# PARAMETER ESTIMATION TECHNIQUES FOR NONLINEAR DYNAMIC MODELS WITH LIMITED DATA, PROCESS DISTURBANCES AND MODELING ERRORS

by

Hadiseh Karimi

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

(December, 2013)

# Abstract

In this thesis appropriate statistical methods to overcome two types of problems that occur during parameter estimation in chemical engineering systems are studied. The first problem is having too many parameters to estimate from limited available data, assuming that the model structure is correct, while the second problem involves estimating unmeasured disturbances, assuming that enough data are available for parameter estimation. In the first part of this thesis, a model is developed to predict rates of undesirable reactions during the finishing stage of nylon 66 production. This model has too many parameters to estimate (56 unknown parameters) and not having enough data to reliably estimating all of the parameters. Statistical techniques are used to determine that 43 of 56 parameters should be estimated. The proposed model matches the data well. In the second part of this thesis, techniques are proposed for estimating parameters in Stochastic Differential Equations (SDEs). SDEs are fundamental dynamic models that take into account process disturbances and model mismatch. Three new approximate maximum likelihood methods are developed for estimating parameters in SDE models. First, an Approximate Expectation Maximization (AEM) algorithm is developed for estimating model parameters and process disturbance intensities when measurement noise variance is known. Then, a Fully-Laplace Approximation Expectation Maximization (FLAEM) algorithm is proposed for simultaneous estimation of model parameters, process disturbance intensities and measurement noise variances in nonlinear SDEs. Finally, a Laplace Approximation Maximum Likelihood Estimation (LAMLE) algorithm is developed for estimating measurement noise variances along with model parameters and disturbance intensities in nonlinear SDEs. The effectiveness of the proposed algorithms is compared with a maximum-likelihood based method. For the CSTR examples studied, the proposed algorithms provide more accurate estimates for the parameters. Additionally, it is shown that the performance of LAMLE is superior to the performance of FLAEM. SDE models and associated parameter estimates obtained using the proposed

techniques will help engineers who implement on-line state estimation and process monitoring schemes.

# Co-Authorship

The material in Chapter 2 has been published in Macromolecular Reaction Engineering and the material in Chapter 3 has been accepted for publication in the Canadian Journal of Chemical Engineering. The material in Chapter 4 has been submitted to Industrial and Engineering Chemistry Research and the material in Chapter 5 has been prepared for submission to a journal. 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

I would like to thank my great supervisor Dr. Kim McAuley for her excellent support and encouragement during my PhD studies. This work would not have been possible without her continuous guidance and outstanding supervision. I will forever be thankful for her continuous support, encouraging meetings and wise advice.

I would also like to express my gratitude to Dr. E.K. Marchildon and Dr. Saeed Varzir for their technical advice.

I gratefully acknowledge financial support provided by Mprime, NSERC (the Natural Sciences and Engineering Research Council of Canada), Hatch and the School of Graduate Studies, Queen's University.

I would like to thank my dear friends and colleagues especially Barb Lawson at Queen's University for their support.

I would like to thank my husband (Ali Moallem) for his unconditional love and support. My parents (Touba and Fathollah) and my sister (Hakimeh) deserve a great deal of thanks for their absolute love and support.

# Table of Contents

# List of Figures

# List of Tables

**Nomenclature for Chapter 2**

| | |
|---|---|
| $a$ | empirical model parameter for Equation (d.5), $mmol^{-0.5} \cdot kg^{0.5}$ |
| $[A]$ | concentration of amine end groups, $mmol \cdot kg^{-1}$ |
| $[\hat{A}]$ | calculated concentration of basic end groups using parameter estimates, |
| $[A]_{meas}$ | concentration of amine and other basic end groups determined by titration, |
| $[A2]$ | concentration of secondary amine groups, $mmol \cdot kg^{-1}$ |
| $[BHMT]$ | concentration of bis(hexamethylene) triamine (BHMT), $mmol \cdot kg^{-1}$ |
| $[\hat{BHMT}]$ | calculated concentration of bis(hexamethylene) triamine (BHMT) using |
| $[C]$ | concentration of carboxylic acid end groups, $mmol \cdot kg^{-1}$ |
| $[\hat{C}]$ | calculated concentration of carboxylic acid end groups using parameter |
| $[CPK]$ | concentration of cyclopentanone, $mmol \cdot kg^{-1}$ |
| $E_c$ | polycondensation activation energy , $kJ \cdot mol^{-1}$ |
| $E_i$ | activation energy of reactions (b.2) to (b.6) and (b.8) to (b.10), (i=2, 3, 4, 5, |
| $\Delta H$ | enthalpy of polycondensation reactions (a.1) and (a.8), $kJ \cdot mol^{-1}$ |
| $\Delta H_2$ | enthalpy of reaction (a.2), $kJ \cdot mol^{-1}$ |
| $\Delta H_6$ | enthalpy of reaction (a.6), $kJ \cdot mol^{-1}$ |
| $j$ | counter used in objective function Equation (4) |
| $J$ | objective function value, dimensionless |
| $k_c$ | polycondensation rate constant, $kg \cdot mmol^{-1} \cdot h^{-1}$ |
| $k_{cr}$ | polycondensation rate constant at reference temperature ($T_r$=549.15 K), |
| $K_a$ | apparent polycondensation equilibrium constant at $T_r$=549.15 K |
| $K_{ar}$ | apparent polycondensation equilibrium constant at reference temperature |
| $k_i$ | rate constant for reactions (b.2), (b.3) and (b.4) (i=2, 3 and 4), $h^{-1}$ and |
| $k_{io}$ | rate constant of reactions (b.2), (b.3), (b.4) (i=2, 3 and 4), $h^{-1}$ and reactions |
| $(k_La)_{CPK}$ | volumetric liquid-phase mass-transfer coefficient for cyclopentanone, $h^{-1}$ |
| $(k_La)_w$ | volumetric liquid-phase mass-transfer coefficient for water, $h^{-1}$ |
| $K_{2eq}$ | equilibrium constant of reaction (a.2) |
| $K_{6eq}$ | equilibrium constant of reaction (a.6) |
| $k_{2R}$ | rate constant of the inverse of reaction (a.2), $h^{-1}$ |
| $k_{6R}$ | rate constant of the inverse of reaction (a.6), $kg \cdot mmol^{-1} \cdot h^{-1}$ |
| $[L]$ | concentration of amide links, $mmol \cdot kg^{-1}$ |
| $[L2]$ | concentration of tertiary amide branch point, $mmol \cdot kg^{-1}$ |
| $MW_i$ | molecular weight of species i, $mol \cdot g^{-1}$ |
| $N_i$ | evolution rates for gaseous species (i=$NH_3$ and i=$CO_2$), $mmol \cdot kg^{-1} \cdot h$ |

| | |
|---|---|
| $\hat{N}_i$ | calculated evolution rates for gaseous species (i=$NH_3$ and i=$CO_2$) using |
| $n_i$ | number of measured values of species i |
| $P_c$ | 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 |
| [*P244*] | concentration of pyridine 244 ends, mmol·kg$^{-1}$ |
| *P244int* | intermediate product that results in P244s in hydrostate |
| $R$ | ideal gas constant, 8.3145×10$^{-3}$ kJ·mol$^{-1}$· K$^{-1}$ |
| $R_i$ | rate of reaction i, mmol·kg$^{-1}$·h |
| $R_{m,i}$ | rate of mass-transfer of species i (i=w (water) and i=CPK (cyclopentanone)) |
| [*SB1*] | concentration of Schiff base groups, mmol·kg$^{-1}$ |
| [*SB2*] | concentration of Schiff base links, mmol·kg$^{-1}$ |
| [*SB3*] | concentration of Schiff base branch points, mmol·kg$^{-1}$ |
| [*SE*] | concentration of stabilized end groups, mmol·kg$^{-1}$ |
| $t$ | time, h |
| $T$ | temperature, K |
| $T_c$ | critical temperature of water, 647.3 K |
| $T_o$ | reference temperature used in the Arrhenius expressions (Equation (d.6) and |
| $T_r$ | reference temperature for polycondensation rate constant and equilibrium |
| [*W*] | concentration of water in molten polymer, mmol·kg$^{-1}$ |

**Greek Letters**

| | |
|---|---|
| $\gamma_w$ | activity coefficient for water in the nylon melt |
| $\gamma_{wr}$ | activity coefficient for water in the nylon melt at $T_r$=549.15 K |
| $\sigma_i$ | standard deviation of measurements for species i, mmol·kg$^{-1}$ |

**Subscripts**

| | |
|---|---|
| eq | equilibrium |
| w | water |
| 0 | initial value |
| 0run,i | initial value at run number i (i=1 to 6) |
| $CO_2$ | carbon dioxide |
| $NH_3$ | ammonia |
| CPK | cyclopentanone |

## Nomenclature for Chapter 3

*Abbreviations*

| | |
|---|---|
| AEM | approximate expectation maximization |
| AMLE | approximate maximum likelihood estimation |
| CSTR | continuous stirred tank reactor |
| CTSM | continuous time stochastic modeling |
| EKF | extended Kalman filter |
| EM | expectation maximization |
| IQR | interquartile range |
| MCMC | Markov chain Monte Carlo |
| MIMO | multi-input multi-output |
| MLE | maximum likelihood estimation |
| SDE | stochastic differential equation |

*Roman letters*

| | |
|---|---|
| $a$ | CSTR model parameter relating heat-transfer coefficient to coolant flow |
| $b$ | CSTR model exponent relating heat-transfer coefficient to coolant flow |
| $c_s$ | number of B-spline coefficients for $s$th state trajectory |
| $C_A$ | concentration of reactant A ($kmol \cdot m^{-3}$) |
| $C_{A0}$ | feed concentration of reactant A ($kmol \cdot m^{-3}$) |
| $c_p$ | heat capacity of reactor contents ($J \cdot kg^{-1} \cdot K^{-1}$) |
| $c_{pc}$ | coolant heat capacity ($J \cdot kg^{-1} \cdot K^{-1}$) |
| $C_1$ | constant in Equations (17) and (18) |
| cov{.} | covariance |
| det | determinant |
| E{.} | expected value |
| $E/R$ | activation energy divided by the ideal gas constant (K) |
| $\mathbf{f}$ | $X$-dimensional nonlinear mapping on the right-hand side of the SDE |
| $F$ | reactant volumetric flow rate ($m^3 \cdot min^{-1}$) |
| $F_c$ | coolant volumetric flow rate ($m^3 \cdot min^{-1}$) |
| $\mathbf{g}$ | $Y$-dimensional vector of nonlinear mappings on the right hand side of |
| $\Delta H_{rxn}$ | enthalpy of reaction ($J \cdot kg^{-1} \cdot K^{-1}$) |
| $I$ | identity matrix |
| $j_1$ and $j_2$ | positive integers in Equation (5) |
| $J_{AEM}$ | AEM objective function defined in Equations (17) and (18) |
| $J_{AEM,CSTR}$ | AEM objective function for CSTR model defined in Equation (25) |

| | |
|---|---|
| $J_{AMLE}$ | AMLE objective function defined in Equation (11) |
| $J_{I,AMLE}$ | AMLE objective function for CSTR model |
| $J_{O,AMLE}$ | Outer AMLE objective function for CSTR model |
| $k_{\text{ref}}$ | kinetic rate constant at temperature $T_{ref}$ ($\text{min}^{-1}$) |
| $k_r$ | rate constant defined in Equation (22) |
| $M$ | order of B-spline basis functions |
| $n$ | number of measurements |
| $n_C$ | number of measurements for concentration of reactant A |
| $N_r$ | number of measurements for $r$th response |
| $n_T$ | number of measurements for temperature |
| $P$ | number of unknown model parameters |
| $p(.)$ | probability density function |
| $q$ | number of discretization points for SDE model (Equation (1)) |
| $q_C$ | number of discretization points for concentration SDE (Equation (21a)) |
| $q_T$ | number of discretization points for temperature SDE (Equation (21b)) |
| $\mathbf{Q}$ | diagonal power spectral density function |
| $\mathbf{Q_d}$ | vector of disturbance intensities as $\mathbf{Q_d}=[Q_1,\ldots,Q_X]$ |
| $Q_C$ | process disturbance intensity for concentration ($\text{kmol}^2\cdot\text{m}^{-6}\cdot\text{min}^{-1}$) |
| $Q_T$ | process disturbance intensity for temperature ($\text{K}^2\cdot\text{min}^{-1}$) |
| $Q_s$ | process disturbance intensity for state $s$ |
| $R(\boldsymbol{\zeta},\boldsymbol{\zeta}_k)$ | expected value defined in Equation (13) |
| $S_T^2$ | measurement noise variance for $T(0)$ |
| $\mathbf{S_{m0}}$ | covariance matrix for measured initial states $\mathbf{x}_0$ |
| $t$ | time (min) |
| $t_0$ | initial time (min) |
| $t_i$ | times used for discretizing SDEs (min) |
| $t_{mj}$ | $j$th measurement time (min) |
| $t_{m\,r,j}$ | $j$th measurement time at for the $r$th response (min) |
| $t_q$ | final time (min) |
| $T$ | temperature of reactor contents (K) |
| $T_0$ | reactant feed temperature (K) |
| $T_{in}$ | inlet temperature of coolant (K) |
| $T_{ref}$ | reference temperature (K) |
| $\Delta t$ | sampling time used for discretizing SDEs and disturbances (min) |
| $\Delta t_m$ | sampling interval for the case where all states are measured frequently |
| $\mathbf{u}$ | $U$-dimensional vector of input variables for SDE model |
| $UA$ | heat transfer coefficient defined in Equation (23) |
| $\mathbf{U_m}$ | stacked vector of input values at the measurement times |

| | |
|---|---|
| $u_s$ | $s$th input variable |
| $V$ | volume of the reactor ($m^3$) |
| $\mathbf{x}$ | state vector |
| $X$ | dimension of state vector |
| $\mathbf{x}_0$ | state vector at the initial time $t_0$ |
| $\mathbf{x}_{m0}$ | vector of measured values of initial conditions |
| $\mathbf{X_m}$ | stacked vector of state values at measurement times |
| $x_s$ | $s$th state variable |
| $\mathbf{X}_q$ | stacked vector of state values at discrete times |
| $\mathbf{x}_{\sim}$ | B-spline approximation of the vector of state trajectories $\mathbf{x}$ |
| $\dot{\mathbf{x}}_{\sim}(t)$ | time derivative for vector $\mathbf{x}_{\sim}$ |
| $y_C$ | vector of concentration measurements |
| $y_T$ | vector of temperature measurements |
| $\mathbf{y}$ | $Y$-dimensional output vector |
| $Y$ | dimension of output vector |
| $\mathbf{Y_m}$ | vector of stacked measured values at the measurement times |
| $y_r$ | $r$th measured output |
| $z_{\alpha/2}$ | $(1-\alpha/2)$th quantile value of the standard Gaussian distribution |

*Greek letters*

| | |
|---|---|
| $\alpha$ | significance level for confidence intervals |
| $\mathbf{B}$ | stacked vector of B-spline coefficients |
| $\beta_{s,l}$ | $l$th B-spline coefficient for $s$th state trajectory |
| $\beta_{c,1}$ | first B-spline coefficient for concentration state trajectory |
| $\boldsymbol{\beta}_s$ | vector of spline coefficients corresponding to the $s$th state trajectory |
| $\gamma$ | constant defined in Equation (24) |
| $\delta(.)$ | Dirac delta function |
| $\boldsymbol{\varepsilon}$ | $Y$-dimensional vector of zero-mean random variables |
| $\varepsilon_C$ | measurement noise for concentration($kmol \cdot m^{-3}$) |
| $\varepsilon_T$ | measurement noise for temperature (K) |
| $\varepsilon_r$ | normally distributed measurement noise for $r$th measured state |
| $\boldsymbol{\varepsilon_m}$ | stacked vector of measurement noise values at measurement times |
| $\boldsymbol{\zeta}$ | vector of unknown parameters defined as $\boldsymbol{\zeta} = [\boldsymbol{\theta}^T, \mathbf{x_0}^T, \mathbf{Q}^T]^T$ |
| $\boldsymbol{\eta(t)}$ | $X$-dimensional continuous zero-mean stationary white-noise process |
| $\boldsymbol{\eta_d(t)}$ | $X$-dimensional discrete zero-mean stationary white-noise process |
| $\eta_C(t)$ | continuous zero-mean stationary white-noise process for concentration |
| $\eta_T(t)$ | continuous zero-mean stationary white-noise process for temperature |
| $\boldsymbol{\theta}$ | vector of model parameters |

| | |
|---|---|
| $v$ | number of samples from the joint probability distribution |
| $\rho$ | density of reactor contents ($kg \cdot m^{-3}$) |
| $\rho_c$ | coolant density ($kg \cdot m^{-3}$) |
| $\mathbf{\Sigma}$ | covariance matrix for measurement errors defined in Equation (3) |
| $\mathbf{S}_{m0}$ | covariance matrix for state variables at the initial time |
| $\sigma_r^2$ | measurement noise variance for *r*th response |
| $\sigma_C^2$ | measurement noise variance for concentration |
| $\sigma_T^2$ | measurement noise variance for temperature |
| $\boldsymbol{\tau}$ | stacked vector of model and disturbance intensity parameters and the B- |
| $\boldsymbol{\varphi}_s(t)$ | vector of B-spline basis function for *s*th state trajectory |
| $\varphi_{s,l}(t)$ | *l*th B-spline basis function for *s*th state trajectory |
| $\mathbf{\Phi}(t)$ | matrix of B-spline functions defined in Equation (8) |

*Subscripts*

| | |
|---|---|
| *i* | index of times used for discretizing SDE |
| *j* | index of number of measurements |
| *k* | index of iterations used in the EM algorithm for estimating $\boldsymbol{\zeta}$ |
| *l* | index for B-spline coefficients |
| *r* | index for response variable |
| *s* | index for state variables |
| ~ | subscript used to indicate smoothed state trajectories estimated using B- |

*Superscripts*

| | |
|---|---|
| (*l*) | index of *l*th random value of state vector sampled from $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ |
| T | transpose |

## Nomenclature for Chapter 4

*Abbreviations*

| | |
|---|---|
| AEM | approximate expectation maximization |
| AE | algebraic equation |
| AMLE | approximate maximum likelihood estimation |
| AMPL | a modeling language for mathematical programming |
| CSTR | continuous stirred tank reactor |
| CTSM | continuous time stochastic modeling |
| EKF | extended Kalman filter |
| EM | expectation maximization |
| FLA | fully Laplace approximation |
| FLAEM | fully Laplace approximation expectation maximization |
| IQR | interquartile range |
| MCMC | Markov chain Monte Carlo |
| MIMO | multi-input multi-output |
| ML | maximum likelihood |
| MLE | maximum likelihood estimation |
| ODE | ordinary differential equation |
| SDE | stochastic differential equation |

*Roman letters*

| | |
|---|---|
| $a$ | CSTR model parameter relating heat-transfer coefficient to coolant flow rate |
| $b$ | CSTR model exponent relating heat-transfer coefficient to coolant flow rate |
| $c_s$ | number of B-spline coefficients for sth state trajectory |
| $C_A$ | concentration of reactant A ($kmol \cdot m^{-3}$) |
| $C_{A0}$ | feed concentration of reactant A ($kmol \cdot m^{-3}$) |
| $\hat{C}_{A\sim}$ | estimated state trajectory corresponding to estimated B-splines coefficients $\hat{\boldsymbol{\beta}}_C$ |
| $\hat{C}^S_{A\sim}$ | estimated state trajectory corresponding to estimated B-splines coefficients $\hat{\boldsymbol{\beta}}^S_C$ |
| $\hat{C}^Z_{A\sim}$ | estimated state trajectory corresponding to estimated B-splines coefficients $\hat{\boldsymbol{\beta}}^Z_C$ |
| $c_p$ | heat capacity of reactor contents ($J \cdot kg^{-1} \cdot K^{-1}$) |
| $c_{pc}$ | coolant heat capacity ($J \cdot kg^{-1} \cdot K^{-1}$) |
| $C_1$ | constant in Equation (4.A.1) |
| cov{.} | covariance |

| | |
|---|---|
| det | determinant |
| $e$ | sum of the squared relative errors defined in Equation (4.57) |
| E{.} | expected value |
| $E/R$ | activation energy divided by the ideal gas constant (K) |
| $f_r$ | nonlinear function on the right-hand side of the SDE model for the $r$th state |
| $\mathbf{f}$ | $X$-dimensional nonlinear mapping on the right-hand side of the SDE model |
| $F$ | reactant volumetric flow rate ($m^3 \cdot min^{-1}$) |
| $F_c$ | coolant volumetric flow rate ($m^3 \cdot min^{-1}$) |
| $g_r$ | nonlinear functions on the right hand side of the $r$th output equation |
| $\mathbf{g}$ | $Y$-dimensional vector of nonlinear mappings on the right hand side of Equation |
| $G$ | positive scalar function |
| $\mathbf{G}$ | matrix defined in Equation (4.A.26) |
| $\mathbf{H}_{\mathbf{B}}$ | Hessian matrix defined in Equation (4.27) |
| $\mathbf{H}_{\mathbf{B}}^{S}$ | Hessian matrix defined in Equation (4.29) |
| $\mathbf{H}_{\mathbf{B}}^{Z}$ | Hessian matrix defined in Equation (4.28) |
| $\mathbf{H}_{\boldsymbol{\beta}C}$ | Hessian matrix defined in Equation (4.48) |
| $\mathbf{H}_{\boldsymbol{\beta}C}^{S}$ | Hessian matrix defined in Equation (4.52) |
| $\mathbf{H}_{\boldsymbol{\beta}C}^{Z}$ | Hessian matrix defined in Equation (4.50) |
| $\mathbf{H}_{\boldsymbol{\beta}T}$ | Hessian matrix defined in Equation (4.49) |
| $\mathbf{H}_{\boldsymbol{\beta}T}^{S}$ | Hessian matrix defined in Equation (4.53) |
| $\mathbf{H}_{\boldsymbol{\beta}T}^{Z}$ | Hessian matrix defined in Equation (4.51) |
| $\mathbf{H}_{\boldsymbol{\beta}_r}$ | Hessian matrix defined in Equation (4.38) |
| $\mathbf{H}_{\boldsymbol{\beta}_r}^{S}$ | Hessian matrix defined in Equation (4.40) |
| $\mathbf{H}_{\boldsymbol{\beta}_r}^{Z}$ | Hessian matrix defined in Equation (4.39) |
| $\mathbf{H}_{\mathbf{B},\mathrm{CSTR}}$ | Hessian matrix defined in Equation (4.54) |
| $\mathbf{H}_{\mathbf{B},\mathrm{CSTR}}^{S}$ | Hessian matrix defined in Equation (4.56) |
| $\mathbf{H}_{\mathbf{B},\mathrm{CSTR}}^{Z}$ | Hessian matrix defined in Equation (4.55) |
| $\mathbf{H}_{\hat{\mathbf{x}}}$ | Hessian matrix defined in Equation (4.A.12) |
| $\mathbf{H}_{\hat{\mathbf{x}}}^{S}$ | Hessian matrix defined in Equation (4.A.14) |
| $\mathbf{H}_{\hat{\mathbf{x}}}^{Z}$ | Hessian matrix defined in Equation (4.A.13) |
| $\mathbf{H}_{\hat{\mathbf{x}}_{\sim}}$ | Hessian matrix defined in Equation (4.A.20) |
| $\mathbf{H}_{\hat{\mathbf{x}}_{\sim}}^{S}$ | Hessian matrix defined in Equation (4.A.22) |
| $\mathbf{H}_{\hat{\mathbf{x}}_{\sim}}^{Z}$ | Hessian matrix defined in Equation (4.A.21) |

| | |
|---|---|
| $\Delta H_{rxn}$ | enthalpy of reaction ($J \cdot kg^{-1} \cdot K^{-1}$) |
| $j_1$ and $j_2$ | positive integers in Equation (4.5) |
| $J$ | objective function defined in Equation (4.A.15) |
| $J_{\text{AMLE}}$ | objective function defined in Equation (4.14) |
| $J_{\text{AEM,CSTR}}$ | AEM objective function for CSTR model defined in Equation (4.59) |
| $J_{\text{AMLE,CSTR}}$ | AMLE objective function for CSTR model defined in Equation (4.38) |
| $J^S$ | objective function defined in Equation (4.A.17) |
| $J^Z$ | objective function defined in Equation (4.A.16) |
| $J^S_{\sim}$ | objective function defined in Equation (4.31) |
| $J^S_{\text{CSTR}}$ | objective function defined in Equation (4.39) |
| $J^Z_{\sim}$ | objective function defined in Equation (4.30) |
| $J^Z_{\text{CSTR}}$ | objective function defined in Equation (4.40) |
| $k_{\text{ref}}$ | kinetic rate constant at temperature $T_{ref}$ ($min^{-1}$) |
| $k_r$ | rate constant defined in Equation (4.33) |
| $L$ | likelihood function |
| $M$ | order of B-spline basis functions |
| $n$ | number of measurements |
| $n_C$ | number of measurements for concentration of reactant A |
| $N_r$ | number of measurements for $r$th response |
| $n_T$ | number of measurements for temperature |
| $P$ | number of unknown model parameters |
| $p(.)$ | probability density function |
| $q$ | number of discretization points for SDE model (Equation (4.13) |
| $\mathbf{Q}$ | diagonal power spectral density matrix |
| $\mathbf{Q_d}$ | vector of disturbance intensities as $\mathbf{Q_d}=[Q_1,\ldots,Q_X]^T$ |
| $Q_C$ | process disturbance intensity for concentration ($kmol^2 \cdot m^{-6} \cdot min^{-1}$) |
| $Q_T$ | process disturbance intensity for temperature ($K^2 \cdot min^{-1}$) |
| $Q_s$ | process disturbance intensity for state $s$ |
| $R$ | expected value defined in Equation (4.15) |
| $S$ | sum of squared error terms defined in Equation (4.20) |
| $S_T$ | measurement noise variance for $T(0)$ |
| $\mathbf{S_{m0}}$ | covariance matrix for measured initial states $\mathbf{x}_0$ |
| $t$ | time (min) |
| $t_0$ | initial time (min) |

| | |
|---|---|
| $t_i$ | times used for discretizing SDEs (min) |
| $t_{m\,C,j}$ | $j$th measurement time for concentration (min) |
| $t_{m\,T,j}$ | $j$th measurement time for temperature (min) |
| $t_{m\,r,j}$ | $j$th measurement time for the $r$th response (min) |
| $t_{nC}$ | final time for concentration SDE (min) |
| $t_{nT}$ | final time for temperature SDE (min) |
| $t_q$ | final time (min) |
| $T$ | temperature of reactor contents (K) |
| $T_0$ | reactant feed temperature (K) |
| $\hat{T}_{\sim}$ | estimated state trajectory corresponding to estimated B-splines coefficients $\hat{\boldsymbol{\beta}}_T$ |
| $\hat{T}_{\sim}^{S}$ | estimated state trajectory corresponding to estimated B-splines coefficients $\hat{\boldsymbol{\beta}}_T^{S}$ |
| $\hat{T}_{\sim}^{Z}$ | estimated state trajectory corresponding to estimated B-splines coefficients $\hat{\boldsymbol{\beta}}_T^{Z}$ |
| $T_{in}$ | inlet temperature of coolant (K) |
| $T_{ref}$ | reference temperature (K) |
| $\Delta t$ | sampling interval used for discretizing SDEs and disturbances (min) |
| $\mathbf{u}$ | $U$-dimensional vector of input variables for the SDE model |
| $UA$ | heat transfer coefficient defined in Equation (4.34) |
| $U$ | dimension of the input vector |
| $\mathbf{U_m}$ | stacked vector of input values at measurement times with different sampling interval |
| $V$ | volume of the reactor ($m^3$) |
| $\mathbf{x}$ | state vector |
| $X$ | dimension of state vector |
| $\mathbf{x}_0$ | state vector at the initial time $t_0$ |
| $\mathbf{x}_{m0}$ | vector of measured values of initial conditions |
| $\mathbf{X_m}$ | stacked vector of state values at measurement times |
| $x_s$ | $s$th state variable |
| $\mathbf{X}_q$ | stacked vector of state values at discrete times |
| $\hat{\mathbf{X}}_{\mathbf{m}}$ | stacked vector of state values at measurement times evaluated at $\hat{\mathbf{x}}$ |
| $\hat{\mathbf{X}}_{\mathbf{m}}^{S}$ | stacked vector of state values at measurement times evaluated at $\hat{\mathbf{x}}^{S}$ |
| $\hat{\mathbf{X}}_{\mathbf{m}}^{Z}$ | stacked vector of state values at measurement times evaluated at $\hat{\mathbf{x}}_{\sim}^{Z}$ |
| $\mathbf{x}_{\sim}$ | B-spline approximation of the vector of state trajectories $\mathbf{x}$ |
| $\dot{\mathbf{x}}_{\sim}(t)$ | time derivative for vector $\mathbf{x}_{\sim}$ |

| | |
|---|---|
| $\hat{\mathbf{x}}_{\sim}$ | estimated state trajectory corresponding to $\hat{\mathbf{B}}$ |
| $\hat{\mathbf{x}}_{\sim}^{S}$ | estimated state trajectory corresponding to $\hat{\mathbf{B}}^{S}$ |
| $\hat{x}_{\sim s}^{S}$ | estimated state trajectory for the *s*th state corresponding to $\hat{\boldsymbol{\beta}}_{s}^{S}$ |
| $\hat{\mathbf{x}}_{\sim}^{Z}$ | estimated state trajectory corresponding to $\hat{\mathbf{B}}^{Z}$ |
| $\hat{x}_{\sim s}^{Z}$ | estimated state trajectory for the *s*th state corresponding to $\hat{\boldsymbol{\beta}}_{s}^{Z}$ |
| $y_C$ | concentration measurements |
| $y_r$ | *r*th measured output |
| $y_T$ | temperature measurements |
| $\mathbf{y}$ | *Y*-dimensional output vector |
| $Y$ | dimension of output vector |
| $\mathbf{Y_m}$ | vector of stacked measured values at the measurement times |
| $Z$ | penalty term defined in Equation (4.19) |
| $z_{\alpha/2}$ | $(1-\alpha/2)$th quantile value of the standard Gaussian distribution |

*Greek letters*

| | |
|---|---|
| $\alpha$ | significance level for confidence intervals |
| $\mathbf{B}$ | stacked vector of B-spline coefficients |
| $\hat{\mathbf{B}}$ | vector of spline coefficients that minimize $J_{AMLE}$ |
| $\mathbf{B}_{\text{CSTR}}$ | stacked vector of B-spline coefficients for the CSTR model |
| $\hat{\mathbf{B}}^{S}$ | vector of spline coefficients that minimize $J_{\sim}^{S}$ |
| $\hat{\mathbf{B}}_{CSTR}^{S}$ | vector of spline coefficients that minimize $J_{\text{CSTR}}^{S}$ |
| $\hat{\mathbf{B}}^{Z}$ | vector of spline coefficients that minimize $J_{\sim}^{Z}$ |
| $\hat{\mathbf{B}}_{CSTR}^{Z}$ | vector of spline coefficients that minimize $J_{\text{CSTR}}^{Z}$ |
| $\beta_{s,l}$ | *l*th B-spline coefficient for *s*th state trajectory |
| $\beta_{C,1}$ | first B-spline coefficient for concentration state trajectory |
| $\beta_{T,1}$ | first B-spline coefficient for temperature state trajectory |
| $\boldsymbol{\beta}_s$ | vector containing $c_s$ B-spline coefficients for the *s*th state trajectory |
| $\gamma$ | constant defined in Equation (4.35) |
| $\delta$ (.) | Dirac delta function |
| $\boldsymbol{\varepsilon}$ | *Y*-dimensional vector of zero-mean random variables |
| $\varepsilon_C$ | measurement noise for concentration ($\text{kmol}\cdot\text{m}^{-3}$) |
| $\varepsilon_T$ | measurement noise for temperature (K) |

| | |
|---|---|
| $\varepsilon_r$ | normally distributed measurement noise for *r*th measured state |
| $\boldsymbol{\varepsilon}_{\mathbf{m}}$ | stacked vector of measurement noise values at measurement times |
| $\boldsymbol{\zeta}$ | vector of unknown parameters defined as $\boldsymbol{\zeta} = [\boldsymbol{\theta}^{\mathrm{T}}, \mathbf{x}_{\mathbf{0}}^{\mathrm{T}}, \mathbf{Q}_{\mathbf{d}}^{\mathrm{T}}, \boldsymbol{\Sigma}_{\mathbf{d}}^{\mathrm{T}}]^{\mathrm{T}}$ |
| $\boldsymbol{\zeta}_{\mathrm{CSTR}}$ | complete vector of parameters in the CSTR case study |
| $\boldsymbol{\eta}(t)$ | *X*-dimensional continuous zero-mean stationary Gaussian white-noise process |
| $\boldsymbol{\eta}_{\mathbf{d}}(t)$ | *X*-dimensional discrete zero-mean stationary white-noise process |
| $\eta_C(t)$ | continuous zero-mean stationary Gaussian white-noise process for concentration |
| $\eta_T(t)$ | continuous zero-mean stationary Gaussian white-noise process for temperature |
| $\boldsymbol{\theta}$ | vector of model parameters |
| $\boldsymbol{\theta}_{\mathrm{CSTR}}$ | vector of model parameters for the CSTR model . |
| $\rho$ | density of reactor contents (kg· $m^{-3}$) |
| $\rho_c$ | coolant density (kg· $m^{-3}$) |
| $\boldsymbol{\Sigma}$ | covariance matrix for measurement errors defined in Equation (4.3) |
| $\rho_{\mathbf{d}}$ | diagonal elements of the covariance matrix defined as $\rho_{\mathbf{d}}=[\sigma_1^2,\ldots,\sigma_Y^2]^{\mathrm{T}}$ |
| $\sigma_r^2$ | measurement noise variance for *r*th response |
| $\sigma_C^2$ | measurement noise variance for concentration |
| $\sigma_T^2$ | measurement noise variance for temperature |
| $\boldsymbol{\varphi}_s(t)$ | vector of B-spline basis function for *s*th state trajectory |
| $\varphi_{s,l}(t)$ | *l*th B-spline basis function for *s*th state trajectory |
| $\boldsymbol{\Phi}(t)$ | matrix of B-spline functions defined in Equation (4.8) |
| $\boldsymbol{\chi}$ | vector argument in FLA integrals in Equation (4.17) |
| $\hat{\boldsymbol{\chi}}$ | vector that maximize $\psi$ |
| $\hat{\boldsymbol{\chi}}^*$ | vector that maximize $\psi^*$ |
| $\boldsymbol{\Psi}$ | matrix of B-spline basis functions defined in Equation (A.33) |
| $\cdot$ | scalar function |
| $\cdot^*$ | scalar function defined in Equation (4.18) |

*Subscripts*

| | |
|---|---|
| *i* | index of times used for discretizing SDE |
| *j* | index for the sampling times |
| *k* | index of iterations used in the EM algorithm for estimating $\boldsymbol{\zeta}$ |
| *l* | index for B-spline coefficients |
| *m* | measurement |
| *r* | index for response variable |

| | |
|---|---|
| $s$ | index for state variables |
| ~ | subscript used to indicate smoothed state trajectories estimated using B-splines |

*Superscripts*

| | |
|---|---|
| T | transpose |

## Nomenclature for Chapter 5

*Abbreviations*

| | |
|---|---|
| AEM | approximate expectation maximization |
| AMLE | approximate maximum likelihood estimation |
| CSTR | continuous stirred tank reactor |
| CTSM | continuous time stochastic modeling |
| EKF | extended Kalman filter |
| IQR | interquartile range |
| LA | Laplace approximation |
| LAMLE | Laplace approximation likelihood method |
| MCMC | Markov chain Monte Carlo |
| MIMO | multi-input multi-output |
| ML | maximum likelihood |
| MLE | maximum likelihood estimation |
| SDE | stochastic differential equation |
| SML | simulated maximum likelihood |

*Roman letters*

| | |
|---|---|
| $a$ | CSTR model parameter relating heat-transfer coefficient to coolant flow rate |
| $b$ | CSTR model exponent relating heat-transfer coefficient to coolant flow rate |
| $c_s$ | number of B-spline coefficients for sth state trajectory |
| $C_A$ | concentration of reactant A ($kmol \cdot m^{-3}$) |
| $C_{A0}$ | feed concentration of reactant A ($kmol \cdot m^{-3}$) |
| $c_p$ | heat capacity of reactor contents ($J \cdot kg^{-1} \cdot K^{-1}$) |
| $c_{pc}$ | coolant heat capacity ($J \cdot kg^{-1} \cdot K^{-1}$) |
| $C_1$ | constant in Equation (5.15) |
| $C_2$ | constant in Equation (5.A.7) |
| $C_3$ | constant in Equation (5.B.6) |

| | |
|---|---|
| cov{.} | covariance |
| $\mathbf{D}$ | function of $\mathbf{f}$ and its derivatives shown in Equation (B.16) |
| det | determinant |
| dim | dimension of a vector |
| E{.} | expected value |
| $E/R$ | activation energy divided by the ideal gas constant (K) |
| $\mathbf{f}$ | $X$-dimensional nonlinear mapping on the right-hand side of the SDE model |
| $F$ | reactant volumetric flow rate ($m^3 \cdot min^{-1}$) |
| $F_c$ | coolant volumetric flow rate ($m^3 \cdot min^{-1}$) |
| $\mathbf{g}$ | $Y$-dimensional vector of nonlinear mappings on the right hand side of Equation |
| $\mathbf{G}$ | derivative of $J_1$ defined in Equation (5.B.29) |
| $g_r$ | nonlinear function on the right hand side of Equation (5.1.c) for $r$th |
| $\Delta H_{rxn}$ | enthalpy of reaction ($J \cdot kg^{-1} \cdot K^{-1}$) |
| $\mathbf{H}_{\mathbf{X}\sim}$ | Hessian matrix of the $-\ln p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ with respect to $\mathbf{X_q}$ evaluated at $X_{\mathbf{q}\sim}$ |
| $\mathbf{H_B}$ | Hessian matrix of the $-\ln p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ with respect to B-spline basis |
| $\mathbf{I}$ | identity matrix |
| $j_1$ and $j_2$ | positive integers in Equation (5.3) |
| $J_{AMLE,CSTR}$ | AEM objective function for CSTR model defined in Equation (5.25) |
| $J_{AMLE}$ | AMLE objective function defined in Equation (5.11) |
| $J_1$ | objective function defined in Equation (5.A.2) |
| $J_d$ | objective function defined in Equation (5.B.6) |
| $k_{ref}$ | kinetic rate constant at temperature $T_{ref}$ ($min^{-1}$) |
| $k_r$ | rate constant defined in Equation (5.22) |
| $M$ | order of B-spline basis functions |
| $n$ | number of measurements |
| $n_C$ | number of measurements for concentration of reactant A |
| $N_r$ | number of measurements for $r$th response |
| $n_T$ | number of measurements for temperature |
| $P$ | number of unknown model parameters |
| $p(.)$ | probability density function |
| $q$ | number of discretization points for SDE model (Equation (5.1)) |
| $\mathbf{Q}$ | diagonal power spectral density function |
| $\mathbf{Q_d}$ | vector of disturbance intensities as $\mathbf{Q_d}=[Q_1,\ldots,Q_X]$ |
| $Q_C$ | process disturbance intensity for concentration ($kmol^2 \cdot m^{-6} \cdot min^{-1}$) |

| | |
|---|---|
| $Q_T$ | process disturbance intensity for temperature ($K^2 \cdot min^{-1}$) |
| $Q_s$ | process disturbance intensity for state *s* |
| $R(\boldsymbol{\zeta}, \boldsymbol{\zeta}_k)$ | expected value defined in Equation (5.13) |
| $S_T^2$ | measurement noise variance for $T(0)$ |
| $\mathbf{S_{m0}}$ | covariance matrix for measured initial states $\mathbf{x}_0$ |
| $t$ | time (min) |
| $t_0$ | initial time (min) |
| $t_i$ | times used for discretizing SDEs (min) |
| $t_{mj}$ | *j*th measurement time (min) |
| $t_{m\,r,j}$ | *j*th measurement time at for the *r*th response (min) |
| $t_q$ | final time (min) |
| $T$ | temperature of reactor contents (K) |
| $T_0$ | reactant feed temperature (K) |
| $T_{cin}$ | inlet temperature of coolant (K) |
| *tr* | trace |
| $T_{ref}$ | reference temperature (K) |
| $\Delta t$ | sampling time used for discretizing SDEs and disturbances (min) |
| $\mathbf{u}$ | *U*-dimensional vector of input variables for SDE model |
| $U$ | dimension of the vector of the input variables |
| $UA$ | heat transfer coefficient defined in Equation (5.23) |
| $\mathbf{U_m}$ | stacked vector of input values at measurement times with different sampling |
| $u_s$ | *s*th input variable |
| $V$ | volume of the reactor ($m^3$) |
| $\mathbf{w}$ | function defined in Equation (5.B.7) |
| $\mathbf{x}$ | state vector |
| $X$ | dimension of state vector |
| $\mathbf{x}_0$ | state vector at the initial time $t_0$ |
| $\mathbf{x}_{m0}$ | vector of measured values of initial conditions |
| $\mathbf{X_m}$ | stacked vector of state values at measurement times |
| $x_s$ | *s*th state variable |
| $\mathbf{X}_q$ | stacked vector of state values at discrete times |
| $\hat{\mathbf{X}}_\mathbf{m}$ | mode of $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ at measurement times |
| $\hat{\mathbf{X}}_\mathbf{q}$ | mode of $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ evaluated as shown in Equation (5.A.5) |
| $\mathbf{x}_{\sim}$ | B-spline approximation of the vector of state trajectories $\mathbf{x}$ |
| $\dot{\mathbf{x}}_{\sim}(t)$ | time derivative for vector $\mathbf{x}_{\sim}$ |

| | |
|---|---|
| $\hat{\mathbf{x}}(t)$ | state vector at the mode of $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ |
| $y_C$ | concentration measurements |
| $y_T$ | temperature measurements |
| $\mathbf{y}$ | $Y$-dimensional output vector |
| $Y$ | dimension of output vector |
| $\mathbf{Y_m}$ | vector of stacked measured values at the measurement times with |
| $\mathbf{Y_n}$ | vector of stacked measured values at the measurement times with |
| $y_r$ | $r$th measured output |
| $z_{\alpha/2}$ | (1−$\alpha$/2)th quantile value of the standard Gaussian distribution |

*Greek letters*

| | |
|---|---|
| $\alpha$ | significance level for confidence intervals |
| $\mathbf{B}$ | stacked vector of B-spline coefficients |
| $\beta_{s,l}$ | $l$th B-spline coefficient for $s$th state trajectory |
| $\beta_{c,1}$ | first B-spline coefficient for concentration state trajectory |
| $\boldsymbol{\beta}_s$ | vector of spline coefficients corresponding to the $s$th state trajectory |
| $\gamma$ | constant defined in Equation (5.24) |
| $\delta$ (.) | Dirac delta function |
| $\boldsymbol{\varepsilon}$ | $Y$-dimensional vector of zero-mean random variables |
| $\varepsilon_C$ | measurement noise for concentration(kmol·m$^{-3}$) |
| $\varepsilon_T$ | measurement noise for temperature (K) |
| $\varepsilon_r$ | normally distributed measurement noise for $r$th measured state |
| $\boldsymbol{\varepsilon_m}$ | stacked vector of measurement noise values at measurement times |
| $\boldsymbol{\zeta}$ | vector of unknown parameters defined as $\boldsymbol{\zeta} = [\boldsymbol{\theta}^{\mathrm{T}}, \mathbf{x}_0^{\mathrm{T}}, \mathbf{Q_d}^{\mathrm{T}}, \boldsymbol{\Sigma}_{\mathbf{d}}^{\mathbf{T}}]^{\mathrm{T}}$ |
| $\boldsymbol{\eta}(t)$ | $X$-dimensional continuous zero-mean stationary white-noise process |
| $\boldsymbol{\eta_d}(t)$ | $X$-dimensional discrete zero-mean stationary white-noise process |
| $\eta_C(t)$ | continuous zero-mean stationary white-noise process for concentration SDE |
| $\eta_T(t)$ | continuous zero-mean stationary white-noise process for temperature SDE |
| $\boldsymbol{\theta}$ | vector of model parameters |
| $\rho$ | density of reactor contents (kg· m$^{-3}$) |
| $\rho_c$ | coolant density (kg· m$^{-3}$) |
| $\boldsymbol{\Sigma}$ | covariance matrix for measurement errors defined in Equation (5.3) |
| $\mathbf{p_d}$ | diagonal elements of the covariance matrix of measurements |
| $\mathbf{S}_{m0}$ | covariance matrix for state variables at the initial time |
| $\sigma_r^2$ | measurement noise variance for $r$th response |

| | |
|---|---|
| $\sigma_C^2$ | measurement noise variance for concentration |
| $\sigma_T^2$ | measurement noise variance for temperature |
| $\boldsymbol{\varphi}_s(t)$ | vector of B-spline basis function for *s*th state trajectory |
| $\varphi_{s,l}(t)$ | *l*th B-spline basis function for *s*th state trajectory |
| $\boldsymbol{\Phi}(t)$ | matrix of B-spline functions defined in Equation (5.8) |
| V | matrix of spline functions defined as $\boldsymbol{\Psi} = \left[\boldsymbol{\Phi}(t_1) \quad \boldsymbol{\Phi}(t_2) \quad \ldots \quad \boldsymbol{\Phi}(t_q)\right]^{\mathrm{T}}$ |

*Subscripts*

| | |
|---|---|
| *i* | index of times used for discretizing SDE |
| *j* | index of number of measurements |
| *k* | index of iterations used in the EM algorithm for estimating $\boldsymbol{\zeta}$ |
| *l* | index for B-spline coefficients |
| *r* | index for response variable |
| *s* | index for state variables |
| ~ | subscript used to indicate smoothed state trajectories estimated using B-splines |

*Superscripts*

| | |
|---|---|
| T | transpose |

# Chapter 1

# Introduction

## 1.1 Introduction

The goal of this thesis is to develop and test accurate and easy-to-use techniques for estimating parameters in models of chemical engineering systems. Statistical methods are developed to overcome two types of problems that commonly occur during parameter estimation in chemical engineering models. The first problem occurs when complicated mechanistic models are developed and there are too many parameters to estimate from limited available data.[1-3] In this situation, we assume that the model structure is sufficiently complicated so that it could describe the underlying process behaviour accurately, if only accurate parameter values were available. The second problem involves parameter estimation in simplified dynamic models of chemical processes where the model structure may be imperfect and where unmeasured disturbances can influence the process behavior.[4,5] In this situation, there is often a considerable amount of data available for parameter estimation, but model imperfections and stochastic disturbances should be taken into account.[5]

The model equations associated with the first type of problem are often complicated ordinary differential equations (ODEs) that are nonlinear in the model parameters.[1,6] As a result, the objectives for the first part of the thesis are to test the effectiveness of recently developed parameter ranking and selection methods using a detailed mathematical model for nylon 66 production containing 14 ODEs and 56 parameters.

In the second type of problem, the model equations tend to be simpler, but stochastic terms are introduced on the right-hand side of the differential equations to account for disturbances and

model mismatch.[4,5] The resulting equations are called stochastic differential equations (SDEs). Accounting for model mismatch and disturbances during parameter estimation can lead to improved parameter estimates and model predictions. Knowledge about the magnitude of the mismatch and disturbances, may be helpful when implementing dynamic process models on-line, using extended Kalman filters or other state estimators.[4] Therefore, the objective for the second part of this thesis is to propose practical methods for addressing difficulties that arise during parameter estimation of nonlinear SDE models. Some of the benefits of the proposed methods that may be attractive to developers of fundamental dynamic models are: i) simplicity of implementation, ii) reliable estimates of model parameters, initial conditions and disturbance intensities, iii) efficient handling of unknown initial states, iv) ability to handle unmeasured state variables. In particular, estimates of disturbance intensities can provide modelers with information about the degree of mismatch and the magnitude of unmeasured disturbances in their models. This information will be helpful when implementing on-line state and parameter estimation schemes for process monitoring and control. Three proposed techniques are developed to address computational and implementation issues that are common using existing methods for parameter estimation in SDEs. The proposed approaches are illustrated and tested using a continuous stirred-tank reactor (CSTR) model with two stochastic differential equations and four model parameters.[5,7]

### 1.1.1 Problem Definition

Effective mathematical modeling for chemical engineering processes involves building a system of equations that is sufficiently complex to be able to simulate physical reality, yet simple enough to give real insight into the process.[1,8] In chemical engineering, fundamental models are derived using material, energy and momentum balances, along with chemical and physical principals including thermodynamics, kinetics and transport phenomena.

Fundamental models require and lead to a better understanding of the process compared to empirical models. Parameters in these models have physical meaning, which lead to a better assessment and interpretation of system behaviour and can help to further process analysis.[1] Fundamental dynamic models of chemical processes may involve the use of algebraic equations, ODEs or partial differential equations (PDEs). These models contain a number of parameters whose values are not known a priori. The determination of suitable values of model parameters is the objective of parameter estimation.[9] Fundamental models have a wide variety of application. They can be used for simulation, design and optimization of chemical processes, training of operators, model predictive control, increasing the quality of products, improving process safety, process scheduling and production planning, reducing costs and obtaining better understanding of process behaviour.[1,6] Therefore, accurate models and reliable parameter estimation techniques are vital for chemical engineers.

All steps for constructing fundamental models of chemical engineering systems are challenging, but perhaps the most difficult task is estimation of model parameters. Parameter estimation difficulties grow with the size and complexity of the system.[10] Chemical engineering models may have many reactions, many kinetic parameters, and many mass-transfer and thermodynamic constants. Consequently, models that can fully describe a chemical process are usually nonlinear and complex and have many unknown parameters.[1] Another challenge in modeling chemical processes is that experiments and measurements are often limited due to cost or inherent inability to measure certain variables. Performing further experiments is expensive and may not be feasible.[1] As a result, the number of data values for parameter estimation may be limited and some of the states are not measured. Some of the parameters may have little influence on the model predictions making them impossible to estimate while some of the parameters may have correlated effects with other parameters.[11,12] Additionally, it is sometimes not clear whether all of the model parameters can be estimated reliably from the available data, due to problems with parameter estimability and identifiability.[1,12] Thus, it is advisable to follow appropriate statistical

procedures to overcome these problems. Chou and Voit[10] reviewed parameter estimation techniques and their challenges and suggested methods for overcoming data-related issues, model-related issues and mathematical issues occur during parameter estimation in complicated models.

In chemical processes, external fluctuations (or disturbances) can influence process behaviour. Theses external disturbances reflect the random character of the environment and of system inputs. For example, variations in external temperature or in impurity levels in feed streams to a continuous stirred tank reactor are examples of common disturbances. Additionally, many chemical engineering processes are not fully understood and are too complex to be modeled in a completely deterministic fashion. Imperfections in the structure of a mathematical model are sometimes treated using stochastic disturbance terms.[13] Often, a full fundamental model would be too complex for the intended use and it would be too expensive to do all of the experiments required to estimate all of the parameters that would appear in the corresponding model equations. Sometimes model users would like their models to include only the most important phenomena, since simple models sometimes give better predictions than complicated models[2,12] and simple models are more portable and easy to use. Thus, many systems of practical interest to chemical engineer are better modeled using simplified fundamental models, especially models that account for stochastic disturbances.

SDEs are differential equations in which the influence of various random disturbances appears explicitly.[8] To account for the effect of unmeasured disturbances, imperfections and model mismatch, stochastic terms in ODEs are introduced. These disturbances can enter the model equations nonlinearly, but additive linear disturbances are most widely used in models for chemical processes.[14]

Consider the Multi-Input Multi-Output (MIMO) SDE model of the following form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\eta}(t) \qquad (1.1.a)$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{1.1.b}$$

$$\mathbf{y}(t_{mj}) = \mathbf{g}(\mathbf{x}(t_{mj}), \mathbf{u}(t_{mj}), \boldsymbol{\theta}) + \boldsymbol{\varepsilon}(t_{mj}) \tag{1.1.c}$$

where $\mathbf{x}=[x_1,\ldots,x_X]$ is an $X$-dimensional state vector, $\mathbf{f}=[f_1,\ldots,f_X]$ is an $X$-dimensional vector of nonlinear mappings, $\mathbf{u}=[u_1,\ldots,u_U]$ is a $U$-dimensional vector of input variables, $\boldsymbol{\theta}=[\theta_1,\ldots,\theta_P]$ is a $P$-dimensional vector of unknown parameters, $\mathbf{y}=[y_1,\ldots,y_Y]$ is a $Y$-dimensional output vector with $Y \leq X$ . Y<X is the case that some of the sates are not measured. Assume that each response variable ($r$=1…$Y$) is measured $n$ times during the set of dynamic experiments. The set of times at which measurements are available for the $r$th response ($r$=1…$Y$) is denoted by $t_{mj}$ ($j$ = 1…$n$). $\mathbf{g}=[g_1,\ldots,g_Y]$ is a $Y$-dimensional vector of nonlinear mappings and $\boldsymbol{\varepsilon}=[\varepsilon_1,\ldots,\varepsilon_Y]$ is a $Y$-dimensional vector of zero-mean random variables. Assume that these measurement errors are independent so that their covariance matrix is:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \pm_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \pm_Y^2 \end{bmatrix} \tag{1.2}$$

and $\boldsymbol{\eta}(t)$ is an $X$-dimensional continuous zero-mean stationary Gaussian white-noise process with covariance matrix $E\{\boldsymbol{\eta}(t_1)\boldsymbol{\eta}(t_2)\}=\mathbf{Q}\,\delta(t_2-t_1)$, where $\mathbf{Q}$ is the corresponding diagonal power spectral density function with dimension $X \times X$ :

$$\mathbf{Q} = \begin{bmatrix} Q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_X \end{bmatrix} \tag{1.3}$$

The diagonal elements of $\mathbf{Q}$ are sometimes referred to as disturbance intensities (i.e. $\mathbf{Q_d}=[Q_1,\ldots,Q_X]^T$).[5] $\delta(.)$ is the Dirac delta function. A large disturbance intensity corresponds either to large random shocks that influence the process or large mismatch between the true process and the behavior of the differential equations. Note that the model (1.1) is different from a traditional ODE model for a chemical process because it accounts for two types of noise: 1) the stochastic disturbances indicated by $\boldsymbol{\eta}(t)$ and the traditional measurement noise indicated by

$\boldsymbol{\varepsilon}(t_{mj})$, which appears in ODE models. The proposed methods in this thesis were developed assuming that the covariance matrixes of measurement noise $\boldsymbol{\Sigma}$ and power spectral density function $\mathbf{Q}$ are diagonal. However the proposed methods might also be used for cases where $\boldsymbol{\Sigma}$ and $\mathbf{Q}$ are not diagonal since derivations of the proposed objective functions do not require the matrixes $\boldsymbol{\Sigma}$ and $\mathbf{Q}$ to be diagonal. However, we have not performed any simulations to test whether it would be difficult, in practice, to obtain reliable estimates of off-diagonal parameters in these matrices. The proposed methods in this thesis were developed assuming that the covariance matrixes of measurement noise $\boldsymbol{\Sigma}$ and power spectral density function $\mathbf{Q}$ are diagonal. However the proposed methods might also be used for cases where $\boldsymbol{\Sigma}$ and $\mathbf{Q}$ are not diagonal since derivations of the proposed objective functions do not require the matrixes $\boldsymbol{\Sigma}$ and $\mathbf{Q}$ to be diagonal. However, we have not performed any simulations to test whether it would be difficult, in practice, to obtain reliable estimates of off-diagonal parameters in these matrices.

In this thesis, parameter estimation techniques for continuous time SDE models (Equation 1) are proposed. The reason for choosing continuous time models instead of discrete time models is that chemical engineering models appear in continuous form. Additionally, parameters in continuous models have physical meaning. It is also easier to interpret the model terms and parameters in continuous models.

When simulating processes described by Equation (1.1), it is common to use a discrete-time white-noise process to approximate the continuous stochastic process $\boldsymbol{\eta}(t)$, where the corresponding discrete process is a series of random step functions with a short sampling interval $\Delta t$, and covariance matrix:[15]

$$\mathrm{E}\{\boldsymbol{\eta}(j_1\Delta t)\boldsymbol{\eta}(j_2\Delta t)\} = \begin{cases} \dfrac{\mathbf{Q}}{\Delta t} & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases} \tag{1.4}$$

where $j_1$ and $j_2$ are positive integers.

A typical discrete-time white-noise process is shown in Figure 1.1.[5] The continuous-time white-noise process $\boldsymbol{\eta}(t)$ is the limiting case of a discrete-time white-noise process where $\Delta t \rightarrow 0$. Since the intensity of the model disturbance **Q** is not usually known a priori by the modeler, it should be estimated along with the model parameters.

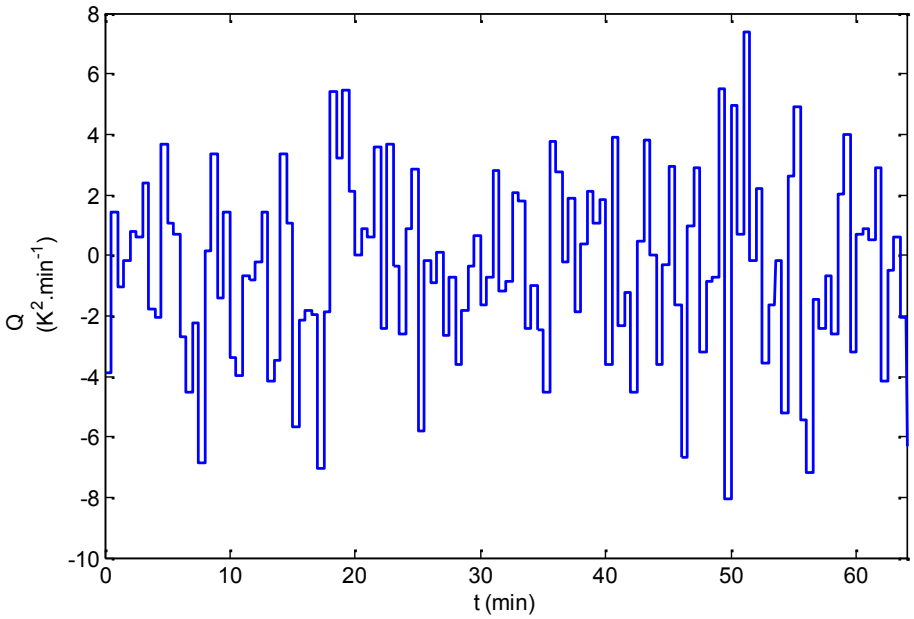



**Figure 1.1 Typical discrete stochastic process disturbance associated with a stochastic energy balance, obtained using $Q_1$= 4 $K^2$min-1 and Δt= 0.5 min**

The existence of a solution of an SDE is ensured when globally Lipschitz, linear growth and boundedness conditions are satisfied. These assumptions are described in detail by Liptser[16] and Bishwal.[17]

An alternative structure for expressing SDEs is:[16,17]

$$\mathrm{d}\,\mathbf{x} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})dt + \mathbf{Q}d\mathbf{W} \qquad (1.1.\mathrm{d})$$

Since the stochastic variable $\mathbf{W}(t)$ has a mathematical interpretation ($\mathbf{W}(t)$ is a Wiener process), SDEs are often written in the differential form shown in Equation 4.[16,17] Mathematicians regard white noise as the time derivative of a Wiener process (or Brownian motion).[17]

Denote $p_d$ as a vector of the diagonal elements of the noise covariance matrix (i.e., $p_d=[\pm_1^2,\ldots,\pm_Y^2]^T$). Let $\zeta=[\boldsymbol{\theta}^T,\mathbf{Q}_\mathbf{d}^T,\boldsymbol{\Sigma}_\mathbf{d}^T]^T$ be the vector of unknown parameters in the SDE model, which includes the model parameters $\boldsymbol{\theta}$, along with disturbance intensities $\mathbf{Q}$ and the unknown noise variance $p$. Information about the measurement noise variance is often available from repeated measurements or from sensor suppliers, but knowledge about the size of the imperfections and disturbances is not usually available to modelers. In some of the SDE parameter estimation studies performed in this thesis, $\boldsymbol{\Sigma}_\mathbf{d}$ will be assumed to be known a priori so that $\zeta=[\boldsymbol{\theta}^T,\mathbf{Q}_\mathbf{d}^T]^T$.

A survey of approaches to parameter estimation in dynamic models is presented below. This review is divided into two parts. The first part provides a literature review of parameter estimation in ODE models with too many parameters to estimate from limited data. The second part discusses available approaches for parameter estimation in SDE models.

## 1.2 Literature Review

### 1.2.1 Selecting Parameters for Estimation

When the number of data values available for parameter estimation is limited, modelers may face estimability issues.[1,10,18] Model parameters are estimable if their values can be estimated uniquely from the available data. Therefore, for complex models, a parameter estimability assessment can be helpful for assessing whether all of the model parameters can be estimated from the available data.[1-3,19] Simplifying the model by leaving some parameters at their initial values can resolve estimabilty issues.[1,10] Recently, statistical techniques have been developed to aid modelers in

selecting important parameters to estimate in complex models.[2,3,11,20,21] McLean and McAuley[1] provide a comprehensive review of estimability and identifiability analysis techniques.

Estimability analysis[10-12,22-26] ranks parameters from the most important to the least important. Parameters are ranked based on the influence of each parameter on the predicted model outputs, correlation between the effects of parameters, and uncertainties in initial parameter values.[11] Estimating too many parameters from the available data leads to poor model predictions due to increased variance, while estimating too few parameters and leaving the remaining parameters at their incorrect initial guesses leads to poor model predictions due to increased bias.[2,3] A mean square error-based (MSE) criterion was developed by Wu et al.[2,3] to select an optimal number of parameters to estimate to obtain the best tradeoff between bias and variance. Use of this methodology will be illustrated using the ODE model for nylon 66 production in Chapter 2.

### 1.2.2 Parameter Estimation in SDE Models

In this section, the problem of parameter estimation in SDE models is described. When estimating parameters in SDE models, it is common to have differential equations that are sufficiently simple and data sets that are sufficiently rich in information so that all of the parameters can be estimated. The difficulty in estimating the model parameters arises due to the vector of stochastic errors $\boldsymbol{\eta}(t)$ and the unknown disturbance intensities $\mathbf{Q_d}$.

A relatively simple approach to estimate parameters in nonlinear SDE models is based on Extended Kalman Filters (EKFs).[27-29] Traditional Kalman filters combine process measurements with dynamic model predictions for linear systems and track unmeasured states in the presence of process disturbances and measurement noise.[30,31] EKFs approximate nonlinear models by a series of updated linear models, and use traditional Kalman filter calculations for the linearized models.[28,29] EKFs can fail to converge to satisfactory estimates of parameters and states when the SDE model is highly nonlinear and the measurement times are far apart.[32]

Other more complicated methods for parameter estimation in SDEs using discrete observations fall into two main categories.[33] In the first category, a moment-matching method is used, while in the second category an approximate maximum likelihood is used.[34] Moment-matching methods do not depend on assumptions about the probability density functions of the measurements given the parameters. A problem that is common with moment-matching estimators is that they require calculation of higher-order moments in order to be efficient. As a result, the focus in this thesis is on the development and implementation of approximate ML methods.

Discretization of the SDE model is helpful when developing parameter estimation techniques for SDE models using ML methods. Using an Euler approximation Equations (1.1.a) and (1.1.b) become:

$$\mathbf{x}(t_{i-1}+\Delta t)=\mathbf{x}(t_i)=\mathbf{x}(t_{i-1})+f(\mathbf{x}(t_{i-1}),\mathbf{u}(t_{i-1}),\boldsymbol{\theta})\Delta t+\boldsymbol{\eta}_{\mathbf{d}}(t_{i-1})\Delta t \tag{1.5a}$$

$$\mathbf{x}(t_0)=\mathbf{x}_0 \tag{1.5b}$$

where $\mathbf{x}(t_i)$ is the value of the state variable at $q$ uniformly-spaced time points $t_i$ , $i=0,..,q$ and $\boldsymbol{\eta}_{\mathbf{d}}$ is a discrete white noise vector. The values of all $X$ state variables at the $q$ discretization times can be stacked in a vector denoted by $\mathbf{X_q}=[\mathbf{x}(t_1)^T,\cdots,\mathbf{x}(t_q)^T]^T$. The stacked vector of measurements at the measurement times is denoted by $\mathbf{Y_m}=[\mathbf{y}(t_{m1})^T,\cdots,\mathbf{y}(t_{mn})^T]$. The corresponding state values, inputs and random errors at the measurement times are $\mathbf{X_m}=[\mathbf{x}(t_{m1})^T,\cdots,\mathbf{x}(t_{mn})^T]$ and $\mathbf{U_m}=[\mathbf{u}(t_{m1})^T,\cdots,\mathbf{u}(t_{mn})^T]$. Note that measurements and states are assumed to be sampled at different times. Measurements are sampled at $t_{mj}$ ($j$ = 1…$n$) while states are sampled at $t_i$ ($i=0,..,q$).

The likelihood function of the parameters given the measurements is defined as[35]:

$$L(\boldsymbol{\zeta}\,|\,\mathbf{Y_m})=p(\mathbf{Y_m}\,|\,\boldsymbol{\zeta}) \tag{1.6}$$

In ML methods, the likelihood function of the parameters given the observed data $L(\boldsymbol{\zeta} | \mathbf{Y_m})$ is maximized to estimate $\boldsymbol{\zeta}$, which contains the unknown model parameters $\boldsymbol{\theta}$, the measurement noise variances $\rho_d$ and disturbance intensities $\mathbf{Q_d}$:

$$\hat{\boldsymbol{\zeta}} = arg \max_{\zeta} L(\boldsymbol{\zeta} | \mathbf{Y_m}) \tag{1.7}$$

ML estimators can result in biased estimates when data sets are small. ML techniques also have several implementation challenges. A major challenge is that an appropriate expression for the probability density function $p(\mathbf{Y_m} | \boldsymbol{\zeta})$ in Equation (1.7) is difficult to derive when some of the states are not measured (i.e., $Y<X$). When measurements are not available for one or more of the state variables, ML estimation becomes more complicated because of the need to estimate the unmeasured states.[35,36]

Numerous methods have been proposed for approximation of the likelihood function in SDE models. Some of these techniques are: simulated maximum likelihood methods (such as Markov Chain Monte Carlo (MCMC) techniques),[37-40] expansion of the likelihood function using Hermite polynomial basis functions,[41,42] solving the Fokker-Planck equation numerically[43-45] and recursive maximum likelihood parameter estimation using polynomial chaos theory.[46] Benefits and drawbacks of these techniques are summarized by Lindstrom.[34] Simulated maximum likelihood methods (SML) developed by Pederson[47] aim to estimate a probability density function by simulation. Marchov Chain Monte Carlo (MCMC) methods are SML methods that numerically approximate a probability density function by drawing samples from it.[37,38] Several MCMC algorithms have been developed.[37-39,48-53] The most popular one is the Metropolis-Hastings (M-H) algorithm.[54] A drawback of SML methods is that they are computationally expensive.

