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

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

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


1. 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 modelers to obtain reliable parameter estimates, but random process disturbances can be even more problematic because they influence the future behavior 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. Modelers 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 because of disturbances that are not included in the model equations and simplifying assumptions that are made during model development.

Approximate maximum likelihood parameter estimation (AMLE) is a novel parameter estimation algorithm that we recently developed to address the problem of parameter estimation in continuous-time nonlinear dynamic models, in which model discrepancies are significant.4–7 A convenient way to account for modeling 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 into stochastic differential equation (SDE) models (as shown in eq 1, below). Until now, the AMLE algorithm has been tested only using simulated data and simple dynamic models of single-phase continuous stirred-tank reactors.4–7 The purpose of this article is to examine and demonstrate the application of AMLE to parameter and state estimation for a two-phase laboratory-scale nylon reactor model, originally developed by Schaffer et al.1 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 were measured, and these measurements were made at irregular sampling times that were convenient for the experimenters.1,2

We show how AMLE can address frequently encountered parameter estimation difficulties such as working with multiresponse models with different levels of measurement accuracy; extracting information from multiple experimental runs with nonuniform sampling times; and dealing with unmeasured states, unknown initial conditions, and unknown levels of modeling error (due to disturbances and structural imperfections). It is also shown that the AMLE framework readily facilitates the integration of additional steady-state or dynamic information that might be available from different sources. Our objective in developing AMLE was to produce a straightforward parameter estimation algorithm that can help modelers to obtain more-reliable parameter estimates and model predictions that can be used in nonlinear model-based control and optimization schemes.

In section 2, the AMLE algorithm is briefly reviewed. In section 3, the laboratory-scale nylon reactor model is introduced. AMLE is then used for parameter estimation in the proposed overall dynamic model. Conclusions are presented in section 4.

2. 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 because of its desirable asymptotic properties.9,10 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.\textsuperscript{11} 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.\textsuperscript{11–18}

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 piecewise 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. In the following paragraphs, we briefly review the AMLE algorithm. We refer the reader to Varziri et al.\textsuperscript{5,7} 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.\textsuperscript{5}

Consider the following continuous-time stochastic dynamic model

\[
\frac{dx(t)}{dt} = f(x(t), u(t), \theta) + \eta(t) \quad x(t_0) = x_0, \quad y(t_0) = x(t_0) + \epsilon(t_0)
\]

(1)

\(x \in \mathcal{R}\) is the state variable, \(u \in \mathcal{R}\) is the input variable, and \(y \in \mathcal{R}\) is the output variable. \(\theta \in \mathcal{R}\) is the vector of unknown model parameters and \(f: \mathcal{R} \times \mathcal{R} \times \mathcal{R}^m \rightarrow \mathcal{R}\) is a nonlinear function of the state variables, the input variables, and the parameters. We assume that \(f\) satisfies some regularity conditions,\textsuperscript{19} so that eq 1 has a unique solution. \(\epsilon\) is a discrete zero-mean uncorrelated normal random variable with variance \(\sigma_\epsilon^2\).

\(\eta(t)\) is a continuous zero-mean stationary white-noise process with covariance matrix \(E(\eta(t) \eta(t + r)) = Q(t)\), 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 \(\Delta t\), where \(\Delta t \to 0\). We also assume that the process disturbance, \(\epsilon(t)\), and the measurement noise, \(\epsilon(t)\), are not correlated. The set of times at which the measurements are available is denoted by \(t_m\) (\(j = 1\) to \(n\)). The measurement times, \(t_m\), do not need to be uniformly spaced. The vector of outputs at observation times \(y(t_m)\) (\(j = 1\) to \(n\)) and its corresponding state vector of true values \(x(t_m)\) (\(j = 1\) to \(n\)) are denoted by \(y_m\) and \(x_m\), respectively. [Note that this dynamic model can be more rigorously written as \(dx(t) = f(x(t), u(t), \theta)\) \(dt + d\eta(t)\), where \(d\eta(t)\) is a Wiener process.\textsuperscript{20}]

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. Known measurement noise variance, \(\sigma_\epsilon^2\), but unknown process disturbance intensity \(Q\), Varziri et al.\textsuperscript{7} used an ML argument\textsuperscript{21} 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, \(\hat{\sigma}_\epsilon^2\), is close to the known value of \(\sigma_\epsilon^2\). The two-step optimization scheme proposed by Varziri et al.\textsuperscript{7} can be summarized as follows

