OBTAINING THE BEST MODEL PREDICTIONS AND PARAMETER ESTIMATES USING LIMITED DATA

by

Kevin McLean

A thesis submitted to the Department of Chemical Engineering

In conformity with the requirements for the degree of

Master of Applied Science

Queen's University

Kingston, Ontario, Canada

(September, 2011)

Abstract

Engineers who develop fundamental models for chemical processes are often unable to estimate all of the model parameters due to problems with parameter identifiability and estimability. The literature concerning these two concepts is reviewed and techniques for assessing parameter identifiability and estimability in nonlinear dynamic models are summarized. Modellers often face estimability problems when the available data are limited or noisy. In this situation, modellers must decide whether to conduct new experiments, change the model structure, or to estimate only a subset of the parameters and leave others at fixed values. Estimating only a subset of important model parameters is a technique often used by modellers who face estimability problems and it may lead to better model predictions with lower mean squared error (MSE) than the full model with all parameters estimated. Different methods in the literature for parameter subset selection are discussed and compared.

An orthogonalization algorithm combined with a recent MSE-based criterion has been used successfully to rank parameters from most to least estimable and to determine the parameter subset that should be estimated to obtain the best predictions. In this work, this strategy is applied to a batch reactor model using additional data and results are compared with computationally-expensive leave-one-out cross-validation. A new simultaneous ranking and selection technique based on this MSE criterion is also described. Unfortunately, results from these parameter selection techniques are sensitive to the initial parameter values and the uncertainty factors used to calculate sensitivity coefficients. A robustness test is proposed and applied to assess the sensitivity of the selected parameter subset to the initial parameter guesses. The selected parameter subsets are compared with those selected using another MSE-based method proposed

by Chu et al. (2009). The computational efforts of these methods are compared and recommendations are provided to modellers.

Co-Authorship

The literature review, Chapter 2, has been accepted for publication in the *Canadian Journal of Chemical Engineering*. A journal article based on the results presented in Chapter 3 is in preparation and will be submitted to *Industrial & Engineering Chemistry Research*. I prepared the first and subsequent drafts for all of these coauthored manuscripts, performed all of the simulations, and produced all of the Figures and Tables. Dr. Kim McAuley is a co-author of both journal articles. She helped to formulate research objectives, provide technical advice, and edit the articles and suggest revisions. Dr. Shaohua Wu is a co-author of the second journal article. He developed the initial ideas for ranking parameters using the MSE-based r_{CC} criterion, which were outlined in an unpublished report.

Acknowledgements

There are many people I would like to acknowledge who played an important role in the development of this thesis. First and foremost, I would like to thank my supervisor Dr. Kim McAuley, for her support and guidance during my graduate studies. I am fortunate to have such a caring supervisor who made herself available at key junctures.

Thank you to Shell Global Solutions and MITACS for their financial support of my research.

I would like to thank Eugene Theobald and Jesse Black at Shell Global Solutions for the opportunity to develop my skills working on an industrial problem, and I am most grateful for their helpful support and guidance over the course of the project. My time at Shell proved to be a valuable learning experience and the knowledge and abilities gained during this time will be very useful in the years to come.

Thanks are also necessary for the useful and timely discussions I had with Shaohua Wu, who previously worked on parameter subset selection techniques. Many thanks to my lab-mates (Jon, Angelica, John, Wei, Hadis, Zahra, Yutian, Hui) for their camaraderie.

I would also like to mention my appreciation for the chemical engineering administration staff for helpful guidance on admin/travel issues.

Last, but not least, I would like to thank my family for their continuous support, encouragement, and love. Without this, this thesis would not have been possible.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Chemical engineers constantly face new challenges in the development of new processes or the improvement of existing ones. In both situations, mathematical models are valuable tools. A fundamental model is not only important for a better understanding of the underlying phenomena, but also as a tool for simulating the behaviour under different operating conditions, designing and optimizing process improvements, training operators, and developing and testing control systems (Foss et al., 1998; Maria, 2004).

Unfortunately, the development of a useful fundamental model can be a difficult challenge in itself. Processes of interest to chemical engineers often involve many reactions, with many associated kinetic parameters. The number of model parameters further increases when mass-transfer terms and thermodynamic parameters are required to describe the process. Values of model parameters are often unknown or have a high associated uncertainty. In either case, modellers use nonlinear optimization algorithms to estimate the unknown or uncertain parameter values (Maria, 2004).

In practical situations, it is often impossible for the modeller to estimate all of the model parameters. Modellers often have to work with limited and/or noisy data (Vajda et al., 1989; Brun et al., 2001; Raue et al., 2009), and obtaining additional data from new experiments may not be feasible. Some parameters may have little influence on model predictions, making it impossible to obtain accurate estimates for these parameters (see e.g., Reichert et al., 1995; Brun et al., 2002). The effects of some parameters are frequently correlated with the effects of others (see

e.g., Holmberg, 1982; Peterson et al., 2001), and several sets of parameter values may give identical or nearly identical model predictions. These problems are categorized as identifiability or estimability issues. Identifiability issues arise from the structure of the model and the type of measurements that are available (see e.g., Bellman and Åström, 1970; Vajda et al., 1989; Walter and Pronzato, 1996). Estimability problems result when it is not possible to estimate some model parameters with the available data or new data from a proposed set of experiments (see e.g., Holmberg, 1982; Petersen et al., 2001; Jimenez-Hornero et al., 2009). Many modellers of chemical processes face problems with inestimable parameters. In this situation, the modeller must decide whether to alter the model structure, collect new data to improve the estimates of model parameters, or estimate only a subset of the model parameters. With limited data available, fixing some parameters at initial values and estimating only a subset of parameters can give better predictions, in terms of mean squared error (MSE), than estimating all of the model parameters (Hocking, 1976; Wu et al., 2007).

Many techniques have been proposed to aid modellers in deciding which parameters to estimate and which to leave at fixed nominal values. These methods are summarized in Chapter 2. An orthogonalization algorithm (Yao et al., 2003; Lund and Foss, 2008; Thompson et al., 2009) has been used to rank parameters in several chemical and biochemical processes (see e.g., Gadkar et al., 2005; Kou et al., 2005a, b; Yue et al., 2006; Jayasankar et al., 2009; Koeva et al., 2009; Thompson et al., 2009; Littlejohns et al., 2010; Srinath and Gunawan, 2010; Woloszyn and McAuley, 2011) to determine which parameters are important for estimation. A recent technique was developed by Wu et al. (2007; 2011a, b) to determine the number of parameters to select from a ranked list based on a critical MSE-based ratio. Wu et al. (2011b) demonstrated this technique using the Dow batch reactor model (Biegler et al., 1986) with only a portion of the

available data. Wu's attempts to perform parameter subset selection using all of the Dow data failed due to numerical difficulties. The results of Wu's study using the limited data set agreed with brute-force leave-one-out cross-validation and Wu's strategy required much less computation time. However, Wu's technique depends on the values of the initial parameter guesses and scaling factors used in generating parametric sensitivity coefficients. As a result, modellers who are uncertain about their initial assumptions and their influence on the sensitivity coefficients may question the validity of the selected parameter set using this method.

In this thesis, Chapter 2 summarizes literature definitions for identifiability and estimability, and describes techniques for assessing the identifiability and estimability of parameters in dynamic models. Methods for selecting subsets of estimable parameters are also reviewed and recommendations are made concerning which techniques are effective and practical in different situations. In Chapter 3, Wu's MSE-based criterion is applied to the Dow batch reactor model and the example is expanded to include all of the available data. Leave-one-out cross-validation is used to confirm the validity of the results. The sensitivity of the selected parameter subset to initial parameter values is addressed using Monte Carlo simulations and Wu's MSE-based method. A new MSE-based ranking and selection strategy is also described, and the two techniques are compared with a third MSE-based strategy recently developed by Chu et al. (2009). The computational requirements are computed and compared and recommendations are provided to model users.

1.2 References

Bellman, R. and K. J. Åström, "On Structural Identifiability," Math. Biosci. **7**(3-4), 329-339 (1970).

Biegler, L. T., J. J. Damiano and G. E. Blau, “Nonlinear Parameter Estimation: A Case Study Comparison,” AIChE J. **32**(1), 29-45 (1986).

Brun, R., P. Reichert and H. R. Künsch, "Practical Identifiability Analysis of Large Environmental Simulation Models," Water Resour. Res. **37**(4), 1015-1030 (2001).

Brun, R., M. Kühni, H. Siegrist, W. Gujer and P. Reichert, "Practical Identifiability of ASM2d Parameters - Systematic Selection and Tuning of Parameter Subsets," Water Res. **36**(16), 4113-4127 (2002).

Chu, Y., Z. Huang and J. Hahn, "Improving Prediction Capabilities of Complex Dynamic Models via Parameter Selection and Estimation," Chem. Eng. Sci. **64**(19), 4178-4185 (2009).

Foss, B. A., B. Lohmann and W. Marquardt, "A Field Study of the Industrial Modeling Process," J. Process Control. **8**(5-6), 325-338 (1998).

Gadkar, K. G., R. Gunawan and F. J. Doyle III, "Iterative Approach to Model Identification of Biological Networks," BMC Bioinformatics. **6**(155), (2005).

Hocking, R. R., "Analysis and Selection of Variables in Linear Regression," Biometrics **32**(1), 1-49 (1976).

Holmberg, A., "On the Practical Identifiability of Microbial Growth Models Incorporating Michaelis-Menten Type Nonlinearities," Math. Biosci. **62**(1), 23-43 (1982).

Jayasankar, B. R., A. Ben-Zvi and B. Huang, "Identifiability and Estimability Study for a Dynamic Solid Oxide Fuel Cell Model," Comput. Chem. Eng. **33**(2), 484-492 (2009).

Jiménez-Hornero, J. E., I. M. Santos-Dueñas and I. García-García, "Optimization of Biotechnological Processes. The Acetic Acid Fermentation. Part II: Practical Identifiability Analysis and Parameter Estimation," Biochem. Eng. J. **45**(1), 7-21 (2009).

Koeva, V. I., S. Daneshvar, R. J. Senden, A. H. M. Imam, L. J. Schreiner and K. B. McAuley, "Mathematical Modeling of PAG- and NIPAM-Based Polymer Gel Dosimeters Contaminated by Oxygen and Inhibitor," Macromol. Theory Simul. **18**(9), 495-510 (2009).

Kou, B., K. B. McAuley, J. C. C. Hsu and D. W. Bacon, "Mathematical Model and Parameter Estimation for Gas-Phase Ethylene/Hexene Copolymerization with Metallocene Catalyst," Macromol. Mater. Eng. **290**(6), 537-557 (2005a).

Kou, B., K. B. McAuley, C. C. Hsu, D. W. Bacon and K. Z. Yao, "Mathematical Model and Parameter Estimation for Gas-Phase Ethylene Homopolymerization with Supported Metallocene Catalyst," Ind. Eng. Chem. Res. **44**(8), 2428-2442 (2005b).

Littlejohns, J. V., K. B. McAuley and A. J. Daugulis, "Model for a Solid-Liquid Stirred Tank Two-Phase Partitioning Bioscrubber for the Treatment of BTEX," J. Hazard. Mater. **175**(1-3), 872-882 (2010).

Lund, B. F. and B .A. Foss, "Parameter Ranking by Orthogonalization - Applied to Nonlinear Mechanistic Models," Automatica. **44**(1), 278-281 (2008).

Maria, G., "A Review of Algorithms and Trends in Kinetic Model Identification for Chemical and Biochemical Systems," Chem. Biochem. Eng. Q. **18**(3), 195-222 (2004).

Petersen, B., K. Gernaey and P. A. Vanrolleghem, "Practical Identifiability of Model Parameters by Combined Respirometric-Titrimetric Measurements," Water Sci. Technol. **43**(7), 347-355 (2001).

Raue, A., C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller and J. Timmer, "Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood," Bioinformatics. **25**(15), 1923-1929 (2009).

Reichert, P., R. von Schulthess and D. Wild, "The Use of Aquasim for Estimating Parameters of Activated Sludge Models," Water Sci. Technol. **31**(2), 135-147 (1995).

Srinath, S. and R. Gunawan, "Parameter Identifiability of Power-Law Biochemical System Models," J. Biotechnol. **149**(3), 132-140 (2010).

Thompson, D. E., K. B. McAuley and P. J. McLellan, "Parameter Estimation in a Simplified MWD Model for HDPE Produced by a Ziegler-Natta Catalyst," Macromol. React. Eng. **3**(4), 160-177 (2009).

Vajda, S., H. Rabitz, E. Walter and Y. Lecourtier, "Qualitative and Quantitative Identifiability Analysis of Nonlinear Chemical Kinetic-Models," Chem. Eng. Commun. **83**(1), 191-219 (1989).

Walter, E. and L. Pronzato, "On the Identifiability and Distinguishability of Nonlinear Parametric Models," Math. Comput. Simul. **42**(2-3), 125-134 (1996).