An approximate ML method that is computationally less expensive than MCMC methods was proposed by Kristensen et al.[55] In Kristensen's method, a Gaussian distribution is assumed for the likelihood function and the mean and variance of the likelihood function are estimated using an Extended Kalman Filter (EKF). Easy-to-use CTSM (continuous-time stochastic modeling)

software was developed based on Kristensen's method and is used in simulation studies later in this thesis.[27]

Poyton et al.[56] developed an iterative principal differential analysis (iPDA) method to estimate model parameters $\boldsymbol{\theta}$ in ODE models and later showed that these methods are appropriate for SDES if the disturbance intensities and noise variances are known.[5] iPDA uses B-splines to approximate the trajectories for the measured and unmeasured states. In the first step of iPDA, the B-spline coefficients $\boldsymbol{\beta}$ are estimated by fitting the B-splines to the observations using the most recent estimate of $\boldsymbol{\theta}$ and the following objective function:

$$\min_{\mathbf{B}} \sum_{j=1}^{n}[y_1(t_{mi}) - x_{1\sim}(t_{mi})]^2 + \ldots + \sum_{j=1}^{n}[y_Y(t_{mi}) - x_{Y\sim}(t_{mi})]^2 + \alpha_1 \int_{t_0}^{t_q}[\frac{dx_{1\sim}}{dt} - f_1(x_{1\sim}, \boldsymbol{\theta})]^2 \, \mathrm{d}t + \ldots + \alpha_X \int_{t_0}^{t_q}[\frac{dx_{X\sim}}{dt} - f_X(x_{X\sim}, \boldsymbol{\theta})]^2 \, \mathrm{d}t \tag{1.8}$$

where $\alpha_1,\ldots,\alpha_Y$ are positive weighting factors. $x_{\sim s}(t)$ is the approximate sth state trajectory computed from[56,57]:

$$x_{\sim s}(t) = \sum_{l=1}^{c_s} \beta_{s,l}\varphi_{s,l}(t) \quad \text{for } s=1,\ldots,X \tag{1.9}$$

The subscript ~ is used throughout this thesis to indicate a state trajectory that is approximated using B-splines. $\beta_{s,l}$ is a B-spline coefficient and $\varphi_{s,l}(t)$ is the corresponding B-spline basis function. In matrix form, Equation (1.9) is:

$$\mathbf{x}_{\sim}(t) = \boldsymbol{\Phi}(t)\mathbf{B} \tag{1.10}$$

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

$$\boldsymbol{\Phi}(t) = \begin{bmatrix} \boldsymbol{\varphi}_1^T(t) & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}_2^T(t) & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\varphi}_X^T(t) \end{bmatrix} \tag{1.11}$$

and

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_X \end{bmatrix} \tag{1.12}$$

In the second step of iPDA, the model parameters $\boldsymbol{\theta}$ are estimated by minimizing the objective function in Equation (1.13) using the most recent values of _:

$$\min_{\boldsymbol{\theta}} \quad \alpha_1 \int_{t_0}^{t_q} \left[\frac{dx_{1\sim}}{dt} - f_1(x_{1\sim}, \boldsymbol{\theta})\right]^2 \mathrm{d}t + \ldots + \alpha_X \int_{t_0}^{t_q} \left[\frac{dx_{X\sim}}{dt} - f_X(x_{X\sim}, \boldsymbol{\theta})\right]^2 \mathrm{d}t \tag{1.13}$$

iPDA iterates between the smoothing step (Equation (1.8) with the most recent values of $\boldsymbol{\theta}$) and the estimation step (Equation (1.13) with the most recent values of _) until the solution converges to good spline fits and good parameter values. Varziri et al.[5] showed that the iterative minimization of Equations (1.8) and (1.13) can be performed using a single step by simultaneously minimizing objective function (1.8) using the joint vector of the model parameters and the B-spline coefficients as decision variables. The iPDA objective function in Equation (1.8) contains model penalty terms in addition to the usual sum-of-squared-error terms. The model penalty terms account for model mismatch and process disturbances. In addition to controlling the smoothness of the spline functions, the penalty terms in Equation (1.8) ensure that the fitted B-spline curve will approximate the behavior of the SDE.

The main difficulty of iPDA is determining appropriate values for the weighting factors, which Poyton et al.[56] selected by trial and error. Varziri et al. [5]demonstrated that maximizing $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\theta})$ is equivalent to minimizing the iPDA objective function proposed by Poyton et al.,[56] when model mismatch results from additive stochastic white-noise disturbance inputs. Moreover, Varziri showed that the corresponding weighting factors in the iPDA objective function should be:

$$\alpha_s = \frac{\sigma^2_{m,s}}{Q_s} \tag{1.14}$$

When a particular state is not measured, the corresponding sum-of-squared-error term does not appear in objective function (1.8) and a very large value of $\alpha_s$ is appropriate for the corresponding model-based penalty term, because $\sigma^2_{m,s}$ becomes infinite. Using a large weighting factor ensures that the optimizer will select the spline coefficients so that the differential equation for the unmeasured state is satisfied with only a small amount of error. For the measured states, the values of $\alpha_s$ from Equation (1.14) ensure an appropriate tradeoff between B-spline trajectories that match the data and that match the behavior of the differential equations. This extended iPDA technique was referred to as approximate maximum likelihood estimation (AMLE) by Varziri et al.[5] since they maximized an approximation to the likelihood function.

The main drawback of AMLE is that its application requires knowledge of the value of the stochastic process disturbance intensities $\mathbf{Q_d}$, which are not usually known to the modeler. Varziri et al.[58] modified the formulation of AMLE using a technique developed by Heald and Stark [59] for cases in which measurement-noise variances $\mathbf{\Sigma_d}$ are known to the modeler but the process disturbance intensities $\mathbf{Q_d}$ are not known, and proposed a two-step optimization algorithm to obtain $\boldsymbol{\theta}$ and $\mathbf{Q_d}$. In this algorithm, the inner optimization problem minimizes the AMLE objective function with respect to the model parameters and B-spline coefficients using an assumed value of $\mathbf{Q_d}$. An outer optimization problem selects $\mathbf{Q_d}$ to ensure that the estimated measurement noise variances $\hat{\sigma}^2_r$ are close to the known values $\sigma^2_r$ as follows:

$$\hat{\mathbf{Q}} = \arg\min_{\mathbf{Q}} \left\{ \frac{\hat{\sigma}^2_1(Q_1)}{\sigma^2_1} - 1 \right\}^2 + \ldots + \left\{ \frac{\hat{\sigma}^2_Y(Q_Y)}{\sigma^2_Y} - 1 \right\}^2 \tag{1.15}$$

or equivalently:

$$\hat{\mathbf{Q}} = \arg\min_{\mathbf{Q}} \left\{ \frac{\mathbf{\Sigma}^{-1}}{n} [(\mathbf{Y}_m - \mathbf{X}_{\sim m}(\mathbf{Q}))^T (\mathbf{Y}_m - \mathbf{X}_{\sim m}(\mathbf{Q})) + Trace(\mathbf{H}^{-1})] - \mathbf{I} \right\}^2 \tag{1.16}$$

where **H** is the Hessian matrix of the AMLE objective function with respect to the state variables evaluated at $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\theta}}$ obtained from the inner optimization and **I** is the identity matrix. The main drawback of this method is that it cannot be used for cases where some of the measurement noise variances are unknown. Furthermore, objective function (1.15) was selected arbitrarily by Varziri et al. to ensure that the estimated noise variances are close to the assumed noise variances. No theoretical justification for Equation (1.15) was provided.

An early ML algorithm for parameter estimation in SDE models is the Expectation Maximization (EM) algorithm proposed by Dempster et al.[60]. The idea underlying the EM algorithm is that calculating the probability density function of the complete data $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ is easier than calculating the probability density function of observed data given the parameters $p(\mathbf{Y_m} \mid \boldsymbol{\zeta})$. The EM algorithm has two steps. In the first step, referred to as expectation or E step, the expected value of the probability density of the complete data is estimated based on current values of the parameters (from the *k*th iteration or the initial guesses if *k*=0):[61]

$$R(\boldsymbol{\zeta}, \boldsymbol{\zeta}_k) \approx \underset{\mathbf{X_q}}{E}\{\ln[p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})] \mid \mathbf{Y_m}, \boldsymbol{\zeta}_k\} = \int \ln[p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})] p(\mathbf{X_q} \mid \mathbf{Y_m}, \boldsymbol{\zeta}_k) \, \mathrm{d}\, \mathbf{X_q} \tag{1.17}$$

In Equation (1.17), $\mathbf{X_q}$ below the expectation symbol indicates that this expected value is computed over all possible values of the discretized state variables.

In the second step, referred to as the maximization or M step, the expected value evaluated in the first step is maximized using:

$$\hat{\boldsymbol{\zeta}}_{k+1} = \arg \max_{\boldsymbol{\zeta}} \; R(\boldsymbol{\zeta}, \boldsymbol{\zeta}_k) \tag{1.18}$$

In linear cases with Gaussian noise, explicit recursive equations for computing the required density functions have been developed.[62] However, for nonlinear SDE models, no explicit solutions for the E step and M step are available.[50] Several approaches have been proposed by researchers to approximate the E and M steps of the EM algorithm. For example, extended

Kalman filters (EKFs) have been used for approximating both the E and M steps and [63-66] Sequential Monte Carlo (SMC) methods, also known as particle filter methods, have been used for approximating the E step.[37,49,50,53,67-69]

SMC methods are effective parameter estimation tools that do not require assumptions about the form of the density functions, but the number of mathematical operations required in each optimization step for a typical SMC method is $8qn(X^3+Y^3)$.[48] The complexity of the optimization problem increases rapidly as the number of states, measurements, and parameters increases. Hence, the optimization procedure can be very slow and computationally prohibitive.[50,69,70] Imtiaz et. al.[71] discuss some implementation issues for SMC methods and Kantas et al.[72] present an overview of a variety of SMC methods and discuss their advantages and disadvantages. Recently, Chitralekha et al.[73] compared the performance of three SMC-based EM algorithms: the particle smoother, the unscented Kalman smoother and the extended Kalman smoother. Linearization-based EKF methods for approximating the EM algorithm are beneficial because they do not require MC sampling from probability density functions, but they can give biased parameter estimates in situations where the nonlinearities are strong.[32]

### 1.2.3 Summary of Literature Review

In summary, one type of difficult parameter estimation problem encountered by chemical engineers results from complicated mechanistic models with too many parameters and limited available data. In these situations, statistical techniques have been developed to aid modelers in selecting important parameters to estimate.[1] Techniques that have been developed for parameter ranking[11,12] and for determining the appropriate number of parameters to estimate from the ranked parameter list so that the best possible predictions can be obtained. These methods have been used for a variety of chemical and biochemical process models, but they need further testing on larger-scale practical models with >10 differential equations and ~50 parameters. This is one of the objectives of the modeling and parameter estimation work described in Chapter 2.

A second type of difficult parameter estimation problem occurs when stochastic terms are included in differential equation models to account for possible modeling imperfections and process disturbances. Currently, the most popular methods for estimating the parameters of these SDE models are relatively simple EKF methods and more complex ML methods. For highly nonlinear models, state and parameter estimation using EKFs may perform poorly due to bias and lack of convergence. Maximum likelihood estimation (MLE) has been adopted because they can provide asymptotically consistent and efficient estimates of the model parameters.[36] However, mathematical expressions for the conditional density functions in ML algorithms typically include high-dimensional integrals that are not amenable to analytical simplifications. Hence, an approach that approximates the high-dimensional integrals through an iterative Markov Chain Monte Carlo (SMC) method is sometimes used. The difficulties of parameter estimation in SDE models grow with the number of states and parameters in the system. None of the current ML methods are straightforward, and even for systems of modest size. All can lead to problems with slow algorithmic progress toward the minimum, lack of convergence, and computational complexity.[36,48-50,71] The recently developed AMLE method of Varziri et al. uses B-spline basis functions to approximate the state trajectories when performing approximate ML parameter estimation. Extension of these methods to more difficult problems (i.e., when the noise variances and disturbance intensities are unknown) is explored in Chapters 3 to 5.

The material in Chapter 2 has been published in Macromolecular Reaction Engineering and the material in Chapter 3 has been accepted for publication in the Canadian Journal of Chemical Engineering. The material in Chapter 4 has been submitted to Industrial and Engineering Chemistry Research and the material in Chapter 5 has been prepared for submission to a journal. To keep each chapter self-contained and to keep the presentation in this thesis close to the original manuscripts, some information repeated appears in multiple chapters. Most the repetition is associated with the defining the SDE model equations and associated notation and in presenting the CSTR model used in the simulation studies. Note that the abstracts for these

chapters have been revised slightly when compared with the original manuscripts in an effort to better describe the links between the various chapters.

## 1.3 References


1. McLean KA, McAuley KB. Mathematical modelling of chemical processes- obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *The Canadian Journal of Chemical Engineering*. 2012.

2. Wu S, McAuley K, Harris T. Selection of simplified models: I. Analysis of model-selection criteria using mean-squared error. *The Canadian Journal of Chemical Engineering*. 2011;89:148-158.

3. Wu S, McAuley K, Harris T. Selection of simplified models: II. Development of a model selection criterion based on mean squared error. *The Canadian Journal of Chemical Engineering*. 2011;89:325-336.

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

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

6. Maria G. A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chemical and Biochemical Engineering Quarterly*. 2004;18:195-222.

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

8. King R. Applications of Stochastic Differential Equations to Chemical-Engineering Problems-An Introductory Review. *Chemical Engineering Communication*. 1974;1:221-237.

9. Englezos P, Kalogerakis N. *Applied parameter estimation for chemical engineers*. CRC Press, 2000.

10. Chou I, Voit EO. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Mathematical Bioscience*. 2009;219:57-83.

11. Thompson DE, McAuley KB, McLellan PJ. Parameter Estimation in a Simplified MWD Model for HDPE Produced by a Ziegler-Natta Catalyst. *Macromolecular Reaction Engineering*. 2009;3:160-177.

12. Yao KZ, Shaw BM, Kou B, McAuley KB, Bacon D. Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design. *Polymer Reaction Engineering*. 2003;11:563-588.

13. Gagnon L, MacGregor J. State estimation for continuous emulsion polymerization. *The Canadian Journal of Chemical Engineering*. 2009;69:648-656.

14. Érdi P, Tóth J. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models.* Manchester University Press, 1989.

15. Maybeck PS. *Stochastic models, estimation and control.* Academic Pr, 1982.

16. Liptser RS, Shiryaev AN. *Statistics of Random Processes: II. Applications.* Springer, 2000.

17. Bishwal JPN. *Parameter estimation in stochastic differential equations*. Springer, 2008;

18. Huang JZ, Wu CO, Zhou L. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*. 2004;14:763-788.

19. Zhou M. Fully exponential Laplace approximation EM algorithm for nonlinear mixed effects models.PhD Thesis, University of Nebraska-Lincoln,USA, 2009.

20. Li R, Henson MA, Kurtz MJ. Selection of model parameters for off-line parameter estimation. *Control Systems Technology, IEEE Transactions on*. 2004;12:402-412.

21. Lillacci G, Khammash M. Parameter estimation and model selection in computational biology. *PLoS computational biology*. 2010;6:e1000696.

22. Chu Y, Hahn J. Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters. *Industrial Engineering Chemistry & Research*. 2008;48:6000-6009.

23. Littlejohns JV, McAuley KB, Daugulis AJ. Model for a solid–liquid stirred tank two-phase partitioning bioscrubber for the treatment of BTEX. *J Hazard Mater*. 2010;175:872-882.

24. Schittkowski K. Experimental design tools for ordinary and algebraic differential equations. *Industrial Engineering Chemistry and Research*. 2007;46:9137-9147.

25. Quaiser T, Mönnigmann M. Systematic identifiability testing for unambiguous mechanistic modeling–application to JAK-STAT, MAP kinase, and NF-κB signaling pathway models. *BMC systems biology*. 2009;3:50.

26. Lund BF, Foss BA. Parameter ranking by orthogonalization—applied to nonlinear mechanistic models. *Automatica*. 2008;44:278-281.

27. Kristensen NR, Madsen H. Continuous time stochastic modelling. CTSM 2.3: User's Guide, Technical University of Denmark. 2003.

28. MacGregor JF, Penlidis A, Hamielec AE. Control of Polymerization Reactors-a Review. *Polymer Process Engineering*. 1984;2:179.

29. Kozub DJ, MacGregor JF. State estimation for semi-batch polymerization reactors. *Chemical Engineering Science*. 1992;47:1047-1062.

30. Jones R, MacGregor J, Murphy K. State estimation in wastewater engineering: Application to an anaerobic process. *Environment Monitoring Assessments*. 1989;13:271-282.

31. Kalman RE. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*. 1960;82:35-45.

32. Chen T, Morris J, Martin E. Particle filters for state and parameter estimation in batch processes. *Journal of Process Control*. 2005;15:665-673.

33. Nielsen JN, Madsen H, Young PC. Parameter estimation in stochastic differential equations: an overview. *Annual Reviews in Control*. 2000;24:83-94.

34. Lindström E. Estimating parameters in diffusion processes using an approximate maximum likelihood approach. *Annals of Operations Research*. 2007;151:269-288.

35. Jazwinski AH. *Stochastic processes and filtering theory*. Academic press, 1970.

36. Ljung L. *System identification.* Wiley Online Library, 1999.

37. Doucet A, De Freitas N, Gordon N. *Sequential monte carlo methods.* Springer, 2001.

38. Doucet A, Tadić VB. Parameter estimation in general state-space models using particle methods. *Annals of the institute of Statistical Mathematics*. 2003;55:409-422.

39. Jang SS, De la Hoz H, Ben-zvi A, McCaffrey WC, Gopaluni RB. Parameter estimation in models with hidden variables: An application to a biotech process. *The Canadian Journal of Chemical Engineering*. 2011;90:690-702.

40. Jaeger J, Lambert P. *Bayesian generalized profiling estimation in hierarchical linear dynamic systems*. Universite Catholique de Louvain, Belgium, 2011.

41. Aït-Sahalia Y. Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. *Econometrica*. 2003;70:223-262.

42. Singer H. Moment equations and Hermite expansion for nonlinear stochastic differential equations with application to stock price models. *Computational Statistics*. 2006;21:385-397.

43. Singer H. Parameter estimation of nonlinear stochastic differential equations: simulated maximum likelihood versus extended Kalman filter and Itô-Taylor expansion. *Journal of Computational and Graphical Statistics*. 2002;11:972-995.

44. Lo AW. Maximum Likelihood Estimation of Generalized Itˆo Processes with Discretely Sampled Data. *Econometric Theory*. 1988;4(2):231-247.

45. Christensen B, Poulsen R, Sørensen M. *Optimal inference in diffusion models of the short rate of interest*. CAF, Centre for Analytical Finance. 2001.

46. Pence BL, Fathy HK, Stein JL. A maximum likelihood approach to recursive polynomial chaos parameter estimation. American Control Conference. 2010:2144-2151.

47. Pedersen AR. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian journal of statistics*. 1995:55-71.

48. Gopaluni RB. Nonlinear system identification under missing observations: The case of unknown model structure. *Journal of Process Control*. 2010;20:314-324.

49. Gopaluni RB. Identification of non-linear processes with known model structure under missing observations. The International Federation of Automatic Control Seoul, Korea, July 6-11 2008;11.

50. Gopaluni R. A particle filter approach to identification of nonlinear processes under missing observations. *The Canadian Journal of Chemical Engineering*. 2008;86:1081-1092.

51. Haario H, Laine M, Mira A, Saksman E. DRAM: efficient adaptive MCMC. *Statistics and Computing*. 2006;16:339-354.

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

53. Schön TB, Wills A, Ninness B. System identification of nonlinear state-space models. *Automatica*. 2011;47:39-49.

54. Metropolis AW, Rosenbluth MN, Rosenbluth AH. Equations of state calculations by fast computing machine. *Journal of Chemical Physics*. 1953;21:1087-1091.

55. Kristensen NR, Madsen H, Jørgensen SB. Parameter estimation in stochastic grey-box models. *Automatica*. 2004;40:225-237.

56. Poyton A, Varziri MS, McAuley KB, McLellan P, Ramsay JO. Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers and Chemical Engeering*. 2006;30:698-708.

57. Ramsay J, Silverman B. Functional data analysis. Springer. 2005.

58. Varziri M, McAuley K, McLellan J. P. Parameter and state estimation in nonlinear stochastic continuous-time dynamic models with unknown disturbance intensity. *The Canadian Journal of Chemical Engineering*. 2008;86:828-837.

59. Heald J, Stark J. Estimation of noise levels for models of chaotic dynamical systems. *Physical Review Letters*. 2000;84:2366-2369.

60. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Sta Soc Series B,* 1977:1-38.

61. Anderson BDO, Moore JB. *Optimal filtering.* Prentice-hall Englewood Cliffs, NJ, 1979.

62. Shumway RH, Stoffer DS. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of time series analysis*. 2008;3:253-264.

63. Duncan S, Gyongy M. Using the EM algorithm to estimate the disease parameters for smallpox in 17th century London. 2006:3312-3317.

64. Roweis S, Ghahramani Z. An EM algorithm for identification of nonlinear dynamical systems. University College London. 2000.

65. Goodwin GC, Aguero JC. Approximate EM algorithms for parameter and state estimation in nonlinear stochastic models. 2005:368-373.

66. Ghahramani Z, Roweis ST. Learning nonlinear dynamical systems using an EM algorithm. *Advances in neural information processing systems*. 1999:431-437.

67. Poyiadjis G, Doucet A, Singh SS. Maximum likelihood parameter estimation in general state-space models using particle methods. Proceeding of the American Statistical Association 2005.

68. Schön TB, Wills A, Ninness B. *Maximum likelihood nonlinear system estimation.* Department of Electrical Engineering, Linköpings universitet, 2005.

69. Andrieu C, Doucet A, Singh SS, Tadic VB. Particle methods for change detection, system identification, and control. *Proceeding IEEE*. 2004;92:423-438.

70. Golightly A, Wilkinson DJ. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus*. 2011;1:807-820.

71. Imtiaz SA, Roy K, Huang B, Shah SL, Jampana P. Estimation of states of nonlinear systems using a particle filter. University of Alberta, Canada. 2006:2432-2437.

72. Kantas N, Doucet A, Singh SS, Maciejowski JM. An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. 2009.

73. Chitralekha SB, Prakash J, Raghavan H, Gopaluni R, Shah SL. A comparison of simultaneous state and parameter estimation schemes for a continuous fermentor reactor. *Journal of Process Control*. 2010;20:934-943.

# Chapter 2

# Kinetic Model for Non-Oxidative Thermal Degradation of Nylon 66

H. Karimi, M.A. Schaffer, K. B. McAuley

## 2.1 Abstract

In this section, the use of some traditional parameter estimation and selection techniques are illustrated in a complex ODE model when there is insufficient data to estimate all the parameters and unknown initial conditions. An improved kinetic model was developed for thermal degradation of molten nylon 66 to illustrate and test estimation strategies for estimating parameters in complicated models with too many parameters and limited data available for parameter estimation. One objective of this modeling work was to determine whether all of the parameters could be estimated from available data and if not, to obtain the best possible estimates of the parameters that should be estimated. Elimination of well-known or unimportant model parameters avoids numerical problems during parameter estimation. Estimating too many parameters from the available data can lead to poor model predictions due to increased variance. Estimating too few parameters, and leaving the remaining parameters at their incorrect initial guesses, can lead to poor model predictions due to increased bias. The proposed model, which describes the effect of melt-phase water concentration on degradation, matches the data well with typical errors of 6.1% and 2.9%, respectively, for amine ends (A) and carboxyl ends (C) and 4.3%, 27.2% and 29.4%, respectively, for evolution rates of $CO_2$, $NH_3$ and cyclopentanone (CPK). This chapter has been published as a journal paper in *Macromolecular Reaction Engineering* 6, 93-109, 2012.

## 2.2 Introduction

Nylon 66 is an important thermoplastic polymer. Because of its physical properties such as strength, toughness, stiffness, processability and resistance to heat, it has a wide range of applications from resin to fiber and film.[1] Nylon 66, or poly(hexamethylene adipamide), is produced from the condensation of adipic acid and hexamethylene diamine. The tendency of adipic acid residue segments to cyclize is one of the most important factors that causes thermal degradation in nylon polymerization reactors.[2] Thermal degradation reactions are unwanted side reactions that result in changes in the balance of reactive end groups, evolution of gaseous degradation products, and branching, which can lead to gelation.These reactions tend to occur at the high temperatures used during the final finishing stages of industrial nylon 66 production, where high temperatures and low water concentrations are used to achieve high molecular weight.[2] Degradation eventually affects final product quality by changing important properties such as processability, dyeability, physical properties and colour of the nylon product.[1] Note that this research is concerned with thermal degradation, rather than oxidative degradation, because oxygen was carefully excluded from the lab-scale reactor (see Figure 2.1) used to obtain the data[4-6] for this model study and is excluded from industrial nylon polymerization reactors.

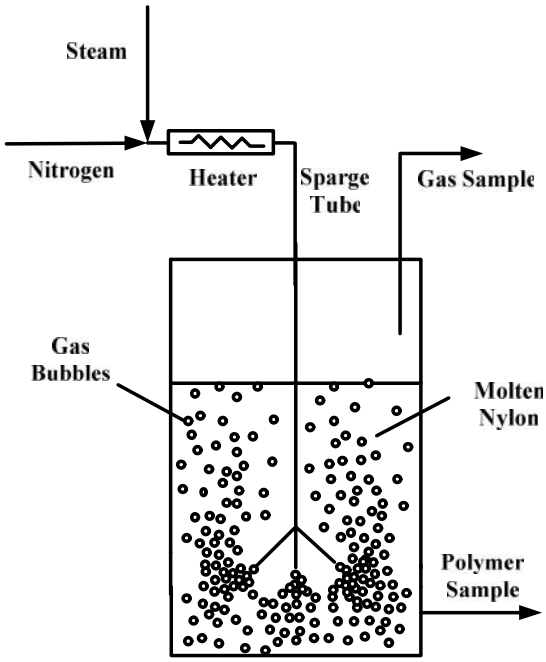
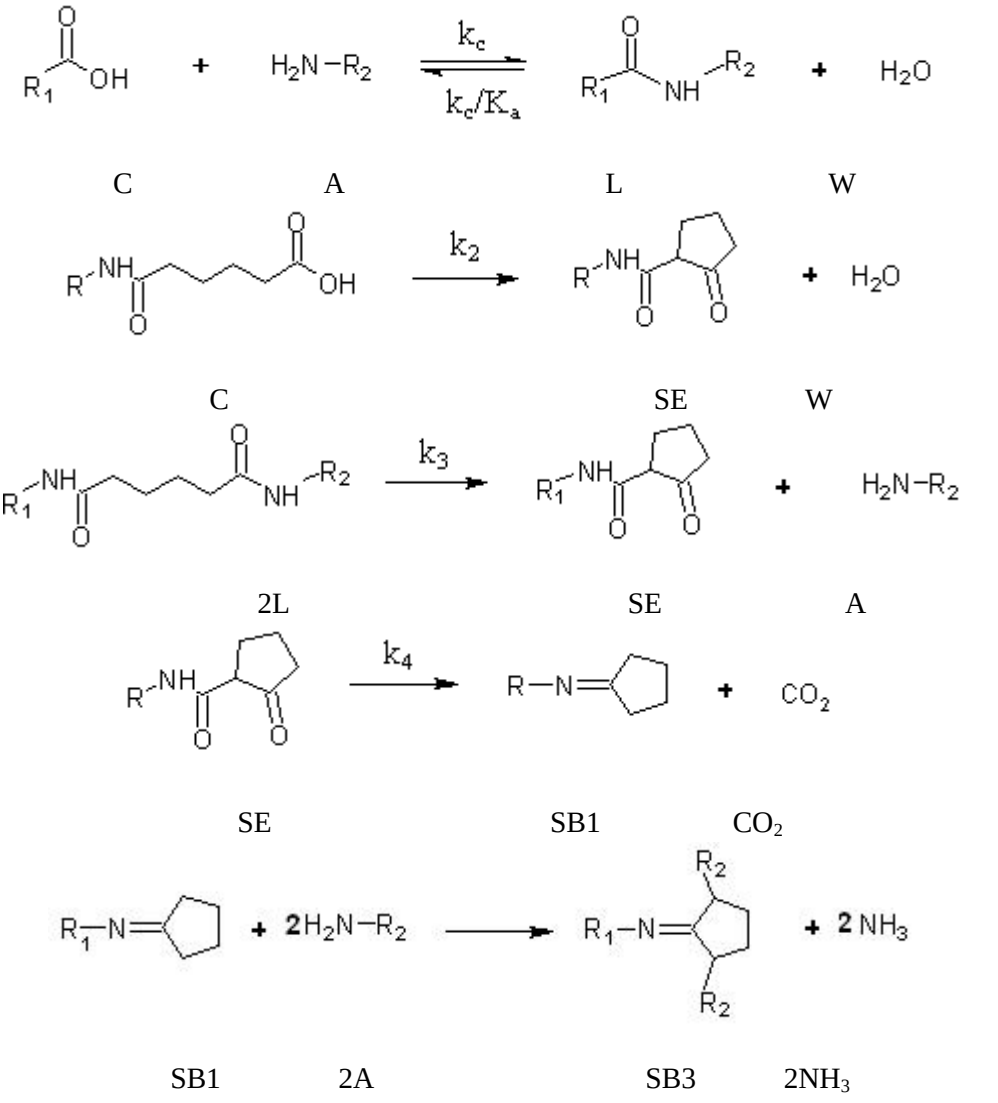



**Figure 2.1 Simplified schematic diagram of polymerization/degradation reactor system. Details concerning temperature control, mixing and polymer sampling are provided by Schaffer et al.[3]**

For each run, approximately 1.8 kg of additive-free nylon 66 polymer pellets were used as received from DuPont Canada with an amine end-group concentration of 58.8 $mmol \cdot kg^{-1}$, and a carboxylic acid end-group concentration of 94.5 $mmol \cdot kg^{-1}$. The pellets were melted via simultaneous heating of the reactor walls and circulation of the preheated mixture of nitrogen and steam through the reactor vessel for approximately 1 h. After the temperature in the center of the reactor had reached a value near 200 °C, impeller drive was started with a set-point value of 20 rpm. After the melt temperature reached the desired setpoint value (approximately 3 h), collection of polymer and gas samples for subsequent analysis began. Temperatures were measured in two places: in the center of the reactor vessel and at the reactor wall. The differences in these two

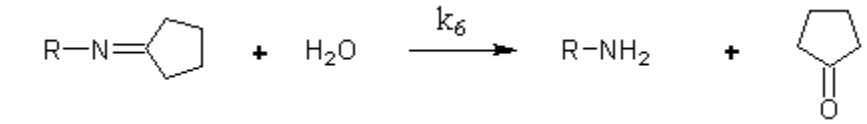
measurements varied between 2 and 11 ˚C in different runs. The reported run temperatures (see Table 2.4) are averages of these two temperatures. Steppan et al. [7] developed the only published model for the thermal degradation of nylon 66 and used literature data for parameter estimation. They developed a minimum subset of the degradation reactions, based on their knowledge at that time, and reactions suggested by Wiloth [8] to describe end-group concentration changes and the evolution of $NH_3$ and $CO_2$, and then fitted their kinetic constants to data. Their mechanism, shown in Table 2.1, accounts for a decrease in the concentration of carboxylic acid end groups and an increase in amine end-group concentration with time, as well as polymer chain branching and evolution of $CO_2$ and $NH_3$. Although the model by Steppan et al.[7] is in agreement with the data they used, it does not account for the formation of cyclopentanone and for branches that result from amine end-group condensation (Reactions (a.7) and (a.8) in Table 2.2). Pimentel and Giudici[11] used Steppan's kinetic scheme and rate constants to develop a mathematical model for nylon 66 production in a two-phase continuous tubular reactor, followed by a continuous stirred tank, and validated their model using industrial data.

**Table 2.1 The kinetic scheme proposed by Steppan et al.[7] to account for polyamidation and thermal degradation in melt-phase nylon 66**



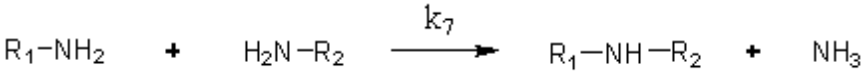

.

**Table 2.2 The supplementary kinetic scheme for thermal degradation in melt phase nylon 66 were proposed by Schaffer[9] and McAuley.[10]**



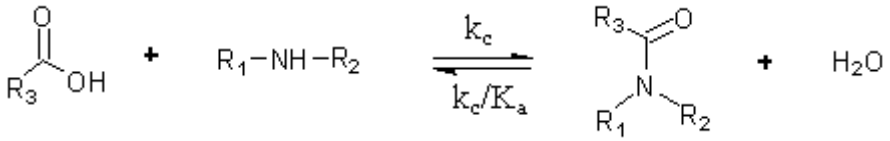
(a.6)

SB1 W A CPK



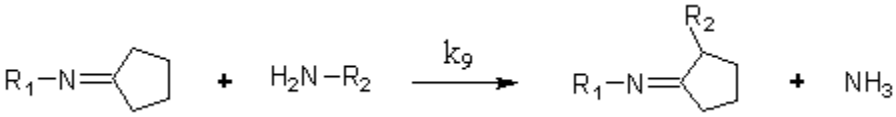
(a.7)

A A A2 $NH_3$



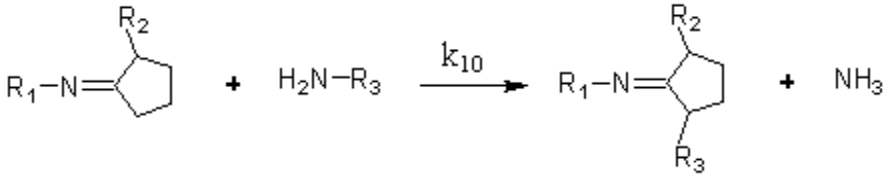
(a.8)

C A2 L2 W



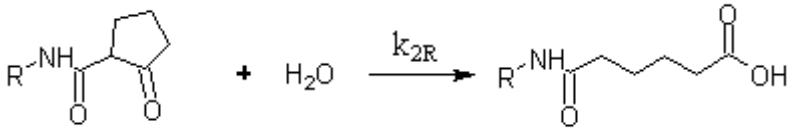
(a.9)

SB1 A SB2 $NH_3$



(a.10)

SB2 A SB3 $NH_3$



(a.2.R)

SE W C

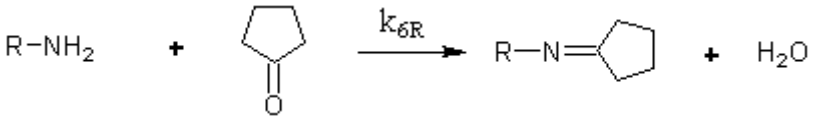
In an effort to learn more about degradation reactions, Schaffer et al.[4] used the reactor system in 2.1 to collect data for end-group concentrations, water content and branch concentrations in molten nylon 66, as well as information about off-gas evolution rates. Schaffer[9] and McAuley[10] used these data to develop a more comprehensive kinetic

scheme for nylon 66 polycondensation under conditions of high temperatures and low water concentrations, as shown in Table 2.2. Values for amidation rate constants were determined from experiments involving nylon 612.[5] Nylon 612, which is produced using dodecanoic acid in place of adipic acid, is more thermally stable than nylon 66. Reactions (a.2) to (a.5) in Table 2.1 can only occur in nylons that are produced using adipic acid. From their nylon 612 data, Schaffer et al.[5] were unable to determine whether the order of the polycondensation reaction was first-order with respect to carboxyl end-groups or second-order as had been suggested by Mallon et al.[12] Kinetic rate constants were fitted twice, using the different assumptions about the reaction order. Schaffer's data[5] could be described using either assumption. Unfortunately, a second-order influence of carboxyl ends on amidation rate was assumed in a subsequent nylon 66 degradation model.[9] Subsequently, Zheng et al.[6] performed additional nylon 612 experiments, using polymer samples with different relative concentrations of carboxyl and amine ends and determined that the polycondensation reaction is first-order in carboxyl ends, rather than second order. As a result, predictions and parameter estimates from Schaffer's model[9] may be unreliable if applied at conditions differing greatly from those at which Schaffer's data were collected. Varziri et al.[13] recently performed an additional parameter estimation study for nylon 612, using the kinetic data of Zheng et al.[6] and additional reaction equilibrium data.[14]

Recently, Schaffer et al.[4] compared their nylon 66 degradation data with predictions from Steppan's model and found that there is a significant discrepancy between Steppan's model predictions and Schaffer's degradation data. As a result, Schaffer et al. proposed additional reactions to be considered in the overall reaction scheme, as shown in Table

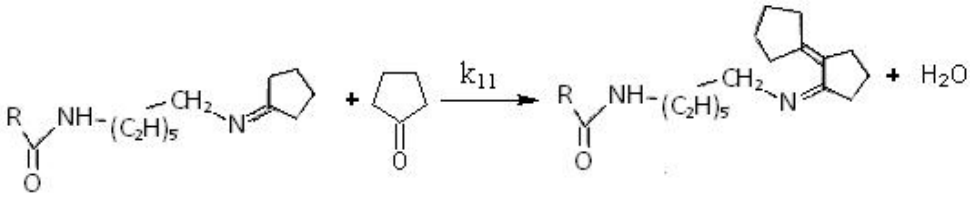
2.2 and Table 2.3. Table 2.2 provides reactions that were included in the original model by Schaffer.[9]

**Table 2.3 Supplementary reactions for thermal degradation of nylon 66 in a melt phase proposed by Schaffer et al.[4]**




(a.6.R)

A CPK SB1 W




(a.11)

SB1 CPK P244int W

Reactions in Table 2.3 are additional reactions required to explain the formation of pyridine 244 ends, which were observed in the nylon 66 degradation products after Schaffer's initial degradation model had been developed.[4] We did not include one of the reactions (SE + W $\rightarrow$ A + CPK + $CO_2$) proposed by Schaffer et al.[4] in the current mechanism because it is the sum of reactions (a.4) and (a.6). As a result of the large number of reactions in Table 2.1, Table 2.2 and Table 2.3, a new model that properly accounts for degradation of nylon 66 will contain a large number of kinetic parameters. Data for estimating these parameters are available from six dynamic experiments that were conducted using the conditions shown in Table 2.4.[4]

**Table 2.4 Experimental run conditions[4]**

| Run | Temperature [˚C] | $P_w$ [kPa] | Time duration [h] |
|---|---|---|---|
| 1 | 281 | 92 | 9.1 |
| 2 | 275 | 57 | 10.6 |
| 3 | 285 | 93 | 6.6 |
| 4 | 292 | 59 | 3.1 |
| 5 | 286 | 23 | 2 |
| 6 | 290 | 101 (lowered to 23 at 3.3 h) | 4.5 |

$P_w$, in column three, is the partial pressure of water in a nitrogen/water vapour mixture that Schaffer bubbled through the molten nylon. He used different temperatures and moisture levels in the six experiments to investigate the influence of water and temperature on degradation rates. A constant water partial pressure was used for the entire duration of Runs 1 to 5. In Run 6, pure steam was fed at the beginning of the experiment ($P_w$=101 kPa) and then a lower water partial pressure was used for the remainder of the run. Literature data from earlier studies[8,15,16] are not suitable for parameter estimation with this full reaction scheme because key information, such as the moisture level in the molten nylon is not provided in these data sets.

It is not clear whether all of the model parameters can be estimated reliably from the available data. Recently, statistical techniques have been developed to aid modelers when estimating parameters in complex models using limited data.[17-21] The objective of these estimability analysis and parameter selection techniques is to aid the modeler in determining which model parameters should be estimated from the available data, and which parameters should be remain at their initial values. First, the model parameters are ranked from most estimable to least estimable, using a sensitivity-based technique.[17,18] Parameters that appear near the top of the ranked list are those that have the most influence on predictions of the data. Parameters appear near the bottom of the list

because they have little influence on the model predictions, because their effects are correlated with those of parameters that appear higher on the list, or because their values are already precisely known. Estimability analysis has been used successfully to rank parameters in models for a number of different chemical and biological reaction systems.[22-28]

In the current work, a mean-squared error criterion is used to determine the appropriate number of parameters to estimate from the ranked list.[19,21] Estimating too many parameters from the available data can lead to poor model predictions due to increased variance. Estimating too few parameters, and leaving the remaining parameters at their incorrect initial guesses, can lead to poor model predictions due to increased bias.[20] One objective of the current modeling work is to determine whether all of the parameters can be estimated from Schaffer's[4] data and, if not, to obtain the best possible estimates of the parameters that should be estimated.

In this work, a dynamic kinetic model is developed to explain Schaffer's experimental data, which were obtained at relatively high temperatures (275 to 292 ˚C) and low water concentrations (10 to 80 mmol $kg^{-1}$).[4] Estimability analysis[17,18,27] and a mean square error-based (MSE) criterion[19,21] are used to determine which parameters can be estimated from the available data. Finally, the parameters and unknown initial conditions are estimated and model predictions are compared with experimental data. The resulting model will be a useful tool for engineers who wish to design improved equipment or operating strategies for the final stages of nylon 66 production.

## 2.3 Model Development

### 2.3.1 Reaction Pathways

Many chemical reactions have been described in the literature to explain the mechanism of thermal degradation of nylon 66.[2,29,30] Inclusion of all or a substantial portion of these proposed reactions in a kinetic model is not feasible because sufficient experimental data are not available to estimate of all of the rate and equilibrium parameters that would be required in such a complex model. We therefore adopt the goal of developing a simplified, reasonable kinetic model that is capable of describing the experimental data of Schaffer et al.[4]

In Tables 2.1, 2.2 and 2.3, convenient symbols are shown below the structural diagrams for the chemical species that are considered in the model (e.g., C for carboxyl end, A for amine end, L for amide link, W for water, SE for stabilized end, CPK for cyclopentanone and SB1, SB2 and SB3, respectively, for Schiff base groups attached to one, two or three chains).

The experimentally-observed phenomena to be accounted for by the model include: decreases in carboxylic acid end-group concentrations with time; increases in amine end-group concentrations with time; evolution of ammonia, carbon dioxide and cyclopentanone; branching of the polymer resulting in 2,5-di(6-aminohexyl) cyclopentantone and bis-hexamethylene triamine (BHMT) in the polymer hydrolysate, and the formation of pyridine 244 ends. Note that the data set of Schaffer et al.[4] includes measurements of gas evolution rates from samples collected over the course of each experiment. End-group and branch point (BHMT only) concentration measurements are also available from polymer samples collected at various times during each dynamic experiment.

The proposed model uses a reaction scheme that consists of reactions (a.1) to (a.4), (a.6) to (a.10), (a.2.R), (a.6.R) and (a.11) shown in Table 2.1 to Table 2.3. Reaction (a.5) in Table 2.1 is not required, because it appears as two separate reactions ((a.9 and (a.10)) in Table 2.2. Reaction (a.1) is the desired step-growth polycondensation reaction, which consumes carboxyl ends and amine ends react to form amide links water. Reaction (a.2) is a cyclization reaction of the carboxylic acid end to form a stabilized end and water. Reaction (a.3) is a similar reaction that occurs when a polymer chain is broken to form a stabilized end an amide end. The stoichiometric coefficient of 2 associated with the amide links in reaction (a.3) does not mean that two amide links actually react with each other. Rather, it reflects the fact that two amide links (with molecular weight (MW) 113 g $mol^{-1}$) are consumed by reaction (a.3) and should be accounted for in the stoichiometry. Reaction (a.4) is an intramolecular reaction wherein a stabilized end group forms a Schiff base end and $CO_2$ is generated. The evolution of significant quantities of cyclopentanone, which was not included in the Steppan model, is accounted for by reaction (a.6). Hydrolysis of Schiff bases is a well-known reaction that has been studied previously.[31] Branching reactions (a.7) and (a.8), are included in the model to account for the appreciable concentrations of BHMT detected in the hydrolyzed polymer samples. BHMT is produced from the hydrolysis of both the secondary amine unit (A2) formed by reaction (a.7) and the tertiary amide branch point (L2) formed in reaction (a.8). Reaction (a.8) is a reversible polycondensation reaction that leads to branch formation. Reactions (a.9) and (a.10) show a two-step branching process in which a Schiff base branch point (SB3) is formed. Reaction (a.11) shows the production of an intermediate species, which results in the formation of pyridine 244 ends. Note that Schaffer et al.[4] used GC-MS (gas

chromatography combined with mass spectrometry) to detect hydrolysis products that are consistent with SB3 branches and with Pyridine 244 ends in the polymer. Unfortunately, no measured concentration values for these species are available for parameter estimation. Rate expressions for the pertinent reactions are shown in Table 2.5. All reactions are assumed to be elementary. Reactivities of functional groups are assumed to be independent of the length of the molecule to which they are attached, in accordance with Flory's[32] equal reactivity hypothesis.

**Table 2.5 Reaction rate expressions for the proposed model**

| | |
|---|---|
| $R_1 = k_c \left( [C][A] - \frac{[L][W]}{K_a} \right)$ | (b.1) |
| $R_2 = k_2 \left( [C] - \frac{[SE][W]}{K_{2eq}} \right)$ | (b.2) |
| $R_3 = k_3 [L]$ | (b.3) |
| $R_4 = k_4 [SE]$ | (b.4) |
| $R_6 = k_6 \left( [SB1][W] - \frac{[A][CPK]}{K_{6eq}} \right)$ | (b.5) |
| $R_7 = k_7 [A]^2$ | (b.6) |
| $R_8 = k_c \left( [C][A2] - \frac{[L2][W]}{K_a} \right)$ | (b.7) |
| $R_9 = k_9 [SB1][A]$ | (b.8) |
| $R_{10} = k_{10} [SB2][A]$ | (b.9) |
| $R_{11} = k_{11} [SB1][CPK]$ | (b.10) |

### 2.3.2 Simplifying Assumptions

In the development of the model equations in Tables 2.6 and 2.7, we assume that the solubilities of $NH_3$ and $CO_2$ are negligible in the molten polymer, and that each species appears in the vapour phase as soon as it is produced. CPK, however, is able to accumulate in the liquid phase where it can either be consumed in reactions (a.6.R) and (a.11) or it can diffuse into the gas phase.

**Table 2.6 Differential equations for melt-phase species concentration changes with time and equations for evolution rates of degradation products for nylon 66**

| | |
|---|---|
| $\frac{d[A]}{dt} = -2R_7 - R_1$ | (c.1) |
| $\frac{d[C]}{dt} = -R_2 - R_8 - R_1$ | (c.2) |
| $\frac{d[L]}{dt} = R_1 - 2R_3$ | (c.3) |
| $\frac{d[W]}{dt} = R_2 - R_6 + R_8 + R_1 + R_{11} - R_{m,W}$ | (c.4) |
| $\frac{d[SE]}{dt} = R_3 + R_2 - R_4$ | ((c.5) |
| $\frac{d[CPK]}{dt} = R_6 - R_{11} - R_{m,CPK}$ | (c.6) |
| $\frac{d[SB1]}{dt} = -R_6 - R_9 - R_{10} + R_4$ | (c.7) |
| $\frac{d[SB2]}{dt} = R_9 - R_{10}$ | (c.8) |
| $\frac{d[SB3]}{dt} = R_{10}$ | (c.9) |
| $\frac{d[A2]}{dt} = R_7 - R_8$ | (c.10) |
| $\frac{d[L2]}{dt} = R_8$ | (c.11) |
| $\frac{d[P244end]}{dt} = R_{11}$ | (c.12) |

Mass-transfer expressions (d.9) and (d.10) in Table 2.7 are used to account for transfer of both water and CPK from the molten polymer to the gas phase. In Equation (d.3) and (d.4) the equilibrium water concentration in the polymer $[W]_{eq}$ that would be in equilibrium with the gas phase, is computed from the water partial pressure using a correlation based on Flory-Huggins theory[34]. Thermal degradation of the polymer is assumed to have no effect on the equilibrium melt-phase water solubility. Since the gas bubbles that are fed to the reactor contain only water and nitrogen, the concentration of CPK in the gas phase is small. As a result, we assume that $[CPK]_{eq}$ in Equation (d.10 ) is zero.

Measurement of the amine end-group concentration [A] was performed by titration. However, since the Schiff base species SB1 and the secondary amine species A2 are basic groups that may also react with acid, we assume that $[A]_{meas}$, the value of the amine end-group concentration determined by titration, is actually the sum [A]+[SB1]+[A2]+[P244]. This assumption is supported by the work of Reimschuessel and Dege[35], Nissen et al.[36] and Curran and Siggia.[37] We assume that species SB2 and SB3 that appear along polymer chains, rather than at the ends, do not influence the titration results, so [SB2] and [SB3] are not included in the expression for $[A]_{meas}$. We further assume that SE groups are not titrated in either the acidic or basic end-group analyses on the polymer samples. As a result the concentrations [SB2], [SB3] and [SE] are predicted by the model, but do not directly influence any of the measurements used for parameter fitting. The measured value of the concentration of BHMT in the hydrolyzed polymer, [BHMT], is assumed to be equal to the sum [L2]+[A2], because both of these species are expected to produce BHMT when the polymer is hydrolyzed.

**Table 2.7 Algebraic expressions required in the model**

| Equation | Source | Equation Number |
|---|---|---|
| $\tau = 1\text{-}\frac{T}{T_C}$ | Poling et al.[33] | (d.1) |
| $\ln\frac{P_w^{sat}}{P_c} = \frac{-7.77224\tau + 1.45684\,\tau^{1.5} - 2.71942\,\tau^3 - 1.41336\,\tau^6}{1-\tau}$ | Poling et al.[33] | (d.2) |
| $\gamma_w = \exp\left[6.390 - \frac{2258}{T}\right]$ | Schaffer et al.[34] | (d.3) |
| $[W]_{eq} = 5.55\times10^4\left(\frac{P_w}{\gamma_w P_w^{sat}}\right)$ | Schaffer et al.[34] | (d.4) |
| $K_a = K_{ar}\left(\frac{1+a\sqrt{[W]_{eq}}}{\gamma_W/\gamma_{Wr}}\right)\exp\left[-\frac{\Delta H}{R}\left(\frac{1}{T}-\frac{1}{T_r}\right)\right]$ | Varziri et al.[13] | (d.5) |
| $k_c = k_{cr}\exp\left[-\frac{E_C}{R}\left(\frac{1}{T}-\frac{1}{T_r}\right)\right]$ | Zheng et al.[6] | (d.6) |
| $K_{6eq} = k_{6eqo}\exp\left[-\frac{\Delta H_6}{R}\left(\frac{1}{T}-\frac{1}{T_o}\right)\right]$ | Zheng et al.[6] | (d.7) |
| $k_i = k_{io}\exp\left[-\frac{E_i}{R}\left(\frac{1}{T}-\frac{1}{T_o}\right)\right]$ where i =2 to 6 and 8 to 10 | Schaffer et al.[9] | (d.8) |
| $R_{m,W} = k_{L,W}a\ ([W]-[W]_{eq})$ | Schaffer et al.[5] | (d.9) |
| $R_{m,CPK} = k_{L,CPK}a\ ([CPK]-[CPK]_{eq})$ | | (d.10) |
| $N_{CO_2} = R_4$ | | (d.11) |
| $N_{NH_3} = R_9 + R_{10} + R_7$ | | (d.12) |

The activation energy for secondary amide formation in reaction (a.9) is assumed to be the same as that for tertiary amide formation in reaction (a.10). This approximation has

been made to reduce the number of model parameters. The melt pool in the reactor is assumed to be well-mixed at all times. Note that no polymer samples were taken during the first 30 minutes of mixing in any experimental run, to ensure that this assumption would be reasonable.

### 2.3.3 Kinetic Model

Table 2.6 provides the 14 ordinary differential equations (one for the melt-phase concentration of each species tracked). Algebraic equations for evolution rates of $CO_2$ and $NH_3$ are provided in Table 2.7, along with other expressions required to solve the model equations. The reference temperature used in the Arrhenius expressions ((d.7) and (d.8)) was chosen to be $T_o$=558.15 K, which is the average temperature for the six experimental runs in Table 2.4. Note that a different reference temperature $T_r$ used in Equation (d.5) and (d.6) was set at 549.15 K, which was the average temperature that Schaffer and Zheng used in their nylon 612 polycondensation studies.[5,6]

### 2.3.4 Initial Conditions

In each experimental run, measured values of [A] and [C] for the first samples were different than the corresponding measurements for the polymer pellets that were initially fed to the reactor. This change indicates that significant thermal degradation and/or polycondensation took place during the heating of the reactor vessel with the polymer held under nitrogen. If only polycondensation and hydrolysis reactions occurred during the heating period, with negligible thermal degradation, the stoichiometry of the system would ensure that the difference between the end-group concentrations, [A]-[C], would be the same as for the initial pellets. Since the value of [A]-[C] obtained for the first sample varied significantly from run to run (i.e., from about –60 to 20 mmolkg$^{-1}$), it is

apparent that degradation during the 3.6 to 4.9 h heat-up period for the various runs should not be neglected.

We have assumed that the measured value of [A] from titration is actually the sum of [A]+[SB1]+[A2]+ [P244], and that SE groups are not detected by titration. Since some thermal degradation and/or polycondensation took place during the heat-up time in each experiment, significant quantities SB1 and A2 may have been formed, and the measured value of [A] by titration would be affected. Since the measured [BHMT] is assumed to be equal to the sum [L2]+[A2], we can calculate the initial values of [L2] using:

$$[L2]_0 = [BHMT]_0 - [A2]_0 \tag{2.1}$$

Moreover, $[L]_o$ can be calculated from a material balance as:

$$[L]_0 = \frac{1}{MW_L}(10^6 - MW_C[C]_0 - MW_A[A]_0 - MW_{SE}[SE]_0 - MW_{SB1}[SB1]_0 - MW_{SB2}[SB2]_0 - MW_{SB3}[SB3]_0 - MW_{A2}[A2]_0) \tag{2.2}$$

Actual initial values of [A], [A2], [SE], [SB1], [SB2], [SB3] and [L2] for each experimental run are unknown, and only the initial value of [C] can be specified with confidence. Some of the unknown initial concentrations were therefore estimated as additional parameters in the model. Efforts were made to estimate all of the unknown initial conditions for [A], [SE], [SB1], [SB2] and [SB3] for each run. The initial value of [C] was set at the first measured value and the initial value for [SB1] was calculated from the initial measurement using:

$$[SB1]_0 = [A]_{meas} - [A]_0 - [P244]_0 - [A2]_0 \tag{2.3}$$

Initial concentrations of pyridine 244 ends and cyclopentanone dissolved in the polymer phase were set to zero for all runs.

## 2.4 Parameter Estimation

In addition to the initial conditions for $[SE]_0$, $[A2]_0$, $[SB2]_0$, $[SB3]_0$ and $[A]_0$ for each experimental run, the model parameters to be estimated are $k_{co}$, $K_{ao}$, $E_c$, $a$, and $\Delta H$ , which are common to nylon 66 and nylon 612 polymerization, and the following 21 parameters, which are related only to nylon 66 degradation: $k_{2o}$, $E_2$, $K_{2eqo}$, $\Delta H_2$, $k_{3o}$, $E_3$, $k_{4o}$, $E_4$, $k_{6o}$, $E_6$, $K_{6eqo}$, $\Delta H_6$, $k_{7o}$, $E_7$, $k_{9o}$, $k_{10o}$, $E_9$, $k_{11o}$, $E_{11}$, $(k_La)_w/(k_La)_{CPK}$ and $(k_La)_w$.

Since we have six experimental runs, the total number of the unknown parameters in the system is 56. Instead of estimating the cyclopentanone mass-transfer coefficient $(k_La)_{CPK}$ directly, we decided to estimate the ratio $(k_La)_w/(k_La)_{CPK}$. Cyclopentanone is a larger molecule than water, and should have a lower diffusivity in molten nylon 66, so it is appropriate to set the lower bound for this ratio at unity.

Estimating the model parameters requires numerical solution of the ODEs in Table 2.6 each time the optimizer selects a new candidate set of parameter values. A fourth-order Runge-Kutta solver (ode45) in MATLAB™ was used to solve the differential equations. The model parameters were estimated using the "lsqnonlin" optimizer in MATLAB™ to find values that minimize a weighted sum of squared errors between the model predictions and the experimental measurements:

$$J = \sum_{j=1}^{n_A}\left[\frac{1}{\pm_A^2}([A]_{meas,i} - [\hat{A}]_i)\right]^2 + \sum_{j=1}^{n_C}\left[\frac{1}{\pm_C^2}([C]_i - [\hat{C}]_i)\right]^2 + \sum_{j=1}^{n_{BHMT}}\left[\frac{1}{\pm_{BHMT}^2}([BHMT]_i - [\hat{BHMT}]_i)\right]^2$$
$$+ \sum_{j=1}^{n_{CO2}}\left[\frac{1}{\pm_{CO_2}^2}(N_{CO_2,i} - \hat{N}_{CO_2,i})\right]^2 + \sum_{j=1}^{n_{NH3}}\left[\frac{1}{\pm_{NH_3}^2}(N_{NH_3,i} - \hat{N}_{NH_3,i})\right]^2 + \sum_{j=1}^{n_{CPK}}\left[\frac{1}{\pm_{CPK}^2}(R_{m,CPK,i} - \hat{R}_{m,CPK,i})\right]^2 \quad (2.4)$$

Weighting values in the objective function were selected to account for uncertainties in different types of measurements, and are provided in Table 2.8.

**Table 2.8 Measurement uncertainties for nylon 66 and nylon 612**

| Uncertainty Symbol | Measured Response | Unit | Value | % of largest measurement |
|---|---|---|---|---|
| $\sigma_A$ | $[A]_{meas}$ | mol $Mg^{-1}$ | 6.732 | 4 |
| $\sigma_C$ | [C] | mol $Mg^{-1}$ | 11.325 | 5 |
| $\sigma_{BHMT}$ | [BHMT] | mol $Mg^{-1}$ | 2.3955 | 15 |
| $\sigma_{CO2}$ | $N_{CO2}$ | mol $Mg^{-1}$ | 3.805 | 5 |
| $\sigma_{NH3}$ | $N_{NH3}$ | mol $Mg^{-1}$ | 4.065 | 15 |
| $\sigma_{CPK}$ | $R_{CPK}$ | mol $Mg^{-1}$ | 1.1 | 10 |

Uncertainties were assumed to be 4% for [A], 5% for [C], 15% for [BHMT], 5% for the $CO_2$ evolution rate, 15% for $NH_3$ evolution rate and 10% for the CPK evolution rate, based on prior knowledge about reproducibility for the different types of measurements. For example, uncertainty of $[A]_{meas}, \sigma_A$, was set at 6.732, which is 0.04(168.3) because 168.3 mmol·$kg^{-1}$ is the maximum value of $[A]_{meas}$.

The model parameters were estimated twice. In a preliminary estimation study, we used the data shown in Figures 2.2 to 2.7 to estimate the 21 nylon 66 degradation parameters and 30 initial concentrations with the five polycondensation parameters fixed at their initial values. Next, we included all 56 parameters using the same nylon 66 degradation data and the nylon 612 polycondensation data.[4,5,6] When simulating the nylon 612 experiments, the reduced set of differential equations in Table 2.9 was used. Results are described below.