**Outer optimization problem:**

\[
\hat{Q} = \arg \min_{Q} \left( \frac{\sigma^2_m(Q)}{\sigma^2_m} - 1 \right)^2
\]

(2)

**Inner optimization problem:**

\[
\begin{aligned}
\hat{\theta}(t) &= \arg \min_{\theta} \frac{(y_m - x_m)^T(y_m - x_m)}{2\sigma^2_m} + \frac{1}{2Q} \int_{t_0}^{t_f} \left\{ \frac{dx}{dt} - f(x(t), u(t), \theta) \right\}^2 dt \\
\hat{\beta}_i(t) &= \hat{\beta}_i(t) + \frac{c}{y_m - x_m} \int_{t_0}^{t_f} \left\{ \frac{dx}{dt} - f(x(t), u(t), \theta) \right\} dt
\end{aligned}
\]

(3)

where \([t_0, t_f]\) is the time span over which the measurements are taken. The outer optimization minimizes the discrepancy between the estimated and known measurement variances, and the inner optimization minimizes the criterion in eq 3 using the value of \(Q\) obtained from the outer optimization.

Because \(x(t)\) is unknown, minimizing eq 3 over \(x(t)\) and \(\theta\) is an infinite-dimensional optimization problem (a calculus of variations problem) that 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 because of their compact support and other favorable properties\textsuperscript{4,22,23}.

\[
x(t) = x_0(t) = \sum_{i=1}^{c} \beta_i q_i(t) = q^T(t) \beta
\]

(4)

where \(\beta_i, i = 1 - c\), represents B-spline coefficients; \(\beta\) is a vector of \(c\) B-spline coefficients; \(\phi_i(t), i = 1 - c\), represents B-spline basis functions;\textsuperscript{24} and \(q(t)\) is a vector containing the \(c\) basis functions. Note that other basis functions could readily be instead of B-splines.\textsuperscript{23}

By substituting eq 4 into eq 3, we obtain the following finite-dimensional inner optimization problem

\[
\hat{\theta}(t) = \arg \min_{\theta} \frac{(y_m - x_m)^T(y_m - x_m)}{2\sigma^2_m} + \frac{1}{2Q} \int_{t_0}^{t_f} \left\{ \frac{dx}{dt} - f(x(t), u(t), \theta) \right\}^2 dt
\]

(5)

The objective function in eq 5 can be minimized either simultaneously over \(\theta\) and \(\beta\) or in an iterative way as described by Varziri et al.\textsuperscript{5} Minimizing eq 5 provides point estimates for the model parameters and the spline coefficients, given the current estimate of the process disturbance intensity \(\hat{Q}\). The estimated spline coefficients can then be used to determine the estimated state trajectory, \(x(t)\), which, in turn, is used to evaluate \(\hat{\sigma}_m^2\) from eq 6 below. The variance estimate, \(\hat{\sigma}_m^2\), in the objective function of eq 8 is the approximate ML estimator developed by Head and Stark\textsuperscript{21}

\[
\hat{\sigma}_m^2 = \frac{[y_m - x_m(Q)]^T[y_m - x_m(Q)]}{n} + \frac{\text{trace}(A^{-1})}{n} \hat{\theta} (6)
\]

where \(A\) is the Hessian matrix of the objective function in eq 3 with respect to the discrete state variables, evaluated at the converged parameter and state estimates. Varziri et al.\textsuperscript{5} showed that if the SDE model in eq 1 is discretized using Euler’s method with a very small \(\Delta t\) as the discretization interval, minimizing eq 3 (and approximately minimizing eq 5) is equivalent to minimizing \(-\ln p(y_m|x(\theta))\), where \(x\) is the vector of discretized states and \(p\) denotes the probability density function. Because the state trajectory is approximated in the inner optimization step in AMLE using a B-spline expansion, matrix \(A\) can also be approximated in terms of the Hessian matrix of the objective function in eq 5 with respect to the B-spline coefficients \(\beta\) evaluated at \(\theta\) and \(\beta\). This information is often available from
solving the inner optimization problem in eq 5 and does not need to be recalculated.