Woloszyn, J. D. and K. B. McAuley, "Application of Parameter Selection and Estimation Techniques in a Thermal Styrene Polymerization Model," Macromol. React. Eng. (2011). doi: 10.1002/mren.201100021

Wu, S., T. J. Harris and K. B. McAuley, "The Use of Simplified or Misspecified Models: Linear Case," Can. J. Chem. Eng. **85**(4), 386-398 (2007).

Wu, S., K. B. McAuley and T. J. Harris, "Selection of Simplified Models: II. Development of a Model Selection Criterion Based on Mean Squared Error," Can. J. Chem. Eng. 89(2), 325-336 (2011a).

Wu, S., K. A. P. McLean, T. J. Harris and K. B. McAuley, "Selection of Optimal Parameter Set Using Estimability Analysis and MSE-Based Model-Selection Criterion," Int. J. Advanced Mechatronic Systems, **3**(3), 188-197 (2011b).

Yao, K. Z., B. M. Shaw, B. Kou, K. B. McAuley and D. W. Bacon, "Modeling Ethylene/Butene Copolymerization with Multi-Site Catalysts: Parameter Estimability and Experimental Design," Polym. React. Eng. **11**(3), 563-588 (2003).

Yue, H., M. Brown, J. Knowles, H. Wang, D. S. Broomhead and D. B. Kell, "Insights into the Behaviour of Systems Biology Models from Dynamic Sensitivty and Identifiability Analysis: a Case Study of an NF-κB Signalling Pathway," Mol. Biosyst. **2**(12), 640-649 (2006).

Chapter 2

Mathematical Modelling of Chemical Processes – Obtaining the Best Model Predictions and Parameter Estimates using Identifiability and Estimability Procedures[1]

2.1 Summary

Chemical engineers who develop fundamental models often have difficulties estimating all model parameters due to problems with parameter identifiability and estimability. These two concepts are reviewed, as are techniques for assessing identifiability and estimability. When some parameters are not estimable from the data, modellers must decide whether to conduct new experiments, change the model structure, or to estimate only a subset of the parameters and leave the others at fixed values. Estimating a reduced number of parameters can lead to better model predictions with lower mean squared error (MSE). MSE-based techniques for parameter subset selection are discussed and compared.

2.2 Introduction

Fundamental dynamic models are frequently used by chemical engineers to develop new chemical processes and to accurately predict the behaviour of existing industrial processes. For example, models can be used to simulate how a process will behave under new operating conditions, to indicate how new products can be made using an existing plant, to investigate

[1] The work in this Chapter has been accepted for publication in the *Canadian Journal of Chemical Engineering*. K.A.P. McLean and K.B. McAuley were co-authors of this journal article. Note that this thesis has been prepared using a manuscript format, so some nomenclature used is not consistent throughout the entire thesis. Please refer to Section 2.7 for the nomenclature used in this Chapter.

product quality enhancements, and to achieve production rate improvements. Models may also be used in the training of operators, the design and testing of control schemes, and the planning of experiments (Foss et al., 1998; Maria, 2004).

As shown in Figure 2.1, several steps are used in the development of fundamental models. Parameter estimation (Step 4) can be quite difficult for several reasons. The models that chemical engineers use are generally nonlinear in the parameters, so that parameters must be estimated using nonlinear optimization algorithms, requiring initial guesses for parameter values. It is important to use good initial guesses or to use several different sets of initial guesses because multiple optima may be encountered (Jitjareonchai et al., 2006). The modeller may only have a limited number of noisy data points, and further experiments may be expensive or infeasible, making estimation of all of the model parameters impossible. A model that attempts to accurately describe the underlying phenomena in a chemical process may be complex with many parameters, some of which may have little influence on the model predictions. In addition, the effects of some parameters on model predictions may be correlated with the effects of other parameters. For example, a pair of parameters may always appear together as a product in the model equations, making it impossible to obtain unique estimates of both parameters. The associated numerical difficulties that arise during parameter estimation can be attributed to the model structure or the limited experimental data that are available. When deciding whether additional experiments should be performed (Step 6a in Figure 2.1) or if the model equations should be revised (Step 6b), a modeller may want to determine whether such measures will help to alleviate parameter estimation difficulties. The concepts of parameter identifiability and estimability, which are summarized in Table 2.1, are useful ideas that can assist modellers in making these decisions.

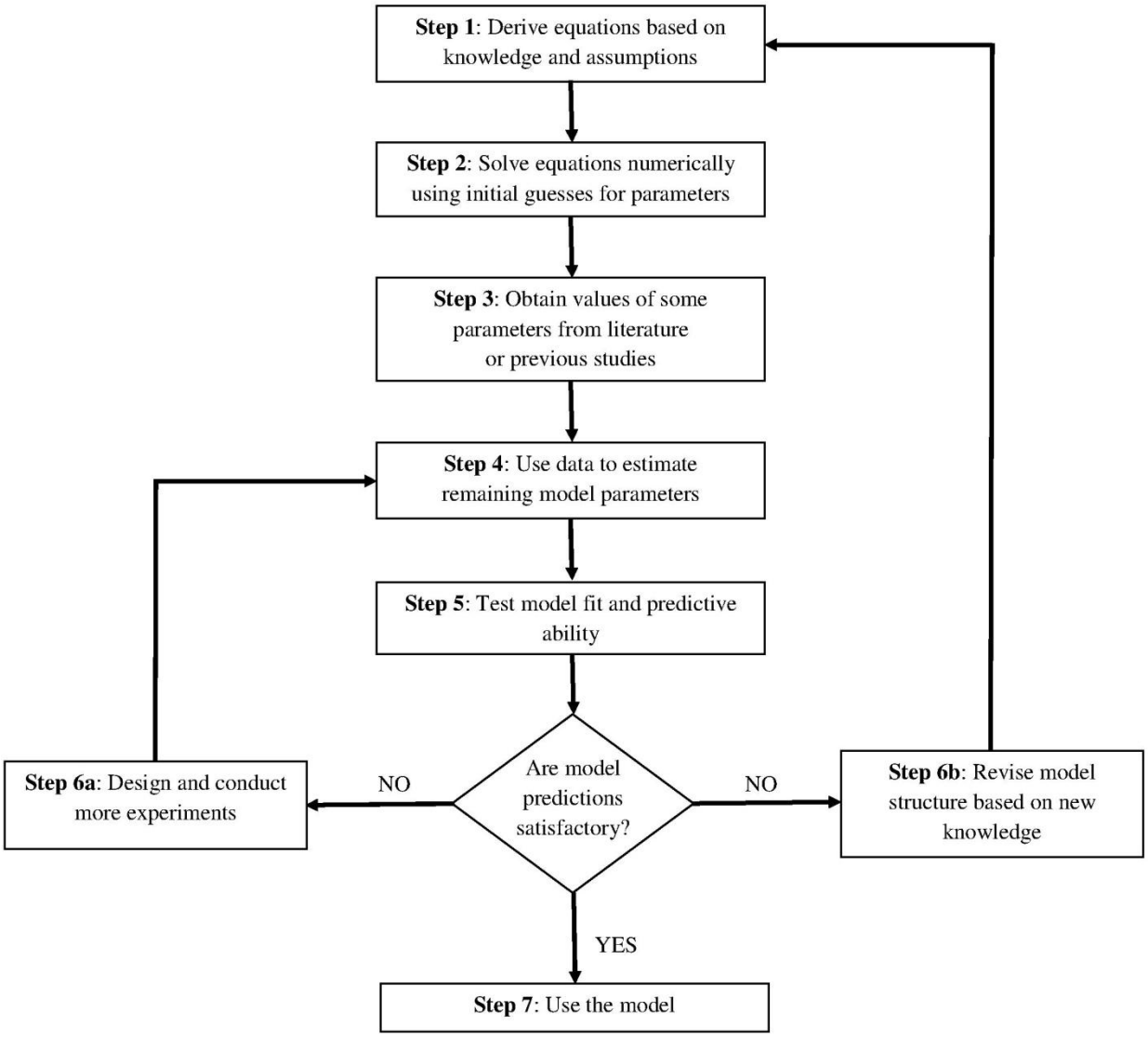

Figure 2.1: Steps in model formulation

Table 2.1: Identifiability and estimability

	Identifiability Analysis	Estimability Analysis
Question Answered	Can different values of model parameters lead to the same input-output behaviour for the model?	Can all parameter values be estimated uniquely from the available experimental data?
Alternative Names	Structural identifiability (Bellman and Åström, 1970) Theoretical identifiability (Pohjanpalo, 1978) Qualitative identifiability (Vajda et al., 1989b) a priori identifiability (Audoly et al., 2001)	Practical identifiability (Holmberg, 1982) Quantitative identifiability (Vajda et al., 1989b) a posteriori identifiability (Rodriguez-Fernandez et al., 2006)
Information Used	Model equations Types of Measurements	Model equations Types of measurements Experimental settings Initial parameter guesses
Mathematical Techniques	Linearization and Markov parameters (Grewal and Glover, 1976) Taylor series expansion (Pohjanpalo, 1978) Generating series (Walter and Pronzato, 1996) Similarity transformation (Vajda and Rabitz, 1989) Differential algebra (Ljung and Glad, 2004)	Methods based on Fisher Information Matrix (I_F) (Petersen et al., 2001) Graphical or visual inspection (Holmberg, 1982) Correlation matrix (Rodriguez-Fernandez et al., 2006) Collinearity index (Brun et al., 2001) Monte Carlo simulations (Nikerel et al., 2009)
Model type	Algebraic equations Ordinary differential equations (Chappell and Godfrey, 1992; Julien et al., 1998) Differential algebraic equations (Ben-Zvi, 2004; 2006; Xia and Moog, 2003)	Algebraic equations Ordinary differential equations (Vanrolleghem et al., 1995; Yao et al., 2003) Differential algebraic equations (Wu et al., 2011b) Partial differential equations
Model Complexity	Difficult to assess in models with more than ~ 10 parameters and states (Ljung and Glad, 2004; Jiménez-Hornero et al., 2008)	Can assess in models with more than 50 parameters and states (Kou et al. 2005a; Yue et al., 2006)

Assessing identifiability can uncover problems with model structure, whereas estimability is concerned with whether it is possible to estimate the parameters uniquely using the existing experimental data or data from a proposed set of experiments. Many complex models that are derived to describe chemical processes contain unidentifiable or inestimable parameters (e.g., Ben-Zvi et al., 2004; Kou et al., 2005a). If the parameters are inestimable, the modeller can attempt to resolve the problem by either simplifying the model or collecting additional experimental data. If the parameters are unidentifiable, the modeller will need to either simplify the model equations or will need to obtain information from additional output variables (e.g., the concentration of an intermediate chemical species). Performing additional experiments using the same measured outputs will not be sufficient to resolve an identifiability problem. Three common ways to simplify a model are to: 1) lump several parameters that always appear together into a single overall parameter (e.g., Vajda et al., 1989b; Chu and Hahn, 2009); 2) remove terms from model equations that are expected to have little influence on the model predictions (e.g., Maria and Rippin, 1993; Degenring et al., 2004); 3) fix some model parameters at reasonable values, based on prior knowledge about chemical or physical phenomena, thereby reducing the number of parameters that require estimation (e.g, Chu et al., 2009; Quaiser and Mönnigmann, 2009).

Sections 2.3 and 2.4 in the current Chapter discuss identifiability and estimability and describe techniques that can be used to determine whether all parameters in a model are identifiable and estimable. Some estimability analysis techniques can help modellers who cannot estimate of all of their parameters and who want to select an appropriate subset of the parameters to estimate. These subset selection techniques are summarized in Section 2.5. Advice for modellers about how and when to use different techniques is provided in Section 2.6.

2.3 Identifiability

When the parameters in a model cannot be uniquely estimated, the modeller may want to know whether this problem results from inappropriate model structure or insufficient experimental data. Identifiability analysis can be performed to determine if the model structure is the main problem. The concept of structural identifiability was first presented by Bellman and Åström (1970). As noted in Table 2.1, a model is identifiable if there exists a unique input-output behaviour for each set of candidate parameter values, or alternatively, if unique parameter estimates could be obtained using perfect noise-free data. Consider a nonlinear ordinary differential equation (ODE) model of the form:

$$\frac{dx}{dt} = f(x,u,\theta) \tag{2.1}$$

$$y = g(x,u,\theta) \tag{2.2}$$

where x is the vector of model states, u is the vector of (time varying) input trajectories, θ is vector of model parameters, and y is a vector of model predictions. The model and parameters are said to be identifiable if the following holds for any θ_1 and θ_2 belonging to the allowable parameter space:

$$g(x,u,\theta_1) = g(x,u,\theta_2) \Leftrightarrow \theta_1 = \theta_2 \tag{2.3}$$

If Equation (2.3) holds for the entire parameter space, the parameters are said to be globally identifiable, whereas if Equation (2.3) holds only for a neighbourhood of the parameter space the parameters are said to be locally identifiable (Jiménez-Hornero et al., 2008). For ODE models with $f(x,u,\theta)$ and $g(x,u,\theta)$ that are linear in the input and state variables, determining identifiability is relatively straightforward using existing techniques (see e.g., Walter, 1982; Jacquez and Grief,

1985). However, chemical engineers are often concerned with nonlinear ODE models. Assessing identifiability in nonlinear models is more difficult.