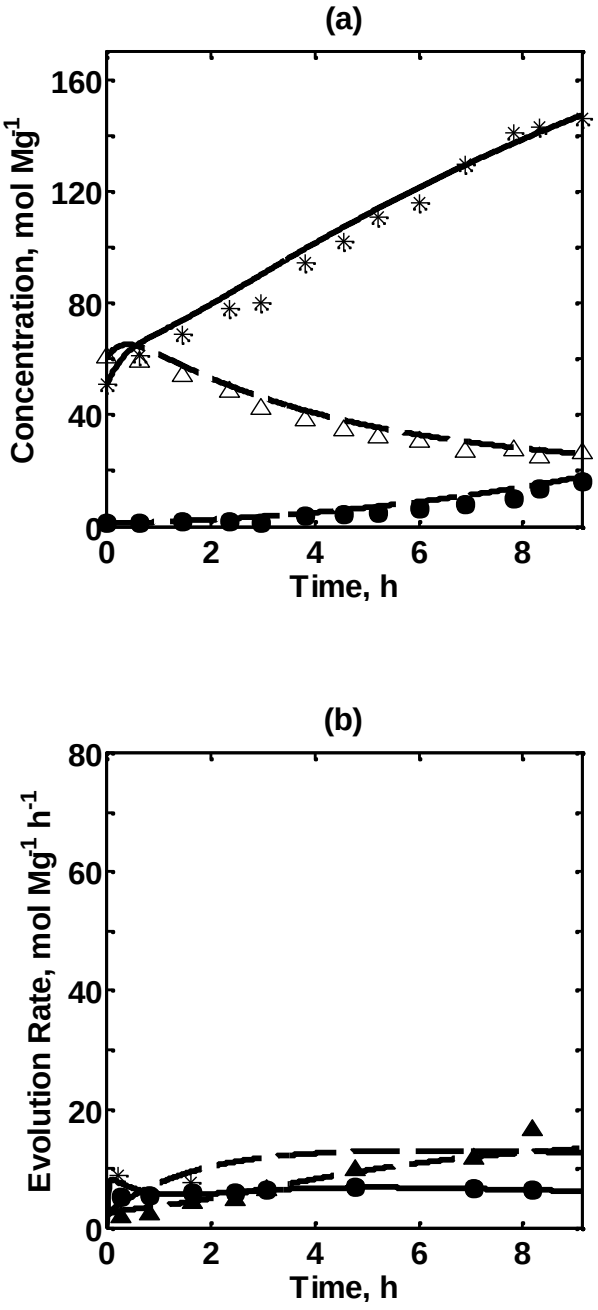



**Figure 2.2 Comparison of model predictions and experimental data for run # 1 at 281˚C using final parameter estimates from Table 2.10; ) Polymer properties: measured $[A]_{tot}$ (*), predicted $[A]_{tot}$ (—), measured [C] (▲), predicted [C] (-·-), measured [BHMT] in hydrolyzed polymer (●), predicted [BHMT] (—·—·); b) Gas evolution rates: measured $CO_2$ (*), predicted $CO_2$ (—·—·), measured $NH_3$ (▲), predicted $NH_3$ (-·-), measured CPK (●), predicted CPK (—)**

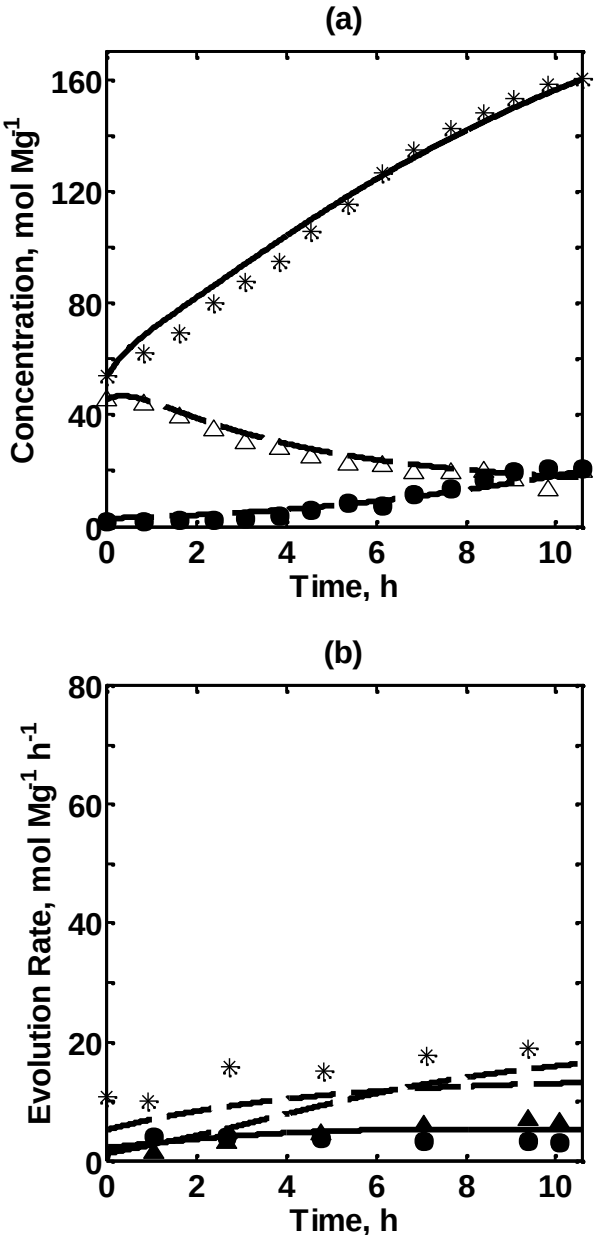



**Figure 2.3 Comparison of model predictions and experimental data for run # 2 at 275˚C using final parameter estimates from Table 2.10; a) Polymer properties: measured $[A]_{tot}$ (*), predicted $[A]_{tot}$ (—), measured [C] (▲), predicted [C] (-·-), measured [BHMT] in hydrolyzed polymer (●), predicted [BHMT] (—·—·); b) Gas evolution rates: measured $CO_2$ (*), predicted $CO_2$ (—·—·), measured $NH_3$ (▲), predicted $NH_3$ (-·-), measured CPK (●), predicted CPK (—)**

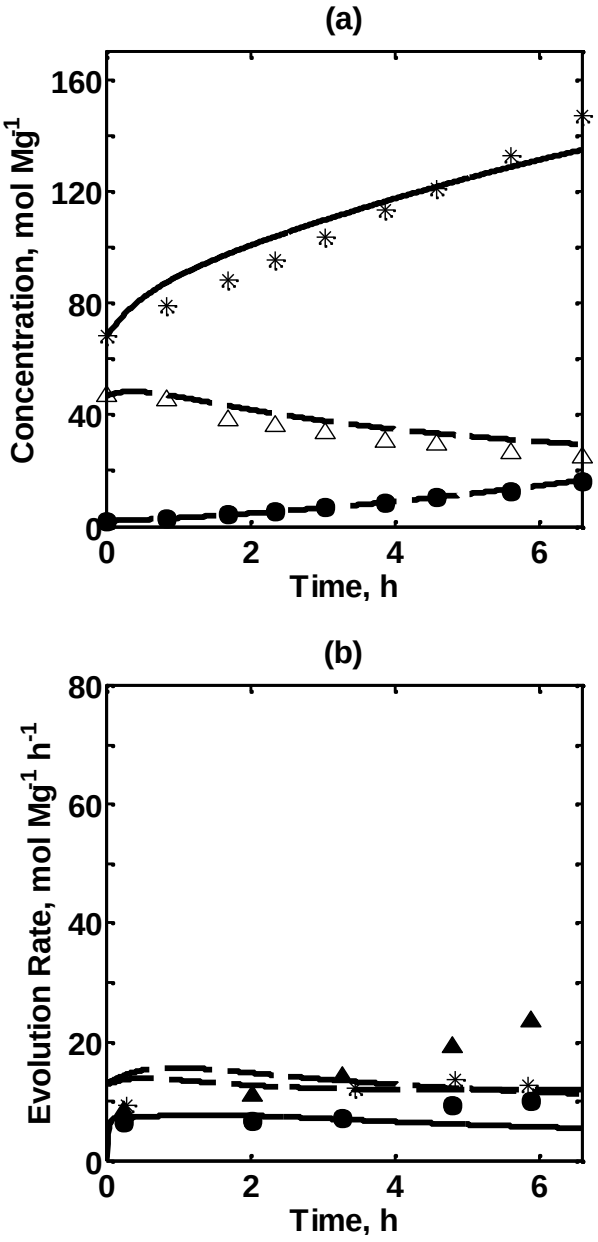



**Figure 2.4 . Comparison of model predictions and experimental data for run # 3 at 285˚C using final parameter estimates from Table 2.10; a) Polymer properties: measured $[A]_{tot}$ (*), predicted $[A]_{tot}$ (—), measured [C] (▲), predicted [C] (-·-·), measured [BHMT] in hydrolyzed polymer (●), predicted [BHMT] (—·—·); b) Gas evolution rates: measured $CO_2$ (*), predicted $CO_2$ (—·—·), measured $NH_3$ (▲), predicted $NH_3$ (-·-·), measured CPK (●), predicted CPK (—)**

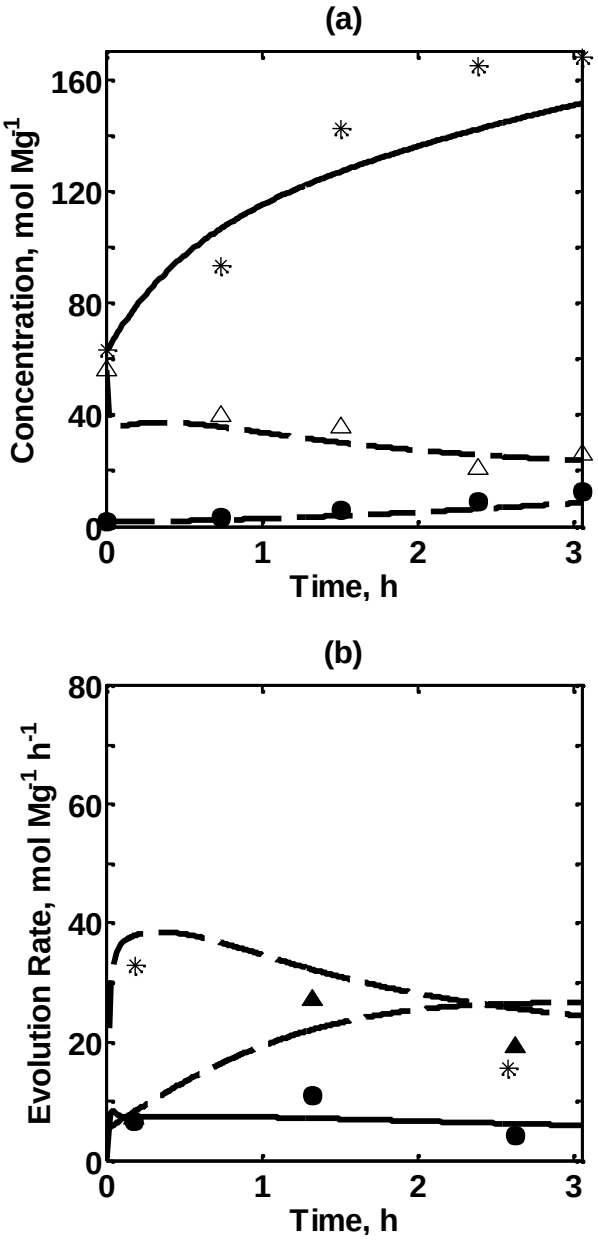



**Figure 2.5 Comparison of model predictions and experimental data for run # 4 at 292˚C using final parameter estimates from Table 2.10; a) Polymer properties: measured $[A]_{tot}$ (*), predicted $[A]_{tot}$ (—), measured [C] (▲), predicted [C] (-·-), measured [BHMT] in hydrolyzed polymer (●), predicted [BHMT] (—·—·); b) Gas evolution rates: measured $CO_2$ (*), predicted $CO_2$ (—·—·), measured $NH_3$ (▲), predicted $NH_3$ (-·-), measured CPK (●), predicted CPK (—)**

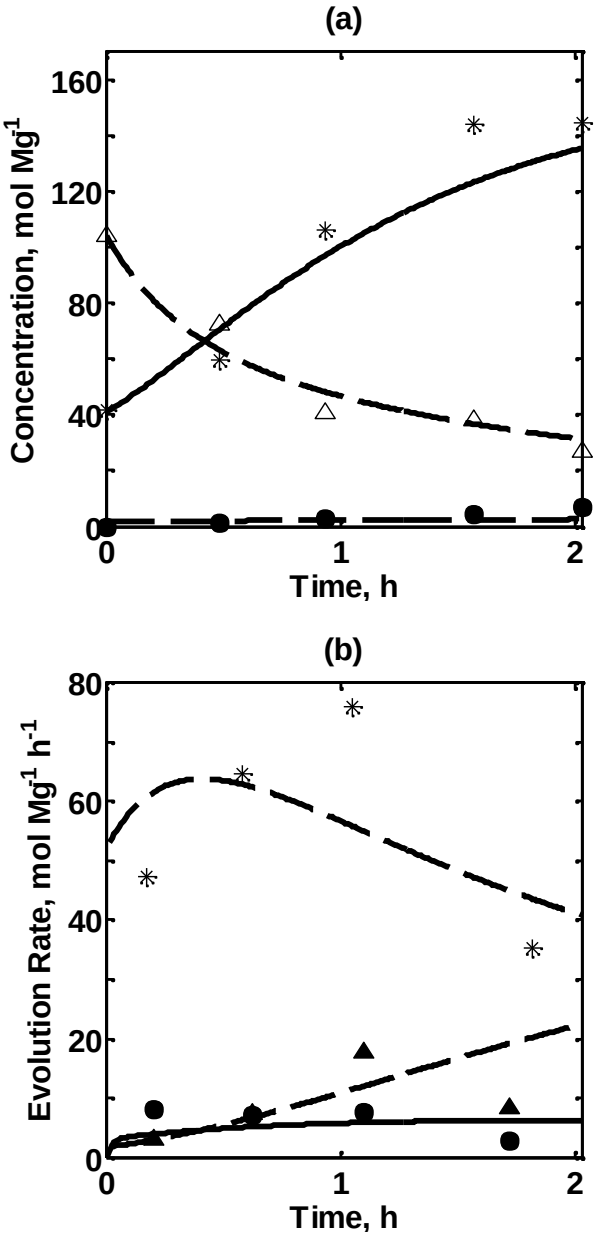



**Figure 2.6 Comparison of model predictions and experimental data for run # 5 at 286˚C using final parameter estimates from Table 2; a) Polymer properties: measured $[A]_{tot}$ (*), predicted $[A]_{tot}$ (—), measured [C] (▲), predicted [C] (-·-), measured [BHMT] in hydrolyzed polymer (●), predicted [BHMT] (—·—·); b) Gas evolution rates: measured $CO_2$ (*), predicted $CO_2$ (—·—·), measured $NH_3$ (▲), predicted $NH_3$ (-·-), measured CPK (●), predicted CPK (—)**

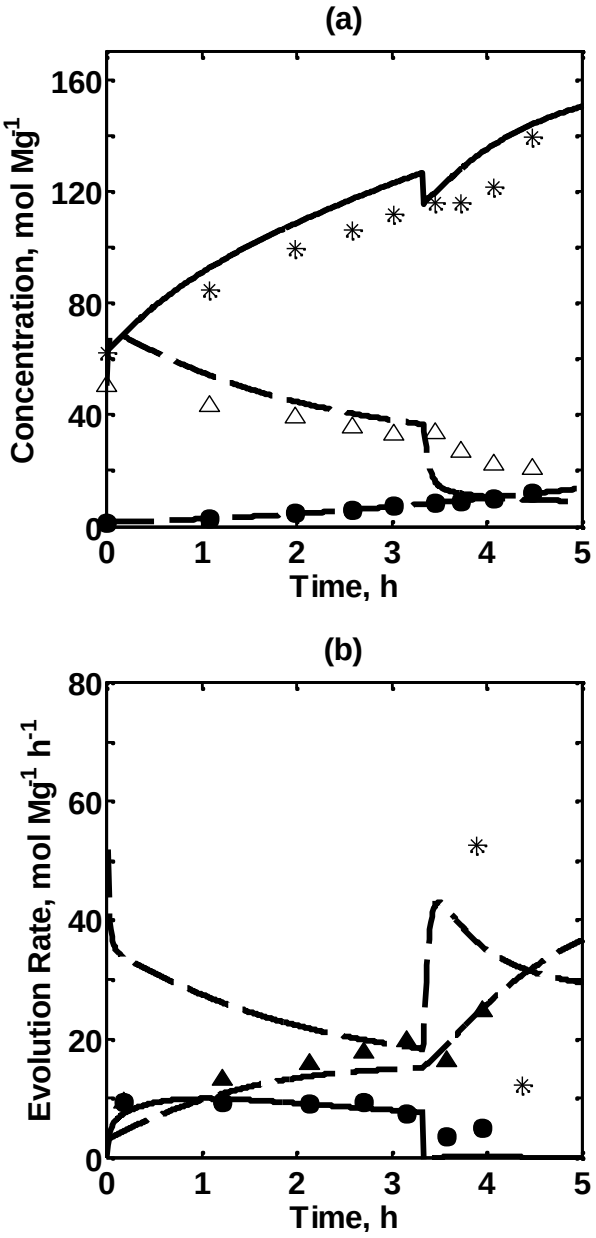



**Figure 2.7 Comparison of model predictions and experimental data for run # 6 at 290˚C using final parameter estimates from Table 2.10; a) Polymer properties: measured $[A]_{tot}$ (*), predicted $[A]_{tot}$ (—), measured [C] (▲), predicted [C] (- - -), measured [BHMT] in hydrolyzed polymer (●), predicted [BHMT] (—·—·); b) Gas evolution rates: measured $CO_2$ (*), predicted $CO_2$ (—·—·), measured $NH_3$ (▲), predicted $NH_3$ (- - -), measured CPK (●), predicted CPK (—)**

**Table 2.9 Differential equations for melt-phase species concentration changes with time and equations for evolution rates of degradation products for nylon 612**

| | |
|---|---|
| $\frac{d[A]}{dt} = -2R_7 - R_1$ | (e.1) |
| $\frac{d[C]}{dt} = -R_8 - R_1$ | (e.2) |
| $\frac{d[L]}{dt} = R_1$ | (e.3) |
| $\frac{d[W]}{dt} = R_8 + R_1 - R_{m,W}$ | (e.4) |
| $\frac{d[A2]}{dt} = R_7 - R_8$ | (e.5) |
| $\frac{d[L2]}{dt} = R_8$ | (e.6) |
| $N_{NH_3} = R_7$ | (e.7) |

### 2.4.1 Initial Parameter Guesses

The initial guesses in Table 2.10 were used to estimate the parameters. Varziri's estimation results for nylon 612 polycondensation were used as initial guesses for the polycondensation parameters ($k_{co}$, $K_{ao}$, $a$, $E_{Co}$ and $\Delta H$).[13] Initial guesses for $k_{2o}$, $E_2$, $k_{3o}$, $E_3$, $k_{4o}$, $E_4$, $k_{6o}$, $K_{6eq}$, $\Delta H_6$, $k_{7o}$, $E_7$, $k_{9o}$, $k_{10o}$, $E_9$, $(k_La)_w/(k_La)_{CPK}$ and $(k_La)_w$ were obtained from the prior modeling work of Schaffer,[9] which neglected the reactions in Table 2.3. A mass-transfer coefficient for water in molten nylon 612 obtained using the same reactor system (Schaffer et al.[5]) was used as the initial guess for $(k_La)_w$. The initial value of $(k_La)_w/(k_La)_{CPK}$ was set at 1.0. Guesses for the unknown initial concentrations of all degraded ends were set to zero because we anticipated only a small level of degradation at the beginning of our experiments.

**Table 2.10 Estimability ranking for preliminary and final studies, initial guesses, bounds, parameter estimates and confidence intervals for the final study**

| Final Rank | Parameter Ranked List | Rank for Preliminary Study | Units | Initial Guess | Lower Bound | Upper Bound | Final Estimate & Approximate 95% Confidence Intervals |
|---|---|---|---|---|---|---|---|
| 1 | $k_{11o}$ | 1 | $Mg \cdot mol^{-1}\ h^{-1}$ | $1.00\times10^{-2}$ | 0 | $500.00\times10^{-2}$ | $58.18\times10^{-2}\pm514.46\times10^{-2}$ |
| 2 | $k_{4o}$ | 2 | $h^{-1}$ | 2 | 0 | 500 | 0.67±0.42 |
| 3 | $k_{7o}$ | 4 | $Mg \cdot mol^{-1}\ h^{-1}$ | $4.48\times10^{-4}$ | 0 | $500.00\times10^{-4}$ | $2.47\times10^{-4}\pm0.42\times10^{-4}$ |
| 4 | $k_{2o}$ | 3 | $h^{-1}$ | 1.58 | 0 | 500 | 0.90±0.79 |
| 5 | $k_{3o}$ | 5 | $h^{-1}$ | $1.86\times10^{-3}$ | 0 | $500.00\times10^{-3}$ | $2.29\times10^{-3}\pm0.62\times10^{-3}$ |
| 6 | $k_{6o}$ | 6 | $Mg \cdot mol^{-1}\ h^{-1}$ | 4.32 | 0 | 500 | 10.39±4.09 |
| 7 | $[SE]_{0run,1}$ | 7 | $mol \cdot Mg^{-1}$ | 0 | 0 | 60.5 | 5.85±18.85 |
| 8 | $k_{9o}$ | 8 | $Mg \cdot mol^{-1}\ h^{-1}$ | $3.82\times10^{-3}$ | 0 | $500.00\times10^{-3}$ | $4.11\times10^{-3}\pm1.33\times10^{-3}$ |
| 9 | $[SE]_{0run,5}$ | 9 | $mol \cdot Mg^{-1}$ | 0 | 0 | 104 | 68.55±55.26 |
| 10 | $[SE]_{0run,2}$ | 10 | $mol \cdot Mg^{-1}$ | 0 | 0 | 45.5 | 25.28±47.13 |
| 11 | $k_{10o}$ | 11 | $Mg \cdot mol^{-1}\ h^{-1}$ | $3.82\times10^{-3}$ | 0 | $500.00\times10^{-3}$ | $10.20\times10^{-3}\pm8.67\times10^{-3}$ |
| 12 | $[A2]_{0run,3}$ | 12 | $mol \cdot Mg^{-1}$ | 0 | 0 | 68.4 | 1.97±1.85 |
| 13 | $[SE]_{0run,4}$ | 13 | $mol \cdot Mg^{-1}$ | 0 | 0 | 56 | 0.00±34.99 |
| 14 | $[A2]_{0run,1}$ | 14 | $mol \cdot Mg^{-1}$ | 0 | 0 | 51.2 | 0.00±82.23 |
| 15 | $[A2]_{0run,2}$ | 15 | $mol \cdot Mg^{-1}$ | 0 | 0 | 54 | 2.95±1.73 |
| 16 | $[SE]_{0run,3}$ | 16 | $mol \cdot Mg^{-1}$ | 0 | 0 | 46.7 | 18.93±23.23 |
| 17 | $[A2]_{0run,5}$ | 17 | $mol \cdot Mg^{-1}$ | 0 | 0 | 41.5 | 1.97±2.05 |
| 18 | $E_7$ | 18 | $kJ\ mol^{-1}$ | 195.66 | 0 | 2092.5 | 119.53±17.75 |
| 19 | $K_{2eqo}$ | 19 | dimensionless | 10 | 0 | 500 | 36.52±22.55 |
| 20 | $E_6$ | 20 | $kJ\ mol^{-1}$ | 41.85 | 0 | 2092.5 | 109.36±51.34 |
| 21 | $E_4$ | 21 | $kJ\ mol^{-1}$ | 92.39 | 0 | 2092.5 | 306.54±79.86 |
| 22 | $K_{ao}$ | | dimensionless | 36.61 | 13.62 | 91.93 | 20.53±20.98 |
| 23 | $[A2]_{0run,3}$ | 22 | $mol \cdot Mg^{-1}$ | 0 | 0 | 68.4 | 16.65±14.3 |
| 24 | $E_9=E_{10}$ | 23 | $kJ\ mol^{-1}$ | 187.28 | 0 | 2092.5 | 71.07±38.14 |
| 25 | $E_{3o}$ | 24 | $kJ\ mol^{-1}$ | 110.45 | 0 | 2092.5 | 0.04±21.37 |
| 26 | $K_{6eqo}$ | 25 | dimensionless | 10 | 0 | 500 | 0.37±3.38 |
| 27 | $(K_La)_w$ | 26 | $h^{-1}$ | 24.3 | 0 | 324 | 46.74±109.02 |
| 28 | $[A]_{0run,3}$ | 27 | $mol \cdot Mg^{-1}$ | 68.4 | 34.2 | 68.4 | 55.17±13.75 |
| 29 | $[A]_{0run,4}$ | 28 | $mol \cdot Mg^{-1}$ | 63 | 31.5 | 63 | 31.51±23.92 |
| 30 | $k_{co}$ | | $kJ \cdot mol^{-1}$ | 0.013 | 0.005 | 0.03 | 0.023±0.01 |
| 31 | $[A]_{0run,2}$ | 29 | $mol \cdot Mg^{-1}$ | 54 | 27 | 54 | 44.96±12.53 |
| 32 | $[A]_{0run,1}$ | 31 | $mol \cdot Mg^{-1}$ | 51.2 | 25.6 | 51.2 | 36.74±80.60 |
| 33 | $[A]_{0run,6}$ | 30 | $mol \cdot Mg^{-1}$ | 62.4 | 31.2 | 62.4 | 54.94±22.44 |

**Table 2.10. Continued, Estimability ranking for preliminary and final studies, initial guesses, bounds, parameter estimates and confidence intervals for the final study**

| Final Rank | Parameter Ranked List | Rank for Preliminary Study | Unit | Initial Guess | Lower Bound | Upper Bound | Final Estimate & Approximate 95% Confidence Intervals |
|---|---|---|---|---|---|---|---|
| 34 | $E_2$ | 32 | kJ $mol^{-1}$ | 110.45 | 0 | 2092.5 | 1833.53±275.92 |
| 35 | $[SB3]_{0run,3}$ | 33 | $mol \cdot Mg^{-1}$ | 0 | 0 | 68.4 | 60.24±845.53 |
| 36 | $[A]_{0run,5}$ | 34 | $mol \cdot Mg^{-1}$ | 41.5 | 20.75 | 41.5 | 20.75±13.39 |
| 37 | a | | $mol^{-0.5} \cdot Mg^{0.5}$ | 0.286 | 0.085 | 0.804 | 0.749±0.93 |
| 38 | $\Delta H_2$ | 35 | kJ $mol^{-1}$ | 41.85 | -209.25 | 837 | -124.39±92.86 |
| 39 | $(K_La)_w/(K_La)$ | 36 | dimensionless | 1 | 1 | 10 | 1.11±10.84 |
| 40 | $E_c$ | | kJ $mol^{-1}$ | 45.9 | 0 | 156.6 | 62.65±122.45 |
| 41 | $E_{11}$ | 37 | $kJ \cdot mol^{-1}$ | 41.85 | 0 | 2092.5 | 398.01±85.10 |
| 42 | $\Delta H$ | | $kJ \cdot mol^{-1}$ | 41.85 | -209.25 | 837 | -20.97±18.41 |
| 43 | $[SE]_{0run,6}$ | 38 | $mol \cdot Mg^{-1}$ | 0 | 0 | 50.5 | 50.47±38.07 |
| 44 | $[SB2]_{0run,1}$ | 39 | $mol \cdot Mg^{-1}$ | 0 | 0 | 51.2 | |
| 45 | $[SB2]_{0run,2}$ | 40 | $mol \cdot Mg^{-1}$ | 0 | 0 | 54 | |
| 46 | $[SB2]_{0run,4}$ | 41 | $mol \cdot Mg^{-1}$ | 0 | 0 | 63 | |
| 47 | $[SB2]_{0run,5}$ | 42 | $mol \cdot Mg^{-1}$ | 0 | 0 | 41.5 | |
| 48 | $[SB2]_{0run,6}$ | 43 | $mol \cdot Mg^{-1}$ | 0 | 0 | 62.4 | |
| 49 | $[A2]_{0run,4}$ | 44 | $mol \cdot Mg^{-1}$ | 0 | 0 | 63 | |
| 50 | $[A2]_{0run,6}$ | 45 | $mol \cdot Mg^{-1}$ | 0 | 0 | 62.4 | |
| 51 | $\Delta H_6$ | 46 | $kJ \cdot mol^{-1}$ | 41.85 | -209.25 | 837 | |
| 52 | $[SB3]_{0run,1}$ | 47 | $mol \cdot Mg^{-1}$ | 0 | 0 | 51.2 | |
| 53 | $[SB3]_{0run,2}$ | 48 | $mol \cdot Mg^{-1}$ | 0 | 0 | 54 | |
| 54 | $[SB3]_{0run,4}$ | 49 | $mol \cdot Mg^{-1}$ | 0 | 0 | 63 | |
| 55 | $[SB3]_{0run,5}$ | 50 | $mol \cdot Mg^{-1}$ | 0 | 0 | 41.5 | |
| 56 | $[SB3]_{0run,6}$ | 51 | $mol \cdot Mg^{-1}$ | 0 | 0 | 62.4 | |

Note that the initial concentration of carboxyl ends was not estimated, but was set at the first measured value. Similarly, the initial concentration of amine ends is calculated from the first value of $[A]_{meas}$ and the initial guesses for the other basic groups that influence that titration measurement. Order-of-magnitude guesses are shown for all other

parameters ($K_{2eq}$, $\Delta H_2$, $E_6$, $K_{6eq}$, $\Delta H_6$, $k_{11o}$ and $E_{11}$) because no published values were available.

### 2.4.2 Bounds for Parameter Estimates

Table 2.10 also provides lower and upper bounds that were used during parameter estimation. Lower bounds for all parameters were set to zero, except for the polycondensation parameters. For these parameters, the lower and upper bounds in Table 2.10 were obtained using approximate 95% confidence intervals reported by Varziri et al.[13] We set the lower and upper bounds at the initial value ± twice the half-width of the corresponding 95% confidence interval to ensure that parameters remained within a reasonable range. We selected twice the half-width because Variziri's confidence intervals were obtained using linearization and might be overly conservative. The upper and lower bounds for $(k_La)_w$ were obtained based on values from the literature.[38,39] Upper bounds for the remaining parameters and initial guesses were selected based on our judgment about physically reasonable values.

## 2.5 Results and Discussion

### 2.5.1 Simplifying the Model using Estimabilty Analysis

Estimability analysis was used to rank the parameters from most estimable to least estimable, as shown in Table 2.10. The results in the first two columns of Table 2.10 are from our final parameter estimation study (obtained using the settings for nylon-66 degradation and nylon-612 polycondensation together). Ranking results in the third column are from the preliminary study (obtained using nylon-66 experimental settings alone). The rank of the various parameters depends on the influence of each parameter on the predicted model outputs, correlation between the effects of parameters, and scaling

factors used to reflect uncertainties in initial parameter values. When estimability analysis was performed using the nylon 66 experimental settings alone (preliminary study) 38 of the 51 parameters could be ranked. Attempts to rank additional parameters resulted in numerical problems. When estimability analysis was performed using all of the experimental settings (final study) 44 of the 56 parameters could be ranked before numerical problems were encountered. Wu's MSE-based criterion[21,19] was used to determine that estimating the top 38 parameters should give the best model predictions for the preliminary study and the top 43 parameters should be estimated using all of the data. The parameters that were *not* selected for estimation are the initial [A2] for runs 4 and 6, initial [SB3] for all runs except run 3, initial [SB2] for all runs except run 3 and $\Delta H_6$. These inestimable parameters were left at their initial guesses shown in the fifth column of Table 2.10.

$k_{11o}$ was selected as the most estimable parameter because it involves degradation products CPK and SB1 (see reaction (a.11)) and this parameter had a large initial uncertainty. The next three parameters in the ranked list are kinetic parameters $k_{4o,}$ $k_{7o}$, and $k_{2o}$, which also had large uncertainty values. These parameters influence the rates of production of important degradation products (i.e., $CO_2$, $NH_3$, BHMT and SE). The estimability ranking algorithm tends to rank parameters near the top of the list when they have a large influence on model predictions and when their initial values are not well known.[18] The initial concentrations for stabilized ends $[SE]_0$ for the various runs are among top parameters in the ranked list because they influence the rates of $CO_2$ and, subsequently, $NH_3$ production (see reaction (a.4)). Although activation energies for the various reactions were assigned large uncertainties, they appear further down the list

because their influence on the model predictions is not as high compared to the influence of the rate constant values at the reference temperature. Polycondensation parameters ($k_{co}$, $k_{ao}$, $a$, $E_c$ and $\Delta H$) appear further down the list because their initial uncertainties were relatively small, since we had prior information about their values from previous nylon 612 parameter estimation studies.[6,13]

Results of the parameter ranking and selection methods that were used depend on the initial parameters values and scaling factors employed.[20,21] To test the robustness of the ranking results in Table 2.10, we investigated the effects of changing the initial guesses and scaling factors. The ranked list changed slightly using different sets of reasonable initial conditions (for example changing initial guesses randomly up and down by 10%). However, the final parameters selected for estimation did not change by changing the initial conditions and scaling factors over a range of reasonable values.

### 2.5.2 Parameter Estimation Results

Figure 2.8 shows the influence of the number of parameters estimated from the ranked list on the objective function values from the preliminary and final parameter estimation studies. As expected, estimating more parameters from the ranked list resulted in a better fit to the available data in both studies. Note that the final objective function value of J= 645 obtained from the final parameter estimation study (using all of the nylon 66 and 612 data) was lower than J=870 obtained from the preliminary study (using only the nylon 66 data), even though the number of terms in the objective function for the full parameter estimation problem was larger (480 data points in the full study and 254 data points in the preliminary study). This result indicates that there was considerable benefit to re-estimating the polycondensation parameters $k_{co}$, $K_{ao}$, $a$, $E_{co}$ and $\Delta H$ because of the additional information about polycondensation kinetics in Schaffer's nylon 66 data set.[4]

Note that the fit to the nylon 612 polycondensation data[6,5] (not shown) using the final parameter estimates in Table 2.10 was similar to the fit using the initial parameter values.[13]

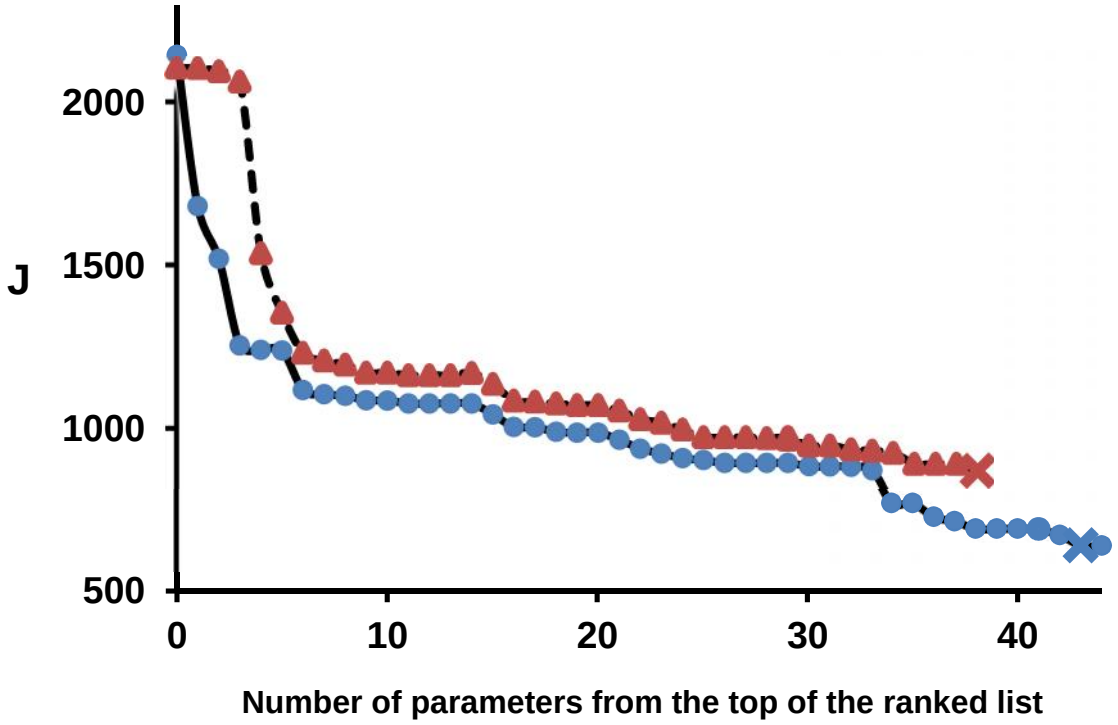



**Figure 2.8 Influence on the number of parameters estimated from the ranked list on the objective function value for preliminary study, (▲), and final parameter estimation study, (●).× indicates the number of parameters selected using Wu's method [19,21] (Wu, 2011)**

As shown in Figure 2.8, in the full study no noticeable improvement was obtained when 44 parameters were estimated instead of 43, which is consistent with the selection of 43 parameters using Wu's method.[19,21] The objective function values in Figure 2.8 were obtained by starting from the initial parameter values in Table 2.10, and estimating the

top-ranked parameter $k_{11o}$, then the top two parameters and so on. The optimization procedure was sensitive to parameter initial guesses; some other sets of initial guesses resulted in convergence to local optima with higher objective function values than those shown in Figure 2.8. Although the final parameter estimates in Table 2.10 give a good fit to the data (see Figures 2.2 to 2.7), these estimation results may also correspond to a local minimum. A comprehensive search of the full parameter space would need to be performed before concluding that the best possible parameter values have been obtained.

Approximate 95% confidence intervals for the true values of the parameters were calculated using linearization methods,[40] and are reported in the final column of Table 2.10. Wide confidence intervals for some parameters were obtained because the data provided little information about these parameters (i.e., the limited number of experimental runs and lack of experimental data for the concentrations [A2], [SE], [SB1], [SB2], [SB3] and [P244]). Also, the experimental data were collected within a relatively narrow temperature range, complicating the estimation of activation energies. Note that the 95% confidence intervals for four of the five polycondensation parameters ($K_{ao}$, $k_{co}$, $E_c$ and $\Delta H$) are narrower than those obtained by Varziri et al.[13] using the nylon 612 data alone, because of the additional information provided by the nylon 66 data. The confidence intervals in Table 2.10 are only approximations due to linearization and because some of the typical least-squares assumptions used to calculate them might not be true. For example, we recognize that our model equations arise from simplifying assumptions, independent variables (such as temperature) are not error-free, and measurement errors within each run may not be independent. Also, the final parameter estimates and confidence intervals in Table 2.10 are conditional on the values of the 8

parameters at the bottom of Table 2.10 that were not estimated. We recommend that additional experiments should be performed if more accurate parameter estimates and model predictions are required. Note that the estimate for equilibrium constant $K_{2eqo}$ is 36.52, which indicates that the equilibrium for reaction (a.2) is far to the right and that the reverse reaction (a.2.R) is not favourable.

The predictions of the proposed model are compared with the experimental data for nylon 66 in Figures 2.2 through 2.7. Note that the same vertical-axis scaling was used for all experimental runs to permit easy comparisons between runs. The model predictions agree well with the majority of the end-group and BHMT branch-point concentration data. Although the fit to evolution rates of gaseous degradation products are not as good as the fit to polymer property data, possibly due to sampling difficulties, the trends of the model predictions for the gasous species are consistent with the data. Note that the poorer fit to these gas evolution data was expected due to their larger measurement uncertainties, which are shown in Table 2.8. In runs 1 to 3, the evolution rates for $CO_2$, $NH_3$ and CPK were relativly low, as predicted. Higher evolution rates were measured and predicted in runs 4 and 6, which were conducted at the highest temperatures (292 and 290 ˚C, respectively) where thermal degradation tends to occur. Run 5 also experienced high gas evolution rates, as predicted, because this experiment was conducted at the lowest water concentration ($P_w$=23 kPa). Reactions (a.2.R) and (a.6) in Table 2.2 show how keeping a desirable concentration of water in the molten nylon can reduce the concentrations of stabilized ends and Schiff base ends that lead to branching. Although, the fit for run 5 in Figure 2.6 is not as good as that for other runs, the model does predict a maximum in the $CO_2$ evolution rate, but the position of the maximum does not match

the model prediction well. Discrepancies between the model predictions and measurements may result from imperfections in the model structure, measurement errors, and fluctuations in reactor operating temperature and sparge-gas composition.

### 2.5.3 Prediction of Gelation Times

One of the potential uses of the proposed model is the prediction of the time required for gelation of nylon 66 due to crosslinking under different operating conditions. Jacobs and Zimmerman[41] state that gelation is predicted to occur when the concentration of trifunctional branch points in the polymer exceeds a critical value of one-third the sum of the end group concentrations. Since there are two types of trifunctional branch points and five types of end groups considered in the present model, gelation would be predicted when [SB3]+[L2] exceeds the value of ([A]+[C]+[SE]+[SB1]+[P244])/3. Experimentally, gelation can be considered to have occurred when the polymer is no longer completely soluble in 90% formic acid. Figure 2.9 shows that when the predicted values of these quantities are plotted versus time for run 3, the predicted gelation time is 6.8 h (the intersection for the two curves), which agrees well 6.9 hours, which is the experimental gelation time for run 3 observed by Schaffer during this experiment.[9] Note that Schaffer observed gelation after 4.9 hours for run 6, and that our model predicts gelation after 4.8 hours (not shown). These results illustrate the utility of the proposed model for predicting gelation using different experimental conditions.

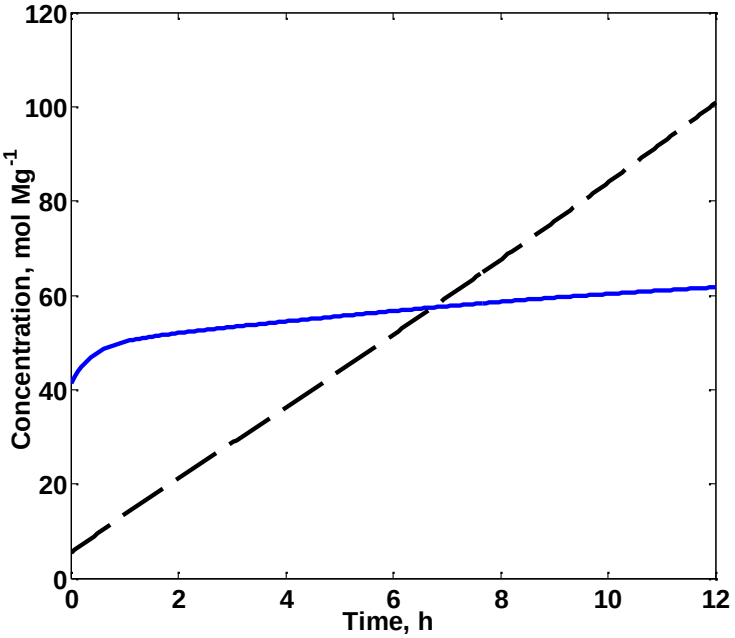



**Figure 2.9 Model prediction of gel formation point for run #3, 285oC, predicted ([SB3]+[L2]) (- - -), predicted ([A]+[C]+[SE]+[SB1]+[P244])/3 (——).**

## 2.6 Conclusion

In this work, an improved kinetic model for thermal degradation of molten nylon 66 was developed. The reactions included in previous models[7,9,10] were augmented with additional reactions that account for the influence of water on degradation rates, the formation of cycopentanone and the formation of Schiff base ends.

Experimental data from dynamic experiments involving nylon 66[4] and nylon 612[5,6] were used to estimate the parameters. Estimability analysis[17,18] and a MSE-criterion[19,21] were used to determine that 43 out of 56 model parameters and unknown initial concentrations should be estimated to provide the best model predictions, given the limited data available for parameter estimation. The model describes the experimental polymer property data well, with typical errors of 6.1% for amine ends, 2.9% for carboxyl ends

and 13.2% for BHMT branch points. Trends in the rates of evolution of gaseous byproducts were predicted reasonably well, with typical errors of 4.3% for $CO_2$, 27.2% for $NH_3$ and 29.4% for cyclopentanone. Although complete mechanistic accuracy in terms of the set of chemical reactions and kinetic rate parameters used in the present model cannot be claimed, the model is a useful tool for the prediction of the rate and extent of thermal degradation under various process conditions of commercial interest. This model can predict degradation rates during the final stages of the nylon 66 production when water concentrations are low (10 to 80 $mmol \cdot kg^{-1}$) and temperatures are high (275 to 292$^{\circ}$C), and can also predict the time required for gelation at different experimental conditions.

## 2.7 References


1. R. E. Putscher, in *Kirk-Othmer Encyclopedia of Chemical Technology*, 3rd Edition, Vol. 18 (Eds. M. Grayson, D. Eckroth), Wiley, New York, USA **1982**, p. 334.

2. M. A. Schaffer, E. K. Marchildon, K. B. McAuley, M. F. Cunningham, *J. Macromol. Sci. Rev. Macromol. Chem. Phys.* **2000**, *C40*, 233.

3. M. A. Schaffer, E. K. Marchildon, K. B. McAuley, M. F. Cunningham, *Chem. Eng. & Tech.* **2001**, *24*, 401.

4. M. A. Schaffer, K. B. McAuley, E. K. Marchildon, M. F. Cunningham, *Macromol. React. Eng.*, **2007**, *1*, 563.

5. M. A. Schaffer, E. K. Marchildon, K. B. McAuley, M. F. Cunningham, *Ind. Eng. Chem. Res*. **2003**, *42*, 2946.

6. W. Zheng, K. B. McAuley, E. K. Marchildon, K. Z. Yao, *Ind. Eng. Chem. Res.* **2005,** *44*, 2675.

7. D. D. Steppan, M. F. Doherty, M. F. Malone, *J. Appl. Polym. Sci*. **1991**, *42*, 1009.

8. F. Wiloth, *Makromol. Chem.* **1971**, *144*, 283.

9. M. A Schaffer, Ph.D. Thesis, Queen's University Canada, September, **2003**.

10. K. Y. Choi and K. B. McAuley, in *Polymer Reaction Engineering*, (Eds. J. M. Asua), Blackwell Publishing, Iowa,USA **2007**, p.273.

11. R. O. Pimentel, R. Giudici, *Ind. Eng. Chem. Res.* **2006**, *45*, 4558.

12. F. K. Mallon, W. H. A Ray, *J. Appl. Polym. Sci*. **1998,** *69*, 1213.

13. M.S. Varziri, K. B. McAuley and P. J. McLellan, *Ind. Eng. Chem. Res.* **2008**, *19*, 7274.

14. W. Zheng, K. B. McAuley, E. K. Marchildon and K. Z. Yao, *Can J Chem Eng.* **2007,** *85*, 180.

15. G. Meacock, *J. Appl. Chem*. **1954**, *4*, 172.

16. B. Kamerbeek, G. H. Kroes, W. Grolle, *Soc. Chem. Ind. Monogr.* **1961**, *13*, 357.

17. Z. K. Yao, K. B. McAuley and E. K. Marchildon, *J. Appl. Polym. Sci*. **2003**, *89*, 3701.

18. D. E. Thompson, K. B. McAuley and P. J. McLellan, *Macromol. React. Eng*. **2009**, *3*, 160.

19. S. Wu, K.B. McAuley and T. J. Harris, *Can. J. Chem. Eng*. **2011**, *89*, 325.

20. S. Wu, K. B. McAuley and T. J. Harris, *Can. J. Chem. Eng.* **2011**, *1*, 148.

21. S. Wu, K.A.P. McLean, T.J. Harris and K.B. McAuley, *Int. J. Advanced Mechatronic Systems*, **2011**, *3*, 188.

22. J. Littlejohns, K. B. McAuley and A. J. Daugulis, *J. Hazard. Mater*. **2010,** *175*, 872.

23. J.E. Puskas, S. Shaikh, K.Z. Yao, K.B. McAuley, G. Kaszas, *Eur. Polym*. J. **2005**, *41*, 1.

24. V.I. Koeva, S. Daneshvar, R.J. Senden, A.H.M. Imam, L.J. Schreiner, K.B. McAuley, *Macromol. Theory Simul.* **2009**, *18*, 495.

25. K.G. Gadkar, J. Varner, F.J. Doyle III, *Syst. Biol*. **2005**, *2*, 17.

26. T. Quaiser, M. Monningmann, *Systems Biol*. **2009**, *3*, 50.

27. B. Kou, K. B. McAuley, C. C. Hsu, D. W. Bacon, K. Z. Yao, *Ind. Eng. Chem. Res*. **2005.** *44*, 2428.

28. B. Kou, K.B. McAuley, J.C.C. Hsu, D.W. Bacon, *Macromol. Mater. Eng.* **2005**, *290*, 537.

29. S. V. Levchik, E. D. Weil, M. Lewin, *Polym. Int.* **1999**, *48*, 532.

30. B. Kamberbeek, G.H. Kroes, W. Grolle, *Society of Chemical Industry*, London, **1961**, 357.

31. A. Bruylants, E. F. Medicis, in *The Chemistry of the Carbon-Nitrogen Double Bond*, (Ed. S. Patai), Wiley and Sons, London, England **1970**, p.468.

32. P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, New York, USA **1953.**

33. B. E. Poling, J. M. Rausnitz. and J. P. OConnell., *The Properties of Gases and Liquids*. Fifth Edition, McGraw-Hill, New York, USA **2001.**

34. M. A. Schaffer, E. K. Marchildon, K. B. McAuley, M. F. Cunningham, *Polym. Eng. Sci.* **2003**, *43*, 639.

35. H. K. Reimschuessel, G. J. Dege, J. Polym. Sci. **1970**, *8*, 3265.

36. D. Nissen, V. Rossbach, H. Zahn, Angew. Chem., Int. Ed. Engl. **1973**, *12*, 602.

37. D. J. Curran, S. Siggia, in *The Chemistry of the Carbon-Nitrogen Double Bond*, (Ed. S. Patai), Wiley, London, England **1970**, p. 156.

38. R. Giudici, C. A. O. Nascimento, I. C. Beiler, N. Scherbakoff, *J. Appl. Polym. Sci*. **1998**, *67*, 1573.

39. K. Nagasubramanian, H. K. Reimschuessel, *J. Appl. Polym. Sci.* **1973**, *17*, 1663.

40. S. G. Nash, A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, New York, USA **1996**.

41. D. B. Jacobs, J. Zimmerman, in *Polymerization Processes*, (Eds. C. E. Schildknecht, I. Skeist), Wiley, New York, USA **1977**, p. 424.

# Chapter 3

# An Approximate Expectation Maximization Algorithm for Estimating Parameters in Nonlinear Dynamic Models with Process Disturbances

## 3.1 Abstract

In chapter 2, a kinetic model that accounts for nylon 66 degradation reactions is developed. The fit to the data for this model have some discrepancies indicating the existence of process disturbances and model mismatch. The goal of the remaining chapters of this thesis is to propose parameter estimation techniques that take into account the process disturbances and model mismatch. Stochastic terms are included in fundamental dynamic models of chemical processes to account for disturbances, input uncertainties and model mismatch. The resulting equations are called stochastic differential equations (SDEs). An Approximate Expectation Maximization (AEM) algorithm using B-splines is developed for estimating parameters in SDE models when the magnitude of the disturbances and model mismatch is unknown. The AEM method is evaluated using a two-state nonlinear CSTR model. The proposed algorithm is compared with two other maximum-likelihood-based methods (CTSM[1,2] and Varziri's extended AMLE[3]). For the CSTR examples studied, the AEM algorithm provides more accurate estimates of model parameters, unknown initial conditions and disturbance intensities. This chapter has been accepted for publication as a journal paper by *Canadian Chemical Engineering Journal*.

## 3.2 Introduction

In chemical processes, external fluctuations or disturbances influence the system inputs.[4,5] Often, simplified fundamental models are used because a detailed fundamental model would be too complex for the intended use (e.g., on-line process monitoring or control) or because it would be too expensive to do all of the experiments required to estimate the large number of kinetic, thermodynamic and transport parameters that would appear in detailed model equations. As a result, modelers develop models that include only the most important phenomena, knowing that simple models sometimes give better predictions than more complicated models[6-8] and that simpler models are more portable and easy to use.

Stochastic terms are added to simplified dynamic models to account for the effects of unmeasured disturbances and model imperfections.[9,10] Stochastic differential equations (SDEs) are differential equations in which the influence of various random disturbances appears explicitly.[5] These disturbances (also called system noise) can enter the model equations nonlinearly, but additive linear disturbances are often used, especially when disturbances arise from multiple sources.[4] Parameter estimation problem in SDE models are usually addressed using probability density functions.[11] There are two main categories for parameter estimation in SDE models based on discrete observations.[12] In the first category, a moment-matching method is used, while in the second category approximate maximum likelihood methods are used. Moment-matching estimators can be difficult to use because they require the calculation of higher-order moments in order to be efficient, and the appropriate order is difficult to determine.[12] Maximum likelihood methods are used widely because of their efficiency. In maximum likelihood estimation

(MLE) methods, the likelihood function of the parameter vector given observed data is maximized to estimate the unknown model and noise variance parameters.[13] In the presence of unmeasured states and missing observations, it is difficult to find a closed form for the likelihood function.[14] Several techniques have been proposed for approximating the likelihood function. Some of these techniques are: simulated maximum likelihood methods (such as Markov Chain Monte Carlo (MCMC) techniques),[14-19] expansion of the likelihood function using Hermite polynomial basis functions,[20,21] solving the Fokker-Planck equation numerically[22-24] and recursive maximum likelihood parameter estimation using polynomial chaos theory.[25] Benefits and drawbacks of these techniques are summarized by Lindstrom.[12] An approximate ML method was proposed by Kristensen et al.[2] In Kristensen et al.'s method, a Gaussian distribution is assumed for the likelihood function and the mean and variance of the likelihood function are estimated using an Extended Kalman Filter (EKF). Easy-to-use CTSM (continuous-time stochastic modeling) software was developed based on Kristensen's method and is used in a simulation study later in this chapter.[1]

Varziri et al. developed an approximate maximum likelihood estimation (AMLE) method for estimating parameters in SDE models when disturbance intensities and noise variance are known. Varziri et al.[3] extended AMLE for estimating disturbance intensities, based on a technique developed by Heald and Stark.[26] Although reasonable results were obtained using this methodology, this technique uses a heuristic outer objective function and may not be consistent with maximum likelihood estimates of the disturbance intensities. It is unclear whether this methodology is robust to imperfect estimates of

noise variances that must be provided by the user. The algorithm developed in the current chapter builds on Varziri's results.

A popular MLE-based algorithm for parameter estimation in SDE models is the expectation maximization (EM) algorithm.[27-29] The EM algorithm is suitable for situations where the likelihood function of the parameters given the measurements does not have a closed form, but the probability density function of the measured data and unmeasured states (i.e., the likelihood of the complete data) does have a closed form.[30,31] In the first step of the EM algorithm (referred to as the Expectation step or E step), the expected value of the likelihood of the complete data is determined using the observations and the most recent estimates of the parameters. In the second step (referred to as the Maximization step or M step), the expected value determined in the first step is maximized with respect to the model and the variance parameters (i.e., the noise variances and the disturbance intensities). These two steps are repeated until parameter convergence is obtained.[32] The two steps of the EM algorithm have explicit recursive solutions for linear systems with Gaussian noise,[27] but they do not have explicit solutions for nonlinear SDEs.[29]

Approximation techniques have been used to find closed-form expressions for the E and M steps of the EM algorithm. For example, extended Kalman filters (EKFs) have been used for approximating both the E and M steps.[33-36] Markov Chain Monte Carlo (MCMC) methods, also known as particle filter methods, have been used for approximating the E step.[15,28,29,37-40] In the MCMC methods, a probability density function is numerically approximated by drawing samples from a target density function.[18,14] MCMC methods are effective parameter estimation tools that do not require assumptions about the form of

the density functions. A drawback of MCMC methods is that they can be computationally expensive for models with a relatively large number of states and parameters.[29,40,41] The complexity of the optimization problem increases rapidly as the number of states, measurements, and parameters increases. Imtiaz et. al.[42] discuss some implementation issues for MCMC methods and Kantas et al.[43] present an overview of a variety of MCMC methods and discuss their advantages and disadvantages. Recently, Chitralekha et al.[44] compared the performance of three EM algorithms: the particle smoother, the unscented Kalman smoother and the extended Kalman smoother. Linearization-based EKF methods for approximating the EM algorithm are beneficial because they do not require MC sampling from probability density functions, but they can give biased parameter estimates in situations where the nonlinearities are strong.[45] Further details about the EM algorithm and its relationship to the AEM algorithm proposed in this chapter are presented in section 2.3.

The goal of this research is to develop an easy-to-use Approximate Expectation Maximization (AEM) method for estimating model parameters and disturbance intensities in SDE models. Some of the benefits of the AEM method are: i) simplicity of implementation, ii) reliable estimates of model parameters, initial conditions and disturbance intensities, iii) efficient handling of unknown initial states, iv)ability to handle unmeasured state variables. First, some notation and background information are presented and a closed form for the E step of the EM algorithm is derived using B-spline approximations for state trajectories. An objective function for the proposed AEM method is determined, and the resulting AEM algorithm is illustrated using a two-state nonlinear CSTR model with unknown disturbance intensities. Results obtained using the

AEM method are compared to results obtained using CTSM[1] and extended AMLE method.[3] It is shown that the proposed AEM method provides more accurate parameter estimates and is less computationally intensive than the alternative methods for the example studied.

## 3.3 Preliminaries

### 3.3.1 Model and Notation

Consider the following Multi-Input Multi-Output (MIMO) model:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\eta}(t) \tag{3.1.a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{3.1.b}$$

$$\mathbf{y}(t_{m\,r,j}) = \mathbf{g}(\mathbf{x}(t_{m\,r,j}), \mathbf{u}(t_{m\,r,j}), \boldsymbol{\theta}) + \boldsymbol{\varepsilon}(t_{m\,r,j}) \tag{3.1.c}$$

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

$$\mathbf{Q} = \begin{bmatrix} Q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_X \end{bmatrix} \tag{4.2}$$

The diagonal elements of $\mathbf{Q}$ are sometimes referred to as disturbance intensities (i.e. $\mathbf{Q_d}=[Q_1,\ldots,Q_X]^T$).[3] $\delta(.)$ is the Dirac delta function, $\mathbf{y}$ is a $Y$-dimensional output vector with $Y \leq X$ . Assume that the $r$th response variable ($r$=1…$Y$) is measured $N_r$ times during the set of dynamic experiments. The set of times at which measurements are available for the $r$th response ($r$=1…$Y$) is denoted by $t_{m\,r,j}$ ($j$ = 1…$N_r$). $\mathbf{g}$ is a $Y$-dimensional vector of

nonlinear mappings and $\boldsymbol{\varepsilon}$ is a $Y$-dimensional vector of zero-mean random variables. Assume that these measurement errors are independent so that their covariance matrix is:

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

(3.3)

Consider the vector $\mathbf{Y_m}$ that contains all of the stacked measured values, where $\mathbf{Y_m} = [y_1(t_{m\,1,1}) \ldots y_1(t_{m\,1,N_1}) \ldots y_Y(t_{m\,Y,1}) \ldots y_Y(t_{m\,Y,N_Y})]^T$ and

$\mathbf{X_m} = [x_1(t_{m\,1,1}) \ldots x_1(t_{m\,1,N_1}) \ldots x_Y(t_{m\,Y,1}) \ldots x_Y(t_{m\,Y,N_Y})]^T$ is a corresponding stacked vector of state values at the measurement times, and $\mathbf{U_m}$ and $\boldsymbol{\varepsilon_m}$ are corresponding stacked vectors for the input variables and random errors:

$$\mathbf{Y_m} = \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon_m} \tag{3.4}$$

A continuous white noise process is defined as the limit of a discrete-time white-noise sequence when the discretization interval $\Delta t$ approaches zero. In simulations of continuous stochastic processes, it is common to implement stochastic disturbances $\boldsymbol{\eta}(t)$ using discrete-time white-noise sequences that are a series of random step functions with a short sampling interval $\Delta t$ and covariance:[46,47]

$$\mathrm{E}\{\boldsymbol{\eta}(j_1 \Delta t)\boldsymbol{\eta}(j_2 \Delta t)\} = \begin{cases} \dfrac{\mathbf{Q}}{\Delta t} & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases} \tag{3.5}$$

where $j_1$ and $j_2$ are positive integers.

Large values of the diagonal elements of $\mathbf{Q}$ correspond to large disturbances or considerable model mismatch.

Three situations can occur for initial conditions of state variables. The initial conditions may be perfectly known, entirely unknown, or they may be measured. In cases where initial conditions are measured, assume that their measured values $\mathbf{x}_{m0}$ were sampled from a normal distribution with $\mathrm{E}\{\mathbf{x}_{m0}\} = \mathbf{x}_0$ and $\mathrm{cov}\{\mathbf{x}_{m0}\} = \mathbf{S}_{m0}$.

Modelers often have knowledge about the accuracy of their measurements (i.e., $\mathsf{p}$ and $\mathbf{S}_{m0}$), but do not have good knowledge about the magnitudes of the imperfections in their material and energy balance models (i.e., values for the disturbance intensities in $\mathbf{Q}$). Let $\boldsymbol{\zeta} = [\boldsymbol{\theta}^{\mathrm{T}}, \mathbf{x}_0^{\mathrm{T}}, \mathbf{Q_d}^{\mathrm{T}}]^{\mathrm{T}}$ be the vector of unknown parameters in the SDE model, which includes the fundamental model parameters $\boldsymbol{\theta}$, along with unknown initial states and the disturbance intensities. In this chapter, a method is proposed for estimating the model parameters $\boldsymbol{\theta}$ and $\mathbf{x}_0$ together with the disturbance intensities in $\mathbf{Q}$ in situations where a good prior estimate of the measurement noise covariances $\mathsf{p}$ and $\mathbf{S}_{m0}$ are available, either from previous experiments or information available from the sensor manufacturer.

### 3.3.2 B-Spline Basis Functions

Any smooth function can be approximated to an arbitrary degree of accuracy using basis functions such as B-splines.[33,48] B-splines basis functions consist of *M*th order (or (*M*-1) degree) piecewise polynomials that are positive within *M* intervals and zero elsewhere.[49] For example, the *s*th state of the multivariate stochastic system shown in Equation (3.1) can be represented by a linear combination of $c_s$ B-splines:[50,51]

$$x_{\sim s}(t) = \sum_{l=1}^{c_s} \beta_{s,l} \varphi_{s,l}(t) \quad \text{for } s=1,\ldots,X \tag{3.6}$$

where $\beta_{s,l}$ is a B-spline coefficient and $\varphi_{s,l}(t)$ is the corresponding B-spline basis function. In matrix form, Equation (3.6) is:

$$\mathbf{x}_{\sim}(t) = \mathbf{\Phi}(t)\mathbf{B} \tag{3.7}$$

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

$$\mathbf{\Phi}(t) = \begin{bmatrix} \boldsymbol{\varphi}_1^T(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}_2^T(t) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\varphi}_X^T(t) \end{bmatrix} \tag{3.8}$$

and

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_X \end{bmatrix} \tag{3.9}$$

A sufficient number of well-placed spline knots must be selected by the modeler to obtain an accurate approximation for the state trajectory.[50,51]

### 3.3.3 Approximate Maximum Likelihood Estimation (AMLE) Algorithm

Varziri et al.[52] proposed an approximate maximum likelihood estimation (AMLE) method for estimating the parameters of the SDE models. When deriving this algorithm, they discretized the SDE in Equation (3.1) using an Euler scheme:

$$\mathbf{x}(t_{i-1} + \Delta t) = \mathbf{x}(t_i) = \mathbf{x}(t_{i-1}) + f(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t + \boldsymbol{\eta}_{\mathbf{d}}(t_{i-1})\Delta t \tag{3.10a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{3.10b}$$

where $\mathbf{x}(t_i)$ is the value of the state variable at $q$ uniformly-spaced time points $t_i$ , $i=0,..,q$ and $\boldsymbol{\eta}_{\mathbf{d}}$ is a discrete white noise vector. Consider $\mathbf{X_q} = [\mathbf{x}^{\mathbf{T}}(t_0)\dots\mathbf{x}^{\mathbf{T}}(t_q)]^{\mathbf{T}}$ as the stacked vector of state values at discrete times. Varziri et al. used this discrete model and B-spline basis functions to derive the following objective function for estimating parameters in SDE models with known noise variances and disturbance intensities:[52]

$$J_{AMLE} = -\ln p(\mathbf{Y_m}, \mathbf{X_{q\sim}}, \mathbf{x}_m(t_0) \mid \boldsymbol{\theta}) = [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{m\sim}}, \mathbf{U_m}, \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{m\sim}}, \mathbf{U_m}, \boldsymbol{\theta})]$$
$$+ (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0})^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0})$$
$$+ \int_{t_0}^{t_q} [\dot{\mathbf{x}}_\sim(t) - \mathbf{f}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta})]^T \mathbf{Q}^{-1} [\dot{\mathbf{x}}_\sim(t) - \mathbf{f}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta})] dt$$

(3.11)

The integral in Equation (3.11) is called an Ito stochastic integral. Since a discrete white noise is used in simulations to approximate the continuous white noise, the Ito integral in Equation (3.11) is approximated by:

$$\sum_{i=1}^{q} [\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t]^T \mathbf{Q}^{-1} \Delta t^{-1} [\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t]$$

where $\Delta t$ is chosen as a small value compare to the dynamic of system.

Because the B-spline expressions in Equation (3.7) can be readily integrated with respect to time, the integrand in the third term on the right-hand side of Equation (3.11) becomes an algebraic expression. As a result, there is no need for numerical solution of differential equations when estimating $\mathbf{B}$ and $\boldsymbol{\theta}$.[50,52] Note that the initial states $\mathbf{x}_0$ do not need to be included as separate decision variables in the optimization because they are approximated using $\mathbf{x}_{\sim 0}$, which can be computed using the B-spline coefficients. When no measurements are available for any of the initial state values, the $(\mathbf{x}_{m0} - \mathbf{x}_{\sim 0})^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0})$ term disappears from the objective function. If the initial value of the *s*th state variable is not measured, but is perfectly known by the modeler, this knowledge can be incorporated by fixing the first B-spline coefficient for the *s*th state at the known value $\beta_{s,1} = x_s(0)$.

A drawback of Varziri's AMLE method is that it requires known values of the stochastic disturbance intensities in $\mathbf{Q}$ when optimizing to find $\boldsymbol{\theta}$ and $\mathbf{B}$.[52] To address this problem,

Varziri et al.[3] extended their AMLE methodology, based on a technique developed by Heald and Stark[26] for estimating parameters in time-series models with unknown noise variance. Varziri et al.[3] proposed a two-step optimization algorithm where $\boldsymbol{\theta}$ and $\mathbf{B}$ are determined in an inner loop, using the objective function in Equation (3.11) and an assumed value of $\mathbf{Q}$. In the second step (outer loop) $\mathbf{Q}$ is updated using the following objective function assuming that $\boldsymbol{\theta}$ and $\mathbf{B}$ are known:[3]

$$\hat{\mathbf{Q}} = \arg\min_{\mathbf{Q}} \left\{ \frac{\hat{\sigma}_1^2(Q_1)}{\sigma_1^2} - 1 \right\}^2 + \ldots + \left\{ \frac{\hat{\sigma}_Y^2(Q_Y)}{\sigma_Y^2} - 1 \right\}^2 \tag{3.12}$$

where the true noise variance $\sigma_j^2$ ($j$=1,…,$Y$) is assumed to be known and the estimate of the noise variance from the data $\hat{\sigma}_j^2$ ($j$=1,…,$Y$) is calculated using residuals between the fitted splines and the data and the approximate Hessian for the AMLE objective function.[3,26]

### 3.3.4 EM Algorithm

In general, the likelihood function of the parameters given the measurements does not have a closed form. However, using the Markov property of the states, the likelihood of the complete data $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ has a closed form. In the $k$th iteration of the E step of the EM algorithm, the expected value of the log-likelihood of the complete data is determined, given the vector of complete measurements and values of the parameters $\boldsymbol{\zeta}_k$ arising from the current (i.e., $k$th) parameter iteration:[32]

$$R(\boldsymbol{\zeta}, \boldsymbol{\zeta}_k) = \underset{\mathbf{X_q}}{\mathrm{E}} \{ \ln[\, p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})] \mid \mathbf{Y_m}, \boldsymbol{\zeta}_k \} \tag{3.13}$$

In Equation (3.13), $\mathbf{X_q}$ below the expectation symbol indicates that this expected value is computed over all possible values of the discretized state variables.

In the M step of the EM algorithm (i.e., the maximization step), the expected value determined in the first step is maximized to generate improved parameter estimates for use in the next iteration:

$$\hat{\boldsymbol{\zeta}}_{k+1} = \arg\max_{\boldsymbol{\zeta}} \; R(\boldsymbol{\zeta}, \boldsymbol{\zeta}_k) \tag{3.14}$$

Wei and Tanner[53] showed that the E step in the EM algorithm can be approximated as:

$$R(\boldsymbol{\zeta}, \boldsymbol{\zeta}_\mathbf{k}) \approx \frac{1}{v}\sum_{l=1}^{v} \ln[p(\mathbf{X_q}^{(l)}, \mathbf{Y_m} \mid \boldsymbol{\zeta})] \tag{3.15}$$

where the superscript ($l$) is used to indicate the $l$th random value of the state vector sampled from $p(\mathbf{X_q} \mid \mathbf{Y_m}, \boldsymbol{\zeta})$ [14,35]. When computing $R(\boldsymbol{\zeta}, \boldsymbol{\zeta}_\mathbf{k})$ in Equation (3.15), it is possible to use $v$=1 (i.e., a single term rather than a summation), if $\mathbf{X_q}^{(l)}$ is replaced by the mode of $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ : [35]

$$R(\boldsymbol{\zeta}_k, \boldsymbol{\zeta}) \approx \ln[p(\mathbf{X}_\mathbf{q}^{(1)}, \mathbf{Y_m} \mid, \boldsymbol{\zeta})] \tag{3.16}$$

where $\mathbf{X}_\mathbf{q}^{(1)}$ is the mode of $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$.

### 3.4 Development of the Proposed AEM Algorithm

In this section, a new Approximate Expectation Maximization (AEM) algorithm is developed for estimating model parameters and process disturbance intensities. The proposed algorithm uses B-spline basis functions in the approximation for the E-step of the EM algorithm. The AEM algorithm is derived by approximating the expectation step of the EM algorithm using Equation (3.16). Appendix 3.A contains the derivation of a closed-form expression for $R(\boldsymbol{\zeta}_k, \boldsymbol{\zeta})$ that is obtained using B-spline basis functions. We first consider the case where there are no missing observations (i.e. all of the states are

measured and measurements are taken frequently). As shown in Appendix 3.A, the corresponding objective function is:

$$\begin{aligned} J_{\text{AEM}} &= [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U_m}, \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U_m}, \boldsymbol{\theta})] \\ &\quad + (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0})^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{\sim m0} - \mathbf{x}_{\sim 0}) - n \ln[\det(\mathbf{Q})] \\ &\quad + \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]^T \mathbf{Q}^{-1} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt \end{aligned} \tag{3.17}$$

where $n$ is the number of measurements. In the case where there are missing observations, a closed form for Equation (3.16) is derived in Appendix 3.B, resulting in:

$$\begin{aligned} J_{\text{AEM}} &= [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U}, \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U_m}, \boldsymbol{\theta})] \\ &\quad + (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0})^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0}) - q \ln[\det(\mathbf{Q})] \\ &\quad + \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]^T \mathbf{Q}^{-1} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt \end{aligned} \tag{3.18}$$

where

$$q = \frac{t_q - t_0}{\Delta t} \tag{3.19}$$

and $\Delta t$ is the sampling time selected by the modeller for the discrete white noise disturbances that are assumed to influence the process behavior. The value of $\Delta t$ should be small compared to the dynamics of the system. Note that the difference between Equations (3.17) and (3.18) is in their third term. In the case where all measurements are available $n \ln[\det(\mathbf{Q})]$ term appears in the objective function (Equation (3.17)). However, in the case where there are missing observations $q \ln[\det(\mathbf{Q})]$ term appears in the objective function (Equation (3.18)).