The converged results of this overall optimization problem (eqs 2 and 5) provide the modeler with \( \hat{\theta}, \hat{\beta}, \) and \( \hat{Q} \). \( \hat{\theta} \) is the desired estimate for the fundamental model parameters. The estimated spline coefficients \( \hat{\beta} \), can be used to obtain the estimated state trajectory \( x^* \), and \( \hat{Q} \) provides information about the magnitude of the model uncertainty and process disturbances. A diagram summarizing the algorithm appears in Figure 1.

Note that the outer optimization (eq 3) is an ad hoc way to ensure that the estimated intensity results in an estimate for the measurement noise variance that is consistent with its known value. Therefore, even though the parameters in the inner optimization problem in eq 5 need to be recalculated, solving the inner optimization problem in eq 5 and does not need to be recalculated.

3. Case Study: Laboratory-Scale Nylon 612 Reactor Model

3.1. Nylon 612 Mathematical Model. Nylons are widely used polymers that 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):

\[
\text{COOH}_C + \text{H}_2\text{N}_A \rightarrow \text{CONH}_L + \text{H}_2\text{O}_W
\]

The forward reaction is a polyamidation reaction in which carboxyl and amine end groups are consumed to form amide links and water. In the reverse hydrolysis reaction, amide links are broken to form carboxyl and amine end groups.

The purpose of the experimental study and modeling of nylon reactions conducted by Shaffer et al.\(^1\) and Zheng et al.\(^2\) was 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.

The following equations describe the dynamic behavior of the liquid contents of a well-stirred melt-phase nylon polymerization reactor (see Figure 2), through which a gaseous mixture of nitrogen and steam is bubbled. Refer to Schaffer et al.\(^1\) and Zheng et al.\(^2\) for detailed information about the reactor equipment and the development of the material-balance equations.

\[
\frac{dL}{dt} = -\frac{dA}{dt} = -\frac{dC}{dt} = k_p \left( CA - \frac{LW}{K_a} \right) = f_1(\theta,C,A,L,W) \tag{7}
\]

\[
\frac{dW}{dt} = k_m \left( CA - \frac{LW}{K_a} \right) - k_m (W - W_{\text{eq}}) = f_2(\theta,C,A,L,W,W_{\text{eq}}) \tag{8}
\]

In eqs 7 and 8 is the temperature-dependent polycondensation rate constant, and \( k_{p0} \) is the polycondensation rate constant at the reference temperature, \( T_0 = 549.15 \text{ K} \), which was chosen to be in the middle of the temperature range over which the experiments were conducted. \( K_a \) is the apparent equilibrium constant, and \( k_m = 24.3 \text{ h}^{-1} \) is a mass-transfer coefficient that was previously estimated\(^4\) for the conditions encountered in the laboratory-scale reactor.

The concentration of water in the polymer melt, \( W_{\text{eq}} \), that would be in equilibrium with the water vapor in the gas bubbles is determined using a Flory–Huggins-based expression\(^26\)

\[
W_{\text{eq}} = 5.55 \times 10^4 \frac{P_{w}}{P_{w}^{\text{sat}}} \exp(-9.624 + 3613/T) \tag{10}
\]

where the saturation vapor pressure of the water in the gas bubbles, \( P_{w}^{\text{sat}} \), can be calculated using the Wagner equation

\[
\ln(P_{w}^{\text{sat}}/P_c) = [-7.77224(1 - T/T_c) + 1.45684(1 - T/T_c)^{1.5} - 2.71492(1 - T/T_c)^{3} - 1.41336(1 - T/T_c)^{6}]/(T/T_c) \tag{11}
\]

where \( P_c \) 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 material balance equation\(^1\)

\[
155.23L = 10^{6} - 115.15C - 58.10A - 18.02W \tag{12}
\]

Note that, because the concentration of the amide linkages \( L \) can be computed algebraically from eq 12, only the differential equations for \( A, C, \) and \( W \) need to be solved.

To allow for possible modeling errors and process disturbances, stochastic terms can be added to the differential equations.
where $\eta_A$, $\eta_C$, and $\eta_W$ are continuous-time Gaussian disturbances with intensities $Q_A$, $Q_C$, and $Q_W$ [(mol Mg\(^{-1}\))\(\cdot\)h], respectively.

