. They have always been there for me when I needed them.

I cannot thank my two great supervisors, Dr. Jim McLellan and Dr. Kim McAuley, enough for the invaluable advice they have given me, be it technical or general, and for their continuous support and encouragement during this journey. This work would not have been possible without their diligent guidance and outstanding supervision. I am truly thankful for having the opportunity to work with Dr. Kim McAuley and Dr. Jim McLellan. They have shown me not only how to be a better engineer, but also above all, how to be a better person.

I would also like to express my gratitude to Dr. Jim Ramsay from McGill University and Dr. David Campbell from Simon Fraser University for technical advice and fruitful discussions.

This journey would not have been as enjoyable, without my friends and fellow graduate students especially Mr. Duncan Thompson, Mr. Shauhoa (Roy) Wu, and Mr. Dean Latham.

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

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Model-based control and process optimization technologies are becoming more commonly used by chemical engineers. These algorithms generally rely on fundamental or empirical models that are frequently described by systems of ordinary differential equations (ODEs) with unknown parameters. The popularity of model-based control and process optimization algorithms has consequently led to an increased interest in developing fundamental models with precise parameter estimates (Biegler and Grossman, 2004; El-Farra and Christofides, 2003; Nagy and Braatz, 2003). Therefore, it is very important for modellers of chemical engineering processes to have access to reliable and efficient tools for parameter estimation in dynamic models. The purpose of this thesis is to develop an efficient and easy-to-use parameter estimation algorithm that can address difficulties that frequently arise when estimating parameters in nonlinear continuous-time dynamic models of industrial processes. This thesis builds on the research work by Poyton et al. (2006), which is described in the next section.

In this chapter the problem of parameter estimation in continuous-time dynamic models is formulated, and the traditional method for approaching this problem is discussed. Disadvantages of the traditional approach are then pointed out to motivate the development of the proposed parameter estimation approach. Finally the overall outline of the thesis and the objectives of the remaining chapters are presented.

1.1 Problem definition

Improving the performance of control systems that are used in chemical engineering processes can increase the quality of the products and safety of the processes while reducing the cost. One way to improve the performance of model-based control systems is to provide them with more accurate process models. When parameters of a dynamic model are efficiently estimated using input-output data, the quality of the model is improved. It is, therefore, very important for modellers of chemical engineering processes to have access to efficient parameter estimation algorithms since this will eventually lead to better product quality, lower cost, and safer processes.

To introduce the parameter estimation problem we consider the following simple first-order single-input single-output (SISO) nonlinear ODE model:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), \boldsymbol{\theta}) \\ x(t_0) &= x_0 \\ \\ y(t_{mj}) &= x(t_{mj}) + \varepsilon(t_{mj}) \end{aligned} \tag{1.1}$$

x is the state variable, u is the input variable and y is the output variable. f is a nonlinear function of the vector of model parameters $\boldsymbol{\theta}$, state variable, and input variable. It is assumed that f is Lipschitz in its first argument, and continuous in its second argument, so that eq. (1.1) has a unique solution (Maybeck, 1979). ε is a zero-mean uncorrelated random variable with variance σ_m^2.

The objective is to find a set of parameter estimates, $\hat{\boldsymbol{\theta}}$, given n measurements of the response outputs at times t_{m1} to t_{mn}, $\mathbf{y} = [y(t_{m1}), \cdots, y(t_{mn})]^T$, so that the output predicted by the model using the estimated parameters, $\hat{\boldsymbol{\theta}}$, is as close as possible to the unknown true process response.

The most commonly used optimality criterion to quantify the closeness of observed and predicted output responses is the sum of squared prediction errors; however, other optimality criteria such as sum of absolute errors can also be used (Seber and Wild, 1989).

A common method for obtaining $\hat{\boldsymbol{\theta}}$ is the Nonlinear Least-Squares (NLS) approach in which the sum of squared deviations of the model predictions from the measured output is minimized:

$$\begin{aligned} &\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{n} \left(y(t_{mi}) - x(t_{mi}) \right)^2 \right\} \\ &\text{subject to} \\ &\dot{x}(t) = f(x(t), u(t), \boldsymbol{\theta}), \qquad x(t_0) = x_0 \end{aligned} \tag{1.2}$$

This method is the same as the method of Maximum Likelihood (ML) if the distribution of the measurement noise is assumed to be independent, identically distributed (IID) Normal (Seber and Wild, 1989). In NLS a nonlinear minimization technique is employed along with an ODE solver to find the optimal set of parameter estimates (Bard, 1974; Bates and Watts, 1988; Seber and Wild, 1989; Ogunnaike and Ray, 1994; Englezos and Kalogerakis, 2001). This method is sometimes referred to as the Single Shooting method (Tanartkit and Biegler, 1995) or the Initial Value Problem approach (Bock, 1983) because it requires the solution of an initial value problem and a single shooting method is used for the forward integration of ODEs. In this thesis, the term (traditional) NLS is used to refer to this algorithm. Note that obtaining the sensitivity information required for the minimization step in NLS may not be easy, and can be computationally very expensive (Leis and Kramer, 1988).

An important problem that can arise in the NLS approach comes from the numerical integration of the ODEs. If the ODEs have exponentially increasing and decreasing modes, the successful integration of the ODE models can be difficult due to error-propagation and stability issues (Bard,

1974; Ascher et al., 1988; Tanartkit and Biegler, 1995; Li et al., 2005). To illustrate this problem we consider the following ODE model (Bock, 1983):

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) & x_1(0) &= 0 \\ \dot{x}_2(t) &= \mu^2 x_1(t) - (\mu^2 + p^2)\sin pt & x_2(0) &= \pi \end{aligned} \qquad t \in [0,1] \tag{1.3}$$

where $\mu = 20$ is a constant and p is the model parameter with the true value of $p = \pi$.

The particular solution for this problem is:

$$\begin{aligned} x_1(t) &= \sin \pi t \\ x_2(t) &= \pi \cos \pi t \end{aligned}$$

We attempted to numerically solve the above set of ODEs using the "ode45" solver in Matlab™, which employs an explicit Runge-Kutta method (Dormand and Prince, 1980), with relative and absolute tolerances set to 10^{-6} and 10^{-12}, respectively. As shown in Figure 1.1, the numerical solution diverges from the true solution significantly.

To understand why this problem occurs, we consider the general solution of (1.3):

$$\begin{aligned} x_1(t) &= \sin \pi t + c_1 \exp(-\mu t) + c_2 \exp(\mu t) \\ x_2(t) &= \pi \cos \pi t - c_1 \exp(-\mu t) + c_2 \exp(\mu t) \end{aligned}$$

where c_1 and c_2 are constants that can be determined using the given initial conditions. Suppose that initial conditions have some errors. Although the true initial conditions for x_1 and x_2 are zero and π, respectively, we believe that $x_1(0) = e_1$ and $x_2(0) = \pi + \mu e_2$ where e_1 and μe_2 are errors. Using the general solution and the erroneous initial conditions, c_1 and c_2 can be obtained as: $c_1 = (e_1 - \mu e_2)/2$ and $c_2 = (e_1 + \mu e_2)/2$. Note that error propagation is exponential with respect to μ and t and, consequently, even small errors can lead to disastrous results (Bock, 1983). Even if there are no errors in the initial conditions, there will always be a (small) error after the

first time step of the Runge-Kutta algorithm, which will propagate to result in large errors in subsequent steps.

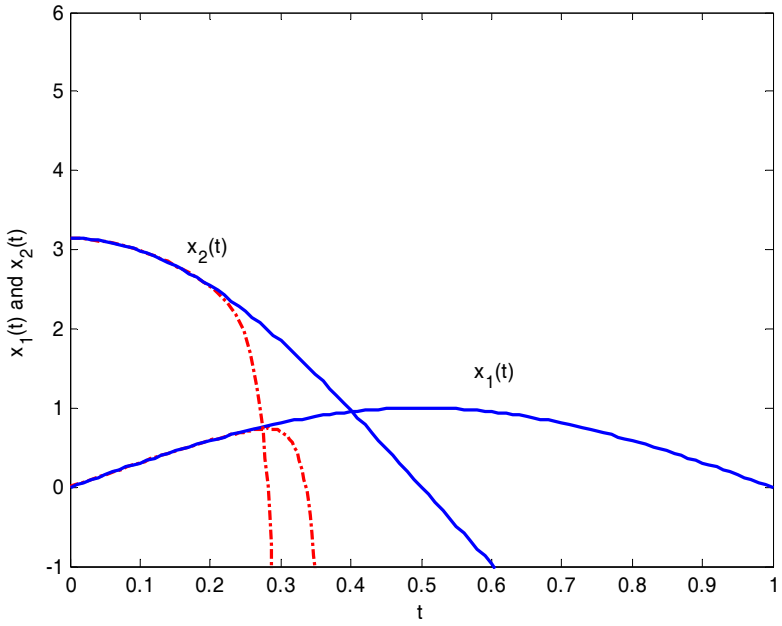

Figure 1.1 Numerical and true solutions of the ODEs in (1.3); (dashed line: numerical solution obtained by single shooting; solid line: true solution). Note that the numerical and true trajectory curves overlap at low times.

This stability issue can be addressed by using a piece-wise polynomial discretization scheme for the ODE solution. In this method, the state trajectories are approximated by piece-wise polynomial functions, such as B-spline basis function expansions (de Boor et al., 1973; de Boor, 2001). The appropriate values of the polynomial coefficients are then determined by forcing the piece-wise polynomial to satisfy the ODE model at some specified points, which are called the collocation points (Ascher et al., 1988). Figure 1.2 shows that using a B-spline collocation

method to approximate the solution of the model in (1.3) eliminates the problem observed in Figure 1.1.

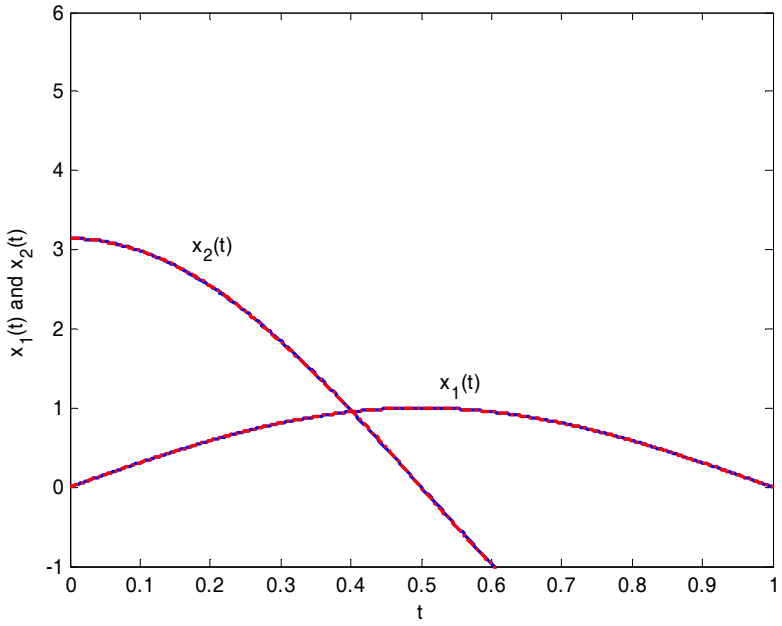

Figure 1.2 B-spline approximation and true solutions of the ODEs in (1.3); (dashed line: B-spline approximation; solid line: true solution). Note that the numerical and true trajectory curves are nearly coincident.

Another important issue in parameter estimation algorithms for dynamic models is that the developed models, which are generally derived based on material, energy, and momentum balances, are not exact. These equations are merely approximations of the physical processes. Model discrepancies can arise from model simplifications, and from input and process disturbances (Gagnon and MacGregor, 1991). Hence, the traditional NLS method, and other methods that disregard model imperfections may not result in the best parameter estimates. Modelling imperfections are very common in chemical engineering applications; therefore,

developing parameter estimation tools that account for possible modelling errors and process disturbances, is of crucial importance. Recently, Poyton et al. (2006) proposed an iterative technique referred to as iteratively-refined Principal Differential Analysis (iPDA) for parameter estimation in continuous-time nonlinear dynamic models. iPDA addresses the error-propagation and stability issues, by approximating the model states using piece-wise polynomial functions (B-splines). Possible modelling discrepancies are also taken into account in iPDA, by allowing the estimated state trajectories to differ from the solution of the model ODEs. A brief account of the iPDA algorithm is given below. A detailed discussion appears in Chapter 3.

In iPDA, the state trajectory is approximated using a B-spline curve of the form:

$$\underset{\sim}{x}(t) = \sum_{i=1}^{c} \beta_i \phi_i \tag{1.4}$$

where β_i, $i = 1 \cdots c$ are B-spline coefficients and $\phi_i(t)\ i = 1 \cdots c$ are B-spline basis functions. As an example, in Figure 1.3 seven 4th order (3rd degree) B-spline basis functions are illustrated. The shown basis functions can be used to approximate *x*(t) for $t \in [0,40]$. The knot sequence used in this example is (0,10,20,30,40). The fourth B-spline function for instance has the following form:

$$\phi_4(t) = \begin{cases} \dfrac{t^3}{6000} & 0 \le t \le 10 \\ -\dfrac{1}{3} + \dfrac{t}{20} + \dfrac{(t-10)^2}{200} - \dfrac{(t-10)^3}{2000} & 10 \le t \le 20 \\ \dfrac{2}{3} - \dfrac{(t-20)^2}{100} + \dfrac{(t-20)^3}{2000} & 20 \le t \le 30 \\ \dfrac{5}{3} - \dfrac{t}{20} + \dfrac{(t-30)^2}{200} - \dfrac{(t-30)^3}{6000} & 30 \le t \le 40 \end{cases} \tag{1.5}$$

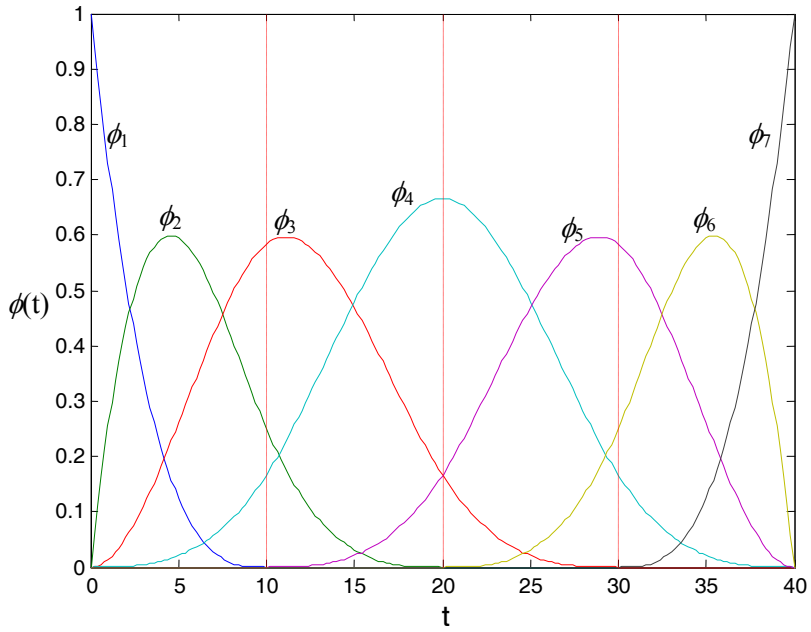

Figure 1.3 Seven B-spline basis functions used to construct spline fit

The iPDA objective function used for parameter estimation, is

$$\sum_{j=1}^{n}(y(t_{mj})-x_{\sim}(t_{mj}))^2+\lambda\int_{t_0}^{t_{mn}}(\dot{x}_{\sim}(t)-f(x_{\sim}(t),u_{\sim}(t),\boldsymbol{\theta}))^2dt \qquad (1.6)$$

where n is the number of data points and t_{mj} is the time at which the jth data point was measured.

iPDA alternates between the following two steps until convergence occurs. In the first step, the objective function shown in (1.6) is minimized with respect to B-spline coefficients, using current parameter estimates. In the second step, the same objective function in (1.6) is minimized with respect to model parameters, using current B-spline coefficients obtained from the first step. Note that the vector of model parameters, $\boldsymbol{\theta}$, appears only in the integrand in (1.6). Hence, the first term in the objective function can be ignored in the second step.

The first term in objective function (1.6), which is a sum of squared errors term, ensures that the B-spline trajectories are close to the measurements, while the second term forces the B-spline curve to satisfy the model ODE to some extent, which is determined by a positive weighting factor λ. In other words, λ determines how the empirical B-spline curve balances between matching the observed data and satisfying the ODE model. In iPDA, the modeller can account for modelling errors and process disturbances by selecting a relatively small λ. A very large λ on the other hand, corresponds to the assumption that the model is perfect with no modelling discrepancies. In the case of a very large λ, iPDA and NLS lead to the same parameter estimates.

This thesis builds on the work presented by Poyton et al. (2006). A survey of existing parameter estimation methods is presented in Chapter 2. In Chapter 3, iPDA is extended for use in stochastic differential equation (SDE) models (Astrom 1970, Maybeck 1979, Brown and Hwang 1992; Kloeden and Platen, 1992), which can be obtained by adding a continuous zero-mean stationary white-noise process, $\eta(t)$, to the right-hand- side of the ODE in (1.1):

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), \boldsymbol{\theta}) + \eta(t) \\ x(t_0) &= x_0 \\ \\ y(t_{mj}) &= x(t_{mj}) + \varepsilon(t_{mj}) \end{aligned} \tag{1.7}$$

where the intensity of $\eta(t)$ is Q. Further information about continuous white-noise processes and SDEs is presented in Chapter 3 where the probabilistic interpretation of the iPDA objective function, shown in (1.6), is also demonstrated. This development naturally leads to an equation for selecting the optimal weighting factor in the iPDA objective function:

$$\lambda_{opt} = \frac{\sigma_m^2}{Q} \tag{1.8}$$

This optimal choice of λ ensures that iPDA properly accounts for the measurement noise and the model imperfections. In Chapter 3, the effectiveness of the iPDA methodology is demonstrated using a simple Continuous Stirred Tank Reactor (CSTR) example. It is shown that the minimization of (1.6), which is performed in an iterative fashion in iPDA, can also be performed simultaneously over the joint vector of the model parameters and the B-spline coefficients. From Chapter 4 onwards, instead of the iterative scheme, the simultaneous minimization approach is used. Therefore, since the modified algorithm does not iterate between the two steps (spline fitting and parameter estimation), and also because the algorithm corresponds to maximizing an approximate likelihood function (as shown in Chapter 3), the modified iPDA algorithm is referred to as Approximate Maximum Likelihood Estimation (AMLE) from Chapter 4 onwards.

In Chapter 4 the applicability of AMLE to parameter estimation in stochastic dynamic models, in which unmeasured states and non-stationary disturbances are present, is demonstrated. Furthermore, theoretical approximate confidence interval expressions for the model parameters are derived and are compared to results from Monte Carlo simulations.

Determining the optimal value for the weighting factor λ in (1.8), requires knowledge of the amount of modelling error (the process disturbance intensity in the model) as well as the value of the measurement noise variance. In engineering applications, knowledge about the quality of the measurements that are available for parameter estimation can be obtained from repeated measurements or from sensor suppliers, but *a priori* knowledge about the quality and accuracy of the model equations is not easily attainable. To address this problem, in Chapter 5, the AMLE algorithm is extended so that it can be applied to parameter estimation in SDE models, in which the process disturbance intensity, $Q_{,,}$ is not known *a priori*. The theoretical results are tested using a nonlinear CSTR simulation study.

The theoretical developments presented in Chapter 3 to Chapter 5, are tested using simple two-state CSTR simulation studies. To examine the performance of AMLE in a practical chemical engineering problem with a larger scale, parameter estimation in a lab-scale nylon reactor model is considered in Chapter 6. First, the structure of the reactor model, initially developed by Schaffer et al. (2003) and subsequently modified by Zheng et al. (2005) and Campbell (2007), is re-evaluated and further improved. Then, AMLE is used to estimate the parameters and states of the revised model using experimental data provided by Schaffer et al. (2003) and Zheng et al. (2005 and 2007). In this practical example, it is shown that AMLE can address practical difficulties such as: different types of measured responses with different levels of measurement noise, measurements taken at irregularly-spaced sampling times, unknown initial conditions for some state variables, unmeasured state variables, and unknown disturbances that enter the process and influence its future behaviour.

The material presented in Chapter 3 to Chapter 6 (sometimes with minor modifications) has been either published, accepted for publication, or submitted to peer-reviewed scientific journals for publication. Therefore, to keep each chapter self contained, and also to keep the presentation in this thesis close to the format of the original manuscripts, some information appears in multiple chapters. Most of the repetition is associated with reviewing the AMLE algorithm at the beginning of each chapter, and introducing the case studies.

1.2 References

U.M. Ascher, R.M.M. Mathheij, R.D. Russell. *Numerical Solutions of Boundary Value Problems for Ordinary Differential Equations.* Prentice Hall, New Jersey, 1988.

K.J. Astrom. *Introduction to Stochastic Control Theory*. Academic Press, Inc., New York, 1970.

Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, Inc., New York, 1974.

D.M. Bates, D.G. Watts. *Nonlinear Regression Analysis and Its applications.* John Wiley & Sons, Inc., New York, 1988.

L.T. Biegler, I.E. Grossman. Retrospective on optimization. *Computers and Chemical Engineering*. 28, 2004, 1169-1192.

H.G. Bock. Recent advances in parameter identification for ordinary differential equations. In *Progress in Scientific Computing*. P. Deuflhard and E. Hairer, editors. Birkhäuser, Boston. 1983, 95-121.

C. de Boor, B. Swartz. Collocation at Gaussian Points. *SIAM Journal of Numerical Analysis* 10 (4), 1973 582-606.

C. de Boor. *A practical Guide to Splines.* Springer, New York, 2001.

R.G. Brown, P.Y.C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering* 2nd edition. John Wiley & Sons, Inc., 1992.

D.A. Campbell. *Bayesian Collocation Tempering and Generalized Profiling for Estimation of Parameters for Differential Equation Models*. Ph.D. Thesis, McGill, University, Montreal, Canada, 2007.

J.R. Dormand, P.J. Prince. A family of embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics*. 6 (1), 1980, 19-26.

N.H. El-Farra, P.D. Christofides. Bounded robust control of constrained multivariable nonlinear processes. *Chemical Engineering Science*. 58, 2003, 3025-3047.

P. Englezos, N. Kalogerakis. *Applied Parameter Estimation for Chemical Engineering.* Marcel Dekker, Inc., New York, 2001.

L. Gagnon, J.F., MacGregor. State estimation for continuous emulsion polymerization. *Canadian Journal of Chemical Engineering.* 69, 1991, 648-656.

P.E. Kloeden, E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, 1992.

J.R. Leis, M.A. Kramer. ALGORITHM 658: ODESSA – an ordinary differential equation solver with explicit simultaneous sensitivity analysis. *ACM Transactions on Mathematical Software.* 14, 1988, 61-67.

Z. Li, M. R. Osborne, T. Pravan. Parameter estimation for ordinary differential equations. *IMA J. of Num. Science*. 25, 2005, 264-285.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 1*. Academic Press, New York, 1979.

Z. K. Nagy, R.D. Braatz. Robust nonlinear model predictive control of batch processes. *AICHE Journal.* 49, 2003, 1776-1786.

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

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. *Computers and Chemical Engineering.* 30, 2006, 698-708.

M.A. Schaffer, K.B. McAuley, M.F. Cunningham, E.K. Marchildon. Experimental study and modeling of nylon polycondensationin the melt phase. *Industrial Engineering and Chemistry Research.* 42, 2003, 2946-2959.

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

P. Tanartkit, L. T. Biegler. Stable decomposition for dynamic optimization. *Industrial Engineering and Chemistry Research.* 34, 1995, 1253-1266.

W. Zheng, K.B. McAuley, E.K. Marchildon, and K.Z. Yao. Effects of end-group balance on melt-phase Nylon 612 polycondensation: experimental study and mathematical model. *Industrial Engineering and Chemistry Research.* 44, 2005, 2675-2686.

W. Zheng, K.B. McAuley, E.K. Marchildon, and K.Z. Yao. Melt-phase Nylon 612 polycondensation kinetics: effects of sodium hypophosphite catalyst. *Canadian Journal of Chemical Engineering.* 85, 2007, 180-187.

Chapter 2

Literature Review

The problem of parameter estimation in dynamic models was introduced in Chapter 1. The NLS method, which is the traditional approach for common parameter estimation problems, was also described. By means of an example, it was illustrated that the numerical integration step in the traditional NLS method can result in stability problems. It was also discussed that if the degree of modelling errors and process disturbances is significant, then using the traditional NLS approach (and in general, methods that disregard model imperfections), may not be appropriate. To address problems associated with the traditional NLS method, researchers have proposed various parameter estimation algorithms. In this chapter, a survey of these alternative approaches to parameter estimation in dynamic models is presented.

This review is divided into two parts. The first part provides a literature review of parameter estimation in deterministic Ordinary Differential Equation ODE models (where model imperfections are usually ignored), while the second part discusses available approaches for parameter estimation in Stochastic Differential Equation (SDE) models, in which modelling errors and process disturbances are explicitly accounted for by adding stochastic disturbance terms to the model equations.

2.1 Parameter estimation in ODE models

The traditional NLS method can be computationally intensive, particularly for complicated models, and also prone to stability and error-propagation issues. Hence, researchers have proposed various modified algorithms for parameter estimation in ODE models to avoid difficulties associated with the traditional approach. Some of these proposed methods are reviewed below.

The solution of the ODE model in (1.1) can be written as:

$$x(t_{mj}) = x(t_0) + \int_{t_0}^{t_{mj}} f(x(t), u(t), \boldsymbol{\theta}) dt \qquad (2.1)$$

To avoid computationally-intensive (for 1967) forward integration (single shooting) of the ODE model, Himmelblau et al. (1967) proposed substituting the noisy state measurements $y(t_{mj})$ (rather than $x(t_{mj})$) and the corresponding inputs into the integrand in (2.1) and calculating the integral using a numerical quadrature method. In this way eq. (2.1) is transformed into an (approximate) algebraic equation, which is regarded as a constraint when parameter estimates are obtained by minimizing the sum of squared residuals with respect to $\boldsymbol{\theta}$. Tang (1971) noted that this method is no longer applicable when only a small number of data points are available for parameter estimation. Therefore, Tang proposed fitting natural cubic splines through calculated values of $f(y(t_{mj}), u(t_{mj}), \boldsymbol{\theta})$ to approximate the integrand in (2.1). Note that spline curves can be integrated analytically, thereby transforming equation (2.1) to an algebraic equation. Vajda et al. (1987) concluded that using this "direct integration" method results in biased parameter estimates, but is computationally more efficient than the traditional NLS approach. They also demonstrated that, in some cases, the direct integration approach can outperform the traditional method and lead to a smaller mean-squared error. Fitting the observations (rather than calculated values of

$f(y(t_{mj}),u(t_{mj}),\boldsymbol{\theta})$) directly by natural cubic splines was proposed by Tang (1971) as an alternative approach, especially if some of the observations are missing.

Swartz and Bremermann (1975) pointed out that the sum of squares of the errors in the predicted *derivatives* of the observations, rather than the traditional sum of squares, can be used as an alternative objective function, for carrying out the parameter estimation problem. For ODE model (1.1), the corresponding objective function that considers deviations in the derivative is:

$$\sum_{i=1}^{n}\left(\dot{x}(t_{mi}) - f(x(t_{mi}),u(t_{mi}),\boldsymbol{\theta})\right)^2 \qquad (2.2)$$

Swartz and Bremermann also noted that small errors in the measured values of the state variables may result in unacceptably large errors in numerical differentiation. Therefore, they proposed fitting a piecewise polynomial to the measurements, which can smooth out much of the noise in the measurements. A similar method was employed by Benson (1979), in which the observations are fitted by piecewise cubic splines and by cubic Hermite polynomials when sharp bends occur in the observed data. As noted by Varah (1982), minimizing the sum of squares of the derivative errors, unfortunately, does not necessarily lead to minimum sum of squared deviations between the observations and the model predictions. Parameter estimates obtained using objective function (2.2) can be biased, inconsistent, and very sensitive to outliers (Maria, 2004).

Swartz and Bremermann and Varah pointed out that using smoothing methods makes the parameter estimates less sensitive to the initial parameter guesses and initial state conditions. The disadvantage of the works of Swartz and Bremermann, Benson, and Varah, however, is that these approaches lack a systematic method for fitting the empirical piece-wise polynomials to data. If the data are noisy, the resulting empirical curves can contain too much rippling, producing unrealistically large derivative values at certain points. If smoother splines are selected to reduce the rippling, then important model information from the data can be lost. A comparative study of

various parameter estimation methods that use a shortcut approach for estimating the integral, like those described above, is provided by Hosten (1979). Hosten concluded that such shortcut methods tend to produce biased parameter estimates. This is because these shortcut methods either treat noisy measurements as true state values (*e.g.*, Himmelblau et al. 1967; Tnag, 1971) or calculate the derivative information based on the noisy measurements (*e.g.*, Swartz and Bremermann, 1975; Benson 1979; Varah, 1982).

An effective approach to parameter estimation in ODE models that does not suffer from error-propagation and stability issues was proposed by Biegler (1984) and discussed by Logsdon and Biegler (1992). In this method, orthogonal collocation is used to discretize the ODE equations and to transform them into (nonlinear) algebraic equations. The parameter estimation problem is then treated as a constrained nonlinear programming problem. Parameter and state estimates can be simultaneously obtained by minimizing the parameter-estimation objective function, subject to the discretized model equations.

Ramsay (1996) recommended using basis functions such as Fourier series or B-spline functions to represent the observations in a method called Principal Differential Analysis (PDA), which is described below. In the first part of this two-step method, basis functions such as B-splines (eq. (1.4)) are used to fit the response empirically, where spline coefficients are obtained from the following optimization problem

$$\sum_{j=1}^{n}(y(t_{mj})-x_{\sim}(t_{mj}))^2+\lambda\int_{t_0}^{t_{mn}}\left(\frac{d^p x_{\sim}(t)}{dt^p}\right)^2 dt \qquad (2.3)$$

In the smoothing-spline roughness-penalty approach shown in eq. (2.3), the integral of the squared pth order derivatives is penalized so that the smoothness of the spline fits is ensured. Often p=2 is used as the order of the penalized derivative.

In the second step, Ramsay's PDA uses the integral of the squares of the deviations of the smoothed curves from the fundamental model as the corresponding objective function for parameter estimation (the second step used in the iPDA algorithm described in Sections1.1 and 3.2.1 uses the same objective function):

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \int \left(\dot{x}_{\sim}(t) - f(x_{\sim}(t), u(t), \boldsymbol{\theta}) \right)^2 dt \right\} \tag{2.4}$$

In other words, Ramsay's PDA approach is based on minimizing the integral squared prediction error of the derivatives, which is convenient because derivatives can be readily computed and integrated from the smoothing spline representation of the state trajectories. Note that in PDA (and iPDA), and unlike the traditional NLS method, the Jacobian matrix required for the minimization step can readily be derived analytically.

Poyton et al. (2006) demonstrated the application of PDA with a second-order derivative penalty using a continuous stirred tank reactor (CSTR) example with one input and two unknown parameters, and reported that in the case of sparse and noisy data, the parameter estimates from PDA were imprecise and biased. The reason is that the penalty term in objective function (2.3) forced the second derivative of the splines to be smaller than the true second derivates, so that the penalty term and the fundamental model structure were inconsistent. Therefore, Poyton et al. (2006) introduced iPDA, which as previously discussed, addresses these problems by using a model-based penalty in the smoothing step, and also by iterating between the smoothing and estimation steps several times until convergence occurs.

Ramsay et al. (2007) proposed a parameter estimation algorithm based on a Generalized Smoothing scheme (GS). Similar to PDA and iPDA, GS is also a collocation-based algorithm that uses B-spline basis functions to discretize the estimated state trajectories. GS consists of outer and inner optimization problems. In the outer, an appropriate objective function, (usually the

weighted sum of squared errors, similar to that in nonlinear least-squares) is minimized with respect to the model parameters. The objective function in the inner optimization problem is the same objective function with model-based roughness penalty as used in iPDA (eq. (1.6)). Note that the parameter estimates obtained using GS and iPDA are different, due to the different outer (or step 2) objective functions. The relationship between the GS and AMLE (or iPDA) schemes is described further in Chapter 6.

Problems faced in traditional NLS can also be addressed by the method of multiple shooting (Bock, 1981&1983; Baake et al., 1992; Kallrath et al., 1993; Timmer 1998; Timmer et al., 2000; Horbelt et al., 2002). In this method the fitting interval is partitioned into several subintervals. The ODEs are then solved numerically over each of the subintervals. Since initial values for each interval are not precisely known, they are included as additional parameters in the algorithm to be estimated along with the fundamental model parameters. Initial guesses for these additional parameters are simply chosen as observations at those points; therefore, the generated trajectories are close to the observations, making multiple shooting numerically more stable than traditional NLS. A disadvantage of multiple shooting is that it can be computationally very expensive since generally a large number of ODEs with unknown initial conditions is integrated numerically. It should be noted that a collocation-based method can be used to solve the ODEs in the sub-intervals.

Among the various parameter estimation approaches that were discussed in this section, those based on multiple shooting (Bock, 1983) and orthogonal collocation (Biegler, 1984), have received considerable attention because of their sound stability properties. These methods however, do not take possible modelling discrepancies into account.

iPDA (Poyton et al., 2006) and GS (Ramsay et al., 2007) methods also posses desirable stability properties that stem from discretizing state trajectories using piece-wise polynomial functions. In addition, the form of the objective function (1.6) that is used in iPDA and GS allows for taking possible modelling errors into account, since the estimated state trajectories can deviate from the solution of the ODEs to some extent. One disadvantage of these methods, however, is that they require specification of the weighting factor, λ, which is usually selected by trial and error. This disadvantage is addressed in Chapters 3 and 5 of this thesis, where methods for selecting appropriate values of λ are discussed.

2.2 Parameter estimation in SDE models

As previously discussed, mathematical models of physical processes are not exact and modelling discrepancies are inevitable. A common approach to account for modelling errors and imperfections is to consider random process disturbances as additional inputs to the model, thereby, transforming the model equations to stochastic differential equations (SDEs) (*e.g.,* eq. (1.7)). In a general SDE frame-work, stochastic disturbances can enter the model in a nonlinear fashion; however, for tractability issues, the most widely-used models consider additive stochastic disturbances (Maybeck, 1982). Our goal in this section is not to discuss stochastic systems in any detail, but rather to lay out the difficulties that arise when estimating parameters in SDE models, and to briefly review the most widely used approaches to parameter estimation in SDEs. Before proceeding, it should be mentioned that the first equation in (1.7) can be more rigorously written as $dx(t) = f(x(t), u(t), \boldsymbol{\theta})dt + dw(t)$ where dw(t) is the increment of a Wiener process or a Brownian motion. The reason for this representation is that the continuous white noise, $\eta(t)$, cannot be integrated since it has infinite variance (Maybeck, 1979). Note that the covariance matrix of $\eta(t)$ is $E\{\eta(t)\eta(t+\tau)\} = Q\delta(\tau)$, where Q is the corresponding power

spectral density and $\delta(.)$ is the Dirac delta function. Furthermore, the solution of (1.7) can then be written as (Maybeck, 1979; Kloeden and Platen, 1992):

$$x(t_f) = x(t_0) + \int_{t_0}^{t_f} f(x(t), u(t), \boldsymbol{\theta}) dt + \int_{t_0}^{t_f} dw(t) \tag{2.5}$$

where *dw*(t) is the increment of a Wiener process or a Brownian motion.

The Maximum Likelihood (ML) method is a popular parameter estimation method for standard nonlinear regression problems where model discrepancies are ignored, due to the desirable statistical properties of the ML parameter estimates (Rao, 1973; Timmer, 2000). In the ML approach, model parameters are estimated, so that the likelihood of the measurements (*i.e.*, the joint density function of the measured states given the model parameters) is maximized. Obtaining the density function of the measurements, however, requires integrating out the (unobserved) states. In SDE models, difficulties arise since the states and stochastic disturbances undergo a nonlinear mapping. Therefore, obtaining the corresponding density functions becomes very difficult, and in general requires the solution of a set of partial differential equations (*e.g.*, Jazwinski, 1970; Maybeck, 1982). If the SDE is linear, the time-varying state covariance matrix can be obtained by solving a set of ordinary differential equations (Maybeck, 1979).

In engineering applications, Kalman-filter-related algorithms for estimating parameters in SDE models are popular. In linear models, the likelihood function can be obtained by solving the standard Kalman filter equations. In nonlinear models, the Extended Kalman filter, which is based on iteratively linearizing the nonlinear SDE model, can be used to obtain an approximate likelihood function (Kristensen et al., 2004; Voss et al., 2004). 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 (Sitz et al., 2002; Julier and Uhlmann,

2004) or ensemble averaging (Evensen, 2003). These methods, however, are computationally intensive.

An alternative method to maximize the ML criterion in SDE models is the Expectation Maximization (EM) method (Roweis and Ghahramani, 2001). EM is an iterative algorithm that continuously solves the following two steps until convergence occurs. In the first step a smoothing problem is solved, in which optimal states, $x(t_i)$, are estimated given measurements $y(t_i)$, and a current estimate for the model parameters. In the second step, the optimal states obtained from the first step, along with the measurements, are used to re-estimate the model parameters. New parameter estimates are obtained by maximizing the expected log-likelihood of the joint data (states and measurements), where the expectation is taken over the distribution of the states given the measurements and the current estimate of the model parameters. EM is an elegant method; however, calculating the expectation in the second step makes EM very computationally intensive. Note that iPDA and EM are very similar. Nonetheless, there are two main distinctions. The first difference is in the first step. In the EM method presented by Roweis and Ghahramani (2001), the smoothing problem is solved by employing the Extended Kalman smoother, whereas, as discussed in Chapter 3, iPDA transforms the smoothing problem into a nonlinear programming problem by deriving the pertinent likelihood function using a continuous-time SDE model, and then discretizing the likelihood function using piece-wise polynomials. The second difference between EM and iPDA is associated with the parameter-estimation objective function in the second step. In iPDA the log-likelihood of the joint data is maximized whereas, as discussed above, in the second step of EM the *expected* value of the joint data is maximized. Comprehensive reviews of articles related to parameter estimation in SDEs are provided by Young (2000), Bohlin and Graebe (1995), and Nielsen et al. (2000).

In this chapter, two aspects of the parameter estimation problem in dynamic models were considered. In the first part of the chapter, various approaches for parameter estimation in ODE models were reviewed. Several methods that address the computational issues associated with integrating ODEs were discussed. Parameter estimation in SDE models that account for possible modelling imperfections was reviewed in the second part of the chapter. It was demonstrated that addition of stochastic process disturbances can significantly increase the difficulties of the parameter estimation problem. The remainder of this thesis is devoted to developing a novel and easy-to-use parameter estimation technique that can address difficulties that are encountered when estimating parameters in imperfect ODE models and SDE models.

2.3 References

E. Baake, M. Baake, H.G. Bock, K.M. Briggs. Fitting ordinary differential equations to chaotic data. *Physical Review A*. 45 (8), 1992, 5524-5529.

L. T. Biegler. Short note solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering.* 8, 1984, 243-248.

M. Benson. Parameter fitting in dynamic models, *Ecological Modelling.* 6, 1979, 97-115.

H.G. Bock. Numerical treatment of inverse problems in chemical reaction kinetics, in *Modelling of Chemical Reaction Systems*. K.H. Ebert, P. Deuflhard, and W. Jäger, editors. Springer, Berlin. Chapter 8, 1981, 102-125.

H.G. Bock. Recent advances in parameter identification for ordinary differential equations. In *Progress in Scientific Computing*. P. Deuflhard and E. Hairer, editors. Birkhäuser, Boston. 1983, 95-121.

T. Bohlin, S.F. Graebe. Issues in nonlinear stochastic grey-box identification. *International Journal of Adaptive Control and Signal Processing.* 9, 1995, 465-490.

G. Evensen. The ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics.* 53, 2003, 343-367.

D.M. Himmelblau, C.R. Jones, K.B. Bischoff. Determination of rate constants for complex kinetics models. *Industrial and Engineering Chemistry Fundamentals.* 6, 1967, 539-543.

W. Horbelt, J. Timmer, H. U. Voss. Parameter estimation in nonlinear delayed feedback systems from noisy data. *Physics Letters A*. 299, 2002, 513-521

L. H. Hosten. A comparative study of short cut procedures for parameter estimation in differential equations. *Computers and Chemical Engineering.* 3, 1979, 117-126.

A.H. Jazwinski. *Stochastic Processes and Filtering theory*. Academic Press, New York, 1970.

S.J. Julier, J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE.* 92 (3), 2004, 401-422.

J. Kallrath, J.P. Schloder, H.G. Bock. Least squares parameter estimation in chaotic differential equations. *Celestial Mechanics and Dynamical Astronomy*. 56, 1993, 353-371.

P.E. Kloeden, E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, 1992.

N.R. Kristensen, H. Madsen, S.B. Jorgensen. Parameter estimation in stochastic grey-box models. *Automatica*. 40, 2004, 225-237.

J.S. Logsdon, L.T. Biegler. Decomposition strategies for large-scale dynamic optimization problems. *Chemical Engineering Science.* 47, 1992, 851-864.

G. Maria. A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chemical and Biochemical Engineering*. 18 (3), 2004, 195-222.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 1*. Academic Press, New York, 1979.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 2*. Academic Press, New York, 1982.

J.N. Nielsen, H. Madsen, P.C. Young. Parameter estimation in stochastic differential equations: an overview. *Annual Reviews in Control.* 24, 2000, 83-94.

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. *Computers and Chemical Engineering.* 30, 2006, 698-708.

J.O. Ramsay. Principal Differential Analysis: Data reduction by differential operators. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*. 58, 1996, 495-508.

J.O. Ramsay, G. Hooker, D. Campbell, J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of Royal Statistical Society*: *Series B* (*Statistical Methodology*). 69 (5), 2007, 741-796.

C.R. Rao. *Linear Statistical Inference and Applications*, 2nd edition. John Wiley and Sons, 1973.

S. Roweis, Z. Ghahramani. Learning nonlinear dynamical systems using the Expectation-Maximization algorithm. In *Kalman Filtering and Neural Networks*. S. Haykin, editor. John Wiley & Sons, New York. 2001, 175-220.

A. Sitz, U.Schwarz, J. Kurths, H.U. Voss. Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Physical Review E*. 66, 2002, 016210-1-016210-9.

J. Swartz and H. Bremermann. Discussion of parameter estimation in biological modelling: algorithms for estimation and evaluation of estimates. *Journal of Mathematical Biology.* 1, 1975, 241-275.

Y.P. Tang. On the estimation of rate constants for complex kinetic models. *Industrial and Engineering Chemistry Fundamentals*. 10, 1971, 321-322.

J. Timmer. Modeling noisy time series: physiological tremor. *International Journal of Bifurcation and Chaos*. 8, 1998, 1505-1516.

J. Timmer. Parameter estimation in nonlinear stochastic differential equations. *Chaos, Solitons and Fractals*. 11, 2000, 2571-2578.

J. Timmer, H. Rust, W. Horbelt, H.U. Voss. Parametric, nonparametric and parametric modelling of chaotic circuit time series. *Physics Letters A*. 274, 2000, 123-134.

S. Vajda, P. Valko, K.R. Godfrey. Direct and indirect least squares methods in continuous-time parameter estimation. *Automatica*, 23 (6), 1987, 707-718.

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

H.U. Voss, J. Timmer, J. Kurths. Nonlinear dynamical system identification from uncertain and indirect measurements. *International Journal of Bifurcation and Chaos.* 14 (6), 2004, 1905-1933.