Minimizing $J_{\text{AEM}}$ with respect to $\boldsymbol{\theta}$, $\mathbf{Q}$ and $\mathbf{B}$ provides approximate maximum likelihood estimates for the model parameters $\boldsymbol{\theta}$ and the disturbance intensities $\mathbf{Q}$:

$$\hat{\boldsymbol{\theta}}, \hat{\mathbf{Q}}, \hat{\mathbf{B}} = \arg\min_{\boldsymbol{\theta},\mathbf{Q},\mathbf{B}} J_{\text{AEM}} \tag{3.20}$$

Note that spline coefficients $\mathbf{B}$ can be added to the parameter vector $\boldsymbol{\zeta}$ since other nuisance parameters (i.e., measurement noise variances and process disturbance intensities) are incorporated.

### 3.5 Illustrative Example: Nonlinear Two-State CSTR Model

In this section, the SDE model for a CSTR is used to illustrate the use of the AEM objective function (Equation (3.18)) for parameter estimation. Simulations are used to compare parameter estimation results obtained from the proposed AEM algorithm to results obtained from the extended AMLE method[3] and from CTSM software[1]. The SDE model is a two-state CSTR example from Marlin[54] with additional stochastic disturbance terms:[3]

$$\frac{\mathrm{d}C_A(t)}{\mathrm{d}t} = \frac{F(t)}{V}(C_{A0}(t) - C_A(t)) - k_r(T(t))C_A(t) + \eta_C(t) \tag{3.21a}$$

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{F(t)}{V}(T_0(t) - T(t)) + UA(T(t) - T_{in}(t)) + ¥k_r(T(t))C_A(t) + ¦_T(t) \tag{3.21b}$$

$$y_C(t_{m\,1,j}) = C_A(t_{m\,1,j}) + \varepsilon_C(t_{m\,1,j}) \tag{3.21c}$$

$$y_T(t_{m\,1,j}) = T(t_{m\,1,j}) + \varepsilon_T(t_{m\,1,j}) \tag{3.21d}$$

where $C_A$ is the concentration of reactant A and $T$ is the temperature. The rate constant for the exothermic reaction is:

$$k_r(T) = k_{ref} \exp\left(-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \tag{3.22}$$

*UA* is a heat transfer coefficient that depends on the cooling water flow rate and heat capacity:

$$UA(F_c) = \frac{aF_c^{b+1}}{V\rho c_p\left(F_c + \frac{aF_c^b}{2\rho_c c_{pc}}\right)} \tag{3.23}$$

and ¥ is related to the enthalpy of reaction:

$$¥ = \frac{(-\Delta H_{rxn})}{°c_p} \tag{3.24}$$

The initial steady-state values of the states are $C_A(0)$= 1.569 kmol·m$^{-3}$ and $T(0)$=341.37 K. To illustrate the methodology, $C_A(0)$ is assumed to be perfectly known, while $T(0)$ is measured with known variance of $S_T^2$ =5.0 K$^2$. As a result, the unknown true value of the initial temperature is $T(0)$=341.37 K, but initial temperature measurements deviate from this value. The AEM, extended AMLE and CTSM algorithms are used to estimate the model parameters $\boldsymbol{\theta} = [k_{ref}, E/R, a, b]^T$ and disturbance intensities $Q_C$ and $Q_T$. When AEM and extended AMLE are used for parameter estimation, an estimate of the initial temperature $T(0)$ is determined from the B-spline coefficients.

The model inputs are the feed flow rate *F*, the inlet concentration $C_{A0}$, the inlet temperature $T_0$, the coolant inlet temperature $T_{in}$ and the flow rate of coolant to the cooling coil *Fc*. The known constants for this CSTR model, provided by Marlin[54] are shown in Table 3.1.

**Table 3.1 Model constants** [45]

| Model Constants | Value | Units |
|---|---|---|
| $c_p$ | 4186.8 | $J \cdot kg^{-1} \cdot K^{-1}$ |
| $c_{pc}$ | 4186.8 | $J \cdot kg^{-1} \cdot K^{-1}$ |
| $T_{ref}$ | 350 | K |
| $V$ | 1 | $m^3$ |
| $\rho$ | 1000 | $kg \cdot m^{-3}$ |
| $\Delta H_{rxn}$ | $-544.154 \times 10^3$ | $J \cdot kmol^{-1}$ |
| $\sigma_C^2$ | $4 \times 10^{-4}$ | $kmol^2 \cdot m^{-6}$ |
| $\sigma_T^2$ | 0.64 | $K^2$ |

During the dynamic experiments, the concentration $C_A$ is measured $n_C$ times and the temperature $T$ is measured $n_T$ times. $\varepsilon_C(t_{m\,1,j})\quad j=1\ldots n_C$ and $\varepsilon_T(t_{m\,2,j})\quad j=1\ldots n_T$ are the corresponding white-noise sequences with variances $\sigma_C^2$ and $\sigma_T^2$, respectively. We also assume that $\eta_C$, $\eta_T$, $\varepsilon_C$ and $\varepsilon_T$ are independent. The ODE45 solver in MATLAB™ was used to generate simulated data affected by Gaussian measurement errors and stochastic process disturbances, using the input variable trajectories shown in Figure 3.1. The duration of each simulation is 64 min. These same input trajectories were used by Varziri et al.[3,52] when testing their AMLE algorithm. Simulated data were generated using the true parameter values from Marlin[54] shown in Table 3.2. Measurement noise variances $\sigma_C^2$ and $\sigma_T^2$, provided in Table 3.3, are assumed to be known, whereas process disturbance intensities $Q_C$ and $Q_T$ are unknown.

**Table 3.2 Estimates and 95% Confidence intervals for AEM parameter estimates from one of the 100 Monte Carlo simulations**

| Parameter | Unit | True Value | Initial Guess | Estimate ± 95% Confidence Interval |
|---|---|---|---|---|
| $k_{ref}$ | $min^{-1}$ | 0.461 | 0.293 | 0.436 ± 0.007 |
| $(E/R)/ 10^3$ | K | 8.3301 | 4.2012 | 8.1504 ± 0.0952 |
| $a/10^6$ | | 1.678 | 3.236 | 1.568 ± 0.239 |
| $b$ | | 0.500 | 0.760 | 0.494 ± 0.054 |
| $T(0)$ | K | 341.38 | 347.12 | 342.24 ± 0.43 |
| $Q_C$ | $kmol^2 \cdot m^-$ | 0.010 | 0.015 | 0.010 ± 0.0009 |
| $Q_T$ | $K^2 \cdot min^{-1}$ | 4.0 | 6.1 | 4.8 ± 0.01 |

The random process disturbances (see Figure 3.2) were generated using a discrete white-noise process with a sampling interval $\Delta t$=0.5 min which is approximately 10 times smaller than the dominant time constant of the CSTR system.

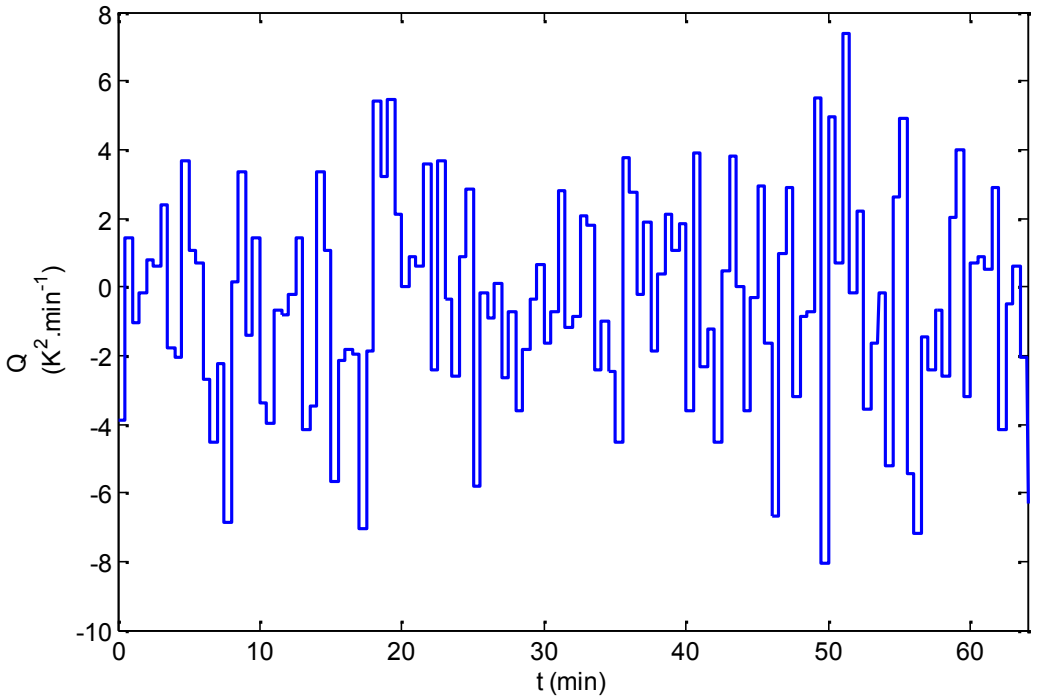



**Figure 3.1. Typical discrete stochastic process disturbances obtained using $Q$= 4 $K^2 \cdot min^{-1}$ and $\Delta t$= 0.5 min**

The AEM objective function for estimating parameters in the CSTR model is:

$$J_{AEM,CTSR} = \frac{1}{\pm_C^2}\sum_{j=1}^{n_C}\left(y_C(t_{m1,j}) - C_{A\sim}(t_{m1,j})\right)^2 + \frac{1}{\pm_T^2}\sum_{j=1}^{n_T}\left(y_T(t_{m2,j}) - T_\sim(t_{m2,j})\right)^2$$
$$+ \frac{(T_m(0) - T_\sim(0))^2}{S_T^2} + q_C \ln(Q_C) + q_T \ln(Q_T)$$
$$+ \frac{1}{Q_C}\int_{t_0}^{t_q}\left(\frac{\mathrm{d}C_{A\sim}(t)}{\mathrm{d}t} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r(T_\sim(t))C_{A\sim}(t)\right)^2 \mathrm{d}t +$$
$$+ \frac{1}{Q_T}\int_{t_0}^{t_q}\left(\frac{\mathrm{d}T_\sim(t)}{\mathrm{d}t} - \frac{F(t)}{V}(T_0(t) - T_\sim(t)) - UA[T_\sim(t) - T_{in}(t)] - ¥k_r(T_\sim(t))C_{A\sim}(t)\right)^2 \mathrm{d}t \quad (3.25)$$

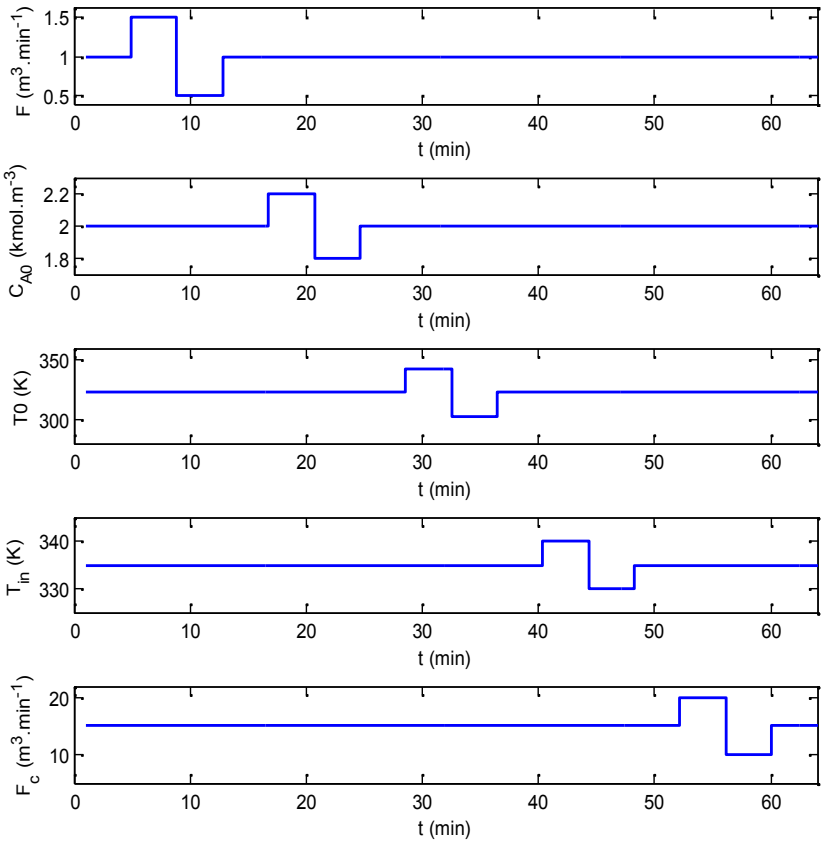



**Figure 3.2 Input trajectories for nonlinear CSTR[50]**

In Equation (3.25), $q_C$ = $q_T$ =128 are the number of discrete random shocks used to generate the disturbance sequences. Note that the discrete white noise is added before solving the dynamic equations of concentration and temperature. The final term on the right-hand side appears because the initial temperature $T(0)$ is assumed to be unknown in this example and must be approximated using $T_{\sim}(0)$. Since $C_A(0)$ is perfectly known, there is no corresponding concentration term in Equation (3.25) and the corresponding first spline coefficient $\beta_{c,1}$ is set at 1.569 and is not used as a decision variable by the optimizer.

For comparison, the corresponding extended AMLE objective function[52] used to estimate the model parameters and B-spline coefficients when $Q_C$ and $Q_T$ are assumed known is:

$$J_{I,AMLE} = \frac{1}{\pm_C^2}\sum_{j=1}^{n_C}\left(y_C(t_{m1,j}) - C_{A\sim}(t_{m1,j})\right)^2 + \frac{1}{\pm_T^2}\sum_{j=1}^{n_T}\left(y_T(t_{m2,j}) - T_{\sim}(t_{m2,j})\right)^2 + \frac{(T_m(0) - T_{\sim}(0))^2}{S_T^2} + \frac{1}{Q_C}\int_{t_0}^{t_q}\left(\frac{\mathrm{d}\,C_{A\sim}(t)}{\mathrm{d}\,t} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r(T_{\sim}(t))C_{A\sim}(t)\right)^2 \mathrm{d}\,t + \frac{1}{Q_T}\int_{t_0}^{t_q}\left(\frac{\mathrm{d}\,T_{\sim}(t)}{\mathrm{d}\,t} - \frac{F(t)}{V}(T_0(t) - T_{\sim}(t)) - UA[T_{\sim}(t) - T_{in}(t)] - ¥k_r(T_{\sim}(t))C_{A\sim}(t)\right)^2 \mathrm{d}\,t \quad (3.26)$$

The only differences between the AEM and AMLE objective functions are the two logarithm terms that appear in Equation (3.25). The outer objective function used for estimating $Q_C$ and $Q_T$ in the extended AMLE method is:[3]

$$J_{O,AMLE} = \frac{1}{n_C \sigma_C^2}\left(\sum_{j=1}^{n_C}\left(y_C(t_{m\,1,j}) - C_{A\sim}(t_{m\,1,j})\right)^2 + Trace\left(\left(\frac{\partial^2 J_{I,AMLE}}{\partial^2 C_{A\sim}^2}\right)^{-1}\right)\right) - 1$$
$$+ \frac{1}{n_T \sigma_T^2}\left(\sum_{j=1}^{n_T}\left(y_T(t_{m\,2,j}) - T_{\sim}(t_{m\,2,j})\right)^2 + Trace\left(\left(\frac{\partial^2 J_{I,AMLE}}{\partial^2 T_{\sim}^2}\right)^{-1}\right)\right) - 1 \quad (3.27)$$

Estimating the parameter vector $[k_{ref}, E/R, a, b, Q_C, Q_T]^T$ using extended AMLE required a two-step optimization procedure wherein $k_{ref}$, $E/R$, $a$, $b$ and the B-spline coefficients were estimated in an inner loop using assumed values of $Q_C$ and $Q_T$. Updated values of $Q_C$ and $Q_T$ were estimated in an outer loop using objective function (3.27). The outer objective function contains the Hessians $\partial^2 J_{I,AMLE}/\partial \boldsymbol{\beta}_C^2$ and $\partial^2 J_{I,AMLE}/\partial \boldsymbol{\beta}_T^2$ where $\boldsymbol{\beta}_C$ and $\boldsymbol{\beta}_T$ are vectors of B-spline coefficients that correspond to approximate state trajectories $C_{A\sim}$ and $T_{\sim}$, respectively. These Hessians were computed using the "gjh" function in IPOPT.

The AEM and extended AMLE optimization problems were solved using the IPOPT optimization code from Wachter and Biegler[55] with model information provided using AMPL™.[56] The outer loop of the extended AMLE algorithm was optimized using the "lsqnonlin" function in MATLAB™. Termination tolerances for lsqnonlin were set at $10^{-3}$ for changes in the value of the objective function and at $10^{-6}$ for parameter values. The maximum numbers of iterations and function evaluations were set to 1000 and 2000, respectively. Default values of optimization settings in IPOPT were used when optimizing objective functions (3.25) and (3.26). Reducing the tolerances used in lsqnonlin and IPOPT by a factor of 10 did not produce significant changes (i.e., in the first three significant figures) in the estimated parameter values. 4th order (cubic) B-splines were used in both the AEM and AMLE parameter estimation studies with one

knot at every sampling interval used to generate the disturbances. This knot placement frequency was chosen based on results from a preliminary study involving several single-state SDE models.[57] Estimates for the complete vector of parameters $\boldsymbol{\zeta}$ were obtained for different sets of simulated data under several different scenarios, as shown in Table 3.3. In each scenario, parameter estimation problems were repeated 100 times with 100 different sets of random initial guesses for the parameters (chosen from a uniform distribution between 50% and 150% of their true values). Different random sequences for Gaussian process disturbances and measurement noise variables were used to generate the 100 data sets. To compare the quality of the parameter estimates obtained in the different scenarios, medians and interquartile ranges (IQRs) are reported in Table 3.3 when parameters are estimated using the simulated data sets and corresponding initial guesses.

In scenario A (the base case scenario) $C_A$ and $T$ were measured once every 0.5 min, so that 128 concentration measurements and 128 temperature measurements are available for parameter estimation. Estimated trajectories ($C_{A\sim}$ and $T_{\sim}$) for one of the simulated data sets are shown in Figure 3.3, along with the true state trajectories and the corresponding data values when AEM is used for parameter estimation. As expected, the estimated state trajectories follow the true trajectories closely.

**Table 3.3 True parameter values, median values and IQRs for the estimates based on 100 Monte Carlo runs for different scenarios.**

| **Parameter** | | | $k_{ref}$ | $(E/R)/10^3$ | $a/10^6$ | $b$ | $T(0)$ | $Q_C$ | $Q_T$ |
|---|---|---|---|---|---|---|---|---|---|
| **Unit** | | | $min^{-1}$ | K | | | K | $kmol^2 \cdot m^{-6} \cdot min^{-1}$ | $K^2 \cdot min^{-1}$ |
| **True Value** | | | **0.461** | **8.3301** | **1.678** | **0.500** | **341.38** | **0.010** | **4.0** |
| **Scenario** | | | | | | | | | |
| A | AEM | Median | 0.432 | 8.2308 | 1.502 | 0.497 | 342.32 | 0.009 | 4.9 |
| | | IQR | 0.020 | 0.2062 | 0.465 | 0.104 | 1.21 | 0.004 | 1.4 |
| | AMLE | Median | 0.420 | 8.1931 | 1.315 | 0.512 | 342.42 | 0.056 | 24.2 |
| | | IQR | 0.024 | 0.2394 | 0.545 | 0.118 | 1.33 | 0.615 | 40.2 |
| | CTSM | Median | 0.460 | 8.2800 | 1.575 | 0.508 | 341.00 | 0.092 | 1.5 |
| | | IQR | 0.015 | 0.1650 | 0.625 | 0.131 | 1.00 | 0.011 | 0.4 |
| B | AEM | Median | 0.440 | 8.2949 | 1.589 | 0.483 | 342.29 | 0.003 | 3.8 |
| | | IQR | 0.026 | 0.2777 | 0.618 | 0.146 | 1.16 | 0.003 | 0.9 |
| | AMLE | Median | 0.432 | 8.3071 | 1.558 | 0.478 | 342.39 | 0.009 | 9.4 |
| | | IQR | 0.037 | 0.3313 | 0.676 | 0.132 | 1.36 | 0.134 | 11.0 |
| C | AEM | Median | 0.444 | 8.2548 | 1.589 | 0.493 | 342.26 | 0.003 | 4.0 |
| | | IQR | 0.020 | 0.1819 | 0.349 | 0.073 | 1.18 | 0.002 | 0.4 |
| | AMLE | Median | 0.428 | 8.2344 | 1.442 | 0.509 | 342.37 | 0.021 | 8.7 |
| | | IQR | 0.025 | 0.2148 | 0.364 | 0.096 | 1.37 | 0.869 | 9.6 |
| D | AEM | Median | 0.437 | 8.1891 | 1.561 | 0.484 | 342.97 | 0.005 | 3.7 |
| | | IQR | 0.022 | 0.2187 | 0.560 | 0.126 | 1.64 | 0.005 | 0.7 |
| | AMLE | Median | 0.423 | 8.1709 | 1.392 | 0.501 | 343.06 | 0.018 | 13.3 |
| | | IQR | 0.030 | 0.2998 | 0.645 | 0.132 | 1.76 | 0.251 | 14.9 |
| E | AEM | Median | 0.436 | 8.2525 | 1.581 | 0.507 | 343.31 | 0.010 | 5.1 |
| | | IQR | 0.024 | 0.2481 | 0.603 | 0.157 | 2.12 | 0.009 | 21.3 |
| | AMLE | Median | 0.415 | 8.1952 | 1.276 | 0.527 | 343.53 | 0.487 | 38.5 |
| | | IQR | 0.023 | 0.2797 | 0.518 | 0.145 | 2.70 | 1.326 | 77.1 |
| F | AEM | Median | 0.427 | 8.1984 | 1.500 | 0.498 | 342.36 | 0.012 | 6.6 |
| | | IQR | 0.023 | 0.2508 | 0.530 | 0.112 | 1.64 | 0.006 | 2.8 |
| | AMLE | Median | 0.413 | 8.1601 | 1.262 | 0.535 | 343.59 | 0.633 | 44.4 |
| | | IQR | 0.027 | 0.2829 | 0.530 | 0.168 | 2.97 | 4.700 | 151.2 |
| G | AEM | Median | 0.416 | 8.3696 | 1.541 | 0.469 | 342.31 | - | 5.3 |
| | | IQR | 0.042 | 0.4701 | 0.978 | 0.225 | 1.31 | - | 2.1 |
| | AMLE | Median | 0.397 | 8.3023 | 1.323 | 0.500 | 342.49 | - | 25.5 |
| | | IQR | 0.053 | 0.5391 | 0.859 | 0.212 | 1.36 | - | 32.4 |

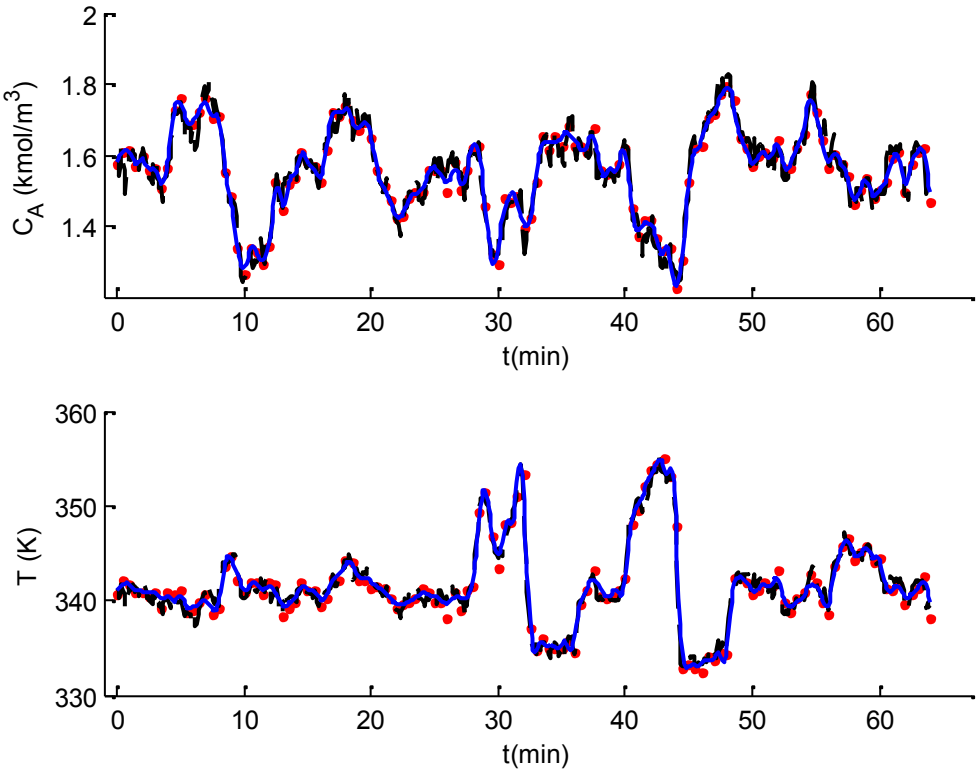



**Figure 3.3 Measured, true, and predicted concentration and temperature responses for the AEM method in scenario A, using simulated data from one of the 100 Monte Carlo simulations. Corresponding initial guesses and parameter estimates are provided in Table 3 (• simulated data, ----- response with true parameters and true stochastic noise, ___AEM response)**

Approximate confidence intervals for the corresponding parameters are provided in Table 3.2. The model parameter and disturbance intensity estimates from the AEM algorithm should be asymptotically Gaussian with covariance equal to the inverse of the Hessian matrix,[58] due to the central limit theorem. Thus, assuming that the data set is sufficiently large, approximate 100(1-α)% confidence intervals can be determined by approximating the covariance matrix using the Hessian evaluated at the estimates of parameters and B-spline coefficients:[52,58-60]

$$\hat{\boldsymbol{\tau}} \pm z_{\alpha/2}\sqrt{\left.\mathrm{diag}(\partial^2 J_{AEM}/\partial \boldsymbol{\tau}^2)^{-1}\right|_{\hat{\boldsymbol{\tau}}}} \tag{3.28}$$

where $\boldsymbol{\tau}=[\boldsymbol{\zeta}^T,\mathbf{B}^T]^T$ is a stacked vector of model parameters, disturbance intensities and B-spline coefficients. Note that Varziri et al. used a similar method to obtain approximate confidence intervals for $k_{ref}, E/R, a, b$ and $T(0)$ from the extended AMLE algorithm. However, they were not able to obtain approximate confidence intervals for $Q_C$ and $Q_T$ because these disturbance intensity parameters were estimated using their outer objective function (Equation (3.27)), which was not derived using Maximum Likelihood methods. The capability to generate approximate confidence intervals for disturbance intensities is one of the benefits of the AEM algorithm compared to extended AMLE.

Box plots summarizing the parameter estimates from all 100 Monte Carlo simulations are shown in Figures 3.4 and 3.5 to provide a comparison of the effectiveness of AEM, extended AMLE and CTSM methods for the base case (scenario A) estimation problem. Default values of optimization settings were used when implementing the CTSM software. Use of CTSM requires upper bounds and lower bounds for all of the parameters. Lower bounds of parameters were set at zero and upper bounds were set at 10 times the true parameter values. Using CTSM, successful parameter estimation was only obtained for 36 of the 100 Monte Carlo cases attempted; the CTSM box plots in Figures 3.4 and 3.5 (and the medians and IQRs in Table 3.3) were constructed using only these 36 sets of parameter values. The remaining 64 estimation attempts experienced convergence failures. Several repeated CTSM parameter estimation attempts with revised tolerances and convergence settings did not result in improved convergence. However, when improved initial guesses (between 75 and 100% of the true values) were used, a higher success rate was obtained. Both the AEM and extended AMLE algorithms

converged for all of the 100 Monte Carlo cases attempted, and the corresponding results are shown in Figures 3.4 and 5.

In scenario A, the AEM algorithm converged to the neighbourhood of the true parameter values using 80 of the 100 simulated data sets. For the remaining 20 data sets, the algorithm converged to parameter estimates with noticeably larger values of $Q_C$ and $Q_T$, which correspond to outlier in the box plots in Figure 3.5. These results do not seem be caused by convergence to local minima, because re-starting the parameter estimation at the true values rather than the random initial guesses for these data sets resulted in the same large estimates of $Q_C$ and $Q_T$. When extended AMLE was used for parameter estimation, these was negligible bias in the model parameter estimates ($k_{ref}$, $E/R$, $a$, $b$ and $T(0)$), but the estimates of $Q_C$ and $Q_T$ were biased toward large values. Also, the IQRs for estimates of $Q_C$ and $Q_T$ obtained using extended AMLE are very wide when compared with those obtained using AEM and CTSM (e.g., for $Q_C$, the IQR is 0.004 for AEM, which is much smaller than 0.615 for extended AMLE and 0.011 for CTSM). The CTSM estimates of $Q_C$ and $Q_T$ have more bias than the estimates obtained using AEM.

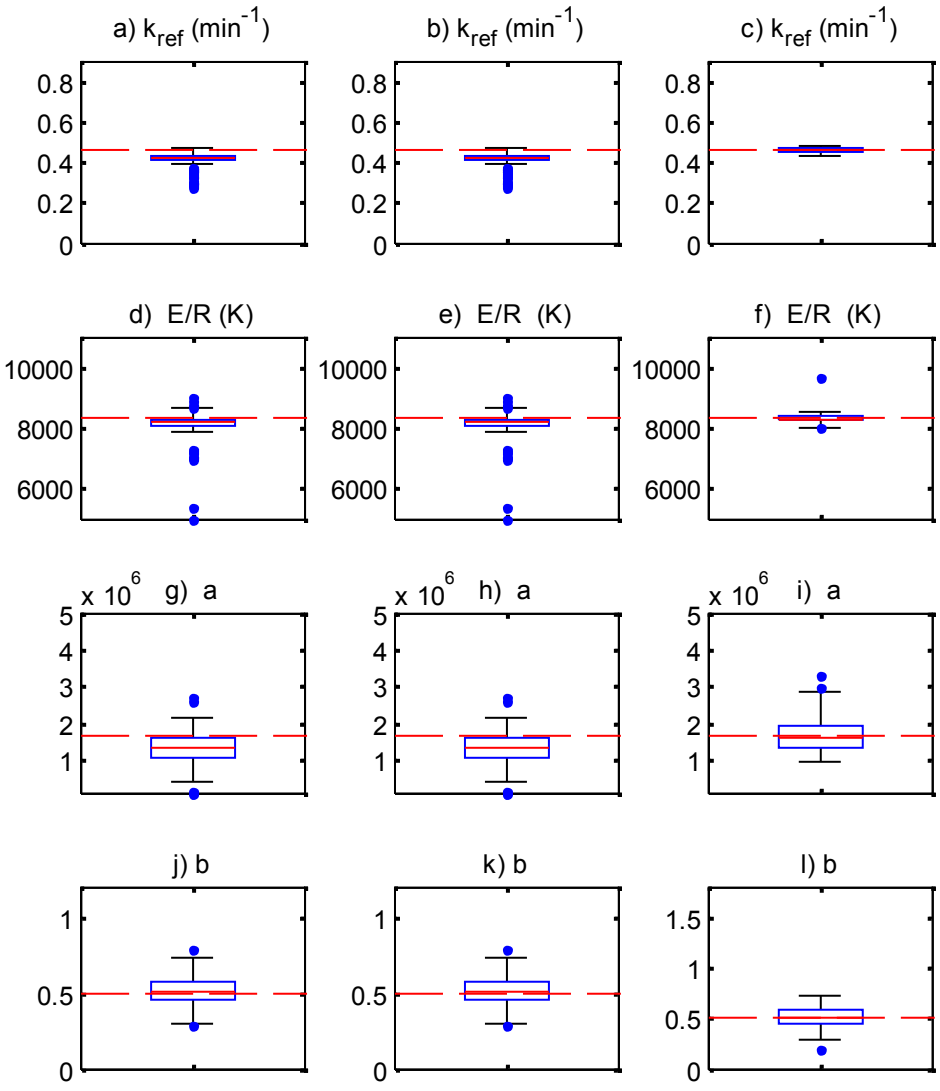



**Figure 3.4 Box-plots for estimates of model parameters using the AEM method extended AMLE method and CTSM in scenario A. The dashed horizontal lines show the true values used to generate the simulated data.**

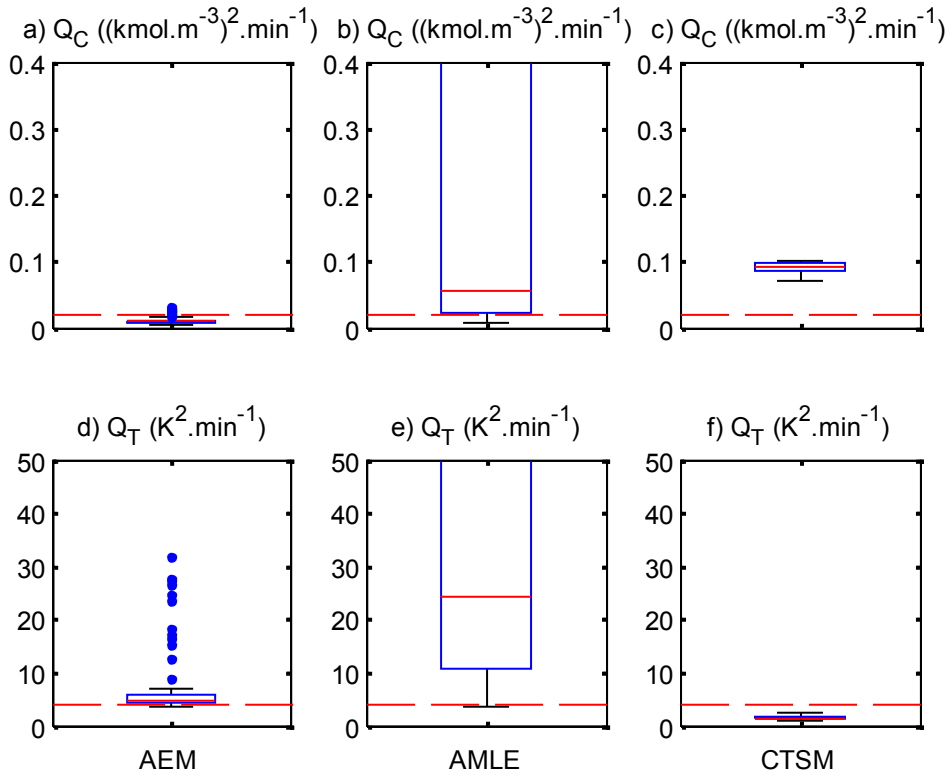



**Figure 3.5 Box-plots for disturbance intensity estimates obtained using the AEM method, extended AMLE method and CTSM in scenario A. The dashed horizontal lines show the true values used to generate the simulated data.**

Note that parameter estimation using CTSM was not studied for the other scenarios in Table 3.3 because of the convergence difficulties encountered for scenario A. On average, each parameter estimation conducted using AEM required approximately 0.2 min of computation time on a laptop computer with Intel® Core™ 2, Duo CPU, 1.86 GHz, 2.96 GB of RAM, which is about 5 times faster than each extended AMLE parameter estimation. The typical computation time for each parameter estimation that converged using CTSM was 3 min.

Scenario B in Table 3.3 was used to study the influence of worse initial guesses on the quality of the parameter estimates. Using initial guesses that were chosen randomly

between 50% and 800% of the true parameter values had no significant influence on the majority of the AEM parameter estimation results. Note, however, that 11 sets of simulated data resulted in convergence to local minima when the worse initial guesses were specified, leading to large IQRs and larger median values for the estimates of $Q_C$ and $Q_T$. The extended AMLE parameter estimates in scenario B have larger variability and are even more biased than those obtained using the good initial values in scenario A. The results obtained using AEM are significantly better than those obtained using AMLE.

In scenario C in Table 3.3, $C_A$ and $T$ were measured less frequently than in scenario A, (i.e., only 64 concentration measurements and 64 temperature measurements instead of 128 for scenario A). As expected, the parameter estimates have larger variability due to the smaller data set in scenario C (e.g., the IQR for $k_{ref}$ increases from 0.020 to 0.026 $min^{-1}$ using AEM and from 0.024 to 0.037 using extended AMLE). The $Q_C$ estimates obtained from AEM in scenario C are more biased toward small values than they were in scenario A. Conversely, the extended AMLE estimates for $Q_C$ and $Q_T$ appear to be better than in scenario A when the smaller data set is used. The reason for this behaviour is unclear. In both scenarios A and C, the AEM estimates of $Q_C$ and $Q_T$ are better than the corresponding extended AMLE estimates.

In scenario D the values of $Q_C$ and $Q_T$ used to generate the simulated data were decreased to the half of their original values (i.e., $Q_C$ =0.005 $(kmol.m^{-3})^2 \cdot min^{-1}$ and $Q_T$ =2 $K^2 \cdot min^{-1}$) used in scenario A. The base case number of measurements (128) was used when estimating the parameters. Because the disturbances are smaller, improved estimates of parameters $k_{ref}$, $E/R$, $a$, $b$ and $T(0)$ are obtained using both AEM and extended AMLE, as

expected. However, the estimates for $Q_C$ and $Q_T$ are slightly worse from AEM than when the larger disturbances were used, presumably because it is more difficult to accurately detect the influence of the disturbances relative to the measurement noise. The AEM estimates of $Q_C$ and $Q_T$ are better than the corresponding AMLE estimates in scenario D.

In scenario E the known true measurement noise variances were increased to twice as their original values (i.e., $\sigma_C^2 = 8\times10^{-4}$ kmol$^2$·m$^{-6}$ and $\sigma_T^2 = 12.8\times10^{-1}$ K$^2$), with all other settings held at those from scenario A. These new true values were used in objective functions (3.25) and (3.26) when estimating the parameters. As expected, using larger noise variances resulted in slightly worse parameter estimates from both AEM and AMLE. Also, the AEM estimates for $Q_C$ and $Q_T$ became slightly worse than in the scenario A. The AEM estimates for $Q_C$ and $Q_T$ in scenario E are better than the corresponding AMLE estimates.

Scenario F was used to investigate the robustness of the AEM and AMLE algorithms to imperfect knowledge about the measurement noise variances $\sigma_C^2$ and $\sigma_T^2$. In this scenario, the modeler believes $\sigma_C^2$ and $\sigma_T^2$ to be twice their respective true values and uses this incorrect assumption in the objective functions. The AEM estimates of the model parameters $k_{ref}$, $E/R$, $a$, $b$ and $T(0)$ are nearly the same as those obtained using the base case scenario (scenario A), but the estimates of $Q_C$ and $Q_T$ tend to be larger and more biased than in the base case. The corresponding extended AMLE parameter estimates in scenario F are more biased and have a larger variability than those obtained using the base case. The simulation results from scenario F suggest that AEM is more robust than AMLE to imperfect knowledge about measurement noise variances.

We also attempted to use the AEM algorithm for a more-difficult parameter estimation problem (scenario G) wherein parameters were estimated using only the temperature measurements, with the concentration as an unmeasured state. The estimation worked well if $Q_C$ was assumed known (results shown in Table 3.3), but poor estimates were obtained for $Q_C$ when both of the disturbance intensities were estimated together using the limited data (not shown).

The results in Table 3.3 and Figures 3.4 and 3.5 reveal that the AEM parameter estimation algorithm was more effective than extended AMLE for the CSTR example studied. Although AEM did encounter some difficulties with convergence to local minima, the parameter values obtained are more reliable than those from extended AMLE and CTSM. Because AEM uses a single objective function to estimate the model parameters and disturbance intensities, computation times tend to be shorter than for extended AMLE. Results in Table 3.3 suggest that AEM is more robust than extended AMLE to imperfect knowledge about noise variances and that AEM has superior performance when there are unmeasured state variables.

One shortcoming of the AEM method compared to more complex MCMC-based MLE algorithms[38,39,48] is that AEM introduces additional approximations when computing the likelihood function. The main approximations correspond to using only the mode rather than many particles and using B-spline approximations for the state trajectories in the expectation step of the EM algorithm. These approximations save computational effort, but may cause some of the bias observed in Figure 3.5 (a and d). The bias may also be due to the limited number of data points used for parameter estimation. Since maximum likelihood methods only result in unbiased parameters in the limit when the data size

goes to infinity, it will be important to compare AEM with MCMC methods.[13] AEM should require considerably less computation time than MCMC-based EM techniques since AEM does not require drawing samples from high-dimensional and complex probability density functions.[18] The current AEM method was developed assuming the measurement noise variance is perfectly known. It will be beneficial to develop more advanced approximate maximum likelihood methods that can estimate measurement noise variances along with disturbance intensities, for situations where the modeler has rich dynamic data, but limited knowledge of measurement noise.

### 3.6 Summary and Conclusions

A method for estimating parameters and process disturbance intensities in nonlinear SDE models is proposed. This approximate maximum likelihood method is an extension of the AMLE method previously proposed by Varziri et al.[3,52] The new Approximate Expectation Maximization (AEM) objective function permits modelers to estimate process disturbance intensities along with the model parameters. The AEM method efficiently approximates the expectation step of the EM algorithm using B-spline state trajectories.

A two-state nonlinear CSTR model with stochastic disturbances and measurement noise was used to test the AEM methodology. Four fundamental model parameters, an initial state and two process disturbance intensities were estimated. Parameter and disturbance intensity estimates were compared with those from Extended AMLE and the CTSM software of Kristensen using simulated data. The resulting AEM parameter estimates are less biased and more precise than the corresponding estimates obtained using Extended

AMLE and CTSM. For the example studied, AEM is also more robust to poor initial guesses than AMLE and CTSM. AEM performed relatively well using imperfect knowledge about measurement noise variances and when only one of the states was measured.

The AEM method was easier to set up and converged faster than extended AMLE because the AEM method does not require successive optimizations using inner and outer objective functions. Since the model parameters and disturbance intensities are estimated using a common Maximum Likelihood objective function, it is relatively easy to obtain approximate confidence intervals for disturbance intensities using AEM.

Some of the benefits of the AEM method that may be attractive to developers of dynamic models are: i) simplicity of implementation, ii) reliable estimates of model parameters, initial conditions and disturbance intensities, iii) efficient handling of unknown initial states, iv)ability to handle unmeasured state variables. In particular, estimates of disturbance intensities can provide modelers with information about the degree of mismatch and the magnitude of unmeasured disturbances in their simplified models. This information will be helpful when implementing on-line state and parameter estimation schemes for process monitoring and control, because the assumptions about measurement errors and process disturbances are consistent with assumptions used in Kalman filtering.[61]

In future, it will be desirable to extend the AEM algorithm so that unknown measurement noise variances can be estimated when sufficient data are available, and to test the AEM methodology using larger-scale estimation problems. It will also be important to compare AEM results with those from recently developed MLE-based

methods that use MCMC techniques for parameter estimation in stochastic differential equation models.[35,38,39] It is expected that AEM computation times will be significantly lower than the times required using MCMC methods, particularly for larger-scale problems, because AEM does not require sampling from high dimensional probability density functions. It will be interesting to determine whether the additional B-spline and mode approximations used to develop AEM will result in any significant degradation in the quality of parameter estimates when compared with MCMC methods. In future, it will be desirable to investigate the convergence of AEM algorithm.

### 3.7 References


1. N. R. Kristensen, H. Madsen, **2003**.

2. N. R. Kristensen, H. Madsen, S. B. Jørgensen, *Automatica.* **2004**, *40*, 225-237.

3. M. Varziri, K. McAuley, P. McLellan, *Can. J. Chem. Eng.* **2008**, *86*, 828-837.

4. P. Érdi, J. Tóth, *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*, Manchester University Press, Manchester **1989**. p.146 .

5. R. King, *Chem.Eng.Commun.* **1974**, *1*, 221-237.

6. K. A. McLean, K. B. McAuley, *Can. J. Chem. Eng.* **2012**, *90*, 351-366.

7. S. Wu, K. McAuley, T. Harris, *Can. J. Chem. Eng.* **2011**, *89*, 325-336.

8. S. Wu, K. McAuley, T. Harris, *Can. J. Chem. Eng.* **2011**, *89*, 148-158.

9. L. Gagnon, J. MacGregor, *Can. J. Chem. Eng.* **2009**, *69*, 648-656.

10. R. Jones, J. MacGregor, K. Murphy, *Environ. Monit. Assess.* **1989**, *13*, 271-282.

11. J. N. Nielsen, H. Madsen, P. C. Young, *Annl. Rev. Control.* **2000**, *24*, 83-94.

12. E. Lindström, *Ann. Oper. Res.* **2007**, *151*, 269-288.

13. L. Ljung, *System identification*, Prentice-Hall, NJ **1987**, p.182 .

14. S. Jang, R. Gopaluni, *Chem. Eng. Sci.* **2011**, *66*, 2774-2787.

15. A. Doucet, N. De Freitas, N. Gordon, *An introduction to sequential Monte Carlo methods*. Springer, New York **2001**, p. 14.

16. A. Doucet, V. B. Tadić, *Ann. Inst. Statis. Math.* **2003**, *55*, 409-422.

17. J. B. Rawlings, B. R. Bakshi, *Comput. Chem. Eng.* **2006**, *30*, 1529-1541.

18. S. S. Jang, H. De la Hoz, A. Ben-zvi, W. C. McCaffrey, R. B. Gopaluni, *Can. J. Chem. Eng.* **2011**, *90*, 690-702.

19. J. Jaeger, P. Lambert, AP Technical report 11001, Institut de Statistique, Biostatistique et Sciences Actuarielles, **2011**.

20. Y. Aït-Sahalia, *Econometrica.* **2003**, *70*, 223-262.

21. H. Singer, *Comput. Stat.* **2006**, *21*, 385-397.

22. H. Singer, *J. Comp. Graph. Stat.* **2002**, *11*, 972-995.

23. A. W. Lo, *Econometric Theory.* **1988**, *4(2)*, 231-247.

24. B. Christensen, R. Poulsen, M. Sørensen, CAF, Centre for Analytical Finance, **2001**.

25. B. L. Pence, H. K. Fathy, *American Control Conference (ACC)*, IEEE, USA, June, **2010**, 2144-2151.

26. J. Heald, J. Stark, *Phys. Rev. Lett.* **2000**, *84*, 2366-2369.

27. R. H. Shumway, D. S. Stoffer, *J. Time Ser. Anal.* **2008**, *3*, 253-264.

28. R. B. Gopaluni, Proceedings of the 17th World Congress, *The International Federation of Automatic Control*, Korea, July, **2008**.

29. R. Gopaluni, *Can. J. Chem. Eng.* **2008**, *86*, 1081-1092.

30. B. Delyon, M. Lavielle, E. Moulines, *Ann. Stat.* **1999**, 94-128.

31. S. Donnet, A. Samson, *ESAIM: Probability and Statistics.* **2008**, *12*, 196-218.

32. A. P. Dempster, N. M. Laird, D. B. Rubin, *J. R. Stat. Soc.* **1977**, 1-38.

33. S. Duncan, M. Gyongy, Proceedings of the IEEE International Conference on Control Applications, Germany, October , **2006**, 3312-3317.

34. S. Roweis, Z. Ghahramani, *Learning Nonlinear Dynamical Systems Using the Expectation-Maximization Algorithm*, in *Kalman Filtering and Neural Networks*, S. Haykin, Ed., John Wiley & Sons, New York **2001**, p. 175.

35. G. C. Goodwin, J. C. Aguero, Proceedings of the 44th IEEE Conference on Decision and Control, **2005**, 368-373.

36. Z. Ghahramani, S. T. Roweis, *Advances in neural information processing systems.* **1999**, 431-437.

37. G. Poyiadjis, A. Doucet, S. S. Singh, *Proc of the American Stat. Assoc*, **2005**.

38. T. B. Schön, A. Wills, B. Ninness, 14th IFAC Symposium on System Identification, Australia, **2005**.

39. T. B. Schön, A. Wills, B. Ninness, *Automatica*. **2011**, *47*, 39-49.

40. C. Andrieu, A. Doucet, S. S. Singh, V. B. Tadic, *Proc IEEE*. **2004**, *92*, 423-438.

41. A. Golightly, D. J. Wilkinson, *Interface Focus.* **2011**, *1*, 807-820.

42. S. A. Imtiaz, K. Roy, B. Huang, S. L. Shah, P. Jampana, *Industrial Technology, ICIT* International Conference IEEE. **2006**, 2432-2437.

43. N. Kantas, A. Doucet, S. S. Singh, J. M. Maciejowski, *15th IFAC Symposium on System Identification*. **2009**, *15*, 2432-2437,

44. S. B. Chitralekha, J. Prakash, H. Raghavan, R. Gopaluni, S. L. Shah, *J. Process Control.* **2010**, *20*, 934-943.

45. T. Chen, J. Morris, E. Martin, *J. Process Control.* **2005**, *15*, 665-673.

46. A. H. Jazwinski, *Stochastic processes and filtering theory*, Academic press, New York **1970**, p. 84.

47. P. S. Maybeck, *Stochastic models, estimation and control*, Academic press, New York **1982**. p.76.

48. R. B. Gopaluni, *J. Process Control.* **2010**, *20*, 314-324.

49. C. De Boor, "A Practical guide to splines" *in Applied Mathematical Sciences*, Springer, New York **1978**.

50. A. Poyton, M. S. Varziri, K. B. McAuley, P. McLellan, J. O. Ramsay, *Comput. Chem. Eng.* **2006**, *30*, 698-708.

51. B. W. Silverman, J. O. Ramsay, *Functional Data Analysis*. Springer, New York **2005**, p. 49.

52. M. Varziri, A. Poyton, K. McAuley, P. McLellan, J. Ramsay, *Comput. Chem. Eng.* **2008**, *32*, 3011-3022.

53. G. C. Wei, M. A. Tanner, *J. Am. Statist. Assoc.* **1990**, *85*, 699-704.

54. T. E. Marlin, T. Marlin, *Process control: designing processes and control systems for dynamic performance*, McGraw-Hill, New York **1995**, p.76.

55. A. Wächter, L. T. Biegler, *Math. Program.* **2006**, *106*, 25-57.

56. R. Fourer, D. M. Gay, B. W. Kernighan, *A Modeling Language for Mathematical Programming*, 2nd ed., Thomson Brooks/Cole, Florence, KY **2003**.

57. H. Karimi, *Parameter Estimation Techniques for Fundamental Dynamic Models with Limited Data*, *Process Disturbances and Modeling Errors*, Ph.D Thesis, Queen's University, Canada, **2013**.

58. S. M. Kay, *Fundamentals of Statistical Signal Processing*, Prentice Hall, Vol. I, **1998**.

59. M. S. Varziri, K. B. McAuley, P. J. McLellan, *Ind. Eng. Chem. Res.* **2008**, *47*, 380-393.

60. J. Z. Huang, C. O. Wu, L. Zhou, *Statistica. Sinica.* **2004**, *14*, 763-788.

61. K. McAuley, J. MacGregor, *AICHE J.* **1991**, *37*, 825-835.

# Chapter 4

# An Approximate Expectation Maximization Algorithm for Estimating Parameters, Noise Variances and Stochastic Disturbance Intensities in Nonlinear Dynamic Models

## 4.1 Abstract

In Chapter 3, AEM was developed for estimating model parameters and disturbance intensities for nonlinear dynamic systems that are described by Stochastic Differential Equations (SDEs), assuming measurement noise variances are known. In this chapter, an algorithm is proposed for simultaneous estimation of model parameters, process disturbance intensities and measurement noise variances for SDE models. The proposed Fully-Laplace Approximation Expectation Maximization (FLAEM) algorithm uses an iterative approach wherein, in the first step, the model parameters are estimated using the Approximate Maximum Likelihood Estimation (AMLE) objective function developed by Varziri et al.[1], assuming disturbance intensities and noise variances are known. In the second step, process disturbance intensities and measurement noise variance estimates are updated using expressions that rely on the Fully-Laplace Approximation (FLA) in the Expectation Maximization (EM) algorithm. The proposed FLAEM method is illustrated using a nonlinear two-state Continuous Stirred Tank Reactor (CSTR) example. The effectiveness of the FLAEM algorithm is compared with a maximum-likelihood based method proposed by Kristensen et al.[2] For the CSTR example studied, FLAEM provides

more accurate parameter estimates and is more robust to poorly known initial guesses of parameters and to smaller data sets. This chapter has been submitted as a journal paper to *Industrial Engineering and Chemistry Research.*

## 4.2 Introduction

Many chemical processes are modeled using ordinary differential equations (ODEs) or algebraic equations (AEs) arising from fundamental laws of physics and chemistry.[3-5] However, some chemical engineering processes are better modeled using stochastic differential equations (SDEs) that account for possible modeling imperfections and stochastic process disturbances.[3,6] Stochastic terms that are included in SDE models can result in improved model predictions due to decreased bias in parameter estimates.[3,7] Parameter estimates obtained using SDE models are suitable for on-line process monitoring applications because SDE models account for measurement errors and stochastic process disturbance, the two types of random errors that are accounted for by Extended Kalman Filters (EKFs) and related state estimators.[8,9]

A common method for estimating parameters in SDEs is the maximum likelihood estimation method via the expectation maximization (EM) algorithm.[10-14] The EM algorithm is summarized in section 2.4 of this chapter. In nonlinear systems, the EM algorithm becomes difficult to use because of problems related to the finding the required expected value of the likelihood of the parameters given the states and measurements.[13,15] Approximation methods have been used to simplify the expectation and maximization steps of the EM algorithm. Some of these methods involve: using an EKF,[16-19] Markov Chain Monte Carlo (MCMC) methods that are also known as particle filter methods [13-15,20-26] and approximations using spline-based methods.[1,27] Linearization-based EKF

methods are computationally attractive, but can give biased parameter estimates when there are strong nonlinearities in the system.[28] MCMC methods are asymptotically efficient and consistent and do not require assumptions about the form of the density function.[29] When MCMC methods are used, the required probability density functions are approximated by drawing samples from a target density function.[30] MCMC methods tend to be computationally intensive because a large number of particles may be required to obtain good approximations, especially when the number of states and parameters is large.[15,24] An overview of MCMC techniques and some implementation issues are presented by Kantas et al.[31] and Imtiaz et. al.[32]

Varziri et al.[1] developed an Approximate Maximum Likelihood Estimation (AMLE) method for estimating model parameters in SDEs when both the process disturbance intensity and the measurement noise variance are known. Because modelers often have poor knowledge about the magnitudes of their model mismatch and the size of the stochastic disturbances that will be encountered, Varziri et al.[33] extended their algorithm for estimating stochastic disturbance intensity along with model parameters. They assumed that measurement noise variances are known and used this variance information in a somewhat arbitrary objective function to estimate the disturbance intensities. In our recent work,[27] we derived a more rigorous method for estimating disturbance intensities using an Approximate Expectation Maximization (AEM) objective function. Unfortunately, this AEM methodology requires the measurement variances to be known by the modeler.

In this chapter, we propose a computationally efficient algorithm that can be used to estimate unknown noise variances along with the model parameters and disturbance

intensities. This technique relies on the Fully-Laplace Approximation (FLA) for approximating the multidimensional integrals of the likelihood function required in the EM algorithm. [34,35] Previously, the FLA has also been used for approximating posterior moments and marginal densities[35] and for approximating posterior distributions in Bayesian methods.[34] The FLA has also has been used for joint modeling of survival and longitudinal data via the EM algorithm,[36] for estimating [37] parameters in generalized linear latent variable models[37] and for parameter estimation in nonlinear mixed-effects models.[38] To our knowledge, the FLA has not been used until now for parameter estimation in SDE models. Details regarding the FLA are provided in section 2.5.

The remainder of this chapter is organized as follows. First, necessary notation and background information are presented. Next, the EM algorithm and FLA are presented and the FLAEM algorithm is developed. The FLAEM algorithm is then tested using a stochastic nonlinear CSTR simulation study. The estimation results obtained using the FLAEM algorithm are compared with results obtained using the ML-based Continuous-Time Stochastic Modeling (CTSM) method proposed by Kristensen et al.[39] Finally, the performance of FLAEM is tested for the simpler situation when measurement noise variances are known and the FLAEM results are shown to be superior to both CTSM and AEM.

## 4.3 Preliminaries

### 4.3.1 SDE Model and Notation

In this chapter, we illustrate the proposed parameter estimation method using a Multi-Input Multi-Output (MIMO) nonlinear SDE model of the following form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\eta}(t) \tag{4.1.a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{4.1.b}$$

$$\mathbf{y}(t_{m\,r,j}) = \mathbf{g}(\mathbf{x}(t_{m\,r,j}), \mathbf{u}(t_{m\,r,j}), \boldsymbol{\theta}) + \boldsymbol{\varepsilon}(t_{m\,r,j}) \tag{4.1.c}$$

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

$$\mathbf{Q} = \begin{bmatrix} Q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_X \end{bmatrix} \tag{4.2}$$

and $\delta(.)$ is the Dirac delta function.[7] The diagonal elements of $\mathbf{Q}$ are referred to as disturbance intensities (i.e., $\mathbf{Q}_{\mathrm{d}} = [Q_1, \ldots, Q_{\mathrm{X}}]^{\mathrm{T}}$). In Equation (4.1.c , $\mathbf{y} \in R^Y$ is vector of measured output variables. Measurement times for the $r$th response ($r$=1…$Y$) are denoted by $t_{m\,r,j}$ ($j$ = 1…$N_r$) and $N_r$ is the number of measurements of $r$th response variable. $\mathbf{g} \in R^Y$ is a vector of nonlinear mappings and $\boldsymbol{\varepsilon} \in R^Y$ is the zero-mean random measurement error. Assume that errors in measurements made at any sampling time $t_{m\,r,j}$ ($j$ = 1…$N_r$) are independent so that their covariance matrix is:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_Y^2 \end{bmatrix} \tag{4.3}$$

In some dynamic models, the initial conditions, $\mathbf{x}_0$, for the state variables are poorly known because they have some associated measurement noise. We assume that the

measurements for initial conditions of the state variables are contained in vector $\mathbf{x}_{m0}$ and that these measurements are normally distributed with mean $\mathbf{E}\{\mathbf{x}_{m0}\} = \mathbf{x}_0$ and $\text{cov}\{\mathbf{x}_{m0}\} = \mathbf{S}_{m0}$.

Consider a vector $\mathbf{Y_m}$ that contains all of the stacked measured values: $\mathbf{Y_m} = [y_1(t_{m1,1}) \ldots y_1(t_{m1,N_1}) \ldots y_Y(t_{mY,1}) \ldots y_Y(t_{mY,N_Y})]^{\mathrm{T}}$.

Similarly, $\mathbf{X_m} = [x_1(t_{m1,1}) \ldots x_1(t_{m1,N_1}) \ldots x_Y(t_{mY,1}) \ldots x_Y(t_{mY,N_Y})]^{\mathrm{T}}$ is a stacked vector of state values at the measurement times, and $\mathbf{U_m}$ and $\boldsymbol{\varepsilon_m}$ are corresponding stacked vectors for the input variables and random errors:

$$\mathbf{Y_m} = \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon_m} \tag{4.4}$$

In Equation (4.1), a discrete-time white-noise sequence is used to approximate and implement the continuous stochastic disturbances $\boldsymbol{\eta}(t)$, where the corresponding discrete process is a series of random step functions with a sampling interval $\Delta t$ and covariance:[7,40]

$$\mathrm{E}\{\boldsymbol{\eta}(j_1\Delta t)\boldsymbol{\eta}^{\mathrm{T}}(j_2\Delta t)\} = \begin{cases} \dfrac{\mathbf{Q}}{\Delta t} & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases} \tag{4.5}$$

where $j_1$ and $j_2$ are positive integers corresponding to the times at which the independent random shocks occur.

Denote $\mathrm{p_d}$ as the diagonal elements of the covariance matrix (i.e., $\mathrm{p_d} = [\sigma_1^2, \ldots, \sigma_Y^2]^{\mathrm{T}}$). Let $\boldsymbol{\zeta} = [\boldsymbol{\theta}^{\mathrm{T}}, \mathbf{x_0}^{\mathrm{T}}, \mathbf{Q_d}^{\mathrm{T}}, \boldsymbol{\Sigma_d}^{\mathrm{T}}]^{\mathrm{T}}$ be the vector of unknown parameters in the SDE model, which includes the model parameters $\boldsymbol{\theta}$ and the unknown initial conditions, along with disturbance intensities $\mathbf{Q}$ and the unknown noise variances $\mathrm{p}$. To simplify the notation,

derivations in this chapter are developed assuming that $n$ measurements are available for each response. However, a derivation for the more general case where $N_r$ measurements are available for the $r$th response is also shown in Appendix 3.A.

### 4.3.2 B-Spline Basis Functions

B-splines basis functions are used to approximate continuous functions and variables. $M$th order B-splines basis functions are piecewise polynomials that are positive within $M$ intervals and zero elsewhere.[41-43]

The $s$th state of the SDE model in Equation (4.1) can be approximated by a linear combination of $c_s$ B-splines:[41,44]

$$x_{\sim s}(t) = \sum_{l=1}^{c_s} \beta_{s,l}\varphi_{s,l}(t) \quad \text{for } s=1,\ldots,X \tag{4.6}$$

where $\beta_{s,l}$ is a B-spline coefficient and $\varphi_{s,l}(t)$ is the corresponding B-spline basis function. The subscript ~ is used to indicate that the state trajectories are being approximated using empirical spline curves. In matrix form, Equation (4.6) is:

$$\mathbf{x}_{\sim}(t) = \mathbf{\Phi}(t)\mathbf{B} \tag{4.7}$$

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

$$\mathbf{\Phi}(t) = \begin{bmatrix} \boldsymbol{\varphi}_1^T(t) & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}_2^T(t) & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\varphi}_X^T(t) \end{bmatrix} \tag{4.8}$$

and

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_X \end{bmatrix} \tag{4.9}$$

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

$$\boldsymbol{\beta}_s = [\ _{s,1}, \ldots, \ _{s,c_s}]^{\mathrm{T}} \qquad \text{for } s=1,\ldots,X \tag{4.10}$$

An advantage of using B-spline basis functions for approximating the state variables in dynamic models is that they can be easily differentiated with respect to time:

$$\dot{x}_{\sim s}(t) = \sum_{l=1}^{c_s} \beta_{s,l} \dot{\varphi}_{s,l}(t) \qquad \text{for } s=1,\ldots,X \tag{4.11}$$

where $\dot{\varphi}_{s,l}(t)$ is a simple polynomial expression. As a result, B-splines can be used to convert differential equations to algebraic equations.[41,44] For example, when B-spline approximations are used, Equation (4.1.a) becomes:

$$\dot{\boldsymbol{\Phi}}(t)\mathbf{B} = \mathbf{f}(\boldsymbol{\Phi}(t)\mathbf{B}, \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\eta}(t) \tag{4.12}$$

#### 4.3.3 Approximate Maximum Likelihood Estimation (AMLE) Algorithm

Varziri et al.[1] discretized the SDE in Equation (4.1) to develop an AMLE method for estimating model parameters $\boldsymbol{\theta}$ in SDE models. The discretized form of Equation (4.1) using an Euler approximation is:

$$\mathbf{x}(t_{i-1} + \Delta t) = \mathbf{x}(t_i) = \mathbf{x}(t_{i-1}) + f(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t + \boldsymbol{\eta}_{\mathbf{d}}(t_{i-1})\Delta t \tag{4.13.a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{4.13.b}$$

where $\mathbf{x}(t_i)$ is the value of the state variable at $q$ uniformly-spaced time points $t_i$ , $i=0,..,q$ and $\boldsymbol{\eta}_{\mathbf{d}}(t_{i-1})$ is the discrete-time white-noise process at $q$ uniformly-spaced time points $t_{i-1}$. Consider $\mathbf{X}_{\mathbf{q}} = [\mathbf{x}^{\mathrm{T}}(t_0), \mathbf{x}^{\mathrm{T}}(t_1) \ldots, \mathbf{x}^{\mathrm{T}}(t_q)]^{\mathrm{T}}$ as the stacked vector of state values at the

discrete times. Varziri et al.[1] assumed that $\mathbf{Q}$ and $\boldsymbol{\Sigma}$ and perfectly known and derived the following analytical expression for the likelihood $-\ln p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\theta})$, while approximating state trajectories by B-spline basis functions:[1]

$$\begin{aligned} J_{\text{AMLE}} = -\ln p(\mathbf{Y_m}, \mathbf{X_{q\sim}} \mid \boldsymbol{\theta}) = & [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{m\sim}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{m\sim}}, \mathbf{U_m}, \boldsymbol{\theta})] \\ & + (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0})^{\mathrm{T}} \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0}) \\ & + \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})] \mathrm{d}t \end{aligned} \tag{4.14}$$

$\mathbf{x}_{\sim}(t)$ and its time derivative $\dot{\mathbf{x}}_{\sim}(t)$ in Equation (4.14) result in an objective function that depends explicitly on the B-spline coefficients $\mathbf{B}$ and model parameters $\boldsymbol{\theta}$. Optimal approximate maximum likelihood estimates for the model parameters $\boldsymbol{\theta}$ can be determined by finding values of $\boldsymbol{\theta}$ and $\mathbf{B}$ that minimize $J_{\text{AMLE}}$.[1] In the current chapter, this AMLE objective function is used as part of a more complicated algorithm for estimating unknown values of $\boldsymbol{\theta}$, $\mathbf{B}$, $\mathbf{x}_0$, $\mathbf{Q}$ and $\mathrm{p}$. This new algorithm is useful for estimating parameters when the modeler does not have prior knowledge of $\mathbf{Q}$ and $\mathrm{p}$.