Several methods have been developed for assessing identifiability in nonlinear ODE models. Grewal and Glover (1976) showed that if the linearized version (i.e., linear with respect to states and inputs) of a nonlinear model is identifiable then the original nonlinear model is also identifiable. Confirming the identifiability of the linearization of a nonlinear model is a sufficient condition for the nonlinear model being identifiable. Their method has been used in several studies, e.g., Ben-Zvi et al. (2006), Dochain et al. (1995) and Julien et al. (1998), and is the easiest method for confirming identifiability of a nonlinear ODE or differential algebraic equation (DAE) model. Unfortunately, non-identifiability of the linearized version of a nonlinear model does not guarantee that the original nonlinear model is also non-identifiable (Vajda et al., 1989b).

Several researchers have developed techniques to address this problem, i.e., methods that can be used to confirm that a nonlinear model is not identifiable, which are described in a recent review article by Miao et al. (2011). Pohjanpalo (1978) proposed a Taylor-series-expansion approach, used by Dochain et al. (1995) and Petersen et al. (2003), requiring higher-order derivatives of system outputs with respect to time. The number of derivatives required is at least equal to the number of parameters, but has no guaranteed upper bound. Walter and Lecourtier (1982) developed a generating-series approach that uses simpler mathematical expressions than Pohjanpalo's method. Jayasankar et al. (2009) recently applied Walter and Lecourtier's method when assessing identifiability of a solid-oxide fuel-cell model with 17 states and 5 parameters. This model was too complex to be analyzed using the generating-series approach, so the authors divided it into four submodels (each with three states and two parameters) to enable identifiability analysis. One of the submodels, which involved DAEs, could not be assessed using this method.

Vajda and Rabitz (1989) and Vajda et al. (1989a) developed a similarity transformation (or local state isomorphism) approach that has been applied by Chappell and Godfrey (1992) and Julien et al. (1998) for models with 2 states and 5 parameters, and 3 states and 10 parameters, respectively. Vajda's method works well for simple single-input single-output models, and becomes more difficult to implement for models with multiple inputs, outputs and parameters, requiring the generation and solution of partial differential equations, along with checks for controllability and observability conditions (Vajda et al., 1989a). Ljung and Glad (1994) developed a differential algebra technique that is restricted to polynomial nonlinearities and Audoly et al. (2001) later developed an improved algorithm. This method has been applied using symbolic computational packages by Bellu et al. (2007) and Zhang et al. (2010), eliminating the need for the user to perform complicated algebraic calculations by hand. Bellu et al. demonstrated their differential algebra software for a drug metabolism model with 2 states and 4 parameters. Zhang et al. analyzed an enhanced biological phosphorus removal model with 3 states and 5 parameters. Jiménez-Hornero et al. (2008) attempted to use four of these techniques (i.e., Taylor series expansion, generating series, local state isomorphism and differential algebra) to assess identifiability of a nonlinear ODE model for an acetic-acid fermentation process with six states and nine parameters. All four techniques either failed to provide firm conclusions or were too difficult to implement due to the complexity of resulting analytical expressions and/or long computation times. The only nonlinear identifiability method that they were able to implement successfully was a probabilistic semi-numerical technique by Sedoglavic (2002) that makes use of similarities between local observability and structural identifiability. In this method, parameters are considered to be time-invariant state variables (i.e., $d\theta/dt = 0$) and the rank of the Jacobian matrix with respect to the states and parameters is used to determine local observability.

Jiménez-Hornero et al. (2008) concluded that a single generalized approach does not exist for all types of nonlinearities, and it is not possible to determine, in advance, which technique to use for a particular model. Many techniques become difficult to use or are computationally intractable for complex models with many parameters and states. Xia and Moog (2003) introduced an implicit-function-theorem approach, which they demonstrated using an HIV/AIDS model with four states and ten parameters. This method, which was not referenced or tested by Jiménez-Hornero et al., determines identifiability by verifying if $\det(\partial\Phi/\partial\theta) \neq 0$, where Φ is a function of the parameters and higher-order derivatives with respect to inputs and outputs. The rank of the $\partial\Phi/\partial\theta$ matrix is difficult to determine in models with many parameters (Wu et al., 2008).

In summary, several techniques have been developed for assessing the identifiability of nonlinear models. It should be noted that some of these methods can only test for local identifiability (e.g., Grewal and Glover, 1976; Sedoglavic, 2002; Wu et al., 2008), whereas other methods can test for global identifiability (e.g., Pohjanpalo, 1978; Vajda and Rabitz, 1989; Audoly et al., 2001). Methods for testing global identifiability are difficult to implement and are usually restricted to small systems. Aside from linearization-based techniques, none of the proposed methods have been demonstrated on a dynamic model with even moderate complexity (i.e., with more than 15 states and parameters). Furthermore, implementation of all of the proposed identifiability techniques (except for linearization) requires mathematical background and expertise that would be unfamiliar to most chemical engineers who develop fundamental models and estimate parameters from data. The linearization approach, which is the easiest method to use, can confirm that a nonlinear model is identifiable via a test on the linearized model. The main drawback of this method is that if the linearized model is not identifiable, firm conclusions cannot be made regarding the identifiability (or not) of the corresponding nonlinear

model. However, non-identifiability of the linearized model does provide a strong clue to the modeller that identifiability of the original model may be a problem, and that it might be wise to reformulate the model equations.

2.4 Estimability

Parameters that are deemed identifiable based on the model structure may not be estimable in practice due to limited information contained in the available experimental data. Parameters are estimable if their values can be uniquely determined from the available data. There are two reasons why a parameter would not be estimable: 1) the model predictions for the data may be insensitive to the value of the parameter, or 2) the effect of the parameter on the model predictions may be correlated with the effects of one or more other parameters. Changes in model predictions caused by adjusting one parameter can sometimes be counteracted by adjusting the values of other parameters. Therefore, to assess parameter estimability, both the sensitivity of model predictions to changes in parameter values and correlations between parameters need to be considered. Techniques for assessing estimability include: 1) graphical or visual inspection of sensitivity functions, 2) methods based on the Fisher Information Matrix (I_F) and parametric sensitivity matrix and 3) methods that require repeated parameter estimation.

2.4.1 Graphical or Visual Inspection of Sensitivity Functions

Consider a multivariate nonlinear dynamic model of the form:

$$\frac{dx}{dt} = f(x, u, \theta) \tag{2.4}$$

$$y = g(x, u, \theta) + \varepsilon \tag{2.5}$$

This model is the same as that in Equations (2.1) and (2.2), except that y is a vector of measured responses that are influenced by stochastic measurement errors ε. Sensitivities of model predictions to changes in parameter values are essential for assessing estimability in this ODE model. Parametric sensitivity functions can be expressed mathematically as:

$$S_{ij} = \frac{\partial g_i}{\partial \theta_j} \tag{2.6}$$

where the S_{ij} indicates the influence of a perturbation in parameter θ_j on the prediction of ith response g_i.

For simple dynamic models with few parameters and response variables, visual (graphical) inspection of sensitivity functions over the duration of each experimental run can yield information about the estimability of parameters. For example, Holmberg (1982) used a sensitivity plot similar to that in Figure 2.2 to show sensitivity functions for the concentration of microorganisms in a microbial batch growth model. As seen in Figure 2.2, the sensitivity functions for parameters μ_m and K_S for this response variable are proportional, indicating a strong correlation. As a result, Holmberg concluded that it would not be possible to estimate these parameters reliably. On the other hand, the different shapes for the sensitivity functions of parameters Y and K_d reveal that these two parameters could be estimated uniquely using dynamic microorganism concentration data.

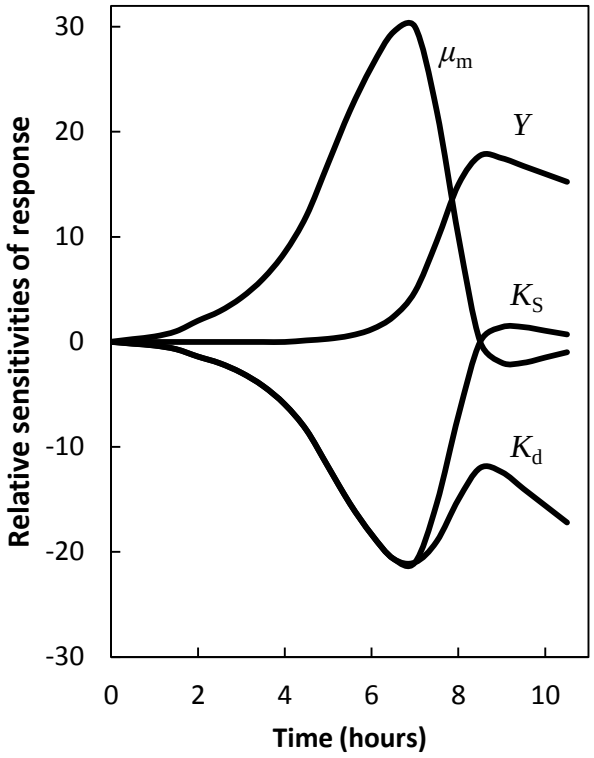

Figure 2.2: Relative sensitivity functions, similar to sensitivity plot in Holmberg (1982)

Visual inspection can also reveal when model predictions are insensitive to changes in a parameter. For example, parameter Y in Figure 2.2 has little influence on model predictions at times less than 5 hours, indicating that data at longer reaction times are required to estimate this parameter. Note that appropriate scaling is important when performing sensitivity analysis to ensure that apparently small sensitivity values do not result from a poor choice of units for the parameters or response variables. Reichert et al. (1995) and Petersen et al. (2001) used visual inspection for estimability analysis in models describing a denitrifying activated sludge process and a two-step nitrification process, respectively. Although this method is easy and intuitive to use for models with a small number of parameters and measured variables, visual inspection

becomes intractable for complex models. Estimability difficulties can arise from interrelationships among three or more parameters. For example, in a single-response model if $\partial g/\partial \theta_1 = 3\partial g/\partial \theta_2 + 4\partial g/\partial \theta_3$ at all times where data are collected, it is impossible to uniquely estimate all three parameters. Such higher-order interrelationships are difficult to detect visually, particularly when the number of parameters and responses to consider is large.

2.4.2 Methods Based on the Fisher Information Matrix and Sensitivity Matrix *S*

The Fisher Information Matrix (I_F) is important for estimability analysis as it contains information about measurement uncertainty and sensitivities of predicted responses to model parameters at all measurement times. Systems describing chemical processes often include multiple types of response variables (e.g., temperatures, concentrations) that are measured many times over the course of several experimental runs. In these systems, the total number of data points available for parameter estimation could be $N=dnr$, where d is the number of response variables, n is the number of sampling times at which measurements are taken, and r is the number of runs. A simple method to express the output vector in this situation is to use a stacked or "rolled-out" format (Seber and Wild, 2003):

$$\begin{pmatrix} y_{111} \\ \vdots \\ y_{1n1} \\ y_{211} \\ \vdots \\ y_{dn1} \\ y_{112} \\ \vdots \\ y_{dn2} \\ \vdots \\ y_{dnr} \end{pmatrix} = \begin{pmatrix} g_{11}(x,u_1,\theta) \\ \vdots \\ g_{1n}(x,u_1,\theta) \\ g_{21}(x,u_1,\theta) \\ \vdots \\ g_{dn}(x,u_1,\theta) \\ g_{11}(x,u_2,\theta) \\ \vdots \\ g_{dn}(x,u_2,\theta) \\ \vdots \\ g_{dn}(x,u_r,\theta) \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \vdots \\ \varepsilon_{1n1} \\ \varepsilon_{211} \\ \vdots \\ \varepsilon_{dn1} \\ \varepsilon_{112} \\ \vdots \\ \varepsilon_{dn2} \\ \vdots \\ \varepsilon_{dnr} \end{pmatrix} \tag{2.7}$$

The associated sensitivities of the predicted responses to the parameters can be placed in a sensitivity matrix, *S*:

$$S = \begin{pmatrix} \left.\frac{\partial g_{11}}{\partial \theta_1}\right|_{u_1} & \cdots & \left.\frac{\partial g_{11}}{\partial \theta_p}\right|_{u_1} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial g_{1n}}{\partial \theta_1}\right|_{u_1} & \cdots & \left.\frac{\partial g_{1n}}{\partial \theta_p}\right|_{u_1} \\ \left.\frac{\partial g_{21}}{\partial \theta_1}\right|_{u_1} & \cdots & \left.\frac{\partial g_{21}}{\partial \theta_p}\right|_{u_1} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial g_{dn}}{\partial \theta_1}\right|_{u_1} & \cdots & \left.\frac{\partial g_{dn}}{\partial \theta_p}\right|_{u_1} \\ \left.\frac{\partial g_{11}}{\partial \theta_1}\right|_{u_2} & \cdots & \left.\frac{\partial g_{11}}{\partial \theta_p}\right|_{u_2} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial g_{dn}}{\partial \theta_1}\right|_{u_2} & \cdots & \left.\frac{\partial g_{dn}}{\partial \theta_p}\right|_{u_2} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial g_{dn}}{\partial \theta_1}\right|_{u_r} & \cdots & \left.\frac{\partial g_{dn}}{\partial \theta_p}\right|_{u_r} \end{pmatrix} \tag{2.8}$$

The dimensions of S are $N \times p$ where each row corresponds to a particular response at a particular time and each column contains derivatives with respect to a particular parameter. If there are missing data values for certain response variables at particular times, the corresponding rows can be removed from the vectors and matrices in Equations (2.7) and (2.8). For models with analytical solutions, it is possible to obtain the derivatives in Equation (2.8) analytically. For more complicated nonlinear ODE models, the elements of S must be computed numerically, either by difference approximations with perturbed parameter values (Saltelli et al., 2000) or by solving sensitivity equations (Leis and Kramer, 1988). Note that initial parameter guesses are required to obtain numerical values for the elements of S, as indicated in Table 2.1.