Schaffer et al.\(^1\) noticed that the apparent polycondensation equilibrium constant, $K_a$, depends on the water concentration as well as the temperature of the molten nylon. Zheng et al.\(^2\) were concerned that activity coefficients for the end groups and amide links might also be influenced by temperature and therefore modified semiempirical $K_a$ model proposed by Schaffer et al.\(^1\) Six parameters were estimated in their nylon reactor model using a weighted nonlinear least-squares (WNLS) approach. Initial values of the states, which are uncertain, were assumed to be perfectly known and were not estimated. In this article, it will be shown that the AMLE algorithm naturally facilitates estimation of unknown initial state conditions, while properly accounting for different measurement variances for measured states. Many of the parameter values estimated from Zheng et al.’s model and the available data showed very high correlations.

Campbell\(^3\) and Ramsay et al.\(^22\) considered parameter estimation in a simplified version of Zheng et al.’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 $\theta$. 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 for manual tuning. In AMLE, the disturbance intensity, $Q$, are estimated using the outer optimization so that the estimated measurement noise variance is consistent with prior knowledge about the quality of the measurements.

To resolve concerns about possible overparameterization in Zheng et al.’s expression for the apparent equilibrium constant, Varziri\(^26\) used additional steady-state data\(^27\) and selected the following semiempirical expression for $K_a$:

$$K_a = \left[ \frac{1 + a_0W_{eq}}{\gamma_W W_{eq}} \right] K_{eq} \exp \left[ -\frac{\Delta H}{R} \left( \frac{1}{T} - \frac{1}{T_0} \right) \right]$$  \hspace{1cm} (14)

where the subscript $eq$ indicates equilibrium concentrations, which are available from steady-state data. From the steady-state data, $K_a$ was calculated and treated as a measured response ($y_k$). The variance $\sigma_{y_k}^2$ of the error, $\epsilon_k$, was estimated to be 94.9. Varziri\(^26\) denoted any postulated semiempirical submodel for $K_a$ by

$$y_{k_2} = f_k(\theta_{k_2} W_{eq}, T) + \epsilon_{k_2}$$  \hspace{1cm} (16)

Eight possible semiempirical submodels for $K_a$ were considered, including those previously proposed.\(^1\)-\(^3\) For each of these candidate submodels, the nonlinear regression problem in eq 16 was solved, using the available steady-state data (see section 3.3), so that estimates for empirical parameters $\theta_{k_2}$ could be obtained. Based on the sum of squared residuals, qualitative analysis of residual plots, and also approximate individual confidence intervals for parameter estimates, the model in eq 14 was concluded to be the most suitable. The estimated parameters are $a = 0.60$ mol\(^{-1}\).Mg\(^{-1}\) and $\gamma_W = 22.01$, and $\Delta H = -39.62$ kJ mol\(^{-1}\), respectively.

### 3.3. Experimental Data

Note that all of the data used in this article result from previous experiments conducted by Schaffer et al.\(^1\) and Zheng et al.\(^2\),\(^27\) Recent steady-state data that were not considered in previous parameter estimation studies\(^1\)-\(^3\) arise from experimental runs conducted to study the effects of sodium hypophosphite catalyst on the polycondensation kinetics.\(^27\)

Of the six dynamic experimental runs that are available for parameter estimation, the first three were conducted at temperatures of 263, 271, and 281 °C, respectively, whereas the last three runs were conducted at 284 °C. The additional steady-state data arise from experimental runs at 290 °C. During each experimental run, the concentrations of $A$ and $C$ are measured at several nonuniformly 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\(^{-1}\). Because 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 an unmeasured dynamic state. The equilibrium water concentration, $W_{eq}$, in the molten nylon can be computed from the input variable $P_W$ using eq 10.

### 3.4. Parameter Estimation Results

In this section, we use the objective function of the form given in eqs 2 and 3 to estimate five parameters, $\theta = [k_{ps}, K_{eq}, E, a, \Delta H]^T$, along with three process disturbance intensities, $Q = [Q_A, Q_C, Q_W]^T$, and three state trajectories in the nylon 612 reactor model described by eqs 9–14.