#### 4.3.4 EM Algorithm

In the EM algorithm, the expected value of the log likelihood of the complete data, given the vector of measurements and values of the parameters $\hat{\boldsymbol{\zeta}}_k$ arising from the current (i.e., *k*th) parameter iteration is calculated in the first step (referred to as the expectation step or E step):[12,45,46]

$$R(\boldsymbol{\zeta}, \hat{\boldsymbol{\zeta}}_k) = \underset{\mathbf{X_q}}{\mathrm{E}} \{\ln[p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta})] \mid \mathbf{Y_m}, \hat{\boldsymbol{\zeta}}_k\} = \int \ln[p(\mathbf{Y_m}, \mathbf{X_q} \mid \mathbf{Y_m}, \boldsymbol{\zeta})] p(\mathbf{X_q} \mid \mathbf{Y_m}, \hat{\boldsymbol{\zeta}}_k) \mathrm{d}\mathbf{X_q} \tag{4.15}$$

Note that the integral in Equation (4.15) is a multidimensional integral with respect to each element of the state vector and that $\hat{\boldsymbol{\zeta}}_k$ contains estimates of $\boldsymbol{\theta}$, $\mathbf{Q}$ and $\rho$ from the previous ($k$th) iteration. In the second step (referred to as the maximization step or M step), this expected value is maximized with respect to $\boldsymbol{\zeta}$:[12,45,46]

$$\hat{\boldsymbol{\zeta}}_{k+1} = \arg\max_{\boldsymbol{\zeta}} R(\boldsymbol{\zeta}, \hat{\boldsymbol{\zeta}}_k) \tag{4.16}$$

Iteration between these two steps continues until convergence is obtained.

#### 4.3.5 Fully-Laplace Approximation

The FLA of the ratio of two related multidimensional integrals is:[34,35,47]

$$\frac{\int G(\boldsymbol{\chi}) \exp\{\cdot(\boldsymbol{\chi})\} d\boldsymbol{\chi}}{\int \exp\{\cdot(\boldsymbol{\chi})\} d\boldsymbol{\chi}} = \frac{\int \exp\{\cdot^*(\boldsymbol{\chi})\} d\boldsymbol{\chi}}{\int \exp\{\cdot(\boldsymbol{\chi})\} d\boldsymbol{\chi}} \approx \left( \frac{\det\left[ \left. \frac{-\partial^2 \cdot(\boldsymbol{\chi})}{\partial\boldsymbol{\chi}\partial\boldsymbol{\chi}^T} \right|_{\boldsymbol{\chi}=\hat{\boldsymbol{\chi}}} \right]}{\det\left[ \left. \frac{-\partial^2 \cdot^*(\boldsymbol{\chi})}{\partial\boldsymbol{\chi}\partial\boldsymbol{\chi}^T} \right|_{\boldsymbol{\chi}=\hat{\boldsymbol{\chi}}^*} \right]} \right)^{1/2} \exp\{\cdot^*(\hat{\boldsymbol{\chi}}^*) - \cdot(\hat{\boldsymbol{\chi}})\} \tag{4.17}$$

where $G(\chi)$ is a positive scalar function, $\cdot(\boldsymbol{\chi})$ is a scalar function and:

$$\cdot^*(\boldsymbol{\chi}) = \ln[G(\boldsymbol{\chi})] + \cdot(\boldsymbol{\chi}) \tag{4.18}$$

In Equation (4.17), $\hat{\boldsymbol{\chi}}$ and $\hat{\boldsymbol{\chi}}^*$ are vectors that maximize $\psi$ and $\psi^*$, respectively.

### 4.4 Development of the FLAEM Algorithm

In this section, an algorithm for estimating the measurement variances $\rho$ and the process intensities $\mathbf{Q}$ along with the model parameters $\boldsymbol{\theta}$ and initial conditions $\mathbf{x}_0$ is developed.

Define:

$$Z = \int_{t_0}^{t_q} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))(\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))^{\mathrm{T}} \, \mathrm{d}\, t \tag{4.19}$$

$$S = (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}))(\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}))^{\mathrm{T}} \tag{4.20}$$

It is shown in Appendix 3.A that, when $\boldsymbol{\theta}$ is assumed to be known, the estimates of the disturbance intensity $\mathbf{Q}$ and the noise variance $\boldsymbol{\Sigma}$ at the $k$+1th iteration are:

$$\mathbf{Q}_{k+1} = \frac{1}{q} \mathrm{E}\{Z \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k\} \tag{4.21}$$

$$\boldsymbol{\Sigma}_{k+1} = \frac{1}{n} \mathrm{E}\{S \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k\} \tag{4.22}$$

The expectations of $Z$ and $S$ conditional on $\mathbf{Y_m}$, $\mathbf{Q}_k$ and $\boldsymbol{\Sigma}_k$ are given by:[38]

$$\mathrm{E}(Z \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k) = \frac{\int Z \; p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta}_k) \, \mathrm{d}\, \mathbf{X_q}}{\int p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta}_k) \, \mathrm{d}\, \mathbf{X_q}} \tag{4.23}$$

$$\mathrm{E}(S \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k) = \frac{\int S \; p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta}_k) \, \mathrm{d}\, \mathbf{X_q}}{\int p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta}_k) \, \mathrm{d}\, \mathbf{X_q}} \tag{4.24}$$

The FLA can be used for calculating the ratios of integrals in Equations (4.23) and (4.24). After substituting the expressions for $\mathrm{E}(Z \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k)$ and $\mathrm{E}(S \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k)$ obtained from the FLA into Equations (4.23) and (4.24), expressions for estimating $\mathbf{Q}$ and $\mathsf{p}$ are (See Appendix 4.A for derivation):

$$\mathbf{Q}_{k+1} = \frac{1}{q}\left(\frac{\det(\mathbf{H_B})}{\det(\mathbf{H_B^Z})}\right)^{1/2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}^{Z}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{Z}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))(\dot{\hat{\mathbf{x}}}_{\sim}^{Z}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{Z}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}}\, \mathrm{d}t$$
$$\exp\{-\frac{1}{2}(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}^{Z}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}^{Z}, \mathbf{U_m}, \boldsymbol{\theta}_k))$$
$$-\frac{1}{2}\int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}^{Z}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{Z}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}_{\sim}^{Z}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{Z}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))\, \mathrm{d}t$$
$$+\frac{1}{2}(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))$$
$$+\frac{1}{2}\int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))\, \mathrm{d}t\} \tag{4.25}$$

$$\boldsymbol{\Sigma}_{k+1} = \frac{1}{n}\left(\frac{\det(\mathbf{H_B})}{\det(\mathbf{H_B^S})}\right)^{1/2} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}^{S}, \mathbf{U_m}, \boldsymbol{\theta}_k))(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}^{S}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}}$$
$$\exp\{-\frac{1}{2}(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}^{S}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}^{S}, \mathbf{U_m}, \boldsymbol{\theta}_k))$$
$$-\frac{1}{2}\int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}^{S}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{S}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}_{\sim}^{S}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{S}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))\, \mathrm{d}t$$
$$+\frac{1}{2}(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))$$
$$+\frac{1}{2}\int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))\, \mathrm{d}t\} \tag{4.26}$$

In Equations (4.25) and (4.26), the Hessian matrices $\mathbf{H_B}$, $\mathbf{H_B^S}$ and $\mathbf{H_B^Z}$ are defined as:

$$\mathbf{H_B} = \left.\frac{\partial^2 J_{\mathrm{AMLE}}}{\partial \mathbf{B} \partial \mathbf{B}^{\mathrm{T}}}\right|_{\mathbf{B}=\hat{\mathbf{B}}} \tag{4.27}$$

$$\mathbf{H_B^Z} = \left.\frac{\partial^2 J_{\sim}^{Z}}{\partial \mathbf{B} \partial \mathbf{B}^{\mathrm{T}}}\right|_{\mathbf{B}=\hat{\mathbf{B}}^{Z}} \tag{4.28}$$

$$\mathbf{H}_{\mathbf{B}}^{S} = \left. \frac{\partial^2 J_{\sim}^{S}}{\partial \mathbf{B} \partial \mathbf{B}^{T}} \right|_{\mathbf{B}=\hat{\mathbf{B}}^{S}} \tag{4.29}$$

$J_{\text{AMLE}}$ in Equation (4.27) is Varziri's AMLE objective function defined in Equation (4.14). $J_{\sim}^{Z}$ and $J_{\sim}^{S}$ in Equations (4.28) and (4.29) are:

$$\begin{aligned} J_{\sim}^{Z} = & -\ln \int_{t_0}^{t_q} (\dot{x}_{\sim 1}(t) - f_1(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^2 \, \mathrm{d}t - \ldots - \ln \int_{t_0}^{t_q} (\dot{x}_{\sim X}(t) - f_X(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^2 \, \mathrm{d}t \\ & + \frac{1}{2} (\mathbf{Y}_{\mathbf{m}} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U}_{\mathbf{m}}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_{\mathbf{m}} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U}_{\mathbf{m}}, \boldsymbol{\theta}_k)) \\ & + \frac{1}{2} \int_{t_0}^{t_q} (\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t \end{aligned} \tag{4.30}$$

$$\begin{aligned} J_{\sim}^{S} = & -\ln \sum_{j=1}^{N_1} [y_1(t_{m1,j}) - g_1(\mathbf{x}_{\sim}(t_{m1,j}), \mathbf{y}(t_{m1,j}), \boldsymbol{\theta}_k)]^2 - \ldots \\ & - \ln \sum_{j=1}^{N_Y} [y_Y(t_{mY,j}) - g_Y(\mathbf{x}_{\sim}(t_{mY,j}), \mathbf{y}(t_{mY,j}), \boldsymbol{\theta}_k)]^2 \\ & + \frac{1}{2} (\mathbf{Y}_{\mathbf{m}} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U}_{\mathbf{m}}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y}_{\mathbf{m}} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U}_{\mathbf{m}}, \boldsymbol{\theta}_k)) \\ & + \frac{1}{2} \int_{t_0}^{t_q} (\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \mathrm{d}t \end{aligned} \tag{4.31}$$

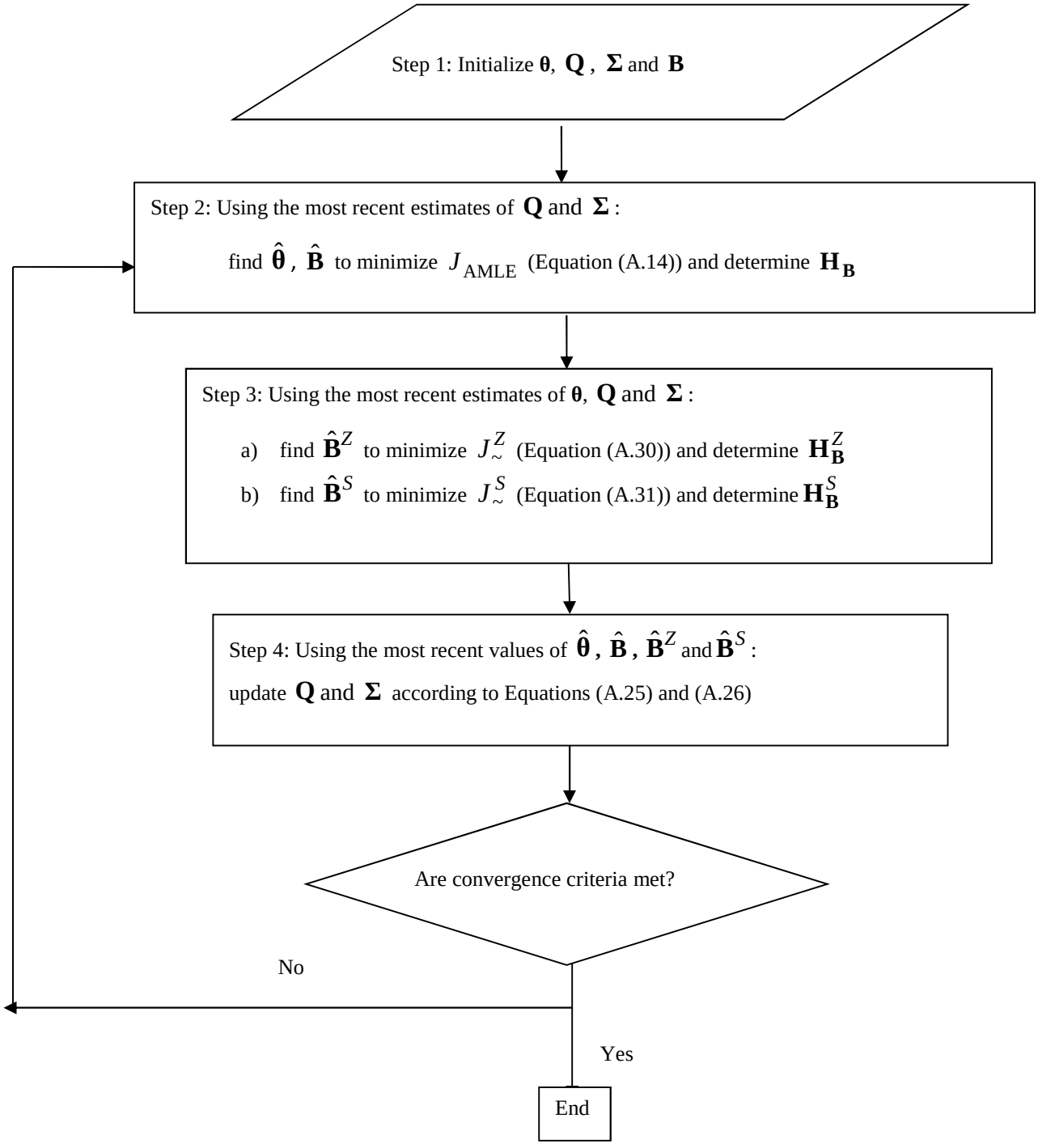
$\hat{\mathbf{B}}$, $\hat{\mathbf{B}}^{Z}$ and $\hat{\mathbf{B}}^{S}$ are vectors of spline coefficients that minimize $J_{\text{AMLE}}$, $J_{\sim}^{Z}$ and $J_{\sim}^{S}$, respectively and $\hat{\mathbf{x}}_{\sim}$, $\hat{\mathbf{x}}_{\sim}^{Z}$ and $\hat{\mathbf{x}}_{\sim}^{S}$ are the corresponding estimated state trajectories.

As shown in Chart 1, an iterative method can be used for estimating all of the parameters ($\boldsymbol{\theta}$, $\mathbf{Q}$, $\boldsymbol{\Sigma}$ and $\mathbf{B}$). Note that the estimate for $\mathbf{x}_0$ is $\mathbf{x}_{\sim 0}$, which can be computed from the estimated spline coefficients. The first step of the FLAEM algorithm is to initialize all of

the parameters ($\boldsymbol{\theta}$, $\mathbf{Q}$, $\boldsymbol{\Sigma}$ and $\mathbf{B}$). The second step is to minimize the AMLE objective function (Equation (4.14)) with respect to $\boldsymbol{\theta}$ and $\mathbf{B}$ to find $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{B}}$, using the fixed values of $\mathbf{Q}$ and $\boldsymbol{\Sigma}$ from their most recent updates. The third step is to minimize Equations (4.30) and (4.31) with respect to $\mathbf{B}$ to find $\hat{\mathbf{B}}^Z$ and $\hat{\mathbf{B}}^S$, using the fixed values of $\boldsymbol{\theta}$, $\mathbf{Q}$ and $\boldsymbol{\Sigma}$ from their most recent updates. The fourth step is to update $\mathbf{Q}$ and $\boldsymbol{\Sigma}$ from Equations (25) and (26), using the most recent values of $\hat{\boldsymbol{\theta}}, \hat{\mathbf{B}}, \hat{\mathbf{B}}^Z$ and $\hat{\mathbf{B}}^S$. The FLAEM algorithm iterates between steps two, three and four until convergence is obtained.

Note that the vector of spline coefficients $\mathbf{B}$ (i.e., $\boldsymbol{\zeta} = [\boldsymbol{\theta}^{\mathrm{T}}, \mathbf{Q}_{\mathbf{d}}^{\mathrm{T}}, \boldsymbol{\Sigma}_{\mathbf{d}}^{\mathrm{T}}, \mathbf{B}^{\mathrm{T}}]^{\mathrm{T}}$) was not added to the parameter vector $\boldsymbol{\zeta}$ before deriving Equations (4.25) and (4.26). If $\mathbf{B}$ is included in the parameter vector, conditions required for approximating the integrals in Equations (4.23) and (4.24) does not hold anymore.[34,45,37]

Chart 1.The FLAEM algorithm



### 4.4 Illustrative Simulation Study: Nonlinear CSTR Stochastic Model

In this section, a two-state nonlinear CSTR model[1,5] is used to illustrate the application of the FLAEM algorithm for parameter estimation in SDEs. The two SDEs that describe dynamic changes in the concentration of reactant A and reactor temperature are:

$$\frac{\mathrm{d}C_A(t)}{\mathrm{d}t} = \frac{F(t)}{V}(C_{A0}(t) - C_A(t)) - k_r(T(t))C_A(t) + \dagger_C(t) \tag{4.32.a}$$

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{F(t)}{V}(T_0(t) - T(t)) + UA(T(t) - T_{in}(t)) + ¥\, k_r(T(t))C_A(t) + \dagger_T(t) \tag{4.32.b}$$

$$y_C(t_{mC,j}) = C_A(t_{mC,j}) + \varepsilon_C(t_{mC,j}) \quad \text{for } j = 1\ldots n_C \tag{4.32.c}$$

$$y_T(t_{mT,j}) = T(t_{mT,j}) + \varepsilon_T(t_{mT,j}) \quad \text{for } j = 1\ldots n_T \tag{4.32.d}$$

$C_A(0)$=1.569 kmol.m$^{-3}$ (4.32.e)

$T(0)$=341.37 K (4.32.f)

where

$$k_r(T) = k_{ref}\ \exp\left(-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \tag{4.33}$$

$$UA(F_c) = \frac{aF_c^{b+1}}{V\rho c_p\left(F_c + \dfrac{aF_c^b}{2\rho_c c_{pc}}\right)} \tag{4.34}$$

$$\gamma = \frac{(-\Delta H_{rxn})}{\rho c_p} \tag{4.35}$$

In Equations (4.32.a) and (4.32.b):

$$\mathrm{E}\{\dagger_C(t_i)\dagger_C(t_j)\} = Q_C¢(t_i - t_j) \tag{4.36}$$

$$\mathrm{E}\{\dagger_T(t_i)\dagger_T(t_j)\} = Q_T¢(t_i - t_j) \tag{4.37}$$

In Equations (4.32.c) and (4.32.d), $£_C(t_{mC,j})\ \ j=1\ldots n_C$ and $£_T(t_{mT,j})\ \ j=1\ldots n_T$ are Gaussian measurement errors with variances $±_C^2$ and $±_T^2$. The concentration $C_A$ is measured $n_C$ times and the temperature $T$ is measured $n_T$ times using equally-spaced sampling intervals. We assume that $\eta_C$, $\eta_T$, $\varepsilon_C$ and $\varepsilon_T$ are independent. The model inputs are: the feed flow rate $F$, the inlet concentration $C_{A0}$, the inlet temperature $T_0$, the coolant inlet temperature $T_{in}$ and the flow rate of coolant to the cooling coil, $F_c$. The known constants for this CSTR model are given in Table 4.1.[5]

**Table 4.1 Model constants[5]**

| Model Constants | Value | Units |
|---|---|---|
| $c_p$ | 4186.8 | $J\cdot kg^{-1}\cdot K^{-1}$ |
| $c_{pc}$ | 4186.8 | $J\cdot kg^{-1}\cdot K^{-1}$ |
| $T_{ref}$ | 350 | K |
| $V$ | 1 | $m^3$ |
| $\rho$ | 1000 | $kg\cdot m^{-3}$ |
| $\Delta H_{rxn}$ | $-544.154\times10^3$ | $J\cdot kmol^{-1}$ |

The handling of known and unknown initial conditions is illustrated in this example by assuming that the initial concentration $C_A(0)$ is perfectly known and the initial temperature $T(0)$ is unknown, but has been measured with a variance of $S_T^2$ =5.0 $K^2$. Since the true value of the initial temperature $T(0)$ is unknown, it must be estimated. However, $T(0)$ does not need to be included explicitly in the list of optimizer decision

variables because the temperature trajectory is computed using the B-spline basis functions so that T(0) corresponds to $\beta_{T,1}$. Since $C_A(0)$ is perfectly known, the first spline coefficient $\beta_{C,1}$ must be fixed at 1.569 kmol·m$^{-3}$.

The model parameters to be estimated are kinetic parameters $k_{ref}$ and $E/R$, and heat-transfer parameters $a$ and $b$. In vector form, $\boldsymbol{\theta}_{\text{CSTR}} = [k_{ref}, E/R, a, b]^T$. In the majority of the situations studied in this chapter, the disturbance intensities $Q_C$ (for the material balance SDE) and $Q_T$ (for the energy balance SDE) and the measurement noise variances $\pm_C^2$ and $\pm_T^2$ are assumed to be unknown. As a result, the complete vector of parameters to be estimated is $\boldsymbol{\zeta}_{\text{CSTR}} = [k_{ref}, E/R, a, b, Q_C, Q_T, \pm_C^2, \pm_T^2]^T$. In a few simulations, however, the case where $\pm_C^2$ and $\pm_T^2$ are perfectly known is also considered to permit comparisons of the FLAEM algorithm with our previously developed AEM technique.

The CSTR model (Equation (4.32)) was simulated in MATLAB using the "ode45" solver. The duration of each simulated experiment is 64 minutes. The corresponding input trajectories are shown in Figure 4.1.[1,44]

The stochastic white noise terms ($\eta_C(t)$ and $\eta_T(t)$) were simulated using band-limited white-noise blocks with a sample time of 0.5 minutes, which is approximately 10 times smaller than the dominant time constant of the CSTR system. Simulated data affected by Gaussian measurement errors and stochastic process disturbances were generated using the true parameter values from Marlin[5] shown in Table 4.2

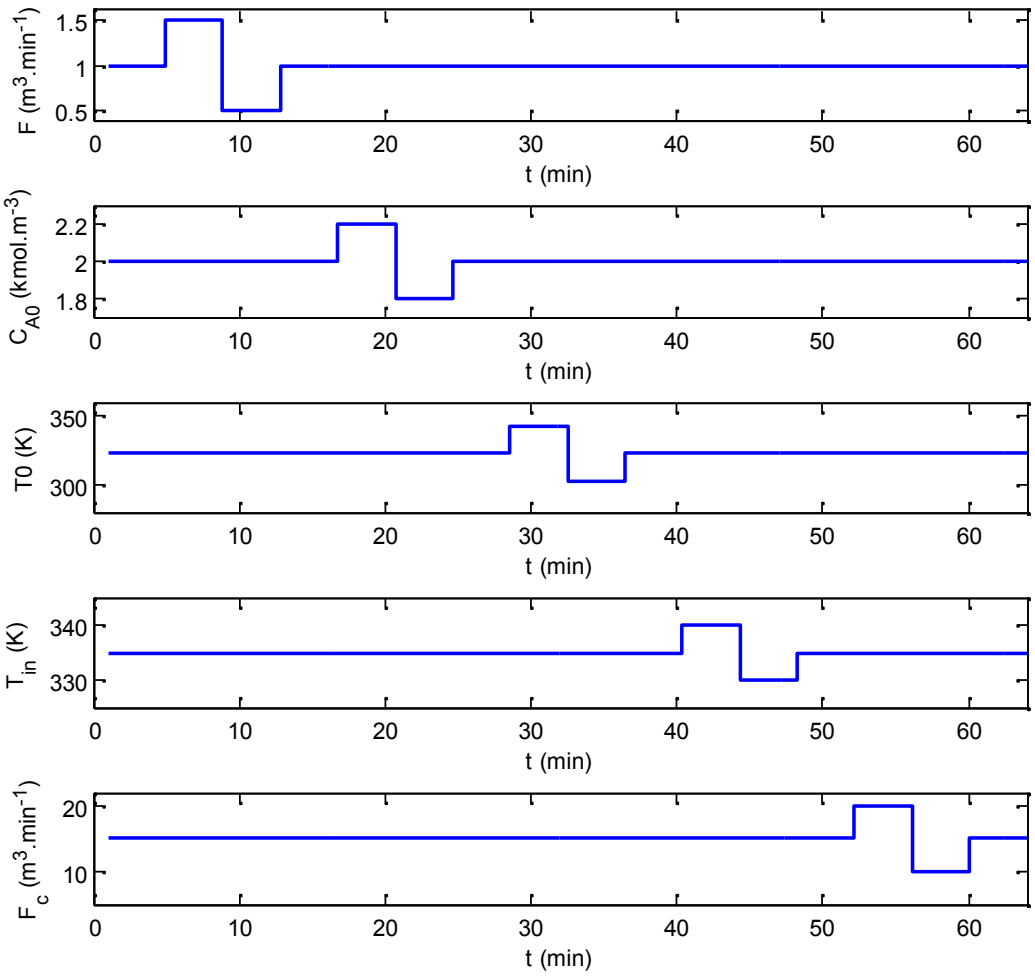



**Figure 4.1 Input trajectories for nonlinear CSTR[44]**

The appropriate objective function for estimating the model parameters $\boldsymbol{\theta}_{CSTR}$ and the B-spline coefficients in the CSTR model is:

$$J_{\text{AMLE,CSTR}} = \frac{1}{\pm_{C,k}^2}\sum_{j=1}^{n_C}\left(y_C(t_{m\,C,j}) - C_{A\sim}(t_{m\,C,j})\right)^2 + \frac{1}{\pm_{T,k}^2}\sum_{j=1}^{n_T}\left(y_T(t_{m\,T,j}) - T_{\sim}(t_{m\,T,j})\right)^2$$
$$+ \frac{(T_m(0) - T_{\sim}(0))^2}{S_T^2}$$
$$+ \frac{1}{Q_{C,k}}\int_{t_0}^{t_{nC}}\left(\frac{\mathrm{d}\,C_{A\sim}(t)}{\mathrm{d}\,t} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r(T_{\sim}(t))C_{A\sim}(t)\right)^2 \mathrm{d}\,t +$$
$$+ \frac{1}{Q_{T,k}}\int_{t_0}^{t_{nT}}\left(\frac{\mathrm{d}\,T_{\sim}(t)}{\mathrm{d}\,t} - \frac{F(t)}{V}(T_0(t) - T_{\sim}(t)) - UA[T_{\sim}(t) - T_{in}(t)] - ¥k_r(T_{\sim}(t))C_{A\sim}(t)\right)^2 \mathrm{d}\,t \tag{4.38}$$

The third term on the right-hand side penalizes deviations of the estimated initial temperature from the corresponding measurement. Note that there is no similar term for the initial concentration, because it is assumed to be perfectly known by the modeler. The first step of the FLAEM algorithm is to initialize all parameters. In the second step, $J_{\text{AMLE,CSTR}}$ is minimized with respect to the model parameters $\boldsymbol{\theta}_{\text{CSTR}}$ and spline coefficients $\mathbf{B}_{\text{CSTR}}$, assuming that the disturbance intensities and noise variances are known:

$$\hat{\boldsymbol{\theta}}_{\text{CSTR}}, \hat{\mathbf{B}}_{\text{CSTR}} = \underset{\boldsymbol{\theta}_{\text{CSTR}}, \mathbf{B}_{\text{CSTR}}}{\arg\min}\; J_{\text{AMLE,CSTR}} \tag{4.39}$$

where $\mathbf{B}_{\text{CSTR}} = [\beta_C, \beta_T]^T$

The appropriate objective functions for the third step of the FLAEM algorithm for the CSTR model are:

$$J^{Z}_{\mathrm{CSTR}} = -\ln[\int_{t_0}^{t_{n_C}} \left( \frac{\mathrm{d}\, C_{A\sim}(t)}{\mathrm{d}\, t} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r(T_\sim(t))C_{A\sim}(t) \right)^2 \mathrm{d}\, t]$$
$$- \ln[\int_{t_0}^{t_{n_T}} \left( \frac{\mathrm{d}\, T_\sim(t)}{\mathrm{d}\, t} - \frac{F(t)}{V}(T_0(t) - T_\sim(t)) - UA[T_\sim(t) - T_{in}(t)] - ¥k_r(T_\sim(t))C_{A\sim}(t) \right)^2 \mathrm{d}\, t]$$
$$+ \frac{1}{\pm_C^2} \sum_{j=1}^{n_C} \left( y_C(t_{m\,C,j}) - C_{A\sim}(t_{m\,C,j}) \right)^2 + \frac{1}{\pm_T^2} \sum_{j=1}^{n_T} \left( y_T(t_{m\,T,j}) - T_\sim(t_{m\,T,j}) \right)^2$$
$$+ \frac{1}{Q_C} \int_{t_0}^{t_{n_C}} \left( \frac{\mathrm{d}\, C_{A\sim}(t)}{\mathrm{d}\, t} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r(T_\sim(t))C_{A\sim}(t) \right)^2 \mathrm{d}\, t$$
$$+ \frac{1}{Q_T} \int_{t_0}^{t_{n_T}} \left( \frac{\mathrm{d}\, T_\sim(t)}{\mathrm{d}\, t} - \frac{F(t)}{V}(T_0(t) - T_\sim(t)) - UA[T_\sim(t) - T_{in}(t)] - ¥k_r(T_\sim(t))C_{A\sim}(t) \right)^2 \mathrm{d}\, t$$

(4.40)

$$J^{S}_{\mathrm{CSTR}} = -\ln[\sum_{j=1}^{n_C} \left( y_C(t_{m\,C,j}) - C_{A\sim}(t_{m\,C,j}) \right)^2] - \ln[\sum_{j=1}^{n_T} \left( y_T(t_{m\,T,j}) - T_\sim(t_{m\,T,j}) \right)^2]$$
$$+ \frac{1}{\pm_C^2} \sum_{j=1}^{n_C} \left( y_C(t_{m\,C,j}) - C_{A\sim}(t_{m\,C,j}) \right)^2 + \frac{1}{\pm_T^2} \sum_{j=1}^{n_T} \left( y_T(t_{m\,T,j}) - T_\sim(t_{m\,T,j}) \right)^2$$
$$+ \frac{1}{Q_C} \int_{t_0}^{t_{n_C}} \left( \frac{\mathrm{d}\, C_{A\sim}(t)}{\mathrm{d}\, t} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r(T_\sim(t))C_{A\sim}(t) \right)^2 \mathrm{d}\, t$$
$$+ \frac{1}{Q_T} \int_{t_0}^{t_{n_T}} \left( \frac{\mathrm{d}\, T_\sim(t)}{\mathrm{d}\, t} - \frac{F(t)}{V}(T_0(t) - T_\sim(t)) - UA[T_\sim(t) - T_{in}(t)] - ¥k_r(T_\sim(t))C_{A\sim}(t) \right)^2 \mathrm{d}\, t$$

(4.41)

In the third step, $J^{Z}_{\mathrm{CSTR}}$ and $J^{S}_{\mathrm{CSTR}}$ are minimized with respect to spline coefficients $\mathbf{B}_{\mathrm{CSTR}}$ assuming that the complete parameter vector $\boldsymbol{\zeta}_{\mathrm{CSTR}}$ is known:

$$\hat{\mathbf{B}}^{Z}_{\mathrm{CSTR}} = \arg\min_{\mathbf{B}_{\mathrm{CSTR}}} J^{Z}_{\mathrm{CSTR}} \tag{4.42}$$

$$\hat{\mathbf{B}}^{S}_{\mathrm{CSTR}} = \arg\min_{\mathbf{B}_{\mathrm{CSTR}}} J^{S}_{\mathrm{CSTR}} \tag{4.43}$$

In the fourth step, $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ are updated:

$$Q_{C,k+1} = \frac{1}{q}\left(\frac{\det(\mathbf{H}_{\boldsymbol{\beta}C})}{\det(\mathbf{H}_{\boldsymbol{\beta}C}^{Z})}\right)^{1/2} \exp[\, J_{\text{CSTR}}^{Z}\big|_{\hat{C}_{A\sim}^{Z},\hat{T}_{\sim}^{Z}} - J_{\text{AMLE,CSTR}}\big|_{\hat{C}_{A\sim},\hat{T}_{\sim}}\,] \tag{4.44}$$

$$\pm_{C,k+1}^{2} = \frac{1}{n}\left(\frac{\det(\mathbf{H}_{\boldsymbol{\beta}C})}{\det(\mathbf{H}_{\boldsymbol{\beta}C}^{S})}\right)^{1/2} \exp[\, J_{\text{CSTR}}^{S}\big|_{\hat{C}_{A\sim}^{S},\hat{T}_{\sim}^{S}} - J_{\text{AMLE,CSTR}}\,\hat{C}_{A\sim},\hat{T}_{\sim}\,] \tag{4.45}$$

$$Q_{T,k+1} = \frac{1}{q}\left(\frac{\det(\mathbf{H}_{\boldsymbol{\beta}T})}{\det(\mathbf{H}_{\boldsymbol{\beta}T}^{Z})}\right)^{1/2} \exp[\, J_{\text{CSTR}}^{Z}\big|_{\hat{C}_{A\sim}^{Z},\hat{T}_{\sim}^{Z}} - J_{\text{AMLE,CSTR}}\big|_{\hat{C}_{A\sim},\hat{T}_{\sim}}\,] \tag{4.46}$$

$$\pm_{T,k+1}^{2} = \frac{1}{n}\left(\frac{\det(\mathbf{H}_{\boldsymbol{\beta}T})}{\det(\mathbf{H}_{\boldsymbol{\beta}T}^{S})}\right)^{1/2} \exp[\, J_{\text{CSTR}}^{S}\big|_{\hat{C}_{A\sim}^{S},\hat{T}_{\sim}^{S}} - J_{\text{AMLE,CSTR}}\big|_{\hat{C}_{A\sim},\hat{T}_{\sim}}\,] \tag{4.47}$$

where $q$ =128 is the number of discrete random shocks used to generate the disturbance sequences. $\hat{C}_{A\sim}$, $\hat{T}_{\sim}$, $\hat{C}_{A\sim}^{Z}$, $\hat{T}_{\sim}^{Z}$, $\hat{C}_{A\sim}^{S}$ and $\hat{T}_{\sim}^{S}$ are estimated state trajectories corresponding to estimated B-splines coefficients $\hat{\boldsymbol{\beta}}_C, \hat{\boldsymbol{\beta}}_T, \hat{\boldsymbol{\beta}}_C^{Z}$, $\hat{\boldsymbol{\beta}}_T^{Z}, \hat{\boldsymbol{\beta}}_C^{S}$ and $\hat{\boldsymbol{\beta}}_T^{S}$, respectively. In Equations (44)-(47) Hessian matrixes $\mathbf{H}_{\boldsymbol{\beta}C}$, $\mathbf{H}_{\boldsymbol{\beta}C}^{Z}$, $\mathbf{H}_{\boldsymbol{\beta}C}^{S}$, $\mathbf{H}_{\boldsymbol{\beta}T}$, $\mathbf{H}_{\boldsymbol{\beta}T}^{Z}$ and $\mathbf{H}_{\boldsymbol{\beta}T}^{S}$ are defined as:

$$\mathbf{H}_{\boldsymbol{\beta}C} = \mathbf{H}_{\mathbf{B},\text{CSTR}}(1:n_C) \tag{4.48}$$

$$\mathbf{H}_{\boldsymbol{\beta}T} = \mathbf{H}_{\mathbf{B},\text{CSTR}}(n_C+1:n_T) \tag{4.49}$$

$$\mathbf{H}_{\boldsymbol{\beta}C}^{Z} = \mathbf{H}_{\mathbf{B},\text{CSTR}}^{Z}(1:n_C) \tag{4.50}$$

$$\mathbf{H}_{\boldsymbol{\beta}T}^{Z} = \mathbf{H}_{\mathbf{B},\text{CSTR}}^{Z}(n_C+1:n_T) \tag{4.51}$$

$$\mathbf{H}_{\boldsymbol{\beta}C}^{S} = \mathbf{H}_{\mathbf{B},\text{CSTR}}^{S}(1:n_C) \tag{4.52}$$

$$\mathbf{H}^{S}_{\boldsymbol{\beta}T} = \mathbf{H}^{S}_{\mathbf{B},\mathrm{CSTR}}\,(n_C + 1 : n_T) \tag{4.53}$$

where the notation $1{:}n_C$ indicates columns 1 to $n_C$ of the Hessian matrix and

$$\mathbf{H}_{\mathbf{B},\mathrm{CSTR}} = \left.\frac{\partial^2 J_{\mathrm{CSTR}}}{\partial \mathbf{B}_{\mathrm{CSTR}} \partial \mathbf{B}^{\mathrm{T}}_{\mathrm{CSTR}}}\right|_{\mathbf{B}=\hat{\mathbf{B}}} \tag{4.54}$$

$$\mathbf{H}^{Z}_{\mathbf{B},\mathrm{CSTR}} = \left.\frac{\partial^2 J^{Z}_{\mathrm{CSTR}}}{\partial \mathbf{B}_{\mathrm{CSTR}} \partial \mathbf{B}^{\mathrm{T}}_{\mathrm{CSTR}}}\right|_{\mathbf{B}=\hat{\mathbf{B}}^{Z}} \tag{4.55}$$

$$\mathbf{H}^{S}_{\mathbf{B},\mathrm{CSTR}} = \left.\frac{\partial^2 J^{S}_{\mathrm{CSTR}}}{\partial \mathbf{B}_{\mathrm{CSTR}} \partial \mathbf{B}^{\mathrm{T}}_{\mathrm{CSTR}}}\right|_{\mathbf{B}=\hat{\mathbf{B}}^{S}} \tag{4.56}$$

IPOPT[48] was used as a solver to optimize objective functions in Equations (4.38), (4.40) and (4.41). AMPL™ [49] was used to define the model for the IPOPT solver. Optimization settings in IPOPT were set at their default values. 4th order (cubic) B-splines were used for simulation studies in this chapter. Several different choices for placement of the spline knots were studied and the corresponding results are presented below. To determine the integrals in Equations (4.38), (4.40) and (4.41), five Gaussian quadrature points were used between every two knots. The "gjh" function in IPOPT was used to determine the required Hessian matrixes.

The iterative procedure in Chart 1 was used for estimating the parameter vector $\boldsymbol{\zeta}_{\mathrm{CTSR}} = [k_{ref}, E/R, a, b, Q_C, Q_T, \pm^2_C, \pm^2_T]^{\mathrm{T}}$. In the first step, the parameter vector $\boldsymbol{\zeta}_{\mathrm{CTSR}}$ and B-spline coefficients are initialized. In the second step, the objective function in Equation (4.38) is minimized with respect to $\boldsymbol{\theta}_{\mathrm{CSTR}}$ and $\mathbf{B}_{\mathrm{CSTR}}$, using the most recent values of $Q_C, Q_T, \pm^2_C$ and $\pm^2_T$. In the third step, Equations (4.40) and (4.41) are minimized with respect to $\mathbf{B}^{Z}_{\mathrm{CSTR}}$ and $\mathbf{B}^{S}_{\mathrm{CSTR}}$ to find $\hat{\mathbf{B}}^{Z}_{\mathrm{CSTR}}$ and $\hat{\mathbf{B}}^{S}_{\mathrm{CSTR}}$, respectively,

using the most recent values of $Q_C, Q_T, \pm_C^2$ and $\pm_T^2$. In the fourth step, updated values of $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ are calculated using Equations (44)-(47), using the most recent values of $\hat{\boldsymbol{\theta}}_{\text{CSTR}}$, $\hat{\mathbf{B}}_{\text{CSTR}}^{S}$, $\hat{\mathbf{B}}_{\text{CSTR}}^{Z}$ and $\hat{\mathbf{B}}_{\text{CSTR}}^{S}$. In step 4 of the FLAEM algorithm, estimates of the disturbance and noise parameters were considered to have converged when the change in the sum of the squared relative errors $e(k)$ is less than $10^{-3}$ where:

$$e(k) = \left(\frac{Q_{C,k} - Q_{C,k-1}}{Q_{C,k}}\right)^2 + \left(\frac{Q_{T,k} - Q_{T,k-1}}{Q_{T,k}}\right)^2 + \left(\frac{\pm_{C,k}^2 - \pm_{C,k-1}^2}{\pm_{C,k}^2}\right)^2 + \left(\frac{\pm_{T,k}^2 - \pm_{T,k-1}^2}{\pm_{T,k}^2}\right)^2 \quad (4.7)$$

### 4.5 Results and Discussions

The FLAEM method was tested using simulated data for 10 different scenarios. In each scenario, 100 simulation runs were performed using different initial parameter guesses and different Gaussian random noise sequences for the disturbances and measurement errors. The initial guesses of the eight parameters in $\boldsymbol{\zeta}$ were chosen randomly between 50% and 150% of the corresponding true values, using uniform probability distributions. The quality of the parameter estimates in different scenarios was compared by determining medians and interquartile ranges (IQRs) for the 100 parameter estimates in each scenario. These medians and IQRs are shown in Table 4.2.

Scenarios 1 and 2 in Table 4.2 were used to study the influence of B-spline knot placement on the quality of parameter estimates obtained using FLAEM. 128 temperature measurements and 128 concentration measurements (once every 0.5 minutes) were available in these simulation studies

In Scenario 1, 128 equally-spaced B-spline knots (one at each measurement time) were used, while in Scenario 2, 256 equally-spaced knots were used for FLAEM algorithm.

For comparison, the parameter vector $\zeta$ was also estimated using an ML-based method proposed by Kristensen et al.[39] In Kristensen's method, a Gaussian distribution is assumed for the likelihood function and the mean and variance of the likelihood function are estimated using an EKF.[2] When CTSM was used to estimate parameters, default values of optimization settings were used. Note that the CTSM software requires parameter bounds be specified by the user. The lower bounds of parameters were set at zero and upper bounds were set at 10 times the true parameter values. Parameter bounds are optional using FLAEM and none were specified when generating the results in this chapter. The CTSM results when there are 128 temperature measurements and 128 concentration measurements are shown at the top of Table 4.2. Twenty-seven simulated data sets encountered convergence failures when using CTSM, wherein the optimizer selected intermediate parameter values where the differential equations could not be solved. Box plots for parameter estimates obtained from the 73 remaining simulated data sets using CTSM and all 100 data sets for FLAEM for Scenarios 1 and 2 are compared in Figures 4.2 and 4.3. These boxplots correspond to the medians and IQRs in the top three rows in Table 4.2.

.

**Table 4.2 True parameter values, median values and IQRs for the estimates using 100 Monte Carlo runs for different scenarios.**

| **Parameter** | | | $k_{ref}$ | $(E/R)/10^3$ | $a/10^6$ | $b$ | $T(0)$ | $Q_C$ | $Q_T$ | $\sigma_C^2$ | $\sigma_T^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Unit** | | | $min^{-1}$ | K | | | K | $kmol^2 \cdot m^{-6} \cdot min^{-1}$ | $K^2 \cdot min^{-1}$ | $kmol^2 \cdot m^{-6}$ | $K^2$ |
| **True Value** | | | **0.461** | **8.3301** | **1.678** | **0.50** | **341.38** | **0.010** | **4.0** | **$4\times10^{-4}$** | **0.640** |
| **Scenario** | | | | | | | | | | | |
| **1** | CTSM | Median | 0.464 | 8.3300 | 1.562 | 0.52 | 341.36 | 0.095 | 0.6 | 0.00000 | 1.026 |
| | | IQR | 0.016 | 0.2243 | 0.518 | 0.11 | 1.05 | 0.008 | 1.3 | 0.00000 | 0.339 |
| | FLAEM | Median | 0.429 | 8.2130 | 1.448 | 0.50 | 341.30 | 0.009 | 4.1 | 0.00037 | 0.660 |
| | | IQR | 0.017 | 0.2061 | 0.424 | 0.09 | 1.08 | 0.006 | 1.8 | 0.00018 | 0.256 |
| **2** | FLAEM | Median | 0.444 | 8.3164 | 1.603 | 0.49 | 341.30 | 0.011 | 4.0 | 0.00036 | 0.637 |
| | | IQR | 0.016 | 0.2265 | 0.472 | 0.09 | 1.10 | 0.004 | 1.7 | 0.00019 | 0.349 |
| **3** | FLAEM | Median | 0.431 | 8.2283 | 1.484 | 0.49 | 341.27 | 0.010 | 5.5 | 0.00024 | 0.523 |
| | | IQR | 0.019 | 0.2273 | 0.504 | 0.11 | 1.08 | 0.005 | 2.6 | 0.00042 | 0.432 |
| **4** | FLAEM | Median | 0.430 | 8.2958 | 1.490 | 0.49 | 341.27 | 0.007 | 4.0 | 0.00031 | 0.538 |
| | | IQR | 0.025 | 0.3233 | 0.501 | 0.13 | 1.04 | 0.004 | 1.4 | 0.00022 | 0.349 |
| **5** | FLAEM | Median | 0.405 | 7.9246 | 1.405 | 0.49 | 341.27 | 0.006 | 4.0 | 0.00028 | 0.411 |
| | | IQR | 0.048 | 0.5480 | 0.870 | 0.19 | 1.04 | 0.003 | 1.5 | 0.00036 | 0.436 |
| **6** | FLAEM | Median | 0.432 | 8.1928 | 1.490 | 0.50 | 341.30 | 0.005 | 2.0 | 0.00037 | 0.651 |
| | | IQR | 0.015 | 0.1801 | 0.325 | 0.08 | 1.05 | 0.002 | 0.9 | 0.00009 | 0.159 |
| **7** | FLAEM | Median | 0.411 | 8.0218 | 1.217 | 0.54 | 341.25 | 0.022 | 8.1 | 0.00037 | 0.658 |
| | | IQR | 0.076 | 1.1953 | 0.929 | 0.28 | 1.07 | 0.010 | 3.5 | 0.00017 | 0.323 |
| **8** | FLAEM | Median | 0.429 | 8.2298 | 1.424 | 0.51 | 341.31 | 0.010 | 3.7 | 0.00019 | 0.324 |
| | | IQR | 0.015 | 0.1856 | 0.432 | 0.09 | 0.79 | 0.005 | 1.0 | 0.00010 | 0.161 |
| **9** | FLAEM | Median | 0.431 | 8.1717 | 1.485 | 0.50 | 341.25 | 0.008 | 4.0 | 0.00059 | 1.052 |
| | | IQR | 0.019 | 0.2152 | 0.504 | 0.12 | 1.49 | 0.003 | 1.2 | 0.00042 | 0.431 |
| **10** | FLAEM | Median | 0.398 | 8.2634 | 1.459 | 0.47 | 341.27 | 0.009 | 8.1 | 0.00041 | 0.654 |
| | | IQR | 0.046 | 0.7269 | 0.909 | 0.23 | 1.18 | 0.015 | 4.7 | 0.00035 | 0.612 |
| **11** | FLAEM | Median | 0.431 | 8.2286 | 1.450 | 0.50 | 341.30 | 0.009 | 4.6 | | |
| | | IQR | 0.018 | 0.2407 | 0.435 | 0.09 | 1.05 | 0.003 | 1.1 | | |
| | AEM | Median | 0.432 | 8.2308 | 1.502 | 0.50 | 342.32 | 0.009 | 4.9 | | |
| | | IQR | 0.020 | 0.2062 | 0.465 | 0.10 | 1.21 | 0.004 | 1.4 | | |
| | CTSM | Median | 0.460 | 8.2800 | 1.575 | 0.51 | 341.00 | 0.092 | 1.5 | | |
| | | IQR | 0.015 | 0.1650 | 0.625 | 0.13 | 1.00 | 0.011 | 0.4 | | |

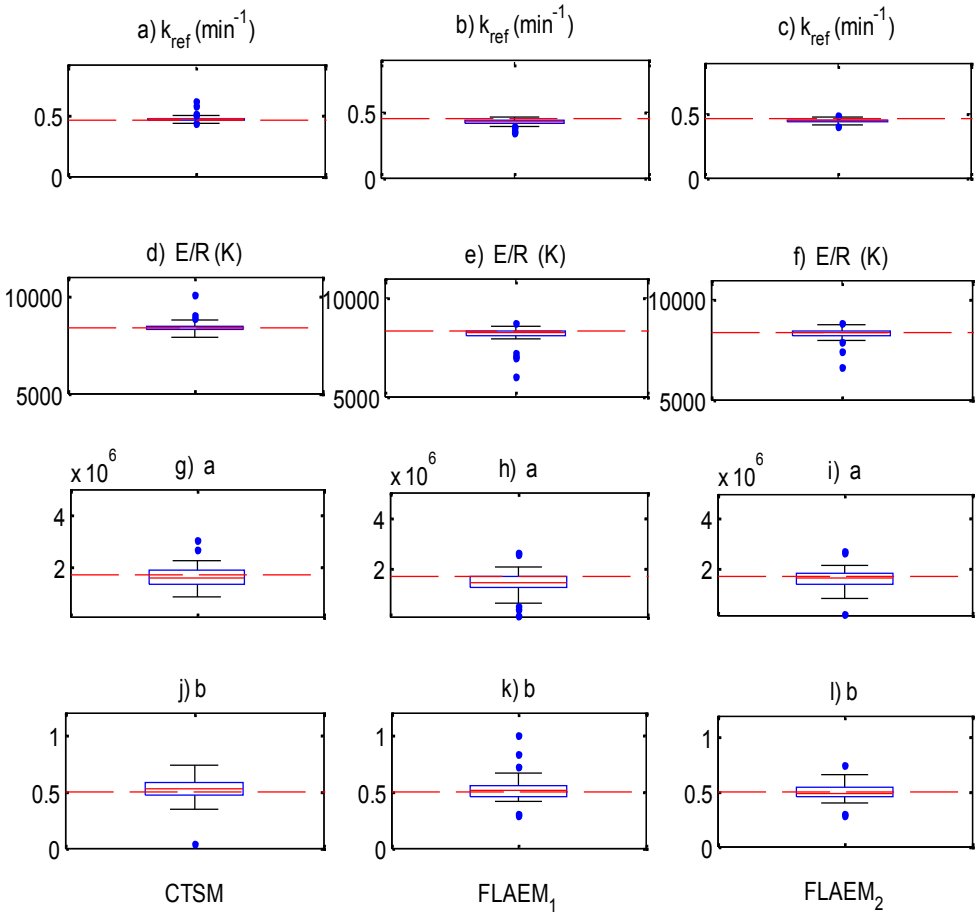



**Figure 4.2. Box-plots for estimates of model parameters using the CTSM and FLAEM methods in Scenarios 1 and 2. The dashed horizontal lines show the true values used to generate the simulated data.**

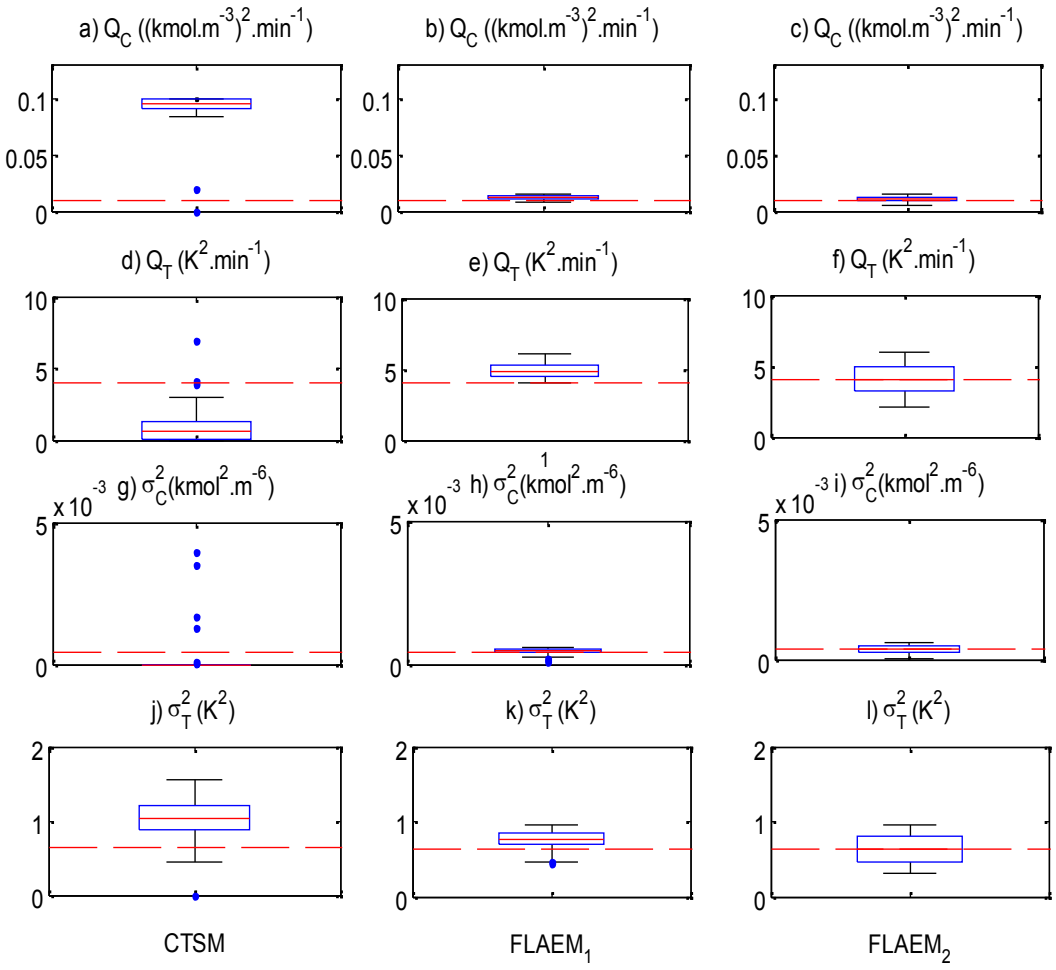



**Figure 4.3 Box-plots for disturbance intensity estimates obtained using the CTSM and FLAEM in Scenarios 1 and 2. The dashed horizontal lines show the true values used to generate the simulated data.**

The estimates of model parameters obtained using CTSM appeared to be unbiased while the estimates of $Q_C$, $Q_T$,$\pm_C^2$ and $\pm_T^2$ obtained using CTSM were noticeably biased. For example, most of the estimates of $\pm_C^2$ obtained using CTSM are nearly zero. The accuracy of the estimates of the model parameters ($k_{ref}$, $E/R$, $a$ and $b$) obtained using CTSM are similar to those obtained using FLAEM in both scenarios 1 and 2. However,

the estimates of noise parameters $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ obtained from FLAEM are less biased than those obtained using CTSM. In fact, no noticeable bias can be observed for any of the model or noise parameters in Scenario 1 (see Figures 4.2 and 4.3), except for some minor bias in $k_{ref}$. The parameter estimation results for one of the simulation studies obtained using FLAMLE (i.e., the first simulated data set) for Scenario 1 are shown in Table 4.3. The results in Figures 4.2 and 4.3 indicate that using 128 spline knots was sufficient and that using additional knots (i.e., 256 knots in Scenario 2) resulted in no improvement in parameter precision (see IQR values in Table 4.2).

Average parameter estimation times for a typical simulated data set are ~3 min for Scenario 1 using FLAEM and ~3 min for CTSM, using a laptop computer with Intel® Core™ 2, Duo CPU, 1.86 GHz. The predicted responses obtained using the FLAEM algorithm for one simulated data set and the corresponding parameter estimates from Scenario 1 are compared with the true responses in Figure 4.4. As expected, the state trajectories from the estimated spline coefficients are close to the true trajectories. The estimated noise parameters for this run are $\hat{Q}_C$ =0.012 kmol$^2\cdot$m$^{-6}\cdot$min$^{-1}$, $\hat{Q}_T$ = 3.7 K$^2\cdot$min$^{-1}$, $\pm_C^2$ = 3.7×10$^{-4}$ kmol$^2\cdot$m$^{-6}$ and $\pm_T^2$ = 0.770 K$^2$.

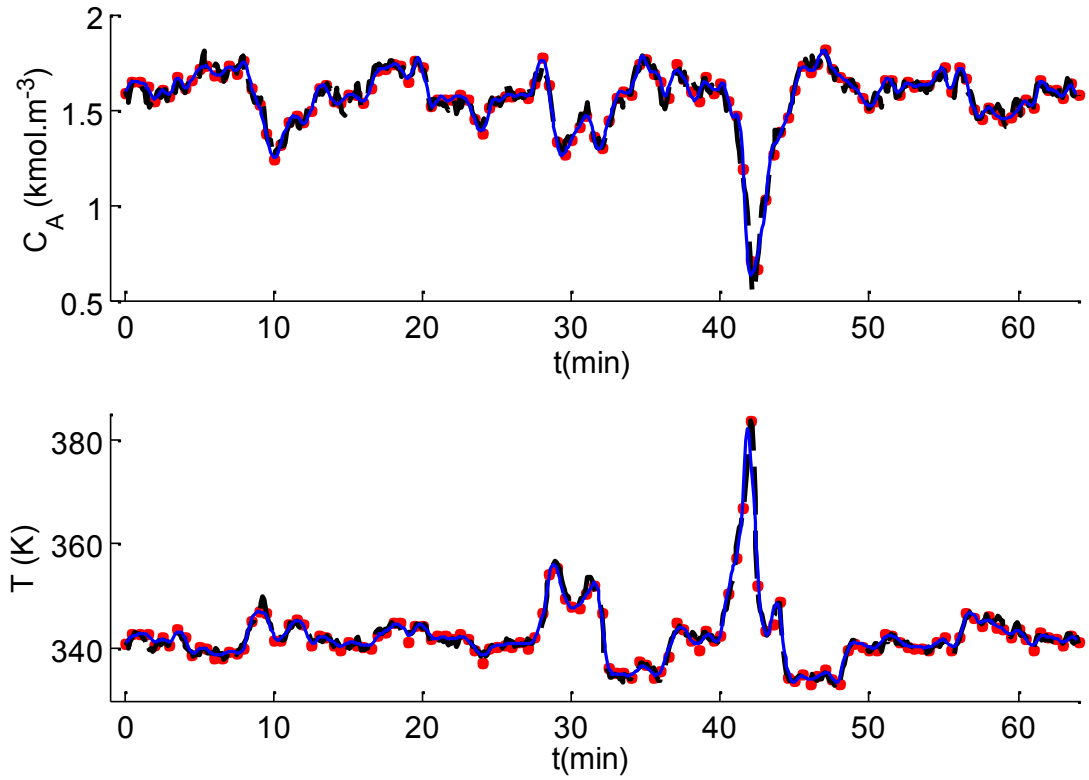



**Figure 4.4. Measured, true, and predicted concentration and temperature responses for the FLAEM method in Scenario 1, using simulated data from one of the 100 Monte Carlo simulations. (• simulated data, ----- response with true parameters and true stochastic noise, ___FLAEM response)**

Point estimates and approximate confidence intervals for the corresponding model parameters are shown in the final column of Table 4.3. These confidence intervals were determined from:[50,51]

$$\hat{\boldsymbol{\theta}} \pm z_{\sim/2}\sqrt{\text{diag}(\partial^2 J_{AMLE}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}})^{-1}\Big|_{\hat{\zeta}}} \tag{4.58}$$

Note that corresponding confidence intervals for $Q_C$, $Q_T$, $\pm\frac{2}{C}$ and $\pm\frac{2}{T}$ are not shown. When FLAEM is used for parameter estimation, numerical values for the elements of the Hessian $\partial^2 J_{AMLE}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}$ are available from IPOPT, assuming that $Q_C$, $Q_T$, $\pm\frac{2}{C}$ and $\pm\frac{2}{T}$

are known. Since estimates of $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ are updated using Equations (4.44) to (4.47), Hessian information for these parameters is not available.

**Table 4.3 Estimates and 95% Confidence intervals for LAMLE parameter estimates from one of the 100 Monte Carlo simulations**

| Parameter | Unit | True Value | Estimate ± 95% Confidence Interval |
|---|---|---|---|
| $k_{ref}$ | $min^{-1}$ | 0.435 | $0.434 \pm 0.019$ |
| $(E/R)/10^3$ | K | 8.2487 | $8.2403 \pm 0.243$ |
| $a/10^6$ | | 1.678 | $1.860 \pm 0.755$ |
| $b$ | | 0.50 | $0.42 \pm 0.15$ |
| $T(0)$ | K | 341.38 | $341.03 \pm 1.02$ |

For all of the remaining scenarios shown in Table 4.2, 128 equally spaced B-spline knots were used for approximating the concentration and temperature trajectories. Scenario 3 in Table 4.2 was used to study the robustness of the FLAEM algorithm to poorer initial guesses of parameters. In Scenario 3, the initial guesses were chosen randomly between 50% and 400% of the true parameter values. Using worse initial guesses had only a small influence on the FLAEM parameter estimation results. The estimates have larger variability than those obtained using the good initial values in Scenario 1. Note that some simulated data sets resulted in convergence to local minima when the worse initial guesses were used, leading to larger IQR values for parameter estimates. 67 of 100 attempts to estimate the parameters in this scenario using CTSM failed (results not shown) suggesting that the use of CTSM requires good values of initial guesses to obtain convergence.

Scenarios 4 and 5 in Table 4.2 were used to investigate the influence of a smaller number of measurements on the quality of the parameter estimates obtained using FLAEM. In Scenario 4, 64 equally-spaced concentration measurements and 64 equally-spaced temperature measurements were available for parameter estimation from the simulated data sets. Knot placement and initial parameter guesses were identical to Scenario 1. As expected, the medians and IQRs for parameter estimates are worse than those in Scenario 1 due to the smaller data sets. CTSM could not provide parameter estimates for any of these data sets, indicating that the use of CTSM requires a relatively larger number of measurements compared to FLAEM. Parameter estimations using CTSM were not attempted for most of the remaining scenarios in Table 4.2.

In Scenario 5, only 22 concentration measurements and 22 temperature measurements were available for parameter estimation. Despite this smaller number of measurements, the estimates of the parameters are still quite good, but the estimates have larger variability than those in Scenarios 1 and 4. Note that the estimates of $k_{ref}$ and $Q_C$ are slightly biased in this scenario. These biases might be related to: i) the finite number of data values used for parameter estimation, which can lead to bias in any ML-based method; ii) approximating the likelihood function $L(\boldsymbol{\theta} \mid \mathbf{Y_m}) = p(\mathbf{Y_m} \mid \boldsymbol{\theta})$ by the likelihood $L(\boldsymbol{\theta} \mid \mathbf{Y_m}, \mathbf{X_q}) = p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\theta})$ in Equation (4.38); iii) the use of the FLA when developing expressions for estimating $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ in Step 4 of the FLAEM algorithm or iv) approximating the state trajectories using B-splines.

Scenarios 6 and 7 were used to study the influence of larger and smaller disturbance intensities on the quality of the parameter estimates. In Scenarios 6 and 7, the values of $Q_C$ and $Q_T$ were changed to the 50% and 200% of their values from Scenario 1,

respectively (i.e., true values of $Q_C$ are 0.005 and 0.02 $kmol^2 \cdot m^{-6}\ min^{-1}$ respectively, and true values of $Q_T$ are 2.0 and 8.0 $K^2 \cdot min^{-1}$, respectively). The number of measurements and all of the other settings were the same as those in Scenario 1. As expected, in Scenario 6, the widths of the IQRs for all parameters are smaller than those obtained in Scenario 1 because smaller stochastic disturbances were encountered. Since larger disturbances occurred in Scenario 7, wider IQRs were obtained in this scenario.

Scenarios 8 and 9 were studied to examine the influence of small and large measurement noise variances on the effectiveness of FLAEM. The true values of the measurement noise variances $\pm_C^2$ and $\pm_T^2$ were changed to 50% and 200% of their values from Scenario 1. All other settings are the same as those in Scenario1. Since smaller measurement noise variances were used in Scenario 8, smaller IQR values were obtained for parameter estimates. Similarly, in Scenario 9 wider IQRs were obtained for the parameters due to the noisier data.

In Scenario 10, parameters were estimated using only the temperature measurements, with the concentration as an unmeasured state. All other settings were held at those from Scenario 1. $\pm_C^2$ was not estimated because no concentration data were obtained and the corresponding terms did not appear in the objective functions. On average, parameter estimates have larger variability than those in Scenario 1 because fewer data values were available.

In Scenario 11, the values of $\pm_C^2$ and $\pm_T^2$ are assumed to be perfectly known to permit comparisons of the FLAEM algorithm with our previously developed AEM technique. The knot placements, number of measurements and initial parameter guesses are the same as those in Scenario 1. The AEM objective function for estimating the model

parameters $\boldsymbol{\theta}_{\text{CTSR}} = [k_{ref}, E/R, a, b]^T$ and disturbance intensities $Q_C$ and $Q_T$ in the CSTR model is:

$$J_{\text{AEM,CTSR}} = \frac{1}{\pm_C^2}\sum_{j=1}^{n_C}\left(y_C(t_{m\,C,j}) - C_{A\sim}(t_{m\,C,j})\right)^2 + \frac{1}{\pm_T^2}\sum_{j=1}^{n_T}\left(y_T(t_{m\,T,j}) - T_{\sim}(t_{m\,T,j})\right)^2$$
$$+ \frac{(T_m(0) - T_{\sim}(0))^2}{S_T^2} + q\ln(Q_C) + q\ln(Q_T)$$
$$+ \frac{1}{Q_C}\int_{t_0}^{t_{nC}}\left(\frac{\mathrm{d}\,C_{A\sim}(t)}{\mathrm{d}\,t} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r(T_{\sim}(t))C_{A\sim}(t)\right)^2 \mathrm{d}\,t +$$
$$+ \frac{1}{Q_T}\int_{t_0}^{t_{nT}}\left(\frac{\mathrm{d}\,T_{\sim}(t)}{\mathrm{d}\,t} - \frac{F(t)}{V}(T_0(t) - T_{\sim}(t)) - UA[T_{\sim}(t) - T_{in}(t)] - ¥k_r(T_{\sim}(t))C_{A\sim}(t)\right)^2 \mathrm{d}\,t \tag{4.59}$$

The AEM objective function is similar to the AMLE objective function in Equation (4.38) but it has two additional terms $q\ln(Q_C)$ and $q\ln(Q_T)$. This objective function can be used for estimating model parameters and disturbance intensities in a single step. Attempts were also made to estimate the parameters using CTSM. As expected, the parameter estimates obtained from FLAEM and AEM have negligible bias. However, the AEM parameter estimates in this scenario have slightly larger variability than those obtained using FLAEM. Recall that FLAEM uses the FLA for approximating the E step of the EM algorithm. As explained in our previous work, AEM uses the mode of the expected value of the E step in the EM algorithm.[27] The results from this case study suggest that FLAEM uses a better approximation. Using CTSM, successful parameter estimation was only obtained for 36 of the 100 Monte Carlo cases attempted for this scenario. The remaining 64 cases experienced convergence difficulties due to parameter values that made numerical solution of the differential equations unstable. As shown in

Table 4.2, the 36 CTSM estimates of $Q_C$ and $Q_T$ have more bias than the estimates obtained using FLAEM and AEM.