The sensitivity matrix in Equation (2.8) is related to I_F by:

$$I_F = S^T Q S \tag{2.9}$$

where Q is a square weighting matrix (i.e., the inverse of the variance-covariance matrix for the response variables) that reflects the degree of uncertainty in different types of response variables (Munack, 1991). Note that I_F has dimension $p \times p$, where p is the number of model parameters. The elements of Q provide measurement noise information, obtained either from replicate experimental runs or from making assumptions about the accuracies of different types of measurements.

A simple way to assess estimability is to check the rank of I_F. If the rank is less than p (i.e., I_F is singular) then the parameters are not all estimable (Petersen et al., 2001). Due to numerical uncertainties, it can be difficult in many cases to decide whether I_F has full rank or not. The condition number, which is the ratio of the largest eigenvalue to the smallest eigenvalue of I_F, can also be used to assess parameter estimability; large condition numbers may indicate inestimable or nearly inestimable parameters (Dochain and Vanrolleghem, 2001).

If the model parameters are inestimable, the modeller may choose to design and conduct additional experiments (Step 6a in Figure 2.1). In this situation (i.e., sequential experimental design), the S matrix can be augmented with additional rows corresponding to the proposed experimental conditions (Thompson et al., 2010). These new experimental conditions can be selected to optimize different scalar measures of the size of I_F. For example (Munack, 1991):

- A-optimal design criterion: $\min\left[\mathrm{tr}\left(I_F^{-1}\right)\right]$ (2.10)
- D-optimal design criterion: $\max\left[\det\left(I_F\right)\right]$ (2.11)
- E-optimal design criterion: $\max\left[\lambda_{\min}\left(I_F\right)\right]$ (2.12)

- Modified E-optimal design criterion: $$\min\left[\frac{\lambda_{\max}(I_F)}{\lambda_{\min}(I_F)}\right] \quad (2.13)$$

where $\lambda_{min}(I_F)$ and $\lambda_{max}(I_F)$ are the smallest and largest eigenvalues of I_F. Atkinson et al. (2007) describe the theory of optimal design and the application of design criteria. These criteria, which were used for optimal experimental design by Vanrolleghem et al. (1995), can also be used to aid in selecting a subset of estimable parameters, as will be shown in Section 2.5.1. The modified E-optimality criterion is the condition number of I_F. Assessing estimability using only the condition number can be problematic, as this measure is associated only with the shape of the linearized joint parameter confidence region, and not the size. The values of the measures above depend on the units selected for the parameters, which can lead to different conclusions about estimability for a fixed threshold value when the parameter units are changed (De Pauw, 2005).

The Fisher Information matrix becomes singular when the effects of some parameters on model predictions are highly correlated with the effects of others. A correlation matrix can help to identify linear dependencies between two columns of S (i.e., linear dependencies between parameters). According to the Cramèr-Rao theorem (Ljung, 1999), assuming that the measurement noise is uncorrelated and from a normal distribution with mean zero, the covariance matrix is related to I_F by:

$$C \geq I_F^{-1} \quad (2.14)$$

where C is the parameter estimation error covariance matrix, which can be used to compute elements of the correlation matrix:

$$R_{jl} \cong \frac{C_{jl}}{\sqrt{C_{jj}C_{ll}}} \quad (2.15)$$

C_{jl} is the covariance between the jth and lth parameter estimates, and $\sqrt{C_{jj}}$ and $\sqrt{C_{ll}}$ are standard deviations for the jth and lth parameter estimates, respectively. Correlation values close to ±1 indicate highly correlated parameters; pairs of parameters with correlations equal to ±1 are exactly linearly correlated. The correlation matrix approach has been applied to assess estimability in complex biological models. For example, Rodriguez-Fernandez et al. (2006) and Yue et al. (2006) identified inestimable parameters using a cut-off correlation value of ±0.99. A major shortcoming of the correlation matrix method is that this method does not consider the magnitudes of the sensitivity coefficients, but only directions of the sensitivity vectors (i.e, the columns of S). For example, a parameter may have low correlation with all other parameters, but will not be estimable if it has little influence on the model predictions over the range of permissible parameter values.

In similar work, Brun et al. (2001) proposed a collinearity index, γ, to detect correlations between the effects of two or more parameters. This index is based on a scaled sensitivity matrix, Z, wherein the elements of S are multiplied by scaling factors to ensure dimensional consistency:

$$Z_{ij} = S_{ij} \frac{s_{\theta_j 0}}{s_{yi}} \tag{2.16}$$

$s_{\theta j0}$ is a scaling factor that reflects uncertainty in the initial guess for parameter θ_j and s_{yi} is a scaling factor (e.g., a standard error) that reflects uncertainty in the ith measured value (Thompson et al, 2009). The uncertainty in the initial parameter guess can be estimated using information from previous parameter estimation studies or judgments about the physically reasonable ranges for the particular parameter. Values for s_{yi} are the square roots of the diagonal elements of the weighting matrix Q in Equation (2.9). Using the scaling shown in Equation (2.16), $Z^{\mathrm{T}}Z$ is a Fisher Information matrix, assuming that Q is diagonal and that the model is

expressed in terms of scaled dimensionless parameters $\theta_j/s_{\theta j0}$. Brun et al. computed a normalized matrix, $\tilde{Z}$, whose elements are:

$$\tilde{Z}_{ij} = \frac{Z_{ij}}{\left\| Z_j \right\|} \tag{2.17}$$

where $\left\| Z_j \right\|$ is the Euclidean norm (i.e., square root of sum of squared elements) for the *j*th column of Z. The collinearity index for the complete parameter set is:

$$\gamma = \frac{1}{\sqrt{\tilde{\lambda}}} \tag{2.18}$$

where $\tilde{\lambda}$ is the smallest eigenvalue of $\tilde{Z}^T\tilde{Z}$. If the value of γ is equal to one, then all of the columns are orthogonal, indicating that all parameters can be estimated together. As columns become linearly dependent, the value of γ approaches infinity. The parameters are categorized as inestimable if γ exceeds an arbitrary threshold or cut-off value. γ is closely related to the condition number and Brun et al. (2001) noted that there are only slight differences between results determined using the two measures. The collinearity index can be used in combination with other sensitivity measures to compare the estimability of subsets of parameters, as will be described in Section 2.5.2.

While correlation matrix and collinearity index methods consider only the interactions among parameters and neglect magnitudes of sensitivities, other methods consider only the sizes of sensitivities and neglect parameter interactions. For example, Sidoli et al. (2005) developed a parameter perturbation method that evaluates output sensitivities with respect to individual parameters. A set of proposed experimental settings is first specified. Experiments with these settings are simulated with parameters perturbed one at a time using a higher and lower bound (chosen by Sidoli et al. as ±25% of the nominal parameter values). The influence of each

parameter perturbation on each output variable is quantified using a normalized "output separation" variable. This separation variable is an integral of the difference between the output using high and low values of the parameter, calculated over the duration of each simulated experiment. If the value of the separation variable exceeds an arbitrary threshold tolerance for at least one of the model outputs, then the parameter is categorized as estimable. Sidoli et al. demonstrated this technique using a single-cell mammalian-cell-culture model with 357 parameters (Sidoli et al., 2005) and an extended population balance model with over 700 parameters (Sidoli et al., 2006).

In summary, many estimability techniques rely on parametric sensitivity coefficients and the resulting Fisher Information matrix. The main problems associated with these techniques are: i) computation of sensitivity coefficients requires initial parameter guesses that may be inaccurate, ii) all methods require a somewhat arbitrary cut-off value for categorizing the model parameters as estimable or inestimable, iii) unless parameters are scaled appropriately, different conclusions may be drawn when the model is formulated using different units. Correlation and collinearity index techniques may be helpful to modellers who want to simplify their inestimable model by lumping some parameters together. However, methods that rely solely on a correlation matrix, a collinearity index or a condition number for deciding parameter estimability may result in poor conclusions due to lack information about the magnitudes of sensitivities. Poor conclusions can also result from methods that focus solely on sensitivities for individual parameters and do not include information about correlation. All of the estimability techniques described above require model equations, a set of experimental settings and knowledge about when the various outputs were or will be measured. Actual experimental data are not needed, permitting estimability analysis to be used prior to actual experimentation. If a modeller is concerned about misleading

results due to poor initial parameter guesses, the estimability analysis can be repeated using different sets of candidate parameter values.

2.4.3 Methods That Require Repeated Parameter Estimation

A Monte Carlo method for estimability analysis was developed by Nikerel et al. (2009) who were interested in whether measurement noise would cause parameter estimates to be unreliable. Their approach involves generating a large number of simulated data sets (e.g., 500 data sets were generated using a metabolic reaction network model) with a different sequence of random measurement errors for each simulation. All of the simulations are conducted using the same set of potential parameter values. Parameter estimation is then attempted for each data set, and uncertainty levels for each parameter are estimated from the resulting reference set of parameter estimates. Unacceptably high variation in the values obtained for some (or all) parameters indicate that these parameters are not estimable. To use this method, cut-off values for parameter uncertainty levels are required to distinguish between estimable and inestimable parameters. Nikerel et al. (2009) used their method to assess estimability in an ODE model containing 14 parameters. This method is easy to use, but unfortunately it is computationally expensive to implement as it requires many parameter estimations. Also, misleading results can be obtained if the parameter values used to generate the simulation results are too far from the "true" parameter values for the process that will generate the real data.

The Jackknife and the bootstrap are two methods that are also commonly used to analyze the uncertainty of parameter estimates. Unlike the Monte Carlo approach described above, these methods involve resampling from a single data set and do not make assumptions about parameter values. The theory behind these techniques was summarized by Efron (1982). The jackknife technique relies on a sequential approach where N jackknife samples of size N-1 are created by

leaving out one data point at a time. The jackknife technique can also be applied by leaving out groups of data points instead to reduce computation time. For each jackknife sample, the parameters are estimated and the mean parameter values and the variances are calculated from the set of parameter estimates. For example, Zhang and Huck (1996) used the jackknife technique to obtain confidence intervals for four parameters in a model describing biofilm processes in water treatment, and showed that only one parameter could be reliably estimated with the limited data available. The bootstrap method involves generating a specified number of samples of size *N* by sampling with replacement from the original data set. In this method, a certain bootstrap sample may contain one of the original data points more than once or not at all. Using many bootstrap samples can improve the precision of the estimates, but also increases computation time. Duchesne and MacGregor (2001) developed a criterion based on the jackknife technique for model identification and concluded that the jackknife method is preferred to the bootstrap method due to lower computational requirements and the similarity of its approach with cross-validation.