P. Young. Parameter estimation for continuous-time models-a survey. *Automatoca*. 17 (1), 1981, 23-39.

Chapter 3

Selecting Optimal Weighting Factors in iPDA for Parameter Estimation in Continuous-Time Dynamic Models

3.1 Abstract

As briefly discussed in Chapter 1, iteratively-refined Principal Differential Analysis (iPDA) (Poyton et al., 2006) is a spline-based method for estimating parameters in ordinary differential equation (ODE) models. In this chapter we extend iPDA for use in differential equation models with stochastic disturbances, and we demonstrate the probabilistic basis for the iPDA objective function using a maximum likelihood argument. This development naturally leads to a method for selecting the optimal weighting factor in the iPDA objective function. The effectiveness of iPDA is demonstrated using a simple two-output continuous-stirred-tank-reactor example. Monte Carlo simulations are used to show that iPDA parameter estimates are superior to those obtained using traditional nonlinear least squares techniques, which do not account for stochastic disturbances.

This chapter has been accepted for publication as a journal paper by *Computers and Chemical Engineering*.

Selecting Optimal Weighting Factors in iPDA for Parameter Estimation in Continuous-Time Dynamic Models

M.S. Varziri, A.A. Poyton, K. B. McAuley and P. J. McLellan, Department of Chemical Engineering, Queen's University, Kingston, ON, Canada K7L 3N6

J. O. Ramsay, Department of Psychology, McGill University, Montreal, PQ, Canada H3A 1B1

3.2 Introduction

Parameter estimation in mathematical models is an important, difficult, and ubiquitous problem in chemical engineering and in many other areas of applied science. Fundamental process models can be exploited by many process optimization and control technologies (Biegler and Grossman, 2004), but it is important that appropriate parameter values are used so that model predictions match the underlying process behaviour. Obtaining good parameter values requires informative data for parameter estimation, as well as reliable parameter estimation techniques.

It is particularly difficult to estimate parameters in ordinary differential equation (ODE) models. The weighted sum of squared prediction errors is the usual minimization criterion for parameter estimation, and evaluating this criterion requires (numerical) solution of the ODEs. Sensitivity information, used by gradient-based parameter-estimation techniques, requires the solution of sensitivity equations (Leis and Kramer, 1988) or numerous additional simulations using perturbed parameter values. Numerical overflow and stability problems can arise when poor initial or intermediate parameter values are used during the course of parameter estimation (Biegler and Grossman, 2004).

A variety of ODE parameter estimation techniques have been used, ranging from traditional least-squares methods combined with repeated solution of differential equations (Bard, 1974; Bates and Watts, 1988; Seber and Wild, 1989; Ogunnaike and Ray, 1994;) to multiple shooting (Bock, 1981&1983), collocation-based methods (Biegler, 1984), and algorithms that use spline functions (Varah, 1982; Vajda and Valko, 1986). Biegler and Grossman (2004) provide a detailed survey of the existing methods.

In an attempt to develop an efficient and easy-to-use algorithm for estimating parameters in ODE models, Poyton et al. (2006) proposed iteratively refined Principal Differential Analysis (iPDA), which builds upon ideas from Principal Differential Analysis (PDA). PDA is a functional data analysis tool that was proposed by Ramsay (1996) for empirical modelling using linear ODEs. PDA makes use of basis functions (usually B-splines) for estimating ODE parameters (Ramsay, 1996; Ramsay and Silverman, 2005; Poyton et al., 2006).

In this paper we will address some of the issues raised by Poyton et al. (2006) regarding iPDA. Most importantly we extend the use of iPDA to cases in which process noise (unmeasured disturbances that pass through the process) and measurement noise are both present, and we propose a criterion for selecting optimal weighting factors in the iPDA algorithm. We begin with a brief review of the iPDA algorithm and its advantages and shortcomings. Then we describe how iPDA can be used to estimate parameters in differential equation models with process disturbances (stochastic ODE models (Maybeck, 1979)). Finally, we use a simple continuous-stirred-tank-reactor example to compare parameter estimates obtained using iPDA with those obtained using a traditional Nonlinear Least Squares (NLS) approach (Ogunnaike and Ray, 1994), which does not account for the stochastic process disturbances.

3.2.1 iPDA algorithm

To explain the iPDA algorithm, we will use the following simple first-order single-input single-output (SISO) nonlinear ODE model:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), \boldsymbol{\theta}) \\ x(t_0) &= x_0 \end{aligned} \qquad (3.1)$$

$$y(t_{mj}) = x(t_{mj}) + \varepsilon(t_{mj})$$

x is the state variable, u is the input variable and y is the output variable (which is the same as the state variable in this case). f is a nonlinear function of the vector of model parameters $\boldsymbol{\theta}$, state variables, and input variables. ε is a zero-mean uncorrelated random variable with variance σ_m^2.

The *first step* in iPDA is to fit a B-spline curve to the observed data. This empirical B-spline model is of the form:

$$x_{\sim}(t) = \sum_{i=1}^{c} \beta_i \phi_i \qquad (3.2)$$

where β_i, $i = 1 \cdots c$ are B-spline coefficients and $\phi_i(t)\ i = 1 \cdots c$ are B-spline basis functions, for which a knot sequence must be specified. Please refer to Poyton et al. (2006) for a short introduction to B-splines and to de Boor (2001) for a detailed treatment. Note that eq. (3.2) can be written in matrix form:

$$x_{\sim}(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\beta} \qquad (3.3)$$

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

$$\dot{x}_{\sim}(t) = \frac{d}{dt}\left(\sum_{i=1}^{c} \beta_i \phi_i(t) \right) = \sum_{i=1}^{c} \beta_i \dot{\phi}_i(t) = \dot{\boldsymbol{\varphi}}^T \boldsymbol{\beta} \qquad (3.4)$$

The empirical function $x_{\sim}(t)$ is determined by selecting the spline coefficients $\boldsymbol{\beta}$ that minimize the following objective function, given the most recent estimates for the fundamental model parameters $\boldsymbol{\theta}$:

$$\min_{\boldsymbol{\beta}}\left\{\sum_{j=1}^{n}(y(t_{mj})-x_{\sim}(t_{mj}))^2+\lambda\int_{t_0}^{t_{mn}}(\dot{x}_{\sim}(t)-f(x_{\sim}(t),u_{\sim}(t),\boldsymbol{\theta}))^2dt\right\} \tag{3.5}$$

where n is the number of data points and t_{mj} is the time at which the jth data point was measured. We will refer to the first term $\sum(y(t_{mj})-x_{\sim}(t_{mj}))^2$ in the objective function as SSE (the sum of squared prediction errors) and to the second term $\int(\dot{x}_{\sim}(t)-f(x_{\sim}(t),u_{\sim}(t),\boldsymbol{\theta}))^2dt$ as PEN (the model-based penalty). PEN is a measure of how well the empirical curve satisfies the ODE model. In the initial iteration of iPDA, initial guesses for the fundamental parameters $\boldsymbol{\theta}$ are required to compute PEN. Optimal B-spline coefficients $\boldsymbol{\beta}$ are obtained by considering the measured observations and also the extent to which the empirical curve satisfies the model. The model-based penalty prevents B-spline curves that make the SSE small, but are inconsistent with the behaviour of the fundamental model. The positive weighting factor λ determines how the empirical B-spline curve balances between matching the observed data and satisfying the ODE model. A small λ is appropriate when we believe the measurements more than the model and a large λ is appropriate when we have confidence in our model but our measured observations are noisy. Poyton et al. (2006) pointed out that proper selection of λ is of crucial importance and showed that the quality of the B-spline curves (and also the parameter estimates $\hat{\boldsymbol{\theta}}$ obtained using iPDA) depends very much on the value of λ. In this paper we propose a means of determining an optimal λ given some knowledge about measurement noise and model disturbances.

The *second step* in iPDA is to estimate the vector of fundamental model parameters $\hat{\boldsymbol{\theta}}$, using fixed values of the B-spline coefficients $\hat{\boldsymbol{\beta}}$ (and hence fixed $x_{\sim}$) from step one. The fundamental model parameters are selected to minimize the following objective function:

$$\min_{\boldsymbol{\theta}} \int_{t_0}^{t_{mn}} (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta}))^2 \mathrm{dt} \tag{3.6}$$

Next, we return to the first step and re-estimate the B-spline coefficients using $\hat{\boldsymbol{\theta}}$ obtained from step two. iPDA iterates between step one and step two until convergence, as shown in Figure 3.1.

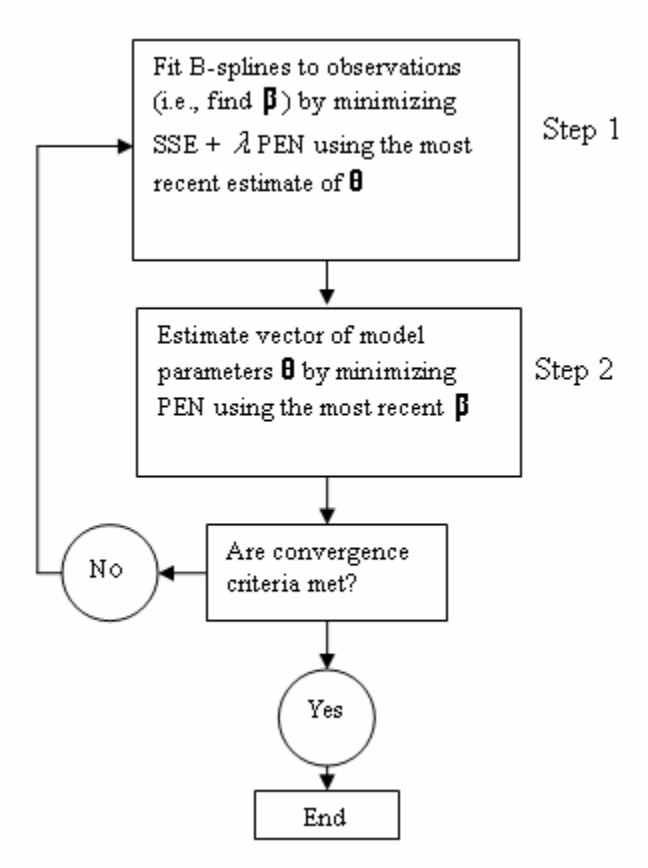

Figure 3.1. iPDA algorithm

The original PDA algorithm (Ramsay, 1996) is not iterative, and it uses penalties on higher-order derivatives (*e.g.*, $\int (\ddot{x}(t))^2 dt$) to prevent over-fitting of the data, instead of the model-based penalty of iPDA. PDA has been used in various areas such as handwriting analysis (Ramsay, 2000), analysis of movement of the lips during speech (Lucero, 2002; Ramsay and Munhall, 1996), economic modelling (Ramsay and Ramsey, 2002) and meteorological modelling (Ramsay and Silverman, 2005). Poyton et al. (2006) showed that poor spline fits from the first step of the original PDA algorithm give misleading derivative information, which results in inaccurate parameter estimates from the second step. They showed that inclusion of the model-based penalty in the first step of iPDA, along with its iterative nature, eliminates this problem.

Since $\sum (y(t_{mj}) - x_{\sim}(t_{mj}))^2$ is constant when $\hat{\boldsymbol{\beta}}$ is fixed, step two of iPDA is equivalent to minimizing:

$$\min_{\boldsymbol{\theta}} \left\{ \sum_{j=1}^{n} (y(t_{mj}) - x_{\sim}(t_{mj}))^2 + \lambda \int (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta}))^2 dt \right\} \tag{3.7}$$

As a result, when the estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ have converged, the following overall objective function is minimized:

$$\min_{\boldsymbol{\beta},\boldsymbol{\theta}} \left\{ \sum_{j=1}^{n} (y(t_{mj}) - x_{\sim}(t_{mj}))^2 + \lambda \int (\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta}))^2 dt \right\} \tag{3.8}$$

This optimization problem can be solved simultaneously for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, instead of using the iterative two-step procedure described above. Since the vector of spline coefficients $\boldsymbol{\beta}$ is generally of high dimension, eq. (3.8) is the objective function for a large-scale nonlinear minimization problem. One benefit of the crude iterative approach shown in Figure 3.1 is that it can simplify this large-scale nonlinear problem: If the ODE is linear in the inputs and outputs, then the first step of iPDA is a large *linear* least-squares problem, and the second step of iPDA,

which is a nonlinear least-squares problem is of much smaller dimension. Nonetheless, we believe that alternative approaches for solving large-scale minimization methods (Biegler, 1984) should be investigated for obtaining the iPDA parameter estimates.

3.2.2 iPDA advantages, shortcomings and the purpose of current paper

Since the empirical B-spline curve that is fitted to the observations can be easily differentiated with respect to time, iPDA circumvents the need for repeated numerical solution of ODEs, which is required by traditional NLS methods (Ogunnaike and Ray, 1994). Solving ODEs numerically during NLS estimation can sometimes lead to numerical overflow and instabilities, especially when the initial guesses (or along-the-way estimates) of model parameters are poor, or if the dynamic model contains unstable modes (Bard, 1974; Ascher et al., 1988; Tanartkit and Biegler, 1995; Li et al., 2005). These problems are not encountered by iPDA.

Another advantage of iPDA arises from the form of the objective function for the parameter estimation step (step two). Since the integral of the squared deviation for the *differentiated* form of the model is minimized in eq. (3.6), as opposed to the sum of squared prediction errors as in traditional NLS, sensitivity information is readily available. Analytical derivatives can be used because $\dot{x}_{\sim}(t) - f(x_{\sim}(t), u_{\sim}(t), \boldsymbol{\theta})$ can easily be differentiated with respect to $\boldsymbol{\theta}$. Thus, unlike traditional NLS, there is no need to numerically integrate sensitivity equations (Bard, 1974; Bates and Watts, 1988; Seber and Wild, 1989). Objective function (3.6) has a further advantage in that the nonlinearity of parameters is often less severe in the differentiated form of the model than in the integrated solution (*e.g.*, kinetic rate constants and heat-transfer coefficients often appear linearly on the right-hand side of the ODE, but would appear in exponential terms in the integrated response). Therefore, iPDA may be less susceptible to problems associated with poor initial parameter guesses than are traditional methods.

iPDA is particularly well suited to parameter estimation in ODE models in which some or all of the initial conditions for the states are unknown. When using iPDA, there is no need to repeatedly solve the ODEs with different guesses for the initial conditions. Estimates for initial conditions are provided automatically by $\underset{\sim}{x}(0)$. As we will show using the examples in this article, it is straightforward to incorporate known initial conditions in the B-spline curve-fitting step by fixing the value of one of the spline coefficients. Parameter estimation problems involving state-variable constraints (Biegler and Grossman, 2004) could also be readily handled using iPDA.

Using basis functions, either to empirically fit the observations or to approximate the solution of the ODEs (collocation-based methods), during parameter estimation, is not exclusive to iPDA (*e.g.*, Tang, 1971; Swartz and Bremermann, 1975; Benson, 1979; Varah, 1982; Vajda and Valko, 1986; Biegler, 1984; Logsdon and Biegler 1992). Several types of basis functions have been used for discretizing ODE models during parameter estimation (Biegler and Grossmann, 2004). For example, Biegler (1984) used Lagrange interpolating polynomials because they facilitate providing bounds and starting points for coefficients that are to be estimated. B-spline functions were selected for our iPDA algorithm (and for original PDA) because they are bounded polynomials that are non-zero only over a finite interval. B-splines provide "compact support" for the empirical curve (de Boor, 2001), which leads to banded matrices that are numerically attractive for smoothing and inverse problems (O'Sullivan, 1986; Eilers and Marx, 1996; Ramsay and Silverman, 2005).

A further benefit of iPDA over other basis-function methods is that the model-based penalty (PEN) in the B-spline fitting objective function (eq. (3.5)) regularizes the fitted curve and prevents it from having unrealistic features that are not consistent with the model. Because of this

property, iPDA is an algorithm in the class of regularization methods, which are widely used for solving linear and nonlinear inverse problems (O'Sullivan, 1986; Kirsch, 1996, Binder et al., 2002). Note that the model-based penalty (PEN) in iPDA is not a hard constraint, but rather a soft constraint that is only satisfied to some extent, which is determined by the value of the weighting factor λ. In other collocation-based methods the parameter estimation problem is posed as a hard-constrained minimization problem, where the sum of squared prediction errors (SSE) is minimized subject to the discretized ODE (Biegler and Grossmann, 2004). As suggested by Poyton et al., (2006), imposing the discretized ODE as a soft constraint in iPDA may be particularly advantageous for estimating parameters in models in which unmeasured stochastic disturbances influence dynamic process behaviour. In this paper we consider these types of models and we demonstrate how iPDA readily addresses the resulting parameter estimation problem.

Another issue raised by Poyton et al. (2006) is whether or not iPDA can be used for problems in which some of the states are not measured. Our recent investigations confirm that the answer is yes; iPDA can readily handle estimation problems with unmeasured states (Chapter 4) so long as certain observability conditions are met, which are analogous to the conditions required for estimation of unmeasured states using an extended Kalman filter. It should be noted that the observability of the dynamic model depends on the model parameters which are not known. If the dynamic model is not observable, either changes in some of the states will not affect the responses, or effects of some of the states on the responses are not distinguishable. This will lead to an ill-conditioned estimation problem.

Shortcomings of iPDA as listed by Poyton et al. (2006) are as follows.

- iPDA requires an appropriate B-spline knot sequence. The quality of the parameter estimates $\hat{\boldsymbol{\theta}}$ depends on the empirical B-spline curve which, in turn, relies on the selected knot sequence. Optimal knot placement is currently under investigation in our research work. However, it seems that if enough data points are available, placing one knot at each observation point (as will be shown in the case study) can lead to satisfactory results. Additional knots maybe required when there are sharp changes in the output. Using too many knots can lead to long computational times.

- iPDA parameter estimates depend on the weighting factor λ in eq. (3.5). Heuristically, the weighting factor should depend on measurement uncertainties and model disturbances. The more uncertain the measurements are, the larger λ should be and the more uncertain the model is, the smaller λ should be. The uncertainty in the model (due to unmodelled disturbances or other model imperfections) can have several sources. One possible source of model disturbances is uncertainty in the inputs to the system. Traditional NLS assumes that the inputs to the system are perfectly known, however in practice, due to flaws in measurement devices and valves and also because of external unmeasured or unmodelled disturbances, this is rarely the case. Finally, model uncertainty arises because there are some physical phenomena that have not been included in the model; in other words there may be missing or improperly specified terms in the model. In the current article we add a stochastic term to the right-hand side of the ODE in eq. (3.1) to account for model uncertainties. We then show that minimizing the objective function in eq. (3.5) leads to optimal B-spline coefficients for smoothing the observations, taking into account the levels of process noise and measurement noise in the system. This development naturally leads to an expression for the optimal value of the weighting factor λ.

- No confidence interval expressions have been developed to assist modellers in making inferences about the quality of parameter estimates and model predictions obtained using iPDA. We will address this issue in our ongoing iPDA research (Chapter 4). Also, iPDA has not yet been used for parameter estimation in larger-scale parameter estimation problems (Chapter 6). Further research will be required to determine efficient algorithms for obtaining iPDA parameter estimates.

3.3 iPDA in presence of model disturbance

To keep the notation compact we use a single-input single-output (SISO) nonlinear model; the multi-input multi-output (MIMO) case is presented in the Appendix 3.6.

Consider the following continuous-time stochastic dynamic model (Astrom 1970, Maybeck 1979, Brown and Hwang 1992):

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), \boldsymbol{\theta}) + \eta(t) \\ x(t_0) &= x_0 \end{aligned} \qquad (3.9)$$

$$y(t_{mj}) = x(t_{mj}) + \varepsilon(t_{mj})$$

The initial condition, x_0 is a Normally distributed random variable with mean $E\{x_0\}$ and variance σ_0^2. $\eta(t)$ 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 power spectral density and $\delta(.)$ is the Dirac delta function. For the discrete-time white-noise process (Maybeck, 1979):

$$E\{\eta(j_1\Delta t)\eta(j_2\Delta t)\} = \begin{cases} \dfrac{Q}{\Delta t} & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases} \qquad (3.10)$$

where j_1 and j_2 are integers and Δt is the sampling period. A discrete random white noise sequence as shown in Figure 3.2 is a series of step functions with sampling interval Δt, where the variance of the white noise, $\sigma_p^2 = Q/\Delta t$. In the limiting case where $\Delta t \to 0$ we get the same behaviour as in the continuous case (using the Dirac delta function). All the other terms in (3.9) remain the same as those in (3.1). We also assume that the process noise $\eta(t)$ and the measurement noise $\varepsilon(t)$ are not correlated. In the next section we use maximum likelihood arguments to justify the use of the B-spline-fitting objective function in (3.5) and we prescribe a method for selecting an optimal λ given Q and σ_m^2.

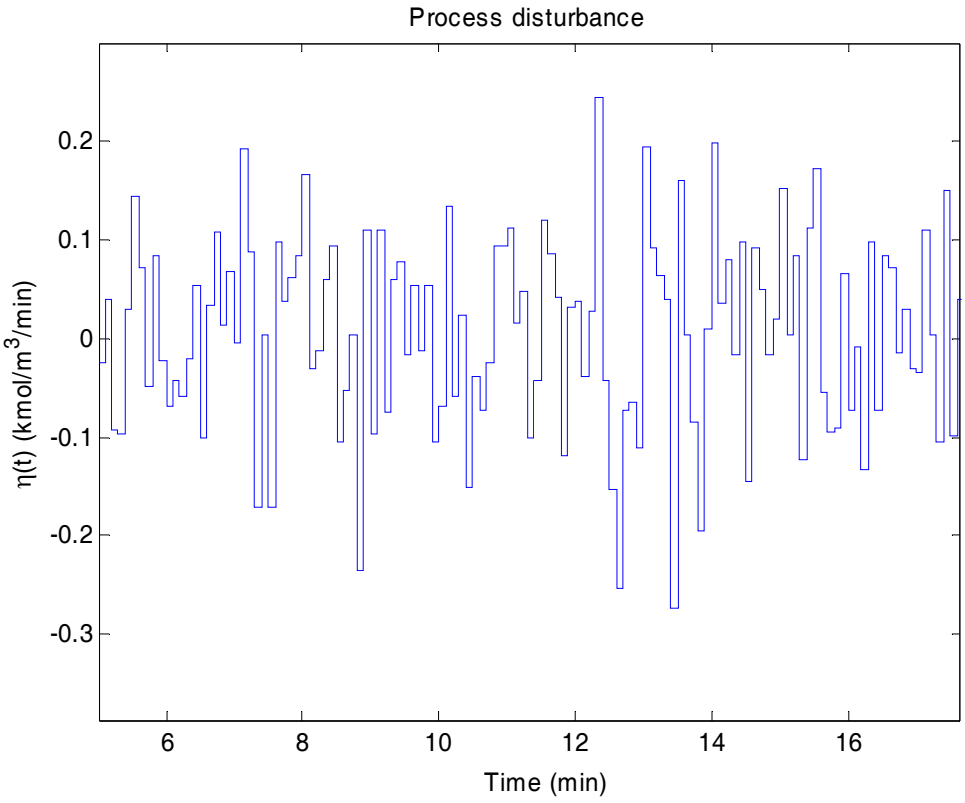

Figure 3.2. Process disturbance

3.3.1 Selecting the optimal weighting factor

We will show that

$$\lambda_{opt} = \frac{\sigma_m^2}{Q} \tag{3.11}$$

where the subscript "opt" indicates the optimal weighting factor, which will lead to maximum *a posteriori* parameter estimates for $\boldsymbol{\beta}$, given $\boldsymbol{\theta}$. The resulting optimal spline curves $x_{\sim}$ will lead to optimal values of $\hat{\boldsymbol{\theta}}$ (the estimate of $\boldsymbol{\theta}$) after the algorithm converges.

Before we begin the mathematical derivations, it is helpful to outline the approach that we will use. The state variable x in (3.9) is a random variable (due to the stochastic input $\eta(t)$), which evolves in time. Values of *x,* sampled at various times, have a prior joint distribution, which is multivariate normal due to the assumptions about $\eta(t)$. Once the measured data become available, the posterior joint distribution of the sampled values of x and the measured response, y, can be obtained. We will consider the general case when the initial value x_0 is not perfectly known, and then we will demonstrate that maximizing the likelihood of the joint distribution of the sampled state values and measured observations given the parameter values, is equivalent to minimizing the iPDA objective function (3.5) in the restricted case when we assume that x_0 in (3.9) is perfectly known.

At the discrete time $t_i = t_{i-1} + \Delta t$, where the sampling interval Δt is small, eq. (3.9) can be written using the following Euler approximation.

$$x(t_{i-1} + \Delta t) = x(t_i) = x(t_{i-1}) + f(x(t_{i-1}), u(t_{i-1}), \boldsymbol{\theta})\Delta t + \eta(\mathrm{t_{i\text{-}1}})\Delta t \tag{3.12}$$

Consider $x(t_i)$ at q+1 uniformly spaced time points, t_i, $i = 0 \cdots q$ 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 $x_i = x(t_i)$.

Please note that the set of times at which the measurements are available is a subset of $t_i\,(i=0\cdots q)$ and is denoted by $t_{mj}\,(j=1\cdots n)$. The measurement times t_{mj} do not need to be uniformly spaced. The vector of outputs at observation times $y(t_{mj})\,(j=1\cdots n)$ and its corresponding state vector of true values $x(t_{mj})\,(j=1\cdots n)$ and measurement noise vector $\varepsilon(t_{mj})\,(j=1\cdots n)$ are denoted by $\mathbf{y}_m$, $\mathbf{x}_m$, and $\boldsymbol{\varepsilon}_m$ respectively.

From Bayes' rule, the joint probability distribution of $x_0,\ldots x_q$ and the vector of observations $\mathbf{y}_m$, given the model parameters can be written as

$$p(x_0,\ldots,x_q,\mathbf{y}_m\mid\boldsymbol{\theta})=p(\mathbf{y}_m\mid x_0,\ldots,x_q,\boldsymbol{\theta})\times p(x_0,\ldots,x_q\mid\boldsymbol{\theta}) \tag{3.13}$$

Due to the Markov property (Maybeck, 1979; Gong et al., 1998):

$$p(x_0,\ldots,x_q\mid\boldsymbol{\theta})=p(x_q\mid x_{q-1},\boldsymbol{\theta})\times\cdots\times p(x_1\mid x_0,\boldsymbol{\theta})\times p(x_0\mid\boldsymbol{\theta}) \tag{3.14}$$

Substituting (3.14) into (3.13) gives:

$$\begin{aligned}&p(x_0,\ldots,x_q,\mathbf{y}_m\mid\boldsymbol{\theta})=\\&p(\mathbf{y}_m\mid x_0,\ldots,x_q,\boldsymbol{\theta})\times p(x_q\mid x_{q-1},\boldsymbol{\theta})\times\cdots\times p(x_1\mid x_0,\boldsymbol{\theta})\times p(x_0\mid\boldsymbol{\theta})\end{aligned} \tag{3.15}$$

We now evaluate each term on the right-hand side of (3.15). From: $y(t_{mj})=x(t_{mj})+\varepsilon(t_{mj})$

$$\mathbf{y}_m=\mathbf{x}_m+\boldsymbol{\varepsilon}_m \tag{3.16}$$

where $\mathbf{x}_m$ is the vector of true state values at the measurement times. Therefore, $p(\mathbf{y}_m\mid x_0,\ldots,x_q,\boldsymbol{\theta})$, is a multivariate Gaussian distribution:

$$p(\mathbf{y}_m\mid x_0,\ldots,x_q,\boldsymbol{\theta})=\frac{1}{(2\pi)^{n/2}\sigma_m^n}\exp\left\{\frac{-(\mathbf{y}_m-\mathbf{x}_m)^T(\mathbf{y}_m-\mathbf{x}_m)}{2\sigma_m^2}\right\} \tag{3.17}$$

with mean $E\{\mathbf{y}_m\mid x_0,\ldots,x_q,\boldsymbol{\theta}\}=\mathbf{x}_m$ and covariance matrix , $\operatorname{cov}\{\mathbf{y}_m\mid x_0,\ldots,x_q\}=\sigma_m^2\mathbf{I}_{n\times n}$.

From, eq. (3.12), $p(x_i \mid x_{i-1})$ is a Gaussian distribution:

$$p(x_i \mid x_{i-1}, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sqrt{Q\Delta t}} \exp\left\{ \frac{-\left(x_i - E\{x_i \mid x_{i-1}, \boldsymbol{\theta}\}\right)^2}{2Q\Delta t} \right\} \tag{3.18}$$

with

$$\begin{aligned} E\{x_i \mid x_{i-1}, \boldsymbol{\theta}\} &= x_{i-1} + f(x_{i-1}, u_{i-1}, \boldsymbol{\theta})\Delta t \text{ and} \\ \operatorname{cov}\{x_i \mid x_{i-1}, \boldsymbol{\theta}\} &= \operatorname{cov}\{\eta_{i-1}\Delta t\} = \left(\frac{Q}{\Delta t}\right)\Delta t^2 = Q\Delta t \end{aligned} \tag{3.19}$$

We assume that the initial condition x_0 has a Gaussian distribution:

$$p(x_0 \mid \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{ \frac{-\left(x_0 - E\{x_0 \mid \boldsymbol{\theta}\}\right)^2}{2\sigma_0^2} \right\} \tag{3.20}$$

Therefore, from equations (3.15), (3.17), (3.18), and (3.20):

$$\begin{aligned} &p(x_0, \ldots, x_q, \mathbf{y}_m \mid \boldsymbol{\theta}) \propto \\ &\exp\left\{ \frac{-(\mathbf{y}_m - \mathbf{x}_m)^T (\mathbf{y}_m - \mathbf{x}_m)}{2\sigma_m^2} \right\} \times \prod_{i=1}^{q} \exp\left\{ \frac{-\left(x_i - E\{x_i \mid x_{i-1}, \boldsymbol{\theta}\}\right)^2}{2Q\Delta t} \right\} \times \exp\left\{ \frac{-\left(x_0 - E\{x_0 \mid \boldsymbol{\theta}\}\right)^2}{2\sigma_0^2} \right\} \end{aligned} \tag{3.21}$$

The optimal state and parameter estimates are denoted $\hat{\mathbf{x}} = \hat{x}_0, \ldots, \hat{x}_q$ and $\hat{\boldsymbol{\theta}}$ respectively and minimize:

$$\frac{(\mathbf{y}_m - \mathbf{x}_m)^T (\mathbf{y}_m - \mathbf{x}_m)}{2\sigma_m^2} + \sum_{i=1}^{q} \frac{1}{2Q} \left(\frac{x_i - E\{x_i \mid x_{i-1}, \boldsymbol{\theta}\}}{\Delta t} \right)^2 \Delta t + \frac{\left(x_0 - E\{x_0 \mid \boldsymbol{\theta}\}\right)^2}{2\sigma_0^2} \tag{3.22}$$

which is the natural logarithm of the right-hand side of eq. (3.21) multiplied by -1.

In the limiting case where $\Delta t \to 0$ (Jazwinski, 1970), eq. (3.19) implies

$$\frac{x_i - E\{x_i \mid x_{i-1}, \boldsymbol{\theta}\}}{\Delta t} \to \dot{x}_{i-1} - f(x_{i-1}, u_{i-1}, \boldsymbol{\theta}) \tag{3.23}$$

and assuming that $E\{x_0 \mid \boldsymbol{\theta}\} = E\{x_0\}$ eq. (3.22) becomes

$$\frac{(\mathbf{y}_m - \mathbf{x}_m)^T (\mathbf{y}_m - \mathbf{x}_m)}{2\sigma_m^2} + \frac{1}{2Q}\int_{t_0}^{t_q} \left(\dot{x}(t) - f(x(t), u(t), \boldsymbol{\theta})\right)^2 dt + \frac{\left(x_0 - E\{x_0\}\right)^2}{2\sigma_0^2} \tag{3.24}$$

If the initial condition x_0 is perfectly known, the last term in eq. (3.24) vanishes. If we assume that $x(t)$ can be approximated by B-spline curves so that $x(t) \cong \underset{\sim}{x}(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\beta}$, then minimizing (3.24) is equivalent to minimizing:

$$\sum_{j=1}^{n} (y(t_{mj}) - \underset{\sim}{x}(t_{mj}))^2 + \frac{\sigma_m^2}{Q}\int_{t_0}^{t_q} \left(\dot{\underset{\sim}{x}}(t) - f(\underset{\sim}{x}(t), u(t), \boldsymbol{\theta})\right)^2 dt \tag{3.25}$$

Eq. (3.25) is the same as the iPDA objective function for B-spline fitting, eq. (3.5), with $\lambda = \frac{\sigma_m^2}{Q}$.

As a result, we have shown that optimal B-spline coefficients, which result in $\underset{\sim}{x}$ approximating the true curve for x, are obtained using the iPDA weighting coefficient $\lambda_{opt} = \frac{\sigma_m^2}{Q}$.

If the integral in (3.25) is approximated by discrete sums:

$$\sum_{j=1}^{n} (y(t_{mj}) - \underset{\sim}{x}(t_{mj}))^2 + \frac{\sigma_m^2}{\sigma_p^2}\sum_{i=1}^{q} \left(\dot{\underset{\sim}{x}}(t_i) - f(\underset{\sim}{x}(t_i), u(t_i), \boldsymbol{\theta})\right)^2 + \frac{\sigma_m^2}{\sigma_0^2}\left(x_0 - E\{x_0\}\right)^2 \tag{3.26}$$

where $\sigma_p^2 = \frac{Q}{\Delta t}$ is the process noise variance, so that the optimal weighting factor using a discretized model and discrete process noise is

$$\lambda_{opt-discrete} = \frac{\sigma_m^2}{\sigma_p^2} \tag{3.27}$$

If the modeller has knowledge about σ_m^2 from replicate measurements of the outputs, and about σ_p^2 from uncertainties in input settings, and the types and magnitudes of anticipated disturbances,

then a reasonable value of λ can be selected. The information required for optimal selection of λ is analogous to the information required to tune a Kalman filter (Gagnon and MacGregor, 1991).

Traditional NLS parameter estimation corresponds to optimizing the iPDA objective function in (3.8) in the limiting case when $\lambda \to \infty$, because traditional NLS methods assume that $\sigma_p^2 = 0$. In this case, the B-spline curve tends to satisfy the differential equation model perfectly (assuming there are sufficient spline knots), and the SSE is minimized through optimal selection of the fundamental parameter estimates $\hat{\boldsymbol{\theta}}$ (which influences $x_{\sim}$) as the iterations proceed.

In the Appendix 3.6, we extend these results to the general multivariate case. To illustrate what happens to the iPDA objective function in a simple multivariate estimation problem, we consider the following nonlinear dynamic system with two measured outputs:

$$\begin{cases} \dot{x}_1(t) = f_1(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + \eta_1(t) \\ \dot{x}_2(t) = f_2(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + \eta_2(t) \\ y_1(t_{m1j}) = x_1(t_{m1j}) + \varepsilon_1(t_{mj}) \\ y_2(t_{m2j}) = x_2(t_{m2j}) + \varepsilon_2(t_{m2j}) \end{cases} \tag{3.28}$$

For this system, the matrices defined in eq. (3.36) of the appendix are

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \quad \mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} \sigma_{m1}^2 & 0 \\ 0 & \sigma_{m2}^2 \end{bmatrix}, \quad \text{and}$$

$\mathbf{Q} = \begin{bmatrix} Q_{p1} & 0 \\ 0 & Q_{p2} \end{bmatrix}$. The resulting iPDA objective function for optimal spline fitting (from eq. (3.47) in the appendix) is:

$$
\begin{aligned}
&\frac{1}{2\sigma_{m1}^2}\sum_{j=1}^{N_1}\left(y_1(t_{m1j})-x_{\sim 1}(t_{m1j})\right)^2+\\
&\frac{1}{2\sigma_{m2}^2}\sum_{j=1}^{N_2}\left(y_2(t_{m2j})-x_{\sim 2}(t_{m2j})\right)^2+\\
&\frac{1}{2Q_{p1}}\int\left(\dot{x}_{\sim 1}(t)-f_1(x_{\sim 1}(t),x_{\sim 2}(t),u_1(t),u_2(t),\boldsymbol{\theta})\right)^2dt+\\
&\frac{1}{2Q_{p2}}\int\left(\dot{x}_{\sim 2}(t)-f_2(x_{\sim 1}(t),x_{\sim 2}(t),u_1(t),u_2(t),\boldsymbol{\theta})\right)^2dt
\end{aligned}
\tag{3.29}
$$

where N_1 and N_2 are the number of available measurements for outputs y_1 and y_2, respectively. This multivariate iPDA objective function readily accommodates outputs that are measured using different sampling rates and measurements that are made using irregularly spaced sampling times. The two SSE terms that appear in this objective function are each weighted by the reciprocal of the variance for their respective measurements. The objective function also contains two PEN terms corresponding to the integral of the squared deviations in the two stochastic differential equations. Because the development in the appendix assumes that the noise variables $\eta_1(t)$ and $\eta_2(t)$ are independent, no cross-product term (involving both $\dot{x}_1$ and $\dot{x}_2$) appears in the objective function. Each of the PEN terms is weighted by reciprocal of its respective process noise variance, which becomes more readily apparent if the integrals are approximated by sums:

$$
\begin{aligned}
&\frac{1}{2\sigma_{m1}^2}\sum_{i=1}^{N_1}\left(y_1(t_{mi})-x_{\sim 1}(t_{mi})\right)^2+\\
&\frac{1}{2\sigma_{m2}^2}\sum_{i=1}^{N_2}\left(y_2(t_{mi})-x_{\sim 2}(t_{mi})\right)^2+\\
&\frac{1}{2\sigma_{p1}^2}\sum_{i=0}^{q}\left(\dot{x}_{\sim 1}(t_i)-f_1(x_{\sim 1}(t_i),x_{\sim 2}(t_i),u_1(t_i),u_2(t_i),\boldsymbol{\theta})\right)^2+\\
&\frac{1}{2\sigma_{p2}^2}\sum_{i=0}^{q}\left(\dot{x}_{\sim 2}(t_i)-f_2(x_{\sim 1}(t_i),x_{\sim 2}(t_i),u_1(t_i),u_2(t_i),\boldsymbol{\theta})\right)^2
\end{aligned}
\tag{3.30}
$$

where $\sigma_{p1}^2=\frac{Q_{p1}}{\Delta t}$ and $\sigma_{p2}^2=\frac{Q_{p2}}{\Delta t}$.

Since each SSE term in the objective function is inversely proportional to its corresponding measurement error variance and each PEN term is inversely proportional to its corresponding model disturbance variance, it is straightforward to write an appropriate iPDA objective function for any multivariate problem. The main difficulty lies in obtaining estimates for these variances (or their ratios) in cases where all that is available is a dynamic model and some data. We are hopeful that the literature on tuning of Kalman filters (Maybeck, 1979) will provide some insight into this problem.

Objective functions in (3.29) and (3.30) reveal how iPDA can be conducted when some of the states are unmeasured (Chapter 4). For example, if no measurements are available for output y_1 so that N_1=0, the first term in the objective functions disappears, and the spline curves $x_{1\sim}$ and $x_{1\sim}$ are fitted simultaneously using the remaining SSE term and the two PEN terms. Then, step 2 of the iPDA algorithm involves selecting $\boldsymbol{\theta}$ to minimize the sum of the two PEN terms (Chapter 4).

3.4 Case study

In this case study we test our results for optimal selection of weighting factors in the iPDA objective function using two examples: a linear SISO continuous stirred tank reactor (CSTR), and a nonlinear MIMO CSTR. Note that we have also demonstrated the applicability of iPDA to more complicated problems including a nonlinear CSTR in which unmeasured states and non-stationary stochastic disturbances are present (Varziri et al., 2008a; Chapter 4), and a nylon polymerization reactor model (Varziri et al., 2008b; Chapter 6). The algorithm has also been extended so that it can be applied to cases in which the process disturbance intensity, Q, is unknown (Varziri et al., 2008b, 2008c) (Chapter 5 and Chapter 6).

3.4.1 Linear SISO CSTR

First, we use a simple linearized continuous CSTR example from Poyton et al., (2006), with a stochastic term $\eta(t)$ added to the original model:

$$\begin{aligned} &\frac{dC_A^{'}(t)}{dt} = w_c C_A^{'}(t) + w_T T^{'}(t) + \eta(t) \\ &C_A^{'}(0) = 0 \\ &y(t_{mj}) = C_A^{'}(t_{mj}) + \varepsilon(t_{mj}) \end{aligned} \tag{3.31}$$

where $C_A^{'} = C_A - C_{As}$ and $T^{'} = T - T_s$ are the concentration of reactant A and the temperature, respectively, in deviation variables. The discrete process disturbance η(t) used in our simulations (see Figure 3.2) is a series of random step inputs whose duration ($\Delta t = 0.1\,\text{min}$) is very short compared to the simulation time (24 minutes) and the process time constant. For this disturbance sequence in Figure 3.2, $\sigma_p^2 = 2\times10^{-3}\ (\text{kmol/m}^3/\text{min})^2$. The measurement noise $\varepsilon(t_{mj})\quad j = 1..n$ is a white noise sequence with variance $\sigma_m^2 = 4\times10^{-4}\ (\text{kmol/m}^3)^2$.

The stochastic differential equation in eq. (3.31) was obtained by linearizing the following nonlinear stochastic differential equation:

$$\frac{dC_A}{dt} = \frac{F}{V}\left(C_{A0} - C_A\right) - k_{ref}\exp\left(\frac{-E}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right)C_A + \eta(t) \tag{3.32}$$

Thus, the coefficients w_C and w_T in (3.31) depend on the unknown parameters k_{ref} (a kinetic rate constant) and E/R (an activation energy parameter), as well as the steady-state operating conditions:

$$w_C = -\frac{F_s}{V} - k_{ref}\exp\left(-\frac{E}{R}\left(\frac{1}{T_s} - \frac{1}{T_{ref}}\right)\right),\ w_T = -k_{ref}\frac{E}{R}\frac{C_{As}}{T_s^2}\exp\left(-\frac{E}{R}\left(\frac{1}{T_s} - \frac{1}{T_{ref}}\right)\right).$$

We assume that $k_{ref} = 0.461 \text{ min}^{-1}$ and $E/R = 8330.1 \text{ K}$ are the true values of the parameters, that the feed rate is steady at $F_s = 0.05 \text{ m}^3 \text{ min}^{-1}$, the inlet reactant concentration C_{A0} is constant at 2.0 kmol m^{-3}, the reactor volume is V = 1.0 m^3 and the reference temperature is $T_{ref} = 350 \text{ K}$. When the reactor is operated at a constant temperature, $T_s = 332 \text{ K}$, the resulting (expected) steady-state concentration of reactant A is $C_{As} = 0.567 \text{ kmol m}^{-3}$. We assume that the initial concentration in the reactor is C_{As} and that this initial concentration is known to the modeller. We also assume that the temperature control system is very effective, so that step changes in the temperature set point result in nearly instantaneous step changes in the reactor temperature.

The parameters k_{ref} and E/R were estimated using the step change in temperature shown in Figure 3.3, which produces the concentration response shown in Figure 3.4. The true response of the concentration (obtained using the true parameter values and the true stochastic disturbance sequence η(t)) is shown as the dashed line in Figure 3.4. The noisy measurements (241 equally spaced concentration measurements, with ten measurements per minute) and the spline fit $C_{A\sim}(t)$ are also shown.

The objective function used to fit $C_{A\sim}(t)$ was:

$$\sum_{j=1}^{241}\left(C_A(t_{mj}) - C_{A\sim}(t_{mj})\right)^2 + \lambda \int_{t=0}^{t=24}\left(\frac{dC'_{A\sim}(t)}{dt} - w_C C'_{A\sim}(t) - w_T T'(t)\right)^2 dt \qquad (3.33)$$

where $C_{A\sim}(t)$ is the B-spline fit and the optimal value of λ is $\lambda_{opt} = \sigma_m^2 / Q = \sigma_m^2 / \sigma_p^2 \Delta t = 2.0 \text{ min}$ (from eq. (3.11)). In our objective function evaluations, we approximated the integral in eq. (3.33) by discrete sums with $\Delta t = 0.1 \text{ min}$, so that $\lambda_{opt-discrete} = \sigma_m^2 / \sigma_p^2 = 0.2$ min^2. In our simulations of the true response, we solved the

linearized stochastic differential equation model using a Runge-Kutta method (ode45 in Simulink with relative tolerance 1e-3 and automatic absolute tolerance). In our subsequent example, the true plant is a more realistic nonlinear CSTR.