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 eq 4

$$A_i(t) = q_{A_i}(t)^T \beta_{A_i}, \hspace{1cm} C_i(t) = q_{C_i}(t)^T \beta_{C_i}, \hspace{1cm} W_i(t) = q_{W_i}(t)^T \beta_{W_i}$$  \hspace{1cm} (17)

Because $W$ is not measured, there is no SSE term associated with this component. The AMLE objective function corresponding to the $i$th run for this multiresponse model with an unmeasured state\(^6\) becomes
\[
J_i = \frac{1}{2g_C} \sum_{j=1}^{N_C} [y_C(t_j) - C_j(t_j)]^2 + \frac{1}{2g_C} \int_{t_{c0}}^{t_{c1}} \left\{ \frac{dC_j(t)}{dt} + f_k[A_j(t),C_j(t),W_j(t),L_j(t),\theta] \right\}^2 dt + \frac{1}{2g_A} \int_{t_{c0}}^{t_{c1}} \left\{ \frac{dA_j(t)}{dt} + f_k[A_j(t),C_j(t),W_j(t),L_j(t),\theta] \right\}^2 dt + \frac{1}{2g_W} \int_{t_{c0}}^{t_{c1}} \left\{ \frac{dW_j(t)}{dt} - f_2[A_j(t),C_j(t),W_j(t),W_{eq}(t),L_j(t),\theta] \right\}^2 dt
\]

where \(N_A\) and \(N_C\) are the numbers of \(A\) and \(C\) measurements, respectively, in the \(i\)th experimental run.

Because 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.

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

\[
J_{inner} = \sum_{i=1}^{6} J_i
\]

Using the objective function in eq 19, the extra steady-state information available from the three experimental runs with catalyst would be disregarded. To include this extra information, the objective function in eq 19 can be modified as follows

\[
J_{inner} = \sum_{i=1}^{6} J_i + \frac{1}{\sigma_K^2} \sum_{j=1}^{3} \left\{ y_{ki} - f_k(\theta_{ki},W_{eq}(T)) \right\}^2
\]

where \(f_k(\theta_{ki},W_{eq}(T))\) is defined by the right-hand side of eq 14. This modification arises naturally from the ML development when \(y_{ki}\) is considered as an additional nondynamic measured response.

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

\[
\hat{\sigma}_x^2 = \frac{1}{\sum_{j=1}^{N_x}} \left( \frac{1}{N_x} \sum_{j=1}^{N_x} (y_{xi}(t_j) - x_{xi}(t_j))^2 \right) + \frac{\text{trace}(A_{x^{-1}})}{x = A,C}
\]

where \(A_{x^{-1}}\) is the inverse of the Hessian of the inner objective function, \(J_{inner}\), with respect to the discrete state variables. Denoting \(\tau_x = 1/\hat{\sigma}_x^2 \text{trace}(A_{x^{-1}})\), the second term on the right-hand side of eq 21 can be written as \(\tau_x \hat{\sigma}_x^2 \sum_{j=1}^{N_x} N_x\). By moving this term to the left-hand side of eq 21 and factoring out \(\hat{\sigma}_x^2\), eq 21 can be rearranged\(^{21}\) to

\[
\hat{\sigma}_x^2 = \frac{1}{\sum_{j=1}^{N_x}} \left( \frac{1}{N_x} \sum_{j=1}^{N_x} (y_{xi}(t_j) - x_{xi}(t_j))^2 \right) x = A,C
\]

In this article, the known values of the measurement noise variance, \(\sigma_x^2\), is used in place of \(\hat{\sigma}_x^2\) to calculate \(\nu_x\). The denominator in eq 22 is the degrees of freedom, \(\text{DOF}_x = \sum_{j=1}^{N_x} N_x\).

The outer optimization problem can then be written as

\[
\hat{Q} = \arg \min_{Q} \left\{ \text{DOF}_{A} \left[ \frac{\sigma_A^2(Q)}{\sigma_A^2} - 1 \right] + \text{DOF}_{C} \left[ \frac{\sigma_C^2(Q)}{\sigma_C^2} - 1 \right] \right\}
\]

Note that the terms in the objective function of eq 23 are weighted by \(\text{DOF}_{A}\) and \(\text{DOF}_{C}\) to account for the different numbers of available measurements for the amine and carboxyl end groups.