Raue et al. (2009) developed a graphical profile-likelihood method to assess parameter estimability. Figure 2.3 shows two plots that are similar to graphs generated by Raue et al. when analyzing the estimability of a signalling pathway model with 7 parameters. The vertical axis on these plots corresponds to values of a weighted least-squares objective function, *J*, used for parameter estimation:

$$J = \sum_{i=1}^{d} \sum_{m=1}^{n} \sum_{q=1}^{r} \left(\frac{y_{imq} - \hat{y}_{imq}}{s_{yi}} \right)^2 \tag{2.19}$$

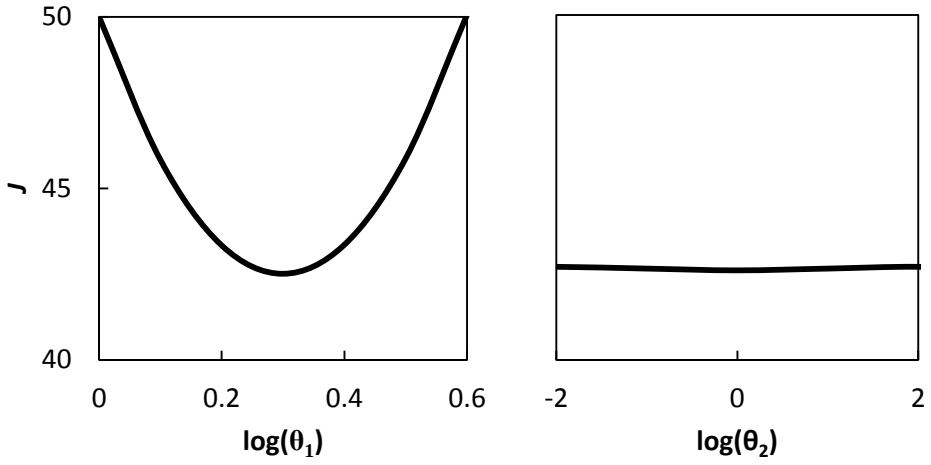

Figure 2.3: Profile-likelihood estimability plots similar to those generated by Raue et al. (2009).

Using Raue's method, numerous parameter estimation calculations are required wherein the marginal importance of each parameter on the objective function is considered. A range of candidate values is selected for each parameter of interest. For different candidate values from the range, the remaining parameters are adjusted to minimize the value of J (conditional on the perturbed value of the parameter of interest) to construct plots like those shown in Figure 2.3. If the parameter of interest has a negligible effect on J (e.g., θ_2 in Figure 2.3), the parameter is deemed inestimable. Conversely, if the value of J is strongly dependent on the value of the parameter of interest (e.g., θ_1 in Figure 2.3), the parameter is estimable. Note that judgment is required to select a threshold value for changes in J to categorize parameters as estimable or inestimable.

The techniques discussed above can be used to ascertain if all the parameters present in a model are estimable based on the data available. If all the parameters are estimable, then they are also identifiable, and the modeller will be able obtain unique parameter estimates. However, in

practice, fundamental models developed by chemical engineers are often complex and over-parameterized, and as a result, they contain unidentifiable or poorly estimable parameters (e.g., Reichert and Vanrolleghem, 2001; Kou et al., 2005a, b; Chu et al., 2009; Jimenez-Hornero et al., 2009). In these situations, modellers may decide to leave some parameters at nominal values and estimate only a subset of the parameters. Important questions for the modeller are: "Is this particular subset of the model parameters estimable?" and "Which subset of the parameters should I select for estimation?"

2.5 Selecting a Subset of Parameters to Estimate

As model parameters are not all estimable in many practical situations, it is often useful to determine which parameters can and should be estimated from the available or proposed data. When modellers consider estimating a subset of their model parameters, they generate a set of simplified models that are nested within the full model (Wu et al., 2011a). Different simplified models have different parameter subsets that will be estimated from the data, with the other parameters fixed at their nominal values. These simplified models can sometimes give better predictions, in terms of mean squared error (MSE), than the full model (Hocking, 1976; Wu et al., 2007), especially when the data available for parameter estimation are limited. MSE is the sum of the squared bias and variance, and is a convenient measure for the quality of the model predictions. Omitting parameters from the estimation problem leads to an increase in bias in the parameter estimates and model predictions, and also leads to a reduction in the variance of parameter estimates and model predictions. A simplified model gives better predictions than the original model when the variance reduction outweighs the bias.

Some techniques for selecting appropriate subsets for estimation involve comparing all possible parameter subsets. For example, a modeller with 54 parameters in his or her model could

use one of the optimal design criteria in Equations (2.10) – (2.13) to compare all possible subsets of 37 parameters, to determine the best 37 parameters to estimate. The next step is to determine the appropriate number of parameters to estimate (i.e., will better predictions be obtained by estimating the best 38 parameters than the best 37 parameters?). These combinatorial parameters subset selection techniques are computationally expensive because of the large number of possible subsets to compare (e.g., $\binom{54}{37} = 4.7 \times 10^{13}$). To avoid combinatorial problems, many researchers employ suboptimal techniques wherein parameters are ranked one at a time (i.e., from most estimable to least estimable (forward selection) or from least estimable to most estimable (backward selection)). After ranking the parameters, an appropriate number of parameters to estimate is determined. The main techniques for selecting estimable subsets are summarized in Table 2.2 and are discussed below.

Table 2.2: Methods for selecting estimable parameter subsets

Technique	Refs.	Approach	Comments
Methods based on Fisher Information matrix characteristics	Weijers and Vanrolleghem (1997) Machado et al. (2009)	Determine estimable subsets using scalar measures of I_F	Problems with combinatorial complexity
Correlation and collinearity methods	Jacquez and Grief (1985) Brun et al. (2002) Quaiser and Mönnigmann (2009) Chu and Hahn (2009)	Determine inestimable parameters through linear dependencies in columns of S	Considers pairs of parameters and directions of sensitivities, but not size of sensitivities
Principal component analysis (PCA) and eigenvalue based methods	Jolliffe (1972) Degenring et al. (2004) Vajda et al. (1989b) Schittkowski (2007) Quaiser and Mönnigmann (2009)	Determine subsets of estimable parameters based on eigenvectors and eigenvalues of S^TS or I_F	Difficult to combine three separate measures in PCA technique. Difficult to match eigenvalues with specific parameters.
Orthogonalization method	Yao et al. (2003) Kou et al. (2005a, b) Lund and Foss (2008) Thompson et al. (2009)	Determine parameter ranking based on parameter influence and correlations using forward selection approach	Easy-to-use method, provides definite ranking, but may be suboptimal compared with combinatorial approaches.

2.5.1 Methods Based on Scalar Measures of the Fisher Information Matrix

Weijers and Vanrolleghem (1997) developed a three-step technique based on measures of I_F for selecting estimable parameters in an activated sludge model. In the first step, a sensitivity analysis and arbitrary cut-off value were used to remove parameters with little influence on the model predictions. For example, the following mean sensitivity measure can be calculated to rank parameters from most to least influential (Brun et al., 2002):

$$\delta_j^{msqr} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} S_{ij}^2} \tag{2.20}$$

Small values of δ_j^{msqr} correspond to parameters with small influence on the model predictions, making these parameters difficult to estimate. Weijers and Vanrolleghem reduced their original 23-parameter model to 11 parameters. In their second combinatorial step, I_F was constructed for each possible subset containing 2 to 8 parameters and the D-optimality (determinant) and modified-E criterion (condition number) from Equations (2.11) and (2.13) were calculated and compared. In this step, the best two parameters for optimization were determined, followed by the best three parameters, then the best four, and so on up to eight. The determinant and condition number criteria sometimes resulted in selection of different optimal subsets of parameters. Despite the removal of unimportant parameters in the first step, the computational cost of this second step was still high (e.g., there were 1969 D- and E-optimal values that needed to be computed). This combinatorial problem becomes much more severe for problems with larger numbers of parameters.

The third step is to select the number of parameters that should be estimated. Weijers and Vanrolleghem did not use a clear criterion to select the appropriate number of parameters to estimate, but used their judgment to select a reasonably large number of parameters that would not cause numerical problems during estimation. Freni et al. (2009) used the method of Weijers and Vanrolleghem to study a 19-parameter urban drainage model. They reduced the number of parameters of interest to 14 in the first step, and then used the modified E-optimality criterion and a determinant measure to select optimal parameter subsets of each size. They considered subsets of size 5 to 14, which involved comparing approximately 14 000 combinations. The authors then used a determinant measure:

$$\rho_k = \det\left(X_k^T X_k\right)^{1/2k} \tag{2.21}$$

to determine the appropriate number of parameters to estimate, where X_k is the matrix containing columns from the scaled sensitivity matrix Z corresponding to the k candidate parameters for estimation. Freni et al. chose to estimate 7 parameters, because ρ_7 was the largest value obtained for their determinant measure.

Machado et al. (2009) used a different criterion to select the number of parameters to estimate in a 51-parameter activated sludge model:

$$RDE = \frac{\det(FIM)\left(\sum_{j=1}^{k}\theta_j^2\right)}{\min\left[\frac{\lambda_{\max}(FIM)}{\lambda_{\min}(FIM)}\right]} \tag{2.22}$$

This RDE criterion is a ratio between a determinant criterion and a modified E criterion. We do not recommend this RDE criterion, because appropriate scaling has not been used to ensure dimensional consistency. For example, the initial parameter guesses that appear in the numerator of Equation (2.22) may have numerical values with very different relative magnitudes that will change based on the units used when formulating the model equations.

2.5.2 Correlation and Collinearity Methods

The collinearity index γ was used by Brun et al. (2001) to detect linear dependencies between columns of a sensitivity matrix, as described in Section 2.4.2. Later, Brun et al. (2002) proposed an algorithm based on this measure to determine the best subset of parameters to estimate in their activated sludge model. Similar to the method proposed by Weijers and Vanrolleghem (1997), an initial sensitivity analysis step was performed to reduce the total number of parameters. Brun et al. used δ^{msqr} values from Equation (2.20) to obtain an initial ranked list for their 75 model parameters. They repeated this ranking step using two different sets of plausible parameter guesses and obtained a slightly different ranked parameter list. The authors then used their

judgment to decide which of the top 30 parameters they thought it would be possible to estimate based on their specified experimental design, and they selected 17 parameters for further analysis.

In a second step, Brun et al. used a combinatorial approach wherein they compared all possible parameter subsets of size 2 to 17. For each size, they selected the best parameter subset by calculating and comparing the determinant measure in Equation (2.21) and the collinearity measure in Equation (2.18). For a subset of *k* selected parameters to be potentially estimable, the collinearity measure, γ_k , needed to be less than their arbitrary threshold value of 10. They considered 6 to 9 parameters for estimation because subsets containing up to 9 parameters satisfied the collinearity constraint. Finally, six parameters were selected for estimation because unreasonable estimates for some parameters were obtained when more than six were estimated. The collinearity index has also been used to select subsets of parameters for estimation in an industrial hydrometallurgical leaching reactor model with 31 parameters (Díez et al., 2006), a biofilm model with 10 parameters (Brockmann et al., 2008), and a 9-parameter tree water flow and storage model (De Pauw et al., 2008).

Freni et al. (2009) also used the method proposed by Brun et al. in their study of an urban drainage model. Instead of using a computationally expensive combinatorial approach, Freni et al. employed a forward selection technique using the collinearity index, wherein parameters were sequentially added to the estimable subset to obtain the lowest possible value of γ_k. For example, starting with the parameter having the highest sensitivity measure, combinations of two parameters containing this parameter and each of the remaining parameters were considered. The second parameter chosen minimized the increase in the collinearity measure. Using this selection approach greatly reduced the computational cost; 91 sensitivity matrices were required using this sequential method compared with the more than 14 000 possible parameter subsets that were

considered using the combinatorial approach. Unfortunately, the next parameter selected depends on the previously selected parameters, and this approach may be suboptimal compared with combinatorial approaches.

The correlation matrix, described in Section 2.4.2, has also been used to aid in selecting subsets of estimable parameters. However, there are two problems associated with using this method (Quaiser and Mönnigmann, 2009): 1) there is no clear criterion to decide which parameter in a correlated pair is more estimable, and 2) when several pairs of parameters result in high correlation, there is no clear criterion to determine which of the pairs is more important for estimation. To address the first problem, Daun et al. (2008) identified all of the correlated pairs in a model for acute inflammatory response in rats. As long as the response variables were not highly sensitive to both parameters, the less influential parameter (based on normalized sensitivity coefficients) of the pair was fixed at its nominal value. Using this strategy, correlation values for 1035 pairs were calculated and 163 pairs were found to be highly correlated. The original 46 model parameters were reduced to an 18-parameter subset. Quaiser and Mönnigmann (2009) introduced the concept of total correlation to provide a means of ranking parameters from least estimable to most estimable, thereby addressing both problems. The highest total correlation, as calculated below, determines the next parameter to fix at its nominal value:

$$c_j^{tot}(L) = \sum_{l \in L, l \neq j} R_{jl}^* \tag{2.23}$$

where

$$R_{jl}^* = \begin{cases} \left|R_{jl}\right|, \text{if} \left|R_{jl}\right| \geq 1 - \xi_c \\ 0, \text{if} \left|R_{jl}\right| < 1 - \xi_c \end{cases} \tag{2.24}$$

L contains all of the parameters that have not yet been selected and ξ_c is a tuning parameter. Absolute values of correlations greater than 1-ξ_c are considered highly correlated. This ξ_c value is

initially set near zero and increased gradually to obtain a ranking of parameters. Quaiser and Mönnigmann (2009) demonstrated this technique using three large signal transduction pathway models with up to 98 parameters and 103 states.