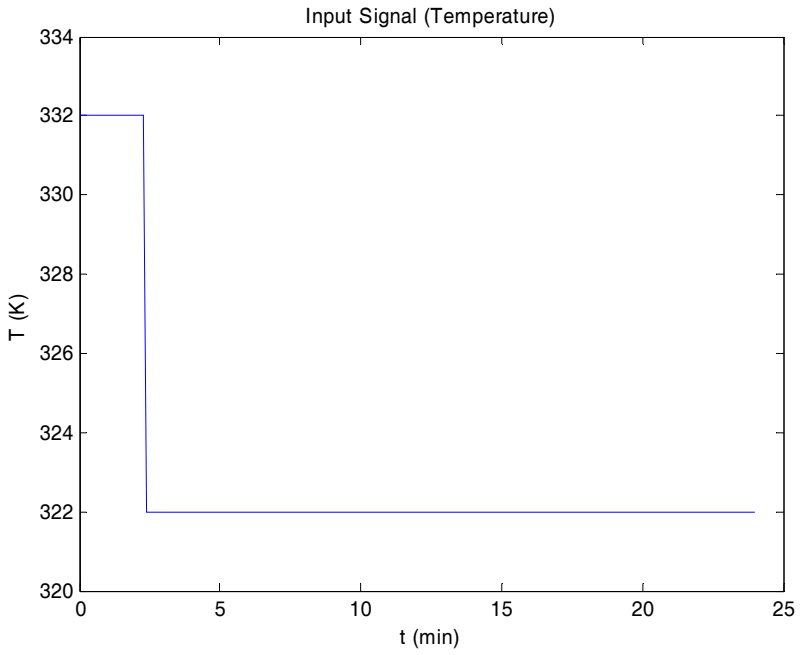

Figure 3.3. Input Signal

We obtained optimal parameter values in two different ways: i) using the iterative procedure shown in Figure 3.1 iPDA and ii) by simultaneously estimating the combined vector of two fundamental model parameters and B-spline coefficients $\boldsymbol{\tau} = \left[\boldsymbol{\theta}', \boldsymbol{\beta}'\right]'$ in eq. (3.33) using "lsqnonlin" routine in Matlab. The parameter estimates from the simultaneous approach were on average better than the iterated method and hence preferred. When fitting the B-splines we used one knot at each observation time. We found that due to the stochastic disturbance, using coincident knots at the time when the step change in T occurred did not improve the overall fit to the observed data. Another reason is that placing one knot at each observation point in this

example provides B-spline curves that are flexible enough to fit the observed data. The effect of coincident knots is more obvious when fewer knots are used.

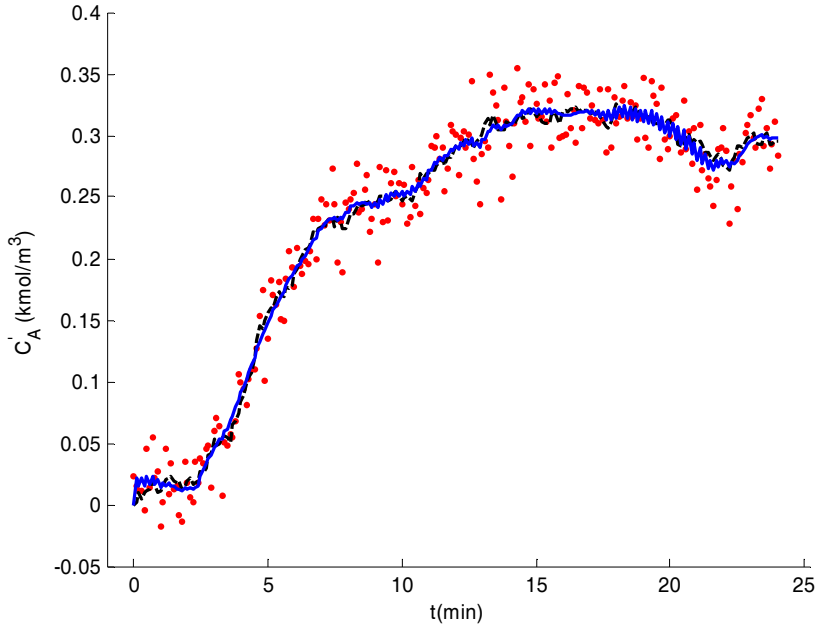

Figure 3.4. Measured, true, and fitted response for Linear SISO CSTR MODEL Obtained using iPDA with λopt-discrete =0.2 min^2 . (● , simulated data;-----, response of the system with true parameters and true stochastic noise; — , iPDA response)

Using the simulated data shown in Figure 3.4, model parameters were also estimated using traditional nonlinear least squares estimation (ode45 in Simulink was used to repeatedly solve the ODE model with η(t)=0 and the associated sensitivity equations). The fitted response is shown along with the data and the true process behaviour in Figure 3.5. Note that the fitted model response obtained from traditional NLS is much further from the true response curve than is the iPDA state response curve $C_{A\sim}$(t), which is shown in Figure 3.4. When iPDA is used for parameter estimation, two options are available for estimating (or smoothing) the state values.

One possibility, is to use the iPDA parameter estimates, $\hat{\boldsymbol{\theta}}$, in the model and solve the differential equations numerically, ignoring the stochastic disturbance. The other way is to use $x_{\sim}(t)$ as a state estimate (like a Kalman smoother) as is shown in Figure 3.4.

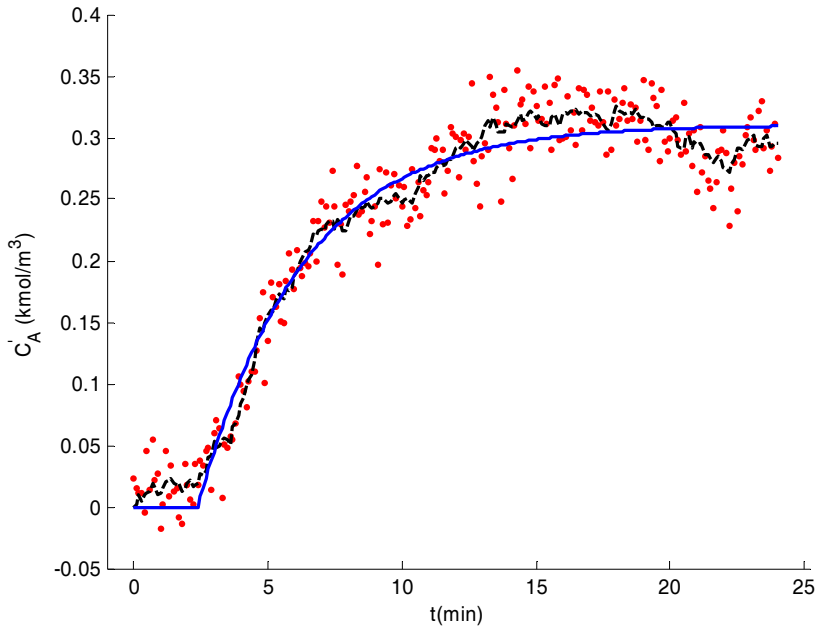

Figure 3.5. Observed, true, and predicted response for NLS, linear SISO CSTR (• , simulated data;-----, response of the system with true parameters and true stochastic noise; — , NLS response)

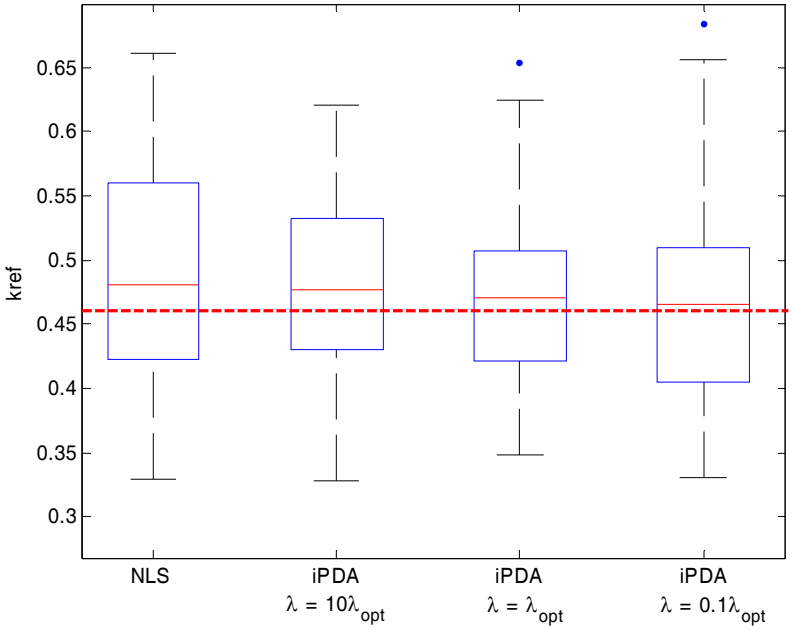

Figure 3.6. Box-plots for k_{ref} using NLS and iPDA linear SISO CSTR

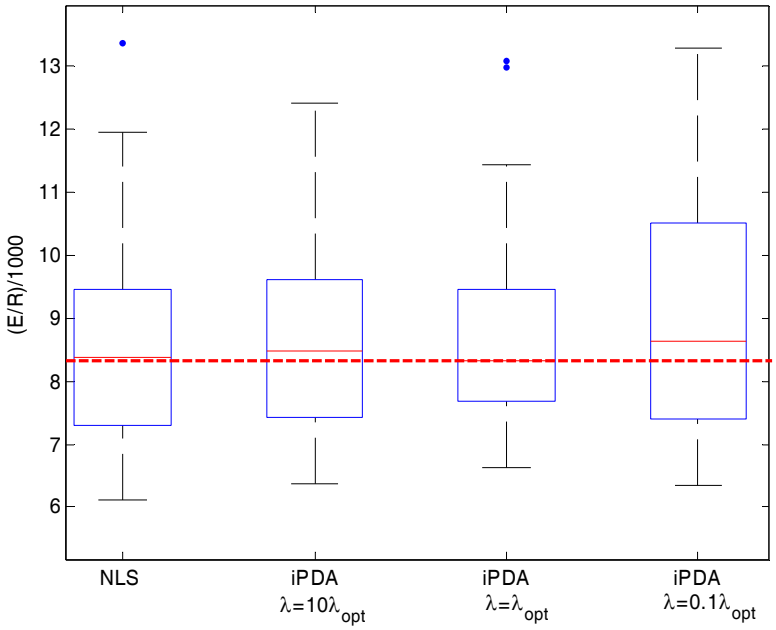

Figure 3.7. Box-plots for $\frac{E}{R}$ using NLS and iPDA, linear SISO CSTR

To compare the parameter estimates obtained using iPDA and traditional NLS, and to examine the effect of the iPDA weighting factor on the quality of the parameter estimates, parameters k_{ref} and E/R were estimated using iPDA with three different weighting factors ($0.1\lambda_{opt}$, λ_{opt}, $10\lambda_{opt}$) as well as traditional NLS, for 50 simulated data sets. The initial parameter guesses were 50% of the true parameter values. iPDA (NLS) iterations were stopped, when the change in $\|\boldsymbol{\tau}\|$ ($\|\boldsymbol{\theta}\|$ in case of NLS) was less than 1E-12, or if a maximum of 1000 iterations was reached. Figure 3.6 and Figure 3.7 show the corresponding box plots for the estimates of k_{ref} and E/R, respectively. Both iPDA and traditional NLS produced good and comparable estimates. $iPDA_{opt}$ (iPDA with λ_{opt}) gave the most precise estimates.

When $\lambda = 10\lambda_{opt}$ was used, the iPDA parameter estimates were closer to those obtained traditional NLS than the estimates using $\lambda = \lambda_{opt}$. This result was expected, because traditional NLS assumes that $\eta(t)$ can be neglected because $\sigma_p^2 \to 0$, which corresponds to very large λ. When we attempted iPDA estimation with very large weighting factors (*e.g.*, $\lambda_{opt-discrete} = 100$ min^2) the iterations stopped before reaching the NLS parameter estimates. Unfortunately, the spline knot sequence that we specified was too coarse for the empirical B-spline curve to be able to solve the ODE arbitrarily well. As a result, the PEN term $\lambda \int \left(\frac{dC'_{A\sim}(t)}{dt} - w_C C'_{A\sim}(t) - w_T T'(t) \right)^2 dt$ in the iPDA objective function was not able to approach zero, and remained large, relative to the SSE term. We anticipate that a large number of spline knots (and perhaps long computational times) will be required to obtain accurate iPDA parameter estimates using very large values of λ. This is not a serious problem for modellers,

because in situations where a large λ is appropriate, traditional NLS or collocation-based methods provide good parameter estimates, and iPDA would not be particularly beneficial. We advocate that iPDA be used in situations where there are significant process disturbances, uncertainties in input variables, or an imperfect or simplified ODE model, so that traditional least-squares assumptions do not apply. In such situations, very large values of λ are not required. To confirm that we had used a sufficient number of spline knots to obtain good iPDA parameter estimates for our stochastic CSTR problem, we re-estimated the parameters using twice as many spline knots and obtained almost exactly the same results.

To confirm that iPDA becomes even more beneficial when there are significant process disturbances, we generated additional simulated data sets using the same measurement noise variance as in Figure 3.4, but ten times the amount of process noise variance (making the optimal choice of λ ten times smaller). As expected, because of the larger process noise, both iPDA and traditional NLS parameter estimates worsened, but the traditional NLS parameter estimates deteriorated more than iPDA parameter estimates.

3.4.2 Nonlinear MIMO CSTR

The second case study is a non-isothermal CSTR. The model equations consist of material and energy balances (Marlin, 2000) with additional stochastic terms:

$$\begin{aligned}
&\frac{dC_A(t)}{dt} = \frac{F(t)}{V}\left(C_{A0}(t) - C_A(t)\right) - g(T(t))C_A(t) + \eta_1(t) \\
&\frac{dT(t)}{dt} = \frac{F(t)}{V}\left(T_0(t) - T(t)\right) + \beta_1\left(T(t) - T_{cin}(t)\right) + \beta_2 g(T(t))C_A(t) + \eta_2(t) \qquad (3.34) \\
&C_A(0) = 1.569\,(\text{kmol m}^{-3}) \\
&T(0) = 341.37\,(\text{K})
\end{aligned}$$

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

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

$E\{\eta_1(t_i)\eta_1(t_j)\} = \sigma_{p1}^2 \delta(t_i - t_j)$, $E\{\eta_2(t_i)\eta_2(t_j)\} = \sigma_{p2}^2 \delta(t_i - t_j)$, $\varepsilon_1(t_{mj})$ $j = 1..N_1$ 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.

As in the previous example, C_A is the concentration of the reactant A, T is the reactor temperature, V is the volume and T_{ref}= 350 K is the reference temperature. The parameters to be estimated and their true values are the same as those of the SISO case study: E/R= 8330.1 K, k_{ref} = 0.461 min^{-1}. The initial parameter guesses were set at 50% of the true parameter values. 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, and the coolant flow rate F_c. Parameters a=1.678E6, $b = 0.5$ which we assume to be known from previous heat-transfer experiments, account for the effect of F_c on the heat transfer coefficient. Values for the various other known constants (Marlin, 2000) are as follows: $V = 1.0$ m^3, $C_p = 1$ cal g^{-1}K^{-1}, $\rho =$ 1E6 gm^{-3}, $C_{pc} = 1$ cal g^{-1}K^{-1}, $\rho_c =$ 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. The linear SISO case study used a simplified form of this model that assumed perfect temperature control.

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

We assume that concentration can be measured once per minute and temperature can be measured once every 0.3 minutes. The duration of the simulation is 64 minutes, so that there are 64 concentration measurements and 213 temperature measurements. The noise variances 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 are $Q_{p1} = 4\times10^{-3}\ (\text{kmol/m}^3)^2/\text{min}$ and $Q_{p2} = 4\ \text{K}^2/\text{min}$. Since $\sigma_{p1}^2 = \frac{Q_{p1}}{\Delta t}$ and $\sigma_{p2}^2 = \frac{Q_{p2}}{\Delta t}$, for $\Delta t = 1\,\text{min}$ the corresponding process noise variances are $\sigma_{p1}^2 = 4\times10^{-3}\ (\text{kmol/m}^3/\text{min})^2$ and $\sigma_{p2}^2 = 4\ (\text{K/min})^2$.

From eq. (3.29), the iPDA objective function is:

$$
\begin{aligned}
&\sum_{j=1}^{64}\left(C_A(t_{m1j}) - C_{A\sim}(t_{m1j})\right)^2 + \\
&\lambda_1 \int_{t=0}^{64}\left(\frac{dC_{A\sim}(t)}{dt} - \frac{F(t)}{V}\left(C_{A0}(t) - C_{A\sim}(t)\right) + g(T_\sim(t))C_{A\sim}(t)\right)^2 dt + \\
&\sum_{j=1}^{213}\left(T(t_{m2j}) - T_\sim(t_{m2j})\right)^2 + \\
&\lambda_2 \int_{t=0}^{64}\left(\frac{dT_\sim(t)}{dt} - \frac{F(t)}{V}\left(T_0(t) - T_\sim(t)\right) - \beta_1\left(T_\sim(t) - T_{cin}(t)\right) + \beta_2 g(T_\sim(t))C_{A\sim}(t)\right)^2 dt
\end{aligned}
\tag{3.35}
$$

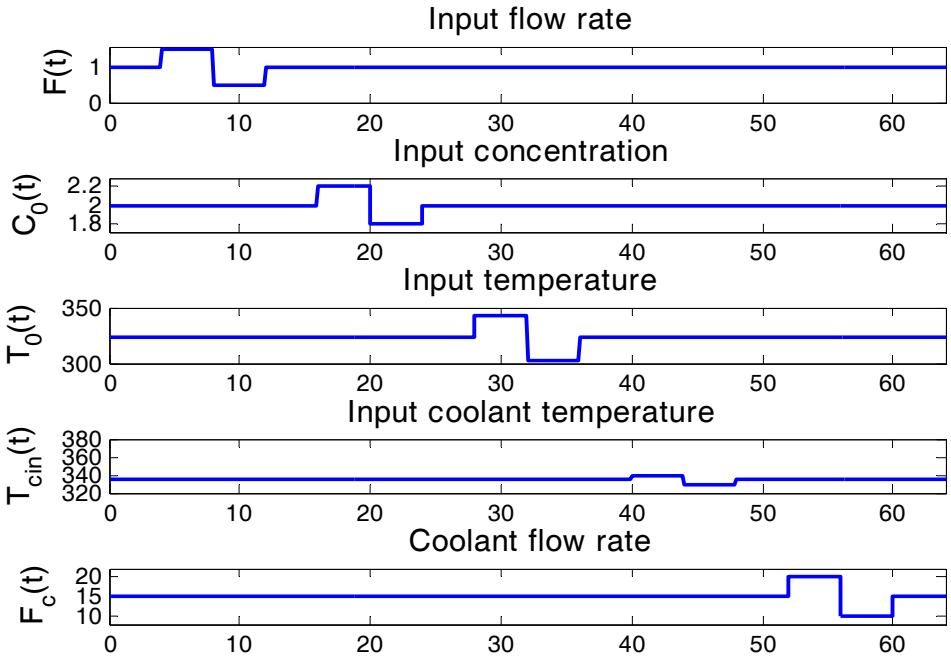

Figure 3.8. Input scheme for MIMO nonlinear CSTR

From (3.11), the optimal weighting factors in this case are $\lambda_1 = 0.1\,\text{min}$ and $\lambda_2 = 0.16$ min. We used two different knot sequences to fit B-spline curves to concentration and temperature observations, with the knots placed at observation times. Again, we obtained optimal parameter values using the iterative procedure shown in Figure 3.1 iPDA and also by simultaneously estimating the B-spline coefficients along with the two parameters in eq. (3.35) using "lsqnonlin" routine in Matlab. However, in this example the simultaneous approach proved to be much slower (due to the existence of two states and considerably fewer concentration observations), so that the iterative approach was preferred. To compare the sampling behaviour of the iPDA and NLS parameter estimates, 50 sets of concentration and temperature measurements were generated using different measurement and process noise sequences. Gaussian quadrature was used to calculate the integrals in eq. (3.33). Box-plots for the parameter estimates are shown in Figure

3.9 and Figure 3.10. Both iPDA and traditional NLS produced reasonable estimates; the iPDA parameter estimates are better, on average, than the traditional NLS estimates. The iPDA and NLS predicted responses are compared against the true responses in Figure 3.11, Figure 3.12, Figure 3.13, andFigure 3.14. The iPDA predicted responses closely follow the true trajectories.

3.5 Summary and Conclusions

Parameters were estimated in differential equation models with stochastic disturbances using iPDA. By considering the joint probability distribution of the states and observations, given the parameters, we demonstrated that optimal model parameters and B-spline coefficients can be obtained by minimizing the iPDA objective function. We also demonstrated that the optimal value of the weighting factor λ is proportional to the measurement noise variance, and inversely proportional to the model disturbance variance, so that tuning the weighting factor in iPDA resembles tuning the Kalman gain in Kalman filtering applications. For parameter estimation in MIMO dynamic models, the overall iPDA objective function includes a sum-of-squared errors term for each response, with the reciprocal of the corresponding measurement variance as a weighting factor, and a model-based penalty term for each differential equation, with the reciprocal of the process noise variance as a weighting factor. Optimizing the iPDA objective function, using either a simple iterative two-step procedure (in which spline coefficients are estimated in the first step and fundamental parameters are estimated in the second) or simultaneous estimation of spline coefficients and model parameters, produces maximum likelihood estimates for the parameters, conditional on the data and knowledge about the measurement and disturbance variances.

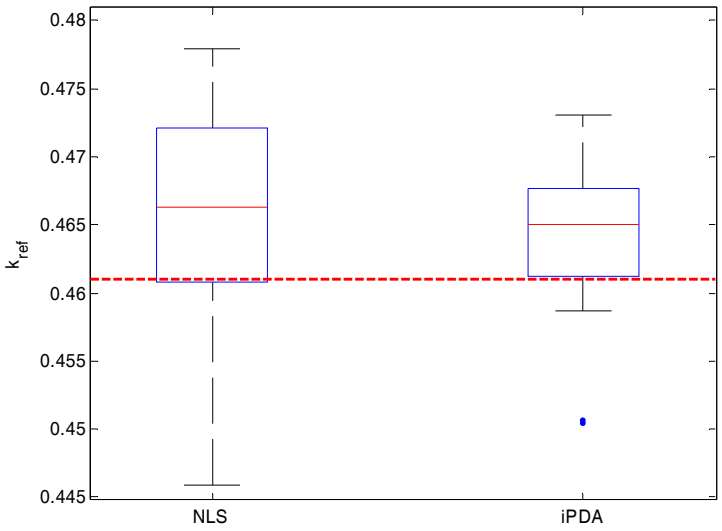

Figure 3.9. Box-plots for k_{ref} using NLS and iPDA nonlinear MIMO CSTR

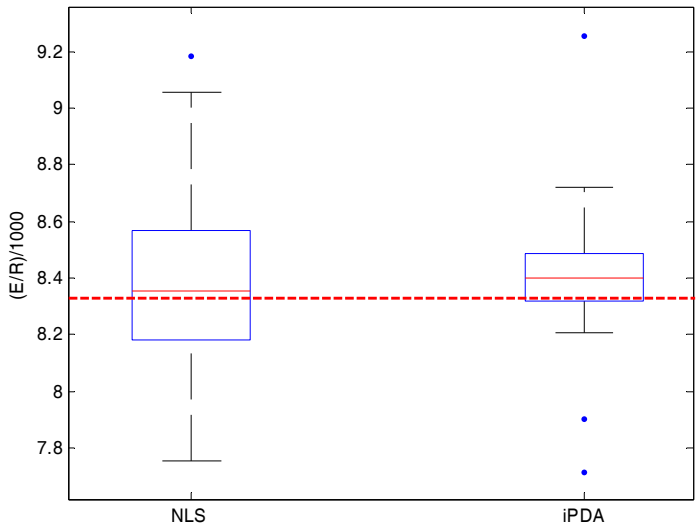

Figure 3.10. Box-plots for $\frac{E}{R}$ using NLS and iPDA nonlinear MIMO CSTR

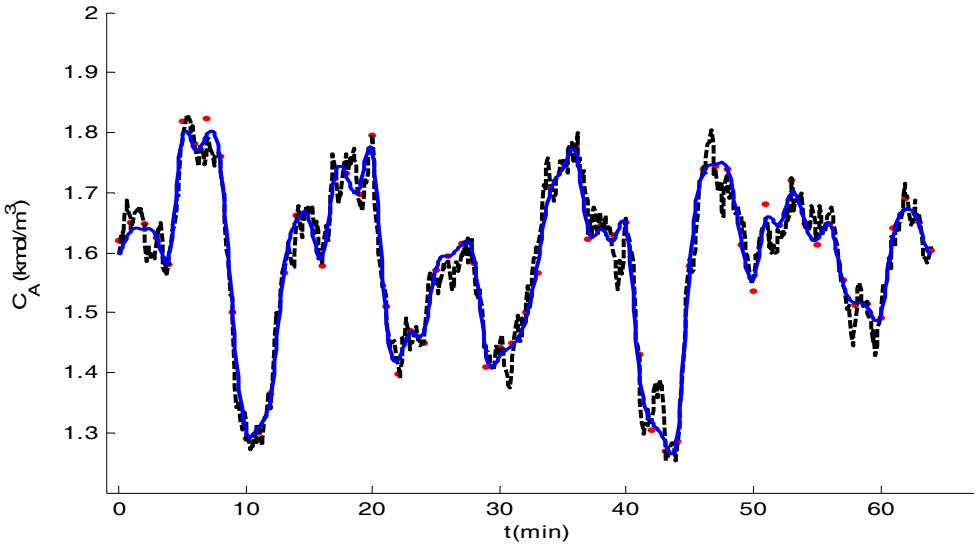

Figure 3.11. Measured, true, and predicted concentration response for iPDA for the nonlinear MIMO CSTR Example (• , simulated data;-----, response of the system with true parameters and true stochastic noise; — , iPDA response)

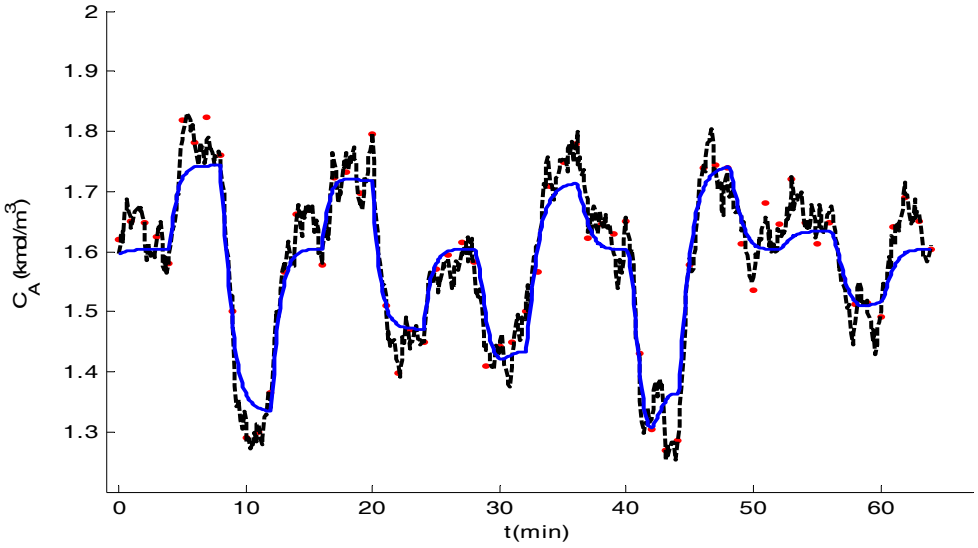

Figure 3.12. Measured, true, and predicted concentration responses using NLS for the nonlinear MIMO CSTR Example (• , simulated data;-----, response of the system with true parameters and true stochastic noise; — , NLS response)

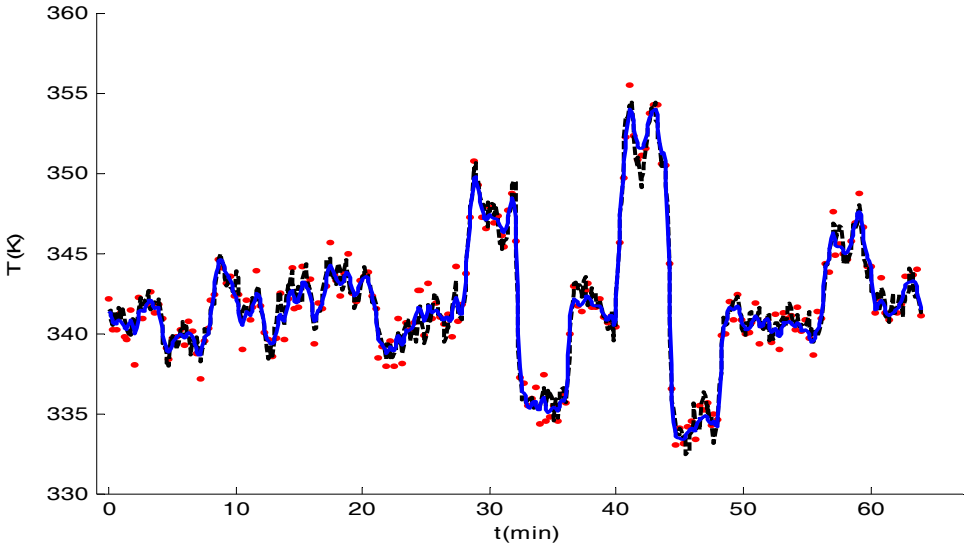

Figure 3.13. Measured, true, and predicted temperature response for iPDA for the nonlinear MIMO CSTR Example (• , simulated data;-----, response of the system with true parameters and true stochastic noise; — , iPDA response)

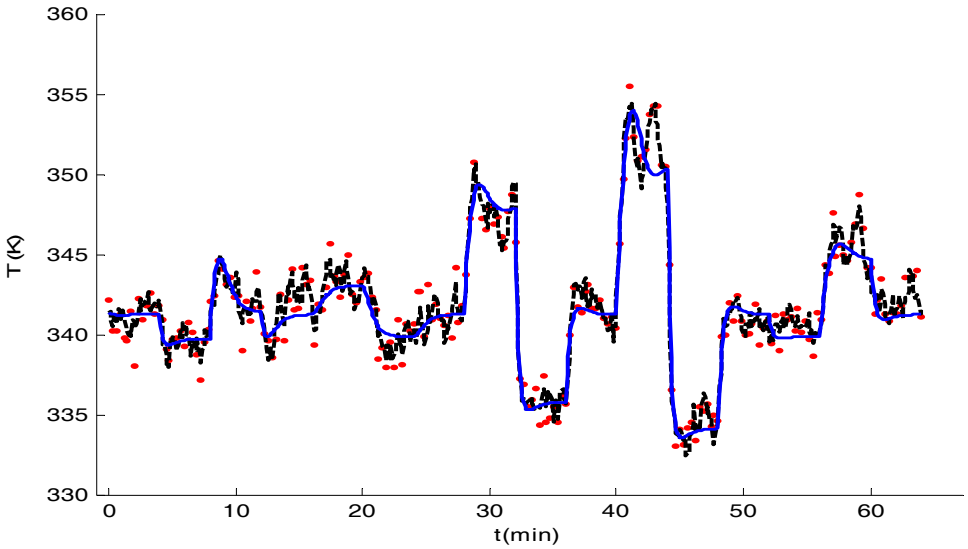

Figure 3.14. Measured, true, and predicted temperature response for NLS for the nonlinear MIMO CSTR Example (• , simulated data;-----, response of the system with true parameters and true stochastic noise; — , NLS response)

Two examples were used to validate the results. In the first, parameters were estimated in a linearized SISO differential equation model with a stochastic disturbance using both traditional NLS and iPDA. iPDA parameter estimates were superior to those obtained using traditional NLS, and iPDA was effective in reconstructing the true underlying response trajectory. When iPDA was used with λ larger than the optimal value, parameter estimates were similar to those obtained using traditional NLS because large values of λ are appropriate when the traditional NLS assumptions (*i.e.*, negligible stochastic disturbance compared to the measurement noise) hold. In the second case study, two parameters in a nonlinear MIMO CSTR with a stochastic disturbance were estimated. Similar to the first case study, iPDA performed better than traditional NLS in estimating parameters and reconstructing the response trajectory.

From the case studies we see that iPDA (like a Kalman filter) can be used as a state smoother. Recent simulation studies (Chapter 4) have confirmed that iPDA can also be used to observe unmeasured states and to estimate parameters when some of the states are unmeasured.

In our case studies, the measurement noise variance and the model disturbance variance were assumed known. In practical situations however, this is not the case. Although knowledge about measurement variances can be obtained from replicate measurements, it is difficult to obtain *a priori* knowledge about the model disturbance variance. One objective of our ongoing work is to establish a means of estimating the weighting factor without the limiting assumptions for the model parameters. This issue is discussed in Chapter 5.

In iPDA the objective function is minimized by an iterative approach but other approaches for minimizing the same objective function are also possible. In particular, nonlinear and quadratic programming techniques that can handle large-scale minimization problems are potential candidates. These methods have been applied to problems in which orthogonal collocation on

finite elements is used to discretize the ODEs (Tjoa and Biegler, 1991; Tanartkit and Biegler, 1995; Biegler and Grossman, 2004). In the remaining chapters, we show that efficient nonlinear programming tools can be used effectively to reduce the computational requirements of iPDA.

3.6 Appendix

In this section, the results of Section 3.3 are repeated for a multivariate case.

We consider a Multi-Input Multi-Output first order, nonlinear system defined as follows:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\eta}(t) \\ \mathbf{x}(t_0) &= \mathbf{x}_0 \\ \\ \mathbf{y}(t_j) &= \mathbf{C}\mathbf{x}(t_j) + \boldsymbol{\varepsilon}(t_j) \end{aligned} \tag{3.36}$$

where:

$\mathbf{x}$ is the vector of n state variables, $\mathbf{F}$ (nonlinear function) is an $n \times 1$ vector, $\mathbf{u}$ is the vector of u input variables, $\mathbf{y}$ is the vector of m output variables, and $\mathbf{C}$ is a $m \times n$ matrix. $\mathbf{x}_0$ is a multivariate normal random variable with mean $E\{\mathbf{x}_0\}$ and $\text{cov}(\mathbf{x}_0) = \boldsymbol{\Sigma}_0$.

$\boldsymbol{\eta}(t)$ is a continuous zero-mean stationary white noise process with covariance matrix $E\{\boldsymbol{\eta}(t_1)\boldsymbol{\eta}(t_2)^T\} = \mathbf{Q}\delta(t_2 - t_1)$, where $\mathbf{Q}$ is the corresponding power spectral density and $\delta(.)$ is the Dirac delta function. For the discrete time white noise process:

$$E\{\boldsymbol{\eta}(j_1\Delta t)\boldsymbol{\eta}^T(j_2\Delta t)\} = \begin{cases} \dfrac{\mathbf{Q}}{\Delta t} & j_1 = j_2 \\ \\ 0 & j_1 \neq j_2 \end{cases} \tag{3.37}$$

where j_1 and j_2 are integers and Δt is the sampling period. $\boldsymbol{\varepsilon}(t_j) = [\varepsilon_{1j}, ..., \varepsilon_{mj}]^T$ is a vector of m zero-mean random variables. We assume that measurements of different responses from the same experimental run are independent.

$$E\left\{\boldsymbol{\varepsilon}(t_i)\boldsymbol{\varepsilon}^T(t_j)\right\} = \begin{cases} \mathbf{E} & i = j \\ 0 & i \neq j \end{cases} \tag{3.38}$$

$\mathbf{E}$ is diagonal with $\text{diag}(\mathbf{E}) = [\sigma_{1j}^2, ..., \sigma_{mj}^2]^T$ for the jth run. We also assume that measurements from different experimental runs are independent. Whenever the above assumptions are not appropriate, they can be made more general. However, the results in that case maybe more complex depending on the covariance matrix structure.

We assume that the response of the above stochastic system can be approximated by a linear combination of some B-splines:

$$x_{i\sim}(t) = \varphi_i^T(t)\boldsymbol{\beta}_i \text{ for } i = 1...n \tag{3.39}$$

where $\varphi_i(t)$ is a vector containing c_i basis functions, $\boldsymbol{\beta}_i$ is vector of c_i spline coefficients and $\omega_i(t)$ is the stochastic term. Therefore

$$\mathbf{x}_{\sim}(t) = \begin{bmatrix} \varphi_1^T(t) & 0 & \cdots & 0 \\ 0 & \varphi_2^T(t) & \cdots & 0 \\ 0 & 0 & \cdots & \varphi_n^T(t) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_n \end{bmatrix} = \boldsymbol{\Phi}(t)\boldsymbol{\beta}_c \tag{3.40}$$

where $\boldsymbol{\Phi}(t) = \begin{bmatrix} \varphi_1^T(t) & 0 & \cdots & 0 \\ 0 & \varphi_2^T(t) & \cdots & 0 \\ 0 & 0 & \cdots & \varphi_n^T(t) \end{bmatrix}$ and $\boldsymbol{\beta}_c = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_n \end{bmatrix}$ is the concatenated vector of spline coefficients.

Again we consider eq. (3.13), from the same arguments that were used for the SISO case, we will get the same expression for $p(\mathbf{x}_0, \ldots, \mathbf{x}_q \mid \boldsymbol{\theta})$ as before:

$$
\begin{aligned}
&p(\mathbf{x}_0,\ldots,\mathbf{x}_q \mid \boldsymbol{\theta}) \propto \\
&\exp\left\{-\int \left(\dot{\mathbf{x}}(t)-\mathbf{F}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})\right)^T \mathbf{Q}^{-1}\left(\dot{\mathbf{x}}(t)-\mathbf{F}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})\right)dt\right\}\times \\
&\exp\left\{-\left(\mathbf{x}_0-E\{\mathbf{x}_0 \mid \boldsymbol{\theta}\}\right)^T \boldsymbol{\Sigma}_0^{-1}\left(\mathbf{x}_0-E\{\mathbf{x}_0 \mid \boldsymbol{\theta}\}\right)\right\}
\end{aligned} \tag{3.41}
$$

Suppose the ith response ($i = 1\cdots m$)is measured N_i times observations. Let $\mathbf{y}_c = [y_{11}\cdots y_{1N_1}\cdots y_{m1}\cdots y_{mN_m}]^T$ and $\boldsymbol{\varepsilon}_c = [\varepsilon_{11}\cdots\varepsilon_{1N_1}\cdots\varepsilon_{m1}\cdots\varepsilon_{mN_m}]^T$ is zero mean with $\operatorname{cov}(\boldsymbol{\varepsilon}_c) = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2\mathbf{I}_{N_1} & \cdots & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \cdots & \sigma_m^2\mathbf{I}_{N_m} \end{bmatrix}$. Considering $\mathbf{x}_c = [x_{11}\cdots x_{1N_1}\cdots x_{m1}\cdots x_{mN_m}]^T$

$$
\mathbf{y}_c = \mathbf{C}_c\mathbf{x}_c + \boldsymbol{\varepsilon}_c \tag{3.42}
$$

$$
E\{\mathbf{y}_c \mid \mathbf{x}_c,\boldsymbol{\theta}\} = \mathbf{x}_c \quad \text{and } \operatorname{cov}\{\mathbf{y}_c \mid \mathbf{x}_c,\boldsymbol{\theta}\} = \boldsymbol{\Sigma} \tag{3.43}
$$

where $\mathbf{C}_c = \begin{bmatrix} c_{11}\mathbf{I}_{N_1} & \cdots & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \cdots & c_{mm}\mathbf{I}_{N_m} \end{bmatrix}$. From (3.43) :

$$
p(\mathbf{y}_c \mid \mathbf{x}_0,\ldots,\mathbf{x}_q,\boldsymbol{\theta}) \propto \exp\left\{-(\mathbf{y}_c-\mathbf{C}_c\mathbf{x}_c)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}_c-\mathbf{C}_c\mathbf{x}_c)\right\} \tag{3.44}
$$

From (3.13), (3.41), and (3.44):

$$
\begin{aligned}
&p(\mathbf{x}_0,\ldots,\mathbf{x}_q,\mathbf{y}_c \mid \boldsymbol{\theta}) = \\
&\exp\left\{-(\mathbf{y}_c-\mathbf{C}_c\mathbf{x}_c)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}_c-\mathbf{C}_c\mathbf{x}_c)\right\}\times \\
&\exp\left\{-\int \left(\dot{\mathbf{x}}(t)-\mathbf{F}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})\right)^T \mathbf{Q}^{-1}\left(\dot{\mathbf{x}}(t)-\mathbf{F}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})\right)dt\right\}\times \\
&\exp\left\{-\left(\mathbf{x}_0-E\{\mathbf{x}_0 \mid \boldsymbol{\theta}\}\right)^T \boldsymbol{\Sigma}_0^{-1}\left(\mathbf{x}_0-E\{\mathbf{x}_0 \mid \boldsymbol{\theta}\}\right)\right\}
\end{aligned} \tag{3.45}
$$

Therefore the optimal state and parameter estimates, $\hat{\mathbf{x}},\hat{\boldsymbol{\theta}}$ minimize

$$
\begin{aligned}
&(\mathbf{y}_c-\mathbf{C}_c\mathbf{x}_c)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}_c-\mathbf{C}_c\mathbf{x}_c) + \\
&\int \left(\dot{\mathbf{x}}(t)-\mathbf{F}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})\right)^T \mathbf{Q}^{-1}\left(\dot{\mathbf{x}}(t)-\mathbf{F}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})\right)dt + \\
&\left(\mathbf{x}_0-E\{\mathbf{x}_0 \mid \boldsymbol{\theta}\}\right)^T \boldsymbol{\Sigma}_0^{-1}\left(\mathbf{x}_0-E\{\mathbf{x}_0 \mid \boldsymbol{\theta}\}\right)
\end{aligned} \tag{3.46}
$$

If we assume that the initial condition is perfectly known and that the states can be approximated by eq. (3.40) we have:

$$\begin{aligned}&(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_{\sim c})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_{\sim c}) + \\ &\int \left(\dot{\mathbf{x}}_{\sim}(t) - \mathbf{F}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})\right)^T \mathbf{Q}^{-1}\left(\dot{\mathbf{x}}_{\sim}(t) - \mathbf{F}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})\right) dt\end{aligned} \tag{3.47}$$

If the integral in (3.47) is discretized:

$$\begin{aligned}&(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_{\sim c})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y}_c - \mathbf{C}_c\mathbf{x}_{\sim c}) + \\ &\sum_{i=0}^{q} \left(\dot{\mathbf{x}}_{\sim}(t_i) - \mathbf{F}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})\right)^T \boldsymbol{\Sigma}_p^{-1}\left(\dot{\mathbf{x}}_{\sim}(t_i) - \mathbf{F}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})\right)\end{aligned} \tag{3.48}$$

Where $\boldsymbol{\Sigma}_p = \dfrac{\mathbf{Q}}{\Delta t}$.