In summary, the results in Table 4.2 and Figures 4.2 and 4.3 suggest that the FLAEM parameter estimates are less biased and more accurate than corresponding estimates obtained using CTSM for the CSTR example studied. Since the FLAEM algorithm is an approximate MLE method, some bias in parameter estimates was expected in situations involving sparse data sets. Some of the minor bias observed in Figures 4.2 and 4.3 and also Table 4.2 may also be due to the B-spline approximations and the FLA. Since computationally intensive MCMC-based MLE techniques are asymptotically efficient and consistent estimators that do not make B-spline or fully-Laplace approximations, we recommend that the performance of the FLAEM algorithm should be compared to several MCMC methods. The FLAEM computational times encountered in the CSTR examples in this chapter are modest (~ 3 minutes using a laptop computer with Intel® Core™ 2, Duo CPU, 1.86 GHz.) and are expected to be significantly shorter than the corresponding MCMC computing times. The relative computational benefits of the FLAEM algorithm are expected to become more important for larger-scale problems, because FLAEM does not require the drawing of numerous samples from high-dimensional probability density functions.[52] As a result, the performance of FLAEM and MCMC should be compared using a larger-scale example problem than the illustrative CSTR problem used in the current chapter.

## 4.6 Conclusions

In this paper, the Fully-Laplace-Approximation Expectation Maximization (FLAEM) algorithm is presented for estimating parameters, stochastic disturbance intensities and measurement noise variances for nonlinear stochastic differential equation (SDE) models. In the first stage of the FLAEM algorithm, model parameters $\boldsymbol{\theta}$ are estimated by minimizing Varziri's AMLE objective function, assuming that the disturbance and noise parameters are known.[1] In the second stage, disturbance intensities and noise variance estimates are updated. The expressions used to obtain these noise parameters were derived by approximating the E-step of the EM algorithm using the FLA and B-spline basis functions. The proposed FLAEM algorithm iterates between these two stages until convergence is obtained. The effectiveness of the FLAEM algorithm was tested using a two-state nonlinear stochastic CSTR model. The FLAEM algorithm showed good performance for estimating model parameters, disturbance intensities and measurement noise variances and was more robust than CTSM, which uses a classical ML-based method[39], particularly when the number of measurements was relatively small or when initial guesses of parameters were relatively poor. For the cases where CTSM was able to converge, the resulting parameter values tended to be less accurate than the corresponding parameter estimates from FLAEM. Although the FLAEM algorithm was developed for situations where model parameters, disturbance intensities and noise variances must all be estimated from the data, a few simulations were also performed assuming that noise variances were known. These simulation results suggest that FLAEM performs better than our previous AEM algorithm. Implementation of the FLAEM algorithm is relatively easy. The user must supply information about the knot location for

the B-spline basis functions, along with the initial parameter guesses. If the user is not certain about the number of knots that are required, additional knots can be added until the resulting parameter estimates and estimated state trajectories do not change appreciably when additional knots are used.

In future, it will be important to test the FLAEM algorithm using larger-scale parameter estimation problems and to compare FLAEM results with MLE-based methods that use Markov Chain Monte Carlo (MCMC) algorithms for parameter estimation.[13-15,26,53]

It will also be interesting to investigate whether other potential approximations (e.g., a Laplace approximation[54] for the likelihood function) can lead to further improvements in parameter estimates without resorting to computationally intensive MCMC-based techniques. In future, it will be desirable to investigate the convergence of FLAEM algorithm.

## 4.7 References

(1) Varziri, M.; Poyton, A.; McAuley, K.; McLellan, P.Ramsay, J. Selecting Optimal Weighting Factors in iPDA for Parameter Estimation in Continuous-Time Dynamic Models. *Comput. Chem. Eng.* **2008**, *12*, 3011.

(2) Kristensen, N. R.; Madsen, H. Jørgensen, S. B. Parameter Estimation in Stochastic Grey-Box Models. *Automatica* **2004**, *2*, 225.

(3) Érdi, P.; Tóth, J. *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models*; Manchester University Press: 1989.

(4) King, R. Applications of Stochastic Differential Equations to Chemical-Engineering Problems-an Introductory Review. *Chem. Eng. Commun.* **1974**, *5*, 221.

(5) Marlin, T. E.; Marlin, T. *Process Control: Designing Processes and Control Systems for Dynamic Performance*; McGraw-Hill New York: 1995.

(6) Jones, R.; MacGregor, J. Murphy, K. State Estimation in Wastewater Engineering: Application to an Anaerobic Process. *Environ. Monit. Assess.* **1989**, *2*, 271.

(7) Jazwinski, A. H. *Stochastic Processes and Filtering Theory*; Academic press: 1970.

(8) Lima, F. V.; Rawlings, J. B. Nonlinear Stochastic Modeling to Improve State Estimation in Process Monitoring and Control. *AICHE J.* **2011**, *4*, 996.

(9) Gagnon, L.; MacGregor, J. State Estimation for Continuous Emulsion Polymerization. *The Canadian Journal of Chemical Engineering* **1991**, *3*, 648.

(10) Dempster, A. P.; Laird, N. M. Rubin, D. B. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J. R. Sta. Soc. Series B.* **1977**, *37*, 1.

(11) Shumway, R. H.; Stoffer, D. S. *Time Series Analysis and its Applications: With R Examples*; Springer: 2011.

(12) Gibson, S.; Ninness, B. Robust Maximum-Likelihood Estimation of Multivariable Dynamic Systems. *Automatica.* **2005**, *10,* 1667.

(13) Gopaluni, R. A Particle Filter Approach to Identification of Nonlinear Processes Under Missing Observations. *Can. J. Chem. Eng.* **2008**, *6,* 1081.

(14) Schön, T. B.; Wills, A.Ninness, B. System Identification of Nonlinear State-Space Models. *Automatica.* **2011**, *1,* 39.

(15) Gopaluni, R. B. Identification of Non-Linear Processes with Known Model Structure under Missing Observations. *Proceedings of the IFAC 17th World Congress*, Seoul, Korea, July 6, **2008**, Vol. 1.

(16) Lillacci, G.; Khammash, M. Parameter Estimation and Model Selection in Computational Biology. *PLoS Comput. Biol.* **2010**, *3,* e1000696.

(17) Duncan, S.; Gyongy, M. Using the EM Algorithm to Estimate the Disease Parameters for Smallpox in 17th Century London. *International Conference on Control Applications.* IEEE, **2006**, 3312.

(18) Roweis, S.; Ghahramani, Z. An EM Algorithm for Identification of Nonlinear Dynamical Systems. 2000.

(19) Goodwin, G. C.; Aguero, J. C. Approximate EM Algorithms for Parameter and State Estimation in Nonlinear Stochastic Models. *European Control Conference*. IEEE, **2005**, 368.

(20) Chen, W.; Bakshi, B. R.; Goel, P. K. Ungarala, S. Bayesian Estimation via Sequential Monte Carlo Sampling: Unconstrained Nonlinear Dynamic Systems. *Ind. Eng. Chem. Res.* **2004**, *14,* 4012.

(21) Poyiadjis, G.; Doucet, A.; Singh, S. S. Maximum Likelihood Parameter Estimation in General State-Space Models using Particle Methods. Proc. of the American Stat. Assoc. Citeseer, **2005**.

(22) Doucet, A.; De Freitas, N.; Gordon, N. *Sequential Monte Carlo Methods;* Springer: New York, 2001.

(23) Doucet, A.; Tadić, V. B. Parameter Estimation in General State-Space Models using Particle Methods. *Ann. I. Stat. Math.* **2003**, *2*, 409.

(24) Andrieu, C.; Doucet, A.; Singh, S. S. Tadic, V. B. Particle Methods for Change Detection, System Identification, and Control. *Proc. IEEE*. **2004**, *3*, 423.

(25) Schön, T. B.; Wills, A.; Ninness, B. Maximum Likelihood Nonlinear System Estimation; Department of Electrical Engineering, Linköpings universitet: 2005.

(26) Gopaluni, R. B. Nonlinear System Identification Under Missing Observations: The Case of Unknown Model Structure. *J. Process Control.* **2010**, *3*, 314.

(27) Karimi, H.; McAuley, K. B. An Approximate Expectation Maximization Algorithm for Estimating Parameters in Nonlinear Dynamic Models with Process Disturbances. *Can. J. Chem. Eng.* **2013**, Accepted.

(28) Chen, T.; Morris, J. Martin, E. Particle Filters for State and Parameter Estimation in Batch Processes. *J. Process Control.* **2005**, *6*, 665.

(29) Haario, H.; Laine, M.; Mira, A. Saksman, E. DRAM: Efficient Adaptive MCMC. *Stat. Comput.* **2006**, *4*, 339.

(30) Klaas, M.; Briers, M.; De Freitas, N. Doucet, A.; Maskell, S.; Lang, D. Fast Particle Smoothing: If I Had a Million Particles. *Proceedings of the 23rd international conference on Machine learning*. ACM, **2006**; 481.

(31) Kantas, N.; Doucet, A.; Singh, S. S. Maciejowski, J. M. An Overview of Sequential Monte Carlo Methods for Parameter Estimation in General State-space Models. *Proc. IFAC Symposium on System Identification*. **2009**.

(32) Imtiaz, S. A.; Roy, K.; Huang, B.; Shah, S. L. Jampana, P. Estimation of States of Nonlinear Systems Using a Particle Filter. *ICIT IEEE.* **2006**, 2432.

(33) Varziri, M.; McAuley, K.McLellan, P. Parameter and State Estimation in Nonlinear Stochastic continuous-time Dynamic Models with Unknown Disturbance Intensity. *Can. J. Chem. Eng.* **2008**, *5*, 828.

(34) Tierney, L.; Kadane, J. B. Accurate Approximations for Posterior Moments and Marginal Densities. *J. Am. Statist. Assoc.* **1986**, *393*, 82.

(35) Tierney, L.; Kass, R. E. Kadane, J. B. Approximate Marginal Densities of Nonlinear Functions. *Biometrika* **1989**, *3*, 425.

(36) Rizopoulos, D.; Verbeke, G.Lesaffre, E. Fully Exponential Laplace Approximations for the Joint Modelling of Survival and Longitudinal Data. *J. R. Stat. Soc.: Series B.* **2009**, *3*, 637.

(37) Bianconcini, S.; Cagnone, S. Estimation of Generalized Linear Latent Variable Models Via Fully Exponential Laplace Approximation. *J. Multivariate. Anal.* **2012**, *112*, 183.

(38) Zhou, M. Fully exponential Laplace approximation EM algorithm for nonlinear mixed effects models, PhD Thesis, University of Nebraska-Lincoln, 2009.

(39) Kristensen, N. R.; Madsen, H. Continuous Time Stochastic Modelling: CTSM 2.3 User's Guide, Technical University of Denmark, 2003.

(40) Maybeck, P. S. *Stochastic Models, Estimation and Control*; Academic Pr: 1982.

(41) Ramsay, J. *Functional Data Analysis*; Wiley Online Library: 2005.

(42) De Boor, C. *A Practical Guide to Splines*; Springer: 1978.

(43) Ramsay, J.; Silverman, B. Functional Data Analysis; John Wiley & Sons: 2005.

(44) Poyton, A.; Varziri, M. S.; McAuley, K. B.; McLellan, P. Ramsay, J. O. Parameter Estimation in Continuous-Time Dynamic Models using Principal Differential Analysis. *Comput. Chem. Eng.* **2006**, *4*, 698.

(45) Dempster, A. P.; Laird, N. M. Rubin, D. B. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J. R. Sta. Soc. Series B,* **1977**, 1.

(46) Anderson, B. D. O.; Moore, J. B. *Optimal Filtering*; Prentice-hall Englewood Cliffs, NJ: 1979.

(47) Tierney, L.; Kass, R. E. Kadane, J. B. Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions. *J. Am. Statist. Assoc.* **1989**, *407*, 710.

(48) Wächter, A.; Biegler, L. T. On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming. *Math. Program.* **2006**, *1*, 25.

(49) Fourer, R.; Gay, D. M.; Kernighan, B. W. *A Modeling Language for Mathematical Programming*; INFORMS: 1990.

(50) Kay, S. M. *Fundamentals of Statistical Signal Processing, Vol. I and Vol. II*; Prentice Hall: 1998.

(51) Varziri, M. S.; McAuley, K. B. McLellan, P. J. Parameter Estimation in Continuous-Time Dynamic Models in the Presence of Unmeasured States and Nonstationary Disturbances. *Ind. Eng. Chem. Res.* **2008**, *2*, 380.

(52) Jang, S. S.; De la Hoz, H.; Ben-zvi, A.; McCaffrey, W. C. Gopaluni, R. B. Parameter Estimation in Models with Hidden Variables: An Application to a Biotech Process. *Can. J. Chem. Eng.* **2011**, *3*, 690.

(53) Gopaluni, R. B. Nonlinear System Identification Under Missing Observations: The Case of Unknown Model Structure. *J. Process Control.* **2010**, *3*, 314.

(54) Barndorff-Nielsen, O. E.; Cox, D. R. *Asymptotic Techniques for use in Statistics*; Chapman and Hall London: 1989.

(55) Lancaster, P.; Tismenetsky, M. *Theory of Matrices*; Academic press New York: 1969.

# Chapter 5

# A Maximum-Likelihood Method for Estimating Parameters, Stochastic Disturbance Intensities and Measurement Noise Variances in Nonlinear Dynamic Models with Process Disturbances

## 5.1 Abstract

An improved approximate maximum likelihood algorithm is developed for estimating measurement noise variances along with model parameters and disturbance intensities in nonlinear stochastic differential equation (SDE) models. This algorithm uses a Laplace approximation and B-spline basis functions for approximating the likelihood function of the parameters given the measurements. The resulting Laplace approximation maximum likelihood estimation (LAMLE) algorithm is tested using a nonlinear continuous stirred tank reactor (CSTR) model. Estimation results for four model parameters, two process disturbance intensities and two measurement noise variances are obtained using LAMLE and are compared with results from two other maximum-likelihood-based methods, the continuous-time stochastic method (CTSM) of Kristensen and Madsen [1] and the Fully Laplace Approximation Estimation Method (FLAEM).[2] Parameter estimations using 100 simulated data sets reveal that the LAMLE estimation results tend to be more precise and less biased than corresponding estimates obtained using CTSM and FLAEM.

This chapter will be submitted as a journal paper to *Computers and Chemical Engineering Journal*.

## 5.2 Introduction

Fundamental models based on mass and energy balances are imperfect representations of process behavior due to simplifying assumptions and approximations that ignore complex interactions.[3] Model uncertainties may also arise from random disturbances associated with feed streams and the environment conditions.[4,5] As a result, some modelers add stochastic terms to their dynamic fundamental models to account for model mismatch and process disturbances, resulting in systems of stochastic differential equations (SDEs).[6,7] Parameter estimation in these models is difficult because two different sources of uncertainty are accounted for: i) stochastic process disturbances and ii) measurement noise. Maximum Likelihood (ML) methods are commonly used to estimate model parameters and uncertainty parameters in these systems.[8,9] ML methods provide reliable asymptotically unbiased parameter estimates for SDE models.[9] ML methods maximize the likelihood function of the unknown model and uncertainty parameters given the measurements.[9] When some states are unmeasured, a closed form for this likelihood function is difficult to derive since it implicitly depends on the values of the state variables.[9,10] Several different approximation methods have been developed to find a closed form for the likelihood function so that parameters can be estimated in SDE models. For example, simulated maximum likelihood (SML) methods such as Markov Chain Monte Carlo (MCMC) techniques[10-18] estimate the likelihood function by drawing samples from a target density function. MCMC methods do not require assumptions about the form of the density function. Benefits and drawbacks of MCMC methods are summarized by Imtiaz et. al.[19] and Kantas et al.[18]. Other techniques that can be used for approximating the likelihood function are: Hermite expansions,[20,21] solution of the

Fokker-Planck equation[22,23] and polynomial chaos.[24] Comprehensive reviews of these techniques are given by Lindstrom[25] and Hurn et al.[26] Unfortunately, the aforementioned methods are computationally demanding especially when the dimension of the model and or the number of parameters increases.[15,17,27,28]

Varziri et al.[29] developed a computationally efficient Approximate Maximum Likelihood Estimation (AMLE) method for estimating parameters in SDE models when disturbance intensities and noise variances are known. This technique was developed by finding a closed form for the likelihood function of the parameters given the joint vector of measurements and states, using B-spline basis functions to approximate the state trajectories.[29] The AMLE methodology has been extended so that unknown disturbance intensities can be estimated along with the model parameters in cases where the measurement noise variances are known.[2,30]

Linearization-based ML methods have also been used for approximating the likelihood function. For example, an extended Kalman filter (EKF) is used for approximating the likelihood function in the CTSM software developed by Kristensen et al.[31] Although linearization-based techniques are computationally efficient, parameter estimates from these methods can be biased when strong nonlinearities are present in the model.[11]

In ML methods, a closed form for the likelihood function (i.e., the probability density function of measurements given the parameters) can be found by integrating the likelihood of the complete data over all of the values of states, where the complete data is a joint vector of the measurements and the states. Calculating this multidimensional integral is challenging, particularly when the number of state variables is large.[10,32] Multidimensional integrals can be approximated using a second-order Taylor series

expansion, known as the Laplace Approximation (LA).[33,34] The LA has been used for approximating the likelihood of mixed algebraic models that contain deterministic and stochastic components[35-42] and for approximating posterior density functions in Bayesian estimation.[43,44] Heald and Stark[45] developed an iterative ML algorithm based on the LA to estimate the magnitude of the dynamic noise (i.e., the process disturbance intensity) and the measurement noise variance in nonlinear time-series models. To our knowledge, the LA has never been used for estimation of model and disturbance parameters in SDEs.

Recently, we developed an Approximate Expectation Maximization (AEM) algorithm[2] that provides disturbance intensity estimates that are more accurate than those obtained using AMLE[30] and CTSM.[1] Unlike AMLE, the AEM algorithm uses a single ML-based objective function for simultaneous estimation of model parameters and disturbance intensities. One shortcoming of the AEM and AMLE methods is that these algorithms require measurement noise variances to be known a priori. More recently, we developed a Fully-Laplace-Approximation Expectation Maximization (FLAEM) algorithm for simultaneous estimation of model parameter, process disturbance intensities and unknown measurement noise variances in SDE models. In the first stage of this iterative algorithm, model parameters are estimated using Varziri's AMLE objective function,[29] assuming that the disturbance intensities and measurement noise variances are known. In the second stage, disturbance intensity and noise variance estimates are updated. The expressions used to update the disturbance intensities and noise variances were derived using the FLA and B-spline basis functions. The FLAEM algorithm iterates between the two steps until convergence is obtained. Additional details about AMLE, AEM and FLAEM methods are provided in Sections 5.3 and 5.4.

The goals of this chapter are to develop a new Laplace Approximation Maximum Likelihood Estimation (LAMLE) algorithm that makes use of the LA for integrating the likelihood function of the parameters given the complete data, and to test this algorithm using a simple nonlinear two-state SDE model. First, necessary notation and background information are introduced. Next, the objective function for the LAMLE algorithm is derived. The proposed methodology is then tested using a CSTR model and results are compared with those from CTSM and FLAEM.[46] Advantages and limitations of the three methods are discussed, revealing that the proposed LAMLE method is computationally effective and provides more accurate estimates than CTSM and FLAEM for the CSTR example studied.

## 5.3 Preliminaries

### 5.3.1 Model and Notation

Consider a Multi-Input Multi-output (MIMO) nonlinear SDE model of the following form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\eta}(t) \tag{5.1.a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{5.1.b}$$

$$\mathbf{y}(t_{m\,r,j}) = \mathbf{g}(\mathbf{x}(t_{m\,r,j}), \mathbf{u}(t_{m\,r,j}), \boldsymbol{\theta}) + \boldsymbol{\varepsilon}(t_{m\,r,j}) \tag{5.1.c}$$

where $\mathbf{x} \in \mathrm{R}^X$ is the $X$-dimensional vector of state variables, $t$ is time, $\mathbf{u} \in \mathrm{R}^U$ is the $U$-dimensional vector of input variables, $\boldsymbol{\theta} \in \mathrm{R}^P$ is the $P$-dimensional vector of unknown model parameters, $\mathbf{f} : \mathrm{R}^X \times \mathrm{R}^U \times \mathrm{R}^P \to \mathrm{R}^X$ is an $X$-dimensional nonlinear mapping, $\mathbf{y} \in \mathrm{R}^Y$ is the $Y$-dimensional vector of measured outputs and $\boldsymbol{\varepsilon} \in \mathrm{R}^Y$ is the $Y$-dimensional zero-mean measurement error with a diagonal covariance matrix $\rho$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \pm_1^2 \cdots 0 \\ \vdots \ \ddots \ \vdots \\ 0 \ \cdots \ \pm_Y^2 \end{bmatrix} \tag{5.2}$$

Measurement errors for $r$th response at different times $t_{m\,r,j}$ ($j = 1 \ldots N_r$) are assumed to be independent. $N_r$ is the number of measurements for $r$th response. $\boldsymbol{\eta}(t)$ is a continuous zero-mean stationary Gaussian white-noise process with covariance matrix $\mathrm{E}\{\boldsymbol{\eta}(t_1)\boldsymbol{\eta}(t_2)\} = \mathbf{Q}\delta(t_2 - t_1)$, where $\mathbf{Q}$ is the process disturbance intensity matrix for $\boldsymbol{\eta}(t)$:

$$\mathbf{Q} = \begin{bmatrix} Q_1 \ \cdots \ 0 \\ \vdots \ \ddots \ \vdots \\ 0 \ \cdots \ Q_X \end{bmatrix} \tag{5.3}$$

which is also referred to as the power spectral density matrix, and $\delta(.)$ is the Dirac delta function.[47] In cases where the vector of initial conditions is measured and is not perfectly known, measurements for the initial conditions are assumed to have a normal distribution with mean $\mathrm{E}\{\mathbf{x_{m}}_0\} = \mathbf{x}_0$ and $\mathrm{cov}\{\mathbf{x_{m}}_0\} = \mathbf{S_{m}}_0$, which is diagonal.

The parameters that usually require estimation in the SDE model in Equation (5.1) are the model parameters $\boldsymbol{\theta}$, the process disturbance intensity $\mathbf{Q}$ and the measurement noise variance $\rho$. Sometimes modelers may know $\rho$ from information about the measurement device or from replicate measurements, in which case $\rho$ would not require estimation. Modelers usually do not know the magnitude of the stochastic error terms (i.e., $\mathbf{Q}$) in their model. Throughout this chapter, the vector of stacked measurements for the model in Equation (5.1) is written using the notation

$\mathbf{Y_m} = [y_1(t_{m1,1}) \ldots y_1(t_{m1,N_1}) \ldots y_Y(t_{mY,1}) \ldots y_Y(t_{mY,N_Y})]^{\mathrm{T}}$. Also

$\mathbf{X_m} = [x_1(t_{m1,1}) \ldots x_1(t_{m1,N_1}) \ldots x_Y(t_{mY,1}) \ldots x_Y(t_{mY,N_Y})]^T$ is the stacked vector of state values at the measurement times. $\mathbf{U_m}$ and $\boldsymbol{\varepsilon_m}$ are corresponding stacked vectors for the input variables and random errors:

$$\mathbf{Y_m} = \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon_m} \tag{5.4}$$

Denote $\mathbf{Q_d}$ as the vector of diagonal elements of $\mathbf{Q}$ (i.e., $\mathbf{Q_d}=[Q_1,\ldots,Q_X]^T$). Also denote $\rho_\mathbf{d}$ as the diagonal elements of the covariance matrix (i.e., $\rho_\mathbf{d}=[\sigma_1^2,\ldots,\sigma_Y^2]^T$). Let $\boldsymbol{\zeta} = [\boldsymbol{\theta}^T, \mathbf{x}_0^T, \mathbf{Q_d}^T, \boldsymbol{\Sigma}_\mathbf{d}^T]^T$ denote the vector of model parameters, unknown initial states, disturbance intensities and measurement noise variances in the SDE model that should be estimated. To simplify the notation, derivations in this chapter are developed assuming *n* measurements are available for each response. However, derivations for the case where $N_r$ measurements are available for the *r*th response are also shown in Appendix 5.C.

### 5.3.2 Spline Basis Functions

Discrete data can be smoothed using basis functions that are continuous functions of time. B-spline functions are a common and effective choice of basis function due to their compact support and piecewise definitions.[48,49] *M*th order B-splines are positive within *M* intervals and are zero elsewhere.[50] Throughout this paper the subscript ~ is used to indicate states approximated by B-spline basis functions. A linear combination of B-splines can be used to approximate states of the SDE model in Equation (5.1):

$$x_{\sim s}(t) = \sum_{l=1}^{c_s} \beta_{s,l}\varphi_{s,l}(t) \qquad \text{for} \quad s=1,\ldots,X \tag{5.5}$$

where $x_{\sim s}$ is the *s*th state trajectory for the SDE model shown in Equation (5.1), $c_s$ is the number of B-spline basis functions for the *s*th state, $_{s,l}$ is a the *l*th B-spline coefficient for the *s*th state and $\varphi_{s,l}(t)$ is the corresponding B-spline basis function. In matrix form:

$$\mathbf{x}_{\sim}(t) = \mathbf{\Phi}(t)\mathbf{B} \tag{5.6}$$

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

$$\mathbf{\Phi}(t) = \begin{bmatrix} \boldsymbol{\varphi}_1^T(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}_2^T(t) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\varphi}_X^T(t) \end{bmatrix} \tag{5.7}$$

and

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_X \end{bmatrix} \tag{5.8}$$

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

$$\boldsymbol{\beta}_s = [\ \ _{s,1}, \dots, \ \ _{s,c_s}]^{\mathrm{T}} \qquad \text{for} \quad s=1,\dots,X \tag{5.9}$$

### 5.3.3 Approximate Maximum Likelihood Estimation (AMLE) Algorithm

Varziri et al.[29] discretized the SDE in Equation (5.1) to develop their AMLE objective function for estimating the model parameters $\boldsymbol{\theta}$ in SDE models. The discretized form of Equation (5.1) using an Euler approximation is:

$$\mathbf{x}(t_{i-1} + \Delta t) = \mathbf{x}(t_i) = \mathbf{x}(t_{i-1}) + \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t + \boldsymbol{\eta}_{\mathbf{d}}(t_{i-1})\Delta t \tag{5.10a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{5.10b}$$

where $\mathbf{x}(t_i)$ is the vector of state variables at $q$+1 uniformly-spaced discrete time points $t_i$ , $i=0,..,q$. $\boldsymbol{\eta}_d$ is a discrete white-noise process that can be used to approximate continuous-time white noise using a small sampling interval $\Delta t$. Consider $\mathbf{X}_{\mathbf{q}} = [\mathbf{x}^{\mathrm{T}}(t_0), \dots, \mathbf{x}^{\mathrm{T}}(t_q)]^{\mathrm{T}}$ as the stacked vector of state values at the discrete times. Varziri et al.[30] developed a closed form expression for $-\ln p(\mathbf{Y_m}, \mathbf{X_{\sim q}} \mid \boldsymbol{\theta})$ using B-splines to represent state trajectories

and used it to develop the following objective function for estimating **θ**, assuming **Q** and p are known:[30]

$$J_{\text{AMLE}} = [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{m\sim}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{m\sim}}, \mathbf{U_m}, \boldsymbol{\theta})] + (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0})^{\mathrm{T}} \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0}) + \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})] \mathrm{d}t \tag{5.11}$$

wherein the state trajectories in **x**(*t*) are approximated by B-spline basis functions.

### 5.3.4 Laplace Approximation

The LA[33] is used for approximating integrals of Gaussian joint density functions of the form $\int \mathrm{e}^{p(\mathbf{x})} \mathrm{d}\,\mathbf{x}$.[42] The LA is based on a second-order Taylor series expansion of $p(\mathbf{x})$ about its mode ($\hat{\mathbf{x}}$):[33]

$$\int \mathrm{e}^{p(\mathbf{x})} \mathrm{d}\,\mathbf{x} \approx \mathrm{e}^{p(\hat{\mathbf{x}})} \int \exp\left\{ \frac{1}{2} (\mathbf{x} - \hat{\mathbf{x}})^{\mathrm{T}} \left( \frac{\partial^2 p}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}} \right)\Bigg|_{\hat{\mathbf{x}}} (\mathbf{x} - \hat{\mathbf{x}}) \right\} \mathrm{d}\,\mathbf{x} \tag{5.12}$$

or equivalently taking the natural algorithm of (5.12) and solving the integral, the LA is:

$$\ln \int \mathrm{e}^{p(\mathbf{x})} \mathrm{d}\,\mathbf{x} \approx \mathrm{p}(\hat{\mathbf{x}}) - \frac{1}{2} \ln \det \left( \left\{ \frac{-\partial^2 p}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}} \right\}\Bigg|_{\hat{\mathbf{x}}} \right) + \frac{1}{2} \dim(\mathbf{x}) \ln(2\text{®}) \tag{5.13}$$

where **x** is a vector of random variables and $\hat{\mathbf{x}}$ is the vector of random variable values that maximizes $p(\mathbf{x})$. The accuracy of Laplace approximation is on the order of $[\dim(\mathbf{x})]^{-1}$.

## 5.4 Development of the Laplace Approximation Maximum Likelihood Estimation Method

In this section, a Laplace Approximation Maximum Likelihood Estimation (LAMLE) algorithm is developed for estimating the model parameters $\boldsymbol{\theta}$, the process disturbance intensities $\mathbf{Q}$ and the measurement noise variances $\rho$ in SDE models of the form shown in Equation (5.1). The likelihood function of the parameters given the measurements (i.e., $L(\boldsymbol{\theta} \mid \mathbf{Y_m}) = p(\mathbf{Y_m} \mid \boldsymbol{\theta})$) can be obtained by integrating the likelihood function of the complete data $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ over the vector of possible values for the state variables $\mathbf{X_q}$:[16]

$$p(\mathbf{Y_m} \mid \boldsymbol{\zeta}) = \int p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta}) \, \mathrm{d}\, \mathbf{X_q} \tag{5.14}$$

The following closed form expression was derived in our previous work:[2]

$$\begin{aligned} p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta}) = {} & C_1 [\det(\boldsymbol{\Sigma})]^{-n} \exp\{-\frac{1}{2}[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]\} \\ & \times [\det(\mathbf{S}_{m0})]^{-\frac{1}{2}} \exp[-\frac{1}{2}(\mathbf{x}_{m0} - \mathbf{x}_0)^{\mathrm{T}} \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_0)] \\ & \times [\det(\mathbf{Q})]^{-\frac{q}{2}} \exp\{-\int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] \, \mathrm{d}\, t\} \end{aligned} \tag{5.15}$$

where $C_1$ is a constant.

The integral in Equation (5.14) is a multi-dimensional integral (over all possible values of the state variables at the discrete times) that can be difficult to evaluate unless approximations are made. Using the LA to find a closed form for $-\ln p(\mathbf{Y_m} \mid \boldsymbol{\zeta})$ gives:

$$J_{\text{LAMLE}} = [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U_m}, \boldsymbol{\theta})] + n[\det(\boldsymbol{\Sigma})]$$
$$+ (\mathbf{x}_{\sim m0} - \mathbf{x}_0)^{\mathrm{T}} \mathbf{S}_{m0}^{-1} (\mathbf{x}_{\sim m0} - \mathbf{x}_0) + q \ln[\det(\mathbf{Q})]$$
$$+ \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})] \mathrm{d}t + \ln[\det(\mathbf{H}_{\mathbf{X}_{\sim}})] \tag{5.16}$$

as shown in the derivation in Appendix 5.A. In Equation (5.16), $\mathbf{H}_{\mathbf{X}\sim}$ is the Hessian matrix of $-\ln p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ with respect to $\mathbf{X_q}$ evaluated at $X_{\mathbf{q}\sim}$

$$\mathbf{H}_{\mathbf{X_q}} = -\frac{\partial^2 \ln p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}}\bigg|_{\mathbf{X_q} = \mathbf{X_{q\sim}}} \tag{5.17}$$

The optimal values of the model parameters $\boldsymbol{\theta}$, the disturbance intensities $\mathbf{Q}$, the noise variances $\mathsf{p}$ and the B-spline coefficients $\mathbf{B}$ can be found by minimizing $J_{\text{LAMLE}}$:

$$\hat{\boldsymbol{\zeta}}, \hat{\mathbf{B}} = \arg\min_{\boldsymbol{\zeta}, \mathbf{B}} J_{\text{LAMLE}} \tag{5.18}$$

An analytical expression for the Hessian matrix in Equation (5.16) is difficult to obtain. Several approximations have been used to avoid complex Hessian expressions.[41,45,51] The method of Heald and Stark, which was developed for determining noise parameters in time-series models, is particularly attractive due to its simplicity and accuracy. In the current chapter, this methodology is adapted for use in SDE models. The corresponding derivation is shown in Appendix 5.B.

Taking the derivative of $J_{\text{LAMLE}}$ with respect to $\mathbf{Q}$ and $\mathsf{p}$ and setting them to zero results in the following useful expressions for updating estimates of $\mathbf{Q}$ and $\mathsf{p}$ (See Appendix 5.B for derivations):

$$\mathbf{Q}_{k+1} = (\,\mathrm{tr}(\boldsymbol{\Psi} \mathbf{H}_{\mathbf{B}}^{-1} \boldsymbol{\Psi}^{\mathrm{T}}) \boldsymbol{\Sigma}^{-1})^{-1} \int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^T [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt \Big|_k \tag{5.19}$$

and

$$\boldsymbol{\Sigma}_{k+1} = n^{-1}\left\{[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U_m}, \boldsymbol{\theta})]\,[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} + tr(\boldsymbol{\Psi}\mathbf{H_B^{-1}}\boldsymbol{\Psi}^{\mathrm{T}})\mathbf{I}\right\}\Big|_k \tag{5.20}$$

where $\boldsymbol{\Psi}$ is:

$$\boldsymbol{\Psi} = \left[\boldsymbol{\Phi}(t_1) \quad \boldsymbol{\Phi}(t_2) \quad \ldots \quad \boldsymbol{\Phi}(t_q)\right]^{\mathrm{T}} \tag{5.21}$$

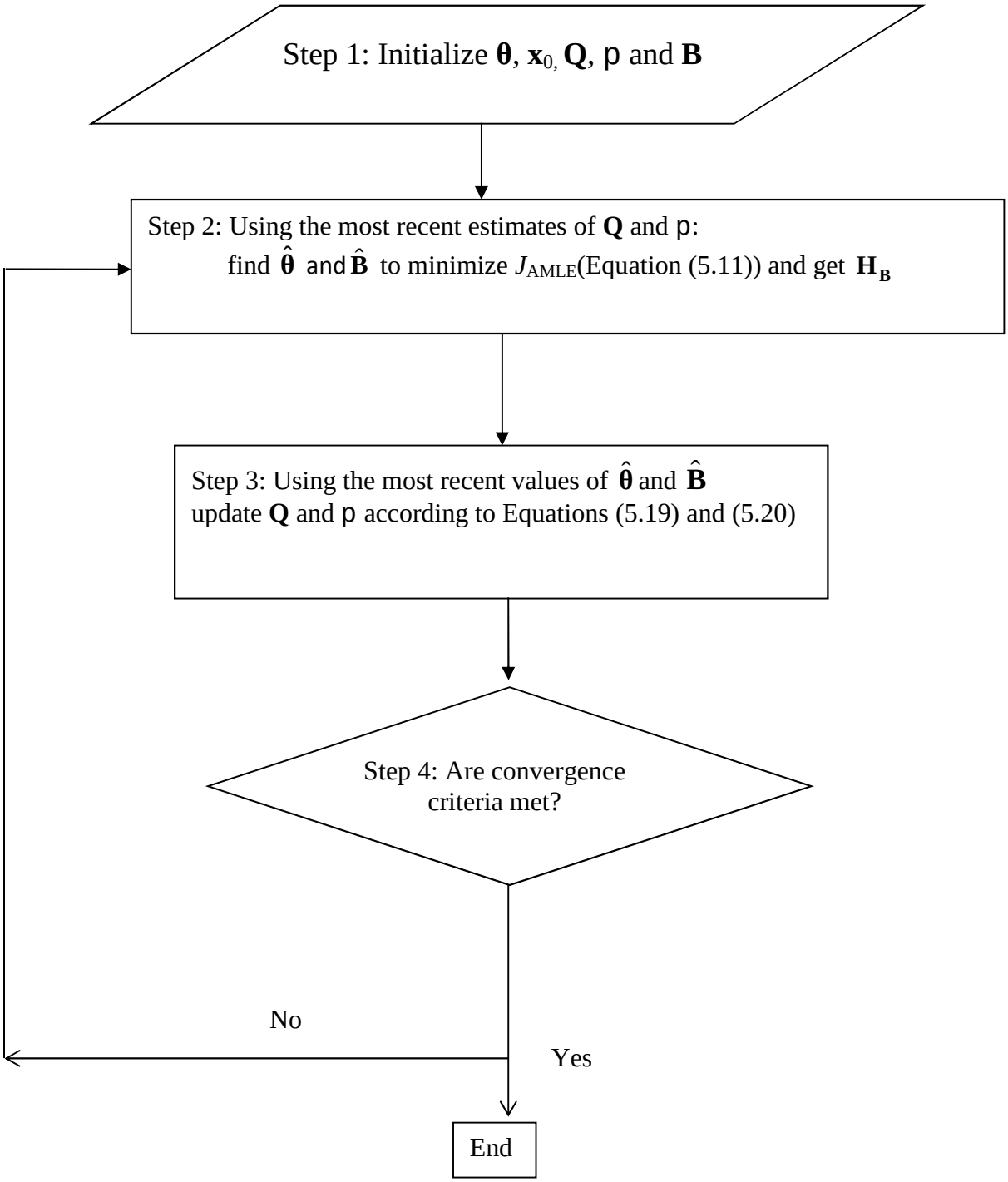
and $\mathbf{H_B}$ is the Hessian matrix with respect to the B-spline coefficients:

$$\mathbf{H_B} = \frac{\partial J_{AMLE}}{\partial \mathbf{B} \partial \mathbf{B}^{\mathrm{T}}} \tag{5.22}$$

As shown in Chart 1, the first step of the proposed LAMLE algorithm involves minimizing $J_{AMLE}$ in Equation (5.11) to estimate $\boldsymbol{\theta}$ and **B,** using initial guesses for **Q** and $\mathsf{p}$. In the second step, the estimated values of $\boldsymbol{\theta}$ and **B** from the first stage are used to calculate updated values of **Q** and $\mathsf{p}$ from Equations (5.19) and (5.20). Iteration between these two steps continues until convergence is obtained. It is common for optimization software to automatically report numerical values of the Hessian matrix with respect to the decision variables (e.g., when the LAMLE algorithm is implemented using AMPL™,[52] and IPOPT[53], the Hessian matrix $\mathbf{H_B}$ is reported by the "jgh" function in IPOPT after each iteration). Thus, there is no need for analytical calculation of the required Hessian matrix using the proposed LAMLE method.

Note that the vector of spline coefficients **B** (i.e., $\boldsymbol{\zeta} = [\boldsymbol{\theta}^{\mathrm{T}}, \mathbf{Q_d^{\mathrm{T}}}, \boldsymbol{\Sigma}_{\mathbf{d}}^{\mathrm{T}}, \mathbf{B}^{\mathrm{T}}]^{\mathrm{T}}$) was not added to the parameter vector $\boldsymbol{\zeta}$ before deriving a closed form for the likelihood function $L(\boldsymbol{\zeta} \mid \mathbf{Y_m})$. If **B** is included in the parameter vector, conditions required for approximating the integral in Equation (5.14) does not hold anymore.[33,42]

Chart 1.The LAMLE algorithm



## 5.5 Illustrative Example: Nonlinear Two-State CSTR Model

In this section, parameters in a two-state nonlinear CSTR model[54] with stochastic disturbance terms[29] are estimated to illustrate the use of the proposed LAMLE algorithm. Parameter estimation results obtained using LAMLE method are also compared with those obtained from CTSM[1] and FLAEM[46]. The two SDEs, which describe dynamic changes in the concentration of reactant A and the reactor temperature are:

$$\frac{\mathrm{d}C_A(t)}{\mathrm{d}t} = \frac{F(t)}{V}(C_{A0}(t) - C_A(t)) - k_r(T(t))C_A(t) + ¦_C(t) \tag{5.23.a}$$

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{F(t)}{V}(T_0(t) - T(t)) + UA(T(t) - T_{cin}(t)) + \gamma k_r(T(t))C_A(t) + \eta_T(t) \tag{5.23.b}$$

$$y_C(t_{m1,j}) = C_A(t_{m1,j}) + \varepsilon_C(t_{m1,j}) \qquad \text{for } j = 1 \ldots n_C \tag{5.23.c}$$

$$y_T(t_{m1,j}) = T(t_{m1,j}) + \varepsilon_T(t_{m1,j}) \qquad \text{for } j = 1 \ldots n_T \tag{5.23.d}$$

$C_A(0)=1.569$ (kmol.m$^{-3}$) (5.23.e)

T(0)=341.37 (K) (5.23.f)

where $k_r$ is the reaction rate constant:

$$k_r(T) = k_{ref}\ \exp\left(-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \tag{5.24}$$

*UA* is a heat transfer coefficient:

$$UA(F_c) = \frac{aF_c^{b+1}}{V\rho c_p\left(F_c + \dfrac{aF_c^b}{2\rho_c c_{pc}}\right)} \tag{5.25}$$

and ¥ is:

$$\gamma = \frac{(-\Delta H_{rxn})}{\rho c_p} \tag{5.26}$$

In Equations (5.23.a) and (5.23.b):

$$E\{\dagger_C(t_i)\dagger_C(t_j)\} = Q_C\delta(t_i - t_j) \tag{5.27}$$

$$E\{\dagger_T(t_i)\dagger_T(t_j)\} = Q_T\delta(t_i - t_j) \tag{5.28}$$

In Equations (5.23.c) and (5.23.d), $\pounds_C(t_{mj})\ \ j=1\ldots n_C$ and $\pounds_T(t_{mj})\ \ j=1\ldots n_T$ are white noise sequences with variances $\pm_C^2$ and $\pm_T^2$, respectively. The concentration $C_A$ is measured $n_C$ times and the temperature $T$ is measured $n_T$ times at equally spaced sample times. We assume that $\eta_C$, $\eta_T$, $\varepsilon_C$ and $\varepsilon_T$ are independent. The model inputs are: the feed flow rate $F$, the inlet concentration $C_{A0}$, the inlet temperature $T_0$, the coolant inlet temperature $T_{cin}$ and the flow rate of coolant to the cooling coil, $Fc$. The known constants for this CSTR model are shown in Table 5.1.[54]

**Table 5.1 Model constants[54]**

| **Model Constants** | **Value** | **Units** |
|---|---|---|
| $c_p$ | 4186.8 | $J\cdot kg^{-1}\cdot K^{-1}$ |
| $c_{pc}$ | 4186.8 | $J\cdot kg^{-1}\cdot K^{-1}$ |
| $T_{ref}$ | 350 | K |
| $V$ | 1 | $m^3$ |
| $\rho$ | 1000 | $kg\cdot m^{-3}$ |
| $\Delta H_{rxn}$ | $-544.154\times10^3$ | $J\cdot kmol^{-1}$ |

Simulated experiments with measurement noise and stochastic disturbances were performed using the "ode45" solver in MATLAB™ and the input trajectories shown in Figure 5.1.

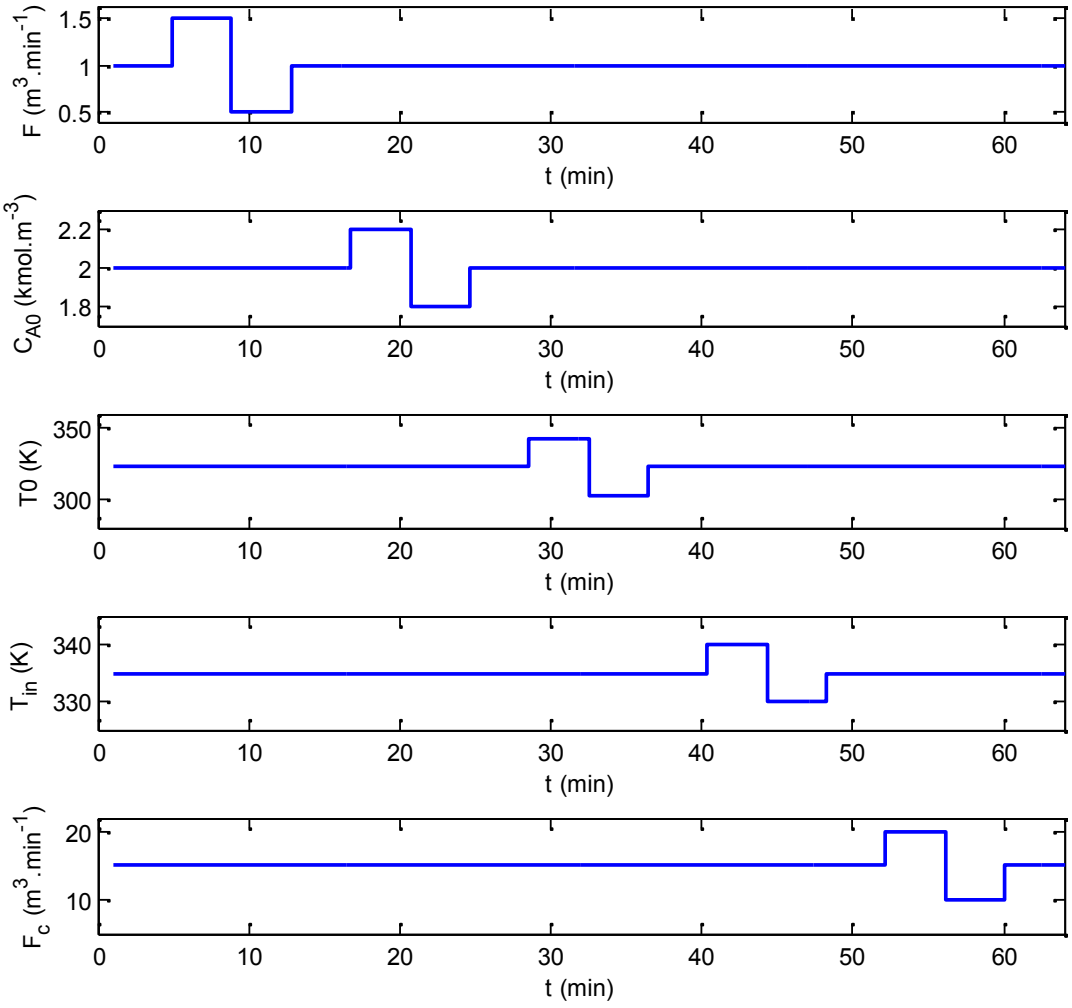



**Figure 5.1 Input trajectories for nonlinear CSTR[49]**

In the simulations, the continuous white noise disturbances were approximated using discrete white noise sequences with a sampling interval of $\Delta t$=0.5 min which is

approximately 10 times smaller than the dominant time constant of the CSTR system. To illustrate the handling of initial conditions that are perfectly known and also measured initial conditions, it is assumed that the initial concentration $C_A(0)$ is perfectly known, but the initial temperature $T(0)$ is measured with a known variance of $S_T^2$ =5.0 $K^2$. Since the true value of the initial temperature is unknown, it must be estimated. But there is no need to include $T(0)$ explicitly in the list of decision variables selected by the optimizer because $T(0)$ can be calculated from the estimated B-spline coefficients for the temperature trajectory. The perfectly known value $C_A(0)$=1.569 kmol.$m^{-3}$ can be enforced by setting the corresponding first spline coefficient $\beta_{C,1}$ =1.569.[29] We assume that measurement noise variances $\pm_C^2$ ,$\pm_T^2$ and process disturbance intensities $Q_C$ , $Q_T$ are unknown so that the complete list of parameters that needs to be estimated includes the model parameters $\boldsymbol{\theta}_{\text{CSTR}} = [k_{ref}, E/R, a, b]^T$, the disturbance intensities $Q_C$, $Q_T$, the measurement noise variances $\pm_C^2$, $\pm_T^2$ and the B-spline coefficients _ (except for $\beta_{C,1}$). The true values of parameters, which were used to generate the simulated data, are shown at the top of Table 5.2.

**Table 5.2 True parameter values, median values and IQRs for the estimates based on 100 Monte Carlo runs for different scenarios.**

| **Parameter** | | | $k_{ref}$ | $(E/R)/10^3$ | $a/10^6$ | $b$ | $T(0)$ | $Q_C$ | $Q_T$ | $\pm\frac{2}{C}$ | $\pm\frac{2}{T}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Unit** | | | $min^{-1}$ | K | | | K | $kmol^2 \cdot m^{-6} \cdot min^{-1}$ | $K^2 \cdot min^{-1}$ | $kmol^2 \cdot m^{-6}$ | $K^2$ |
| **True Value** | | | **0.461** | **8.3301** | **1.678** | **0.50** | **341.38** | **0.010** | **4.0** | **$4\times10^{-4}$** | **0.64** |
| **Scenario** | | | | | | | | | | | |
| I | CTSM | Median | 0.464 | 8.3300 | 1.560 | 0.52 | 341.36 | 0.095 | 0.6 | 0.00000 | 1.03 |
| | | IQR | 0.016 | 0.2240 | 0.520 | 0.11 | 1.05 | 0.008 | 1.3 | 0.00000 | 0.34 |
| | FLAEM | Median | 0.429 | 8.2130 | 1.448 | 0.50 | 341.30 | 0.009 | 4.1 | 0.00037 | 0.66 |
| | | IQR | 0.017 | 0.2061 | 0.424 | 0.09 | 1.08 | 0.006 | 1.8 | 0.00018 | 0.26 |
| | LAMLE | Median | 0.434 | 8.2420 | 1.536 | 0.50 | 342.07 | 0.013 | 10.7 | 0.00050 | 0.86 |
| | | IQR | 0.020 | 0.2175 | 0.458 | 0.09 | 0.91 | 0.003 | 2.2 | 0.00022 | 0.29 |
| II | FLAEM | Median | 0.444 | 8.3164 | 1.603 | 0.49 | 341.30 | 0.011 | 4.0 | 0.00036 | 0.64 |
| | | IQR | 0.016 | 0.2265 | 0.472 | 0.09 | 1.10 | 0.004 | 1.7 | 0.00019 | 0.35 |
| | LAMLE | Median | 0.444 | 8.3102 | 1.598 | 0.49 | 341.67 | 0.011 | 6.3 | 0.00039 | 0.63 |
| | | IQR | 0.022 | 0.2275 | 0.480 | 0.09 | 0.86 | 0.003 | 1.4 | 0.00014 | 0.16 |
| III | FLAEM | Median | 0.449 | 8.3090 | 1.657 | 0.49 | 341.30 | 0.010 | 4.3 | 0.00041 | 0.63 |
| | | IQR | 0.020 | 0.2547 | 0.489 | 0.09 | 1.12 | 0.005 | 2.0 | 0.00018 | 0.32 |
| | LAMLE | Median | 0.448 | 8.3273 | 1.652 | 0.49 | 341.30 | 0.010 | 5.1 | 0.00039 | 0.64 |
| | | IQR | 0.021 | 0.2403 | 0.493 | 0.10 | 1.07 | 0.003 | 1.1 | 0.00011 | 0.16 |

**Table 5.2 Continued, True parameter values, median values and IQRs for the estimates based on 100 Monte Carlo runs for different scenarios.**

| **Parameter** | | | $k_{ref}$ | $(E/R)/10^3$ | $a/10^6$ | $b$ | $T(0)$ | $Q_C$ | $Q_T$ | $\pm\frac{2}{C}$ | $\pm\frac{2}{T}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Unit** | | | $min^{-1}$ | K | | | K | $kmol^2 \cdot m^{-6} \cdot min^{-1}$ | $K^2 \cdot min^{-1}$ | $kmol^2 \cdot m^{-6}$ | $K^2$ |
| **True Value** | | | **0.461** | **8.3301** | **1.678** | **0.50** | **341.38** | **0.010** | **4.0** | **$4\times10^{-4}$** | **0.64** |
| **Scenario** | | | | | | | | | | | |
| IV | FLAEM | Median | 0.448 | 8.3402 | 1.527 | 0.50 | 341.20 | 0.004 | 2.1 | 0.00008 | 0.16 |
| | | IQR | 0.025 | 0.2576 | 0.696 | 0.16 | 1.15 | 0.002 | 0.8 | 0.00013 | 0.20 |
| | LAMLE | Median | 0.449 | 8.3311 | 1.626 | 0.50 | 341.29 | 0.010 | 5.2 | 0.00040 | 0.60 |
| | | IQR | 0.021 | 0.2506 | 0.470 | 0.10 | 1.08 | 0.003 | 1.0 | 0.00001 | 0.16 |
| V | FLAEM | Median | 0.447 | 8.3240 | 1.523 | 0.50 | 341.26 | 0.009 | 4.0 | 0.00037 | 0.65 |
| | | IQR | 0.020 | 0.2308 | 0.620 | 0.14 | 1.06 | 0.005 | 2.3 | 0.00019 | 0.32 |
| | LAMLE | Median | 0.450 | 8.3406 | 1.577 | 0.50 | 341.27 | 0.010 | 5.7 | 0.00037 | 0.62 |
| | | IQR | 0.022 | 0.2206 | 0.648 | 0.14 | 1.06 | 0.004 | 1.2 | 0.00012 | 0.17 |
| VI | FLAEM | Median | 0.439 | 8.2012 | 1.512 | 0.50 | 341.27 | 0.009 | 4.1 | 0.00043 | 0.63 |
| | | IQR | 0.046 | 0.5383 | 0.822 | 0.17 | 1.06 | 0.005 | 2.0 | 0.00020 | 0.31 |
| | LAMLE | Median | 0.449 | 8.3702 | 1.540 | 0.50 | 341.27 | 0.009 | 10.2 | 0.00037 | 0.61 |
| | | IQR | 0.051 | 0.5393 | 0.932 | 0.16 | 1.07 | 0.003 | 2.5 | 0.00016 | 0.21 |
| VII | FLAEM | Median | 0.448 | 8.2883 | 1.526 | 0.51 | 341.31 | 0.005 | 2.1 | 0.00038 | 0.67 |
| | | IQR | 0.015 | 0.2035 | 0.296 | 0.07 | 1.07 | 0.003 | 0.9 | 0.00018 | 0.32 |
| | LAMLE | Median | 0.450 | 8.3246 | 1.567 | 0.50 | 341.30 | 0.005 | 3.8 | 0.00036 | 0.61 |
| | | IQR | 0.017 | 0.1860 | 0.323 | 0.07 | 1.06 | 0.001 | 0.7 | 0.00013 | 0.12 |

**Table 5.2 Continued, True parameter values, median values and IQRs for the estimates based on 100 Monte Carlo runs for different scenarios.**

| **Parameter** | | | $k_{ref}$ | $(E/R)/\ 10^3$ | $a/10^6$ | $b$ | $T(0)$ | $Q_C$ | $Q_T$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Unit** | | | $min^{-1}$ | K | | | K | $kmol^2 \cdot m^{-6} \cdot min^{-1}$ | $K^2 \cdot min^{-1}$ | $kmol^2 \cdot m^{-6}$ | $K^2$ |
| **True Value** | | | **0.461** | **8.3301** | **1.678** | **0.50** | **341.38** | **0.010** | **4.0** | **$4\times10^{-4}$** | **0.64** |
| **Scenario** | | | | | | | | | | | |
| VII | FLAEM | Median | 0.445 | 8.3152 | 1.639 | 0.48 | 341.28 | 0.017 | 6.2 | 0.00036 | 0.57 |
| | | IQR | 0.023 | 0.4083 | 0.658 | 0.16 | 1.12 | 0.008 | 3.8 | 0.00025 | 0.42 |
| | LAMLE | Median | 0.447 | 8.3294 | 1.657 | 0.47 | 341.30 | 0.024 | 8.2 | 0.00037 | 0.66 |
| | | IQR | 0.022 | 0.3636 | 0.664 | 0.14 | 1.14 | 0.016 | 2.2 | 0.00012 | 0.18 |
| IX | FLAEM | Median | 0.448 | 8.3414 | 1.622 | 0.49 | 341.32 | 0.009 | 3.8 | 0.00020 | 0.35 |
| | | IQR | 0.017 | 0.1947 | 0.458 | 0.09 | 0.77 | 0.004 | 1.9 | 0.00010 | 0.17 |
| | LAMLE | Median | 0.448 | 8.3409 | 1.631 | 0.49 | 341.32 | 0.010 | 4.9 | 0.00020 | 0.32 |
| | | IQR | 0.018 | 0.2201 | 0.457 | 0.09 | 0.79 | 0.003 | 1.0 | 0.00006 | 0.09 |
| X | FLAEM | Median | 0.448 | 8.2810 | 1.610 | 0.50 | 341.26 | 0.009 | 3.7 | 0.00079 | 1.36 |
| | | IQR | 0.020 | 0.2371 | 0.538 | 0.11 | 1.44 | 0.005 | 2.0 | 0.00039 | 0.69 |
| | LAMLE | Median | 0.447 | 8.2784 | 1.651 | 0.49 | 341.27 | 0.010 | 5.4 | 0.00074 | 1.25 |
| | | IQR | 0.024 | 0.2512 | 0.583 | 0.12 | 1.45 | 0.003 | 1.3 | 0.00024 | 0.30 |
| XI | FLAEM | Median | 0.436 | 8.5984 | 1.707 | 0.47 | 341.27 | 0.001 | 6.4 | | 0.23 |
| | | IQR | 0.055 | 0.3768 | 0.912 | 0.20 | 1.17 | 0.001 | 2.8 | | 0.15 |
| | LAMLE | Median | 0.442 | 8.6068 | 1.708 | 0.47 | 341.27 | 0.002 | 5.0 | | 0.70 |
| | | IQR | 0.052 | 0.4320 | 0.965 | 0.21 | 1.15 | 0.002 | 1.7 | | 0.28 |

The objective function used by LAMLE for estimating the model parameters **θ** and the B-spline coefficients in the CSTR model is:

$$J_{\text{AMLE,CSTR}} = \frac{1}{\pm_{C,k}^2}\sum_{j=1}^{n_C}\left(y_C(t_{m\,1,j}) - C_{A\sim}(t_{m\,1,j})\right)^2 + \frac{1}{\pm_{T,k}^2}\sum_{j=1}^{n_T}\left(y_T(t_{m\,2,j}) - T_{\sim}(t_{m\,2,j})\right)^2$$
$$+ \frac{(T_m(0) - T_{\sim}(0))^2}{S_T^2}$$
$$+ \frac{1}{Q_{C,k}}\int_{t_0}^{t_q}\left(\frac{\mathrm{d}C_{A\sim}(t)}{\mathrm{d}t} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r(T_{\sim}(t))C_{A\sim}(t)\right)^2 \mathrm{d}t +$$
$$+ \frac{1}{Q_{T,k}}\int_{t_0}^{t_q}\left(\frac{\mathrm{d}T_{\sim}(t)}{\mathrm{d}t} - \frac{F(t)}{V}(T_0(t) - T_{\sim}(t)) - UA[T_{\sim}(t) - T_{cin}(t)] - ¥k_r(T_{\sim}(t))C_{A\sim}(t)\right)^2 \mathrm{d}t \tag{5.29}$$

The associated expressions for updating disturbance intensities $Q_C$, $Q_T$ and measurement noise variances $\pm_C^2$, $\pm_T^2$ in the CSTR model are:

$$\pm_{C,k+1}^2 = \frac{\sum_{j=1}^{n_C}\left(C_A(t_{m\,C,j}) - C_{A\sim}(t_{m\,C,j})\right)^2 + tr(\mathbf{H}_C{}^{-1})}{n_C} \tag{5.30}$$

$$Q_{C,k+1} = \frac{\int_{t_0}^{t_q}\left(\frac{dC_{A\sim}(t)}{dt} - \frac{F(t)}{V}(C_{A0}(t) - C_{A\sim}(t)) + k_r(T_{\sim}(t))C_{A\sim}(t)\right)^2 dt}{\frac{1}{\pm_{C,k}^2}tr(\mathbf{H}_C{}^{-1})} \tag{5.31}$$

$$\pm_{T,k+1}^2 = \frac{\sum_{j=1}^{n_C}\left(T(t_{m\,T,j}) - T_{\sim}(t_{m\,T,j})\right)^2 + tr(\mathbf{H}_T{}^{-1})}{n_T} \tag{5.32}$$

$$Q_{T,k+1} = \frac{\int_{t_0}^{t_q}\left(\frac{dT_{\sim}(t)}{dt} - \frac{F(t)}{V}(T_0(t) - T_{\sim}(t)) - UA(T_{\sim}(t) - T_{cin}(t)) - ¥k_r g(T_{\sim}(t))C_{A\sim}(t)\right)^2 dt}{\frac{1}{\pm_{T,k}^2}tr(\mathbf{H}_T{}^{-1})}$$

(5.33)

where $\mathbf{H}_C$ and $\mathbf{H}_T$ are Hessian matrixes:

$$\mathbf{H}_C = \frac{\partial^2 J_{AMLE,CSTR}}{\partial \boldsymbol{\beta}_C^2} \tag{5.34}$$

$$\mathbf{H}_T = \frac{\partial^2 J_{AMLE,CSTR}}{\partial \boldsymbol{\beta}_T^2} \tag{5.35}$$

In Equations (5.42) and (5.43), $\boldsymbol{\beta}_C$ and $\boldsymbol{\beta}_T$ are the B-spline coefficients corresponding to the estimated state trajectories $C_{A\sim}$ and $T_{\sim}$, respectively. Hessians were computed using the "gjh" function in IPOPT. "ghj" compute Hessians using a second order difference. Objective function (5.29) was optimized using the IPOPT solver[53] with model information provided by AMPL™.[52] Optimization settings in IPOPT were set at their default values. Default values for CTSM optimization settings were also employed when CTSM was used to estimate the parameters. For all simulation studies, cubic (4th order) B-splines were used.

Estimating the parameter vector $\boldsymbol{\zeta}_{\text{CSTR}} = [k_{ref}, E/R, a, b, Q_C, Q_T, \pm_C^2, \pm_T^2]^T$ using LAMLE required a two-step procedure wherein $k_{ref}$, $E/R$, $a$, $b$ and the B-spline coefficients were estimated by optimizing objective function 29 using assumed values of $Q_C$ ,$Q_T$, $\pm_C^2$ and $\pm_T^2$. Updated values of $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ were then calculated using Equations (5.30) to (5.33). These two steps were repeated until convergence was obtained.

Estimates of the disturbance and noise parameters were deemed to have converged when the change in the relative error $e(k)$ was less than $10^{-3}$ where $e(k)$ is defined as:

$$e(k) = \left(\frac{Q_{C,k} - Q_{C,k-1}}{Q_{C,k}}\right)^2 + \left(\frac{Q_{T,k} - Q_{T,k-1}}{Q_{T,k}}\right)^2 + \left(\frac{\pm^2_{mC,k} - \pm^2_{mC,k-1}}{\pm^2_{C,k}}\right)^2 + \left(\frac{\pm^2_{mT,k} - \pm^2_{mT,k-1}}{\pm^2_{T,k}}\right)^2 \quad (5.53)$$

Convergence failure was noted when the maximum numbers of iterations reached 1000.

The parameter vector $\boldsymbol{\zeta}$ was estimated under a variety of scenarios. Each scenario was repeated 100 times using 100 simulated data sets. These data sets were generated using different sets of initial guesses and random sequences for the Gaussian process disturbances and measurement noise. Medians and interquartile ranges (IQRs) for the parameter estimates from the various scenarios were calculated and are shown in Table 5.2. Initial guesses for the parameter values were selected randomly from uniform distributions between 50% and 150% of their true values.

Scenarios I, II and III were studied to compare the LAMLE results with those obtained using FLAEM and CTSM, and to show the effect of B-spline knot placement on the quality of parameter estimates obtained using LAMLE and FLAEM. In the simulated experiments used in scenarios I, II and III, 128 measurements were available for concentration and 128 measurements were available for temperature (once every 0.5 minutes), but different B-spline knot placement sequences were used to implement LAMLE and FLAEM. Use of the CTSM method does not require B-spline knots, so the CTSM results at the top of Table 5.2 can be compared with the LAMLE and FLAEM results from all four scenarios. Upper bounds and lower bounds must be set for all of parameters when CTSM is used. The lower bounds for the parameters were set at zero and the upper bounds were set at 10 times the true parameter values.

Scenario I was implemented using 128 equally spaced B-spline knots (one at each measurement time). Additional spline knots were used to construct the state trajectories

in scenarios II and III. In scenario II, 256 equally spaced knots were used for the B-spline basis functions (one knot at each measurement time and one additional knot between every two measurements). The results in Table 5.2 confirm that the LAMLE parameter estimates for scenario II are better than those in scenario I. For example, the median for $Q_T$ obtained from LAMLE decreased from 10.7 $K^2 \cdot min^{-1}$ in scenario I to a value of 6.3 $K^2 \cdot min^{-1}$ in Scenario II, which is much closer to the true value of 4.0 $K^2 \cdot min^{-1}$ used to generate the simulated data sets. Also, the width of the IQR for $Q_T$ obtained from LAMLE decreased from 2.2 $K^2 \cdot min^{-1}$ to 1.4 $K^2 \cdot min^{-1}$. Note that the widths of the IQRs obtained using FLAEM are larger than those obtained using LAMLE in both Scenarios I and II. For example, $\pm\frac{2}{T}$ has an IQR of 0.159 $K^2$ for Scenario II using LAMLE, whereas the corresponding IQR from FLAEM is 0.349 $K^2$.