Chu and Hahn (2009) developed a similarity measure to determine the angle, ϕ_{jl}, between two columns in a sensitivity matrix:

$$\cos\phi_{jl} = \frac{\left|S_j^T S_l\right|}{\left\|S_j\right\|_2 \left\|S_l\right\|_2} \tag{2.25}$$

This measure was used to cluster correlated parameters into groups. Chu and Hahn then determined the most influential parameter from each group using lengths of sensitivity vectors. Chu and Hahn (2009) and Huang et al. (2010) applied this technique when studying signal transduction pathway models with 115 and 118 parameters, respectively, and selected 8 and 19 parameters, respectively, for estimation. This method has also been applied to a 19-parameter activated sludge model (Kim et al., 2010). A drawback of the correlation and similarity methods of Quaiser and Mönnigmann and Chu and Hahn is their focus on pair-wise correlations only. It can be important to consider correlated effects between three or more parameters when deciding which parameters are estimable. For example, if one column in the *Z* matrix in Equation (2.16) is the sum of two other columns, it will be impossible to estimate all three parameters from the data, even though the correlations between any two of the columns may not appear to be large. Quaiser and Mönnigmann indicated that false positives can be obtained using their method. They used an example to illustrate that a pair of parameters with sensitivity vectors that are not linearly dependent sometimes results in a correlation of 1.0 or -1.0.

2.5.3 Methods Based on Principal Component Analysis (PCA) and Eigenvalues

Degenring et al. (2004) developed a technique based on principal component analysis (PCA) that uses a sensitivity matrix different from the one defined in Equation (2.8). Here, separate matrices, S^i, are formulated *for each of the d response variables*, where *i* is a superscript corresponding to the particular response variable measured at *n* different times in *r* experimental runs:

$$S^i = \begin{pmatrix} \left.\frac{\partial g_{i1}}{\partial \theta_1}\right|_{u_1} & \cdots & \left.\frac{\partial g_{i1}}{\partial \theta_p}\right|_{u_1} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial g_{in}}{\partial \theta_1}\right|_{u_1} & \cdots & \left.\frac{\partial g_{in}}{\partial \theta_p}\right|_{u_1} \\ \left.\frac{\partial g_{i1}}{\partial \theta_1}\right|_{u2} & \cdots & \left.\frac{\partial g_{i1}}{\partial \theta_p}\right|_{u_2} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial g_{in}}{\partial \theta_1}\right|_{u_r} & \cdots & \left.\frac{\partial g_{in}}{\partial \theta_p}\right|_{u_r} \end{pmatrix} \tag{2.26}$$

The elements of each S^i matrix are scaled to produce Z^i matrices with dimensionless elements. Note that the Z^i matrix used by Degenering et al. was scaled using nominal values of the *i*th response variable and the *j*th parameter, rather than uncertainties as in Equation (2.16). Eigenvalues and eigenvectors are obtained for $Z^{iT}Z^i$, and the eigenvalues are ordered from smallest to largest:

$$\left|\lambda_1^i\right| < \left|\lambda_2^i\right| < \cdots < \left|\lambda_p^i\right| \tag{2.27}$$

The corresponding eigenvectors are columns in the matrix:

$$\Gamma^i = \left(v_1^i, v_2^i, ..., v_p^i\right) = \begin{pmatrix} v_{1,1}^i & \cdots & v_{1,p}^i \\ \vdots & \ddots & \vdots \\ v_{p,1}^i & \cdots & v_{p,p}^i \end{pmatrix} \tag{2.28}$$

Degenring et al. (2004) used the following three methods described by Jolliffe (1972) to select parameter subsets for each Z^i:

1) Starting with the smallest eigenvalue, the parameters are ranked from least estimable to most estimable (backward selection) by selecting the parameter with the largest weight (i.e., absolute value) in the corresponding eigenvector. In this way, the least-estimable parameter is selected, followed by the second least-estimable and so on, until all of the parameters are ranked. If the highest weighting corresponds to a parameter that has already been ranked, then the parameter with the next-highest weighting is selected.
2) Parameters are ranked from least estimable to most estimable (backward selection) by comparing row-wise sums of squares (for parts of rows) from Γ^i. The first parameter selected is the one with the largest squared value in the first eigenvector (i.e., the same parameter selected by method 1). The *j*th-least estimable parameter is selected by comparing the row-wise sums of squares of first *j* elements in each row of Γ^i. The parameter corresponding to the largest sum of squares (for parameters that have not already been selected) becomes the *j*th parameter on the list.
3) This forward selection method is the opposite of method 1. Starting with the eigenvector $v_p{}^i$ (corresponding to the largest eigenvalue) the most-estimable parameter is selected by finding the element with the largest absolute value in the corresponding eigenvector. The second most estimable parameter is selected by examining eigenvector $v_{p-1}{}^i$, and so on.

Different rankings may be obtained using each method for a particular Z^i matrix, and all three require an arbitrary threshold value. The main problem with this approach, however, is that the

authors are able to rank the parameters according to which ones are most estimable using a particular output but, using their three techniques, it is difficult to know which parameters can and should be estimated using all of the existing or proposed measurements simultaneously.

Degenring et al. (2004) used this PCA approach to reduce the number of parameters to estimate from 122 to 79 in a model that describes cellular metabolism for a type of *E. coli*. Parameters that were not included in the estimable subset are those that were determined to be inestimable using all three methods and all 10 types of measurements. Recently, Quaiser and Mönnigmann (2009) attempted to apply this PCA-based method in their case study of three large signal transduction pathway models and found it to be inferior to eigenvalue methods of Schittkowski (2007) and Quaiser and Mönnigmann (2009), and the orthogonalization method of Yao et al (2003), which are described below.

Vajda et al. (1989b) introduced an eigenvalue-based method that requires manual inspection of the eigenvalues and eigenvectors of S^TS to group together lumped parameters (e.g., of the form θ_1/θ_2 and $\theta_1 \times \theta_2$). Because such a manual approach only works well for small models, Schittkowski (2007) and Quaiser and Mönnigmann (2009) independently developed similar, improved algorithms based on this method that involve automatic analyses to rank parameters. In each iteration of the improved method, one parameter is added to a subset of inestimable parameters (backward selection), which are fixed at their nominal values for the estimation step. The corresponding column is removed from S, forming an X_k matrix containing the potentially estimable parameters. The eigenvalues and eigenvectors of $X_k^{\mathrm{T}}X_k$ are then calculated and the eigenvalues are ordered such that $\lambda_1 < \lambda_2 < ... < \lambda_k$. An arbitrary small cut-off value, ξ, is set by the user. If the smallest eigenvalue is less than ξ then a parameter is selected for removal. The parameter that is removed is the one with largest weight in the eigenvector corresponding to the

smallest eigenvalue. In this way, parameters are ranked from least to most estimable, until all of the small eigenvalues are removed, so that the remaining unranked parameters form the estimable subset.

Schittkowski (2007) demonstrated this technique using an isothermal batch reactor model with 10 states and 9 parameters (Biegler et al., 1986). Quaiser and Mönnigmann (2009) demonstrated this technique using their three large signal transduction pathway models. The eigenvalue method and the orthogonalization method, which will be discussed next, were the most successful of the four methods tested and both techniques produced identical results for all three models.

2.5.4 Orthogonalization Method

The orthogonalization method was originally proposed by Yao et al. (2003) for selecting estimable parameters in a complex dynamic reactor model for ethylene-butene copolymerization with 50 parameters. Yao et al. used average values of response variables and initial parameter values for scaling. Subsequently, Thompson et al. (2009) improved the algorithm using uncertainty-based scaling factors (as shown in Equation (2.16)) which permit the modeller to incorporate information about the reliability of measurements and initial parameter guesses. The algorithm, shown below in Table 2.3, assesses the "net influence" of a parameter, taking into account its influence on measured responses and correlations with other parameters. Many researchers have used this forward-selection method to rank parameters in complex chemical and biological models (e.g., Gadkar et al., 2005; Kou et al., 2005a, b; Yue et al., 2006; Jayasankar et al., 2009; Koeva et al., 2009; Thompson et al., 2009; Littlejohns et al., 2010; Srinath and Gunawan, 2010). A similar method was later proposed by Lund and Foss (2008) wherein the successive orthogonalization technique to rank parameters in an industrial dynamic simulation model for submerged arc silicon furnaces.

Table 2.3: Orthogonalization Algorithm (Wu et al., 2011b)

1. Calculate the magnitude (i.e., the Euclidean norm) of each column in the scaled sensitivity matrix Z. The most estimable parameter corresponds to the column in Z with the largest magnitude. Set k=1.
2. Put the k columns from Z that correspond to parameters that have been ranked into matrix X_k.
3. Use X_k to predict columns in Z using ordinary least-squares: $$\hat{Z}_k = X_k \left(X_k^T X_k \right)^{-1} X_k^T Z \quad (2.29)$$ and calculate the residual matrix $$R_k = Z - \hat{Z}_k \quad (2.30)$$
4. Calculate the magnitude of each column in R_k. The $(k+1)^{th}$-most estimable parameter corresponds to the column in R_k with the largest magnitude.
5. Increase k by 1, and put the columns corresponding to the k+1 parameters that have been ranked in matrix X_k
6. Advance the iteration counter (subscripts for X and R) and repeat Steps 3 to 5, until all parameters are ranked or until it is impossible to perform the least-squares prediction of Z in Step 3 due to matrix singularity.

Modellers who have used this ranking algorithm have used a variety of techniques to determine the appropriate number of parameters to select for estimation from the ranked list. Yao et al. (2003), Kou et al. (2005a, b), and Jayasankar et al. (2009) used arbitrary cut-off values for the magnitude of the largest column in R_k. Littlejohns et al. (2010) estimated parameters sequentially (the top parameter, then the top two parameters, and so on) until there was negligible improvement in the value of the weighted least-squares objective function for parameter estimation. Thompson et al. (2009) used cross-validation, which is computationally expensive, to determine the optimal number of parameters to ensure that the model had the best predictive ability. Lin et al. (2010) used the orthogonalization approach from Yao et al. (2003) and Lund and Foss (2008) to initially reduce the number of parameters in their styrene polymerization model. Starting with no parameters in a ranked list, they estimated the increase in variance resulting from the addition of each parameter, and stopped when the final selected parameter resulted in a variance increase that is larger than an arbitrary threshold. The authors then

attempted to simultaneously estimate the k parameters initially selected. If no unique solution was obtained, then the (k-1)th parameter was removed and parameter estimation was attempted again. After a unique solution was obtained, additional parameters were removed from the estimable subset if they had large confidence intervals compared to the size of the parameter estimate. This step also required an arbitrary cut-off value to decide whether the uncertainty is sufficiently large to warrant removal of one or more parameters.

Wu et al. (2007; 2011a, b) developed a simple criterion to determine the number of parameters to estimate from a ranked parameter list based on MSE. The approach, outlined in Table 2.4, requires repeated parameter estimation to determine the optimal value of J_k (for k = 1...p), which is the objective function in Equation (2.19) with only the top k parameters estimated. This approach, like those of Thompson et al. (2009), Raue et al. (2009), Littlejohns et al. (2010) and Lin et al. (2010), requires experimental data and parameter estimation to determine the appropriate number of parameters to estimate. Other parameter subset selection techniques that have been discussed require experimental settings but no real data (e.g., Holmberg, 1982; Weijers et al., 1996; Brun et al., 2001). Poor numerical conditioning may arise in some cases, and it may not be possible to compute the optimal value of J_p by estimating all p parameters. In such situations, the best possible value of J_p can be estimated using a sufficiently large number of parameters. Wu et al. used the values of J_k and J_p to calculate a critical ratio $r_{C,k}$, which is a ratio of the squared bias introduced by estimating only k parameters divided by the variance reduction due to estimating k rather than p parameters. The optimal number of parameters to estimate in the ranked list corresponds to the lowest value of $r_{CC,k}$, which is a corrected critical ratio that accounts for the different number of parameters being estimated in each nested model. Wu et al. used their method to determine how many parameters to select for estimation in a 9-parameter model for a

batch reactor (Wu et al., 2011b; Biegler et al., 1986) using limited data. They showed that when data are available at only one operating temperature, 5 parameters are estimable. Wu's method has also been used in a styrene polymerization model to show that 9 of the 40 model parameters are estimable from the available literature data (Woloszyn and McAuley, 2011), and in a nylon 66 degradation model to determine that 48 of the 56 parameters should be estimated (Karimi et al., 2011).