3.7 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 $g^{-1}K^{-1}$
$\mathbf{C}$	Output matriz	
$E\{.\}$	Expected value	
$\mathbf{E}$	Covariance matrix of the measurement noise	
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	
$\mathbf{F}$	Vector of all $\mathbf{f}_i$s	
k_{ref}	Kinetic rate constant at temperature *Tref*	min^{-1}
n	Number of measurements	
$p(.)$	Probability density function	
q	Number of discretization points for differential equations	

Q_{pi}	Process noise intensity of stochastic differential equation for state *i*	
$\mathbf{Q}$	Matrix of process noise intensities	
t_{mj}	*j*-th measurement time	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
x,	State variables	
$x_{\sim}$	B-spline approximation of the state	
$x_{i\sim}$	B-spline approximation of the *i*-th state	
$\mathbf{x}$	Vector of state variables	
$\mathbf{x}_c$	Concatenated vector of state variables corresponding to $\mathbf{y}_c$	
y, $\mathbf{y}$	Noisy output measurements	
$\mathbf{y}_m$	Stacked vector of measured outputs for one state	
$\mathbf{y}_c$	Concatenated vector of measures outputs for all states	
β_i	*i*-th B-spline coefficient	
$\boldsymbol{\beta}$	Vector of B-spline coefficients	

$\boldsymbol{\beta}_i$	Vector of B-spline coefficients of the *i*-th state	
$\boldsymbol{\beta}_c$	Concatenated vector of B-spline coefficients for all states	
$\delta(.)$	Dirac delta function	
ΔH_{rxn}	Enthalpy of reaction	cal g^{-1} K^{-1}
ε_i	Normally distributed measurement noise for state *i*	
$\boldsymbol{\varepsilon}_m$	Vector of measurement noise for one state	
$\boldsymbol{\varepsilon}_c$	Concatenated vector of measurement noise for all states	
η	White Gaussian process disturbance entering model differential equation	
$\boldsymbol{\eta}$	Vector of White Gaussian process disturbances	
$\boldsymbol{\theta}$	Vector of model parameters	
$\hat{\boldsymbol{\theta}}$	Vector of parameter estimates	
λ	Weighting factor in iPDA objective function	
λ_{opt}	Optimal weighting factor in iPDA objective function	
ρ	Density of reactor contents	g m^{-3}
ρ_c	Coolant density	g m^{-3}
σ_{mi}^2	Measurement noise variance for the *i*-th state	
σ_{pi}^2	Process noise variance of the discrete-time process for the *i*-th state	
$\boldsymbol{\Sigma}$	Measurement noise covariance matrix for $\boldsymbol{\varepsilon}_c$	
$\boldsymbol{\tau}$	Combined vector of model parameters and spline coefficients	
ϕ_i	*i*-th B-spline basis function	
φ	Vector of B-spline basis functions	

φ_i	Vector of B-spline basis functions for the *i*-th state
$\mathbf{\Phi}$	Matrix containing all φ_i

CSTR	Continuous Stirred Tank Reactor
diag	Diagonal elements of matrix
iPDA	iteratively-refined Principal Differential Analysis
MIMO	Multi-Input Multi-Output
ODE	Ordinary Differential Equation
PEN	Model-based penalty
SISO	Single-Input Single-Output
SSE	Sum of Squared Errors
NLS	Nonlinear Least-Squares

3.8 References

U.M. Ascher, R.M.M. Mathheij, R.D. Russell. *Numerical Solutions of Boundary Value Problems for Ordinary Differential Equations.* Prentice Hall, New Jersey, 1988.

K.J. Astrom. *Introduction to Stochastic Control Theory*. Academic Press, Inc., New York, 1970.

Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, Inc., New York, 1974.

D.M. Bates, D.G. Watts. *Nonlinear Regression Analysis and Its applications.* John Wiley & Sons, Inc., New York, 1988.

M. Benson. Parameter fitting in dynamic models, *Ecological Modelling.* 6, 1979, 97-115.

L.T. Biegler, I.E. Grossman. Retrospective on optimization. *Computers and Chemical Engineering*. 28, 2004, 1169-1192.

L.T. Biegler. Short note solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering.* 8, 1984, 243-248.

T. Binder, L. Blank, W. Dahmen, W. Marquardt. On the regularization of dynamic data reconciliation problems. *Journal of Process Control*. 12, 2002, 557-567.

H.G. Bock. Numerical treatment of inverse problems in chemical reaction kinetics, in *Modelling of Chemical Reaction Systems*. K.H. Ebert, P. Deuflhard, and W. Jäger, editors. Springer, Berlin. Chapter 8, 1981, 102-125.

H.G. Bock. Recent advances in parameter identification for ordinary differential equations. In *Progress in Scientific Computing*. P. Deuflhard and E. Hairer, editors. Birkhäuser, Boston. 1983, 95-121.

C. de Boor. *A practical Guide to Splines.* Springer, New York, 2001.

R.G. Brown, P.Y.C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering* 2nd edition. John Wiley & Sons, Inc., 1992.

P.H.C. Eilers, B.D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*. 11 (2), 1996, 89-121.

L. Gagnon, J.F., MacGregor. State estimation for continuous emulsion polymerization. *Canadian Journal of Chemical Engineering.* 69, 1991, 648-656.

J.Gong, G. Wahba, D.R. Johnson, J. Tribbia. Adaptive tuning of numerical weather prediction models: simultaneous estimation of weighting, smoothing, and physical parameters. *Monthly Weather Review*. 126, 1998, 210-231.

A.H. Jazwinski. *Stochastic Processes and Filtering theory*. Academic Press, New York, 1970.

A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Springer, 1996.

J.R. Leis, M.A. Kramer. ALGORITHM 658: ODESSA – an ordinary differential equation solver with explicit simultaneous sensitivity analysis. *ACM Transactions on Mathematical Software.* 14, 1988, 61-67.

Z. Li, M. R. Osborne, T. Pravan. Parameter estimation for ordinary differential equations. *IMA J. of Num. Science*. 25, 2005, 264-285.

J.S. Logsdon, L.T. Biegler. Decomposition strategies for large-scale dynamic optimization problems. *Chemical Engineering Science.* 47, 1992, 851-864.

J.C. Lucero. Identifying a differential equation for lip motion. *Medical Engineering and Physics*. 24, 2002, 521-528.

T.E. Marlin. *Process Control: Designing Processes and Control Systems for Dynamic Performance, 2nd Edition*. McGraw-Hill, 2000.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 1*. Academic Press, New York, 1979.

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

F. O'Sullivan. A statistical perspective on ill-posed inverse problems. *Statistical Science*. 1 (4), 1986, 502-518.

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

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. *Computers and Chemical Engineering.* 30, 2006, 698-708.

J.O. Ramsay. Principal Differential Analysis: Data reduction by differential operators. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*. 58, 1996, 495-508.

J. O. Ramsay. Functional components of variation in handwriting. *Journal of the American Statistical Association.* 95, 2000, 9-16.

J.O. Ramsay, K. G. Munhall. Functional data analyses of lip motion. *Journal of the Acoustical Society of America*. 99, 1996, 3718-3727.

J. O. Ramsay, J. B. Ramsey. Functional data analysis of the dynamics of the monthly index of nondurable goods production. *Journal of Econometrics*. 107, 2002, 327-344.

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

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

J. Swartz and H. Bremermann. Discussion of parameter estimation in biological modelling: algorithms for estimation and evaluation of estimates. *Journal of Mathematical Biology.* 1, 1975, 241-275.

P. Tanartkit, L. T. Biegler. Stable decomposition for dynamic optimization. *Industrial Engineering and Chemistry Research.* 34, 1995, 1253-1266.

Y.P. Tang. On the estimation of rate constants for complex kinetic models. *Industrial and Engineering Chemistry Fundamentals*. 10, 1971, 321-322.

I.B. Tjoa, L.T. Biegler. Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems. *Industrial Engineering and Chemistry Research*. 30, 1991, 376-385.

S. Vajda, P. Valko. A direct-indirect procedure for estimation of kinetic parameters. *Computers and Chemical Engineering.* 10 (1), 1986, 49-58.

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

M.S. Varziri, K.B. McAuley, P.J. McLellan. Parameter estimation in continuous-time dynamic models in the presence of unmeasured states and non-stationary disturbances. *Industrial Engineering and Chemistry Research*. 47, 2008a, 380-393.

M.S. Varziri, K.B. McAuley, P.J. McLellan. Approximate maximum likelihood parameter estimation for nonlinear dynamic models: application to a lab-scale nylon reactor model. *Industrial Engineering and Chemistry Research.* 2008b, submitted for publication

M.S. Varziri, K.B. McAuley, P.J. McLellan. Parameter estimation in nonlinear stochastic continuous-time dynamic models with unknown disturbance intensity, *Canadian Journal of Chemical Eng*ineering. 2008c, submitted for publication.

Chapter 4

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

4.1 Abstract

In Chapter 3 (Varziri et al.,2008), the iPDA algorithm (Poyton et al., 2006) was reviewed in detail. By considering a stochastic differential equation model, and also using probabilistic principles, it was shown that the iPDA objective function is an approximate maximum likelihood criterion. In this chapter, we investigate the applicability of the iPDA, as an Approximate Maximum Likelihood Estimation (AMLE) algorithm, to parameter estimation in nonlinear stochastic dynamic models, in which some of the states are unmeasured. We also 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.

This chapter has been published as a journal paper in *Industrial Engineering and Chemistry Research*, 47, 2008, 380-393.

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

M.S. Varziri, K. B. McAuley and P. J. McLellan, Department of Chemical Engineering, Queen's University, Kingston, ON, Canada K7L 3N6

4.2 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 (Seber and Wild, 1989; Bates and Watts, 1988; Bard, 1974; Ogunnaike and Ray, 1994). 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 (Leis and Kramer, 1988), 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 (Biegler and Grossman, 2004).

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 (Poyton et al., 2006; Tang, 1971; Swartz and Bremermann, 1975; Varah, 1982, Vajda and Volko, 1986). 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 (Hosten, 1979). Biegler and coworkers (Biegler, 1984; Biegler and Grossman, 2004) 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 (Wächter and Biegler, 2006).

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. (2006) 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. (2007) 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 (Varziri et al., 2008) (Chapter 3). 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 (Gagnon and MacGregor, 1991). 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., (2006), and Varziri et al., (2008) (presented in Chapter 3). 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.

4.2.1 Review of the AMLE algorithm

Varziri et al., (2008) derived the AMLE objective function using probabilistic principles 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:

$$\begin{cases} \dot{x}_1(t) = f_1(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + \eta_1(t) \\ \dot{x}_2(t) = f_2(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + \eta_2(t) \\ \\ x_1(0) = x_{10} \\ x_2(0) = x_{20} \\ \\ y_1(t_{m1j}) = x_1(t_{m1j}) + \varepsilon_1(t_{m1j}) \\ y_2(t_{m2j}) = x_2(t_{m2j}) + \varepsilon_2(t_{m2j}) \end{cases} \tag{4.1}$$

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; bounded growth) 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 state trajectories are approximated using linear combinations of B-spline basis functions (de Boor, 2001; Ramsay and Silverman, 2005):

$$x_{\sim k}(t) = \sum_{i=1}^{c_k} \beta_{ki}\phi_{ki} \qquad k = 1,2 \tag{4.2}$$

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 *k*th state. Eq. (4.1) can be written in matrix form:

$$x_{\sim k}(t) = \varphi_k^T(t)\boldsymbol{\beta}_k \qquad k = 1,2 \tag{4.3}$$

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:

$$\dot{x}_{\sim k}(t) = \frac{d}{dt}\left(\sum_{i=1}^{c_k} \beta_{ki}\phi_{ki}(t)\right) = \sum_{i=1}^{c_k} \beta_{ki}\dot{\phi}_{ki}(t) = \dot{\varphi}_k^T\boldsymbol{\beta}_k \qquad k = 1,2 \tag{4.4}$$

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:

$$\begin{aligned}
&\frac{1}{2\sigma_{m1}^2}\sum_{j=1}^{N_1}(y_1(t_{m1j}) - x_{\sim 1}(t_{m1j}))^2 + \\
&\frac{1}{2\sigma_{m2}^2}\sum_{j=1}^{N_2}(y_2(t_{m2j}) - x_{\sim 2}(t_{m2j}))^2 + \\
&\frac{1}{2Q_{p1}}\int\left(\dot{x}_{\sim 1}(t) - f_1(x_{\sim 1}(t), x_{\sim 2}(t), u_1(t), u_2(t), \boldsymbol{\theta})\right)^2 dt + \\
&\frac{1}{2Q_{p2}}\int\left(\dot{x}_{\sim 2}(t) - f_2(x_{\sim 1}(t), x_{\sim 2}(t), u_1(t), u_2(t), \boldsymbol{\theta})\right)^2 dt
\end{aligned} \tag{4.5}$$

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 (4.5) 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. (2006)) 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 likelihood argument to show that, in the dynamic model described in (4.1), 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 (4.5) was derived (Varziri et al., 2008) (Chapter 3). 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 (addressed in Chapter 5).

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 4.4 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 function 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.

4.3 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:

$$\frac{1}{2\sigma_{m2}^2}\sum_{j=1}^{N_2}(y_2(t_{m2j})-x_{\sim 2}(t_{m2j}))^2 + \frac{1}{2Q_{p1}}\int\left(\dot{x}_{\sim 1}(t)-f_1(x_{\sim 1}(t),x_{\sim 2}(t),u_1(t),u_2(t),\boldsymbol{\theta})\right)^2 dt + \frac{1}{2Q_{p2}}\int\left(\dot{x}_{\sim 2}(t)-f_2(x_{\sim 1}(t),x_{\sim 2}(t),u_1(t),u_2(t),\boldsymbol{\theta})\right)^2 dt \quad (4.6)$$

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 (4.6) are evaluated. Parameter estimation in a nonlinear MIMO CSTR with an unmeasured state is studied in Section 4.5.2 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 (4.7) below, which is a slightly modified version of the model in (4.1). 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 (4.7), $d_1(t)$ is a non-stationary disturbance driven by $\eta_{d1}(t)$ which is a continuous zero-mean stationary white-noise process with intensity $Q_{pdl.}$ $f_{d1}(.)$ is a function defining the non-stationary noise model and $\boldsymbol{\theta}_{d1}$ is the vector of noise model parameters (Maybeck, 1979). If $\boldsymbol{\theta}_{d1}$ is not

known (which generally is the case) it can be estimated along with the vector of fundamental model parameters $\boldsymbol{\theta}$.

$$\begin{cases} \dot{x}_1(t) = f_1(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + d_1(t) + \eta_1(t) \\ \dot{x}_2(t) = f_2(x_1(t), x_2(t), u_1(t), u_2(t), \boldsymbol{\theta}) + \eta_2(t) \\ \dot{d}_1(t) = f_{d1}(d_1(t), \boldsymbol{\theta}_{d1}) + \eta_{d1}(t) \\ \\ x_1(0) = x_{10} \\ x_2(0) = x_{20} \\ d_1(0) = 0 \\ \\ y_1(t_{m1j}) = x_1(t_{m1j}) + \varepsilon_1(t_{m1j}) \\ y_2(t_{m2j}) = x_2(t_{m2j}) + \varepsilon_2(t_{m2j}) \end{cases} \tag{4.7}$$

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:

$$\begin{aligned} &\frac{1}{2\sigma_{m1}^2}\sum_{j=1}^{N_1}(y_1(t_{m1j}) - x_{\sim 1}(t_{m1j}))^2 + \\ &\frac{1}{2\sigma_{m2}^2}\sum_{j=1}^{N_2}(y_2(t_{m2j}) - x_{\sim 2}(t_{m2j}))^2 + \\ &\frac{1}{2Q_{p1}}\int \left(\dot{x}_{\sim 1}(t) - f_1(x_{\sim 1}(t), x_{\sim 2}(t), d_{\sim 1}(t), u_1(t), u_2(t), \boldsymbol{\theta})\right)^2 dt + \\ &\frac{1}{2Q_{p2}}\int \left(\dot{x}_{\sim 2}(t) - f_2(x_{\sim 1}(t), x_{\sim 2}(t), u_1(t), u_2(t), \boldsymbol{\theta})\right)^2 dt + \\ &\frac{1}{2Q_{d1}}\int \left(\dot{d}_{\sim 1}(t) - f_{d1}(d_{\sim 1}(t), \boldsymbol{\theta}_{d1})\right)^2 dt \end{aligned} \tag{4.8}$$

where, $d_{\sim 1}(t) = \varphi_{d1}^T(t)\boldsymbol{\beta}_{d1}$. Note that (4.8) 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 4.5.3.

4.4 Theoretical confidence intervals

Varziri et al. (2008) (Chapter 3) considered the model in (4.1) 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 (1982) 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 $L(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y}_m)$:

$$L(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y}_m) = \ln\left(p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta})\right) \tag{4.9}$$

$$\hat{\boldsymbol{\theta}}, \hat{\mathbf{x}} = \arg\min_{\boldsymbol{\theta}, \mathbf{x}} \left\{ - L(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y}_m) \right\} \tag{4.10}$$

Analogous objective functions have been used in nonlinear filtering problems (Mortensen, 1968; Jazwinski, 1970; Evensen et al., 1998). Classical maximum likelihood parameter estimates (which are obtained by maximizing $p(\mathbf{y}_m \mid \boldsymbol{\theta})$), under general regularity conditions, are consistent, asymptotically unbiased, asymptotically Normally-distributed and asymptotically efficient (Rao, 1973; Kay, 1993). Barshalom (1972), proved the asymptotic efficiency of parameter estimates obtained from the minimization problem in (4.10) in case of a discrete-time linear dynamic model. However, as noted by Yeredor (2000), this may not be true for a general nonlinear dynamic model.

Assuming that *x* can be represented using B-splines as in (4.3), the parameter estimation problem reduces to:

$$\min_{\boldsymbol{\theta}, \boldsymbol{\beta}} \left\{ - L(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y}_m) \right\} \tag{4.11}$$

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

$$\min_{\boldsymbol{\tau}} \{ - L(\boldsymbol{\tau}, \mathbf{y}_m) \} \tag{4.12}$$

We assume that L satisfies the required regularity conditions (*e.g.*, Kay (1993), Appendix 7B, page 212). Then, from the asymptotic properties of maximum likelihood estimators (*e.g.*, Kay (1993), Theorem 7.3, page 183) we have:

$$\hat{\boldsymbol{\tau}} \sim N(\boldsymbol{\tau}, \mathbf{I}^{-1}(\boldsymbol{\tau})) \tag{4.13}$$

where $\mathbf{I}(\boldsymbol{\tau})$ is the Fisher information matrix (Seber and Wild, 1989) evaluated at the true values of $\boldsymbol{\tau}$:

$$\mathbf{I}(\boldsymbol{\tau}) = -E\left\{ \frac{\partial^2 L(\boldsymbol{\tau}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}^2} \right\} = E\left\{ \frac{\partial L(\boldsymbol{\tau}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}}^T \frac{\partial L(\boldsymbol{\tau}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}} \right\} \tag{4.14}$$

In the examples of Section 4.5, we have approximated $\mathbf{I}(\boldsymbol{\tau})$ by equation (4.25) in Appendix 4.8.1 evaluated at $\hat{\boldsymbol{\tau}}$.

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

$$\boldsymbol{\tau} = \hat{\boldsymbol{\tau}} \pm z_{\alpha/2} \times \sqrt{\mathrm{diag}(\mathbf{I}^{-1}(\hat{\boldsymbol{\tau}}))} \tag{4.15}$$

The confidence intervals obtained from (4.15) are approximate because:

1. The model is nonlinear with respect to the parameters and spline coefficients;
2. The parameter estimates obtained from maximizing the prescribed likelihood function may not necessarily be asymptotically Normal and even if they are, asymptotic results are generally violated when small data samples are used;

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 (4.15), rather than a value from student's t distribution because we are assuming that the noise variance is known.

4.5 Case study

In this section we examine our results using a MIMO nonlinear CSTR model (Marlin, 2000). In Section 4.5.1 we estimate 4 parameters in the CSTR model, assuming that both temperature and concentration can be measured. In Section 4.5.2, we assume that only temperature can be measured and concentration is unmeasured. In Section 4.5.3, 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 IPOPT solver (Wächter and Biegler, 2006) 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 IPOPT, 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 IPOPT (Wächter and Biegler, 2006) are essential for the successful implementation of these algorithms, especially for larger problems.

4.5.1 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 (Marlin, 2000) with additional stochastic disturbance terms:

$$
\begin{aligned}
&\frac{dC_A(t)}{dt} = \frac{F(t)}{V}\left(C_{A0}(t) - C_A(t)\right) - gC_A(t) + \eta_1(t) \\
&\frac{dT(t)}{dt} = \frac{F(t)}{V}\left(T_0(t) - T(t)\right) + \beta_1\left(T(t) - T_{cin}(t)\right) + \beta_2 gC_A(t) + \eta_2(t) \\
&C_A(0) = 1.569\,(\text{kmol m}^{-3}) \\
&T(0) = 341.37\,(\text{K}) \\
&y_1(t_i) = C_A(t_i) + \varepsilon_1(t_i) \\
&y_2(t_j) = C_A(t_j) + \varepsilon_2(t_j)
\end{aligned}
\tag{4.16}
$$

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

where $E\{\eta_1(t_i)\eta_1(t_j)\} = \sigma_{p1}^2\delta(t_i - t_j)$, $E\{\eta_2(t_i)\eta_2(t_j)\} = \sigma_{p2}^2\delta(t_i - t_j)$ ($\delta(.)$ 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 (Marlin, 2000) are as follows: $V = 1.0$ m^3, $C_p = 1$ cal g^{-1}K^{-1}, $\rho = 1\text{E}6$ g m^{-3}, $C_{pc} = 1$ cal g^{-1}K^{-1}, $\rho_c = 1\text{E}6$ g m^{-3}, and $-\Delta H_{rxn} = 130\text{E}6$ cal kmol^{-1}. The initial steady-state operating point is: $C_{As} = 1.569$ kmol m^{-3} and $T_s = 341.37$ K.

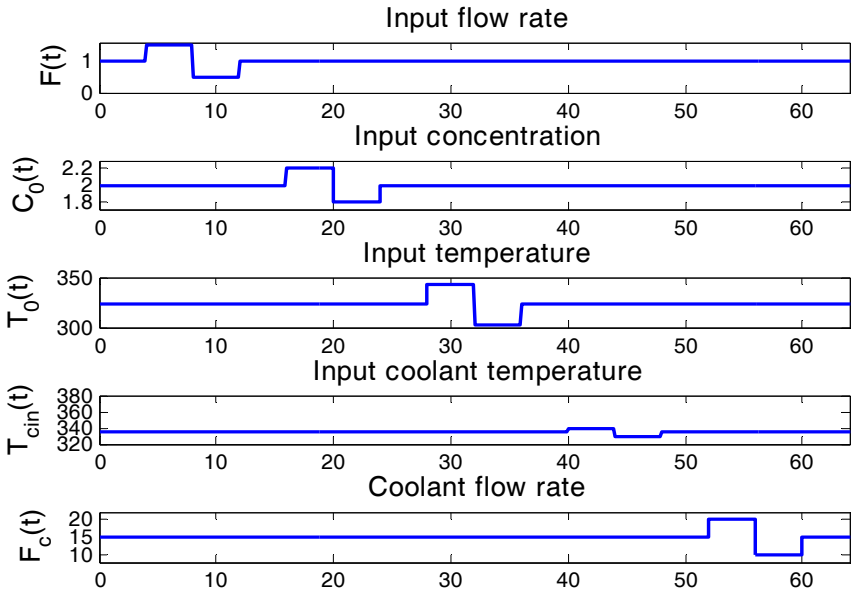

Figure 4.1. Input scheme for MIMO nonlinear CSTR

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 4.1 (Poyton, 2005). 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 (4.5) , the AMLE objective function is:

$$
\begin{aligned}
&\frac{1}{2\sigma_{m1}^2}\sum_{j=1}^{64}\left(y_1(t_{m1j})-C_{A\sim}(t_{m1j})\right)^2+\\
&\frac{1}{2\sigma_{m2}^2}\sum_{j=1}^{213}\left(y_2(t_{m2j})-T_{\sim}(t_{m2j})\right)^2+\\
&\frac{1}{2Q_{p1}}\int_{t=0}^{64}\left(\frac{dC_{A\sim}(t)}{dt}-\frac{F(t)}{V}\left(C_{A0}(t)-C_{A\sim}(t)\right)+gC_{A\sim}(t)\right)^2dt+\\
&\frac{1}{2Q_{p2}}\int_{t=0}^{64}\left(\frac{dT_{\sim}(t)}{dt}-\frac{F(t)}{V}\left(T_0(t)-T_{\sim}(t)\right)-\beta_1\left(T_{\sim}(t)-T_{cin}(t)\right)-\beta_2 gC_{A\sim}(t)\right)^2dt
\end{aligned}
\tag{4.17}
$$

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 (4.17).

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 Figure 4.2 to Figure 4.5. The AMLE and TNLS predicted responses are with the true responses in Figure 4.6 to Figure 4.9. On average, the AMLE parameter estimates are better 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.

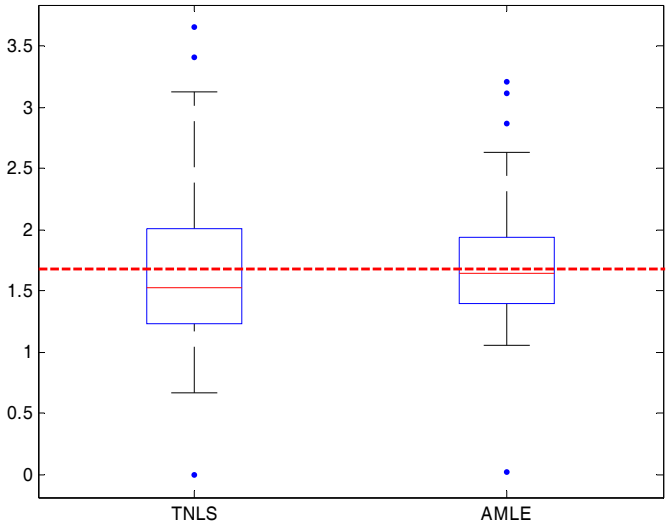

Figure 4.2. Box plots for a using TNLS and AMLE

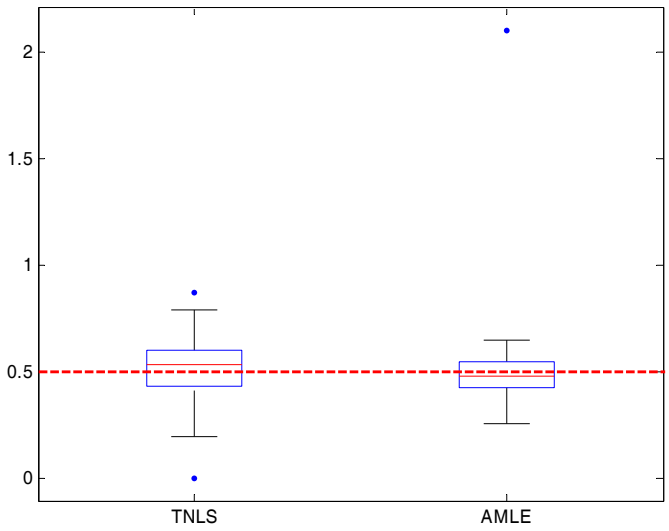

Figure 4.3. Box plots for b using TNLS and AMLE

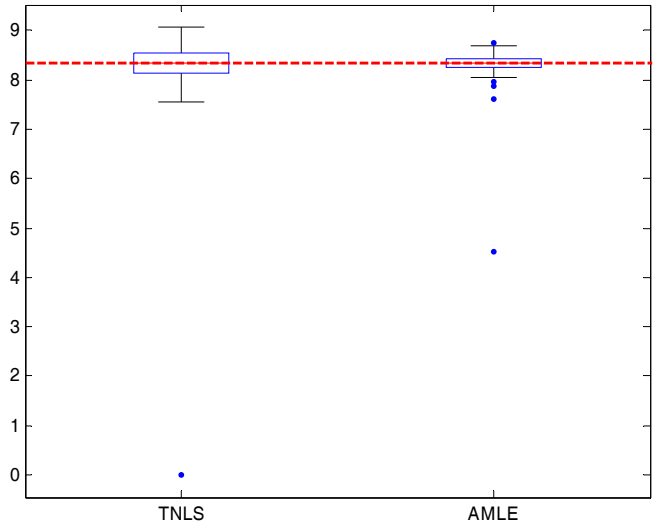

Figure 4.4. Box plots for $\frac{E}{R}$ using TNLS and AMLE

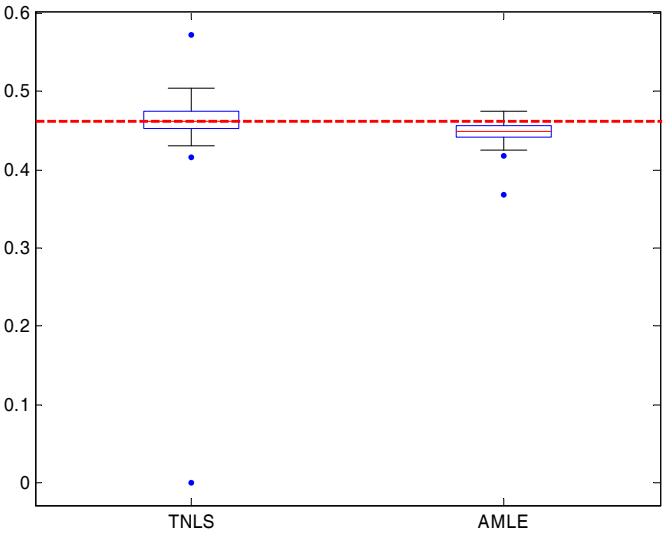

Figure 4.5. Box plots for k_{ref} using TNLS and AMLE

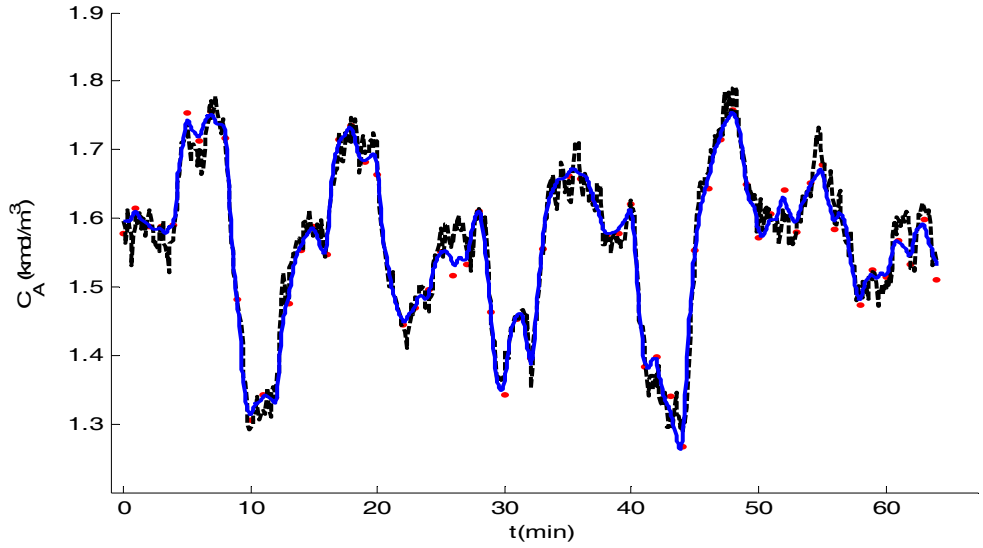

Figure 4.6. Measured, true, and predicted concentration response for AMLE (● simulated data, -- -- response with true parameters and true stochastic noise, — AMLE response)

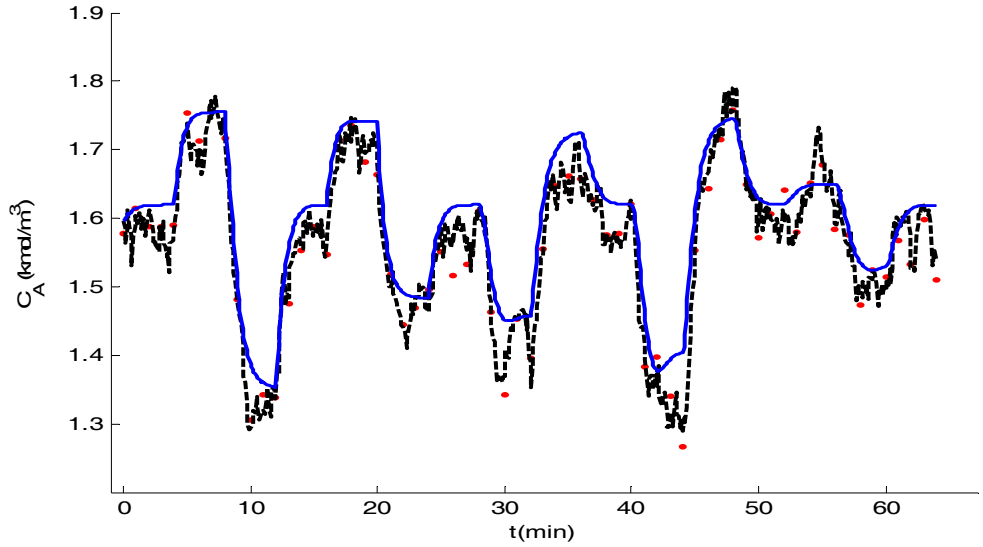

Figure 4.7. Measured, true, and predicted concentration response for TNLS (● simulated data, -- -- response with true parameters and true stochastic noise, — TNLS response)

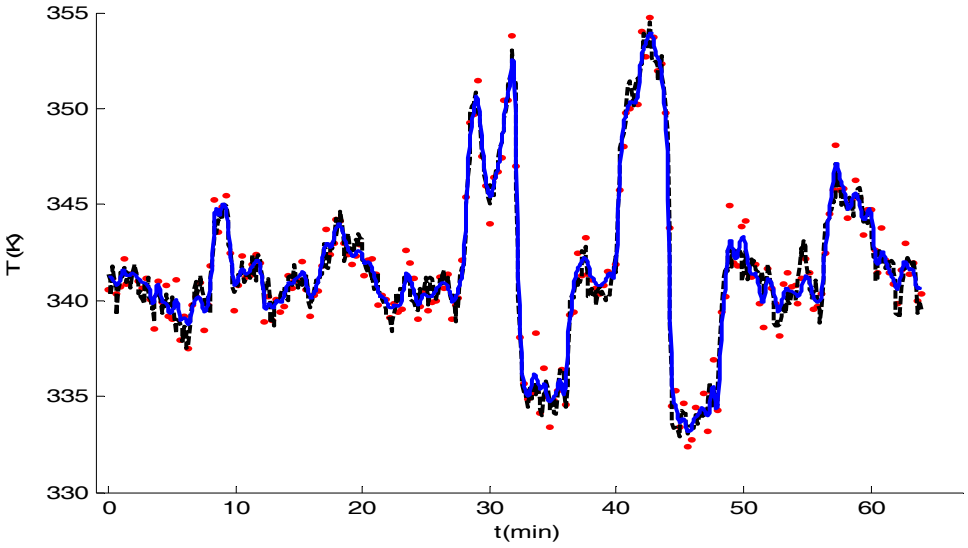

Figure 4.8. Measured, true, and predicted temperature response for AMLE

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

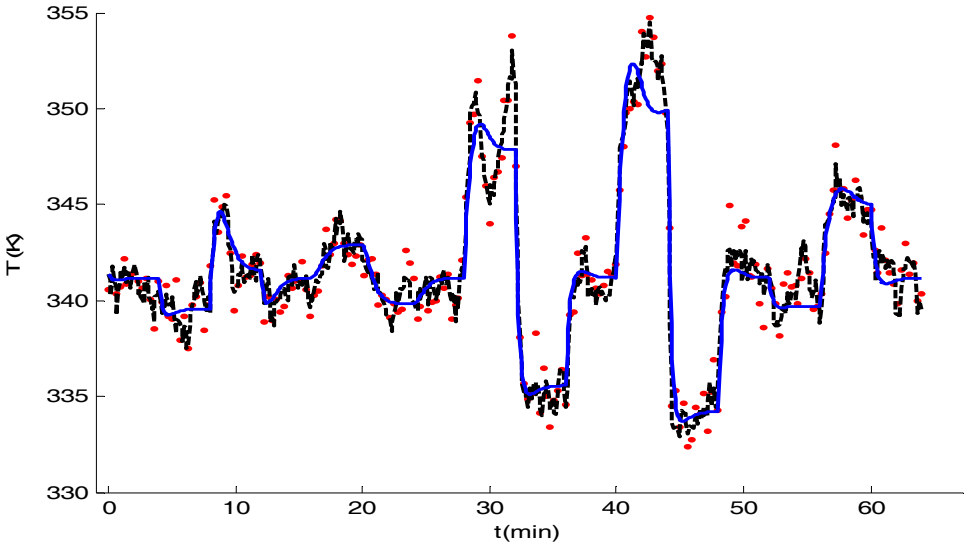

Figure 4.9. Measured, true, and predicted temperature response for TNLS (● simulated data, ---- response with true parameters and true stochastic noise, — TNLS response)

Table 4.1 presents 95% confidence intervals (derived in Section 4.4) 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.

By inspecting the Monte Carlo box plots and the scatter plot matrix given in Appendix 4.8.2, and comparing them to the theoretical results in Table 4.1, we see that the empirical and theoretical confidence intervals are consistent.

Table 4.1. 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

4.5.2 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 (4.6) , the appropriate AMLE objective function is:

$$\frac{1}{2\sigma_{m2}^2}\sum_{j=1}^{213}\left(y_2(t_{m2j})-T_{\sim}(t_{m2j})\right)^2+$$
$$\frac{1}{2Q_{p1}}\int_{t=0}^{64}\left(\frac{dC_{A\sim}(t)}{dt}-\frac{F(t)}{V}\left(C_{A0}(t)-C_{A\sim}(t)\right)+gC_{A\sim}(t)\right)^2dt+ \qquad (4.18)$$
$$\frac{1}{2Q_{p2}}\int_{t=0}^{64}\left(\frac{dT_{\sim}(t)}{dt}-\frac{F(t)}{V}\left(T_0(t)-T_{\sim}(t)\right)-\beta_1\left(T_{\sim}(t)-T_{cin}(t)\right)-\beta_2 gC_{A\sim}(t)\right)^2dt$$

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 Figure 4.10 to Figure 4.13. The AMLE and TNLS predicted responses are compared against the true responses in Figure 4.14 to Figure 4.17. From Figure 4.10 to Figure 4.17 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.

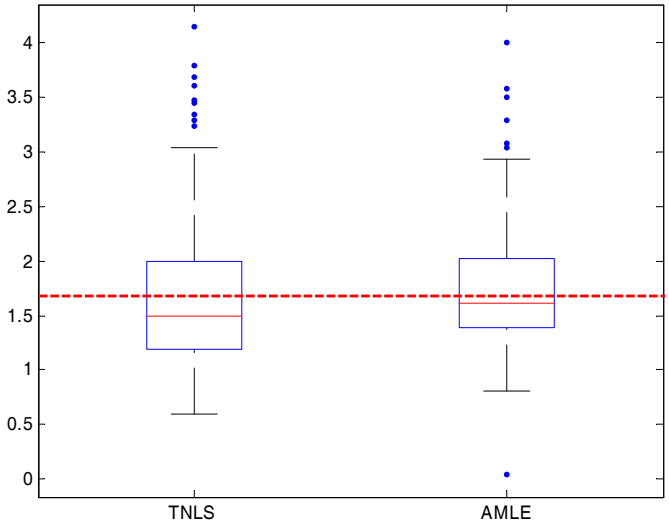

Figure 4.10. Box plots for a using TNLS and AMLE (unmeasured concentration)

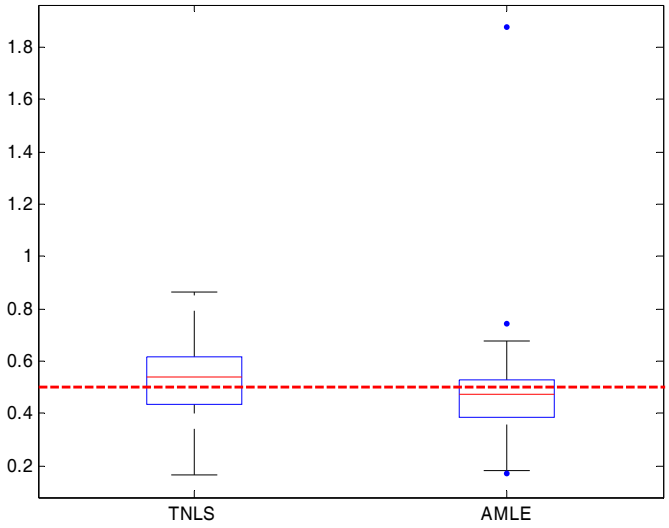

Figure 4.11. Box plots for b using TNLS and AMLE (unmeasured concentration)

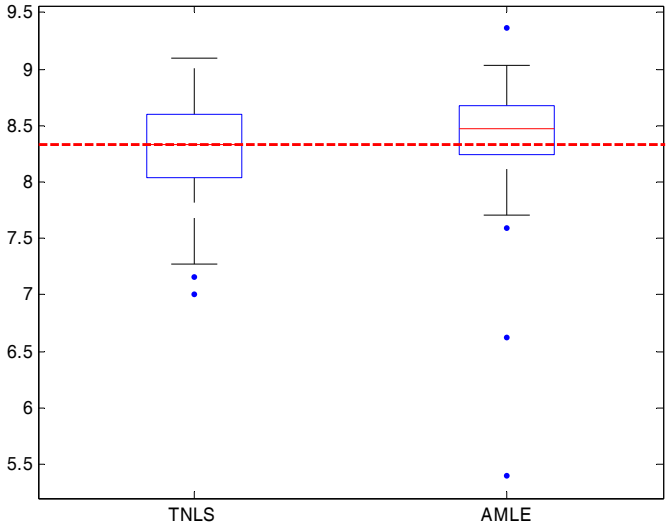

Figure 4.12. Box plots for $\frac{E}{R}$ using TNLS and AMLE (unmeasured concentration)

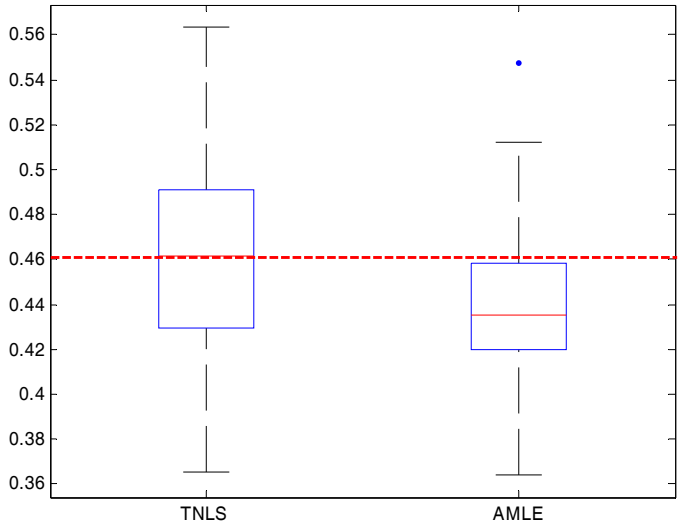

Figure 4.13. Box plots for k_{ref} using TNLS and AMLE (unmeasured concentration)

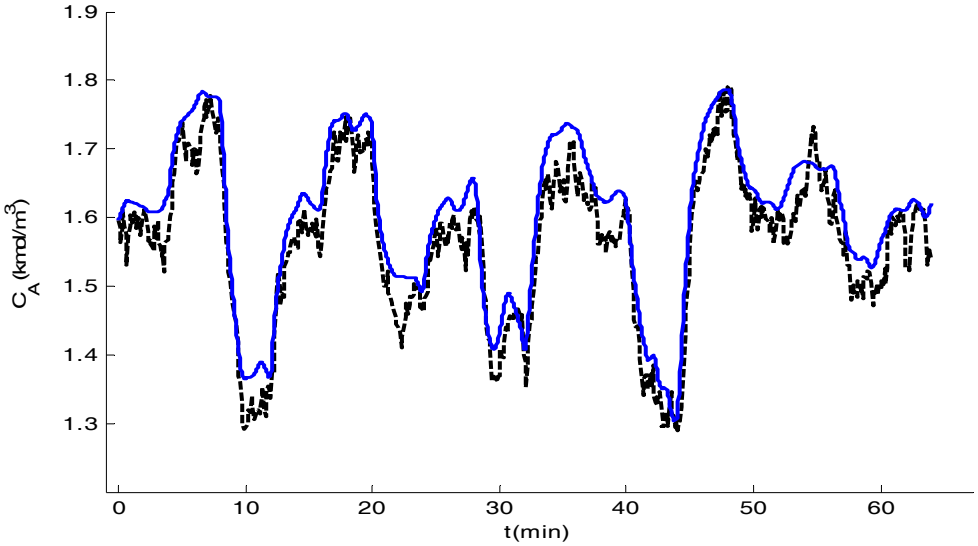

Figure 4.14. Measured, true, and predicted concentration response for AMLE (unmeasured concentration) (---- response with true parameters and true stochastic noise, — AMLE response)

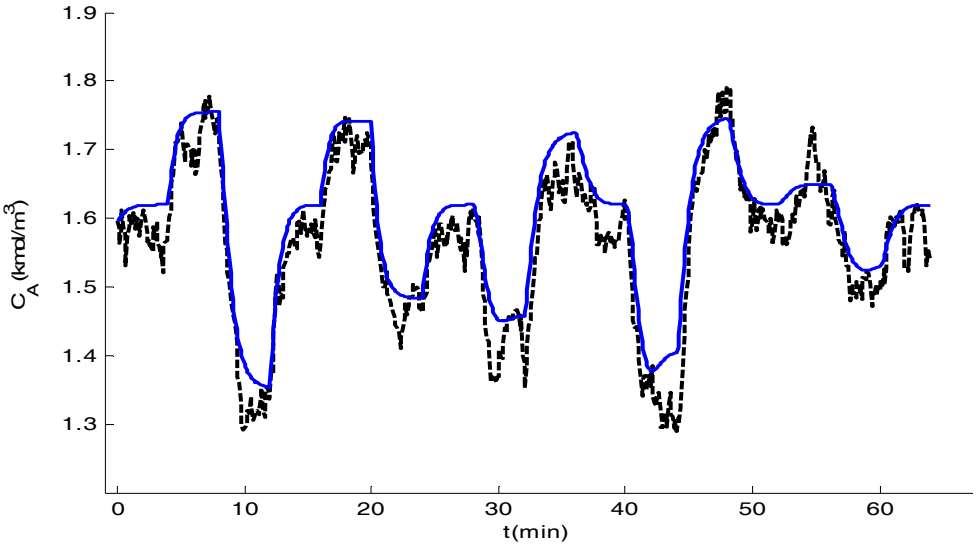

Figure 4.15. Measured, true, and predicted concentration response for TNLS (unmeasured concentration) (---- response with true parameters and true stochastic noise, — TNLS response)

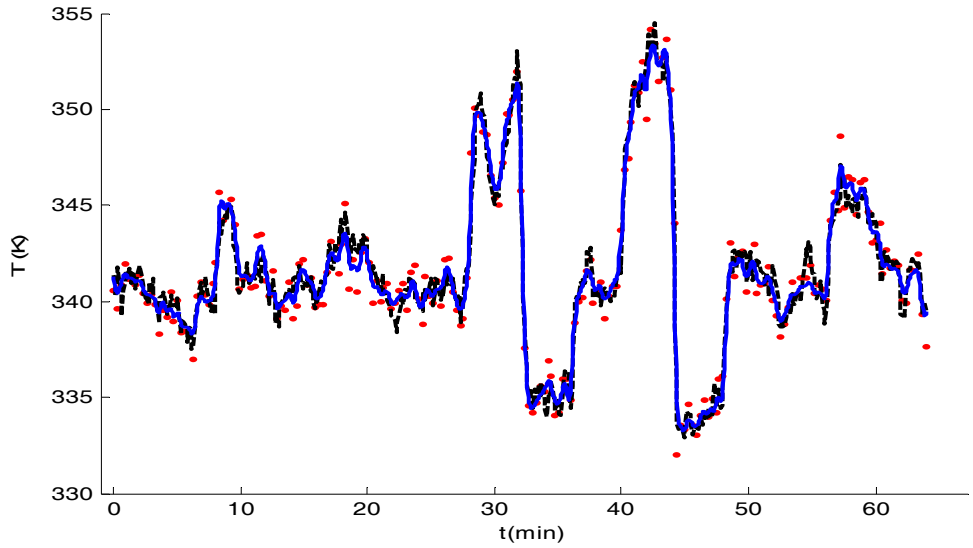

Figure 4.16. Measured, true, and predicted temperature response for AMLE (unmeasured concentration) (• simulated data, ---- response with true parameters and true stochastic noise, — AMLE response)

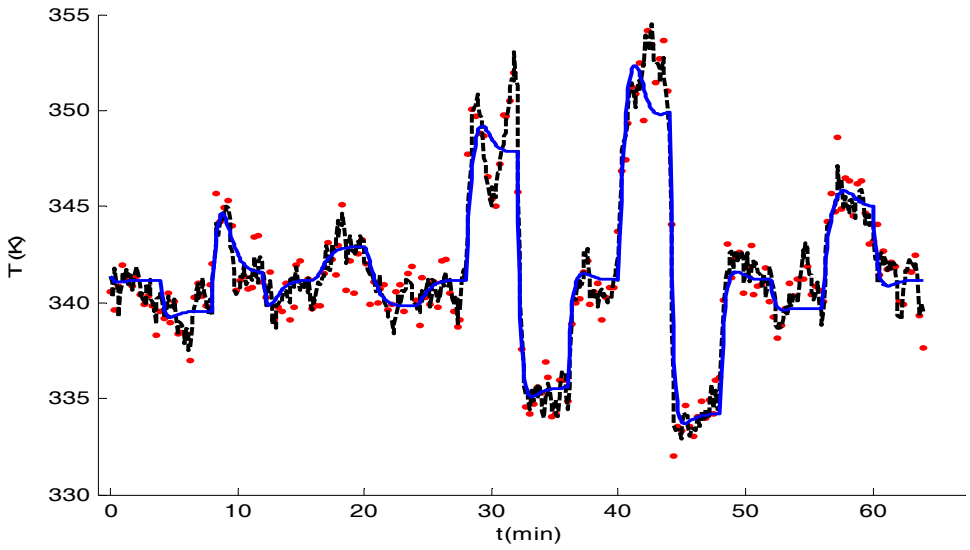

Figure 4.17. Measured, true, and predicted temperature response for TNLS (unmeasured concentration) (• simulated data, ---- response with true parameters and true stochastic noise, — TNLS response)

Theoretical 95% confidence intervals for AMLE parameter estimates are presented in Table 4.2.