Three of the parameters, namely, \(a\), \(K_{do}\), and \(\Delta H\), were estimated by Varziri\(^{26}\) using the steady-state data (section 3.2). These parameters were estimated again using AMLE, and the previously estimated values of \(a = 0.60 \text{ mol}^{-0.5} \text{ Mg}^{0.5}\), \(K_{do} = 22.01\), and \(\Delta H = -39.62 \text{ kJ mol}^{-1}\) were used as initial parameter guesses. For the rest of the parameters, estimated value reported by Zheng et al.\(^{2}\) were used as initial values, \(\theta_{0}\), as shown in Table 2, below.

The initial values for the spline coefficients, \(\beta_A\) and \(\beta_C\), \((i = 1-6)\) were obtained by fitting a smoothing spline\(^{23}\) to the measured data. Because the water concentration is not measured, the initial guesses for the spline coefficients \(\beta_W\), \((i = 1-6)\) were obtained by fitting a smoothing spline to \(W_{eq}\), which was calculated from eq 10. 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 (nonzero) initial guesses for B-spline coefficients is not necessary but can reduce the convergence time. The initial values of the process disturbance intensities were arbitrated set to 1. The estimated intensities, \(\hat{Q}\), are shown in Table 1. The estimated parameters, \(\hat{\theta}\), are reported in Table 2. The results in Tables 1 and 2 were obtained using the following stopping conditions: For the inner problem, the optimizer stopped when the objective function changed by less than \(1 \times 10^{-8}\); the outer optimizer stopped when the value of the objective function changed by less than \(1 \times 10^{-2}\).

The final B-spline fits using the estimated parameters and coefficients are shown in Figure 3. The numerical solution of the model equations, which does not account for the stochastic process disturbances, is shown in Figure 4. Initial values for the output trajectories in Figure 4 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.

**Table 1. Process Disturbance Intensity Estimates**

<table>
<thead>
<tr>
<th>intensity</th>
<th>estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Q_A)</td>
<td>4.37</td>
</tr>
<tr>
<td>(Q_C)</td>
<td>15.56</td>
</tr>
<tr>
<td>(Q_W)</td>
<td>0.88</td>
</tr>
</tbody>
</table>

\(a\) All intensities have units of (mol Mg\(^{-1}\)) h\(^{-1}\).

**Table 2. Point Estimates and Approximate 95% Confidence Intervals for the Nylon Reactor Model Parameters**

<table>
<thead>
<tr>
<th>parameter</th>
<th>initial guess</th>
<th>estimate</th>
<th>lower bound</th>
<th>upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_{do})</td>
<td>0.019</td>
<td>0.0129</td>
<td>0.0106</td>
<td>0.0152</td>
</tr>
<tr>
<td>(K_{do})</td>
<td>22.01</td>
<td>36.605</td>
<td>27.247</td>
<td>45.963</td>
</tr>
<tr>
<td>(E)</td>
<td>4.59</td>
<td>2E-4</td>
<td>-53.434</td>
<td>53.435</td>
</tr>
<tr>
<td>(\Delta H)</td>
<td>-39.62</td>
<td>-51.012</td>
<td>-60.208</td>
<td>-41.815</td>
</tr>
</tbody>
</table>

\(a\) All parameters have units of (mol Mg\(^{-1}\)) h\(^{-1}\).
variance associated with C is larger than that of A ($\sigma_A = 0.6$ mol Mg$^{-1}$ and $\sigma_C = 2.4$ mol Mg$^{-1}$).

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.$^6,8$ We approximate the Fisher information matrix using the Hessian ($H$) of the inner objective function with respect to the model parameters $\theta$ evaluated at the converged values. Approximate 100(1 - $\alpha$)% confidence intervals for the model parameters can be obtained as follows

$$\theta = \hat{\theta} \pm \frac{z_{\alpha/2}}{\sqrt{\text{diag}[H^{-1}(\hat{\theta})]}}$$  \hspace{1cm} (24)

Care should be taken in interpreting the confidence intervals obtained from eq 24; these intervals are approximate and also do not take into account the uncertainty in the estimates of the process disturbance intensities.

Approximate 95% confidence intervals for the parameters are presented in Table 2, and a correlation matrix for the parameter estimates is presented in Table 3. The confidence interval corresponding to the activation energy $E$ contains zero. This is consistent with the findings of Campbell$^3$ and Ramsay et al.$^{22}$ 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 3 indicates that correlations between the parameter estimates are small, except for between $K_{a0}$ and $a$. These two parameters are nearly perfectly negatively correlated.