Table 2.4: MSE-based approach (Wu et al.) to select the number of parameters to estimate (Wu et al., 2011b)

1. Start with a ranked parameter list (i.e., use the algorithm in Table 2.3)
2. Estimate the top-ranked parameter using weighted least-squares regression, with all others fixed at initial guesses. Next, estimate the top two parameters, followed by the top three parameters and so on, until all of the ranked parameters have been estimated. Denote the value of the objective function with the top k parameters estimated and the remaining p-k parameters held fixed as $J_{k..}$
3. Compute the critical ratio $$r_{C,k} = \left(J_k - J_p\right)/\left(p-k\right) \quad (2.31)$$ for k=1...p-1.
4. For each value of k, compute the corrected critical ratio $$r_{CC,k} = \frac{p-k}{N}\left(r_{CKub,k} - 1\right) \quad (2.32)$$ where $$r_{CKub,k} = \max\left(r_{C,k} - 1, \frac{2}{p-k+2} r_{C,k}\right) \quad (2.33)$$
5. Select the value of k corresponding to the lowest value of $r_{CC,k}$ as the appropriate number of parameters to estimate.

Chu et al. (2009) proposed a Monte Carlo method, shown in Table 2.5, wherein another MSE-based approach is used to rank parameters and determine the appropriate number of parameters to estimate. Unlike the method of Wu et al., Chu's method does not require experimental data. Also, it is less prone to problems arising from poor initial parameter guesses because many different sets of possible parameter values are considered, and different sensitivity coefficients are

calculated for each candidate set of potential parameter values. However, Chu's method is computationally expensive and may be prohibitive for large models. For example, when Chu et al. applied this forward-selection method to a 26-parameter signalling pathway model, they recomputed the elements of *S*, the unscaled sensitivity matrix, 10^4 times to estimate the bias associated with incorrect parameter values. In Chu's method, described in Table 2.5, matrix *T* contains sensitivity vectors for parameters that may be included in the estimation and *W* contains parameters that are left out. $\bar{\theta}_U$ is the vector of nominal values for the remaining unselected parameters and θ_U^* is the vector of potential true values for the unselected parameters. Note that the method of Chu et al. is formulated for a model with a single response variable, but it could be extended for use with multi-response models expressed in the "rolled-out" form (Seber and Wild, 2003) used by Wu et al. (2011b).

Table 2.5: MSE-based approach (Chu et al.) to rank and select the number of parameters to estimate

1. Specify initial guesses (nominal values) and parameter ranges (uncertainties) and distributions for all p parameters, as well as the variance σ^2 for the response variable.
2. Randomly select 10^4 possible sets of true parameter values from the specified parameter ranges. For each set of candidate parameter values, evaluate and record the sensitivity matrix S.
3. Starting with no parameters in the ranked list, formulate p candidate parameter subsets each with 1 parameter that will be estimated, so that p-1 parameters that will be left out. For each selected parameter (for $j = 1, \ldots, p$), place its corresponding column from S into T and place the remaining columns in W
4. For $j = 1, \ldots, p$, evaluate the average bias, considering all 10^4 potential sets of true values, caused by selecting parameter j, but leaving the others at their nominal values: $$\left(\bar{\theta}_U - \theta_U^*\right)^T \left[W^T W - W^T T \left(T^T T\right)^{-1} T^T W\right] \left(\bar{\theta}_U - \theta_U^*\right)$$
5. The parameter with the lowest mean value for the bias is ranked first and its corresponding column from S is placed into T for the next steps.
6. With k parameters ranked and k columns in T, formulate p-k candidate parameter subsets, each with k+1 parameters. For each candidate subset, augment T using the corresponding column from S and place the left-over columns in W.
7. For each candidate subset, evaluate the average bias term (considering all 10^4 sets of potential true values) caused by selecting k+1 parameters, but leaving the others at their nominal values using the expression in step 4.
8. The parameter that gives the lowest mean value for the bias (in combination with those that were previously ranked) is ranked next and its corresponding column from S is placed into T.
9. Repeat steps 6-9 until the reduction in bias by adding a new parameter is less than the increase in the total variance caused by including the parameter, which is equal to σ^2.

Most parameter selection techniques are based on the local sensitivity matrix wherein the calculated sensitivity coefficients depend on the initial parameter values used. Repeating these techniques using different initial guesses or scaling factors may lead to different subsets of parameters selected. A high or low ranking for a parameter may be a consequence of poor initial guesses rather than its influence on measured responses or correlations with other parameters. Some modellers have chosen to calculate sensitivity coefficients at new parameter values and repeat the selection technique to avoid this problem (e.g., Brun et al., 2002; Kou et al., 2005 a, b). The method of Chu et al. (2009) should be more robust to poor initial parameter guesses than the

simpler method of Wu et al. (2011a, b) because of the large number of potential parameter values that Chu et al. consider. In Chapter 3, the multi-response batch reactor model of Biegler et al. (1986) is used to compare the effort required for implementing Chu's method, with the effort required to implement Wu's method using multiple sets of potential initial guesses. A shortcoming of both Wu's method and Chu's method is that they consider the MSE for predictions *at the experimental points for parameter estimation*, and they use this anticipated MSE to determine the optimal number of parameters to estimate. Often, the model user is not particularly interested in making predictions the points where data are already available. Instead, accurate predictions may be desired at different sets of operating conditions where the model will be used (i.e., interpolation or extrapolation). It will be helpful if both Wu's and Chu's parameter selection methods can be extended to select the parameters for estimation that will give the best predictions where the modeller needs them.

2.6 Summary and Advice for Modellers

Chemical engineers develop fundamental models following steps shown in Figure 2.1. When developing model equations (Step 1), the modeller must consider all of the important phenomena that need to be described accurately and include the relevant equations and parameters. However, the modeller must also ensure that the model is simple enough for intended users. Initial parameter guesses are required in Steps 2 to 4 and a thorough search of literature and/or industrial information should be conducted to obtain reasonable parameter values, or ranges of values for similar systems. Parameter values that are not already well known must be estimated using a nonlinear optimization algorithm in Step 4. This parameter estimation step often causes the most difficulties for the modeller. Poor initial guesses for parameters can result in optimizer convergence to local optima rather than the best set of parameter values. Other estimation

problems can result from limited or noisy data, parameters with small influence on model predictions, and correlated effects of parameters. Numerical conditioning problems can make estimation of the complete set of parameters impossible. In this situation, the estimability and/or identifiability of model parameters may need to be assessed, model simplification may be undertaken, and additional data collected (where possible).

Before assessing the estimability or identifiability of parameters, even in large complicated models, it is often a good idea to first attempt estimation of all of the parameters, especially when the modeller has a large and well-designed data set. Simultaneous parameter estimation and solution of the model equations, using the method described by Zavala et al. (2008) and Lin et al. (2010) can be helpful for complex systems with large numbers of parameters. Regular sequential parameter estimation, wherein model equations are solved repeatedly for each set of candidate parameters, is effective for smaller models. If physically-reasonable parameter estimates are obtained with narrow confidence intervals, and the model predictions are satisfactory compared with validation data, then no further work is required. It is always a good idea to test the predictive ability of the model (the second part of Step 5 in Figure 2.1), but most modellers do not like to save a portion of their limited data for validation, when they could be using all of the data to obtain improved parameter estimates. Cross-validation techniques are useful in this type of situation (e.g., Thompson et al, 2009).

Unfortunately, in many practical situations, numerical problems, insufficient data and the large number of model parameters make it impossible to reliably estimate of all model parameters in chemical-process models. In these situations, a modeller may be tempted to first try to assess parameter identifiability since identifiability (outlined in the first column of Table 2.1) depends on the model structure and types of measurements and not on the data or experimental settings. If

the model parameters are identifiable, then the modeller can focus attention on obtaining more and better experimental results so that appropriate estimates can be obtained for all of the model parameters. However, an easier and often better choice is to first assess the estimability of the model parameters, given the available and/or proposed experimental data. Many identifiability techniques are complicated and can be computationally intractable when applied to nonlinear fundamental models describing chemical processes with more than about 15 parameters and states. Techniques for determining estimability (the second column in Table 2.1) are easier to use and can readily be applied to models with more than 50 parameters. Estimability should be assessed using one of the techniques that relies on a Fisher Information matrix, since this matrix contains useful information about measurement uncertainties, the magnitude of the influences of model parameters, and correlated effects among parameters. When estimability analysis indicates that some parameters are inestimable, the estimability problem may be due either to the model structure, insufficient experimental data, or an unreliable parameter ranking due to poor initial parameter guesses. Depending on the estimability analysis technique selected, the modeller may elect to repeat the analysis with a few alternative sets of initial parameter guesses. When the modeller has redone the analysis and is convinced that some parameters are inestimable, there are three different choices that can be made: i) the modeller can decide to estimate only a subset of the model parameters while leaving the others fixed at their initial guesses; ii) the modeller can design additional experiments; or iii) the modeller can simplify the model to reduce the number of parameters.

A modeller who decides to fix some parameters and estimate only a subset should use the forward-selection or backward-selection techniques discussed in Section 2.5 to obtain a parameter ranking. The orthogonalization methods of Yao et al. (2003) and Lund and Foss (2008) and the

eigenvalue-based methods of Schittkowski (2007) and Quaiser and Mönnigmann (2009) are easy to use and provide clear parameter rankings. The MSE criterion developed by Wu et al. (2011a, b) can then be used to select the number of parameters to estimate from the ranked list. A benefit of using this MSE criterion is that no arbitrary threshold or cut-off value is needed, and the parameters that are selected for estimation are those that are expected to result in the best model predictions (in terms of mean squared error). Modellers who choose to design additional experiments can augment the sensitivity matrix S to reflect the proposed experiments (see e.g., Thompson et al., 2010). The modeller can then estimate new parameter confidence intervals using the Fisher Information matrix corresponding to the existing and proposed experiments. Modellers interested in simplifying their models can use identifiability and estimability techniques to elucidate problems with the model structure and find combinations of parameters that are difficult to estimate together. Because of the statistical and optimization methods developed over the past decade, modellers of chemical processes now have tools for building effective fundamental models, even when the number of parameters is large and data are limited.

2.7 Nomenclature

c^{tot}	total correlation
d	number of response variables
f	right-hand side of nonlinear ODE model
g	response prediction from nonlinear ODE model
i	index for response variables
j	index for parameters
k	index for number of candidate parameters selected
l	index for parameters
m	index for measurements
n	number of measurements for each response variable
p	number of unknown parameters
q	index for experimental runs
r	number of experimental runs
r_C	critical ratio in MSE-based algorithm
r_{CC}	corrected critical ratio in MSE-based algorithm

r_{CKub} truncated Kubokawa estimator
s_{yi} uncertainty associated with measured response
$s_{\theta j0}$ uncertainty associated with initial parameter guess
t time
u input trajectories
v eigenvectors
x state variables
y response variables
$\hat{y}$ predicted responses using parameter estimates
C parameter estimation covariance matrix
I_F Fisher Information matrix
J objective function in parameter estimation
K_d decay rate coefficient
K_S Michaelis-Menten constant
L subset of parameters that have not been previously selected
N total number of data points
Q square weighting matrix in I_F
R correlation matrix
R_k residual matrix in parameter ranking algorithm
S unscaled sensitivity matrix
T columns from S corresponding to sensitivity vectors for parameters that may be included in the estimation
W columns from S corresponding to sensitivity vectors for parameters that are not included in the estimation
X subset of Z matrix corresponding to selected parameters
Y yield coefficient
Z scaled sensitivity matrix
$\tilde{Z}$ normalized scaled sensitivity matrix

Greek symbols

γ collinearity index
δ^{msqr} sensitivity measure
ε stochastic component
θ unknown parameters
$\bar{\theta}_U$ vector of nominal values for unselected parameters
θ_U^* vector of potential true values for unselected parameters
λ eigenvalues
μ_m maximum specific growth rate
ξ threshold value
ξ_c tuning parameter used in calculation of total correlation
ρ determinant measure
σ^2 noise variance
ϕ angle between sensitivity vectors
Γ^i matrix of eigenvectors

Φ identification function used in implicit-function-theorem approach for identifiability

2.8 References

Atkinson, A. C., A. N. Donev and R. D. Tobias, "Optimum Experimental Designs, with SAS," Oxford University Press (2007), pp. 119-150.

Audoly, S., G. Bellu, L. D'Angiò, M. P. Saccomani and C. Cobelli, "Global Identifiability of Nonlinear Models of Biological Systems," IEEE Trans. Biomed. Eng. **48**(1), 55-65 (2001).

Bellman, R. and K. J. Åström, "On Structural Identifiability," Math. Biosci. **7**(3-4), 329-339 (1970).

Bellu, G., M. P. Saccomani, S. Audoly and L. D'Angio, "DAISY: A New Software Tool to Test Global Identifiability of Biological and Physiological Systems," Comput. Methods Programs Biomed. **88**(1), 52-61 (2007).

Ben-Zvi, A., P. J. McLellan and K. B. McAuley, "Identifiability of Non-Linear Differential Algebraic Systems via a Linearization Approach," Can. J. Chem. Eng. **84**(5), 590-596 (2006).

Ben-Zvi, A., K. McAuley and J. McLellan, "Identifiability Study of a Liquid-Liquid Phase-Transfer Catalyzed Reaction System," AIChE J. **50**(10), 2493-2501 (2004).