Table 4.2. 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 4.8.2.

4.5.3 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:

$$
\begin{aligned}
&\frac{dC_A(t)}{dt} = \frac{F(t)}{V}\left(C_{A0}(t) - C_A(t)\right) - gC_A(t) + d_1(\mathrm{t}) + \eta_1(t) \\
&\frac{dT(t)}{dt} = \frac{F(t)}{V}\left(T_0(t) - T(t)\right) + \beta_1\left(T(t) - T_{cin}(t)\right) + \beta_2 gC_A(t) + \eta_2(t) \\
&\frac{dd_1(t)}{dt} = \eta_3(t) \\
&C_A(0) = 1.569\,(\mathrm{kmol\,m^{-3}}) \\
&T(0) = 341.37\,(\mathrm{K}) \\
&d_1(0) = 0(\mathrm{kmol\,m^{-3}\,min^{-1}}) \\
&y_1(t_i) = C_A(t_i) + \varepsilon_1(t_i) \\
&y_2(t_j) = C_A(t_j) + \varepsilon_2(t_j)
\end{aligned}
\tag{4.19}
$$

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

$E\{\eta_3(t)\eta_3(t-\tau)\} = Q_{p3}\delta(\tau)$ where $Q_{p3} = 6\times10^{-2}$ (kmol/m^3 / min)^2 and everything else remains the same as in Section 4.5.1.

From equation (4.8), the AMLE objective function is:

$$\begin{aligned}
&\frac{1}{2\sigma_{m1}^2}\sum_{j=1}^{64}\left(y_1(t_{m1j}) - C_{A\sim}(t_{m1j})\right)^2 + \\
&\frac{1}{2\sigma_{m2}^2}\sum_{j=1}^{213}\left(y_2(t_{m2j}) - T_{\sim}(t_{m2j})\right)^2 + \\
&\frac{1}{2Q_{p1}}\int_{t=0}^{64}\left(\frac{dC_{A\sim}(t)}{dt} - \frac{F(t)}{V}\left(C_{A0}(t) - C_{A\sim}(t)\right) + gC_{A\sim}(t) - d_{1\sim}(t)\right)^2 dt + \\
&\frac{1}{2Q_{p2}}\int_{t=0}^{64}\left(\frac{dT_{\sim}(t)}{dt} - \frac{F(t)}{V}\left(T_0(t) - T_{\sim}(t)\right) - \beta_1\left(T_{\sim}(t) - T_{cin}(t)\right) - \beta_2 gC_{A\sim}(t)\right)^2 dt + \\
&\frac{1}{2Q_{p3}}\int_{t=0}^{64}\left(\dot{d}_{1\sim}(t)\right)^2 dt
\end{aligned} \tag{4.20}$$

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.

Figure 4.18 to Figure 4.21 show the Monte Carlo box plots for the parameter estimates. Estimated concentration and temperature trajectories are presented (for one data set) in Figure 4.22 to Figure 4.25.

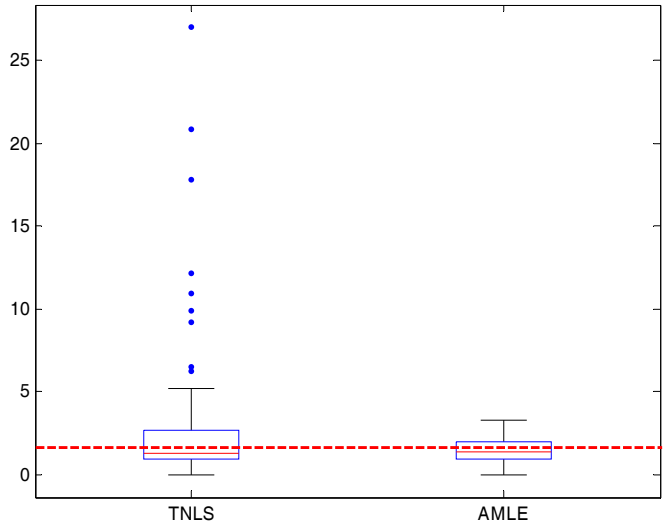

Figure 4.18. Box plots for a using TNLS and AMLE (with non-stationary disturbance)

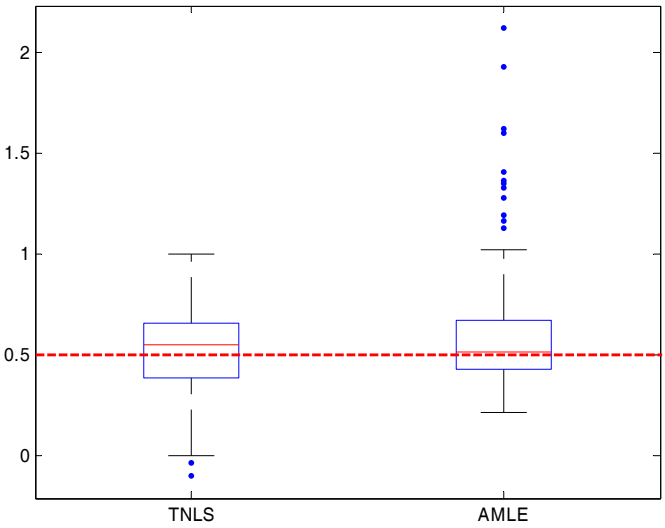

Figure 4.19. Box plots for b using TNLS and AMLE (with non-stationary disturbance)

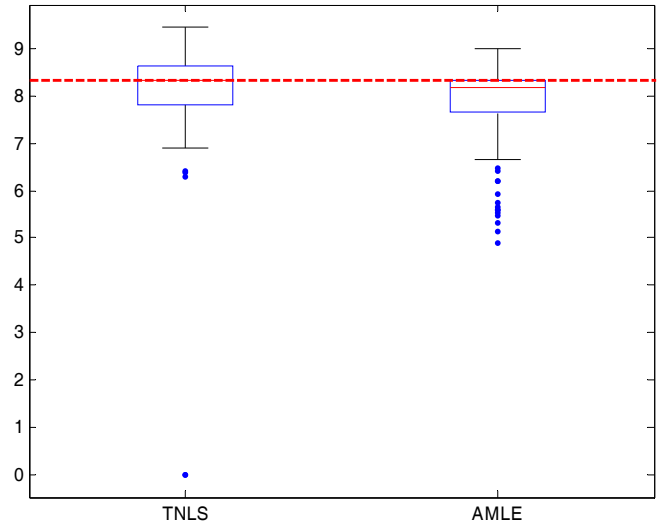

Figure 4.20. Box plots for $\frac{E}{R}$ using TNLS and AMLE (with non-stationary disturbance)

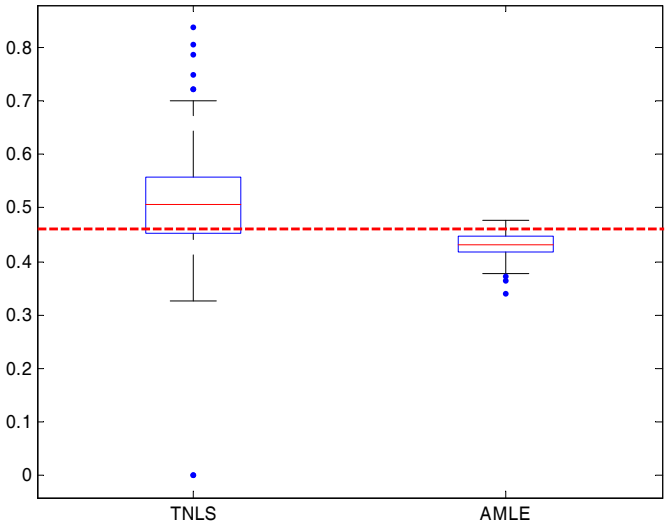

Figure 4.21. Box plots for k_{ref} using TNLS and AMLE (with non-stationary disturbance)

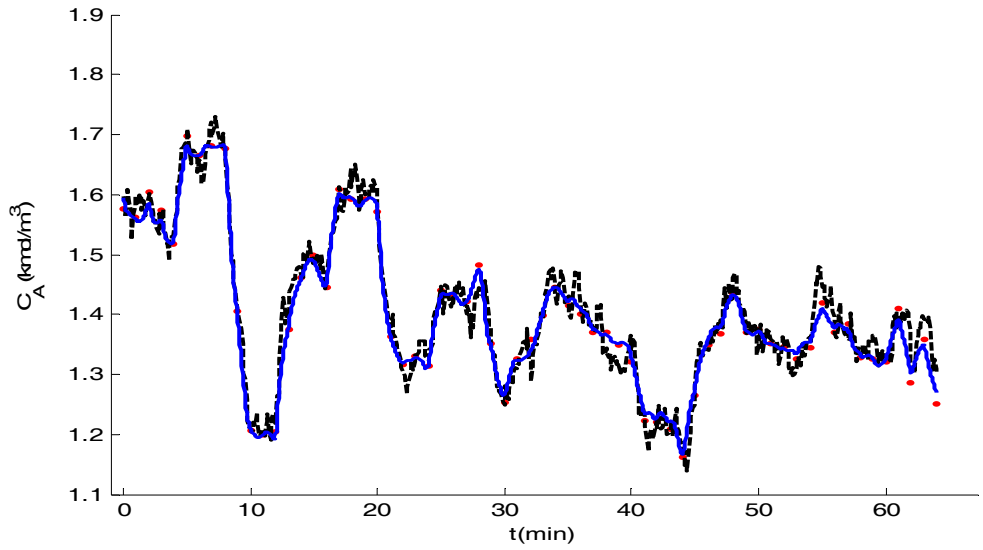

Figure 4.22. Measured, true, and predicted concentration response for AMLE, non-stationary example (• simulated data, ----- response with true parameters and true stochastic noise, — AMLE response)

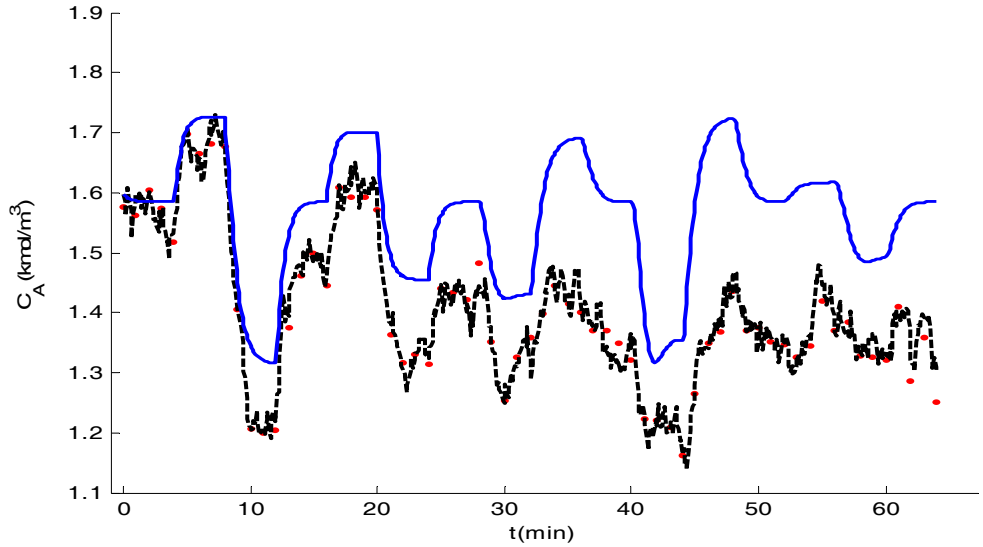

Figure 4.23. Measured, true, and predicted concentration responses using TNLS, non-stationary example (• simulated data, ----- response with true parameters and true stochastic noise, — TNLS response)

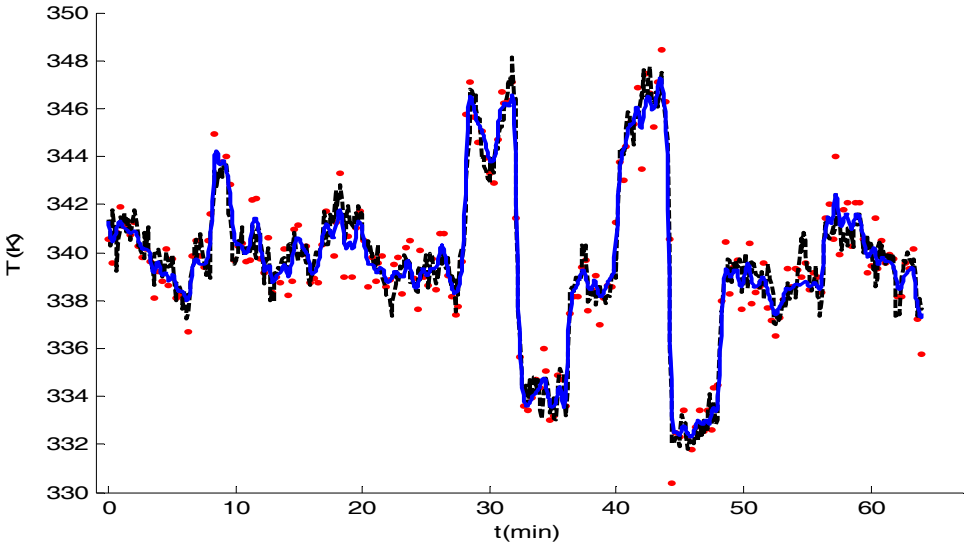

Figure 4.24. Measured, true, and predicted temperature response for AMLE, non-stationary example (• simulated data, ----- response with true parameters and true stochastic noise, — AMLE response)

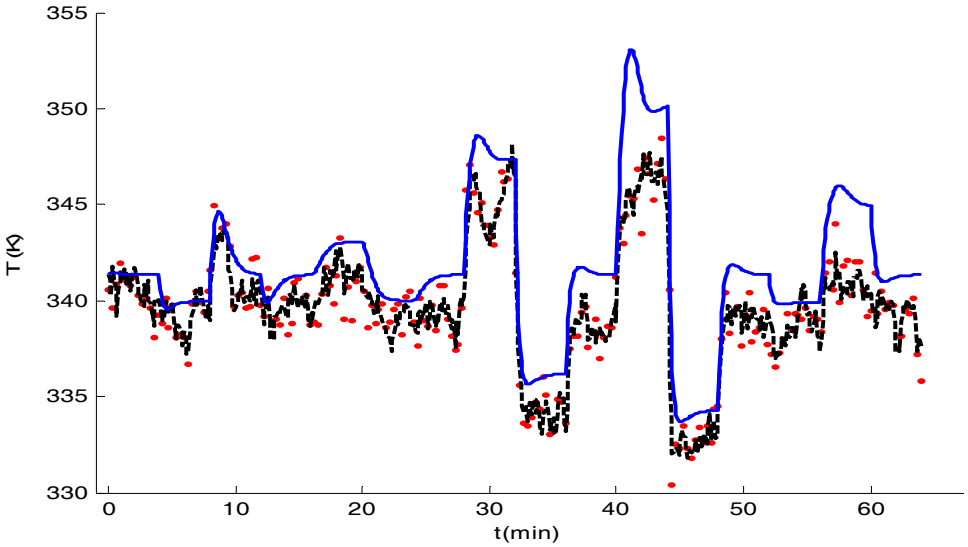

Figure 4.25. Measured, true, and predicted temperature response for TNLS, non-stationary example (• simulated data, ----- response with true parameters and true stochastic noise, — TNLS response)

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 4.26. Please note that the non-stationary disturbance was estimated using different sets of random observations (not shown) and it was confirmed that the apparent bias in Figure 4.26 is not systematic.

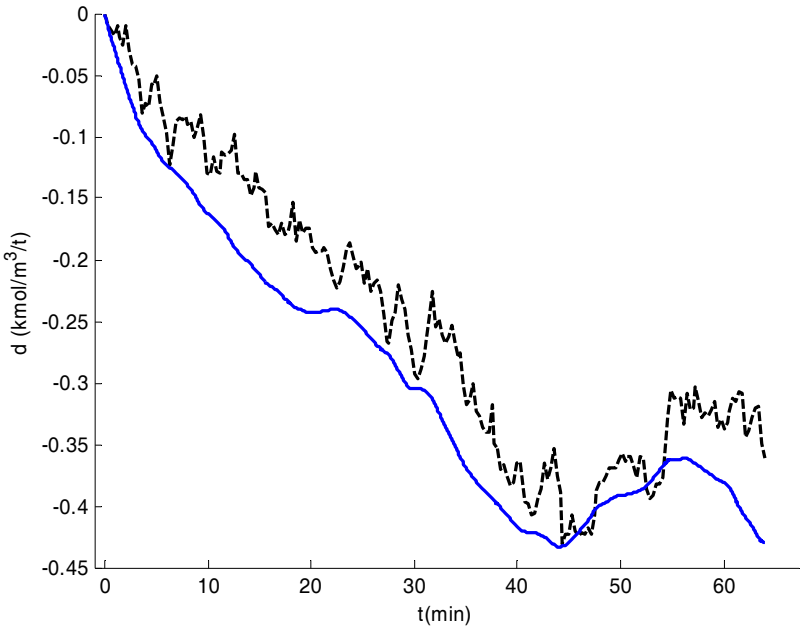

Figure 4.26. True, and estimated non-stationary disturbance using AMLE,

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

Theoretical confidence intervals for the AMLE parameter estimates for the second example are presented in Table 4.3. As expected, these confidence intervals are wider than when there was no meandering disturbance (Table 4.1).

A scatter plot matrix for the parameter estimates is presented in Appendix 4.8.2.

Table 4.3. 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

4.6 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 (addressed in Chapter 5).

4.7 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.

4.8 Appendix

4.8.1 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 (4.5) in the matrix-vector form where we replace the integrals by Gaussian Quadrature sums we have:

$$\begin{aligned}
&\frac{1}{2\sigma_{m1}^2}(\mathbf{y}_1-\mathbf{x}_{\sim 1})^T(\mathbf{y}_1-\mathbf{x}_{\sim 1})+\\
&\frac{1}{2\sigma_{m2}^2}(\mathbf{y}_2-\mathbf{x}_{\sim 2})^T(\mathbf{y}_2-\mathbf{x}_{\sim 2})+\\
&\frac{1}{2Q_{p1}}\left(\dot{\mathbf{x}}_{\sim q1}-\mathbf{f}_1(\mathbf{x}_{\sim q1},\mathbf{x}_{\sim q2},\mathbf{u}_1,\mathbf{u}_2,\boldsymbol{\theta})\right)^T\mathbf{w}_1\left(\dot{\mathbf{x}}_{\sim q1}-\mathbf{f}_1(\mathbf{x}_{\sim q1},\mathbf{x}_{\sim q2},\mathbf{u}_1,\mathbf{u}_2,\boldsymbol{\theta})\right)+\\
&\frac{1}{2Q_{p2}}\left(\dot{\mathbf{x}}_{\sim q2}-\mathbf{f}_2(\mathbf{x}_{\sim q1},\mathbf{x}_{\sim q2},\mathbf{u}_1,\mathbf{u}_2,\boldsymbol{\theta})\right)^T\mathbf{w}_2\left(\dot{\mathbf{x}}_{\sim q2}-\mathbf{f}_2(\mathbf{x}_{\sim q1},\mathbf{x}_{\sim q2},\mathbf{u}_1,\mathbf{u}_2,\boldsymbol{\theta})\right)
\end{aligned} \tag{4.21}$$

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 (4.21) can further be simplified by introducing the following vectors and matrices:

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

$$\boldsymbol{\beta}=\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix},\quad \boldsymbol{\varphi}=\begin{bmatrix}\boldsymbol{\varphi}_1 & \\ & \boldsymbol{\varphi}_2\end{bmatrix}\quad\text{and}\quad \boldsymbol{\varphi}_q=\begin{bmatrix}\boldsymbol{\varphi}_{q1} & \\ & \boldsymbol{\varphi}_{q2}\end{bmatrix}.$$ 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 $2\sigma_{m1}^2$ and $2\sigma_{m2}^2$, respectively, on the diagonals. Then, the AMLE objective function in (4.21) can be written in the following form:

$$L(\boldsymbol{\tau}, \mathbf{y}_m) = -\mathbf{g}(\boldsymbol{\tau}, \mathbf{y}_m)^T \mathbf{W} \mathbf{g}(\boldsymbol{\tau}, \mathbf{y}_m) \tag{4.22}$$

where

$$\mathbf{g}(\boldsymbol{\tau}, \mathbf{y}_m) = \begin{bmatrix} \mathbf{y}_m - \boldsymbol{\varphi}\boldsymbol{\beta} \\ \\ \dot{\boldsymbol{\varphi}}_q \boldsymbol{\beta} - \mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\beta}) \end{bmatrix} \tag{4.23}$$

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

$$diag(\mathbf{W}) = [diag(\boldsymbol{\Sigma}_1^{-1}), diag(\boldsymbol{\Sigma}_2^{-1}), diag(\frac{1}{2Q_{p1}}\mathbf{w}_1), diag(\frac{1}{2Q_{p2}}\mathbf{w}_2)].$$

Since, according to (4.22), the objective function is in a least-squares form, the Hessian matrix can be written as:

$$\frac{\partial^2 L(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} = \frac{\partial \mathbf{g}(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}^T}^T \left(\frac{\partial^2 L}{\partial \mathbf{g} \partial \mathbf{g}^T} \right) \frac{\partial \mathbf{g}(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}} + \sum_i \frac{\partial L}{\partial \mathbf{g}_i} \frac{\partial^2 \mathbf{g}_i}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} \tag{4.24}$$

If the second-order-derivative terms in the summation on the right-hand side of the above expression can be neglected, then equation (4.24) can be further simplified by the following approximation:

$$\frac{\partial^2 L(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} \cong -2 \frac{\partial \mathbf{g}(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}}^T \mathbf{W} \frac{\partial \mathbf{g}(\hat{\boldsymbol{\tau}}, \mathbf{y}_m)}{\partial \boldsymbol{\tau}} \tag{4.25}$$

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(4.23):

$$\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathbf{0} \\ -\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}} \end{bmatrix}, \quad \frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} -\boldsymbol{\varphi} \\ \dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}} \end{bmatrix} \tag{4.26}$$

Hence:

$$\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\tau}}\right)^T \mathbf{W} \left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\tau}}\right) = \begin{bmatrix} \left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right)^T \mathbf{w}_{12} \left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right) & -\left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right)^T \mathbf{w}_{12} \left(\dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right) \\ -\left(\dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right)^T \mathbf{w}_{12} \left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right) & \boldsymbol{\varphi}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varphi} + \left(\dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right)^T \mathbf{w}_{12} \left(\dot{\boldsymbol{\varphi}}_q - \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right) \end{bmatrix} \tag{4.27}$$

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

Note that, to obtain the inverse of the matrix in (4.27), the following property can be used (Seber and Wild, 1989).

If all inverses exist then:

$$\begin{aligned} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} &= \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21} & -\mathbf{B}_{12}\mathbf{B}_{22}^{-1} \\ -\mathbf{B}_{22}^{-1}\mathbf{B}_{21} & \mathbf{B}_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{C}_{11}^{-1} & -\mathbf{C}_{11}^{-1}\mathbf{C}_{12} \\ -\mathbf{C}_{21}\mathbf{C}_{11}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12} \end{bmatrix} \end{aligned} \tag{4.28}$$

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

4.8.2 Scatter plot matrices

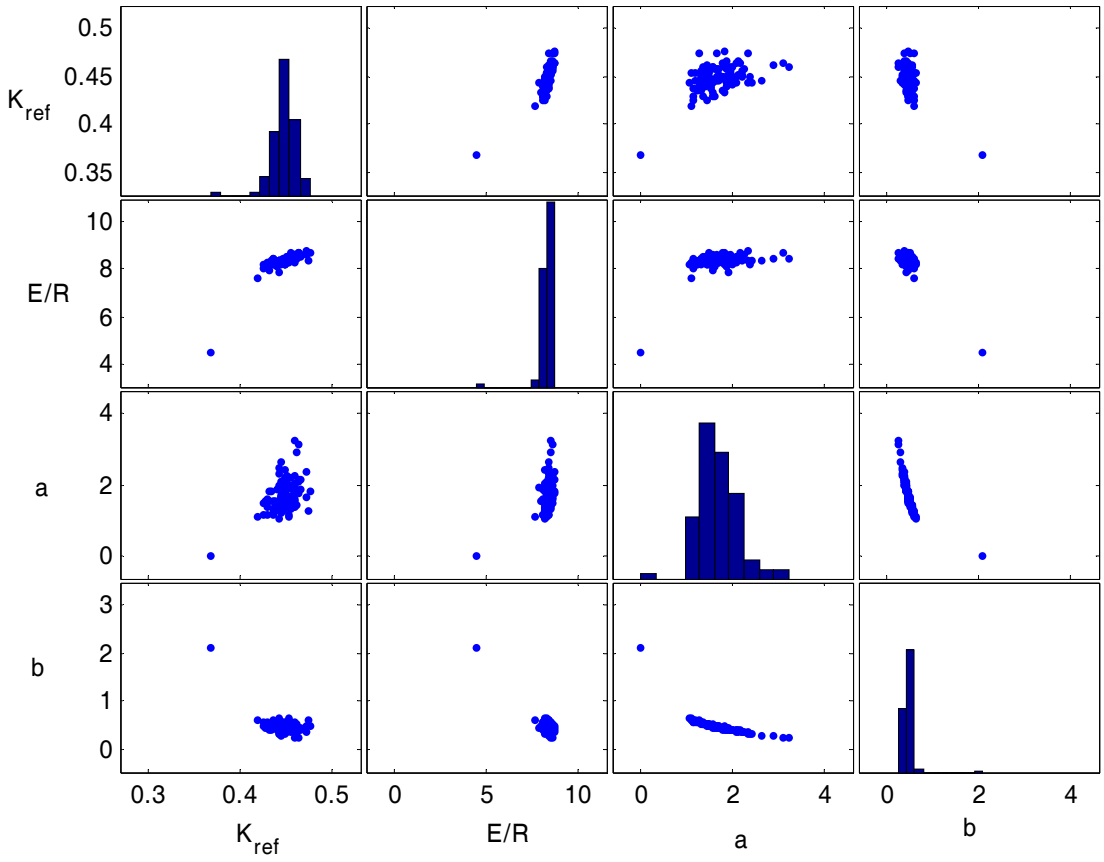

Figure 4.27. Scatter plot matrix for 100 sets of parameter estimates obtained using AMLE with both states measured

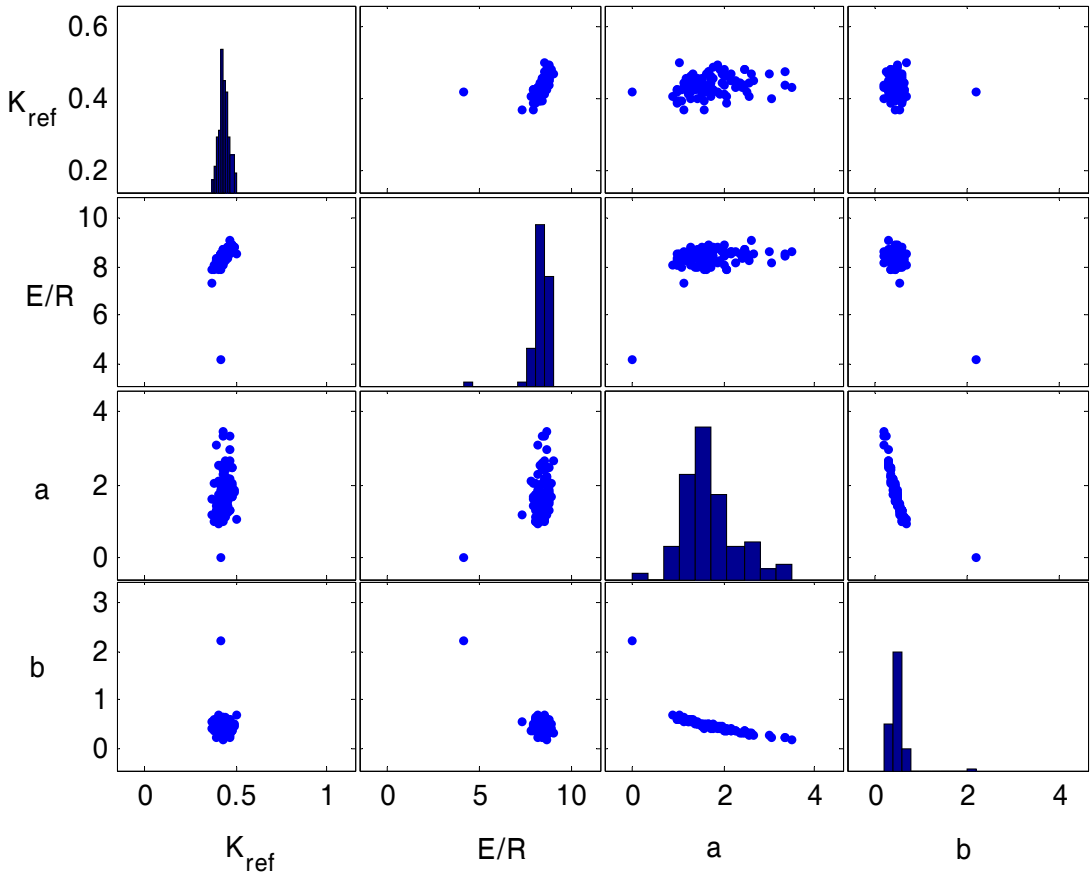

Figure 4.28. Scatter plot matrix for 100 sets of parameter estimates obtained using MLE without concentration measurements

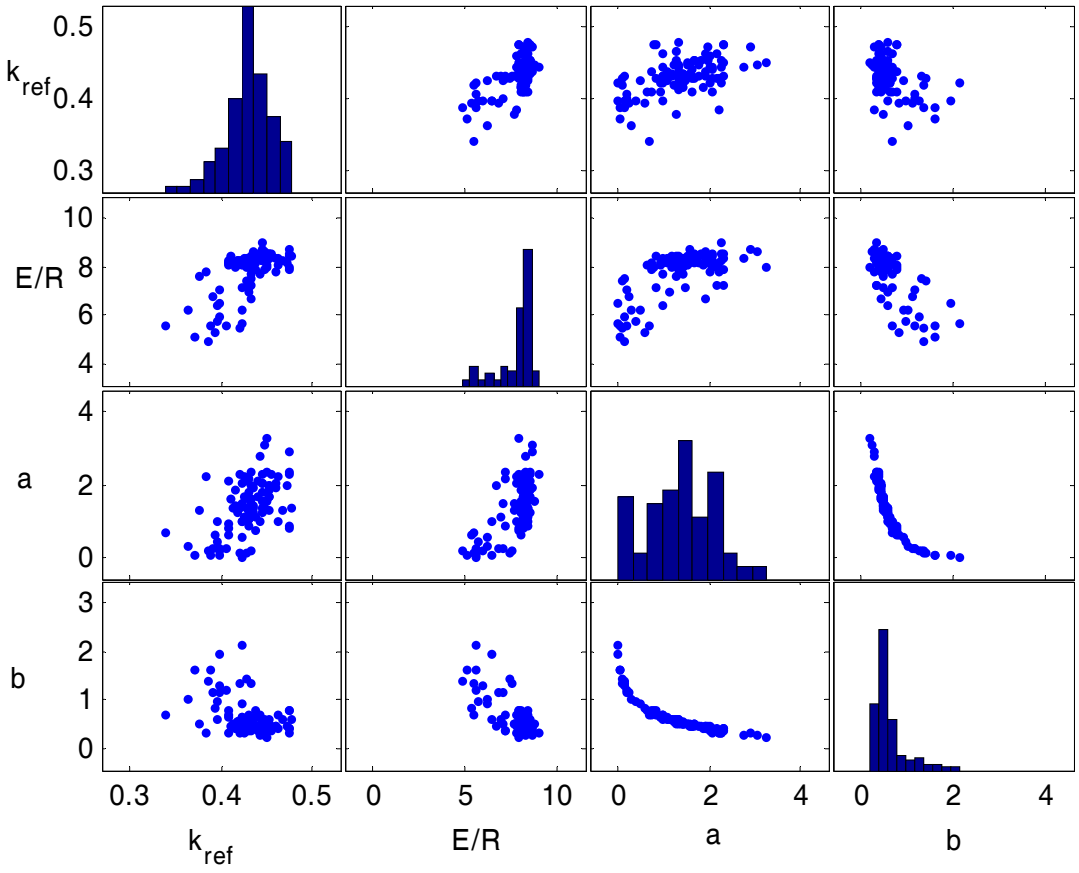

Figure 4.29. Scatter plot matrix for 100 sets of parameter estimates obtained from AMLE with both temperature and concentration measured and a non-stationary disturbance in the material balance

4.9 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 *Tref*	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_{mi}^2	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

4.10 References

Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, Inc., New York, 1974.

Y. Bar-Shalom. Optimal simultaneous state estimation and parameter identification in linear discrete-time systems. *IEEE Transactions on Automatic Control.* AC-17 (3), 1972, 308-319.

D.M. Bates, D.G. Watts. *Nonlinear Regression Analysis and Its applications.* John Wiley & Sons, Inc., New York, 1988.

L.T. Biegler, I.E. Grossman. Retrospective on optimization. *Computers and Chemical Engineering*. 28, 2004, 1169-1192.

L.T. Biegler. Short note solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering.* 8, 1984, 243-248.

C. de Boor. *A practical Guide to Splines.* Springer, New York, 2001.

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

L. Gagnon, J.F., MacGregor. State estimation for continuous emulsion polymerization. *Canadian Journal of Chemical En*gineering. 69, 1991, 648-656.

L. H. Hosten. A comparative study of short cut procedures for parameter estimation in differential equations. *Computers and Chemical Engineering.* 3, 1979, 117-126.

A.H. Jazwinski. *Stochastic Processes and Filtering theory*. Academic Press, New York, 1970.

S.M. Kay. *Fundamentals of Statistical Signal Processing*. Prentice Hall Signal Processing Series, 1993.

J.R. Leis, M.A. Kramer. ALGORITHM 658: ODESSA – an ordinary differential equation solver with explicit simultaneous sensitivity analysis. *ACM Transactions on Mathematical Software.* 14, 1988, 61-67.

T. E. Marlin. *Process Control: Designing Processes and Control Systems for Dynamic Performance, 2nd Edition.* McGraw-Hill, 2000.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 1*. Academic Press, New York, 1979.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 2*. Academic Press, New York, 1982.

R.E. Mortensen. Maximum-Likelihood recursive nonlinear filtering. *Journal of Optimization Theory and Applications.* 2 (6), 1968, 386.

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

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

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. *Computers and Chemical Engineering.* 30, 2006, 698-708.

J.O. Ramsay, G. Hooker, D. Campbell, J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of Royal Statistical Society*: *Series B* (*Statistical Methodology*). 69 (5), 2007, 741-796.

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

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

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

J. Swartz and H. Bremermann. Discussion of parameter estimation in biological modelling: algorithms for estimation and evaluation of estimates. *Journal of Mathematical Biology.* 1, 1975, 241-275.

Y.P. Tang. On the estimation of rate constants for complex kinetic models. *Industrial and Engineering Chemistry Fundamentals*. 10, 1971, 321-322.

S. Vajda, P. Valko. A direct-indirect procedure for estimation of kinetic parameters. *Computers and Chemical Engineering.* 10 (1), 1986, 49-58.

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

M.S. Varziri, A.A. Poyton, K.B. McAuley, P.J. McLellan, J.O. Ramsay. Selecting optimal weighting factors in iPDA for parameter estimation in continuous-time dynamic models. Accepted, to appear in *Computers and Chemical Engineering*. 2008.

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

A. Yeredor. The joint MAP-ML criterion and its relation to ML and extended Least-Squares, *IEEE Transactions on Signal Processing*. 48 (12), 2000, 3484-3492.

Chapter 5

Parameter and State Estimation in Nonlinear Stochastic Continuous-Time Dynamic Models with Unknown Disturbance Intensity

5.1 Abstract

In the development of AMLE in Chapter 3 (Varziri et al., 2008b) and Chapter 4 (Varziri et al. 2008a) it was assumed that the process-disturbance intensities and measurement-noise variances in the stochastic dynamic models are known. In this chapter, 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. Approximate 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.

This chapter has been accepted for publication as a journal paper in *Canadian Journal of Chemical Engineering*.

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, ON, Canada K7L 3N6

5.2 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. 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. Measurement errors, which arise due to sensor inaccuracy and fluctuations, only influence the measured outputs at the current time. Process disturbances, 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; Seber and Wild, 1989; Bates and Watts, 1988; Ogunnaike and Ray, 1994). Robust and efficient methods such as multiple shooting (Bock, 1981 and 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). In engineering applications, Kalman-filter-related algorithms are popular. These methods generally require the time-varying state covariance matrix. If the model is linear, the time evolution of the state covariance matrix can be described by a set of ODEs (Maybeck, 1979). However, if the model is nonlinear, the state covariance matrix satisfies a system of partial differential equations (PDEs) (the forward Kolmogorov equation) that is generally hard to solve. Alternatively, the model can be linearized and then treated as a linear SDE parameter estimation problem. 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 (Evensen, 2003). These methods, however, remain computationally intensive.

Maximum Likelihood (ML) methods have also 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.*, Kristensen et al., 2004). The density function used in the classical ML method is generally hard to obtain because it involves integrating out the unobserved states. It is possible to avoid the integration in the classical ML method by using the Expectation Maximization (EM) method (Roweis and Ghahramani, 2001). 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 still computationally intensive. 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. 2008b, 2008a), (Chapter 3 and Chapter 4), 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 (5.8)) 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, 1998), 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 Section 3.2.1, the AMLE algorithm is briefly reviewed. Estimation of process disturbance intensities is discussed in Section 5.4. In Section 5.5, the nonlinear CSTR case study is presented, followed by the summary and conclusions in Section 5.6.

5.3 AMLE Fitting Criterion

In the following paragraphs we briefly review the AMLE algorithm. 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. 2008b and 2008a) (Chapter 3 and Chapter 4).

Consider the following continuous-time stochastic dynamic model*:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), \boldsymbol{\theta}) + \eta(t) \\ x(t_0) &= x_0 \end{aligned} \qquad (5.1)$$

$$y(t_{mj}) = x(t_{mj}) + \varepsilon(t_{mj})$$

$x \in \Re$ is the state variable, $u \in \Re$ is the input variable and $y \in \Re$ is the output variable. $\boldsymbol{\theta} \in \Re^p$ is the vector of model parameters and $f: \Re \times \Re \times \Re^p \rightarrow \Re$ 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 eq. (3.9) has a unique solution. ε is a zero-mean uncorrelated Normal random variable with variance σ_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 power spectral density and $\delta(.)$ is the Dirac delta function. The random noise trajectory $\eta(t)$ is a series of random steps with a switching time of Δt, where $\Delta t \rightarrow 0$. For the corresponding discrete-time white-noise process (Maybeck, 1979):

---

* This dynamic model can be more rigorously written as: $dx(t) = f(x(t), u(t), \boldsymbol{\theta})dt + d\omega(t)$ where $d\omega(t)$ is the increment of a Wiener process (Maybeck, 1979).

$$E\{\eta(j_1\Delta t)\eta(j_2\Delta t)\} = \begin{cases} \dfrac{Q}{\Delta t} & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases} \tag{5.2}$$

where j_1 and j_2 are integers and Δt is the sampling period. We also assume that the process disturbance $\eta(t)$ and the measurement noise $\varepsilon(t)$ are not correlated.