To examine the robustness of the proposed algorithm to poor initial parameter guesses, the AMLE parameter estimation was repeated using the arbitrary value of 1 as an initial guess for the five model parameters: $\theta_0 = [1, 1, 1, 1, 1]^T$. Despite these poor parameter guesses, the algorithm converged to point estimates that are nearly identical to those shown in Tables 1 and 2 (identical up to the second decimal place) for the model parameters and the disturbance intensities.

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 4. This result is not surprising, because the parameter estimation was started from the optimal WNLS parameter estimates obtained by Zheng et al.$^2$ The main differences between the WNLS approach reported here and that of Zheng et al.$^2$ is that Zheng et al. assumed known initial conditions for A, C, and W and 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 4). The results in Table 4 were obtained using a relative tolerance value of $1 \times 10^{-12}$ for changes in the objective function and for changes in the norm of the parameter values. Attempts to obtain confidence intervals for these parameters using the WNLS algorithm were unsuccessful.

Figure 3. Optimal B-spline trajectories of A, C, and W for six experimental runs (—, B-spline fit; •, measured data).
intervals for the WNLS parameters were unsuccessful because
the Jacobian, evaluated at the converged estimates, was very
ill-conditioned, indicating severe correlations among the pa-
rameter estimates. Note that Zheng et al. were able to compute
approximate confidence intervals for their parameter estimates,
presumably because they assumed perfectly known initial
conditions.

3.5. Implementation Considerations. Choice of the
Knot Sequence. For three state trajectories and six experimental
runs ($3 \times 6 = 18$), B-spline expansions are required; we used
a fourth-order B-spline basis (third-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.
Because the partial pressure of the water in these experimental
runs was adjusted using step changes, the water concentration
in the molten polymer underwent fast changes that were almost
perfect steps. To accommodate the sharp transitions in the
$W$ response, 20 extra knots were uniformly placed in the neighbor-
hood (within 0.5 h) of the times at which step changes occurred.
These extra knots helped remove ripples in the final B-spline
expansion of the $W$ trajectory 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 were not continuous at the times of the step changes.
Overall, 108 B-spline coefficients were used for each of the 18
B-spline expansions.

If the process disturbances and structural model imperfections
are small, the values of the B-spline coefficients are determined
by requiring the estimated trajectory to closely follow the
solution to the differential equations. However, if the process
disturbances and model imperfections are large, the measured
data have a larger effect on the estimated value of the B-spline
coefficients, and having too many coefficients, compared to the
number of data points, can lead to large variances for the
estimated coefficients.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial guess</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{p0}$</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td>$K_{a0}$</td>
<td>22.01</td>
<td>22.0104</td>
</tr>
<tr>
<td>$E$</td>
<td>45.9</td>
<td>45.9</td>
</tr>
<tr>
<td>$\Delta H$</td>
<td>-39.62</td>
<td>-39.62</td>
</tr>
</tbody>
</table>

Table 4. Point Estimates for the Nylon Reactor Model Parameters
Using the WNLS Method
Calculating the Integral in the Inner Minimization. To calculate the integral in eq 18, a quadrature rule was used; four collocation points were placed between every two knots. As noted by Campbell and Ramsay et al., 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 neighborhood (within $1 \times 10^{-4}$ h) around each of the times at which the step changes occur was removed during calculation of the integral in eq 18.

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 $\theta$. 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, which is a nonlinear solver that can be used with AMPL, 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. The Hessian matrix that is used in eq 21 was obtained using the automatic differentiation feature in AMPL.

Unfortunately, we were not able to implement the complete two-level minimization problem as appears in eqs 20 and 23 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 nlson, 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, $\theta_0$, for each iteration should be set to the converged values from the previous iteration or the same initial guesses as used in the first iteration. We tested both methods for this problem and found that both methods led to the same point and interval estimates for the model parameters and disturbance intensities.

4. Summary and Conclusions

In this article, the AMLE parameter estimation algorithm was applied to a laboratory-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 modeler to take modeling 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 was to exploit all of the available data, which consisted of data from 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 multiresponse model with three outputs. Two of the outputs, namely, the concentration of carboxylic acid end groups and the concentration of amine end groups, were measured at different and nonuniform 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, because the form of its inner objective function allows for the incorporation of nonuniform 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 byproduct 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 was to initial parameter guesses, we repeated the parameter estimation algorithm using arbitrary initial parameter guesses that were far from the estimated values. Despite 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.