In Scenario III, 384 knots were used to construct the state trajectories (one knot at each measurement time and two knots between). LAMLE and FLAEM results from Scenario III are slightly better than those from Scenarios I and II, especially for $Q_T$ and $\pm\frac{2}{T}$. No noticeable bias can be observed for any of the model or noise parameters in this scenario (see the boxplots in Figures 5.2 and 5.3). When additional knots were used (not shown), there was no significant improvement compared to the results in Scenario III. As a result, for all of the additional scenarios shown in Table 5.2, the knot placement settings from Scenario III were used. Typical estimation times for a simulated data set are ~1.2 min for Scenario III in LAMLE, ~3 min for FLAEM and 3 min for CTSM, using a laptop computer with Intel® Core™ 2, Duo CPU, 1.86 GHz.

Box plots for parameter estimates obtained using CTSM, FLAEM and LAMLE for scenario III are shown in Figures 5.2 and 5.3 to compare the effectiveness of the three

methods. Convergence was obtained for only 73 of the 100 sets of simulated data and corresponding initial parameter guesses using CTSM. As a result, the CTSM box plots in Figures 5.2 and 5.3 (and the medians and IQR widths in Table 5.2) are based on these 73 sets of parameter estimates. Changing the optimization settings used in CTSM was not effective in solving this problem. All 100 sets of simulated data were used in the boxplots and IQRs obtained using FLAEM and LAMLE. When CTSM was used for parameter estimation, the estimates of model parameters were unbiased. However, estimates of noise parameters $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ are biased as shown in Figures 5.3 a), d), g) and j). In particular, the estimates of $\pm_C^2$ tend to be very close to zero, so it is difficult to see the corresponding box on the box plot. The estimates of the model parameters obtained using LAMLE and FLAEM are unbiased and of similar quality, but the estimates for $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ obtained using FLAEM have more variability than those obtained using LAMLE. This is true for all three scenarios. Note that the distributions of the parameter estimates are broad and somewhat asymmetric.

Figure 5.4 shows the true trajectories of concentration and temperature from one of the simulated experiments, along with the corresponding measurements and estimated state trajectories ($C_{A\sim}$ and $T_{\sim}$) obtained using LAMLE and the settings for Scenario III. The estimated state trajectories are close to the true state trajectories because the model parameter estimates converged to the neighbourhood of the corresponding true values, as shown in Table 5.2.

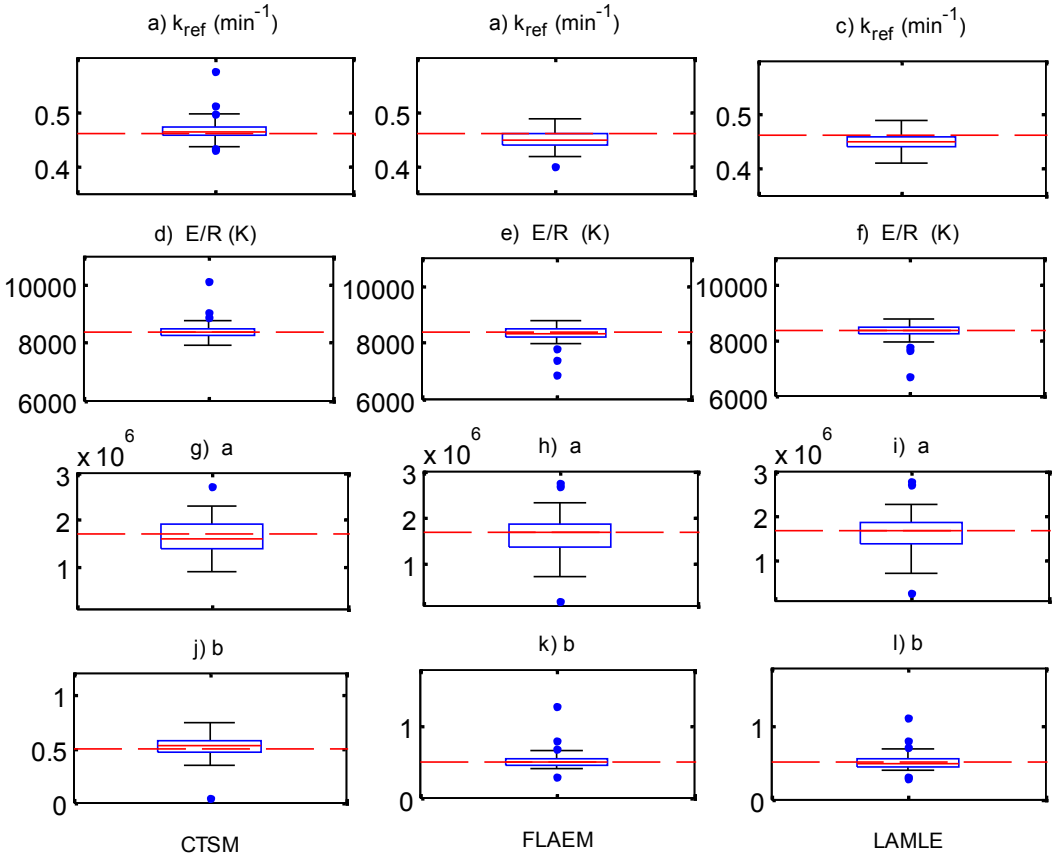



**Figure 5.2 Box-plots for estimates of model parameters using CTSM, the LAMLE and FLAEM methods in scenario I. The red dashed horizontal lines show the true values used to generate the simulated data.**

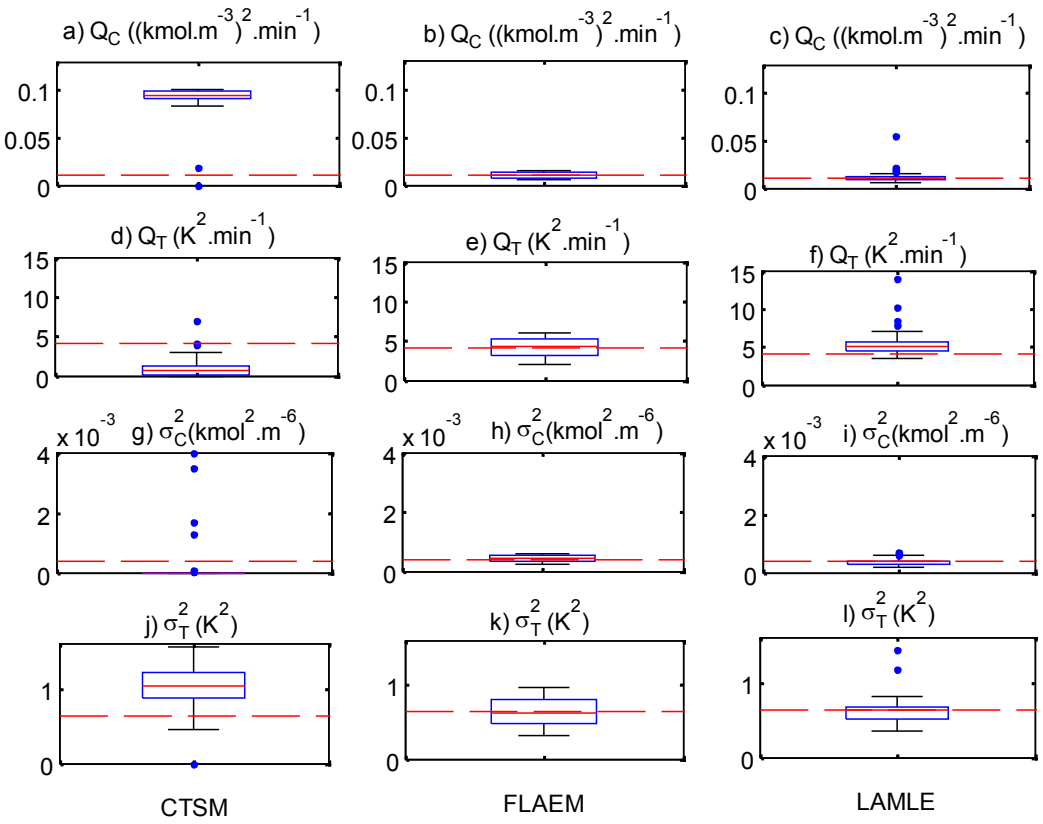



**Figure 5.3 Box-plots for disturbance intensity estimates obtained using the CTSM, the LAMLE and FLAEM methods in scenario III. The red dashed horizontal lines show the true values used to generate the simulated data.**

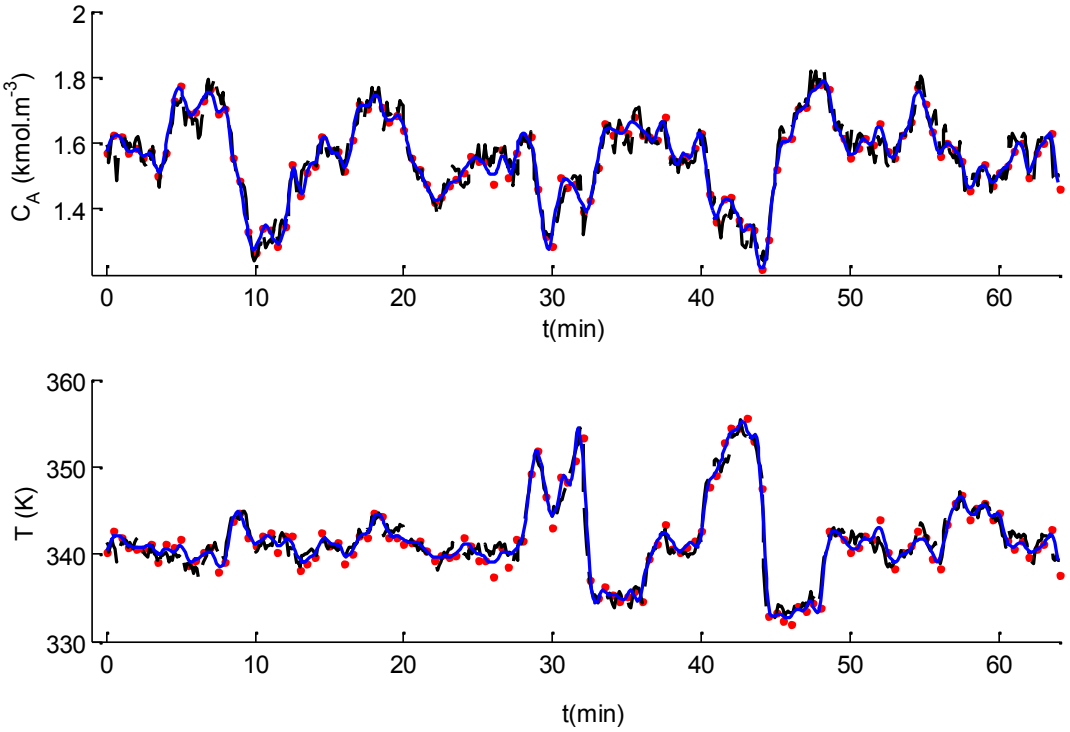



**Figure 5.4 Measured, true, and predicted concentration and temperature responses for the LAMLE method in scenario I using simulated data. Corresponding initial guesses and parameter estimates are provided in Table 5.2 (• simulated data, ----- response with true parameter values and true stochastic noise, ___predicted response using LAMLE parameter values)**

Scenario IV in Table 5.2 was used to investigate the robustness of the LAMLE and FLAEM algorithms to poorer initial guesses of the parameters. In this scenario, the initial guesses were selected randomly from a uniform distribution between 50% and 450% of the true values. The simulated data sets and the LAMLE and FLAEM settings in scenario IV are identical to those in scenario III. IQRs for LAMLE parameter estimates in this scenario are similar to those in scenario III, indicating that using worse initial guesses did not have a significant influence on the quality of the LAMLE parameter estimates. Parameter estimation attempts using CTSM with these poorer initial guesses resulted in numerous convergence failures, indicating that use of CTSM requires good initial values

for the parameters. The FLAEM parameter estimates for $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ are biased. On average, the FLAEM parameter estimates have larger variability and are much more biased than those obtained using the good initial values in Scenario III, because FLAEM converged to an undesirable local minimum for 32 out of 100 of the sets of poor initial guesses. These simulate on results suggest that LAMLE may be less prone than FLAEM to convergence to local minima when poor initial guesses are used.

Scenarios V and VI in Table 5.2 were used to study the influence of smaller data sets on the quality of the parameter estimates. In scenario V, 64 measurement for concentration and 64 measurements for temperature are available (once every minute). All of the LAMLE and FLAEM settings (i.e., knot placement, optimizer tolerances and initial parameter guesses) were the same as for Scenario III (e.g., one knot every 0.33 minutes). The median estimates for $k_{ref}$, $a$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ obtained from LAMLE are slightly worse than those for Scenario III. As expected, the IQRs for the LAMLE parameter estimates are slightly wider than those in Scenario III. Attempts to estimate parameters using the reduced number of measurements in CTSM failed due to lack of convergence in all 100 attempts, suggesting successful use of CTSM requires relatively larger data sets than when LAMLE is used. The quality of the model parameter estimates obtained from LAMLE is nearly the same as for FLAEM. However, the widths of IQRs for $Q_C$, $Q_T$, $\pm_C^2$ and $\pm_T^2$ obtained using FLAEM are noticeably larger than those obtained using LAMLE.

In scenario VI, $C_A$ and $T$ were measured even less frequently than in scenario V (only 22 equally-spaced concentration measurements and 22 temperature measurements are available for each dynamic experiment). As expected, the LAMLE parameter estimates

for Scenario VI have larger variability and are more biased than the corresponding parameter estimates in scenarios III and V due to the smaller data sets used. However, LAMLE parameter estimates from Scenario VI are still reasonably accurate. Note that bias in estimates of parameters using a relatively small number of observations is a well-known problem for all ML-based techniques.[9] CTSM could not provide parameter estimates for this scenario. As a result, use of CTSM was not attempted for the remaining scenarios in Table 5.2. The LAMLE estimates of $Q_T$ for Scenario VI are biased and are worse, on average, than the corresponding FLAEM estimates. However, the LAMLE estimates of $Q_C$, $\pm^2_C$ and $\pm^2_T$ from Scenario VI are better than the corresponding FLAEM estimates.

In scenarios VII and VIII, the values of $Q_C$ and $Q_T$ were changed to the half of their values from Scenario III (i.e., $Q_C$ =0.005 kmol$^2$ m$^{-6}$ min$^{-1}$ and $Q_T$ =2 K$^2$ min$^{-1}$) and twice their values, respectively, to examine the influence of small and large stochastic disturbances on the effectiveness of LAMLE and FLAEM. The numbers of measurements and all other settings are the same as those in Scenario III. Since smaller disturbances were used, smaller IQR values for all of the parameter estimates were obtained in Scenario VII. In scenario VIII wider IQRs were obtained for all parameters, as expected. No noticeable bias was observed for any of the parameters, except for $Q_T$ obtained from LAMLE in Scenario VIII (i.e., the median estimate was 3.8 and the true value was 2.0 K$^2$·min$^{-1}$). Obtaining accurate estimates for disturbance intensities may be more difficult when the size of the stochastic disturbances becomes small, relative to the measurement noise. In both scenarios VII and VIII, the parameter estimates obtained from FLAEM have larger variability than those obtained using LAMLE.

In scenarios IX and X, the measurement noise variances were changed to half of the true values in Scenario III (i.e., $\sigma_C^2 = 2\times10^{-4}$ kmol$^2$·m$^{-6}$ and $\sigma_T^2 = 0.32\times10^{-1}$ K$^2$) and double the values in Scenario III, respectively, when generating the simulated data. All other settings were held constant at those for Scenario III. As expected, using data with less noise led to narrower IQRs for the parameter estimates. Larger noise variances led to wider IQRs. No problems with bias were detected using either FLAEM or LAMLE.

Parameter estimates obtained using LAMLE (Scenario III) and one of the simulated data sets (see Figure 5.4) are shown in Table 5.3 along with approximate confidence intervals. These confidence intervals were obtained from:[55,56]

$$\hat{\boldsymbol{\theta}} \pm z_{\sim/2}\sqrt{\text{diag}(\partial^2 J_{AMLE}/\partial\boldsymbol{\theta}^2)^{-1}}\Big|_{\hat{\boldsymbol{\zeta}}} \tag{5.54}$$

**Table 5.3 Estimates and 95% Confidence intervals for LAMLE parameter estimates from one of the 100 Monte Carlo simulations**

| Parameter | Unit | True Value | Initial | Estimate ± 95% |
|---|---|---|---|---|
| $k_{ref}$ | min$^{-1}$ | 0.461 | 0.320 | 0.444 ± 0.036 |
| $(E/R)/10^3$ | K | 8.3301 | 6.8877 | 8.2351 ± 0.456 |
| $a/10^6$ | | 1.678 | 2.316 | 1.157 ± 0.831 |
| $b$ | | 0.50 | 0.487 | 0.62 ± 0.25 |
| $T(0)$ | K | 341.38 | 344.68 | 340.08 ± 1.43 |
| $QC$ | kmol$^2$·m$^{-6}$·min$^{-1}$ | 0.010 | 0.009 | 0.009 |
| $QT$ | K$^2$·min$^{-1}$ | 4.0 | 2.717 | 4.5 |
| $\sigma_C^2$ | kmol$^2$·m$^{-6}$ | $4\times10^{-4}$ | 0.0003 | 0.0003 |
| $\sigma_T^2$ | K$^2$ | 0.64 | 0.861 | 0.61 |

Note that these approximate confidence intervals are conditional on the values of the estimated noise parameters. Also note, that no confidence intervals for $Q_C$, $Q_T$, $\pm\sigma_C^2$ and

$\pm_T^2$ are shown. In theory, it would be possible to obtain confidence intervals for all of the model and noise parameters from:

$$\hat{\boldsymbol{\zeta}} \pm z_{\sim/2}\sqrt{\text{diag}(\partial^2 J_{LAMLE}/\partial\boldsymbol{\zeta}^2)^{-1}\Big|_{\hat{\boldsymbol{\zeta}}}} \tag{5.55}$$

However, it is very difficult to obtain appropriate numerical or analytical values for the corresponding Hessian in Equation (5.55) because second derivatives of objective function (5.16) with respect to the model parameters would be required. Problems arise due to the complexity of the $\ln(\det(\mathbf{H}_{\mathbf{x}\sim}))$ term. Recall that the LAMLE method does not require the optimizer to use objective function (5.16). Rather, this optimization problem is solved by optimization of objective function (5.10) and iterative updating of the noise parameters using expressions (5.19) and (5.20). As a result, the overall Hessian required in Equation (5.55) cannot be obtained automatically from the optimizer.

In Scenario XI, parameters were estimated using only the temperature measurements, with concentration unmeasured. The knot placements, number of measurements and initial parameter guesses for LAMLE and FLAEM are the same as those in Scenario III. Since no concentration data were obtained, the sum of the squared error terms for concentration and $n_C \ln\pm_C^2$ term do not appear in the LAMLE and FLAEM objective functions. Biased estimates were obtained for $Q_C$ from LAMLE, but estimates for all other parameters are unbiased. As expected, IQRs are wider than in scenario III. Estimates for $k_{ref}$, $Q_C$, $Q_T$ and $\pm_T^2$ obtained from FLAEM have noticeable bias. In general, the results obtained using LAMLE are significantly better than those obtained using FLAEM in this situation where one of the states is unmeasured.

In summary, results in Table 5.2 and Figures 5.2 and 5.3 show that the proposed LAMLE method provides more accurate and reliable parameter estimates than FLAEM and CTSM for the CSTR example studied. Because LAMLE requires optimization of a single objective function, its computation times is shorter than for FLAEM which requires optimization of three objective functions. Results in Table 5.2 suggest that LAMLE is more robust than FLAEM in situations involving poor initial parameter guesses or unmeasured state variables. These promising results using LAMLE suggest that using the LA to approximate the likelihood function provides adequate accuracy for estimating parameters. It is recommended that the proposed LAMLE method should be tested on larger-scale parameter estimation problems and that the performance of LAMLE should be compared with MCMC-based methods. The computation time for LAMLE is expected to be significantly lower since MCMC techniques can require large numbers of particles to approximate high-dimensional probability density functions.[27,57]

## 5.6 Conclusions

A method is proposed for estimating model parameters, process disturbance intensities and measurement variances in stochastic differential equation (SDE) models. The proposed method uses the Laplace Approximation (LA) to approximate the likelihood function and B-spline basis functions to approximate the state trajectories. The proposed Laplace Approximation Maximum Likelihood Estimation (LAMLE) method is tested using a two-state nonlinear CSTR model with 8 unknown parameters (i.e., a kinetic rate constant, an activation energy, two heat-transfer parameters, two stochastic disturbance intensities and two measurement noise variances). The parameter estimates obtained

using LAMLE compared favourably with estimates obtained using a two competing approximate maximum likelihood methods: i) the continuous time stochastic modeling (CTSM) method of Kristensen,[1] which relies on linearization-based Kalman filtering techniques, ii) fully Laplace approximation expectation maximization (FLAEM) method. The LAMLE parameter estimates were less biased and more accurate than corresponding estimates obtained using CTSM and FLAEM. LAMLE did not experience any of the convergence difficulties faced by CTSM when relatively poor initial parameter guesses were used and when measurement data were relatively sparse. LAMLE is also more robust to poor initial guesses than FLAEM. For the example studied, LAMLE is more efficient for the cases where there are unmeasured states than FLAEM and CTSM. The LAMLE method was easier to implement and converged faster than the FLAEM method since LAMLE requires optimizing one objective function rather than optimizing three objective functions required in FLAME. In future, it will be advantageous if the LAMLE algorithm can be extended to account for a wider range of estimation problems involving prior knowledge about some or all of the parameters. In future, it will be desirable to investigate the convergence of LAMLE algorithm.

## 5.7 References

1. Kristensen NR, Madsen H. Continuous time stochastic modelling: CTSM 2.3 User's Guide, Technical University of Denmark. 2003.

2. Karimi H, McAuley KB. An Approximate Expectation Maximization Algorithm for Estimating Parameters in Nonlinear Dynamic Models with Process Disturbances. *Can J Chem Eng*. 2013.

3. Maria G. A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chem Biochem Eng Q*. 2004;18:195-222.

4. Gagnon L, MacGregor J. State estimation for continuous emulsion polymerization. *Can J Chem Eng*. 1991;69:648-656.

5. Srivastava R, Anderson DF, Rawlings JB. Comparison of finite difference based methods to obtain sensitivities of stochastic chemical kinetic models. *J Chem Phys*. 2013;138:074110.

6. King R. Applications of Stochastic Differential Equations to Chemical-Engineering Problems-An Introductory Review. *Chem Eng Commun*. 1974;1:221-237.

7. Érdi P, Tóth J. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models.* Manchester University Press, 1989.

8. Söderström T, Stoica P. *System identification.* Prentice-Hall, Inc., 1988.

9. Ljung L. *System identification.* Wiley Online Library, 1999.

10. Jang S, Gopaluni R. Parameter estimation in nonlinear chemical and biological processes with unmeasured variables from small data sets. *Chem Eng Sci*. 2011;66:2774-2787.

11. Chen T, Morris J, Martin E. Particle filters for state and parameter estimation in batch processes. *J Process Control*. 2005;15:665-673.

12. Chitralekha SB, Prakash J, Raghavan H, Gopaluni R, Shah SL. A comparison of simultaneous state and parameter estimation schemes for a continuous fermentor reactor. *J Process Control*. 2010;20:934-943.

13. Doucet A, De Freitas N, Gordon N. *Sequential monte carlo methods.* , 2001.

14. Doucet A, Tadić VB. Parameter estimation in general state-space models using particle methods. *Annals of the institute of Statistical Mathematics*. 2003;55:409-422.

15. Gopaluni RB. Nonlinear system identification under missing observations: The case of unknown model structure. *J Process Control*. 2010;20:314-324.

16. Gopaluni RB. Identification of nonlinear processes with known model structure under missing observatrions. Proceedings of the 17th World Congress, Korea, July 6-11 2008;11.

17. Gopaluni R. A particle filter approach to identification of nonlinear processes under missing observations. *Can. J. Chem. Eng.* 2008;86:1081-1092.

18. Kantas N, Doucet A, Singh SS, Maciejowski JM. An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. *Proc. IFAC Symposium on System Identification*. 2009.

19. Imtiaz SA, Roy K, Huang B, Shah SL, Jampana P. Estimation of states of nonlinear systems using a particle filter. *ICIT IEEE.* 2006:2432-2437.

20. Aït-Sahalia Y. Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. *Econometrica*. 2003;70:223-262.

21. Singer H. Moment equations and Hermite expansion for nonlinear stochastic differential equations with application to stock price models. *Comput Stat*. 2006;21:385-397.

22. Lo AW. Maximum Likelihood Estimation of Generalized Itˆo Processes with Discretely Sampled Data. *Econom Theor*. 1988;4(2):231-247.

23. Singer H. Parameter estimation of nonlinear stochastic differential equations: simulated maximum likelihood versus extended Kalman filter and Itô-Taylor expansion. *J Comput Graph Stat*. 2002;11:972-995.

24. Pence BL, Fathy HK, Stein JL. A maximum likelihood approach to recursive polynomial chaos parameter estimation. 2010 American Control Conference, USA. 2010:2144-2151.

25. Lindström E. Estimating parameters in diffusion processes using an approximate maximum likelihood approach. *Ann Oper Res.* 2007;151:269-288.

26. Hurn A, Jeisman J, Lindsay K. Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. *J Financ Econ*. 2007;5:390-455.

27. Andrieu C, Doucet A, Singh SS, Tadic VB. Particle methods for change detection, system identification, and control. *Proceeding IEEE*. 2004;92:423-438.

28. Golightly A, Wilkinson DJ. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus*. 2011;1:807-820.

29. Varziri M, Poyton A, McAuley K, McLellan P, Ramsay J. Selecting optimal weighting factors in iPDA for parameter estimation in continuous-time dynamic models. *Comput Chem Eng*. 2008;32:3011-3022.

30. Varziri M, McAuley K, McLellan P. Parameter and state estimation in nonlinear stochastic continuous-time dynamic models with unknown disturbance intensity. *Can J Chem Eng*. 2008;86:828-837.

31. Kristensen NR, Madsen H, Jørgensen SB. Parameter estimation in stochastic grey-box models. *Automatica*. 2004;40:225-237.

32. Schön TB, Wills A, Ninness B. System identification of nonlinear state-space models. *Automatica*. 2011;47:39-49.

33. Barndorff-Nielsen OE, Cox DR. *Asymptotic techniques for use in statistics.* Chapman\ & Hall, 1989.

34. Shun Z, McCullagh P. Laplace approximation of high dimensional integrals. *JRSSB.* 1995:749-760.

35. Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J Comput Graph Stat*. 1995;4:12-35.

36. Pinheiro JC, Chao EC. Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *J Comput Graph Stat*. 2006;15:58-81.

37. McCulloch CE. *Generalized linear mixed models.* Wiley Online Library, 2006.

38. Demidenko E. *Mixed models: theory and applications.* Wiley-Interscience, 2005.

39. Hedeker D, Gibbons RD. *Longitudinal data analysis.* Wiley-Interscience, 2006.

40. Lee Y, Nelder JA, Pawitan Y. *Generalized linear models with random effects: unified analysis via H-likelihood.* Chapman and Hall/CRC, 2006.

41. Picchini U, Ditlevsen S. Practical estimation of high dimensional stochastic differential mixed-effects models. *Comput Stat Data Anal*. 2011;55:1426-1444.

42. Evangelou E, Zhu Z, Smith RL. Estimation and prediction for spatial generalized linear mixed models using high order Laplace approximation. *J Stat Plan Infer*. 2011;141:3564-3577.

43. Tierney L, Kass RE, Kadane JB. Approximate marginal densities of nonlinear functions. *Biometrika*. 1989;76:425-433.

44. Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*. 1986;81:82-86.

45. Heald J, Stark J. Estimation of noise levels for models of chaotic dynamical systems. *Phys Rev Lett*. 2000;84:2366-2369.

46. Karimi H, McAuley KB. An Approximate Expectation Maximization Algorithm for Estimating Parameters, Noise Variances and Stochastic Disturbance Intensities in Nonlinear Dynamic Models. *Comput Chem Eng*. 2013.

47. Jazwinski AH. *Stochastic processes and filtering theory.* Academic press, 1970.

48. Ramsay J. *Functional data analysis.* Wiley Online Library, 2005.

49. Poyton A, Varziri MS, McAuley KB, McLellan P, Ramsay JO. Parameter estimation in continuous-time dynamic models using principal differential analysis. *Comput Chem Eng*. 2006;30:698-708.

50. De Boor C. *A practical guide to splines.* Springer, 1978.

51. Overgaard RV, Jonsson N, Tornøe CW, Madsen H. Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *Journal of pharmacokinetics and pharmacodynamics*. 2005;32:85-107.

52. Fourer R, Gay DM, Kernighan BW. *A modeling language for mathematical programming.* Informs, 1990.

53. Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program*. 2006;106:25-57.

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

55. Kay SM. *Fundamentals of Statistical Signal Processing, Vol. I and Vol. II.* Prentice Hall, 1998.

56. Varziri MS, McAuley KB, McLellan PJ. Parameter estimation in continuous-time dynamic models in the presence of unmeasured states and nonstationary disturbances. *Ind Eng Chem Res*. 2008;47:380-393.

57. Poyiadjis G, Doucet A, Singh SS. Maximum likelihood parameter estimation in general state-space models using particle methods. *Proc of the American Stat. Assoc*. 2005.

58. Bodewig E. *Matrix calculus.* North-Holland Amsterdam, 1959.

# Conclusions

# Chapter 6

## 6.1 Summary

In this thesis, appropriate statistical methods to overcome two types of problems that occur during parameter estimation in chemical engineering models were studied. The first problem is having too many parameters to estimate from limited data while assuming that the model structure is correct. The second problem is estimating unmeasured disturbances and model parameters when the data are sufficient to estimate all of the parameters. In the first part of this thesis, a model for non-oxidative thermal degradation of nylon 66 was developed and used to illustrate the first problem and to test statistical methods that had recently been developed by other students in our research group. In the second part of this thesis, new techniques were proposed for estimating parameters in nonlinear dynamic models with process disturbances and model mismatch. These new techniques were tested and compared with literature methods using a nonlinear two-state SDE model for a CSTR.

In Chapter 2, an improved kinetic model was developed for thermal degradation of molten nylon 66.[1] This nylon 66 degradation model was developed in cooperation with a previous Ph.D. student, Mark Schaffer who performed the experiments that were used in parameter estimation. Schaffer had developed an earlier version of the model as part of his Ph.D. thesis[2] but he was unable to obtain reliable parameter estimates so that he could adequately test the model predictions. Like many other fundamental models of chemical

processes, this model had too many parameters to estimate using the available data set, even though Schaffer had done a large number of dynamic experiments. At the time when the model was developed, some of the current tools did not exist to determine which parameters and how many parameters should be estimated using the data. My objective in this modeling work was to use the recently developed tools to determine which parameters could be estimated from Schaffer's data and to obtain the best possible estimates of the parameters. Using these parameter estimates, the quality of the model predictions was assessed and some of Schaffer's assumptions and the corresponding model equations were revised to achieve a better match between the data and the model. The nylon degradation model contains 14 ordinary differential equations (ODEs). Total number of the unknown parameters and unknown initial conditions in the system is 56. Estimability analysis[3,4] was used to rank the parameters from most estimable to least estimable and a mean-squared error criterion was used to determine the appropriate number of parameters to estimate from the ranked list.[5,6] When estimability analysis was performed 44 of the 56 parameters could be ranked before numerical problems were encountered. Wu's MSE-based criterion[5,6] was used to determine that the top 43 parameters should be estimated using the data to provide the best model predictions, given the limited data available for parameter estimation. Predictions from this revised model agree well with the experimental data. The resulting degradation model will assist industrial nylon producers in selection of operation conditions and design of improved equipment. The fit between the data and the model predictions showed some minor discrepancies which may indicate the existence of model mismatch or unmodelled process disturbances.

After this initial parameter estimation study for the ODE model, three algorithms were developed for estimating model parameters, process disturbance intensities and measurement noise variances in nonlinear SDE models. The three corresponding objective functions for parameter estimation can be used in cases when measurement-noise variance is known, but process-disturbance intensities are not known. Two of these objective functions can also be used for the more complicated situation when the measurement noise variances are also unknown. Some of the benefits of the proposed methods that may be attractive to developers of fundamental dynamic models are: i) simplicity of implementation, ii) reliable estimates of model parameters, initial conditions and disturbance intensities, iii) efficient handling of unknown initial states, iv) ability to handle unmeasured state variables. In particular, estimates of disturbance intensities can provide modelers with information about the degree of mismatch and the magnitude of unmeasured disturbances in their models. This information will be helpful when implementing on-line state and parameter estimation schemes for process monitoring and control.[7,8]

In Chapter 3, an Approximate Expectation Maximization (AEM) method for estimating parameters and process disturbance intensities in nonlinear SDE models was proposed for situations where measurement noise variances are known. The AEM method approximates the expectation step of the EM algorithm using B-spline state trajectories. The effectiveness of the AEM algorithm was tested using a two-state nonlinear stochastic CSTR model. Parameter estimates were compared with those from Extended AMLE[9] and the CTSM software of Kristensen, which relies on a linearization-based maximum likelihood method. For the CSTR example studied, the resulting AEM parameter

estimates are less biased, more precise and more robust to poor initial parameter guesses than those obtained using Extended AMLE and CTSM. The AEM method was easier to set up and converged faster than extended AMLE. Approximate confidence intervals for model parameters and disturbance intensities were determined by approximating the covariance matrix using the Hessian evaluated at the estimates of the parameters and the B-spline coefficients.

In Chapter 4, a Fully-Laplace-Approximation Expectation Maximization (FLAEM) algorithm was developed for parameter estimation in SDE models. In the first stage of this iterative algorithm, model parameters are estimated using Varziri's AMLE objective function, assuming that the disturbance intensities and measurement noise variances are known. In the second stage, disturbance intensity and noise variance estimates are updated. The expressions used to update disturbance intensities and noise variances were derived by approximating the E step of the EM algorithm using the FLA and B-spline basis functions. The FLAEM algorithm iterates between these two steps until convergence is obtained. The FLAEM methodology was tested using the two-state nonlinear CSTR model. More accurate estimates for model parameters, disturbance intensities and measurement noise variances were obtained using FLAEM when compared with CTSM, particularly when the number of measurements was relatively small or when initial guesses of parameters were relatively poor.

In Chapter 5, an even better approximate maximum likelihood algorithm was developed for approximating model parameters, process disturbance intensities and measurement variances in nonlinear SDE models. The proposed Laplace Approximation Maximum Likelihood Estimation (LAMLE) method uses the Laplace Approximation (LA) and B-

spline basis functions to approximate the likelihood function of the parameters given the measurements. The LAMLE method was tested using the CSTR model, revealing that LAMLE parameter estimates are more accurate and less biased than corresponding estimates obtained using CTSM and FLAEM.

The case studies conducted in this thesis suggest that AEM, FLAEM and especially LAMLE are potentially appealing parameter estimation algorithms that should be further studied and tested for more complicated problems. Note that AEM, FLAEM and LAMLE circumvent potential problems associated with finding a closed form for the likelihood function at the expense of solving a large nonlinear programming problem. Fast and efficient nonlinear programming solvers such as IPOPT[10] are essential for the successful implementation of these algorithms, especially for larger problems.

### 6.2 Recommendation for Future Work

1. In Chapter 2, a model for non-oxidative thermal degradation of nylon 66 was developed assuming that the model structure is perfect. The fit to the data showed some discrepancies indicating the possible existence of model mismatch or process disturbances. In future, modeling errors and process disturbances could be considered during parameter estimation in the nylon 66 model to improve the model predictions. This type of analysis would require additional experimental data. In particular, replicate experiments would be useful for assessing whether nonstationary disturbances should be included in some of the differential equations.

2. The AEM, FLAEM and LAMLE techniques were tested successfully using a CSTR case study in Chapters 3-5. In future, it will be desirable to test the proposed

methodologies using larger-scale dynamic models with a larger number of states and parameters.

3. It will also be important to compare AEM, FLAEM and LAMLE results with those from recently developed MLE-based methods that use MCMC techniques for parameter estimation in SDE models.[11,12] It is expected that computation times for the proposed methods in this thesis will be significantly lower than the times required using MCMC methods, particularly for larger-scale problems, because the proposed methods do not require sampling from high dimensional probability density functions. It will be important to determine whether the additional B-spline and mode approximations used to develop the proposed methods result in any significant degradation in the quality of parameter estimates when compared with MCMC methods.

4. In chemical engineering applications, usually some prior information about the parameters is available. In many cases, modelers have some physical insights about physically realistic values of the parameters. In future, it will be advantageous to extend the FLAEM or LAMLE algorithm to account for prior knowledge about some or all of the parameters. Box and Draper[13] introduced the use of Bayesian methods for estimating the parameters of chemical engineering models to include prior knowledge of parameters. In general, in the Bayesian method, the approximate conditional density function of the parameters given the measured data $p(\boldsymbol{\zeta} \mid \mathbf{Y}_m)$ is maximized to estimate the unknown parameters. The conditional density function of the parameters given the observed data $p(\boldsymbol{\zeta} \mid \mathbf{Y}_m)$ is called a posterior density function. The posterior density function is:[14]

$$p(\boldsymbol{\zeta} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\zeta})\, p(\boldsymbol{\zeta})}{p(\mathbf{y})} \tag{6.1}$$

The numerator on the right-hand side is the product of the ML probability density function $p(\mathbf{Y_m}|\boldsymbol{\zeta})$ and the prior distribution of the parameters $p(\boldsymbol{\zeta})$. The denominator, which ensures that the posterior integrates to unity, does not depend on the parameter values. The prior density function $p(\boldsymbol{\zeta})$ represents knowledge about the possible values of $\boldsymbol{\zeta}$ before the data that will be used for parameter estimation have been observed. The prior probability is very important when very little observation data are available and must restrict the parameters to meaningful values.

5. It is suggested to test the proposed methods for estimating the parameters in SDE models that contain additional unmeasured states to account for non-stationary disturbances and to develop statistical methods to assess when and where non-stationary disturbances should be included to improve the quality of model predictions.

6. Obtaining approximate confidence intervals for model parameters and process disturbances for the AEM method was discussed in Chapter 3. When FLAEM and LAMLE are used, the corresponding approximate confidence intervals are conditional on the estimated values of the disturbance intensities and measurement noise variances. In future, it will be important to develop approaches for determining inference regions for all of the noise and model parameter estimates obtained from FLAEM and LAMLE.

7. In future it will be advantageous to investigate the convergence of the proposed methods.

8. It is recommended to show the predictive performance of the proposed methods for predicting state trajectories based on infinite horizon predictions.

## 6.3 References

(1) Karimi, H.; Schaffer, M. A.McAuley, K. B. A Kinetic Model for Non-Oxidative Thermal Degradation of Nylon 66. *Macromolecular Reaction Engineering* **2012**, *2-3*, 93.

(2) Schaffer, M. A., Non-Oxidative Thermal Degradation of Nylon, Queen's University Canada, 2003.

(3) Thompson, D. E.; McAuley, K. B.McLellan, P. J. Parameter Estimation in a Simplified MWD Model for HDPE Produced by a Ziegler-Natta Catalyst. *Macromolecular Reaction Engineering* **2009**, *4*, 160.

(4) Yao, K. Z.; Shaw, B. M.; Kou, B.; McAuley, K. B.Bacon, D. Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design. *Polymer Reaction Engineering* **2003**, *3*, 563.

(5) Wu, S.; McAuley, K.Harris, T. Selection of Simplified Models: II. Development of a Model Selection Criterion Based on Mean Squared Error. *The Canadian Journal of Chemical Engineering* **2011**, *2*, 325.

(6) Wu, S.; McAuley, K.Harris, T. Selection of Simplified Models: I. Analysis of model-selection Criteria using mean-squared Error. *The Canadian Journal of Chemical Engineering* **2011**, *1*, 148.

(7) MacGregor, J. F.; Penlidis, A.Hamielec, A. E. Control of Polymerization Reactors-a Review. *Polym. Proc. Eng.* **1984**, 179.

(8) Jazwinski, A. H. *Stochastic Processes and Filtering Theory*; Academic press: 1970.

(9) Varziri, M.; McAuley, K.McLellan, P. Parameter and State Estimation in Nonlinear Stochastic continuous-time Dynamic Models with Unknown Disturbance Intensity. *The Canadian Journal of Chemical Engineering* **2008**, *5*, 828.

(10) Wächter, A.; Biegler, L. T. On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming. *Math. Program.* **2006**, *1*, 25.

(11) Gopaluni, R. B. Nonlinear System Identification Under Missing Observations: The Case of Unknown Model Structure. *J. Process Control* **2010**, *3*, 314.

(12) Imtiaz, S. A.; Roy, K.; Biao Huang; Shah, S. L.; Jampana, P. In *In Estimation of States of Nonlinear Systems using a Particle Filter;* Industrial Technology, 2006. ICIT 2006. IEEE International Conference on; 2006; pp2432-2437.

(13) Box, M.; Draper, N. R. Estimation and Design Criteria for Multiresponse Non-Linear Models with Non-Homogeneous Variance. *Journal of the Royal Statistical Society.Series C (Applied Statistics)* **1972**, *1*, 13.

(14) Seber, G. A.; Lee, A. J. *Linear Regression Analysis*; Wiley: 2012.

## Appendix 3.A Derivation of the AEM Objective Function with Frequent Measurements

In this section, the AEM objective function is developed for the system shown in Equation (3.1), where $\mathbf{p}$ is known and $\mathbf{Q}$ is unknown. First, we consider the case where all states are measured frequently compared with the dynamics of the system so that Equation (3.1) can be approximated using the following Euler approximation:

$$\mathbf{x}(t_{mj-1} + \Delta t_m) = \mathbf{x}(t_{mj}) = \mathbf{x}(t_{mj-1}) + \mathbf{f}(\mathbf{x}(t_{mj-1}), \mathbf{u}(t_{mj-1}), \boldsymbol{\theta})\Delta t_m + \boldsymbol{\eta}_{\mathbf{d}}(t_{mj-1})\Delta t_m \quad (3.A.1a)$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \quad (3.A.1b)$$

where $\mathbf{x}(t_{mj})$ contains the values of the state variable at the measurement times $t_{mj}$ , $j$=0,..,$n$ and $\Delta t_m$ is the sampling interval. Appendix 3.B considers the case where measurements are available further apart in time and when some states may be unmeasured.

To keep the notation simple, assume that measurements are available at $n$ equally-spaced sample times $t_{mj}$ ($j$=1,…,$n$) for all outputs:

$$\mathbf{Y_m} = [\mathbf{y}^{\mathrm{T}}(t_{m1})\ldots\mathbf{y}^{\mathrm{T}}(t_{mn})]^T = [y_1(t_{m1})\ldots y_1(t_{mn})\ldots y_Y(t_{m1})\ldots y_Y(t_{mn})]^T$$
$$\mathbf{X_m} = [\mathbf{x}^{\mathrm{T}}(t_{m1})\ldots\mathbf{x}^{\mathrm{T}}(t_{mn})]^T = [x_1(t_{m1})\ldots x_1(t_{mn})\ldots x_Y(t_{m1})\ldots x_Y(t_{mn})]^T$$

First, we find a closed form expression for $p(\mathbf{Y_m}, \mathbf{X_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta})$.

By the law of total conditional probability, the density function $p(\mathbf{Y_m}, \mathbf{X_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta})$ can be decomposed to:

$$p(\mathbf{Y_m}, \mathbf{X_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = p(\mathbf{Y_m} \mid \mathbf{X_m}, \mathbf{x_m}(t_0), \boldsymbol{\zeta})\ p(\mathbf{X_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) \quad (3.A.2)$$

Jazwinski[46] showed that the stochastic process $\{x(t_i),\ i<q+1\}$ generated in Equation (3.10a) is a Markov Process. Thus, by the Markov property of the states $p(\mathbf{X_m}\,|\,\boldsymbol{\zeta})$ can be written as:[46,47,52]

$$p(\mathbf{X_m}\,|\,\boldsymbol{\zeta}) = \prod_{j=1}^{n} p(\mathbf{x}(t_{mj})\,|\,\mathbf{x}(t_{mj-1}),\boldsymbol{\zeta}) \tag{3.A.3a}$$

When $\mathbf{x}(t_0)$ is approximated using a measurement at time zero, the following probability expression applies:

$$p(\mathbf{X_m},\mathbf{x_m}(t_0)\,|\,\boldsymbol{\zeta}) = p(\mathbf{x_m}(t_0)\,|\,\boldsymbol{\zeta}) \prod_{j=1}^{n} p(\mathbf{x}(t_{mj})\,|\,\mathbf{x}(t_{mj-1}),\boldsymbol{\zeta}) \tag{3.A.3b}$$

From Equation (3.A.1a) $p(\mathbf{x}(t_{mj})\,|\,\mathbf{x}(t_{mj-1}),\boldsymbol{\zeta})$ has a multivariate Gaussian distribution:[52]

$$\begin{aligned} p(\mathbf{x}(t_{mj})\,|\,\mathbf{x}(t_{mj-1}),\boldsymbol{\zeta}) = {} & \frac{\mathbf{1}}{\sqrt{2®}}[\det(\mathbf{Q})]^{-\frac{1}{2}}\,\Delta t_m{}^{-\frac{1}{2}} \\ & \exp\{-\frac{1}{2}[\mathbf{x}(t_{mj}) - \mathbf{x}(t_{mj-1}) - \mathbf{f}(\mathbf{x}(t_{mj-1}),\mathbf{u}(t_{mj-1}),\boldsymbol{\theta})\Delta t_m]^{\mathrm{T}}\mathbf{Q}^{-1}\Delta t_m{}^{-1} \\ & \times[\mathbf{x}(t_{mj}) - \mathbf{x}(t_{mj-1}) - \mathbf{f}(\mathbf{x}(t_{mj-1}),\mathbf{u}(t_{mj-1}),\boldsymbol{\theta})\Delta t_m]\} \end{aligned} \tag{3.A.4}$$

Assuming a Gaussian distribution for initial measured state $\mathbf{x_m}(t_0)$ gives:

$$p(\mathbf{x_m}(t_0)\,|\,\boldsymbol{\zeta}) = \frac{[\det(\mathbf{S}_{m0})]^{-\frac{1}{2}}}{\sqrt{2\pi}} \exp[-\frac{1}{2}(\mathbf{x}_{m0} - \mathbf{x}_0)^T\,\mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0} - \mathbf{x}_0)] \tag{3.A.5}$$

where $\mathbf{x}_0$ is the vector of initial conditions for the state variables and $\mathbf{x}_{\mathrm{m0}}$ is a vector of multivariate normal random variables with mean $\mathrm{E}\{\mathbf{x}_{m0}\} = \mathbf{x}_0$ and $\mathbf{c}ov\{\mathbf{x}_{m0}\} = \mathbf{S}_{\mathbf{m0}}$.

Substituting Equations (3.A.4) and (3.A.5) into Equation (3.A.3b):

$$p(\mathbf{X_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = \frac{[\det(\mathbf{S}_{m0})]^{-\frac{1}{2}}}{\sqrt{2\pi}} \exp[-\frac{1}{2}(\mathbf{x}_{m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0} - \mathbf{x}_0)] \times$$

$$\prod_{j=1}^{n} \frac{\mathbf{1}}{\sqrt{2\pi}} [\det(\mathbf{Q})]^{-\frac{1}{2}} \Delta t_m^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathbf{x}(t_{mj}) - \mathbf{x}(t_{mj-1}) - \mathbf{f}(\mathbf{x}(t_{mj-1}), \mathbf{u}(t_{mj-1}), \boldsymbol{\theta}) \Delta t_m]^T$$

$$\mathbf{Q}^{-1} \Delta t_m^{-1} \times [\mathbf{x}(t_{mj}) - \mathbf{x}(t_{mj-1}) - \mathbf{f}(\mathbf{x}(t_{mj-1}), \mathbf{u}(t_{mj-1}), \boldsymbol{\theta}) \Delta t_m]\}$$

(3.A.6)

From Equation (3.1c) $p(\mathbf{Y_m} \mid \mathbf{X_m}, \mathbf{x_m}(t_0), \boldsymbol{\zeta})$ has a multivariate Gaussian distribution:

$$p(\mathbf{Y_m} \mid \mathbf{X_m}, \mathbf{x_m}(t_0), \boldsymbol{\zeta}) = \prod_{j=1}^{n} \frac{\mathbf{1}}{\sqrt{2\pi}} [\det(\boldsymbol{\Sigma})]^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]^T \boldsymbol{\Sigma}^{-1}$$

$$[\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]\}$$

(3.A.7)

Substituting Equations (3.A.6) and (3.A.7) into Equation (3.A.1), gives:

$$p(\mathbf{Y_m}, \mathbf{X_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = \prod_{j=1}^{n} \frac{\mathbf{1}}{\sqrt{2\pi}} [\det(\boldsymbol{\Sigma})]^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]^T \boldsymbol{\Sigma}^{-1}\}$$

$$[\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})] \times \frac{[\det(\mathbf{S}_{m0})]^{-\frac{1}{2}}}{\sqrt{2\pi}} \exp[-\frac{1}{2}(\mathbf{x}_{m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0} - \mathbf{x}_0)]$$

$$\times \prod_{j=1}^{n} \frac{\mathbf{1}}{\sqrt{2\pi}} [\det(\mathbf{Q})]^{-\frac{1}{2}} \Delta t_m^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathbf{x}(t_{mj}) - \mathbf{x}(t_{mj-1}) - \mathbf{f}(\mathbf{x}(t_{mj-1}), \mathbf{u}(t_{mj-1}), \boldsymbol{\theta}) \Delta t]^T$$

$$\mathbf{Q}^{-1} \Delta t_m^{-1} \times [\mathbf{x}(t_{mj}) - \mathbf{x}(t_{mj-1}) - \mathbf{f}(\mathbf{x}(t_{mj-1}), \mathbf{u}(t_{mj-1}), \boldsymbol{\theta}) \Delta t_m]\}$$

(3.A.8)

Taking the negative natural logarithm of Equation (3.A.8) and collecting the constant terms into a single term $C_1$:

$$
\begin{aligned}
-\ln\, p(\mathbf{Y_m}, \mathbf{X_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = \\
& C_1 + \sum_{j=1}^{n} [\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})] \\
& + (\mathbf{x}_{m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_0) + n \ln[\det(\mathbf{Q})] \\
& + \sum_{j=1}^{n} [\mathbf{x}(t_{mj}) - \mathbf{x}(t_{mj-1}) - \mathbf{f}(\mathbf{x}(t_{mj-1}), \mathbf{u}(t_{mj-1}), \boldsymbol{\theta}) \Delta t]^T \mathbf{Q}^{-1} \Delta t_m^{-1} \\
& \times [\mathbf{x}(t_{mj}) - \mathbf{x}(t_{mj-1}) - \mathbf{f}(\mathbf{x}(t_{mj-1}), \mathbf{u}(t_{mj-1}), \boldsymbol{\theta}) \Delta t_m]
\end{aligned}
$$

((3.A.9)

Let

$$
J_{\text{AEM}} = -\ln\, p(\mathbf{Y_m}, \mathbf{X_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) - C_1 \tag{3.A.10}
$$

Since $\boldsymbol{\Delta} t_m$ is small, the last term in Equation (3.A.9) can be approximated by an integral:

$$
\begin{aligned}
J_{\text{AEM}} = & \sum_{j=1}^{n} [\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})] \\
& + (\mathbf{x}_{m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_0) + n \ln[\det(\mathbf{Q})] \\
& + \int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^T \mathbf{Q}^{-1} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt
\end{aligned} \tag{3.A.11}
$$

Approximating the states of the system using B-spline basis functions gives:

$$
\begin{aligned}
J_{\text{AEM}} = & \sum_{j=1}^{n} [\mathbf{g}(\mathbf{x}_{\sim}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{g}(\mathbf{x}_{\sim}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})] \\
& + (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0})^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_{\sim 0}) + n \ln[\det(\mathbf{Q})] \\
& + \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]^T \mathbf{Q}^{-1} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt
\end{aligned} \tag{3.A.12}
$$

In order to use Equation (3.16), the mode of $J_{\text{AEM}}$ with respect to states should be found:

$$
\frac{\partial J_{\text{AEM}}}{\partial \mathbf{X_m}} = 0 \tag{3.A.13}
$$

Since B-splines are used to represent the states, and the continuous state trajectory is directly related to the B-spline coefficients, the mode value of $J_{\text{AEM}}$ occurs where

$$\frac{\partial J_{\text{AEM}}}{\partial \mathbf{B}} = 0 \tag{3.A.14}$$

Thus, finding the mode of $J_{\text{AEM}}$ with respect to states is equal to minimizing $J_{\text{AEM}}$ with respect to the B-spline coefficients. Minimizing Equation (3.A.11) with respect to the parameters ensures that $\mathbf{X}_{\mathbf{m}\tilde{}}$ is at the mode of $J_{\text{AEM}}$.

$$\begin{aligned} J_{\text{AEM}} = {} & (\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\mathbf{m}\tilde{}}, \mathbf{U_m}, \boldsymbol{\theta}))^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\mathbf{m}\tilde{}}, \mathbf{U_m}, \boldsymbol{\theta})) \\ & + (\mathbf{x}_{m0} - \mathbf{x}_{\tilde{}0})^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_{\tilde{}0}) - n \ln[\det(\mathbf{Q})] \\ & + \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\tilde{}}(t) - \mathbf{f}(\mathbf{x}_{\tilde{}}(t), \mathbf{u}(t), \boldsymbol{\theta})]^T \mathbf{Q}^{-1} [\dot{\mathbf{x}}_{\tilde{}}(t) - \mathbf{f}(\mathbf{x}_{\tilde{}}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt \end{aligned} \tag{3.17}$$

Thus, using B-spline basis functions to represent state trajectories eliminates the expectation step of the EM algorithm. Equation (3.17) is the corresponding overall AEM objective function that should be minimized by selecting appropriate values of $\boldsymbol{\theta}$, $\mathbf{Q}$ and $\mathbf{B}$.

## Appendix 3.B Derivation of the AEM Objective Function with Missing Data

In this section, an AEM objective function is developed for the system shown in Equation (3.1) for the more complicated case where measurements are only available at longer sampling intervals or not all states are measured at the same times. The situation where some states are not measured at all is also considered.

An Euler scheme with a short discretization interval $\Delta t$, which is shorter than the measurement intervals, is used for discrete-time approximation of the SDE model in Equation (3.1) (shown in Equation (3.10a)). First, we find a closed form for

$p(\mathbf{Y_m}, \mathbf{X_q}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta})$, where the subscripts **q** and **m** indicate that we are interested in values of the state variables at *q* discrete times, but measurements are available at only some of these times.

By the law of total conditional probability, the density function $p(\mathbf{Y_m}, \mathbf{X_q}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta})$ can be decomposed to:[46,47]

$$p(\mathbf{Y_m}, \mathbf{X_q}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = p(\mathbf{Y_m} \mid \mathbf{X_q}, \mathbf{x_m}(t_0), \boldsymbol{\zeta})\ p(\mathbf{X_q}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) \tag{3.B.1}$$

By the Markov property of the states, $p(\mathbf{X_q}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta})$ can be written as:[46,47]

$$p(\mathbf{X_q}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = p(\mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) \prod_{i=1}^{q} p(\mathbf{x}(t_i) \mid \mathbf{x}(t_{i-1}), \boldsymbol{\zeta}) \tag{3.B.2}$$

From Equation (3.10a) $p(\mathbf{x}(t_i) \mid \mathbf{x}(t_{i-1}), \boldsymbol{\zeta})$ has a Gaussian distribution:[52]

$$\begin{aligned} p(\mathbf{x}(t_i) \mid \mathbf{x}(t_{i-1}), \boldsymbol{\zeta}) = {} & \frac{\mathbf{1}}{\sqrt{2®}} [\det(\mathbf{Q})]^{-\frac{1}{2}} \boldsymbol{\Delta}t^{-\frac{1}{2}} \\ & \exp\{-\frac{1}{2}[\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\boldsymbol{\Delta}t]^T \mathbf{Q}^{-\frac{1}{2}} \boldsymbol{\Delta}t^{-\frac{1}{2}} \\ & \times [\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\boldsymbol{\Delta}t]\} \end{aligned} \tag{3.B.3}$$

Assume a Gaussian distribution for the initial measurements of the state variables:[46]

$$p(\mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = \frac{[\det(\mathbf{S}_{m0})]^{-\frac{1}{2}}}{\sqrt{2\pi}} \exp[-\frac{1}{2}(\mathbf{x}_{m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_0)] \tag{3.B.4}$$

where $\mathbf{x}_{m0}$ is a multivariate normal random variable corresponding to the measurements of the state variables at the initial time $t_0$, with mean $\mathrm{E}\{\mathbf{x}_{m0}\} = \mathbf{x}_0$ and $\mathbf{c}ov\{\mathbf{x}_{m0}\} = \mathbf{S_{m0}}$.

Substituting Equations (3.B.3) and (3.B.4) into Equation (3.B.2):

$$p(\mathbf{X_q}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = \frac{[\det(\mathbf{S}_{m0})]^{-\frac{1}{2}}}{\sqrt{2\pi}} \exp[-\frac{1}{2}(\mathbf{x}_{m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0} - \mathbf{x}_0)] \times$$
$$\prod_{i=1}^{q} \frac{\mathbf{1}}{\sqrt{2\pi}} [\det(\mathbf{Q})]^{-\frac{1}{2}} \Delta t^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t]^T \mathbf{Q}^{-1} \Delta t^{-1}$$
$$\times [\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t]\} \tag{3.B.5}$$

From Equation (3.1c), $p(\mathbf{Y_m} \mid \mathbf{X_q}, \mathbf{x_m}(t_0), \boldsymbol{\zeta})$ has a multivariate Gaussian distribution:[47,52]

$$p(\mathbf{Y_m} \mid \mathbf{X_q}, \mathbf{x_m}(t_0), \boldsymbol{\zeta}) = \prod_{j=1}^{n} \frac{\mathbf{1}}{\sqrt{2\pi}} [\det(\boldsymbol{\Sigma})]^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]^T \boldsymbol{\Sigma}^{-1}[\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]\} \tag{3.B.6}$$

Substituting Equations (3.B.5) and (3.B.6) into Equation (3.B.1), gives:

$$p(\mathbf{Y_m}, \mathbf{X_q}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = \prod_{j=1}^{n} \frac{\mathbf{1}}{\sqrt{2\pi}} [\det(\boldsymbol{\Sigma})]^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]^T$$
$$\boldsymbol{\Sigma}^{-1}[\mathbf{g}(\mathbf{x}(t_{mj}, \mathbf{u}(t_{mj}), \boldsymbol{\theta}) - \mathbf{y}(t_{mj})]\} \times \frac{[\det(\mathbf{S}_{m0})]^{-\frac{1}{2}}}{\sqrt{2\pi}}$$
$$\exp[-\frac{1}{2}(\mathbf{x}_{m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0} - \mathbf{x}_0)]$$
$$\times \prod_{i=1}^{q} \frac{\mathbf{1}}{\sqrt{2\pi}} [\det(\mathbf{Q})]^{-\frac{1}{2}} \Delta t^{-\frac{1}{2}} \exp\{-\frac{1}{2}[\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t]^T$$
$$\mathbf{Q}^{-1} \Delta t^{-1} \times [\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t]\} \tag{3.B.7}$$

Taking the negative natural logarithm of Equation (3.B.7) and collecting the constant terms into a single term $C_1$ gives:

$$-\ln p(\mathbf{X_q}, \mathbf{Y_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) = C_1 + (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}))^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}))$$
$$+ (\mathbf{x}_{m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0} - \mathbf{x}_0) + q \ln[\det(\mathbf{Q})]$$
$$+ \sum_{i=1}^{q} [\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta \mathbf{t}]^T \mathbf{Q}^{-1} \Delta t^{-1}$$
$$\times [\mathbf{x}(t_i) - \mathbf{x}(t_{i-1}) - \mathbf{f}(\mathbf{x}(t_{i-1}), \mathbf{u}(t_{i-1}), \boldsymbol{\theta})\Delta t]$$

(3.B.8)

Since $\Delta t$ is small, the last term in Equation (3.B.8) can be written as:

$$\sum_{i=1}^{q}[\mathbf{x}(t_i)-\mathbf{x}(t_{i-1})-\mathbf{f}(\mathbf{x}(t_{i-1}),\mathbf{u}(t_{i-1}),\boldsymbol{\theta})\Delta t]^T\mathbf{Q}^{-1}\Delta t^{-1}[\mathbf{x}(t_i)-\mathbf{x}(t_{i-1})-\mathbf{f}(\mathbf{x}(t_{i-1}),\mathbf{u}(t_{i-1}),\boldsymbol{\theta})\Delta t]$$

$$\approx \int_{t_0}^{t_q}[\dot{\mathbf{x}}(t)-\mathbf{f}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})]^T\mathbf{Q}^{-1}[\dot{\mathbf{x}}(t)-\mathbf{f}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})]dt \tag{3.B.9}$$

Let

$$J_{\text{AEM}} = -\ln\, p(\mathbf{X_q}, \mathbf{Y_m}, \mathbf{x_m}(t_0) \mid \boldsymbol{\zeta}) - C_1 \tag{3.20}$$

Substituting Equation (3.B.9) into Equation (3.B.8):

$$\begin{aligned} J_{\text{AEM}} &= [\mathbf{Y_m}-\mathbf{g}(\mathbf{X_m},\mathbf{U_m},\boldsymbol{\theta})]^T\boldsymbol{\Sigma}^{-1}[\mathbf{Y_m}-\mathbf{g}(\mathbf{X_m},\mathbf{U_m},\boldsymbol{\theta})] \\ &+ (\mathbf{x}_{m0}-\mathbf{x}_{\sim 0})^T\mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0}-\mathbf{x}_{\sim 0}) + q\ln[\det(\mathbf{Q})] \\ &+ \int_{t_0}^{t_q}[\dot{\mathbf{x}}(t)-\mathbf{f}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})]^T\mathbf{Q}^{-1}[\dot{\mathbf{x}}(t)-\mathbf{f}(\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta})]dt \end{aligned} \tag{3.B.10}$$

Using B-splines to represent the states:

$$\begin{aligned} J_{\text{AEM}} &= (\mathbf{Y_m}-\mathbf{g}(\mathbf{X_{\sim m}},\mathbf{U_m},\boldsymbol{\theta}))^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y_m}-\mathbf{g}(\mathbf{X_{\sim m}},\mathbf{U_m},\boldsymbol{\theta})) \\ &+ (\mathbf{x}_{m0}-\mathbf{x}_{\sim 0})^T\mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0}-\mathbf{x}_{\sim 0}) + q\ln[\det(\mathbf{Q})] \\ &+ \int_{t_0}^{t_q}[\dot{\mathbf{x}}_{\sim}(t)-\mathbf{f}(\mathbf{x}_{\sim}(t),\mathbf{u}(t),\boldsymbol{\theta})]^T\mathbf{Q}^{-1}[\dot{\mathbf{x}}_{\sim}(t)-\mathbf{f}(\mathbf{x}_{\sim}(t),\mathbf{u}(t),\boldsymbol{\theta})]dt \end{aligned} \tag{3.18}$$

## Appendix 4.A Derivations

In this appendix, equations for updating process intensities in **Q** and measurement noise variances in ρ are developed when **θ** is assumed to be known. These equations are derived by approximating the E step of the EM algorithm using the FLA and using B-spline basis functions to approximate the state trajectories. Note that the corresponding spline coefficients are also assumed to be known because they are estimated along with **θ**. The likelihood function of complete data $p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta})$ has a closed form, which was derived in our previous work:[27]

$$\begin{aligned} p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta}) = C_1[\det(\boldsymbol{\Sigma})]^{-\frac{n}{2}} \exp\{-\frac{1}{2}[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]\} \\ \times[\det(\mathbf{S}_{m0})]^{-\frac{1}{2}} \exp[-\frac{1}{2}(\mathbf{x}_{m0} - \mathbf{x}_0)^{\mathrm{T}} \mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0} - \mathbf{x}_0)] \\ \times[\det(\mathbf{Q})]^{-\frac{q}{2}} \exp\{-\frac{1}{2}\int_{t_0}^{t_q}[\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1}[\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]\,\mathrm{d}\,t\} \end{aligned} \tag{4.A.1}$$

where $C_1$ is a constant.