At the discrete time t_i, where $t_i = t_{i-1} + \Delta t$ and the sampling interval Δt is small, eq. (5.1) can be written using the following Euler approximation.

$$x(t_{i-1} + \Delta t) = x(t_i) = x(t_{i-1}) + f(x(t_{i-1}), u(t_{i-1}), \boldsymbol{\theta})\Delta t + \eta(\mathrm{t_{i\text{-}1}})\Delta t \tag{5.3}$$

Note that the variance of random noise term $\eta(\mathrm{t_{i\text{-}1}})\Delta t$ is $\mathrm{Q}\,\Delta t$. 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, $i = 0 \cdots q$ 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_is for $i = 0 \cdots q$. 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} ($j = 1 \cdots n$). The measurement times t_{mj} do not need to be uniformly spaced. The vector of outputs at observation times $y(t_{mj})$ ($j = 1 \cdots n$) and its corresponding state vector of true values $x(t_{mj})$ ($j = 1 \cdots n$) and measurement noise vector $\varepsilon(t_{mj})$ ($j = 1 \cdots n$) are denoted by $\mathbf{y}_m$, $\mathbf{x}_m$, and $\boldsymbol{\varepsilon}_m$ respectively.

---

* Please note that $\sigma_p^2 = \mathrm{var}(\eta(\mathrm{t_{i\text{-}1}})\Delta t)$ in this chapter, is different from σ_p^2 used in Chapter 3 which was used to denote $\mathrm{var}(\eta(\mathrm{t_{i\text{-}1}}))$.

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

1. $p(\mathbf{y}_m \mid \boldsymbol{\theta})$
2. $p(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}_m)$
3. $p(\mathbf{x} \mid \mathbf{y}_m, \boldsymbol{\theta})$
4. $p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta})$

The first density function 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 $\boldsymbol{\theta}$ 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:

$$p(\mathbf{x} \mid \mathbf{y}_m, \boldsymbol{\theta}) \times p(\mathbf{y}_m \mid \boldsymbol{\theta}) = p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta}) \tag{5.4}$$

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

Using eqns. (5.1) and (5.3), the density function $p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta})$ can be written as (Varziri et al., 2008b) (Chapter 3):

$$p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta}) = \frac{1}{2\pi^{n/2}\sigma_m^n} \exp\left(-\frac{SSE}{2\sigma_m^2}\right) \times \frac{1}{2\pi^{q/2}\sigma_p^q} \exp\left(-\frac{PEN}{2\sigma_p^2}\right) \tag{5.5}$$

where

$$\begin{aligned} SSE &= (\mathbf{y}_m - \mathbf{x}_m)^T (\mathbf{y}_m - \mathbf{x}_m) \\ PEN &= \sum_{i=1}^{q} \left(x_i - x_{i-1} - f(x_{i-1}, u_{i-1}, \boldsymbol{\theta})\Delta t\right)^2 \\ \sigma_p^2 &= Q\Delta t \end{aligned} \tag{5.6}$$

If Q and σ_m^2 are known, (5.5) can be simplified to

$$p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta}) = C_1 \times \exp\left(-\frac{SSE}{2\sigma_m^2}\right) \times \exp\left(-\frac{PEN}{2\sigma_p^2}\right) \tag{5.7}$$

where C_1 depends on σ_m^2 and σ_p^2, but not on the model parameters or the states. Note that when $\Delta t \to 0$, then $C_1 \to \infty$. To avoid this problem, the probability density functional (Jazwinski, 1970) is defined as: $\lim_{\Delta t \to 0} \frac{p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta})}{C_1}$. It can then be shown that optimal state and parameter estimates are obtained by minimizing the negative of the natural logarithm of the probability density functional as follows:

$$\frac{(\mathbf{y}_m - \mathbf{x}_m)^T (\mathbf{y}_m - \mathbf{x}_m)}{2\sigma_m^2} + \frac{1}{2Q} \int_{t_0}^{t_q} \left(\dot{x}(t) - f(x(t), u(t), \boldsymbol{\theta})\right)^2 dt \tag{5.8}$$

Since $x(t)$ is an unknown function, minimizing (5.8) 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; Varziri et al., 2008a & 2008b, Ramsay et al., 2007):

$$x(t) \approx x_{\sim}(t) = \sum_{i=1}^{c} \beta_i \phi_i \tag{5.9}$$

where β_i, $i = 1 \cdots c$ are B-spline coefficients and $\phi_i(t)$ $i = 1 \cdots c$ are B-spline basis functions (de Boor, 2001) Note that eq. (5.9) can be written in matrix form:

$$x_{\sim}(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\beta} \tag{5.10}$$

where $\boldsymbol{\varphi}(t)$ 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:

$$\dot{x}_{\sim}(t) = \frac{d}{dt}\left(\sum_{i=1}^{c} \beta_i \phi_i(t)\right) = \sum_{i=1}^{c} \beta_i \dot{\phi}_i(t) = \dot{\boldsymbol{\varphi}}^T \boldsymbol{\beta} \tag{5.11}$$

By substituting (5.9) and (5.11) into (5.8) we have the following finite-dimensional optimization problem:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\theta},\boldsymbol{\beta}}{\arg\min}\left\{ \frac{(\mathbf{y}_m - \mathbf{x}_{\sim m})^T (\mathbf{y}_m - \mathbf{x}_{\sim m})}{2\sigma_m^2} + \frac{1}{2Q}\int_{t_0}^{t_q} \left(\dot{x}_{\sim}(t) - f(x_{\sim}(t), u(t), \boldsymbol{\theta})\right)^2 dt \right\} \tag{5.12}$$

Minimizing (5.12) 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., 2008a) (Chapter 4). The AMLE criterion in nonlinear problems may produce 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). The AMLE criterion is also much easier to formulate. Eq. (5.8) was

derived based on the assumption that σ_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, $\boldsymbol{\theta}$.

5.4 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-time dynamic system. As mentioned in the introduction, using the classical ML criterion is not straightforward because the density function to be maximized is: $p(\mathbf{y}_m \mid \boldsymbol{\theta}, \sigma_m, \sigma_p) = \int p(\mathbf{y}_m, \mathbf{x} \mid \boldsymbol{\theta}, \sigma_m, \sigma_p) d\mathbf{x}$ 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:

$$\hat{\sigma}_m^2 = \frac{SSE}{n - \gamma_m} \tag{5.13}$$

where $\gamma_m = (1/\hat{\sigma}_m^2) Trace(\mathbf{A}^{-1})$ and $\mathbf{A}$ is the Hessian matrix of $-\ln p(\mathbf{x} \mid \mathbf{y}_m, \boldsymbol{\theta}, \sigma_m, \sigma_p)$ with respect to $\mathbf{x}$ evaluated at $\arg\min_{\mathbf{x}} \{-\ln p(\mathbf{x} \mid \mathbf{y}_m, \boldsymbol{\theta}, \sigma_m, \sigma_p)\}$. Substituting γ_m in (5.13) and solving for $\hat{\sigma}_m^2$ we have:

$$\hat{\sigma}_m^2 = \frac{SSE}{n} + \frac{Trace(\mathbf{A}^{-1})}{n} \tag{5.14}$$

If we try to obtain an estimate of σ_m^2 by maximizing (5.5) (or by minimizing $-\ln p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta})$) we will get $\hat{\sigma}_m^2 = SSE/n$ which lacks the second term in (5.14); this would have been a good estimate if the true state trajectory, rather than the estimated state trajectory, $\mathbf{x}_{\sim}$, was used to

calculate *SSE*. The term, $Trace(\mathbf{A}^{-1})/n$ can therefore be thought of as a compensation for deviations of $\mathbf{x}_{\sim}$ 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 $\mathbf{x}_{\sim}$. Having equation (5.14) at our disposal and assuming that σ_m^2 is known, we can pick an initial estimate of Q and solve the minimization problem in (5.12) to get $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}$, and therefore $x_{\sim}$. The estimates obtained can then be used to evaluate $\hat{\sigma}_m^2$ using (5.14). Next, a new estimate of Q is obtained to minimize the loss function $\left(\hat{\sigma}_m^2 / \sigma_m^2 - 1\right)^2$. This two-step optimization scheme can be summarized as follows:

Outer optimization problem:

$$\hat{Q} = \arg\min_{Q} \left(\frac{(\mathbf{y}_m - \mathbf{x}_{\sim m}(Q))^T (\mathbf{y}_m - \mathbf{x}_{\sim m}(Q))}{n\sigma_m^2} + \frac{Trace(\mathbf{A}^{-1})}{n\sigma_m^2}\bigg|_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\beta}}} - 1 \right)^2 \tag{5.15}$$

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

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\theta},\boldsymbol{\beta}} \left\{ \frac{(\mathbf{y}_m - \mathbf{x}_{\sim m})^T (\mathbf{y}_m - \mathbf{x}_{\sim m})}{2\sigma_m^2} + \frac{1}{2\hat{Q}} \int_{t_0}^{t_q} \left(\dot{x}_{\sim}(t) - f(x_{\sim}(t), u(t), \boldsymbol{\theta}) \right)^2 dt \right\} \tag{5.16}$$

The overall minimization problem presented in (5.15) and (5.16) 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 (5.12) using an estimate of Q obtained from the outer optimization. The results of this overall optimization problem provide the modeller with $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}$ and $\hat{Q}$. $\hat{\boldsymbol{\theta}}$ is the desired estimate for the fundamental model parameters. The estimated spline coefficients $\hat{\boldsymbol{\beta}}$, 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.

5.5 Simulation Case Study: Nonlinear CSTR

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

$$\begin{aligned}
&\frac{dC_A(t)}{dt} = \frac{F(t)}{V}\left(C_{A0}(t) - C_A(t)\right) - gC_A(t) + \eta_1(t)\\
&\frac{dT(t)}{dt} = \frac{F(t)}{V}\left(T_0(t) - T(t)\right) + \beta_1\left(T(t) - T_{cin}(t)\right) + \beta_2 gC_A(t) + \eta_2(t)\\
&C_A(0) = 1.569\,(\text{kmol m}^{-3})\\
&T(0) = 341.37\,(\text{K})\\
&y_1(t_i) = C_A(t_i) + \varepsilon_1(t_i)\\
&y_2(t_j) = C_A(t_j) + \varepsilon_2(t_j)
\end{aligned} \tag{5.17}$$

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

where $E\{\eta_1(t_i)\eta_1(t_j)\} = Q_{p1}\delta(t_i - t_j)$, $E\{\eta_2(t_i)\eta_2(t_j)\} = Q_{p2}\delta(t_i - t_j)$ ($\delta(.)$ 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. 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. 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_{ref} = 0.461 min^{-1}, a=1.678E6,

$b = 0.5$. 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_{cin}, and the coolant flow rate F_c. Values for the various other known constants are as follows: $V = 1.0$ m^3, $C_p = 1\,\text{cal g}^{-1}\text{K}^{-1}$, $\rho = 1\text{E}6$ g m^{-3}, $C_{pc} = 1\,\text{cal g}^{-1}\text{K}^{-1}$, $\rho_c = 1\text{E}6\,\text{gm}^{-3}$, and $-\Delta H_{rxn} = 130\text{E}6$ 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 5.1 (Poyton, 2005).

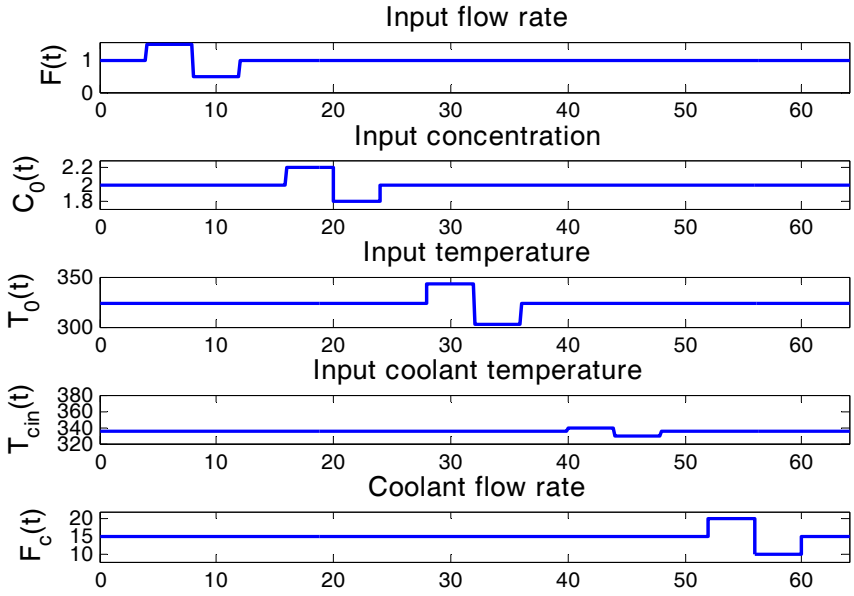

Figure 5.1. Input scheme for MIMO nonlinear CSTR

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., 2008b). 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 used in the simulations are $Q_{p1} = 4\times10^{-3}\ (\text{kmol/m}^3)^2/\text{min}$ and $Q_{p2} = 4\ \text{K}^2/\text{min}$. We assume that measurement noise variances are known but the process disturbance intensities are unknown. From eq. (5.15) the appropriate outer objective function is $\left(J_{1C} + J_{1T}\right)$ shown below in eq. (5.18), which is minimized with respect to Q_{p1} and Q_{p2}, and the inner objective function is $\left(J_{2C} + J_{2T}\right)$ which is minimized with respect to model parameters and states, given estimates for Q_{p1} and Q_{p2}:

$$
\begin{aligned}
J_{1C} &= \frac{1}{64\sigma_{m1}^2}\left(\sum_{j=1}^{64}\left(y_1(t_{m1j}) - C_{A\sim}(t_{m1j})\right)^2 + Trace\left(\left(\frac{\partial^2 J_{2C}}{\partial C_{A\sim}{}^2}\right)^{-1}\right)\right) - 1 \\
J_{2C} &= \frac{1}{2\sigma_{m1}^2}\sum_{j=1}^{64}\left(y_1(t_{m1j}) - C_{A\sim}(t_{m1j})\right)^2 + \\
&\quad \frac{1}{2Q_{p1}}\int\left(\dot{C}_{A\sim}(t) - \left(\frac{F(t)}{V}\left(C_{A0}(t) - C_{A\sim}(t)\right) + gC_{A\sim}(t)\right)\right)^2 dt \\
J_{1T} &= \frac{1}{213\sigma_{m2}^2}\left(\sum_{j=1}^{213}\left(y_2(t_{m2j}) - T_{\sim}(t_{m2j})\right)^2 + Trace\left(\left(\frac{\partial^2 J_{2T}}{\partial T_{A\sim}{}^2}\right)^{-1}\right)\right) - 1 \\
J_{2T} &= \frac{1}{2\sigma_{m2}^2}\sum_{j=1}^{213}\left(y_2(t_{m2j}) - T_{\sim}(t_{m2j})\right)^2 + \\
&\quad \frac{1}{2Q_{p2}}\int\left(\dot{T}_{\sim}(t) - \left(\frac{F(t)}{V}\left(T_0(t) - T_{\sim}(t)\right) - \beta_1\left(T_{\sim}(t) - T_{cin}(t)\right) + \beta_2 gC_{A\sim}(t)\right)\right)^2 dt
\end{aligned}
\tag{5.18}
$$

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™.

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 Figure 5.2, Figure 5.3, and Figure 5.4, respectively. The histograms in Figure 5.3 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 5.4 indicates a strong nonlinear co-dependency between estimates for parameters *a* and *b*. There is noticeable linear co-dependency between estimates for *a* and *kref* 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 *kref* 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-α)% confidence intervals for the model parameters and spline coefficients can be obtained as follows (*e.g.,* Varziri et al. 2008a):

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} \pm z_{\alpha/2} \times \sqrt{\mathrm{diag}(\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}))} \tag{5.19}$$

where $\mathbf{I}$ is the Fisher information matrix for the inner objective function $\left(J_{2C} + J_{2T}\right)$ in eq. (5.18). 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 5.1. The estimated noise intensities for this run are $\hat{Q}_{p1} = 2.43 \times 10^{-3} (\mathrm{kmol/m^3})^2 / \mathrm{min}$ and $\hat{Q}_{p2} = 2.87 \, (\mathrm{K^2/min})$. 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 5.5. 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, $p(\mathbf{y}_m \mid \boldsymbol{\theta})$. The estimation was carried out using the CSTM software developed by Kristensen and Madsen (2003). This software is publicly available and easy to use. The estimated noise intensities obtained using the CSTM package in this case are $\hat{Q}_{p1} = 3.00 \times 10^{-3} (\text{kmol/m}^3)^2 / \text{min}$ and $\hat{Q}_{p2} = 2.63 (\text{K}^2/\text{min})$. The rest of the results are presented in Table 5.2. 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 Ricatti 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.

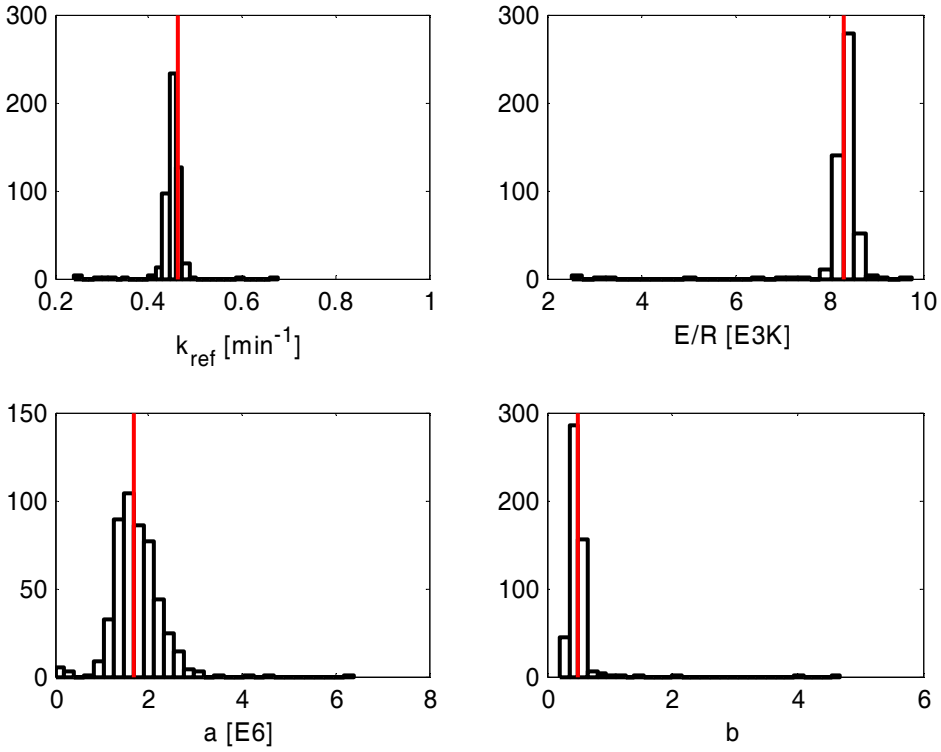

Figure 5.2. Histograms for AMLE parameter estimates with unknown disturbance intensity

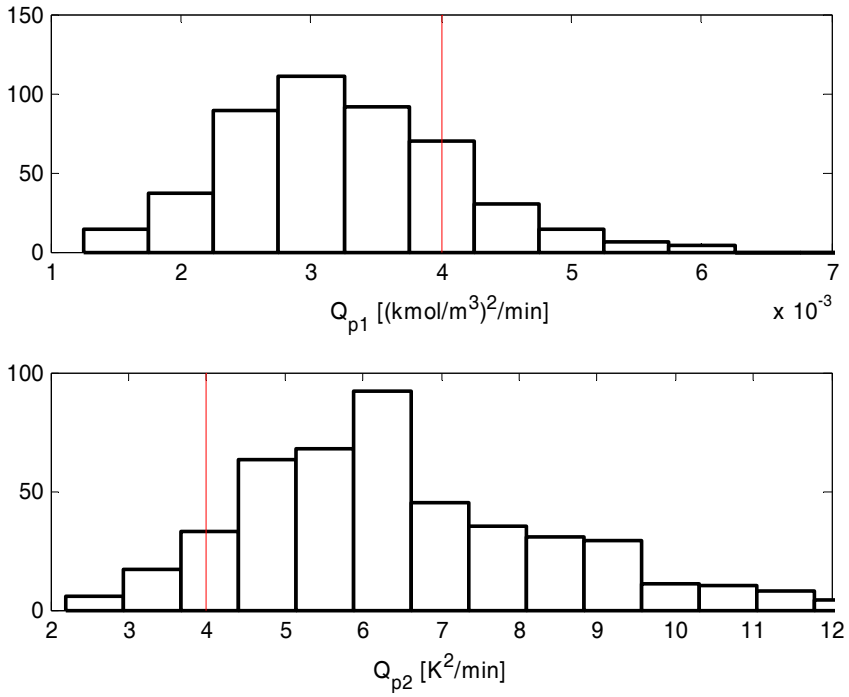

Figure 5.3. Histograms for AMLE process intensity estimates

Table 5.1. 95% Confidence Intervals for AMLE model parameter estimates from one of the 500 Monte Carlo simulations

Parameter	True Value	Estimate	Std. Dev.	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	341.38	340.59	0.54	339.54	341.64

Table 5.2. 95% Confidence Intervals for classical ML model parameter estimates (obtained using CSTM algorithm of Kirstensen et al., (2004))

Parameter	True Value	Estimate	Std. Dev.	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	341.38	340.55	0.79	339.00	342.10

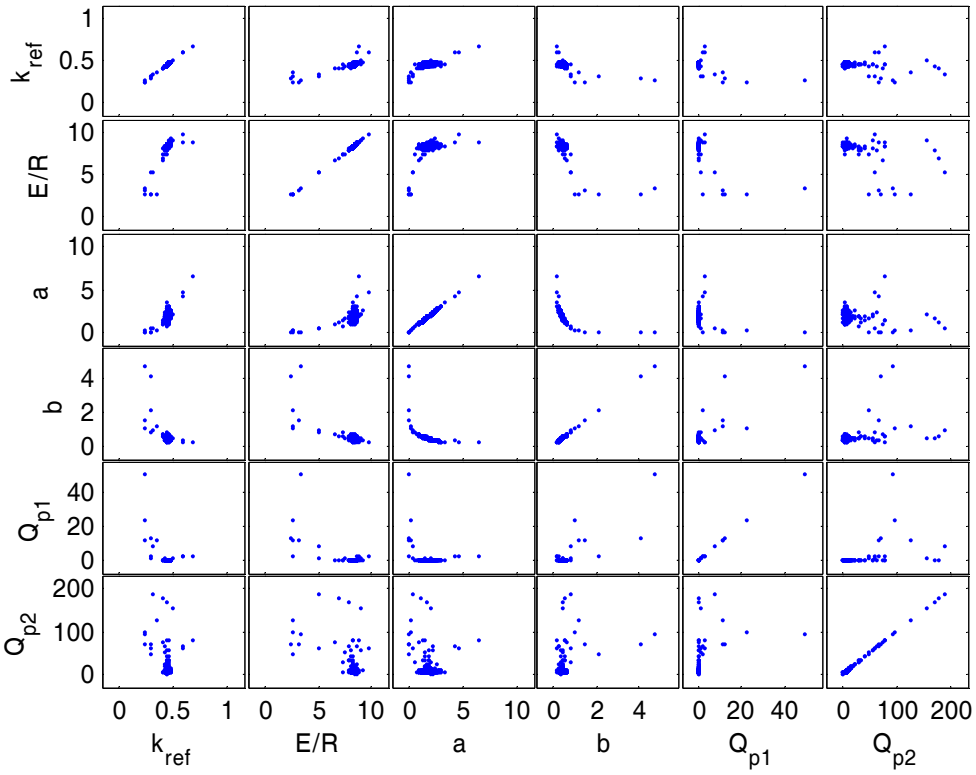

Figure 5.4. Scatter plot matrix for AMLE estimates

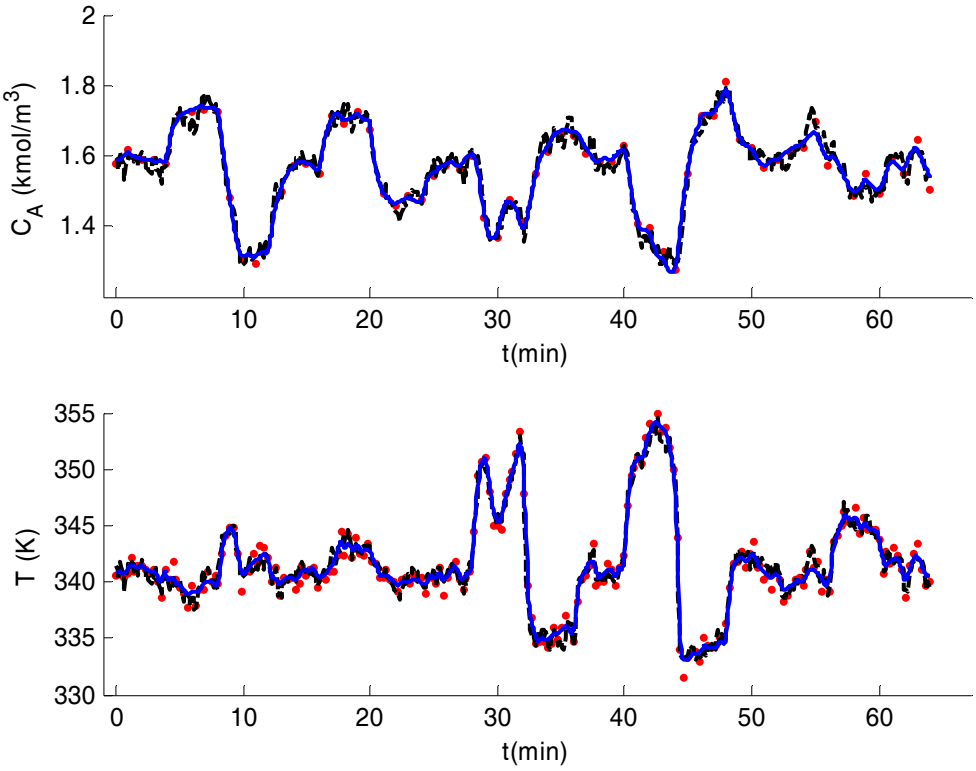

Figure 5.5. Estimated B-Spline trajectories of C_A and T, for one of the 500 considered samples using AMLE (___, B-spline fit; •, measured data; ---- response with true parameters and true stochastic disturbances)

5.6 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. (2008b, 2008a), 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 $\boldsymbol{\theta}$ and states $\mathbf{x}$ by maximizing $p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta})$ given the disturbance intensity Q. Setting up the proposed algorithm is much easier than classical likelihood methods in which $p(\mathbf{y}_m \mid \boldsymbol{\theta})$ is maximized. Using the likelihood corresponding to $p(\mathbf{x}, \mathbf{y}_m \mid \boldsymbol{\theta})$, 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, they can be better in the sense of mean-square-error, than classical ML estimates, especially when a small number of measured data is used. 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. The AMLE algorithm converged faster than 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 non-stationary 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.

5.7 Acknowledgments

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.

5.8 Appendix

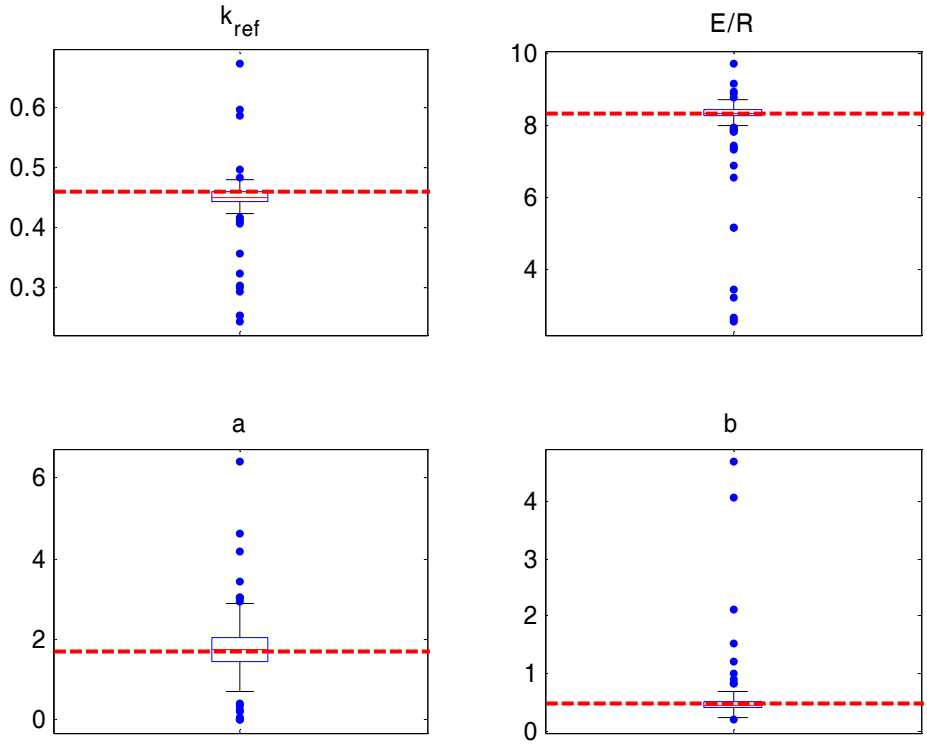

Figure 5.6. Boxplots for AMLE parameter estimates

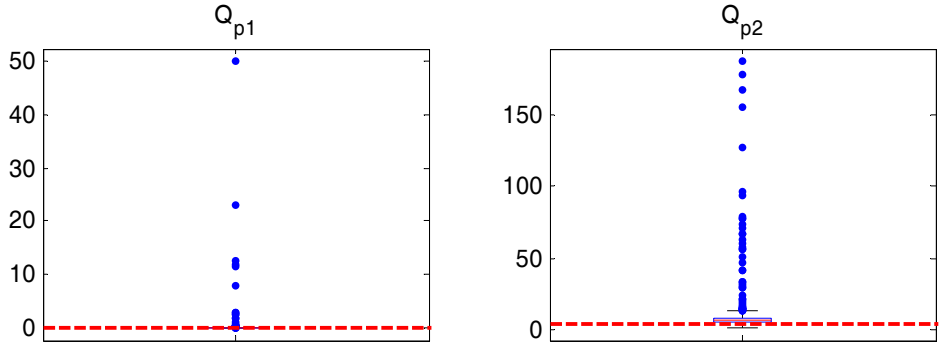

Figure 5.7. Boxplots for AMLE process intensity estimates

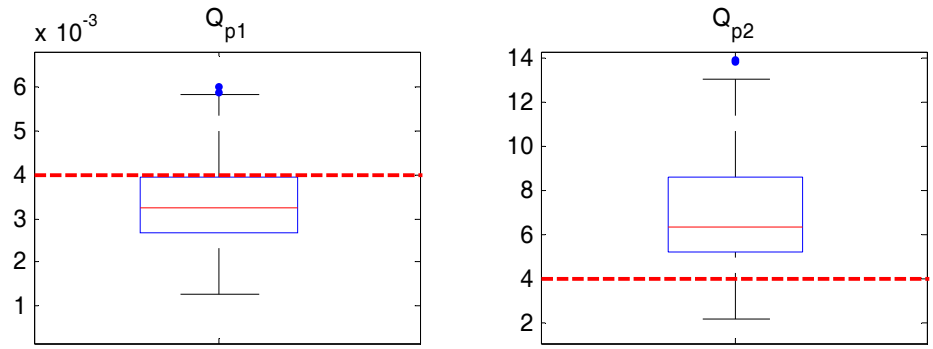

Figure 5.8. Zoomed boxplots for AMLE process intensity estimates

5.9 Nomenclature

a	CSTR model parameter relating heat-transfer coefficient to coolant flow	
b	CSTR model exponent relating heat-transfer coefficient to coolant flow	
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 $g^{-1}K^{-1}$
$E\{.\}$	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	Nonlinear function	
$\mathbf{I}$	Fisher information matrix	
k_{ref}	Kinetic rate constant at temperature *Tref*	min^{-1}
l	Likelihood function	
N_i	Number of observations of state i	
$p(.)$	Probability density function	
t_{j}	j-th measurement time	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}	Input to the differential equation for state i	

V	Volume of the reactor	m^3
x, $\mathbf{x}$	State variables	
$x_{\sim}$	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 $\alpha/2$	
α	Significance level for confidence intervals	
β_i	i-th B-spline coefficient	
$\boldsymbol{\beta}$	Vector of B-spline coefficients	
$\delta(.)$	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	
$\boldsymbol{\theta}$	Vector of model parameters	
ρ	Density of reactor contents	g m^{-3}
ρ_c	Coolant density	g m^{-3}
σ_{mi}^2	Measurement noise variance for the i-th state	
σ_{pi}^2	Process noise intensity of stochastic differential equation for state i	
ϕ_i	i-th B-spline basis function	
$\boldsymbol{\varphi}$	Matrix containing all ϕ_i s	
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

5.10 References

Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, Inc., New York, 1974.

D.M. Bates, D.G. Watts. *Nonlinear Regression Analysis and Its applications.* John Wiley & Sons, Inc., New York, 1988.

L.T. Biegler, I.E. Grossman. Retrospective on optimization. *Computers and Chemical Engineering*. 28, 2004, 1169-1192.

L.T. Biegler. Short note solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering.* 8, 1984, 243-248.

H.G. Bock. Numerical treatment of inverse problems in chemical reaction kinetics, in *Modelling of Chemical Reaction Systems*. K.H. Ebert, P. Deuflhard, and W. Jäger, editors. Springer, Berlin. Chapter 8, 1981, 102-125.

H.G. Bock. Recent advances in parameter identification for ordinary differential equations. In *Progress in Scientific Computing*. P. Deuflhard and E. Hairer, editors. Birkhäuser, Boston. 1983, 95-121.

C. de Boor. *A practical Guide to Splines.* Springer, New York, 2001.

D.P. Dee. On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*. 123 (4), 1995, 1128-1145.

D.P. Dee, A.M. Da Silva. Maximum-Likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Monthly Weather Review*. 127 (8), 1999, 1822-1834.

G. Evensen. The ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics.* 53, 2003, 343-367.

J. Gong, G. Wahba, D.R. Johnson, J. Tribbia. Adaptive tuning of numerical weather prediction models: simultaneous estimation of weighting, smoothing, and physical parameters. *Monthly Weather Review*. 126, 1998, 210-231.

J.P.M. Heald, J. Stark. Estimation of noise levels for models of chaotic dynamical systems. *Phisycal Review Letters*. 84, 2000, 2366-2369.

A.H. Jazwinski. *Stochastic Processes and Filtering theory*. Academic Press, New York, 1970.

S.J. Julier, J.K. Uhlmann. Unscented filtering and nonlinear estimation, *Proceedings of the IEEE.* 92 (3), 2004, 401-422.

P.E. Kloeden, E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, 1992.

N.R. Kristensen, H. Madsen, S.B. Jorgensen. *Continuous time stochastic modelling-CSTM 2.3.* Technical University of Denmark, Lyngby, Denmark. 2003.

N.R. Kristensen, H. Madsen, S.B. Jorgensen. Parameter estimation in stochastic grey-box models. *Automatica*. 40, 2004, 225-237.

D.J.C. Mackay. *Information Theory, Inference, and Learning Algorithms*, version 7. Cambridge University Press, 2004.

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

P.S. Maybeck, Combined Estimation of States and Parameters for On-Line Applications. Ph.D. thesis, MIT, Cambridge, Massachusetts, 1972.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 1*. Academic Press, New York, 1979.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 2*. Academic Press, New York, 1982.

J.N. Nielsen, H. Madsen, P.C. Young. Parameter estimation in stochastic differential equations: an overview. *Annual Reviews in Control.* 24, 2000, 83-94.

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

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

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. *Computers and Chemical Engineering.* 30, 2006, 698-708.

J.O. Ramsay, G. Hooker, D. Campbell, J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of Royal Statistical Society*: *Series B* (*Statistical Methodology*). 69 (5), 2007, 741-796.

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

S. Roweis, Z. Ghahramani. Learning nonlinear dynamical systems using the Expectation-Maximization algorithm. In *Kalman Filtering and Neural Networks*. S. Haykin, editor. John Wiley & Sons, New York. 2001, 175-220.

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

M.E. Solari. The "Maximum Likelihood Solution" of the problem of estimating a linear functional relationship. *Journal of Royal Statistical Society*: *Series B* (*Statistical Methodology*). 31, 1969, 372-375.

V. Solo. A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems. *In Proc. IEEE ICIP96.IEEE, IEEE press.*1996, 89-92.

M.S. Varziri, K.B. McAuley, P.J. McLellan. "Parameter estimation in continuous-time dynamic models in the presence of unmeasured states and non-stationary disturbances". *Industrial Engineering and Chemistry Research*. 47, 2008a, 380-393.

M.S. Varziri, A.A. Poyton, K.B. McAuley, P.J. McLellan, J.O. Ramsay. Selecting optimal weighting factors in iPDA for parameter estimation in continuous-time dynamic models. Accepted, to appear in *Computers and Chemical Engineering.* 2008b.

H.U. Voss, J. Timmer, J. Kurths. Nonlinear dynamical system identification from uncertain and indirect measurements. *International Journal of Bifurcation and Chaos.* 14 (6), 2004, 1905-1933.

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

J.J. Warnes, B.D. Ripley. Problems with likelihood estimation of covariance functions of spatial Gaussian processes. *Biometrika*, 74 (3), 1987, 640-642.

A. Yeredor. The joint MAP-ML criterion and its relation to ML and extended Least-Squares, *IEEE Transactions on Signal Processing*. 48 (12), 2000, 3484-3492.

Chapter 6

Approximate Maximum Likelihood Parameter Estimation for Nonlinear Dynamic Models: Application to a Lab-Scale Nylon Reactor Model

6.1 Abstract

In the previous chapters, AMLE was developed and tested using a simple two-state CSTR simulation study. The main objective of this chapter is to evaluate the performance of AMLE by applying this algorithm to a practical chemical engineering example.

First, a lab-scale nylon 612 reactor model, initially developed by Schaffer et al. (2003b) and subsequently modified by Zheng et al. (2005) and Campbell (2007), is re-evaluated using additional data so that the model structure can be further improved. Then the AMLE algorithm is used for parameter and state estimation in the nylon reactor model. The nylon reactor model equations are represented by stochastic differential equations (SDEs) to account for any modelling errors or unknown process disturbances that enter the reactor system during experimental runs. In this chapter, we demonstrate that AMLE can address difficulties that frequently arise when estimating parameters in nonlinear continuous-time dynamic models of industrial processes. Among these difficulties are: different types of measured responses with different levels of measurement noise, measurements taken at irregularly-spaced sampling times, unknown initial conditions for some state variables, unmeasured state variables, and unknown disturbances that enter the process and influence its future behaviour.

This chapter was submitted as a journal paper to *Industrial Engineering and Chemistry Research*.

Approximate Maximum Likelihood Parameter Estimation for Nonlinear Dynamic Models: Application to a Lab-Scale Nylon Reactor Model

M.S. Varziri, K. B. McAuley and P. J. McLellan, Department of Chemical Engineering, Queen's University, Kingston, ON, Canada K7L 3N6

6.2 Introduction

Parameter estimation in dynamic models that are described by a combination of nonlinear algebraic and differential equations is a challenging problem. The complexity of the problem increases significantly if it is acknowledged that there are two different types of random errors that influence the measurements obtained from dynamic processes: measurement errors and process disturbances. Measurement errors are problematic because they can make it difficult for modellers to obtain reliable parameter estimates, but random process disturbances can be even more problematic, because they influence the future behaviour of the process and therefore future measurements of process outputs. For example, consider an unknown disturbance that influences the temperature in a chemical reactor. The change in temperature can alter the rates of chemical reactions and can influence several different types of process measurements and how they change over time. Modellers often have knowledge about the quality of the measurements that are available for parameter estimation (*e.g.*, good estimates of measurement variance from repeated measurements or from sensor suppliers), but they do not have *a priori* knowledge about the quality of their model equations, which are only approximate representations of the true physical

process due to disturbances that are not included in the model equations and to simplifying assumptions that are made during model development.

Approximate Maximum Likelihood parameter Estimation (AMLE) is a novel parameter estimation algorithm that we have recently developed to address the problem of parameter estimation in continuous-time nonlinear dynamic models, in which model discrepancies are significant (Poyton et al., 2006; Varziri et al., 2008a, 2008b, 2008c). A convenient way to account for modelling errors and process disturbances is to include Gaussian noise terms on the right-hand side of the state equations, thereby converting ordinary differential equation models to stochastic differential equation (SDE) models (as shown in equation (6.1)). Until now, the AMLE algorithm has only been tested using simulated data and simple dynamic models of single-phase continuous stirred-tank reactors (Poyton et al., 2006; Varziri et al., 2008a, 2008b, and 2008c). The purpose of this article is twofold: we examine and demonstrate the application of AMLE to parameter and state estimation for a two-phase lab-scale nylon reactor model, and we improve and re-evaluate the structure of the nylon reactor model, originally developed by Schaffer et al. (2003b), to ensure model adequacy and parsimony. The nylon reactor model has four states, which are described by one algebraic equation and three nonlinear differential equations. Only two of the four states are measured, and these measurements were made at irregular sampling times that were convenient for the experimenters (Schaffer et al., 2003b; Zheng et al., 2005).

Parameter estimation in the nylon reactor model has been discussed in several articles (Schaffer et al., 2003b; Zheng et al., 2005; Ramsay et al., 2007). Researchers have proposed several different versions of the reactor model in an effort to reduce the correlations among parameter estimates and to keep the model parsimonious while maintaining model adequacy. Most of these proposed changes involve a semi-empirical expression for the apparent polycondensation equilibrium constant, K_a (shown in equation (6.18)). A contribution of the current article is to consider several

candidate expressions for K_a and study their performance using steady-state data arising from several different experiments (Schaffer et al., 2003b; Zheng et al., 2005; Zheng et al. 2007). As a result, a more appropriate expression for K_a is selected. Parameters in the overall dynamic model are then estimated using all of the available steady-state and dynamic data. It is shown that the AMLE framework readily facilitates the integration of additional steady-state or dynamic information that might be available from different sources.

We show how AMLE can address frequently-encountered parameter-estimation difficulties such as working with multi-response models with different levels of measurement accuracy, extracting information from multiple experimental runs with non-uniform sampling times, unmeasured states, unknown initial conditions and unknown levels of modelling error (due to disturbances and structural imperfections). Our objective in developing AMLE was to produce a straightforward parameter estimation algorithm that can help modellers to obtain more-reliable parameter estimates and model predictions that can be used in nonlinear model-based control and optimization schemes.

In Section 6.3, the AMLE algorithm is briefly reviewed. In Section 6.4, the lab-scale nylon reactor model is introduced, followed by a model-selection analysis using steady-state data from the literature. AMLE is then used for parameter estimation in the proposed overall dynamic model in Section 6.5. Conclusions are presented in Section 6.7.