Acknowledgment

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.

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
SISO = single-input single-output
SSE = sum of squared errors
WNLS = weighted nonlinear least-squares
$A$ = concentration of amine end groups, mol Mg$^{-1}$
$A$ = 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$^{-1}$
$E[\cdot]$ = expectation operator
$H$ = Hessian matrix of the inner objective function with respect to model parameters
$\Delta H$ = apparent enthalpy of polycondensation, kJ mol$^{-1}$
\( J_i = \) \( i \)th component (\( i \)th experimental run) of the inner objective function  
\( J_{\text{outer}} = \) inner objective function  
\( K_a = \) apparent polycondensation equilibrium constant  
\( K_0 = \) apparent polycondensation equilibrium constant at \( T_0 \) and low water content  
\( L = \) concentration of amide links  
\( N_A = \) number of observations for \( A \) in the \( i \)th experimental run  
\( N_C = \) number of observations for \( C \) in the \( i \)th experimental run  
\( P_c = \) critical pressure of water, kPa  
\( P_w = \) partial pressure of water in the gas phase, kPa  
\( P_{sat} = \) saturation pressure of water in the gas phase, kPa  
\( Q = \) process disturbance intensity  
\( Q_{1i} = \) process disturbance intensity for SDE corresponding to \( A \), (mol Mg\(^{-1}\))\( h^{-1} \)  
\( Q_{2i} = \) process disturbance intensity for SDE corresponding to \( C \), (mol Mg\(^{-1}\))\( h^{-1} \)  
\( Q_{0i} = \) process disturbance intensity for SDE corresponding to \( W \), (mol Mg\(^{-1}\))\( h^{-1} \)  
\( Q = \) vector containing process disturbance intensities  
\( R = \) ideal gas law constant, \( 8.3145 \times 10^{-3} \) 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  
\( f_K = \) nonlinear function for \( K_a \) model  
\( k_m = \) volumetric liquid-phase mass-transfer coefficient for a nylon/water system, \( h^{-1} \)  
\( k_a = \) apparent polycondensation rate constant, Mg mol\(^{-1}\) h\(^{-1} \)  
\( k_{ph} = \) apparent polycondensation rate constant at the reference temperature, \( 549.15 \) K  
\( t_m = \) time of the \( j \)th measurement, h  
\( u(t) = \) input function  
\( x_0 = \) initial state value  
\( x = \) state of the system  
\( x_i(t) = \) spline approximation of the system state  
\( x_m = \) vector of state values at measurement times  
\( y(t_m) = \) measured value at time \( t_m \)  
\( y_{Ki} = K_a \) calculated from the measurements  
\( y_n = \) vector of the measurements  
\( z_{0\alpha} = \) normal random deviate corresponding to an upper tail area of \( \alpha/2 \)  
\( \alpha = \) significance level for confidence intervals  
\( \beta_i = \) \( i \)th B-spline coefficient  
\( \beta = \) vector of B-spline coefficients  
\( \delta_i = \) Dirac delta function  
\( \varepsilon(t_m) = \) normally distributed measurement noise for state \( i \)  
\( \varepsilon_{Ki} = \) approximate error in \( K_a \) calculated using the measurements  
\( \eta_a = \) Gaussian process disturbance for differential equation of state \( A \), mol Mg\(^{-1}\) h\(^{-1} \)  
\( \eta_C = \) Gaussian process disturbance for differential equation of state \( C \), mol Mg\(^{-1}\) h\(^{-1} \)  
\( \eta_W = \) Gaussian process disturbance for differential equation of state \( W \), mol Mg\(^{-1}\) h\(^{-1} \)  
\( \theta = \) vector of model parameters  
\( \theta_0 = \) initial value for the vector of model parameters  
\( \theta_{Ki} = \) vector of \( K_a \) model parameters  
\( \sigma^2 = \) measurement noise variance  
\( \sigma^2_A = \) approximate noise variance for \( K_a \) calculated from measurements  
\( \sigma^2_C = \) measurement noise variance for \( C \), (mol Mg\(^{-1}\))^\(^2 \)  
\( \phi = \) matrix containing all \( \phi \) values

**Literature Cited**


Received for review March 31, 2008
Revised manuscript received June 2, 2008
Accepted July 9, 2008