Taking the negative natural logarithm of Equation (4.A.1) gives the likelihood:

$$\begin{aligned} -\ln\, p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta}) = -\ln C_1 + \frac{n}{2}\ln[\det(\boldsymbol{\Sigma})] \\ + \frac{1}{2}[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})] \\ + \frac{1}{2}\ln[\det(\mathbf{S}_{m0})] + \frac{1}{2}(\mathbf{x}_{m0} - \mathbf{x}_0)^{\mathrm{T}} \mathbf{S}_{m0}^{-1}(\mathbf{x}_{m0} - \mathbf{x}_0) \\ + \frac{q}{2}\ln[\det(\mathbf{Q})] + \frac{1}{2}\int_{t_0}^{t_q}[\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1}[\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]\,\mathrm{d}\,t \end{aligned} \tag{4.A.2}$$

Substituting the log likelihood of the complete data from Equation (4.A.2) into the E-step of the EM algorithm (Equation (4.15)) gives:

$$\begin{aligned}\mathrm{E}\{\ln p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta}) \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k\} = \frac{1}{2}\mathrm{E}\{[&-2\ln C_1 + (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}))^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \\ &(\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})) \\ &+ \ln[\det(\mathbf{S}_{m0})] + (\mathbf{x}_{m0} - \mathbf{x}_0)^{\mathrm{T}} \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_0) \\ &+ \int_{t_0}^{t_q} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))^{\mathrm{T}} \mathbf{Q}^{-1} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))\,\mathrm{d}t \\ &+ n\ln(\det(\boldsymbol{\Sigma})) + q\ln(\det(\mathbf{Q}))] \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k\}\end{aligned} \tag{4.A.3}$$

Taking the partial derivative of Equation (4.A.3) with respect to **Q** gives:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{Q}}\mathrm{E}\{\ln p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta}) \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k\} = \mathrm{E}\{[&-\frac{1}{2}\int_{t_0}^{t_q} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))^{\mathrm{T}} \mathbf{Q}^{-1}\mathbf{Q}^{-1} \\ &\times (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))\,\mathrm{d}t + \frac{q}{2}\mathbf{Q}^{-1}] \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k\}\end{aligned} \tag{4.A.4}$$

Note that development of Equation (4.A.4) relies on the following expressions for the derivative of the determinant of a matrix:[55]

$$\frac{\partial \ln(\det(\mathbf{Q}))}{\partial \mathbf{Q}} = \mathbf{Q}^{-1} \tag{4.A.5}$$

Setting the right-hand side of Equation (4.A.4) to zero and rearranging to solve for **Q**, gives the following expression for **Q**:

$$\mathbf{Q} = \frac{1}{q}\mathrm{E}\{\int_{t_0}^{t_q} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))(\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))^{\mathrm{T}}\,\mathrm{d}t \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k\} \tag{4.A.6}$$

Similarly, setting the partial derivative of Equation (4.A.3) with respect to p to zero and solving for p gives:

$$\mathbf{\Sigma} = \frac{1}{n} \mathrm{E}\{[(\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}))(\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}))^{\mathrm{T}}] \,|\, \mathbf{Y_m}, \mathbf{Q}_k, \mathbf{\Sigma}_k\} \tag{4.A.7}$$

Recall that:

$$Z = \int_{t_0}^{t_q} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))(\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))^{\mathrm{T}} \, \mathrm{d}\, t \tag{4.19}$$

$$S = (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}))(\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}))^{\mathrm{T}} \tag{4.20}$$

Using the definition for *Z* and *S* from Equations (4.19) and (4.20) into Equations (4.A.6) and (4.A.7), the estimates of the disturbance intensity **Q** and the noise variance **Σ** at the *k*+1th iteration are:

$$\mathbf{Q}_{k+1} = \frac{1}{q} \mathrm{E}\{Z \,|\, \mathbf{Y_m}, \mathbf{Q}_k, \mathbf{\Sigma}_k\} \tag{4.21}$$

$$\mathbf{\Sigma}_{k+1} = \frac{1}{n} \mathrm{E}\{S \,|\, \mathbf{Y_m}, \mathbf{Q}_k, \mathbf{\Sigma}_k\} \tag{4.22}$$

To estimate $\mathbf{Q}_{k+1}$ and $\mathbf{\Sigma}_{k+1}$ from Equations (4.21) and (4.22) expressions for $\mathrm{E}\{Z \,|\, \mathbf{Y_m}, \mathbf{Q}_k, \mathbf{\Sigma}_k\}$ and $\mathrm{E}\{S \,|\, \mathbf{Y_m}, \mathbf{Q}_k, \mathbf{\Sigma}_k\}$ should be obtained for use in the *k*th iteration. These expectations are given by:[38]

$$\mathrm{E}(Z \,|\, \mathbf{Y_m}, \mathbf{Q}_k, \mathbf{\Sigma}_k) = \frac{\int Z \; p(\mathbf{Y_m}, \mathbf{X_q} \,|\, \boldsymbol{\zeta}_k) \, \mathrm{d}\, \mathbf{X_q}}{\int p(\mathbf{Y_m}, \mathbf{X_q} \,|\, \boldsymbol{\zeta}_k) \, \mathrm{d}\, \mathbf{X_q}} \tag{4.23}$$

$$\mathrm{E}(S \,|\, \mathbf{Y_m}, \mathbf{Q}_k, \mathbf{\Sigma}_k) = \frac{\int S \; p(\mathbf{Y_m}, \mathbf{X_q} \,|\, \boldsymbol{\zeta}_k) \, \mathrm{d}\, \mathbf{X_q}}{\int p(\mathbf{Y_m}, \mathbf{X_q} \,|\, \boldsymbol{\zeta}_k) \, \mathrm{d}\, \mathbf{X_q}} \tag{4.24}$$

Substituting $p(\mathbf{Y_m}, \mathbf{X_q} \mid \boldsymbol{\zeta}_k)$ from Equation (4.A.1) into Equations (4.A.21) and (4.22) and simplifying to remove the constants $C_1$, the second, the fourth, the fifth and the six terms in Equation (4.A.1) gives:

$$\mathrm{E}(Z \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k) =$$

$$\frac{\int \left\{ \begin{array}{l} \int_{t_0}^{t_q} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))(\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}))^{\mathrm{T}} \, \mathrm{d}t \\ \exp\{-[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)]^{\mathrm{T}} \frac{\boldsymbol{\Sigma}^{-1}}{2} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)]\} \\ \exp\{-\int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)]^{\mathrm{T}} \frac{\mathbf{Q}^{-1}}{2} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)]\} \mathrm{d}t \end{array} \right\} \mathrm{d}\,\mathbf{X_q}}{\int \left\{ \begin{array}{l} \exp\{-[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)]^{\mathrm{T}} \frac{\boldsymbol{\Sigma}^{-1}}{2} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)]\} \\ \exp\{-\int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)]^{\mathrm{T}} \frac{\mathbf{Q}^{-1}}{2} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)]\} \mathrm{d}t \end{array} \right\} \mathrm{d}\,\mathbf{X_q}} \tag{4.A.8}$$

$$\mathrm{E}(S \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k) =$$

$$\frac{\int \left\{ \begin{array}{l} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)][\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)]^{\mathrm{T}} \\ \exp\{-[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)]^{\mathrm{T}} \frac{\boldsymbol{\Sigma}^{-1}}{2} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)]\} \\ \exp\{\int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)]^{\mathrm{T}} \frac{\mathbf{-Q}^{-1}}{2} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)] \mathrm{d}\,t\} \end{array} \right\} \mathrm{d}\,\mathbf{X_q}}{\int \left\{ \begin{array}{l} \exp\{-[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)]^{\mathrm{T}} \frac{\boldsymbol{\Sigma}^{-1}}{2} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)]\} \\ \exp\{\int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)]^{\mathrm{T}} \frac{\mathbf{-Q}^{-1}}{2} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)] \mathrm{d}\,t\} \end{array} \right\} \mathrm{d}\,\mathbf{X_q}} \tag{4.A.9}$$

Note that $\mathbf{x}_0$ is assumed to be known because it is estimated from B-spline coefficients. B-spline coefficients are estimated along with $\boldsymbol{\theta}$. As a result, the fifth term in Equation (4.A.1) is constant. Since $\boldsymbol{\theta}$ is assumed to be known, the FLA can be used to approximate the integrals in Equations A.8 and A.9:

$$
\begin{aligned}
\mathrm{E}(Z \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k) = \left( \frac{\det(\mathbf{H}_{\hat{\mathbf{x}}})}{\det(\mathbf{H}^Z_{\hat{\mathbf{x}}})} \right)^{1/2} & \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}^Z(t) - \mathbf{f}(\hat{\mathbf{x}}^Z(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \\
& (\dot{\hat{\mathbf{x}}}^Z(t) - \mathbf{f}(\hat{\mathbf{x}}^Z(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \, \mathrm{d}t \\
& \exp\{ -\frac{1}{2} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^Z_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^Z_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& - \frac{1}{2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}^Z(t) - \mathbf{f}(\hat{\mathbf{x}}^Z(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}^Z(t) - \mathbf{f}(\hat{\mathbf{x}}^Z(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t \\
& + \frac{1}{2} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& + \frac{1}{2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}(t) - \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}(t) - \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t \}
\end{aligned}
\tag{4.A.10}
$$

$$
\begin{aligned}
\mathrm{E}(S \mid \mathbf{Y_m}, \mathbf{Q}_k, \boldsymbol{\Sigma}_k) = \left( \frac{\det(\mathbf{H}_{\hat{\mathbf{x}}})}{\det(\mathbf{H}^S_{\hat{\mathbf{x}}})} \right)^{1/2} & (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^S_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^S_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \\
& \exp\{ -\frac{1}{2} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^S_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^S_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& - \frac{1}{2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}^S(t) - \mathbf{f}(\hat{\mathbf{x}}^S(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}^S(t) - \mathbf{f}(\hat{\mathbf{x}}^S(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t \\
& + \frac{1}{2} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& + \frac{1}{2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}(t) - \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}(t) - \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t \}
\end{aligned}
\tag{4.A.11}
$$

where

$$\mathbf{H}_{\hat{\mathbf{x}}} = \left. \frac{\partial^2 J}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}} \right|_{\mathbf{X_q} = \hat{\mathbf{X}}_\mathbf{q}} \tag{4.A.12}$$

$$\mathbf{H}^{Z}_{\hat{\mathbf{x}}} = \left. \frac{\partial^2 J^Z}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}} \right|_{\mathbf{X_q} = \hat{\mathbf{X}}^Z_\mathbf{q}} \tag{4.A.13}$$

$$\mathbf{H}^{S}_{\hat{\mathbf{x}}} = \left. \frac{\partial^2 J^S}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}} \right|_{\mathbf{X_q} = \hat{\mathbf{X}}^S_\mathbf{q}} \tag{4.A.14}$$

$J, J^Z$ and $J^S$ in Equations (4.A.12), (4.A.13) and (4.A.14) are defined in Equations (4.A.15), (4.A.16) and (4.A.17).

$$\begin{aligned} J = {} & (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\ & + \int_{t_0}^{t_q} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t \end{aligned} \tag{4.A.15}$$

$$\begin{aligned} J^Z = {} & -\ln \int_{t_0}^{t_q} (\dot{x}_1(t) - f_1(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^2 \, \mathrm{d}t - \ldots - \ln \int_{t_0}^{t_q} (\dot{x}_x(t) - f_X(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^2 \, \mathrm{d}t \\ & + \frac{1}{2} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\ & + \frac{1}{2} \int_{t_0}^{t_q} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t \end{aligned} \tag{4.A.16}$$

$$
\begin{aligned}
J^S = & -\ln \sum_{j=1}^{N_1} [y_1(t_{m1,j}) - g_1(\mathbf{x}(t_{m1,j}), \mathbf{y}(t_{m1,j}), \boldsymbol{\theta}_k)]^2 - \ldots \\
& - \ln \sum_{j=1}^{N_Y} [y_Y(t_{mY,j}) - g_Y(\mathbf{x}(t_{mY,j}), \mathbf{y}(t_{mY,j}), \boldsymbol{\theta}_k)]^2 \\
& + \frac{1}{2} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& + \frac{1}{2} \int_{t_0}^{t_q} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \mathrm{d}t
\end{aligned}
\tag{4.A.17}
$$

Using B-spline basis functions for representing state trajectories in Equations (4.A.10) and (4.A.11) gives:

$$
\begin{aligned}
\mathrm{E}(Z \mid \mathbf{Y_m}, \boldsymbol{\zeta}_k) = & \left( \frac{\det(\mathbf{H}_{\hat{\mathbf{x}}_\sim})}{\det(\mathbf{H}^Z_{\hat{\mathbf{x}}_\sim})} \right)^{1/2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}^Z_\sim(t) - \mathbf{f}(\hat{\mathbf{x}}^Z_\sim(t), \mathbf{u}(t), \boldsymbol{\theta}_k))(\dot{\hat{\mathbf{x}}}^Z_\sim(t) - \mathbf{f}(\hat{\mathbf{x}}^Z_\sim(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathrm{d}t \\
& \exp\{ -\frac{1}{2} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^Z_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^Z_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& - \frac{1}{2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}^Z_\sim(t) - \mathbf{f}(\hat{\mathbf{x}}^Z_\sim(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}^Z_\sim(t) - \mathbf{f}(\hat{\mathbf{x}}^Z_\sim(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \mathrm{d}t \\
& + \frac{1}{2} ((\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& + \frac{1}{2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_\sim(t) - \mathbf{f}(\hat{\mathbf{x}}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}_\sim(t) - \mathbf{f}(\hat{\mathbf{x}}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \mathrm{d}t \}
\end{aligned}
\tag{4.A.18}
$$

$$
\begin{aligned}
\mathrm{E}(S \mid \mathbf{Y_m}, \boldsymbol{\zeta}_k) = & \left( \frac{\det(\mathbf{H}_{\hat{\mathbf{x}}_{\sim}})}{\det(\mathbf{H}^{S}_{\hat{\mathbf{x}}_{\sim}})} \right)^{1/2} ((\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^{S}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^{S}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \\
& \exp\{ -\frac{1}{2} ((\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^{S}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^{S}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& - \frac{1}{2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}^{S}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}^{S}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}^{S}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}^{S}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t \\
& + \frac{1}{2} ((\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& + \frac{1}{2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t \}
\end{aligned}
\tag{4.A.19}
$$

where $\mathbf{H}_{\hat{\mathbf{x}}_{\sim}}$, $\mathbf{H}^{Z}_{\hat{\mathbf{x}}_{\sim}}$ and $\mathbf{H}^{S}_{\hat{\mathbf{x}}_{\sim}}$ are defined in Equations (4.A.20), (4.A.21) and (4.A.22):

$$
\mathbf{H}_{\hat{\mathbf{x}}_{\sim}} = \left. \frac{\partial^2 J}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}} \right|_{\mathbf{X_q} = \hat{\mathbf{X}}_{\sim \mathbf{q}}} \tag{4.A.20}
$$

$$
\mathbf{H}^{Z}_{\hat{\mathbf{x}}_{\sim}} = \left. \frac{\partial^2 J^{Z}}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}} \right|_{\mathbf{X_q} = \hat{\mathbf{X}}^{Z}_{\sim \mathbf{q}}} \tag{4.A.21}
$$

$$
\mathbf{H}^{S}_{\hat{\mathbf{x}}_{\sim}} = \left. \frac{\partial^2 J^{S}}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}} \right|_{\mathbf{X_q} = \hat{\mathbf{X}}^{S}_{\sim \mathbf{q}}} \tag{4.A.22}
$$

The AMPL software used to implement the AMLE algorithm provides $\mathbf{H_B}$ the Hessian matrix with respect to **B**. The relationship between $\mathbf{H}_{\mathbf{x}\sim}$ (i.e., the Hessian matrix with respect to $\mathbf{X}_{\mathbf{q}\sim)}$ and $\mathbf{H_B}$ (i.e., the Hessian matrix with respect to **B**) is derived below.

Using the chain rule for partial derivatives to find the partial derivative of $J_{\mathrm{AMLE}}$ (Equation (4.14)) with respect to matrix of B-spline basis functions **B** gives:

$$\frac{\partial J_{\text{AMLE}}}{\partial \mathbf{B}} = \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0)} \frac{\partial \mathbf{x}_{\sim}(t_0)}{\partial \mathbf{B}} + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1)} \frac{\partial \mathbf{x}_{\sim}(t_1)}{\partial \mathbf{B}} + \cdots + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q)} \frac{\partial \mathbf{x}_{\sim}(t_q)}{\partial \mathbf{B}} \tag{4.A.23}$$

Finding partial derivatives from Equation (4.7):

$$\frac{\partial \mathbf{x}_{\sim}(t_i)}{\partial \mathbf{B}} = \mathbf{\Phi}(t_i) \tag{4.A.24}$$

Substituting partial derivatives from Equation (4.A.24) into Equation (4.A.23) gives:

$$\frac{\partial J_{\text{AMLE}}}{\partial \mathbf{B}} = \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0)} \mathbf{\Phi}(t_0) + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1)} \mathbf{\Phi}(t_1) + \cdots + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q)} \mathbf{\Phi}(t_q) \tag{4.A.25}$$

Let

$$\mathbf{G} = \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{B}} \tag{4.A.26}$$

Using the chain rule for finding the partial derivatives of **G**:

$$\frac{\partial \mathbf{G}}{\partial \mathbf{B}} = \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_0)} \frac{\partial \mathbf{x}_{\sim}(t_0)}{\partial \mathbf{B}} + \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_1)} \frac{\partial \mathbf{x}_{\sim}(t_1)}{\partial \mathbf{B}} + \cdots + \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_q)} \frac{\partial \mathbf{x}_{\sim}(t_q)}{\partial \mathbf{B}} \tag{4.A.27}$$

Substituting partial derivatives from Equation (4.A.24) into Equation (4.A.27):

$$\frac{\partial \mathbf{G}}{\partial \mathbf{B}} = \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_0)} \mathbf{\Phi}(t_0) + \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_1)} \mathbf{\Phi}(t_1) + \cdots + \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_q)} \mathbf{\Phi}(t_q) \tag{4.A.28}$$

Substituting **G** from Equations (4.A.26) and (4.A.25) into Equations (4.A.28) gives the second derivative of $J_{\text{AMLE}}$ with respect to the spline coefficients:

$$\begin{aligned}\frac{\partial \mathbf{G}}{\partial \mathbf{B}} = \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{B} \partial \mathbf{B}^{\text{T}}} = {} & \frac{\partial}{\partial \mathbf{x}_{\sim}(t_0)} \left[ \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0)} \mathbf{\Phi}(t_0) + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1)} \mathbf{\Phi}(t_1) + \cdots + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q)} \mathbf{\Phi}(t_q) \right] \mathbf{\Phi}(t_0) + \\ & \frac{\partial}{\partial \mathbf{x}_{\sim}(t_1)} \left[ \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0)} \mathbf{\Phi}(t_0) + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1)} \mathbf{\Phi}(t_1) + \cdots + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q)} \mathbf{\Phi}(t_q) \right] \mathbf{\Phi}(t_1) + \cdots + \\ & \frac{\partial}{\partial \mathbf{x}_{\sim}(t_q)} \left[ \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0)} \mathbf{\Phi}(t_0) + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1)} \mathbf{\Phi}(t_1) + \cdots + \frac{\partial J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q)} \mathbf{\Phi}(t_q) \right] \mathbf{\Phi}(t_q)\end{aligned} \tag{4.A.29}$$

Simplifying Equation (4.A.29) gives:

$$\frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{B} \partial \mathbf{B}^{\text{T}}} = \boldsymbol{\Phi}(t_0) \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0) \partial \mathbf{x}_{\sim}^{\text{T}}(t_0)} \boldsymbol{\Phi}^{\text{T}}(t_0) + \boldsymbol{\Phi}(t_1) \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0) \partial \mathbf{x}_{\sim}^{\text{T}}(t_1)} \boldsymbol{\Phi}^{\text{T}}(t_0) + \cdots$$
$$+ \boldsymbol{\Phi}(t_q) \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0) \partial \mathbf{x}_{\sim}^{\text{T}}(t_q)} \boldsymbol{\Phi}(t_0) + \boldsymbol{\Phi}(t_0) \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\sim}^{\text{T}}(t_0)} \boldsymbol{\Phi}(t_1) + \boldsymbol{\Phi}(t_1) \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\sim}^{\text{T}}(t_1)} \boldsymbol{\Phi}^{\text{T}}(t_1)$$
$$+ \cdots + \boldsymbol{\Phi}(t_q) \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\sim}^{\text{T}}(t_q)} \boldsymbol{\Phi}^{\text{T}}(t_1) + \cdots + \boldsymbol{\Phi}(t_0) \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\text{T}}(t_0)} \boldsymbol{\Phi}^{\text{T}}(t_q)$$
$$+ \boldsymbol{\Phi}(t_1) \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\text{T}}(t_1)} \boldsymbol{\Phi}^{\text{T}}(t_q) + \cdots + \boldsymbol{\Phi}(t_q) \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\text{T}}(t_q)} \boldsymbol{\Phi}^{\text{T}}(t_q)$$

(4.A.30)

In matrix form, Equation (4.A.30) becomes:

$$\begin{bmatrix} \frac{\partial^2 J_{\text{AMLE}}}{\partial \boldsymbol{\beta}_1^2} & \frac{\partial^2 J_{\text{AMLE}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2} & \cdots & \frac{\partial^2 J_{\text{AMLE}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_{c_s}} \\ \frac{\partial^2 J_{\text{AMLE}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1} & \frac{\partial^2 J_{\text{AMLE}}}{\partial \boldsymbol{\beta}_2^2} & \cdots & \frac{\partial^2 J_{\text{AMLE}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_{c_s}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J_{\text{AMLE}}}{\partial \boldsymbol{\beta}_{c_s} \partial \boldsymbol{\beta}_1} & \frac{\partial^2 J_{\text{AMLE}}}{\partial \boldsymbol{\beta}_{c_s} \partial \boldsymbol{\beta}_2} & \cdots & \frac{\partial^2 J_{\text{AMLE}}}{\partial \boldsymbol{\beta}_{c_s}^2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}(t_1) & \boldsymbol{\Phi}(t_2) & \cdots & \boldsymbol{\Phi}(t_q) \end{bmatrix} \times$$
$$\begin{bmatrix} \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0) \partial \mathbf{x}_{\sim}^{\text{T}}(t_0)} & \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0) \partial \mathbf{x}_{\sim}^{\text{T}}(t_1)} & \cdots & \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0) \partial \mathbf{x}_{\sim}^{\text{T}}(t_q)} \\ \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_0) \partial \mathbf{x}_{\sim}^{\text{T}}(t_1)} & \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\sim}^{\text{T}}(t_1)} & \cdots & \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\sim}^{\text{T}}(t_q)} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\text{T}}(t_0)} & \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\text{T}}(t_1)} & \cdots & \frac{\partial^2 J_{\text{AMLE}}}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\text{T}}(t_q)} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\Phi}(t_1) \\ \boldsymbol{\Phi}(t_2) \\ \vdots \\ \boldsymbol{\Phi}(t_q) \end{bmatrix}$$

(4.A.31)

or equivalently:

$$\mathbf{H}_{\mathbf{B}} = \boldsymbol{\Psi}^{\text{T}} \mathbf{H}_{\mathbf{x}_{\sim}} \boldsymbol{\Psi} \tag{4.A.32}$$

where $\mathbf{\Psi}$ is a matrix of B-spline basis functions defined as.

$$\mathbf{\Psi} = \left[\mathbf{\Phi}(t_1) \quad \mathbf{\Phi}(t_2) \quad \ldots \quad \mathbf{\Phi}(t_q)\right]^{\mathrm{T}} \tag{4.A.33}$$

and $H_B$ is the Hessian matrix with respect to B-spline basis functions :

$$\mathbf{H_B} = \frac{\partial J_{\mathrm{AMLE}}}{\partial \mathbf{B} \partial \mathbf{B}^{\mathrm{T}}} \tag{4.A.34}$$

The ratio of two Hessian matrixes is:

$$\frac{\mathbf{H_B}}{\mathbf{H_B^S}} = \frac{\mathbf{\Psi}^{\mathrm{T}} \mathbf{H_{x\sim}} \mathbf{\Psi}}{\mathbf{\Psi}^{\mathrm{T}} \mathbf{H_{x_\sim}^S} \mathbf{\Psi}} \tag{4.A.35}$$

Since $\mathbf{\Psi}^{\mathrm{T}}$ and $\mathbf{\Psi}$ are constants:

$$\frac{\mathbf{H_{x\sim}}}{\mathbf{H_{x_\sim}^S}} = \frac{\mathbf{H_B}}{\mathbf{H_B^S}} \tag{4.A.36}$$

Thus, in Equations (4.A.18) and (4.A.19), the Hessian matrices with respect to **B** can be used instead of the Hessian matrices with respect to $\mathbf{X_{\sim q}}$ so that:

$$\begin{aligned}
\mathbf{Q}_{k+1} = {} & \frac{1}{q}\left(\frac{\det(\mathbf{H_B})}{\det(\mathbf{H_B^Z})}\right)^{1/2} \int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}^{Z}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{Z}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))(\dot{\hat{\mathbf{x}}}_{\sim}^{Z}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{Z}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \, \mathrm{d}t \\
& \exp\{-\frac{1}{2}(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}^{Z}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}^{Z}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& -\frac{1}{2}\int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}^{Z}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{Z}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}_{\sim}^{Z}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}^{Z}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \mathrm{d}t \\
& +\frac{1}{2}(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k)) \\
& +\frac{1}{2}\int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \mathrm{d}t\}
\end{aligned} \tag{4.25}$$

$$\boldsymbol{\Sigma}_{k+1} = \frac{1}{n}\left(\frac{\det(\mathbf{H_B})}{\det(\mathbf{H_B^S})}\right)^{1/2} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^S_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^S_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}}$$
$$\exp\{-\frac{1}{2}(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^S_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}^S_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))$$
$$-\frac{1}{2}\int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}^S_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}^S_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}^S_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}^S_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))\,\mathrm{d}t$$
$$+\frac{1}{2}(\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))$$
$$+\frac{1}{2}\int_{t_0}^{t_q} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\hat{\mathbf{x}}}_{\sim}(t) - \mathbf{f}(\hat{\mathbf{x}}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))\,\mathrm{d}t\} \tag{4.26}$$

In Equations (4.25) and (4.26), the Hessians $\mathbf{H_B}$, $\mathbf{H_B^Z}$ and $\mathbf{H_B^S}$ are defined as:

$$\mathbf{H_B} = \left.\frac{\partial^2 J_{\mathrm{AMLE}}}{\partial \mathbf{B} \partial \mathbf{B}^T}\right|_{\mathbf{B}=\hat{\mathbf{B}}} \tag{4.27}$$

$$\mathbf{H_B^Z} = \left.\frac{\partial^2 J_{\sim}^Z}{\partial \mathbf{B} \partial \mathbf{B}^T}\right|_{\mathbf{B}=\hat{\mathbf{B}}^{\mathbf{Z}}} \tag{4.28}$$

$$\mathbf{H_B^S} = \left.\frac{\partial^2 J_{\sim}^S}{\partial \mathbf{B} \partial \mathbf{B}^T}\right|_{\mathbf{B}=\hat{\mathbf{B}}^{\mathbf{S}}} \tag{4.29}$$

$J_{\mathrm{AMLE}}$ in Equation (4.27) is Varziri's AMLE objective function defined in Equation (4.14). $J_{\sim}^Z$ and $J_{\sim}^S$ in Equations (4.28) and (4.29) are defined in Equations (4.30) and (4.31).

$$J_{\sim}^{Z} = -\ln \int_{t_0}^{t_q} (\dot{x}_{\sim 1}(t) - f_1(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^2 \, \mathrm{d}t - \ldots - \ln \int_{t_0}^{t_q} (\dot{x}_{\sim X}(t) - f_X(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^2 \, \mathrm{d}t$$
$$+ \frac{1}{2} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))$$
$$+ \frac{1}{2} \int_{t_0}^{t_q} (\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \, \mathrm{d}t$$

(4.30)

$$J_{\sim}^{S} = -\ln \sum_{j=1}^{N_1} [y_1(t_{m1,j}) - g_1(\mathbf{x}_{\sim}(t_{m1,j}), \mathbf{y}(t_{m1,j}), \boldsymbol{\theta}_k)]^2$$
$$- \ldots - \ln \sum_{j=1}^{N_Y} [y_Y(t_{mY,j}) - g_Y(\mathbf{x}_{\sim}(t_{mY,j}), \mathbf{y}(t_{mY,j}), \boldsymbol{\theta}_k)]^2$$
$$+ \frac{1}{2} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta}_k))$$
$$+ \frac{1}{2} \int_{t_0}^{t_q} (\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k))^{\mathrm{T}} \mathbf{Q}_k^{-1} (\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}_k)) \mathrm{d}t$$

(4.31)

and $\hat{\mathbf{B}}$, $\hat{\mathbf{B}}^S$ and $\hat{\mathbf{B}}^Z$ are spline coefficients that maximize $J_{\mathrm{AMLE}}$, $J_{\sim}^{Z}$ and $J_{\sim}^{S}$, respectively. $\hat{\mathbf{x}}_{\sim}$, $\hat{\mathbf{x}}_{\sim}^{S}$ and $\hat{\mathbf{x}}_{\sim}^{Z}$ are the corresponding approximated state trajectories. Equations 25 and 26 can then be used to update $\mathbf{Q}$ and $\boldsymbol{\Sigma}$, using the most recent estimates of the model parameters $\boldsymbol{\theta}$ and spline coefficients **B**.

For the case where the number of measurements for the *r*th response is $N_r$, an expression similar to Equation (4.25) for updating the *r*th measurement noise variance $\pm_r^2$ is:

$$\pm^{2}_{r,k+1} = \frac{1}{N_r}\left(\frac{\det(\mathbf{H}_{\boldsymbol{\beta}_r})}{\det(\mathbf{H}^{S}_{\boldsymbol{\beta}_r})}\right)^{1/2} \exp\{\ln \sum_{j=1}^{N_r}[y_r(t_{m\,r,j}) - g_r(\hat{\mathbf{x}}^{S}(t_{m\,r,j}), \mathbf{y}(t_{m\,1,j}), \boldsymbol{\theta}_k)]^2 -$$

$$\frac{1}{2\pm^{2}_{r,k}} \sum_{j=1}^{N_r}[y_r(t_{m\,r,j}) - g_r(\hat{\mathbf{x}}^{S}(t_{m\,r,j}), \mathbf{y}(t_{m\,1,j}), \boldsymbol{\theta}_k)]^2 - \frac{1}{2Q_k}\int_{t_0}^{t_q}(\dot{\hat{x}}^{S}_{r}(t) - f_r(\hat{\mathbf{x}}^{S}, \mathbf{u}(t), \boldsymbol{\theta}_k))^2 \,\mathrm{d}\,t$$

$$+\frac{1}{2\pm^{2}_{r,k}} \sum_{j=1}^{N_r}[y_r(t_{m\,r,j}) - g_r(\hat{\mathbf{x}}(t_{m\,r,j}), \mathbf{y}(t_{m\,1,j}), \boldsymbol{\theta}_k)]^2 + \frac{1}{2Q_k}\int_{t_0}^{t_q}(\dot{\hat{x}}_{r}(t) - f_r(\hat{\mathbf{x}}\ , \mathbf{u}(t), \boldsymbol{\theta}_k))^2 \,\mathrm{d}\,t\}$$

(4.A.37)

where

$$\mathbf{H}_{\boldsymbol{\beta}_r} = \mathbf{H}_{\mathbf{B}}(N_{r-1}+1 : N_r) \tag{4.A.38}$$

$$\mathbf{H}^{Z}_{\boldsymbol{\beta}_r} = \mathbf{H}^{Z}_{\mathbf{B}}(N_{r-1}+1 : N_r) \tag{4.A.39}$$

$$\mathbf{H}^{S}_{\boldsymbol{\beta}_r} = \mathbf{H}^{S}_{\mathbf{B}}(N_{r-1}+1 : N_r) \tag{4.A.40}$$

Note that other Equations do not change for this case.

## Appendix 5.A Derivation of the LAMLE Objective Function (Equation (5.23 ))

In this section, an approximate closed form for $p(\mathbf{Y_m} \mid \boldsymbol{\zeta})$ is derived using the LA and B-spline basis functions. The likelihood function $L(\boldsymbol{\zeta} \mid \mathbf{Y_m}) = p(\mathbf{Y_m} \mid \boldsymbol{\zeta})$ can be evaluated by integrating $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ over all possible values of $\mathbf{X_q}$ [16]:

$$p(\mathbf{Y_m} \mid \boldsymbol{\zeta}) = \int p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta}) d\mathbf{X_q} \tag{5.14}$$

A closed form for $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ was found in our previous work (see Equation (5.15)). Taking the negative natural logarithm of Equation (5.15) gives:

$$\begin{aligned} -\ln p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta}) = & -\ln C_1 + n \ln[\det(\boldsymbol{\Sigma})] + [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})] \\ & + \ln[\det(\mathbf{S}_{m0})] + (\mathbf{x}_{m0} - \mathbf{x}_0)^{\mathrm{T}} \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_0) \\ & + q \ln[\det(\mathbf{Q})] + \int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] \mathrm{d}t \end{aligned} \tag{5.A.1}$$

Let

$$J_1 = -\ln p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta}) + \ln C_1 \tag{5.A.2}$$

Substituting $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ from Equation (5.15) into Equation (5.14):

$$\begin{aligned} p(\mathbf{Y_m} \mid \boldsymbol{\zeta}) = & \int C_1 [\det(\boldsymbol{\Sigma})]^{-n} \exp\{-\frac{1}{2} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]\} \\ & \times [\det(\mathbf{S}_{m0})]^{-\frac{1}{2}} \exp[-\frac{1}{2} (\mathbf{x}_{m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_0)] \\ & \times [\det(\mathbf{Q})]^{-\frac{q}{2}} \exp\{-\int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^T \mathbf{Q}^{-1} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] \mathrm{d}t\} \mathrm{d}\mathbf{X_q} \end{aligned} \tag{5.A.3}$$

Using the LA:

$$\ln p(\mathbf{Y_m} \mid \boldsymbol{\zeta}) = \ln \int p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta}) d\mathbf{X_q} = \ln p(\hat{\mathbf{X}}_\mathbf{q}, \mathbf{Y_m} \mid \boldsymbol{\zeta}) - \frac{1}{2} \ln \det(\mathbf{H_x}) + \frac{q+1}{2} \ln(2®)$$

(5.A.4)

where $\hat{\mathbf{X}}_\mathbf{q}$ is evaluated from:

$$\hat{\mathbf{X}}_\mathbf{q} = \arg \max_{\mathbf{X_q}} \; p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta}) \tag{5.A.5}$$

and $\mathbf{H_x}$ is the Hessian matrix of $-\ln p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ with respect to $\mathbf{X_q}$ evaluated at $\hat{\mathbf{X}}_\mathbf{q}$:

$$\mathbf{H_x} = -\frac{\ln p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}}\bigg|_{\mathbf{X_q} = \hat{\mathbf{X}}_\mathbf{q}} = \frac{J_1}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}}\bigg|_{\mathbf{X_q} = \hat{\mathbf{X}}_\mathbf{q}} \tag{5.A.6}$$

Substituting $-\ln p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ from Equation (5.A.1) into Equation (5.A.2) and substituting $\mathbf{X_q}$ with $\hat{\mathbf{X}}_\mathbf{q}$:

$$\begin{aligned} -\ln p(\mathbf{Y_m} \mid \boldsymbol{\zeta}) = C_2 &+ [\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_\mathbf{m}, \mathbf{U_m}, \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\hat{\mathbf{X}}_\mathbf{m}, \mathbf{U_m}, \boldsymbol{\theta})] + n[\det(\boldsymbol{\Sigma})] \\ &+ (\mathbf{x}_{m0} - \hat{\mathbf{x}}_0)^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \hat{\mathbf{x}}_0) + q \ln[\det(\mathbf{Q})] \\ &+ \int_{t_0}^{t_q} [\dot{\hat{\mathbf{x}}}(t) - \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \boldsymbol{\theta})]^T \mathbf{Q}^{-1} [\dot{\hat{\mathbf{x}}}(t) - \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \boldsymbol{\theta})] \mathrm{d}t + \ln(\det(\mathbf{H_x})) \end{aligned}$$

(5.A.7)

where $C_2$ is a constant.

Substituting B-spline expressions for $\hat{\mathbf{X}}_\mathbf{m}$, $\dot{\hat{\mathbf{x}}}(t)$ and $\hat{\mathbf{x}}(t)$ gives:

$$\begin{aligned} -\ln p(\mathbf{Y_m} \mid \boldsymbol{\zeta}) = C_2 &+ [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})] + n[\ln(\det(\boldsymbol{\Sigma})] \\ &+ (\mathbf{x}_{\sim m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{\sim m0} - \mathbf{x}_0) + q \ln(\det(\mathbf{Q})) \\ &+ \int_{t_0}^{t_q} [\dot{\mathbf{x}}_\sim(t) - \mathbf{f}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta})]^T \mathbf{Q}^{-1} [\dot{\mathbf{x}}_\sim(t) - \mathbf{f}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta})] \mathrm{d}t + \ln(\det(\mathbf{H}_{\mathbf{x}_\sim})) \end{aligned}$$

(5.A.8)

where

$$\mathbf{H}_{\mathbf{x}_\sim} = \left.\frac{\partial J_1}{\partial \mathbf{X_q} \partial \mathbf{X_q}^{\mathrm{T}}}\right|_{\mathbf{X_q}=\mathbf{X}_{\sim\mathbf{q}}} \tag{5.A.9}$$

Minimizing

$$J_{\mathrm{LAMLE}} = -\ln p(\mathbf{Y_m} \mid \boldsymbol{\zeta}) - C_2 \tag{5.A.10}$$

with respect to _ and $\boldsymbol{\zeta}$ ensures that the resulting values of $\mathbf{X}_{\mathbf{m}\sim}$, $\dot{\mathbf{x}}_\sim(t)$ and $\mathbf{x}_\sim(t)$ correspond to the mode of $p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta})$ with respect to all possible values of the states, which is required for the LA to be valid. Minimizing

$$\begin{aligned} J_{\mathrm{LAMLE}} = {} & n \ln[\det(\boldsymbol{\Sigma})] + [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})] \\ & + (\mathbf{x}_{\sim m0} - \mathbf{x}_0)^T \mathbf{S}_{m0}^{-1} (\mathbf{x}_{\sim m0} - \mathbf{x}_0) + q \ln[\det(\mathbf{Q})] \\ & + \int_{t_0}^{t_q} [\dot{\mathbf{x}}_\sim(t) - \mathbf{f}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} [\dot{\mathbf{x}}_\sim(t) - \mathbf{f}(\mathbf{x}_\sim(t), \mathbf{u}(t), \boldsymbol{\theta})] \mathrm{d}t + \ln[\det(\mathbf{H}_{\mathbf{x}_\sim})] \end{aligned} \tag{5.16}$$

also ensures that the values of $\boldsymbol{\theta}, \mathbf{x}_0, \mathbf{Q}$ and $\boldsymbol{\Sigma}$ contained in $\boldsymbol{\zeta}$ are at approximate ML estimates. The problem with this approach is that an analytical expression for $\mathbf{H}_{\mathbf{x}_\sim}$ would be difficult to derive, especially when the number of states becomes large. An iterative method involving Equations (5.19) and (5.20) can be used to address this problem.

## Appendix 5.B Derivation of Equations (5.19) and (5.20)

Taking partial derivatives of $J_{LAMLE}$ in Equation (5.16) respect to **Q** and p gives:

$$\frac{\partial J_{LAMLE}}{\partial \mathbf{Q}} = q\mathbf{Q}^{-1} - \int_{t_0}^{t_q} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} [\dot{\mathbf{x}}_{\sim}(t) - \mathbf{f}(\mathbf{x}_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta})] \, \mathrm{d}\, t \; \mathbf{Q}^{-1} + tr(\mathbf{H}_{\mathbf{x}\sim}^{-1} \frac{\partial \mathbf{H}_{\mathbf{x}\sim}}{\partial \mathbf{Q}})$$

(5.B.1)

$$\frac{\partial J_{LAMLE}}{\partial \boldsymbol{\Sigma}} = n\boldsymbol{\Sigma}^{-1} + [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim \mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})] \boldsymbol{\Sigma}^{-1} + tr(\mathbf{H}_{\mathbf{x}\sim}^{-1} \frac{\partial \mathbf{H}_{\mathbf{x}\sim}}{\partial \boldsymbol{\Sigma}})$$

(5.B.2)

Note that development of Equations (5.B.1) and (5.B.2) relies on the following expressions for the derivative of the determinant of a matrix:[58]

$$\frac{\partial \det(\mathbf{Q})}{\partial \mathbf{Q}} = \det(\mathbf{Q}) \, \mathbf{Q}^{-1}$$

(5.B.3)

$$\frac{\partial \ln[\det(\mathbf{H}_{\mathbf{x}\sim})]}{\partial \mathbf{Q}} = tr(\mathbf{H}_{\mathbf{x}\sim}^{-1} \frac{\partial \mathbf{H}_{\mathbf{x}\sim}}{\partial \mathbf{Q}})$$

(5.B.4)

The integral term in $J_{LAMLE}$ (Equation (5.16)) can be approximated as:

$$\int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] \, \mathrm{d}\, t \approx \sum_{i=0}^{q} [\mathbf{x}(t_{i+1}) - \mathbf{x}(t_i) - \mathbf{f}(\mathbf{x}(t_i), \mathbf{u}(t_i), \boldsymbol{\theta}) \Delta t]^{\mathrm{T}} \mathbf{Q}^{-1} \Delta t^{-1} [\mathbf{x}(t_{i+1}) - \mathbf{x}(t_i) - \mathbf{f}(\mathbf{x}(t_i), \mathbf{u}(t_i), \boldsymbol{\theta}) \Delta t]$$

(5.B.5)

Substituting Equation (5.B.5) into Equation (5.A.1), the likelihood function of the complete data for a discrete SDE system is:

$$
\begin{aligned}
J_d &= -\ln\, p(\mathbf{X_q}, \mathbf{Y_m} \mid \boldsymbol{\zeta}) = \ln C_3 + n \ln[\det(\boldsymbol{\Sigma})] \\
&\quad + [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})] \\
&\quad + \ln[\det(\mathbf{S}_{m0})] + \ln[(\mathbf{x}_{m0} - \mathbf{x}(t_0))^{\mathrm{T}} \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_0)] + q \ln[\det(\mathbf{Q})] \\
&\quad + \sum_{i=0}^{q} [\mathbf{x}(t_{i+1}) - \mathbf{x}(t_i) - \mathbf{f}(\mathbf{x}(t_i), \mathbf{u}(t_i), \boldsymbol{\theta})\Delta t]^{\mathrm{T}} \mathbf{Q}^{-1} \Delta t^{-1} [\mathbf{x}(t_{i+1}) - \mathbf{x}(t_i) - \mathbf{f}(\mathbf{x}(t_i), \mathbf{u}(t_i), \boldsymbol{\theta})\Delta t]
\end{aligned}
\tag{5.B.6}
$$

For convenience, define **z** and **h** as:

$$
\mathbf{w} = \mathbf{x}(t_{i+1}) - \mathbf{x}(t_i) - \mathbf{f}(\mathbf{x}(t_i), \mathbf{u}(t_i), \boldsymbol{\theta})\Delta t \tag{5.B.7}
$$

$$
\mathbf{h} = \mathbf{w}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{w} \tag{5.B.8}
$$

The second derivation of the symmetric quadratic matrix form **h** with respect to $\mathbf{x}(t_i)$ is:[58]

$$
\frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}(t_i) \partial \mathbf{x}^{\mathrm{T}}(t_i)} = (2\mathbf{w}^{\mathrm{T}} \mathbf{Q}^{-1} \otimes \mathbf{I}) \frac{\partial^2 \mathbf{w}}{\partial \mathbf{x}(t_i) \partial \mathbf{x}^{\mathrm{T}}(t_i)} + 2\left(\frac{\partial \mathbf{w}}{\partial \mathbf{x}(t_i)}\right)^{\mathbf{T}} \mathbf{Q}^{-1} \left(\frac{\partial \mathbf{w}}{\partial \mathbf{x}(t_i)}\right) \tag{5.B.9}
$$

where $\otimes$ is the Kronecker product.

The second derivative of $J_d$ (Equation (5.B.6)) with respect to $\mathbf{x_m}(t_0)$ is:

$$
\begin{aligned}
\left.\frac{\partial^2 J_d}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0) \partial \mathbf{x}_{\mathbf{m}}^{\mathrm{T}}(t_0)}\right|_{\mathbf{x_m}0 = \mathbf{x_m}0\sim} &= 2[(\mathbf{x}_{\sim}(t_1) - \mathbf{x}_{\mathbf{m}\sim}(t_0) - \mathbf{f}(\mathbf{x}_{\mathbf{m}\sim}(t_0), \mathbf{u}(t_0), \boldsymbol{\theta})\Delta t)^{\mathrm{T}} \mathbf{Q}^{-1} \otimes \mathbf{I}]\Delta t^{-1} \\
&\quad \times \left(-\frac{\partial^2 \mathbf{f}(\mathbf{x}_{\mathbf{m}\sim}(t_0), \mathbf{u}(t_0), \boldsymbol{\theta})\Delta t)}{\partial \mathbf{x}_{\mathbf{m}0} \partial \mathbf{x}_{\mathbf{m}}^{\mathrm{T}}(t_0)}\right) + 2\left(-\mathbf{I} - \frac{\partial \mathbf{f}(\mathbf{x}_{\mathbf{m}\sim}(t_0), \mathbf{u}(t_0), \boldsymbol{\theta})\Delta t)}{\partial \mathbf{x}_{\mathbf{m}}(t_0)}\right)^{\mathrm{T}} \\
&\quad \times \mathbf{Q}^{-1} \Delta t^{-1} \left(-\mathbf{I} - \frac{\partial \mathbf{f}(\mathbf{x}_{\mathbf{m}\sim}(t_0), \mathbf{u}(t_0), \boldsymbol{\theta})\Delta t)}{\partial \mathbf{x}_{\mathbf{m}}(t_0)}\right)
\end{aligned}
\tag{5.B.10}
$$

The second derivative of $J_d$ with respect to $\mathbf{x}(t_i)$ where $1 \le i \le q$-1is:

$$\left.\frac{\partial^2 J_d}{\partial \mathbf{x}(t_i)\partial \mathbf{x}^{\mathrm{T}}(t_i)}\right|_{\mathbf{x}=\mathbf{x}_\sim} = \mathbf{Q}^{-1}\Delta t^{-1} + 2([\mathbf{x}(t_{i+1}) - \mathbf{x}_\sim(t_i) - \mathbf{f}(\mathbf{x}_\sim(t_i), u(t_i), \boldsymbol{\theta})\Delta t]^{\mathrm{T}}\mathbf{Q}^{-1} \otimes \mathbf{I})\Delta t^{-1}$$
$$\times(-\frac{\partial^2 \mathbf{f}(\mathbf{x}_\sim(t_i), \mathbf{u}(t_i), \boldsymbol{\theta})\Delta t)}{\partial \mathbf{x}_\sim(t_i)\partial \mathbf{x}_\sim^{\mathrm{T}}(t_i)}) + 2(-\mathbf{I} - \frac{\partial \mathbf{f}(\mathbf{x}_\sim(t_i), \mathbf{u}(t_i), \boldsymbol{\theta})\Delta t)}{\partial \mathbf{x}_\sim(t_i)})^{\mathrm{T}}$$
$$\times\mathbf{Q}^{-1}\Delta t^{-1}(-\mathbf{I} - \frac{\partial \mathbf{f}(\mathbf{x}_\sim(t_i), \mathbf{u}(t_i), \boldsymbol{\theta})\Delta t)}{\partial \mathbf{x}_\sim(t_i)}) \tag{5.B.11}$$

Note that for times when no measurements are available, the first term on the right-hand side of Equation (5.B.11) disappears.

Taking the second derivative of $J_d$ with respect to $\mathbf{x}(t_q)$ gives:

$$\left.\frac{\partial^2 J_d}{\partial \mathbf{x}(t_q)\partial \mathbf{x}^{\mathrm{T}}(t_q)}\right|_{\mathbf{x}=\mathbf{x}_\sim} = \boldsymbol{\Sigma}^{-1} + \mathbf{Q}^{-1}\Delta t^{-1} \tag{5.B.12}$$

The second derivative of $J_d$ for $i \neq j$ and $j \neq j+1$ where $1 \leq i \leq q-1$ is:

$$\left.\frac{\partial^2 J_d}{\partial \mathbf{x}(t_j)\partial \mathbf{x}^{\mathrm{T}}(t_i)}\right|_{\mathbf{x}=\mathbf{x}_\sim} = \mathbf{0} \tag{5.B.13}$$

where $\mathbf{0}$ is the X×X zero matrix.

The second derivative of $J_d$ with respect to $\mathbf{x}(t_i)$ and $\mathbf{x}(t_{i+1})$ where $1 \leq i \leq q-1$ is:

$$\left.\frac{\partial^2 J_d}{\partial \mathbf{x}(t_{i+1})\partial \mathbf{x}^{\mathrm{T}}(t_i)}\right|_{\mathbf{x}=\mathbf{x}_\sim} = 2[-\mathbf{I} - \frac{\partial \mathbf{f}(\mathbf{x}(t_i), \mathbf{u}(t_i), \boldsymbol{\theta})}{\partial \mathbf{x}(t_i)}\Delta t]^{\mathrm{T}}\mathbf{Q}^{-1}\Delta t^{-1} \tag{5.B.14}$$

Equations (5.B.10), (5.B.11), (5.B.12), (5.B.13) and (5.B.14) show that the hessian matrix has the following form:

$$\mathbf{H}_{\mathbf{x}_\sim} = \boldsymbol{\Sigma}^{-1} + \mathbf{Q}^{-1}\mathbf{D} \tag{5.B.15}$$

where $\mathbf{I_H}$ is an identity matrix, $\mathbf{D}$ is a function of $\mathbf{f}$ and its derivatives and $\mathbf{x}_{\sim}(t_0),...,\mathbf{x}_{\sim}(t_q)$, $\mathbf{u}(t_0),...,\mathbf{u}(t_q),\boldsymbol{\theta}$ and $\Delta t$:

$$\mathbf{D} = \mathbf{D}(\mathbf{f}, \mathbf{x}_{\sim}(t_0),..., \mathbf{x}_{\sim}(t_q), \mathbf{u}(t_0),..., \mathbf{u}(t_q), \boldsymbol{\theta}, \Delta t) \tag{5.B.16}$$

Taking the derivative of Equation (5.B.15) with respect to $\mathbf{Q}$ gives:

$$\frac{\partial \mathbf{H}_{\mathbf{x}_{\sim}}}{\partial \mathbf{Q}} = -\mathbf{D}\mathbf{Q}^{-2} \tag{5.B.17}$$

Rearranging Equation (5.B.15) to solve for $\mathbf{D}$ gives:

$$\mathbf{D} = [\mathbf{H}_{\mathbf{x}_{\sim}} - \boldsymbol{\Sigma}^{-1}]\mathbf{Q} \tag{5.B.18}$$

Substituting Equation (5.B.18) into Equation (5.B.17) gives:

$$\frac{\partial \mathbf{H}_{\mathbf{x}_{\sim}}}{\partial \mathbf{Q}} = [-\mathbf{H}_{\mathbf{x}_{\sim}} + \boldsymbol{\Sigma}^{-1}]\mathbf{Q}^{-1} \tag{5.B.19}$$

Substituting Equation (5.B.19) into Equation (5.B.1) gives:

$$\begin{aligned}\frac{\partial J_{LAMLE}}{\partial \mathbf{Q}} = q\mathbf{Q}^{-1} - \int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-2} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt \\ + tr(\mathbf{H}_{\mathbf{x}_{\sim}}^{-1} [-\mathbf{H}_{\mathbf{x}_{\sim}} + \boldsymbol{\Sigma}^{-1}]\mathbf{Q}^{-1})\end{aligned} \tag{5.B.20}$$

Simplifying Equation (5.B.20) gives:

$$\begin{aligned}\frac{\partial J_{LAMLE}}{\partial \mathbf{Q}} = q\mathbf{Q}^{-1} - \int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-2} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt \\ - q\mathbf{Q}^{-1} + tr(\mathbf{H}_{\mathbf{x}_{\sim}}^{-1}) \boldsymbol{\Sigma}^{-1} \mathbf{Q}^{-1}\end{aligned} \tag{5.B.21}$$

Setting Equation (5.B.21) equal to zero and solving for $\mathbf{Q}$ gives:

$$\mathbf{Q} = (\,\mathrm{tr}(\mathbf{H}_{\mathbf{x}_{\sim}}^{-1})\boldsymbol{\Sigma}^{-1})^{-1} \int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{I} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt \tag{5.B.22}$$

Taking the derivative of Equation (5.B.15) with respect to $\boldsymbol{\Sigma}$:

$$\frac{\partial \mathbf{H}_{\mathbf{x}_\sim}}{\partial \boldsymbol{\Sigma}} = -\boldsymbol{\Sigma}^{-2} \tag{5.B.23}$$

Substituting Equation (5.B.23) into Equation (5.B.2) and simplifying gives:

$$\frac{\partial J_{LAMLE}}{\partial \boldsymbol{\Sigma}} = n\boldsymbol{\Sigma}^{-1} - [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})] \boldsymbol{\Sigma}^{-1} - tr(\mathbf{H}_{\mathbf{x}\sim}^{-1}) \boldsymbol{\Sigma}^{-2} \tag{5.B.24}$$

Setting Equation (5.B.24) equal to zero and solving for $\boldsymbol{\Sigma}$ gives:

$$\boldsymbol{\Sigma} = n^{-1} \left\{ [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})] [\mathbf{Y_m} - \mathbf{g}(\mathbf{X}_{\sim\mathbf{m}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} + tr(\mathbf{H}_{\mathbf{x}_\sim}^{-1}) \mathbf{I} \right\} \tag{5.B.25}$$

The Hessian matrix in Equations (5.16), (5.B.22) and (5.B.25) are with respect to state variables. However, IPOPT (and some other types of optimization software) reports the Hessian matrix with respect to the decision variables, which include the B-spline coefficients. The relationship between $\mathbf{H}_{\mathbf{x}\sim}$ (i.e., the Hessian matrix with respect to $\mathbf{X}_{\mathbf{q}\sim}$) and $\mathbf{H_B}$ (i.e., the Hessian matrix with respect to **B**) is derived below.

Using the chain rule for partial derivatives to find the partial derivative of $J_1$ with respect to matrix of B-spline basis functions **B** gives:

$$\frac{\partial J_1}{\partial \mathbf{B}} = \frac{\partial J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)} \frac{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)}{\partial \mathbf{B}} + \frac{\partial J_1}{\partial \mathbf{x}_\sim(t_1)} \frac{\partial \mathbf{x}_\sim(t_1)}{\partial \mathbf{B}} + \cdots + \frac{\partial J_1}{\partial \mathbf{x}_\sim(t_q)} \frac{\partial \mathbf{x}_\sim(t_q)}{\partial \mathbf{B}} \tag{5.B.26}$$

Finding partial derivatives from Equation (5.6):

$$\frac{\partial \mathbf{x}_\sim(t_i)}{\partial \mathbf{B}} = \boldsymbol{\Phi}(t_i) \tag{5.B.27}$$

Substituting partial derivatives from Equation (5.B.27) into Equation (5.B.26) gives:

$$\frac{\partial J_1}{\partial \mathbf{B}} = \frac{\partial J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)} \boldsymbol{\Phi}(t_0) + \frac{\partial J_1}{\partial \mathbf{x}_\sim(t_1)} \boldsymbol{\Phi}(t_1) + \cdots + \frac{\partial J_1}{\partial \mathbf{x}_\sim(t_q)} \boldsymbol{\Phi}(t_q) \tag{5.B.28}$$

Let

$$\mathbf{G} = \frac{\partial J_1}{\partial \mathbf{B}} \tag{5.B.29}$$

Using the chain rule for finding the partial derivatives of **G**:

$$\frac{\partial \mathbf{G}}{\partial \mathbf{B}} = \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)} \frac{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)}{\partial \mathbf{B}} + \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_1)} \frac{\partial \mathbf{x}_{\sim}(t_1)}{\partial \mathbf{B}} + \cdots + \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_q)} \frac{\partial \mathbf{x}_{\sim}(t_q)}{\partial \mathbf{B}} \tag{5.B.30}$$

Substituting partial derivatives from Equation (5.B.27) into Equation (5.B.30):

$$\frac{\partial \mathbf{G}}{\partial \mathbf{B}} = \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)} \mathbf{\Phi}(t_0) + \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_1)} \mathbf{\Phi}(t_1) + \cdots + \frac{\partial \mathbf{G}}{\partial \mathbf{x}_{\sim}(t_q)} \mathbf{\Phi}(t_q) \tag{5.B.31}$$

Substituting **G** from Equations (5.B.28) and (5.B.29) into Equation (5.B.31) gives the second derivative of $J_1$ with respect to the spline coefficients:

$$\begin{aligned}\frac{\partial \mathbf{G}}{\partial \mathbf{B}} = \frac{\partial^2 J_1}{\partial \mathbf{B} \partial \mathbf{B}^{\mathrm{T}}} = {} & \frac{\partial}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)} [\frac{\partial J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)} \mathbf{\Phi}(t_0) + \frac{\partial J_1}{\partial \mathbf{x}_{\sim}(t_1)} \mathbf{\Phi}(t_1) + \cdots + \frac{\partial J_1}{\partial \mathbf{x}_{\sim}(t_q)} \mathbf{\Phi}(t_q)] \mathbf{\Phi}(t_0) + \\ & \frac{\partial}{\partial \mathbf{x}_{\sim}(t_1)} [\frac{\partial J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)} \mathbf{\Phi}(t_0) + \frac{\partial J_1}{\partial \mathbf{x}_{\sim}(t_1)} \mathbf{\Phi}(t_1) + \cdots + \frac{\partial J_1}{\partial \mathbf{x}_{\sim}(t_q)} \mathbf{\Phi}(t_q)] \mathbf{\Phi}(t_1) + \cdots + \\ & \frac{\partial}{\partial \mathbf{x}_{\sim}(t_q)} [\frac{\partial J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0)} \mathbf{\Phi}(t_0) + \frac{\partial J_1}{\partial \mathbf{x}_{\sim}(t_1)} \mathbf{\Phi}(t_1) + \cdots + \frac{\partial J_1}{\partial \mathbf{x}_{\sim}(t_q)} \mathbf{\Phi}(t_q)] \mathbf{\Phi}(t_q)\end{aligned} \tag{5.B.32}$$

Simplifying Equation (5.B.32) gives:

$$\begin{aligned}\frac{\partial^2 J_1}{\partial \mathbf{B} \partial \mathbf{B}^{\mathrm{T}}} = {} & \mathbf{\Phi}(t_0) \frac{\partial^2 J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0) \partial \mathbf{x}_{\mathbf{m}\sim}^{\mathrm{T}}(t_0)} \mathbf{\Phi}^{\mathrm{T}}(t_0) + \mathbf{\Phi}(t_1) \frac{\partial^2 J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_1)} \mathbf{\Phi}^{\mathrm{T}}(t_0) + \cdots \\ & + \mathbf{\Phi}(t_q) \frac{\partial^2 J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_q)} \mathbf{\Phi}(t_0) + \mathbf{\Phi}(t_0) \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\mathbf{m}\sim}^{\mathrm{T}}(t_0)} \mathbf{\Phi}(t_1) + \mathbf{\Phi}(t_1) \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_1)} \mathbf{\Phi}^{\mathrm{T}}(t_1) \\ & + \cdots + \mathbf{\Phi}(t_q) \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_q)} \mathbf{\Phi}^{\mathrm{T}}(t_1) + \cdots + \mathbf{\Phi}(t_0) \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\mathbf{m}\sim}^{\mathrm{T}}(t_0)} \mathbf{\Phi}^{\mathrm{T}}(t_q) \\ & + \mathbf{\Phi}(t_1) \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_1)} \mathbf{\Phi}^{\mathrm{T}}(t_q) + \cdots + \mathbf{\Phi}(t_q) \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_q)} \mathbf{\Phi}^{\mathrm{T}}(t_q)\end{aligned} \tag{5.B.33}$$

In matrix form, Equation (5.B.31) becomes:

$$\begin{bmatrix} \frac{\partial^2 J_1}{\partial \boldsymbol{\beta}_1^2} & \frac{\partial^2 J_1}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2} & \cdots & \frac{\partial^2 J_1}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_{c_s}} \\ \frac{\partial^2 J_1}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1} & \frac{\partial^2 J_1}{\partial \boldsymbol{\beta}_2^2} & \cdots & \frac{\partial^2 J_1}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_{c_s}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J_1}{\partial \boldsymbol{\beta}_{c_s} \partial \boldsymbol{\beta}_1} & \frac{\partial^2 J_1}{\partial \boldsymbol{\beta}_{c_s} \partial \boldsymbol{\beta}_2} & \cdots & \frac{\partial^2 J_1}{\partial \boldsymbol{\beta}_{c_s}^2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}(t_1) & \boldsymbol{\Phi}(t_2) & \ldots & \boldsymbol{\Phi}(t_q) \end{bmatrix} \times$$

$$\begin{bmatrix} \frac{\partial^2 J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0) \partial \mathbf{x}_{\mathbf{m}\sim}^{\mathrm{T}}(t_0)} & \frac{\partial^2 J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_1)} & \cdots & \frac{\partial^2 J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_q)} \\ \frac{\partial^2 J_1}{\partial \mathbf{x}_{\mathbf{m}\sim}(t_0) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_1)} & \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_1)} & \cdots & \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_1) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_q)} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\mathbf{m}\sim}^{\mathrm{T}}(t_0)} & \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_1)} & \cdots & \frac{\partial^2 J_1}{\partial \mathbf{x}_{\sim}(t_q) \partial \mathbf{x}_{\sim}^{\mathrm{T}}(t_q)} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\Phi}(t_1) \\ \boldsymbol{\Phi}(t_2) \\ \vdots \\ \boldsymbol{\Phi}(t_q) \end{bmatrix}$$

(5.B.34)

Or equivalently:

$$\mathbf{H}_{\mathbf{B}} = \boldsymbol{\Psi}^{\mathrm{T}} \mathbf{H}_{\mathbf{X}\sim} \boldsymbol{\Psi} \tag{5.B.35}$$

which can be solved to give $\mathbf{H}_{\mathbf{X}\sim}^{-1}$ :

$$\mathbf{H}_{\mathbf{X}\sim}^{-1} = \boldsymbol{\Psi} \mathbf{H}_{\mathbf{B}}^{-1} \boldsymbol{\Psi}^{\mathrm{T}} \tag{5.B.36}$$

During the parameter estimation process, the spline coefficients $\boldsymbol{\beta}_s$ will change, but the spline function $\boldsymbol{\varphi}_s(t)$ are fixed functions.

Substituting $\mathbf{H}_{\mathbf{X}\sim}^{-1}$ from Equation (5.B.36) into Equations (5.B.22) and (5.B.25) gives:

$$\mathbf{Q} = (\ \mathrm{tr}(\boldsymbol{\Psi} \mathbf{H}_{\mathbf{B}}^{-1} \boldsymbol{\Psi}^{\mathrm{T}}) \boldsymbol{\Sigma}^{-1})^{-1} \int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{I} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] dt \tag{5.19}$$

$$\boldsymbol{\Sigma} = n^{-1}\left\{[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}}\mathbf{I}[\mathbf{Y_m} - \mathbf{g}(\mathbf{X_{\sim m}}, \mathbf{U_m}, \boldsymbol{\theta})] + tr(\boldsymbol{\Psi}\mathbf{H_B^{-1}}\boldsymbol{\Psi}^{\mathrm{T}})\mathbf{I}\right\} \quad (5.20)$$

## Appendix 5.C Derivation of LAMLE Objective Function when there are Different Numbers of Measurements for Different Responses

In this section, LAMLE objective function (Equation (5.16)) and Equations (5.19) and (5.20) are modified for the case where the number of measurements for the *r*th response is $N_r$.

LAMLE objective function for the case where $N_r$ measurements are available for the *r*th response is:

$$\begin{aligned} J_{LAMLE} = {} & N_1 \ln(\pm_1^2) + \ldots + N_Y \ln(\pm_Y^2) + [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} [\mathbf{Y_m} - \mathbf{g}(\mathbf{X_m}, \mathbf{U_m}, \boldsymbol{\theta})] \\ & + \ln[\det(\mathbf{S}_{m0})] + \ln[(\mathbf{x}_{m0} - \mathbf{x}_0)^{\mathrm{T}} \mathbf{S}_{m0}^{-1} (\mathbf{x}_{m0} - \mathbf{x}_0)] \\ & + q \ln[\det(\mathbf{Q})] + \int_{t_0}^{t_q} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})]^{\mathrm{T}} \mathbf{Q}^{-1} [\dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})] \mathrm{d}t \end{aligned} \quad (5.C.1)$$

The equation for obtaining **Q** (Equation (5.19)) does not change in this case. An expression for obtaining the *r*th measurement noise variance $\pm_r^2$ is developed in the following.

Taking partial derivatives of $J_{LAMLE}$ (Equation (5.C.1)) with respect to $\pm_r^2$ gives:

$$\frac{\partial J_{LAMLE}}{\partial \pm_r^2} = \frac{N_1}{\pm_r^2} - \frac{\sum_{i=1}^{N_1} [y_1(t_{m\,1,\,j}) - g_1(\mathbf{x}_{\sim}(t_{m\,1,\,j}), \mathbf{y}(t_{m\,1,\,j}), \boldsymbol{\theta})]^2}{(\pm_r^2)^2} + tr\left(\mathbf{H}_{\mathbf{x}\sim}^{-1} \frac{\partial \mathbf{H}_{\mathbf{x}\sim}}{\partial \pm_r^2}\right) \quad (5.C.2)$$

Taking the derivative of Equation (5.C.2) with respect to $\pm_r^2$ gives:

$$\frac{\partial \mathbf{H}_{\mathbf{x}\sim}}{\partial \pm_r^2} = -\frac{\mathbf{I}}{(\pm_r^2)^2} \tag{5.C.3}$$

Substituting Equation (5.C.3) into Equation (5.C.2) gives:

$$\frac{\partial J_{LAMLE}}{\partial \pm_r^2} = \frac{N_r}{\pm_r^2} - \frac{\sum_{j=1}^{N_r}[y_r(t_{m\,r,j}) - g_r(\mathbf{x}_\sim(t_{m\,r,j}), \mathbf{y}(t_{m\,r,j}), \boldsymbol{\theta})]^2}{(\pm_r^2)^2} - \frac{1}{(\pm_r^2)^2} tr(\mathbf{H}_{\mathbf{x}\sim}^{-1}) \tag{5.C.4}$$

Setting Equation (5.C.4) equal to zero and solving for $\pm_r^2$ gives:

$$\pm_r^2 = N_r \left(\sum_{j=1}^{N_r}[y_r(t_{m\,r,j}) - g_r(\mathbf{x}_\sim(t_{m\,r,j}), \mathbf{y}(t_{m\,1,j}), \boldsymbol{\theta})]^2 + tr(\mathbf{H}_{\mathbf{x}\sim}^{-1})\right)^{-1} \tag{5.C.5}$$

Finally, substituting $\mathbf{H}_{\mathbf{x}\sim}^{-1}$ from Equation (5.B.36) into Equation (5.C.5) gives an expression for estimating the measurement noise variance for the *r*th response:

$$\pm_r^2 = N_r \left(\sum_{j=1}^{N_r}[y_r(t_{m\,r,j}) - g_r(\mathbf{x}_\sim(t_{m\,r,j}), \mathbf{y}(t_{m\,1,j}), \boldsymbol{\theta})]^2 + tr(\boldsymbol{\Psi}\mathbf{H}_{\mathbf{B}}^{-1}\boldsymbol{\Psi}^{\mathrm{T}})\right)^{-1} \tag{5.C.6}$$