6.3 Review of the AMLE Fitting Criterion

Maximum Likelihood (ML) estimation is a very popular method for parameter estimation in a wide variety of model types due to its desirable asymptotic properties (Seber and Wild, 1989; Bates and Watts, 1988). Unfortunately, ML estimation in nonlinear Stochastic Differential Equations (SDEs) is generally very difficult. This difficulty arises because when the initial state

condition with an assumed probability distribution function (PDF) is propagated through a nonlinear mapping, calculating the propagated PDF can be very complex and computationally intensive (Timmer, 2000). Available parameter estimation algorithms try to approximate the mapped PDF using techniques such as local linearization and Extended Kalman filtering, deterministic sampling techniques, or ensemble averaging (Jazwinski, 1970; Maybeck, 1982; Nielsen et al., 2000; Roweis and Ghahramani, 2001; Evensen, 2003; Julier and Uhlmann, 2004; Voss et al., 2004; Kristensen et al., 2004).

AMLE is an approximate ML-based method that maximizes the conditional joint density function of the states and measurements, given the model parameters, while assuming a piece-wise polynomial discretization scheme for the time evolution of the states of the dynamic model.

AMLE transforms the problem of state and parameter estimation in SDEs into a nonlinear minimization problem that does not require the time evolution of the PDF of the model states, due to the convenient form of its objective function (shown in eq. (6.7)). In the following paragraphs we briefly review the AMLE algorithm. We refer the reader to Varziri et al. (2008b & 2008c) (Chapter 5 and Chapter 3) for a more detailed description. 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., 2008c) (Chapter 3).

Consider the following continuous-time stochastic dynamic model*:

---

* This dynamic model can be more rigorously written as: $dx(t) = f(x(t), u(t), \boldsymbol{\theta})dt + d\omega(t)$ where $d\omega(t)$ is the increment of a Wiener process (Maybeck, 1979).

$$\frac{dx(t)}{dt} = f(x(t), u(t), \boldsymbol{\theta}) + \eta(t)$$
$$x(t_0) = x_0 \tag{6.1}$$

$$y(t_{mj}) = x(t_{mj}) + \varepsilon(t_{mj})$$

$x \in \Re$ is the state variable, $u \in \Re$ is the input variable and $y \in \Re$ is the output variable. $\boldsymbol{\theta} \in \Re^p$ is the vector of unknown model parameters and $f: \Re \times \Re \times \Re^p \to \Re$ is a nonlinear function of the state variables, the input variables and the parameters. We assume that f satisfies some regularity conditions (Kloeden and Platen, 1992) so that eq. (6.1) has a unique solution. ε is a zero-mean uncorrelated Normal random variable with variance σ_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 power spectral density and $\delta(.)$ is the Dirac delta function. The random noise trajectory $\eta(t)$ is a series of random steps with a switching time of Δt, where $\Delta t \to 0$. For the corresponding discrete-time white-noise process (Maybeck, 1979):

$$E\{\eta(j_1\Delta t)\eta(j_2\Delta t)\} = \begin{cases} \dfrac{Q}{\Delta t} & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases} \tag{6.2}$$

where j_1 and j_2 are integers and Δt is the sampling period. We also assume that the process disturbance $\eta(t)$ and the measurement noise $\varepsilon(t)$ are not correlated.

The set of times at which the measurements are available is denoted by t_{mj} ($j = 1 \cdots n$). The measurement times t_{mj} do not need to be uniformly spaced. The vector of outputs at observation

times $y(t_{mj})$ ($j = 1 \cdots n$) and its corresponding state vector of true values $x(t_{mj})$ ($j = 1 \cdots n$) are denoted by $\mathbf{y}_m$ and $\mathbf{x}_m$ respectively.

In AMLE, the objective is to find estimates of $x(t)$ and $\boldsymbol{\theta}$ (assuming at this moment that Q and σ_m^2 are known constants) that minimize the following function:

$$\frac{(\mathbf{y}_m - \mathbf{x}_m)^T (\mathbf{y}_m - \mathbf{x}_m)}{2\sigma_m^2} + \frac{1}{2Q} \int_{t_0}^{t_f} \left(\frac{dx(t)}{dt} - f(x(t), u(t), \boldsymbol{\theta}) \right)^2 dt \tag{6.3}$$

where $[t_0, t_f]$ is the time span over which the measurements are taken. Note that minimizing objective function (6.3) corresponds to maximization of a likelihood function, as described by Varziri et al. (2008b, 2008c) (Chapter 5 and Chapter 3). The first term in objective function (6.3) is a sum of squared errors (SSE) term, arising from the sum of squared deviations of the estimated states from their corresponding observations. The second term is a model-based penalty term, which ensures that the estimated state trajectory is an approximate solution of the differential equation. The relative sizes of the measurement noise variance, σ_m^2, and the process disturbance intensity, Q, determine whether the optimization focuses more on minimizing the SSE term versus minimizing the model-based penalty.

Since $x(t)$ is an unknown curve, minimizing (6.3) over $x(t)$ and $\boldsymbol{\theta}$ is an infinite-dimensional optimization problem (a calculus of variations problem) which is generally hard to solve.

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

$$x(t) \approx x_{\sim}(t) = \sum_{i=1}^{c} \beta_i \phi_i \tag{6.4}$$

where β_i, $i = 1 \cdots c$ are B-spline coefficients and $\phi_i(t)$ $i = 1 \cdots c$ are B-spline basis functions (de Boor, 2001). Note that eq. (6.4) can be written in matrix form:

$$x_{\sim}(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\beta} \tag{6.5}$$

where $\boldsymbol{\varphi}(t)$ 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:

$$\frac{dx_{\sim}(t)}{dt} = \frac{d}{dt}\left(\sum_{i=1}^{c} \beta_i \phi_i(t)\right) = \sum_{i=1}^{c} \beta_i \dot{\phi}_i(t) = \dot{\boldsymbol{\varphi}}^T \boldsymbol{\beta} \tag{6.6}$$

Note that other basis functions could readily be used instead of B-splines (Ramsay and Silverman, 2005).

By substituting (6.5) and (6.6) into (6.3) we have the following finite-dimensional optimization problem:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\theta}, \boldsymbol{\beta}}{\arg\min} \left\{ \frac{(\mathbf{y}_m - \mathbf{x}_{\sim m})^T (\mathbf{y}_m - \mathbf{x}_{\sim m})}{2\sigma_m^2} + \frac{1}{2Q} \int_{t_0}^{t_f} \left(\frac{dx_{\sim}(t)}{dt} - f(x_{\sim}(t), u(t), \boldsymbol{\theta}) \right)^2 dt \right\} \tag{6.7}$$

Minimizing (6.7) provides point estimates for the model parameters and the spline coefficients. The spline coefficients can then be used to determine the estimated state trajectory $x_{\sim}(t)$. 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 (Seber and Wild, 1989; Varziri et al., 2008a). In engineering applications, reasonable estimates for the measurement noise variance are usually available either from repeated experimental observations or from the manufacturer of the measurement device. Obtaining a reasonable estimate for the process disturbance intensity, *Q*, however, is very

difficult. Assuming known measurement noise variance, σ_m^2, but unknown process disturbance intensity Q, Varziri et al. (2008b) (Chapter 5) used a ML argument (Heald and Stark, 2000) to propose a two-step optimization scheme that allows for estimating the process disturbance intensities along with model states and parameters. The idea is to select Q to ensure that the estimated measurement noise variance $\hat{\sigma}_m^2$ is close to the known value of σ_m^2. The proposed two-step optimization scheme can be summarized as follows:

Outer optimization problem:

$$\hat{Q} = \arg\min_{Q} \left(\frac{\hat{\sigma}_m^2(Q)}{\sigma_m^2} - 1 \right)^2 \tag{6.8}$$

Inner optimization problem:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\theta},\boldsymbol{\beta}} \left\{ \frac{(\mathbf{y}_m - \mathbf{x}_{\sim m})^T (\mathbf{y}_m - \mathbf{x}_{\sim m})}{2\sigma_m^2} + \frac{1}{2\hat{Q}} \int_{t_0}^{t_q} \left(\frac{dx_{\sim}(t)}{dt} - f(x_{\sim}(t), u(t), \boldsymbol{\theta}) \right)^2 dt \right\} \tag{6.9}$$

The variance estimate, $\hat{\sigma}_m^2$, in objective function (6.8) is the approximate ML estimator developed by Heald and Stark (2000):

$$\hat{\sigma}_m^2 = \frac{(\mathbf{y}_m - \mathbf{x}_{\sim m}(Q))^T (\mathbf{y}_m - \mathbf{x}_{\sim m}(Q))}{n} + \left. \frac{Trace(\mathbf{A}^{-1})}{n} \right|_{\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\beta}}} \tag{6.10}$$

where $\mathbf{A}$ is the Hessian matrix of objective function (6.9) with respect to the state variables, evaluated at $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}$, which are obtained as the solution of the inner optimization problem. The outer optimization minimizes the discrepancy between the estimated and the known measurement variance, while the inner optimization minimizes the criterion in (6.7) using the value of Q obtained from the outer optimization. The converged results of this overall optimization problem provide the modeller with $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}$ and $\hat{Q}$. $\hat{\boldsymbol{\theta}}$ is the desired estimate for the fundamental model

parameters. The estimated spline coefficients $\hat{\boldsymbol{\beta}}$ can be used to obtain the estimated state trajectory $x_{\sim}$, while $\hat{Q}$ provides information about the magnitude of the model uncertainty and process disturbances.

6.4 Case study: Lab-scale Nylon 612 reactor model

6.4.1 Nylon 612 mathematical model

Nylons are widely used polymers which are produced according to the following reaction between carboxylic acid end-groups (*C*) and amine end-groups (*A*) to produce amide linkages (*L*) and water (*W*):

$$\begin{array}{ccccc} \text{-COOH} + \text{H}_2\text{N-} & \leftrightarrow & \text{-CONH-} + \text{H}_2\text{O} \\ C \quad + \quad A & & L \quad + \quad W \end{array}$$

The forward reaction is a polyamidation reaction in which -COOH and -NH_2 are joined to form -CONH- and H_2O and the reverse reaction is a hydrolysis reaction in which the amide link -CONH- is broken.

The purpose of the experimental study and modelling of nylon reactions conducted by Shaffer et al. (2003b) and Zheng et al. (2005) is to gain quantitative knowledge of the kinetics and equilibrium of the polycondensation reaction at the high temperatures and low water contents that are experienced in the final stages of commercial polyamidation processes. Nylon 612 was chosen as the study material because thermal degradation reactions are expected to be negligible (Schaffer et al. 2003b) for this particular type of nylon.

Researchers do not agree unanimously on the kinetic rate order for the above reaction, and both second and third-order reaction rates have been reported under different conditions. Based on their experimental analyses, Zheng et al. (2005) concluded that the process reaction rate is

second-order (first-order with respect to carboxyl ends and first-order with respect to amine ends). The following equations describe the dynamic behaviour of the contents of a well-stirred melt-phase nylon polymerization reactor. Please refer to Schaffer et al. (2003b) and Zheng et al., (2005) for detailed information about the development of the material-balance equations.

$$\frac{dL}{dt} = -\frac{dA}{dt} = -\frac{dC}{dt} = k_{\mathrm{p}}\left(CA - \frac{LW}{K_{\mathrm{a}}}\right) = f_1(\boldsymbol{\theta}, C, A, L, W) \tag{6.11}$$

$$\frac{dW}{dt} = k_p\left(CA - \frac{LW}{K_a}\right) - k_m\left(W - W_{eq}\right) = f_2(\boldsymbol{\theta}, C, A, L, W, W_{eq}) \tag{6.12}$$

where

$$k_{\mathrm{p}} = k_{\mathrm{p0}} \exp\left[-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right] \tag{6.13}$$

is the temperature-dependent polycondensation rate constant and k_{p0} is the polycondensation rate constant at the reference temperature, T_0=549.15 K, which is chosen to be in the middle of the temperature range over which the experiments were conducted. K_{a} is the apparent polycondensation equilibrium constant, and k_{m} is a mass-transfer coefficient that was estimated by Schaffer et al. (2003b) to be k_{m}=24.3 h^{-1} for the conditions encountered in the lab-scale reactor. Zheng et al. (2005) used this estimated value of k_{m} in their work.

From eq. (6.11) we note that the rates of consumption of amine ends and carboxyl ends are the same and hence, their corresponding concentrations A and C differ by a constant only. This relationship was confirmed by the experimental data provided by Zheng et al. (2005).

The equilibrium concentration of water in the polymer melt, W_{eq}, can be approximated using a Flory-Huggins-based expression (Schaffer et al., 2003a):

$$W_{eq} = 5.55 \times 10^4 \frac{P_w}{P_w^{sat}} \exp\left(-9.624 + \frac{3613}{T} \right) \tag{6.14}$$

where the saturation vapour pressure of the water in the gas phase, $P_{\mathrm{w}}^{\mathrm{sat}}$, can be calculated using the Wagner equation:

$$\begin{aligned} \ln(P_{\mathrm{w}}^{\mathrm{sat}} / P_{\mathrm{c}}) = [&-7.77224(1 - T/T_{\mathrm{c}}) + 1.45684(1 - T/T_{\mathrm{c}})^{1.5} - 2.71492(1 - T/T_{\mathrm{c}})^3 \\ &-1.41336(1 - T/T_{\mathrm{c}})^6]/(T/T_{\mathrm{c}}) \end{aligned} \tag{6.15}$$

P_{c} and T_{c} are the critical pressure and temperature of water, respectively.

The concentration of amide linkages, L, in the molten Nylon 612 can be obtained from the following material balance equation (Schaffer et al., 2003b):

$$155.23L = 10^6 - 115.15C - 58.10A - 18.02W \tag{6.16}$$

Note that since the concentration of the amide linkages L can be computed algebraically from eq. (6.16), only the differential equations for A, C, and W need to be solved.

To allow for possible modelling errors and process disturbances, stochastic terms can be added to the differential equations:

$$\begin{aligned} \frac{dA}{dt} &= -f_1(\boldsymbol{\theta}, C, A, L, W) + \eta_A \\ \frac{dC}{dt} &= -f_1(\boldsymbol{\theta}, C, A, L, W) + \eta_C \\ \frac{dW}{dt} &= f_2(\boldsymbol{\theta}, C, A, L, W, W_{eq}) + \eta_W \end{aligned} \tag{6.17}$$

where η_A, η_C, and η_W are continuous-time Gaussian disturbances with intensities Q_A Q_C and Q_W [(mol Mg^{-1})^2/hr], respectively.

Schaffer et al. (2003b) noticed that the apparent polycondensation equilibrium constant K_a depends on the water concentration as well as the temperature. The following semi-empirical expression was therefore proposed:

$$K_a = \left[\frac{b + mW_{eq}}{\gamma_W / \gamma_{W0}}\right] K_{a0} \exp\left[-\frac{\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right] \tag{6.18}$$

where the activity coefficient for water in the molten nylon (Schaffer et al., 2003a) is given by $\gamma_W = \exp\left(9.624 - \frac{3613}{T}\right)$ and γ_{W0}=20.97. The empirical $b + mW_{eq}$ term in the numerator of Schaffer's model accounts for the influence of water on the activity coefficients of amine ends, carboxyl ends and amide links in the polymer melt.

Zheng et al. (2005) were concerned that activity coefficients for the end groups and amide links might also be influenced by temperature, and therefore modified the K_a equation as follows:

$$K_a = \left[\frac{1 + \exp(\alpha + \beta/T)W_{eq}}{\gamma_W / \gamma_{W0}}\right] K_{a0} \exp\left[-\frac{\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right] \tag{6.19}$$

Six parameters, $\boldsymbol{\theta} = [k_{p0}, K_{a0}, E, \alpha, \beta, \Delta H]^T$, were estimated in their model using a Weighted Nonlinear Least-Squares (WNLS) approach similar to that of Schaffer et al. (2003b). A weighted sum of squared errors was minimized while solving the model differential and algebraic equations. Initial values of the states were assumed to be known and were not estimated. The initial value, A_0, used to solve for the amine end-group concentration trajectory was set at the measured value from the first polymer sample taken from the reactor. Because the carboxyl end-group concentration measurements were very noisy, the initial carboxyl end-group measurement was not reliable enough to use as the initial condition, C_0. Instead, the average difference between carboxyl and amine end group concentration measurements during each experimental

run, $\overline{C-A}$, was computed. This average difference was added to A_0 to determine the initial carboxyl concentration ($C_0 \approx A_0 + \overline{C-A}$) for each experimental run. It will be shown that the use of empirical spline functions in the AMLE algorithm naturally facilitates estimation of unknown initial state conditions, while properly accounting for different measurement variances for measured states, so that Zheng's approach for determining C_0 is not required. Many of the parameter values estimated from Zheng's model and the available data showed very high correlations.

Campbell (2007) and Ramsay et al. (2007) considered parameter estimation in Zheng's nylon reactor model using a generalized smoothing (GS) approach. Like AMLE, the GS approach solves a two-step minimization problem. However, there are some important differences that distinguish AMLE from GS. The main differences are in the form of the outer optimization objective function. In the GS approach, the outer objective function is a WNLS objective function that is minimized over the model parameters $\boldsymbol{\theta}$. The inner objective function in GS contains a tuning parameter that is adjusted manually by the user to account for possible model imperfections. A major advantage of the AMLE algorithm over the GS approach is that AMLE eliminates the requirement of manually tuning any weighting factors or process disturbance intensities. In AMLE, the disturbance intensities Q are estimated using the outer optimization so that the estimated measurement noise-variance is consistent with prior knowledge about the quality of the measurements.

Because of the high correlations among the parameter estimates, Campbell (2007) modified the model proposed by Zheng et al. (2005) to produce a six-parameter model with a slightly different expression for K_a. Based on statistical considerations such as analyzing parameter-estimate correlations and estimated confidence intervals, Campbell concluded that the six-parameter

model was over-parameterized. A revised four-parameter model was proposed by removing the temperature dependency from k_p (*i.e.*, setting the activation energy E to 0) and modifying the apparent polycondensation equilibrium constant K_a model as follows:

$$K_a = \left[\frac{1 + \frac{g}{1000} W_{eq}}{\gamma_W / \gamma_{W0}} \right] K_{a0} \exp\left[-\frac{\Delta H}{R} \left(\frac{1}{T} - \frac{1}{T_0} \right) \right] \tag{6.20}$$

In the current article, we use additional steady-state data (Zheng et al., 2007) and nonlinear regression to select an adequate and parsimonious semi-empirical model for K_a.

6.4.2 Model selection for K_a

The proposed strategy for K_a model selection is based on the definition of the apparent polycondensaton equilibrium constant (Zheng et al., 2005):

$$K_a = \frac{L_{eq} W_{eq}}{C_{eq} A_{eq}} \tag{6.21}$$

where the subscript "*eq*" indicates equilibrium concentrations. Steady-state data will be used to provide these equilibrium concentrations. From the steady-state data at our disposal, we can calculate K_a and treat it as a measured response (y_{K_a}). The variance $\sigma^2_{K_a}$ of the error, ε_{K_a}, in this response stems not only from the measurement noise, but also the process disturbances that introduce variability into the values of A_{eq} and C_{eq}. Using repeated measurements of A_{eq} and C_{eq} at the same water concentration W_{eq} and temperature T, a pooled variance estimate for K_a was calculated to be 94.9. This variance estimate accounts for both sources of variability, namely the measurement noise and also the process disturbances. Since the steady-state measurements were obtained at time points that are sufficiently far apart, it is reasonable to assume that the correlation among them is negligible and to treat them as replicates.

If we denote any postulated semi-empirical model for K_a by:

$$y_{K_a} = f_{K_a}(\boldsymbol{\theta}_{K_a}, W_{eq}, T) + \varepsilon_{K_a} \tag{6.22}$$

a nonlinear regression problem can be formulated to estimate the parameters $\boldsymbol{\theta}_{K_a}$ and to study the model adequacy. Looking at the semi-empirical K_a models previously studied (eqs. (6.18), (6.19) and (6.20)), we note that these models share the common structure of:

$$f_{K_a}(\boldsymbol{\theta}_{K_a}, W_{eq}, T) = \frac{\Gamma}{\gamma_W / \gamma_{W0}} K_{a0} \exp\left[-\frac{\Delta H}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right] \tag{6.23}$$

and that they differ only in the expression for $\Gamma = \dfrac{\gamma_C \gamma_A \gamma_{L0}}{\gamma_{C0} \gamma_{A0} \gamma_L}$, which is an unknown ratio of activity coefficients at the current water concentration divided by the same activity coefficients when $W_{eq} \rightarrow 0$. Using this common structure, we have examined eight different empirical expressions for Γ shown in Table 6.1.

Table 6.1. Γ expressions considered for eq. (6.23)

	Γ expression	Sum of Squared Errors
(1)	$\Gamma = a + bW_{eq}$	4058.13
(2)	$\Gamma = (1 + \exp(a + b/T)W_{eq})$	3956.58
(3)	$\Gamma = 1 + aW_{eq}$	4058.13
(4)	$\Gamma = 1$	12955.14
(5)	$\Gamma = 1 + (1 - \exp(-W_{eq}/b))$	5351.68
(6)	$\Gamma = 1 + a(1 - \exp(-W_{eq}/b))$	3573.92
(7)	$\Gamma = a\sqrt{W_{eq}} + 1 + bW_{eq}$	3487.91
(8)	$\Gamma = a\sqrt{W_{eq}} + 1$	3573.14

Expressions (1-3) in Table 6.1 correspond to the models proposed by Schaffer et al., Zheng et al. and Campbell, respectively. Expression (4) assumes that the activity coefficient ratio does not depend on the water concentration or the temperature. Expressions (5-8) were selected because steady-state data (plotted in Figure 6.1) indicate that Γ depends on W_{eq} in a nonlinear fashion.

For each of these eight expressions, the nonlinear regression problem in (6.22) was solved, using the steady-state data in the Appendix 6.9.1. Parameter estimates were obtained using the nonlinear least-squares routine "nlscon" (Nowak and Weimann, 1991) in Matlab. Based on the sum of squared residuals, qualitative analysis of residual plots and also the approximate individual confidence intervals for parameter estimates, we concluded that the eighth expression $\Gamma = a\sqrt{W_{eq}} + 1$ is the most suitable. It provides an adequate fit for the steady-state data while maintaining model parsimony. Even though, expression (7) results in a smaller sum of squared error, because of high correlation among the parameter estimates and also approximate confidence intervals that contained zero, this model was rejected. In this study we used steady-state information from the original six sets of experimental runs conducted by Schaffer et al. (2003b) and Zheng et al. (2005) (Table 6.7 to Table 6.12; the steady-state points are distinguished by placing an asterisk sign "*" next to the corresponding measurement times) and also three new sets of experimental runs that were conducted to study the effects of Sodium Hypophosphite Catalyst (SHP) on the nylon 612 polycondensation kinetics (Zheng et al. 2007) (Table 6.13 to Table 6.15, which only contain steady-state data). Note that the SHP catalyst influences the rate of the forward and reverse polycondensation reactions, but it does not affect the reaction equilibrium.

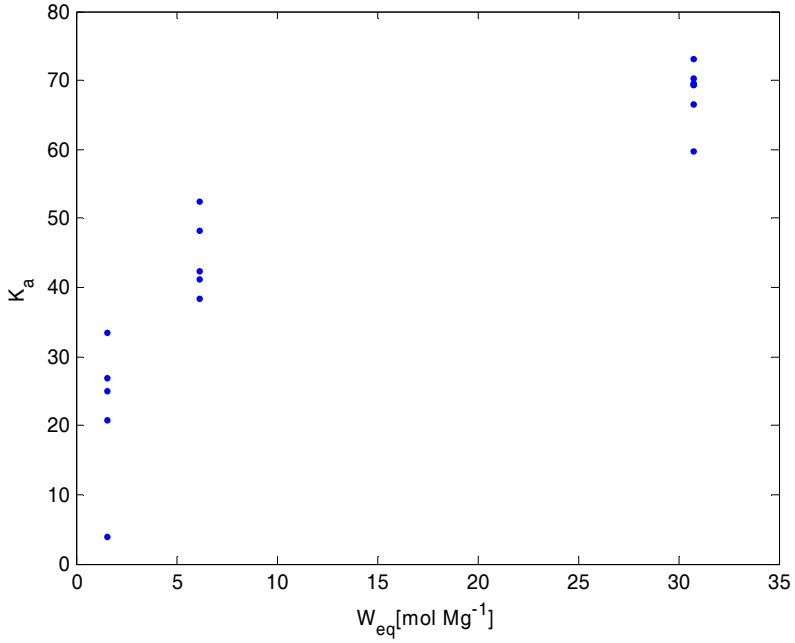

Figure 6.1. K_a versus W_{eq} for T=290 ºC

For the six experimental runs conducted without a catalyst, the first three were conducted at temperatures 263, 271 and 281 ºC, respectively while the last three runs were conducted at 284 ºC. The three experimental runs with the catalyst were conducted at T = 290 ºC. During each experimental run, the concentrations of A and C are measured at several non-uniformly-spaced times. The standard deviations of the A and C concentration measurements are known to be $\sigma_A = 0.6$ mol Mg^{-1}, and $\sigma_C = 2.4$ mol Mg^{-1} respectively (Schaffer et al., 2003b). Since the measured concentration of A is more accurate, more A concentration measurements than C concentration measurements were made during some of the runs. The water concentration W was not measured; therefore, W is treated as an unmeasured dynamic state. However, W_{eq}, which is a steady-state water concentration in the molten nylon can be computed from the input variable, P_W, using eq. (6.14).

Figure 6.2 shows predicted values of K_a along with observed K_a values calculated from steady-state measurements at different temperatures and water concentrations, using the selected expression for Γ. Note that multiple predicted values are shown for K_a at some of the temperatures in Figure 6.2a, because two or three different steady-state water concentrations were achieved in some experimental runs. The values of K_a that were computed from the experimental data are close to the corresponding predicted K_a values. Residual plots (using the selected Γ model, expression 8 in Table 6.1) are presented in Figure 6.3 and Figure 6.4. The predicted K_a values appear to be consistent with the K_a calculated from the measurements. The residual plots show no particular trend and the residuals are scattered randomly about the zero line. Based on these results, we conclude that the following is a reasonable empirical model for the apparent polycondensation equilibrium constant.

$$K_a = \left[\frac{a\sqrt{W_{eq}}+1}{\gamma_W/\gamma_{W0}}\right] K_{a0} \exp\left[-\frac{\Delta H}{R}\left(\frac{1}{T}-\frac{1}{T_0}\right)\right] \tag{6.24}$$

Note that all of the other candidate expressions for K_a obtained using the various expressions for Γ in Table 6.1, either produced approximate confidence intervals for some of the parameters that contained zero, or resulted in a larger sum of squared prediction errors than the selected expression for Γ. The estimated parameter values are presented inTable 6.2. As shown in the next section, these estimated values are used as initial parameter guesses for the overall dynamic parameter estimation problem, where they are estimated along with k_{p0} and E.

Table 6.2. Estimated parameter values for eq. (24) obtained from steady-state nylon 612 polymerization data

a = 0.60	K_{a0} = 22.01	ΔH =-39.62

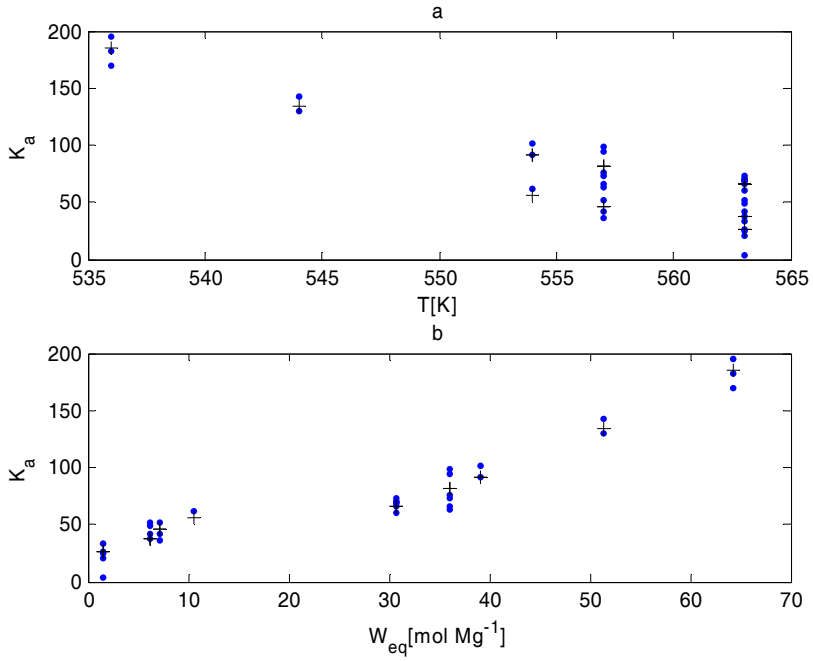

Figure 6.2. Predicted K_a (+) obtained using eq. (24) and observed K_a (•) calculated from experimental data plotted versus the measured reactor temperature T (top plot) and versus the calculated water concentration in the molten polymer W_{eq} (bottom plot) using steady-state data from all nine experimental runs

6.5 Parameter estimation results

In this section, we use the objective function of the form given in (6.8) and (6.9) to estimate five parameters, $\boldsymbol{\theta} = [k_{\mathrm{p0}}, K_{\mathrm{a0}}, E, a, \Delta H]^T$, along with three process disturbance intensities $\mathbf{Q}=[Q_A, Q_C, Q_W]^T$ and three state trajectories in the nylon 612 reactor model described by equations (6.13) to (6.17) and (6.24).

To form the AMLE objective function, we use B-spline expansions to approximate state trajectories. For the i-th experimental run for A, C, and W, from (6.5):

$$A_{\sim i}(t) = \boldsymbol{\varphi}_{\mathrm{A}i}^T(t)\boldsymbol{\beta}_{\mathrm{A}i} \tag{6.25}$$

$$C_{\sim i}(t) = \boldsymbol{\varphi}_{\mathrm{C}i}^{T}(t)\boldsymbol{\beta}_{\mathrm{C}i}$$

$$W_{\sim i}(t) = \boldsymbol{\varphi}_{\mathrm{W}i}^{T}(t)\boldsymbol{\beta}_{\mathrm{W}i}$$

Since W is not measured, there is no SSE term associated with this component. The AMLE objective function corresponding to the i-th run for this multi-response model with an unmeasured state (Varziri et al., 2008a) (Chapter 4) becomes:

$$\begin{aligned} J_i = &\frac{1}{2\sigma_{\mathrm{C}}^2}\sum_{j=1}^{N_{\mathrm{Ci}}}(y_{\mathrm{Ci}}(t_j)-C_{\sim i}(t_j))^2+\frac{1}{2Q_{\mathrm{C}}}\int_{t_{C0}}^{t_{Cf}}\left(\frac{dC_{\sim i}(t)}{dt}+f_1(A_{\sim i}(t),C_{\sim i}(t),W_{\sim i}(t),L_1(t),\boldsymbol{\theta})\right)^2 dt+ \\ &\frac{1}{2\sigma_{\mathrm{A}}^2}\sum_{j=1}^{N_{\mathrm{Ai}}}(y_{\mathrm{Ai}}(t_j)-A_{\sim i}(t_j))^2+\frac{1}{2Q_{\mathrm{A}}}\int_{t_{A0}}^{t_{Af}}\left(\frac{dA_{\sim i}(t)}{dt}+f_1(A_{\sim i}(t),C_{\sim i}(t),W_{\sim i}(t),L_1(t),\boldsymbol{\theta})\right)^2 dt+ \\ &\frac{1}{2Q_{\mathrm{W}}}\int_{t_{W0}}^{t_{Wf}}\left(\frac{d\mathrm{W}_{\sim i}(t)}{dt}-f_2(A_{\sim i}(t),C_{\sim i}(t),W_{\sim i}(t),W_{eqi}(t),L_1(t),\boldsymbol{\theta})\right)^2 dt \end{aligned} \tag{6.26}$$

N_{Ai} and N_{Ci} are the numbers of A and C measurements in the i-th experimental run respectively.

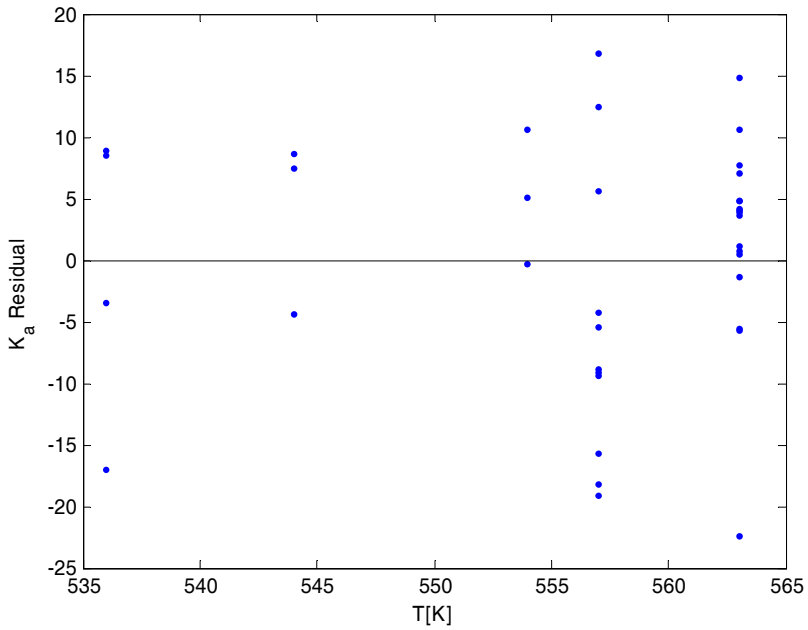

Figure 6.3. Residuals (observed K_a - predicted K_a) versus T for steady-state data from all of the nine experimental runs

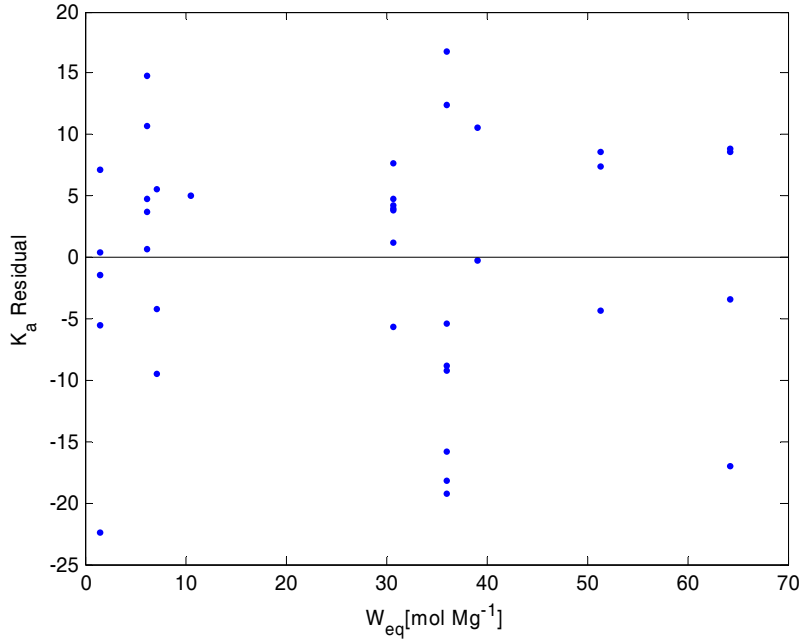

Figure 6.4. Residuals (observed K_a - predicted K_a) versus W_{eq} for steady-state data from all of the nine experimental runs

Since initial conditions for A and C are assumed to be unknown, they are included in the corresponding SSE terms, and the B-spline expansions are not constrained to the initial conditions as they would be if the true initial values were known.

Since only six experimental runs without the catalyst are used in the overall parameter estimation, the inner AMLE objective function for these six runs becomes:

$$J_{inner} = \sum_{i=1}^{6} J_i \tag{6.27}$$

Using the objective function in (6.27), the extra steady-state information available from the three experimental runs with catalyst, which was used to obtain the initial estimates in Table 6.2, would

be disregarded. To include this extra information, objective function (6.27) can be modified as follows:

$$J_{inner} = \sum_{i=1}^{6} J_i + \frac{1}{2\sigma_{K_a}^2} \sum_{i=1}^{3} \left(y_{K_a} - f_{K_a}(\boldsymbol{\theta}_{K_a}, W_{eq}, T) \right)^2 \tag{6.28}$$

where $f_{K_a}(\boldsymbol{\theta}_{K_a}, W_{eq}, T)$ is defined by (6.24). This modification arises naturally from the ML development when y_{K_a} is considered as an additional non-dynamic measured response (Appendix 6.9.2).

To form the objective function for the outer optimization problem, note that the approximate ML estimators for the measurement variances for *A* and *C* can be expressed as:

$$\hat{\sigma}_x^2 = \frac{\sum_{i=1}^{6} \sum_{j=1}^{N_{xi}} \left(y_{xi}(t_j) - x_{\sim i}(t_j) \right)^2}{\sum_{i=1}^{6} N_{xi}} + \frac{Trace(\mathbf{A}_x^{-1})}{\sum_{i=1}^{6} N_{xi}} \qquad x = A, C \tag{6.29}$$

where $\mathbf{A}_x^{-1}$ is the inverse of the Hessian of the inner objective function J_{inner} with respect to $x_{\sim}$. Eq. (6.29) can be rearranged (Heald and Stark, 2000) as

$$\hat{\sigma}_x^2 = \frac{\sum_{i=1}^{6} \sum_{j=1}^{N_{xi}} \left(y_{xi}(t_j) - x_{\sim i}(t_j) \right)^2}{\sum_{i=1}^{6} N_{xi} - \nu_x} \qquad x = A, C \tag{6.30}$$

where $\nu_x = 1/\hat{\sigma}_x^2 Trace(\mathbf{A}^{-1})$. In this article, the known values of the measurement noise variance, σ_x^2, is used in place of $\hat{\sigma}_x^2$ to calculate ν_x. The denominator in (6.30) is the degrees of freedom $DOF_x = \sum_{i=1}^{6} N_{xi} - \nu_x$.

The outer optimization problem can then be written as:

$$\hat{\mathbf{Q}} = \arg\min_{\mathbf{Q}} \left(DOF_A \times \left(\frac{\hat{\sigma}_A^2(\mathbf{Q})}{\sigma_A^2} - 1 \right)^2 + DOF_C \times \left(\frac{\hat{\sigma}_{Ci}^2(\mathbf{Q})}{\sigma_C^2} - 1 \right)^2 \right) \tag{6.31}$$

Note that the terms in objective function (31) are weighted by DOF_A and DOF_C to properly account for the different numbers of available measurements for the amine and carboxyl end groups.

For three of the parameters, namely a, K_{a0}, and ΔH, that were previously estimated using the steady-state data (Section 6.4.2), the values in Table 6.2, were used as initial parameter guesses. For the rest of the parameters, estimated values reported by Zheng et al. (2005) were used as initial values $\boldsymbol{\theta}_0$ as shown in Table 6.4.

The initial values for the spline coefficients $\boldsymbol{\beta}_{\mathrm{A}i}$ and $\boldsymbol{\beta}_{\mathrm{C}i}$, ($i$=1..6) were obtained by fitting a smoothing spline (Ramsay and Silverman, 2005) to the measured data. Since water concentration is not measured, the initial guesses for the spline coefficients $\boldsymbol{\beta}_{\mathrm{W}i}$, ($i$=1..6), were obtained by fitting a smoothing spline to W_{eq} which was calculated from (6.14). An alternative would be to solve the model differential equations and fit a smoothing spline to the solution for W. Based on our experience, obtaining good (non-zero) initial guesses for B-spline coefficients is not necessary but can reduce the convergence time. The initial values of the process disturbance intensities were arbitrarily set to 1. The estimated intensities, $\hat{\mathbf{Q}}$, are shown in Table 6.3. The estimated parameters, $\hat{\boldsymbol{\theta}}$, are reported in Table 6.4. The results in Table 6.3 and Table 6.4 were obtained using the following stopping conditions: for the inner problem, the optimizer stopped when the objective function changed by less than 1E-8; the outer optimizer stopped when the value of the objective function changed by less than 1E-2.

Table 6.3. Process disturbance intensity estimates. All intensities have units of [(mol Mg^{-1})^2/hr]

Intensity	Estimate
Q_A	4.37
Q_C	15.56
Q_W	0.88

The final B-spline fits using the estimated parameters and coefficients are shown in Figure 6.5. The numerical solution of the model equations, which does not account for the stochastic process disturbances, is shown in Figure 6.6. Initial values for the output trajectories in Figure 6 were determined using the AMLE algorithm. Note that the B-spline curves pass very close to the *A* measurements, and they smooth out some of the noise associated with the noisier *C* measurements because we prescribed that the standard deviation of the known noise variance associated with *C* is larger than that of *A* ($\sigma_A = 0.6$ mol Mg^{-1} and $\sigma_C = 2.4$ mol Mg).

For inferences on the uncertainty associated with the parameter estimates, we approximate the Fisher information matrix using the Hessian, (**H**), of the inner objective function with respect to the model parameters $\boldsymbol{\theta}$ evaluated at the converged values. Approximate 100(1-α)% confidence intervals for the model parameters can be obtained as follows (Chapter 4):

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} \pm z_{\alpha/2} \times \sqrt{\text{diag}(\mathbf{H}^{-1}(\hat{\boldsymbol{\theta}}))} \tag{6.32}$$

Care should be taken in interpreting the confidence intervals obtained from (4.15); these intervals arise from linear approximations and also do not take into account the uncertainty in the estimates of the process disturbance intensities.

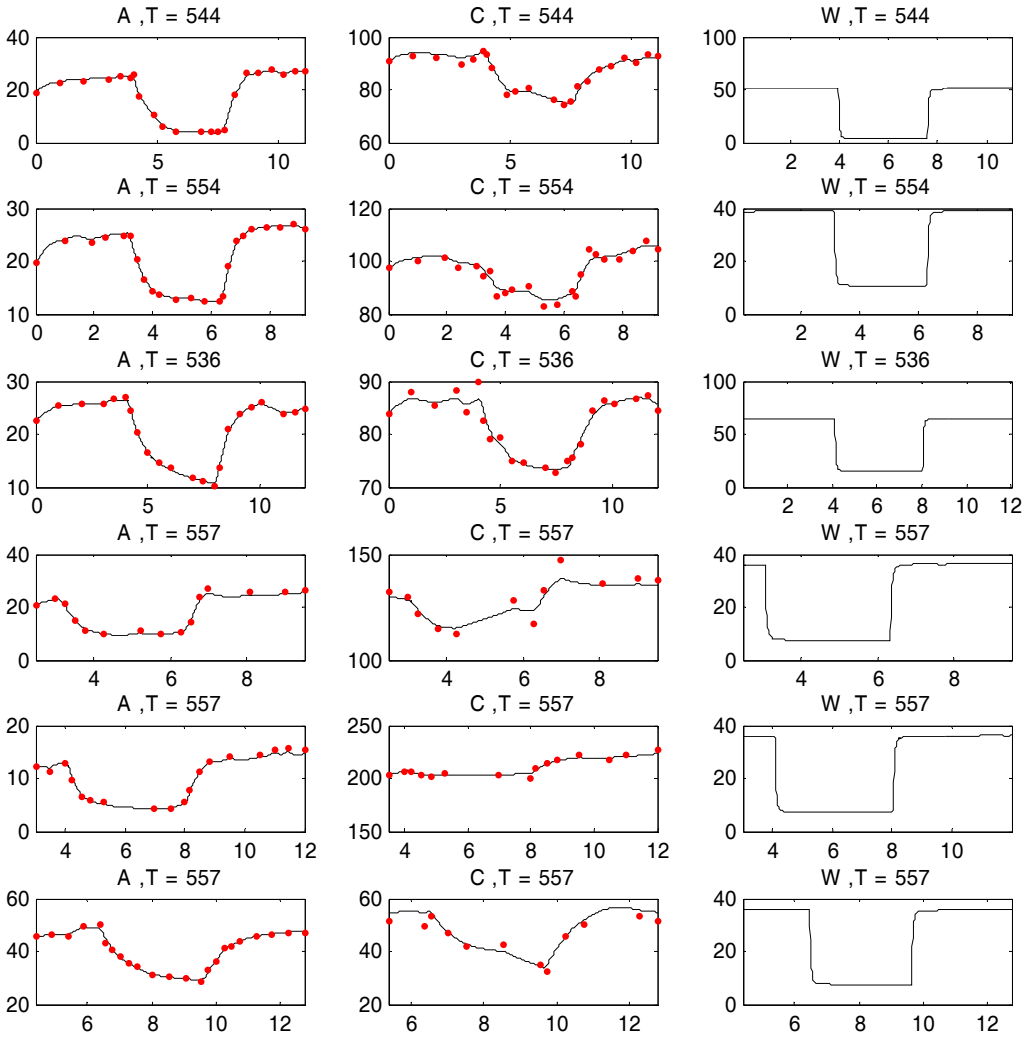

Figure 6.5. Optimal B-spline trajectories of A, C, and W for six experimental runs

(___, B-spline fit; •, measured data)

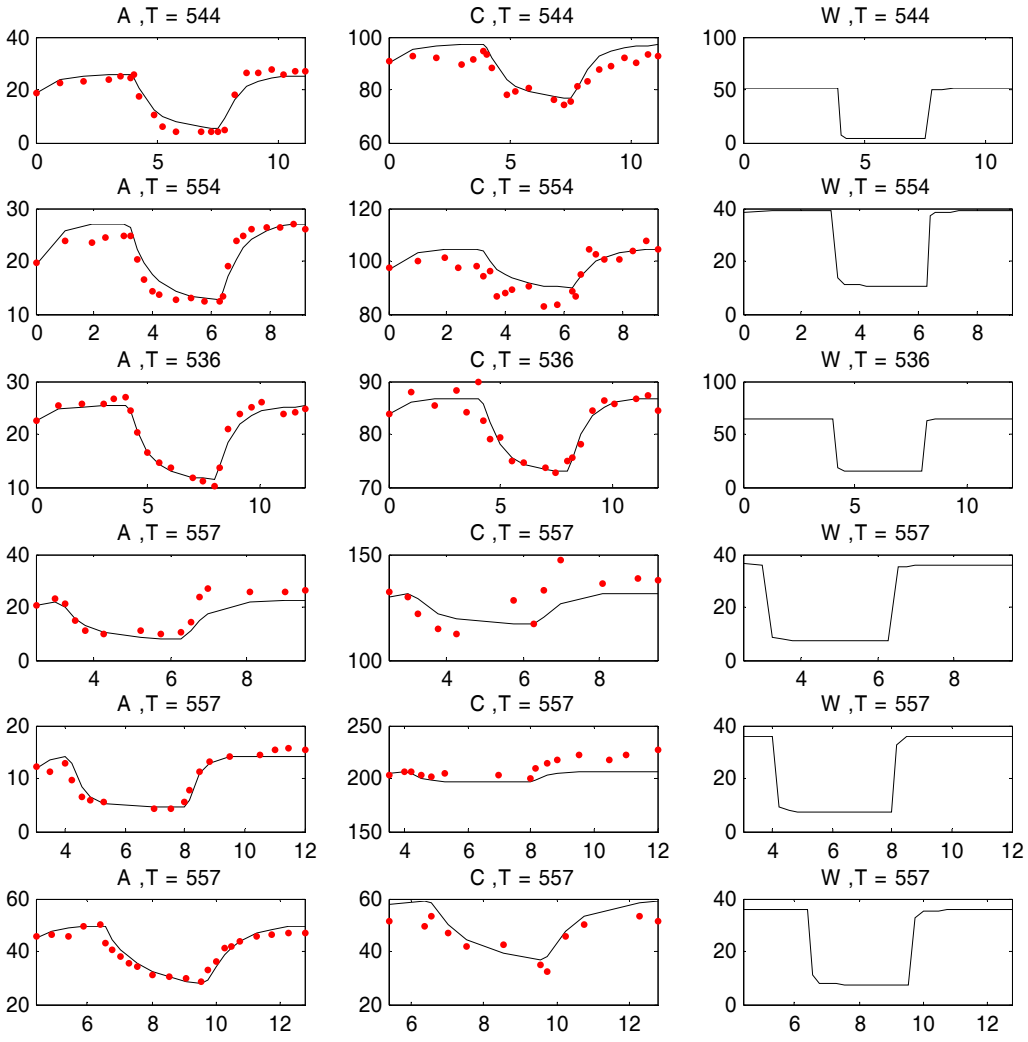

Figure 6.6. Numerical solution of model equations (without stochastic terms) for A, C, and W trajectories for six experimental runs using the estimated parameters. (___, numerical model solution; • , measured data)

Approximate 95% confidence intervals for the parameters are presented in Table 6.4 and a correlation matrix for the parameter estimates is presented in Table 6.5. The confidence interval corresponding to the activation energy E contains zero. This is consistent with the findings of Campbell (2007) and Ramsay et al. (2007). Although this parameter could reasonably be zero from a purely statistical point of view, physical insight tells us that the activation energy is a positive constant. Unfortunately, the dynamic data are not rich enough to provide sufficient information to estimate E reliably. Table 6.5 indicates that correlations between the parameter estimates are small, except for between K_{a0} and a. These two parameters are nearly perfectly negatively correlated.

Table 6.4. Point estimates and approximate 95% confidence intervals for the nylon reactor model parameters

Parameter	Initial Guess	Estimate	Lower	Upper Bound
k_{p0}	0.019	0.0129	0.0106	0.0152
K_{a0}	22.01	36.605	27.247	45.963
E	45.9	2E-4	-53.434	53.435
a	0.6	0.286	0.170	0.402
ΔH	-39.62	-51.012	-60.208	-41.815

To examine the robustness of the proposed algorithm to poor initial parameter guesses, the AMLE parameter estimation was repeated using the arbitrary value of one as an initial guess for the five model parameters: $\boldsymbol{\theta}_0 = [1,\ 1,\ 1,\ 1,\ 1]^T$. Despite these poor parameter guesses, the algorithm converged to point estimates that are nearly identical to those shown in Table 6.3 and Table 6.4 (identical up to the second decimal place) for the model parameters and the disturbance intensities.

Table 6.5. Correlation matrix for AMLE parameter estimates

	K_{a0}	k_{p0}	E	ΔH	a
K_{a0}	1	-0.2478	-0.0564	-0.4858	-0.9935
k_{p0}		1	-0.1309	0.0542	0.2807
E			1	0.1061	0.0480
ΔH				1	0.4659
a					1

For comparison, we also used a standard WNLS approach, which does not account for model imperfections and process disturbances, to estimate the model parameters. To have a fair comparison, we included the additional steady-state data (from the experimental runs with catalyst) in our WNLS analysis and we also estimated 18 initial conditions for A, C and W for the six experimental runs. First we used the same good initial parameter guesses that were used for AMLE. The estimated parameters did not move very far from the initial guesses, producing the parameter estimates in Table 6.6. This result is not surprising, because the parameter estimation was started from the optimal WNLS parameter estimates obtained by Zheng et al. (2005). The main differences between the WNLS approach reported here and that of Zheng et al. is that Zheng assumed known initial conditions for A, C and W and he did not use steady-state information from the experiments performed with catalyst. In a second trial, when we used the poor initial parameter guesses, the WNLS algorithm converged to unrealistic values possibly corresponding to a local minimum (Table 6.6).

The results in Table 6.6 were obtained using a relative tolerance value of 1E-12 for changes in the objective function and for changes in the norm of the parameter values. Attempts to obtain confidence intervals for the WNLS parameters were unsuccessful because the Jacobian, evaluated at the converged estimates, was very ill-conditioned, indicating severe correlations among the

parameter estimates. Note that Zheng et al.(2005) were able to compute approximate confidence intervals for their parameter estimates, presumably because they assumed perfectly known initial conditions.

Table 6.6. Point estimates for the nylon reactor model parameters using WNLS method

	Estimation Results using Good Initial Guesses		Estimation Results using Poor Initial Guesses	
Parameter	Initial Guess	Estimate	Initial Guess	Estimate
k_{p0}	0.019	0.019	1	1
K_{a0}	22.01	22.0104	1	10.62
E	45.9	45.9	1	1
a	0.6	0.611	1	1.123
ΔH	-39.62	-39.62	1	17.81

6.6 Implementation considerations

Choice of the knot sequence

For three state trajectories and 6 experimental runs, $(3 \times 6 = 18)$ B-spline expansions are required; we used a 4^{th} order B-spine basis (3^{rd} degree). Rich knot sequences are required to ensure that the state trajectories are flexible enough to capture all of the features in the dynamic response. To set up the bases, 60 knots were uniformly placed along the time horizon of each state trajectory for each experimental run. Since the partial pressure of the water in these experimental runs was adjusted using step changes, the water concentration in the molten polymer undergoes fast changes that are almost perfect steps. To accommodate the sharp transitions in the *W* response, 20 extra knots were uniformly placed in the neighbourhood (within

0.5 hours) of the times at which step changes occurred. These extra knots helped remove ripples in the final B-spline expansion of the W trajectory (Poyton et al., 2006) that were apparent using the initial coarse knot sequence. In addition to the refined knot sequence, coincident knots were placed at the times of step changes so that first and higher-order derivatives of the state trajectories are not continuous at the times of the step changes. Overall, 108 B-spline coefficients were used for each of the 18 B-spline expansions.

Calculating the integral in the inner minimization

To calculate the integral in (6.26), a quadrature rule was used; 4 collocation points were placed between every two knots. As noted by Campbell (2007) and Ramsay et al. (2007), the step input in W_{eq} results in discontinuous derivatives for all three model outputs at the times of step changes. As a consequence, a small neighbourhood (within 1E-4 hours) around each of the times at which the step changes occur was removed while calculating the integral in (6.26).

Minimization routines

As discussed above, overall, the inner objective function in this problem should be optimized over $108 \times 18 = 1944$ B-spline coefficients as well as five model parameters $\boldsymbol{\theta}$. Every time that the variables in the outer minimization problem (*i.e.*, the process disturbance intensities) are updated, the inner problem should be solved again. Therefore, it is essential for the AMLE algorithm to take advantage of fast and efficient state-of-the-art minimization routines. Based on our experience, IPOPT (Wächter and Biegler, 2006), which is a nonlinear solver that can be used with AMPL ™ (Fourer et al., 2003), provides an excellent tool for solving nonlinear optimization problems. AMPL™ endows IPOPT with automatic differentiation capability, which eliminates the requirement of providing the nonlinear solver with an analytical or a numerical Jacobian.

Unfortunately, we were not able to implement the complete two-level minimization problem as appears in (6.28) and (6.31) using AMPL™, because complicated user-defined functions (such as the solution to an optimization problem) are not permitted in AMPL™ in a straightforward fashion. Note that each iteration of the outer optimization problem requires the solution of another (inner) minimization problem. Instead, we opted to use nlscon (Nowak and Weimann, 1991) which is a very efficient nonlinear solver in Matlab ™ to solve the outer optimization problem and the combination of AMPL™ and IPOPT to solve the larger inner minimization problem by calling AMPL™ from within Matlab™.

Another issue regarding the inner minimization problem is whether the initial values of the parameter guesses $\boldsymbol{\theta}_0$ for each iteration should be set to the converged values from the previous iteration or the same initial guesses used in the first iteration. We tested both methods for this problem and found that both methods lead to the same point and interval estimates for the model parameters and disturbance intensities.

6.7 Summary and Conclusions

There are two main contributions in this article. First, we re-evaluated the equilibrium-constant expression in the nylon 612 reactor model initially developed and studied by Schaffer et al., (2003b) and Zheng et al., (2005). Using nonlinear regression and steady-state data from six experimental runs without catalyst and three runs with catalyst, a parsimonious semi-empirical model for K_a was selected that adequately fits the data. The main contribution is that the AMLE parameter estimation algorithm was applied to the proposed lab-scale nylon 612 reactor model, which is a difficult and practical chemical engineering example, to estimate the states, parameters, and process disturbance intensities. We showed that using AMLE allows the

modeller to take modelling errors into account and to obtain a measure of these discrepancies by estimating the corresponding process noise intensities.

This parameter estimation problem involved several challenges. The first challenge is to exploit all the available data which consist of six dynamic experimental runs without a catalyst and also steady-state data from three additional experimental runs with a catalyst. It was shown that the AMLE objective function can easily be modified to include the additional steady-state information. The nylon reactor is a multi-response model with three outputs. Two of the outputs, namely the concentration of carboxylic acid end-groups and also the concentration of amine end-groups, are measured at different and non-uniform sampling times, with different levels of accuracy. The third output, water concentration, was not measured. It was demonstrated that AMLE can easily cope with these challenges, since the form of its inner objective function allows for incorporating non-uniform observation times and unmeasured states. Another difficulty encountered in the nylon reactor parameter estimation problem is that initial state conditions are not perfectly known and should be treated as unknowns that need to be estimated. These unknown initial conditions were obtained as a by-product of estimating the state trajectories using AMLE, which is facilitated by using B-spline expansions to represent the state trajectories.

AMLE was successfully applied to the nylon reactor problem, and parameters, states and process disturbance intensities were estimated. Approximate confidence intervals were obtained for model parameters. The interval estimate corresponding to the activation energy parameter, E, contained zero, indicating that there was insufficient dynamic information available to obtain a reliable estimate.

To investigate how sensitive the outcome of AMLE is to initial parameter guesses, we repeated the parameter estimation algorithm using arbitrary initial parameter guesses that were far from the

estimated values. Despite using these poor initial guesses, the algorithm converged to the same point estimates as it had for the good initial guesses. A standard weighted nonlinear least squares algorithm failed to converge to reasonable parameter estimates using the poor initial guesses, and converged to parameter estimates that were very close to the corresponding initial values for the good initial guesses.

6.8 Acknowledgements

The authors thank 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 for financial support and Dr. Tom Harris from Queen's University for technical advice.

6.9 Appendix

6.9.1 Experimental data

In Table 6.7 to Table 6.15 the following units are used:

t: hours, *A*: equiv/10^6g, *C*: equiv/10^6g, *PW*: mm Hg

Table 6.7. Experimental run conducted at T=271 ºC

t	A	C	P_W
0	18.9	90.9	760
1.03	22.7	93	760
2	23.3	91.9	760
2.97	23.8	89.7	760
3.48*	24.9	91.5	760
3.9*	24.2	94.9	760
4.05	25.5	93.4	58
4.25	17.3	88.3	58
4.9	10.2	77.9	58
5.23	6	79.3	58
5.8	4.1	80.6	58
6.78	4.1	76.2	58
7.23	3.9	74.2	58
7.5	4	75.3	58
7.8	4.8	81.5	760
8.17	18.4	83.2	760
8.7	26.1	87.5	760
9.18	26.1	89	760
9.68	27.6	92.2	760
10.2	25.9	90	760
10.65	27	93.4	760
11.05*	27.1	92.4	760

Table 6.8. Experimental run conducted at T=281 ºC

t	A	C	P_W
0	19.6	97.5	760
0.98	23.7	99.8	760
1.9	23.5	101.5	760
2.4	24.4	97.4	760
3*	24.8	98.1	760
3.22	24.7	94	205
3.47	20.2	96.3	205
3.7	16.5	86.6	205
3.98	14.4	88	205
4.23	13.7	89.2	205
4.8	12.8	90.6	205
5.3	13	83.1	205
5.75	12.3	83.8	205
6.27*	12.3	88.3	205
6.42	13.3	86.6	760
6.6	18.9	94.7	760
6.85	23.7	104.2	760
7.1	24.8	102.6	760
7.37	25.9	100.8	760
7.88	26.4	100.8	760
8.38	26.4	103.8	760
8.8	27	107.9	760
9.2*	26	104.7	760

Table 6.9. Experimental run conducted at T=263 ºC

t	A	C	P_W
0	22.4	83.7	760
1.03	25.5	87.9	760
2.08	25.7	85.3	760
3.03	25.8	88.3	760
3.53*	26.7	84	760
4.05*	26.9	90	760
4.25	24.3	82.4	181
4.53	20.4	79.2	181
5.02	16.4	79.5	181
5.55	14.5	74.8	181
6.05	13.5	74.5	181
7.05	11.6	73.7	181
7.48	11.2	72.8	181
7.97	10.3	74.8	181
8.23	13.8	75.6	760
8.62	21	78	760
9.15	23.9	84.4	760
9.63	25	86.3	760
10.1	25.9	85.8	760
11.1	23.8	86.8	760
11.6*	24.1	87.3	760
12.02*	24.9	84.4	760

Table 6.10. Experimental run conducted at T=284 ºC

t	A	C	P_W
2.5	20.6	131.8	760
3*	23.1	129.7	760
3.25	21.2	121.7	152
3.53	14.8		152
3.78	11.4	114.5	152
4.25	10	112.6	152
5.23	10.8		152
5.76	10	128	152
6.27*	10.6	117.2	152
6.53	14.3	132.7	760
6.75	23.7		
7	27	147.6	760
8.08*	25.4	136.3	760
9*	26	138.2	760
9.5*	26.5	137.8	760

Table 6.11. Experimental run conducted at T=284 ºC

t	A	C	P_W
3	12.09		760
3.48*	11.35	202.7	760
3.98	12.99	206.1	760
4.2	9.63	206.2	152
4.52	6.65	203.6	152
4.83	5.94	201.3	152
5.25	5.5	204.1	152
6.98*	4.3	203.2	152
7.5	4.35		152
8*	5.4	199.7	152
8.13	7.7	209.6	152
8.5	11.2	213.7	760
8.83	13.2	217	760
9.5*	14	221.7	760
10.5*	14.3	218.1	760
11	15.5	221.9	760
11.48	15.7		760
12	15.4	226.5	760

Table 6.12. Experimental run conducted at T=284 ºC

t	A	C	P_W
4.38	45.6		760
4.88	46.2		760
5.38	45.6	51.3	760
5.88	49.6		760
6.37	50.3	49.8	760
6.55	42.9	53.4	760
6.78	40.5		152
7.03	37.9	46.7	152
7.28	35.3		152
7.53	34.3	42.2	152
8.03	31.1		152
8.53	30.5	42.8	152
9.05	30		152
9.53	28.3	35.1	152
9.73	32.8	32.3	152
10	36.5		760
10.25	41.3	45.8	760
10.48	41.9		760
10.75	43.5	50.3	760
11.25	45.6		760
11.75	46.3		760
12.25	47.2	53.5	760
12.75*	47.3	51.7	760

Table 6.13. Experimental run conducted at T=289 ºC with catalyst (SHP = 113 ppm)

t	*A*	*C*
2.48	24.40	115.5
3	24.70	113.6
5.97	10.10	101.3
7.53	26.60	122.8
9.2	21.00	119.5
11.88	5.6	112.5

Table 6.14. Experimental run conducted at T=289 ºC with catalyst (SHP = 33 ppm)

t	*A*	*C*
5.00	8.50	88
6.02	8.40	96.7
8.00	23.40	120.2
8.98	23.60	113.2
10.98	4.00	91.5
12	4.10	95.9

Table 6.15. Experimental run conducted at T=289 ºC with catalyst (SHP = 249 ppm)

t	*A*	*C*
5.00	10.60	87.4
5.98	10.60	89.7
7.98	26.60	104.7
8.98	27.30	107.5
11.98	5.30	88.9

6.9.2 Incorporating additional information in AMLE objective function

In Section 6.5, it was discussed that if additional information about the model parameters is available, it can easily be incorporated in the AMLE objective function. To show how this is possible, we consider the SDE in (6.1). We also assume that additional measurements corresponding to the following model are available:

$$y_{ad}(t_i) = g(u(t_i), \boldsymbol{\theta}) + \varepsilon_{ad}(t_i) \quad i=1..N_{ad} \tag{6.33}$$

where N_{ad} is the number of additional measurements and $\varepsilon_{ad}(t_i)$ is the measurement noise with variance σ_{ad}^2, associated with the additional measurement taken at time t_i. We assume that $\varepsilon_{ad}(t_i)$ is independent of the other noise sources in (6.1). The vector of additional measurements is denoted by $\mathbf{y}_{ad}$.

To understand how this additional information can be incorporated, we consider the development in Section 3.3.1 in Chapter 3. Note that $p(\mathbf{y}_m \mid x_0,\ldots,x_q,\boldsymbol{\theta})$ in eq. (3.13) is the only term that is affected by the addition of the extra information. This term can be modified to:

$$p(\mathbf{y}_m, \mathbf{y}_{ad} \mid x_0,\ldots,x_q,\boldsymbol{\theta}) = p(\mathbf{y}_m \mid x_0,\ldots,x_q,\boldsymbol{\theta}) \times p(\mathbf{y}_{ad} \mid x_0,\ldots,x_q,\boldsymbol{\theta}) \tag{6.34}$$

because $\mathbf{y}_m$ and $\mathbf{y}_{ad}$ are independent. Since $\mathbf{y}_{ad}$ does not depend on the state variables, $p(\mathbf{y}_{ad} \mid x_0,\ldots,x_q,\boldsymbol{\theta}) = p(\mathbf{y}_{ad} \mid \boldsymbol{\theta})$. Note that

$$p(\mathbf{y}_{ad} \mid \boldsymbol{\theta}) = \frac{1}{(2\pi)^{N_{ad}/2}\sigma_m^{N_{ad}}} \exp\left\{\frac{-\sum_{i=1}^{N_{ad}}(y_{ad}(t_i) - g(u(t_i),\boldsymbol{\theta}))^2}{2\sigma_{ad}^2}\right\} \tag{6.35}$$

Therefore it can be shown (after taking the natural logarithm and multiplying by a negative sign) that the addition of the extra information can be reflected by adding the following term to the AMLE objective function:

$$\frac{\sum_{i=1}^{N_{ad}} (y_{ad}(t_i) - g(u(t_i), \boldsymbol{\theta}))^2}{2\sigma_{ad}^2} \tag{6.36}$$

6.10 Nomenclature

AMLE	Approximate Maximum Likelihood parameter Estimation	
DOF	Degrees of freedom	
GS	Generalized Smoothing	
MIMO	Multi-Input Multi-Output	
ML	Maximum Likelihood	
NLS	Nonlinear Least-Squares	
PDF	Probability Density Function	
SDE	Stochastic Differential Equation	
SHP	Sodium Hypophosphite	
SISO	Single-Input Single-Output	
SSE	Sum of Squared Errors	
WNLS	Weighted Nonlinear Least-Squares	
A	Concentration of amine end groups,	mol Mg^{-1}
$\mathbf{A}$	Hessian of the inner objective function with respect to the system states	
C	Concentration of carboxylic acid end groups	mol Mg^{-1}
E	Activation energy	kJ mol^{-1}
$E\{.\}$	Expectation operator	
$\mathbf{H}$	Hessian matrix of the inner objective function with respect to model parameters	

ΔH	Apparent enthalpy of polycondensation	kJ mol^{-1}
J_i	The *i*-th component (*i*-th experimental run) of the inner objective function	
J_{inner}	The inner objective function	
L	Concentration of amide links	
K_a	Apparent polycondensation equilibrium constant	
Ka_0	Apparent polycondensation equilibrium constant at T_0 and low water content	
NA_i	The number of observations for A in the *i*-th experimental run	
NC_i	The number of observations for C in the *i*-th experimental run	
Q	Process disturbance intensity	
Q_A	Process disturbance intensity for SDE corresponding to *A*	$(mol\ Mg^{-1})^2/hr$
Q_C	Process disturbance intensity for SDE corresponding to *C*	$(mol\ Mg^{-1})^2/hr$
Q_W	Process disturbance intensity for SDE corresponding to *W*	$(mol\ Mg^{-1})^2/hr$
Q	Vector containing process disturbance intensities	
P_c	Critical pressure of water	kPa
P_w	Partial pressure of water in the gas phase	kPa
P_w^{sat}	Saturation pressure of water in the gas phase	kPa
R	Ideal gas law constant, 8.3145E-3	kJ $mol^{-1}\ K^{-1}$
T	Temperature	K
T_0	Reference temperature, 549.15 K	

T_c	Critical temperature of water, K	
W	Concentration of water, mol Mg^{-1}	
X	Exponent in the kinetic expression for the reaction rate	
a	Empirical model parameter	$(mol\ Mg^{-1})^{0.5}$
f	Nonlinear function	
f_{Ka}	Nonlinear function for *Ka* model	
k_m	Volumetric liquid-phase mass-transfer coefficient for a nylon/water system	h^{-1}
k_p	Apparent polycondensation rate constant	$Mg\ mol^{-1}\ h^{-1}$
k_{p0}	Apparent polycondensation rate constant at reference temperature 549.15 K	
t_{mj}	Time of the *j*-th measurement, hr	
u(t)	The input function	
x_0	The initial state value	
x(t)	The state of the system	
x~(t)	The spline approximation of the system state	
xm	Vector of state values at measurement times	
y(tmj)	The measured value at time *tmj*	
y_{Ka}	*Ka* calculated from the measurements	
ym	Vector of the measurements	
$z_{\alpha/2}$	Normal random deviate corresponding to an upper tail area of $\alpha/2$	

α	Significance level for confidence intervals	
β_i	*i*-th B-spline coefficient	
$\boldsymbol{\beta}$	Vector of B-spline coefficients	
$\delta(.)$	Dirac delta function	
$\varepsilon(t_{mi})$	Normally distributed measurement noise for state *i*	
ε_{Ka}	Approximate error in *Ka* calculated using the measurements	
η_A	Gaussian process disturbance for differential equation of state *A*	mol Mg^{-1}/hr
η_C	Gaussian process disturbance for differential equation of state *C*	mol Mg^{-1}/hr
η_W	Gaussian process disturbance for differential equation of state *W*	mol Mg^{-1}/hr
$\boldsymbol{\theta}$	Vector of model parameters	
$\boldsymbol{\theta}_0$	Initial value for the vector of model parameters	
$\boldsymbol{\theta}_{Ka}$	Vector of *Ka* model parameters	
σ_m^2	Measurement noise variance	
σ_A^2	Measurement noise variance for *A*	$(\text{mol Mg}^{-1})^2$
σ_C^2	Measurement noise variance for *C*	$(\text{mol Mg}^{-1})^2$
$\sigma_{K_a}^2$	Approximate noise variance for *Ka* calculated from measurements	
ϕ_i	*i*-th B-spline basis function	
$\boldsymbol{\varphi}$	Matrix containing all ϕ_i s	

6.11 References

D.M. Bates, D.G. Watts. *Nonlinear Regression Analysis and Its applications.* John Wiley & Sons, Inc., New York, 1988.

C. de Boor. *A practical Guide to Splines.* Springer, New York, 2001.

D.A. *Campbell. Bayesian Collocation Tempering and Generalized Profiling for Estimation of Parameters for Differential Equation Models*. Ph.D. Thesis, McGill University, Montreal, Canada, 2007.

G. Evensen. The ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics.* 53, 2003, 343-367.

R. Fourer, D.M. Gay, B.W. Kernighan. *AMPL a Modeling Language for Mathematical Programming,* 2nd edition. Thomson Brooks Cole, 2003.

J.P.M. Heald, J. Stark. Estimation of noise levels for models of chaotic dynamical systems. *Phisycal Review Letters*. 84, 2000, 2366-2369.

A.H. Jazwinski. *Stochastic Processes and Filtering theory*. Academic Press, New York, 1970.

S.J. Julier, J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE.* 92 (3), 2004, 401-422.

P.E. Kloeden, E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, 1992.

N.R. Kristensen, H. Madsen, S.B. Jorgensen. Parameter estimation in stochastic grey-box models. *Automatica*. 40, 2004, 225-237.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 1*. Academic Press, New York, 1979.

P. S. Maybeck. *Stochastic Models, Estimation, and Control, Volume 2*. Academic Press, New York, 1982.

J.N. Nielsen, H. Madsen, P.C. Young. Parameter estimation in stochastic differential equations: an overview. *Annual Reviews in Control.* 24, 2000, 83-94.

U. Nowak, L. Weimann. A family of Newton codes for systems of highly nonlinear equations. Konrad-Zuse-Zentrum fur Informationstechnik, Berlin, report TR-91-10, 1991.

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. *Computers and Chemical Engineering.* 30, 2006, 698-708.

J.O. Ramsay, G. Hooker, D. Campbell, J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of Royal Statistical Society*: *Series B* (*Statistical Methodology*). 69 (5), 2007, 741-796.

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

S. Roweis, Z. Ghahramani. Learning nonlinear dynamical systems using the Expectation-Maximization algorithm. In *Kalman Filtering and Neural Networks*. S. Haykin, editor. John Wiley & Sons, New York. 2001, 175-220.

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

M.A. Schaffer, E.K. Marchildon, K.B. McAuley, M.F. Cunningham. Prediction of water solubility in nylon melts based on Flory-Huggins theory. *Polymymer Engineering and Science.* 43, 2003a, 639-646.

M.A. Schaffer, K.B. McAuley, M.F. Cunningham, E.K. Marchildon. Experimental study and modeling of nylon polycondensationin the melt phase. *Industrial Engineering and Chemistry Research.* 42, 2003b, 2946-2959.

J. Timmer. Parameter estimation in nonlinear stochastic differential equations. *Chaos, Solitons and Fractals*. 11, 2000, 2571-2578.

M.S. Varziri, K.B. McAuley, P.J. McLellan. "Parameter estimation in continuous-time dynamic models in the presence of unmeasured states and non-stationary disturbances". *Industrial Engineering and Chemistry Research*. 47, 2008a, 380-393.

M.S. Varziri, K.B. McAuley, P.J. McLellan, Parameter estimation in nonlinear stochastic continuous-time dynamic models with unknown disturbance intensity, *Canadian Journal of Chemical Eng*ineering. 2008b, submitted for publication.

M.S. Varziri, A.A. Poyton, K.B. McAuley, P.J. McLellan, J.O. Ramsay. Selecting optimal weighting factors in iPDA for parameter estimation in continuous-time dynamic models. Accepted, to appear in *Computers and Chemical Engineering.* 2008c.

H.U. Voss, J. Timmer, J. Kurths. Nonlinear dynamical system identification from uncertain and indirect measurements. *International Journal of Bifurcation and Chaos.* 14 (6), 2004, 1905-1933.

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

W. Zheng, K.B. McAuley, E.K. Marchildon, and K.Z. Yao. Effects of end-group balance on melt-phase Nylon 612 polycondensation: experimental study and mathematical model. *Industrial Engineering and Chemistry Research.* 44, 2005, 2675-2686.

W. Zheng, K.B. McAuley, E.K. Marchildon, and K.Z. Yao. Melt-phase Nylon 612 polycondensation kinetics: effects of sodium hypophosphite catalyst. *Canadian Journal of Chemical Engineering.* 85, 2007, 180-187.

Chapter 7

Conclusions

7.1 Summary and contributions

A novel algorithm has been developed for parameter and state estimation in continuous-time nonlinear dynamic models, in which modelling errors and process disturbances are present. This proposed parameter estimation approach can help modellers to obtain more-reliable parameter estimates and model predictions that can be used in nonlinear model-based control and optimization schemes.

The research work presented in this thesis builds on the work by Poyton et al. (*Computers and Chemical Engineering* 2006, vol. 30, pages 698-708), where iPDA was first developed to address the difficulties associated with traditional parameter estimation methods. iPDA was developed based on the idea that mathematical models of processes are not exact, therefore, the true state trajectories may deviate from the exact solution of the model differential equations. Using this idea and engineering intuition, the iPDA objective function (shown in eq. (1.6)) was proposed by Poyton et al. The weighting factor, λ, in the objective function determines the extent to which the estimated state trajectories are allowed to deviate from the solution of the model differential equations. Consequently, performance of the iPDA depends significantly on the value of λ. Selecting an optimal weighting factor, however, remained a challenge.

In Chapter 3 of this thesis, application of the iPDA algorithm was extended to nonlinear models with stochastic disturbances. It was demonstrated that in the case of SDE models, minimizing the iPDA objective function corresponds to maximizing the joint probability density function of the model states and measurements, given the model parameters. In other words, it was shown that if continuous Gaussian disturbances are used to represent the modelling errors and process

disturbances, then the iPDA objective function is equivalent to an approximate maximum likelihood criterion. Based on this development, it was shown that the optimal value of the weighting factor λ is proportional to the measurement noise variance, and inversely proportional to the process disturbance intensity (eq. (1.8)). The iPDA objective function was then derived for MIMO systems. In this case, the objective function includes a sum-of-squared errors term for each response, weighted by the reciprocals of its corresponding measurement noise variance, and a model-based penalty term for each differential equation, with the reciprocal of the process disturbance intensity as a weighting factor.

Moreover, it was determined that a simultaneous minimization approach, jointly over model parameters and spline coefficients, can be used to minimize the iPDA objective function, as an alternative to the originally proposed iterative scheme. Note that from Chapter 4 onwards, the simultaneous minimization approach was used to solve iPDA problems. Since the modified algorithm is no longer iterative; and also because the minimized objective function is an approximate maximum likelihood criterion, in Chapters 4, 5, and 6, the proposed algorithm is referred to as Approximate Maximum Likelihood Estimation (AMLE).

The theoretical developments and discussions in Chapter 3 were tested using two examples. In the first example, both traditional NLS and iPDA were used to estimate two parameters in a linear SISO SDE model. Using Monte Carlo simulations, it was shown that parameter estimates obtained from iPDA, using the optimal weighting factor, were less biased and more precise than those obtained using traditional NLS. iPDA was very effective in estimating the underlying true state trajectories. When the iPDA algorithm was tested using weighting factors that were either smaller or larger than the prescribed optimal value, the parameter estimates became more biased and less precise. As the value of the weighting factor was increased, the sampling behaviour of the iPDA parameter estimates became closer to that of the traditional NLS. This observation was

anticipated since a large weighting factor in iPDA forces the estimated state trajectories to satisfy the model differential equations, and hence, any modelling discrepancies are disregarded.

In the second case study, parameter estimation in a nonlinear stochastic CSTR was considered. The outcome of the second example was similar to that in the linear case study. Parameter and state trajectory estimates obtained using iPDA were superior to those obtained using the traditional NLS.

In Chapter 4, the simultaneous minimization approach was used for the case studies and, as previously mentioned, the modified iPDA algorithm was referred to by AMLE. It was shown that AMLE can be applied to parameter estimation in SDE models in which some of the states are not measured. It was demonstrated that to accommodate for this problem, the AMLE objective function can be modified by removing the sum-of-squared error terms that correspond to the unmeasured states. As a consequence, AMLE can also be employed for parameter and state estimation in SDE models in which non-stationary disturbances are present; this can be achieved by treating the non-stationary disturbances as unmeasured states.

Theoretical expressions for approximate confidence intervals for model parameters were derived. Since, in AMLE, an approximate maximum likelihood criterion is minimized, the inverse of the Hessian of the objective function, which is an approximation to the corresponding Fisher information matrix, was used as the covariance matrix of the parameter estimates.

To test our results, four parameters in a MIMO (two-state) nonlinear CSTR with stochastic disturbances (same as the nonlinear CSTR used in Chapter 3), were estimated using AMLE and also traditional NLS in three different scenarios. In the first study, both of the states were assumed to be measured, while in the second scenario one of the two CSTR states was assumed to be unmeasured. The AMLE parameter estimates were, on average, less biased and more

precise than the traditional NLS parameter estimates because, unlike NLS, AMLE properly takes into account possible modelling discrepancies. In the third case, an additive non-stationary disturbance was considered as an input to the material-balance differential equation. In this example, AMLE resulted in significantly more-precise and less-biased parameter estimates in comparison to estimates obtained using the traditional NLS method. The unobserved non-stationary disturbance was also successfully estimated using AMLE. In all of the studied cases, theoretical approximate confidence intervals were obtained and compared to empirical confidence intervals generated using Monte Carlo simulations; theoretical and empirical confidence intervals were consistent.

In Chapter 3 an expression for the optimal weighting factor in iPDA and AMLE was obtained. Calculating this optimal weighting factor, however, requires knowledge of measurement noise variances and also process disturbance intensities that affect the response of the dynamic systems. In engineering applications, knowledge about the quality of measurements can usually be obtained either from repeated measurements or from sensor suppliers. Nonetheless, information about the quality of the model is not easily attainable. In Chapter 5, the AMLE algorithm was modified so that it can accommodate parameter estimation in nonlinear SDE models in which process disturbance intensities are not known *a priori*, but information about the measurement noise variances is available. The modified AMLE algorithm is a two-level nonlinear minimization problem. In this algorithm, appropriate process disturbance intensities are selected, such that the discrepancy between the *a priori* known true value and a maximum likelihood estimate of the measurement noise variance is minimized. The outer level of the algorithm is concerned with minimizing the deviation of the true measurement noise variance from its estimated value by adjusting the process disturbance intensity. In the inner level of the algorithm, the regular AMLE objective function is minimized using the process disturbance intensity

obtained from the outer level optimization problem to obtain the corresponding parameter and state estimates. The updated state and parameter estimates, in turn, are used to evaluate the estimate of the measurement noise variance in the outer level.

To evaluate the performance of the modified AMLE algorithm, a nonlinear MIMO CSTR, similar to that used in Chapters 3 and 4, was used. It was assumed that intensities of the process disturbances that enter the differential equations were unknown, but measurement noise variances were known *a priori*. Initial state conditions were assumed to be unknown. Four parameters, as well as two process disturbance intensities and two state trajectories were estimated using the modified AMLE algorithm. From Monte Carlo simulations, it was observed that the bias of the parameter estimates was negligible. The process disturbance intensity estimates, however, were slightly biased. This bias may be due to approximations involved in calculating the maximum likelihood estimate of the measurement noise variance, which can eventually affect the estimates of the process disturbance intensities. Theoretical confidence intervals were also calculated that were in agreement with the empirical confidence intervals obtained from Monte Carlo simulations. For comparison purposes, parameter estimation in the same case study, with identical settings, was performed using a classical maximum likelihood parameter estimation method. The AMLE algorithm converged faster than the classical maximum likelihood approach. The AMLE parameter estimates, based on this case study, seemed to be less biased and more precise. A significant advantage of the AMLE algorithm is that the overall parameter and state estimation problem can readily be formulated and implemented as a nonlinear programming problem, and recursive solution of Riccati equations to obtain the state covariance matrix is not required.

Theoretical developments in Chapters 3 to 5 were examined merely using simulation case studies. It is very important to apply the proposed AMLE algorithm to a practical chemical engineering

parameter estimation problem, so that the performance of the algorithm can be evaluated in a more realistic scenario. For this reason, in Chapter 6, AMLE was applied to parameter and state estimation in a lab-scale reactor model. It was shown that AMLE can address difficulties that frequently arise when estimating parameters in nonlinear continuous-time dynamic models of industrial processes. Parameter estimation in the nylon reactor model involved several challenges. The first challenge was to exploit all of the available data, which consisted of six dynamic experimental runs without a catalyst and also steady-state data from three additional experimental runs *with* a catalyst. It was shown that the AMLE objective function can easily be modified to include the additional steady-state information. The nylon reactor is a multi-response model with three outputs. Two of the outputs, namely the concentration of carboxylic acid end-groups and also the concentration of amine end-groups, are measured at different and non-uniform sampling times, with different levels of accuracy. The third output, water concentration, was not measured. It was demonstrated that AMLE can easily cope with these challenges, since the form of its inner objective function allows for incorporating non-uniform observation times and unmeasured states. Unknown initial state conditions, another difficulty encountered in the nylon-reactor parameter-estimation problem, was also properly handled by AMLE.

Before proceeding with the parameter estimation procedure, the equilibrium-constant expression in the nylon-reactor model was re-evaluated. Using classical nonlinear regression and steady-state data from six experimental runs without catalyst and three runs with catalyst, a parsimonious semi-empirical model for the apparent equilibrium-constant was selected that adequately fits the data.

After re-evaluating the equilibrium-constant, AMLE was successfully applied to the nylon reactor problem, and parameters, states and process disturbance intensities were estimated. Approximate confidence intervals were obtained for model parameters. The interval estimate corresponding to

the activation energy parameter, E, contained zero, indicating that there was insufficient dynamic information available to obtain a reliable estimate.

The parameter estimation algorithm was repeated using arbitrary initial parameter guesses that were far from the estimated values to investigate the robustness of the AMLE parameter estimates to initial parameter guesses. The performance of AMLE did not deteriorate despite using these poor initial guesses. A standard weighted nonlinear least-squares algorithm failed to converge to reasonable parameter estimates using the poor initial guesses.

The analysis and case studies in this thesis suggest that AMLE is a potentially appealing parameter estimation algorithm that should be further studied and tested for more complicated problems.

The contributions from this thesis consist of the following:

1. Development of an approximate maximum likelihood formulation and algorithm for estimating parameters in process models with model uncertainty and process disturbances. The technique can be used to estimate parameters in stochastic differential equations.
2. Extension of AMLE to handle unknown initial conditions, unmeasured states, and non-stationary disturbances, and development of theoretical expressions for approximate confidence intervals for model parameters.
3. Extension of AMLE to estimate process disturbance intensities with prior information about the measurement noise variances.
4. Demonstration of AMLE using simulated CSTR examples.

5. Application of AMLE to a practical estimation problem with complicated data (dynamic and steady-state), non-uniform measurements, unmeasured states and unknown initial conditions.

7.2 Recommendations for future work

1. The application of AMLE to a lab-scale practical chemical engineering parameter estimation problem was successfully demonstrated in Chapter 6. It is important, however, to study the performance of the proposed algorithm in other case studies including larger-scale industrial problems. This will help us to better understand the potential advantages and possible disadvantages of the AMLE algorithm in problems. The problems should have many states and/or many parameters, and should include unmeasured states, non-uniform sampling, unknown initial conditions, and possibly non-stationary disturbances.

2. Obtaining approximate Confidence Intervals (CIs) for model parameters was discussed in Chapter 4. It was noted that these confidence intervals may be inaccurate for several reasons. For instance: *i*) the AMLE parameter estimates are not in general Normally distributed; *ii*) only an approximation of the Fisher information matrix is used in obtaining the CIs; *iii*) proof of the asymptotic Normality of maximum likelihood parameter estimates involves linearization of the derivative of the likelihood function. The accuracy of the approximation will depend on the nonlinearity of the estimation problem, including the model formulation and parameterization, as well as the amount of data available. It is therefore important to consider alternative inference approaches so that a better understanding of the uncertainty in the AMLE parameter estimates can be

achieved. Insight into the inference characteristics can be drawn from Bayesian analysis (examining the Highest Posterior Density region), profiling, and assessment of curvature.

3. In this thesis, B-spline functions are used as the basis function for discretizing the state trajectories. Nonetheless, the choice of the basis functions is by no means limited to B-spline functions. Other bases, for instance monomials, can also be used. A comparison of AMLE results, when different types of basis functions and knot placement schemes are used, can be beneficial in understanding how sensitive the performance of AMLE is to the discretization strategy.

4. In Chapter 5 and Chapter 6, an approximate maximum likelihood estimator for the measurement noise variance (Helad and Stark, *Physical Review Letters*, 2000, vol. 84, pages 2366-2369) was employed (eqs. (5.13) and (6.30)). In Chapter 6 it was noted that as mentioned by Heald and Stark, the denominator of the measurement noise variance estimator can be regarded as the effective number of degrees of freedom (DOF) set by the dynamics. Therefore, in Chapter 6, different degrees of freedom are accounted for by using additional weightings in the objective function in the outer optimization problem (eq. (6.31)). Addition of these weighting factors ensures that more attention is paid to the variance estimator that has a larger DOF. However, this modification is an *ad hoc* approach to account for DOFs and it is not clear whether this method is optimal in any sense. Further research is required to investigate the optimality of this modification and perhaps to propose alternative approaches to account for the DOF.

5. In Section 4.5.2, a CSTR simulation case study was considered in which a non-stationary disturbance enters the right-hand side of the model differential equation. It was demonstrated that by treating the non-stationary disturbance as an unmeasured state, and

adding the corresponding model-based penalty to the AMLE objective function this problem can be addressed. In this *simulation* study, the existence of the non-stationary disturbance was clear. In practical problems however, it is not clear whether extra terms corresponding to possible non-stationary disturbances should be added to the AMLE objective function or not. Further studies are therefore required to develop an approach to examine the necessity of addition of extra terms corresponding to possible non-stationary disturbances to the AMLE objective function. Examining the autocorrelation of the residuals can perhaps be helpful in this investigation.

6. In developing the AMLE objective function, it was assumed that in multi-response models, there is no correlation between the process noise disturbances and also between the measurement noise variances. These assumptions may be violated in industrial-scale problems. Therefore, it is useful to modify the AMLE objective function so that correlated process disturbances and measurement noise variances can be addressed. It will also be important to identify whether there is correlation among process disturbances, given the model and the data. Investigating the cross-correlations among the residuals may prove helpful in this endeavour.