Biegler, L. T., J. J. Damiano and G. E. Blau, "Nonlinear Parameter Estimation: A Case Study Comparison," AIChE J. **32**(1), 29-45 (1986).

Brockmann, D., K.-H. Rosenwinkel and E. Morgenroth, "Practical Identifiability of Biokinetic Parameters of a Model Describing Two-Step Nitrification in Biofilms," Biotechnol. Bioeng. **101**(3), 497-514 (2008).

Brun, R., P. Reichert and H. R. Künsch, "Practical Identifiability Analysis of Large Environmental Simulation Models," Water Resour. Res. **37**(4), 1015-1030 (2001).

Brun, R., M. Kühni, H. Siegrist, W. Gujer and P. Reichert, "Practical Identifiability of ASM2d Parameters – Systematic Selection and Tuning of Parameter Subsets," Water Res. **36**(16), 4113-4127 (2002).

Chappell, M. J. and K. R. Godfrey, "Structural Identifiability of the Parameters of a Nonlinear Batch Reactor Model," Math. Biosci. **108**(2), 241-251 (1992).

Chu, Y. and J. Hahn, "Parameter Set Selection via Clustering of Parameters into Pairwise Indistinguishable Groups of Parameters," Ind. Eng. Chem. Res. **48**(13), 6000-6009 (2009).

Chu, Y., Z. Huang and J. Hahn, "Improving Prediction Capabilities of Complex Dynamic Models via Parameter Selection and Estimation," Chem. Eng. Sci. **64**(19), 4178-4185 (2009).

Daun, S., J. Rubin, Y. Vodovotz, A. Roy, R. Parker and G. Clermont, "An Ensemble of Models of the Acute Inflammatory Response to Bacterial Lipopolysaccharide in Rats: Results from Parameter Space Reduction," J. Theor. Biol. **253**(4), 843-853 (2008).

De Pauw, D. J. W., "Optimal Experimental Design for Calibration of Bioprocess Models: a Validated Software Toolbox," Ph.D. Thesis, University of Ghent (2005), pp. 1-30.

De Pauw, D. J. W., K. Steppe and B. De Baets, "Identifiability Analysis and Improvement of a Tree Water Flow and Storage Model," Math. Biosci. **211**(2), 314-332 (2008).

Degenring, D., C. Froemel, G. Dikta and R. Takors, "Sensitivity Analysis for the Reduction of Complex Metabolism Models," J. Process Control. **14**(7), 729-745 (2004).

Díez, M. D., M. Fjeld, E. Andersen and B. Lie, "Validation of a Compartmental Population Balance Model of an Industrial Leaching Process: The Silgrain® Process," Chem. Eng. Sci. **61**(1), 229-245 (2006).

Dochain, D., P. A. Vanrolleghem, "Dynamical Modelling and Estimation in Wastewater Treatment Processes," IWA Publishing, London, UK (2001), pp. 157-165.

Dochain, D., P. A. Vanrolleghem and M. Van Daele, "Structural Identifiability of Biokinetic Models of Activated Sludge Respiration," Water Res. **29**(11), 2571-2578 (1995).

Duchesne, C. and J. F. MacGregor, "Jackknife and Bootstrap Methods in the Identification of Dynamic Models," J. Process Control. **11**(5), 553-564 (2001).

Efron, B. "The Jackknife, the Bootstrap and Other Resampling Plans," Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA (1982), pp. 5-58.

Foss, B. A., B. Lohmann and W. Marquardt, "A Field Study of the Industrial Modeling Process," J. Process Control. **8**(5-6), 325-338 (1998).

Freni, G., G. Mannina and G. Viviani, "Identifiability Analysis for Receiving Water Body Quality Modelling," Environ. Modell. Softw. **24**(1), 54-62 (2009).

Gadkar, K. G., R. Gunawan and F. J. Doyle III, "Iterative Approach to Model Identification of Biological Networks," BMC Bioinformatics. **6**(155), (2005).

Grewal, M. S. and K. Glover, "Identifiability of Linear and Nonlinear Dynamical Systems," IEEE Trans. Autom. Control. **21**(6), 833-836 (1976).

Hocking, R. R., "Analysis and Selection of Variables in Linear Regression," Biometrics **32**(1), 1-49 (1976).

Holmberg, A., "On the Practical Identifiability of Microbial Growth Models Incorporating Michaelis-Menten Type Nonlinearities," Math. Biosci. **62**(1), 23-43 (1982).

Huang, Z., Y. Chu and J. Hahn, "Model Simplification Procedure for Signal Transduction Pathway Models: An Application to IL-6 Signaling," Chem. Eng. Sci. **65**(6), 1964-1975 (2010).

Jacquez, J. A. and P. Greif, "Numerical Parameter Identifiability and Estimability: Integrating Identifiability, Estimability, and Optimal Sampling Design," Math. Biosci. **77**(1-2), 201-227 (1985).

Jayasankar, B. R., A. Ben-Zvi and B. Huang, "Identifiability and Estimability Study for a Dynamic Solid Oxide Fuel Cell Model," Comput. Chem. Eng. **33**(2), 484-492 (2009).

Jiménez-Hornero, J. E., I. M. Santos-Dueñas and I. García-García, "Optimization of Biotechnological Processes. The Acetic Acid Fermentation. Part II: Practical Identifiability Analysis and Parameter Estimation," Biochem. Eng. J. **45**(1), 7-21 (2009).

Jiménez-Hornero, J. E., I. M. Santos-Dueñas and I. García-García, "Structural Identifiability of a Model for the Acetic Acid Fermentation Process," Math. Biosci. **216**(2), 154-162 (2008).

Jitjareonchai, J. J., P. M. Reilly, T. A. Duever and D. B. Chambers, "Parameter Estimation in the Error-in-Variables Models Using the Gibbs Sampler," Can. J. Chem. Eng. **84**(1), 125-138 (2006).

Jolliffe, I. T., "Discarding Variables in a Principal Component Analysis .1: Artificial Data," J. Roy. Statist. Soc. Ser. C Appl. Statist. **21**(2), 160-172 (1972).

Julien, S., J. P. Babary and P. Lessard, "Theoretical and Practical Identifiability of a Reduced Order Model in an Activated Sludge Process Doing Nitrification and Denitrification," Water Sci. Technol. **37**(12), 309-316 (1998).

Karimi, H., M. A. Schaffer and K. B. McAuley, "A Kinetic Model for Non-Oxidative Thermal Degradation of Nylon 66," submitted to Macromolecular Reaction Engineering, April (2011).

Kim, Y. S., M. H. Kim and C. K. Yoo, "A New Statistical Framework for Parameter Subset Selection and Optimal Parameter Estimation in the Activated Sludge Model," J. Hazard. Mater. **183**(1-3), 441-447 (2010).

Koeva, V. I., S. Daneshvar, R. J. Senden, A. H. M. Imam, L. J. Schreiner and K. B. McAuley, "Mathematical Modeling of PAG- and NIPAM-Based Polymer Gel Dosimeters Contaminated by Oxygen and Inhibitor," Macromol. Theory Simul. **18**(9), 495-510 (2009).

Kou, B., K. B. McAuley, J. C. C. Hsu and D. W. Bacon, "Mathematical Model and Parameter Estimation for Gas-Phase Ethylene/Hexene Copolymerization with Metallocene Catalyst," Macromol. Mater. Eng. **290**(6), 537-557 (2005a).

Kou, B., K. B. McAuley, C. C. Hsu, D. W. Bacon and K. Z. Yao, "Mathematical Model and Parameter Estimation for Gas-Phase Ethylene Homopolymerization with Supported Metallocene Catalyst," Ind. Eng. Chem. Res. **44**(8), 2428-2442 (2005b).

Leis, J. R. and M. A. Kramer, "The Simultaneous Solution and Sensitivity Analysis of Systems Described by Ordinary Differential Equations," ACM Trans. Math. Softw. **14**(1), 45-60 (1988).

Lin, W., L. T. Biegler and A. M. Jacobson, "Modeling and Optimization of a Seeded Suspension Polymerization Process," Chem. Eng. Sci. **65**(15), 4350-4362 (2010).

Littlejohns, J. V., K. B. McAuley and A. J. Daugulis, "Model for a Solid-Liquid Stirred Tank Two-Phase Partitioning Bioscrubber for the Treatment of BTEX," J. Hazard. Mater. **175**(1-3), 872-882 (2010).

Ljung, L, "System Identification: Theory for the User," Prentice-Hall, Englewood Cliffs, NJ (1999), pp. 212-222.

Ljung, L. and T. Glad, "On Global Identifiability for Arbitrary Model Parametrizations," Automatica. **30**(2), 265-276 (1994).

Lund, B. F. and B .A. Foss, "Parameter Ranking by Orthogonalization – Applied to Nonlinear Mechanistic Models," Automatica. **44**(1), 278-281 (2008).

Machado, V. C., G. Tapia, D. Gabriel, J. Lafuente and J. A. Baeza, "Systematic Identifiability Study Based on the Fisher Information Matrix for Reducing the Number of Parameters Calibration of an Activated Sludge Model," Environ. Modell. Softw. **24**(11), 1274-1284 (2009).

Maria, G., "A Review of Algorithms and Trends in Kinetic Model Identification for Chemical and Biochemical Systems," Chem. Biochem. Eng. Q. **18**(3), 195-222 (2004).

Maria, G. and D. W. T. Rippin, "Note Concerning Two Techniques for Complex Kinetic Pathway Analysis," Chem. Eng. Sci. **48**(22), 3855-3864 (1993).

Miao, H., X. Xia, A. S. Perelson and H. Wu, "On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics," SIAM Rev. **53**(1), 3-39 (2011).

Munack, A., "Optimization of Sampling," in "Biotechnology: Vol 4. Measuring, Modelling and Control," K. Schügerl, Ed., VCH, Weinheim (1991), pp. 251-264.

Nikerel, I. E., W. A. van Winden, P. J. T. Verheijen and J. J. Heijnen, "Model Reduction and *a priori* Kinetic Parameter Identifiability Analysis Using Metabolome Time Series For Metabolic Reaction Networks with Linlog Kinetics," Metab. Eng. **11**(1), 20-30 (2009).

Petersen, B., K. Gernaey and P. A. Vanrolleghem, "Practical Identifiability of Model Parameters by Combined Respirometric-Titrimetric Measurements," Water Sci. Technol. **43**(7), 347-355 (2001).

Petersen, B., K. Gernaey, M. Devisscher, D. Dochain and P. A. Vanrolleghem, "A Simplified Method to Assess Structurally Identifiable Parameters in Monod-Based Activated Sludge Models," Water Res. **37**(12), 2893-2904 (2003).

Pohjanpalo, H., "System Identifiability Based on Power Series Expansion of Solution," Math. Biosci. **41**(1-2), 21-33 (1978).

Quaiser, T. and M. Mönnigmann, "Systematic Identifiability Testing for Unambiguous Mechanistic Modeling – Application to JAK-STAT, MAP Kinase, and NF-κB Signaling Pathway Models," BMC Systems Biology. **3**(50), (2009).

Raue, A., C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller and J. Timmer, "Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood," Bioinformatics. **25**(15), 1923-1929 (2009).

Reichert, P. and P. Vanrolleghem, "Identifiability and Uncertainty Analysis of the River Water Quality Model No. 1 (RWQM1)," Water Sci. Technol. **43**(7), 329-338 (2001).

Reichert, P., R. von Schulthess and D. Wild, "The Use of Aquasim for Estimating Parameters of Activated Sludge Models," Water Sci. Technol. **31**(2), 135-147 (1995).

Rodriguez-Fernandez, M., J. A. Egea and J. R. Banga, "Novel Metaheuristic for Parameter Estimation in Nonlinear Dynamic Biological Systems," BMC Bioinformatics. **7**(483), (2006).

Saltelli, A., K. Chan and E. M. Scott, "Sensitivity Analysis," John Wiley & Sons, NY (2000), pp. 15-50.

Schittkowski, K., "Experimental Design Tools for Ordinary and Algebraic Differential Equations," Ind. Eng. Chem. Res. **46**(26), 9137-9147 (2007).

Seber, G. A. F. and C. J. Wild, "Nonlinear Regression," John Wiley & Sons, NY (2003), pp. 529-531.

Sedoglavic, A., "A Probabilistic Algorithm to Test Local Algebraic Observability in Polynomial Time," J. Symb. Comput. **33**(5), 735-755 (2002).

Sidoli, F. R., A. Mantalaris and S. P. Asprey, "Toward Global Parametric Estimability of a Large-Scale Kinetic Single-Cell Model for Mammalian Cell Cultures," Ind. Eng. Chem. Res. **44**(4), 868-878 (2005).

Sidoli, F. R., S. P. Asprey and A. Mantalaris, "A Coupled Single Cell-Population-Balance Model for Mammalian Cell Cultures," Ind. Eng. Chem. Res. **45**(16), 5801-5811 (2006).

Srinath, S. and R. Gunawan, "Parameter Identifiability of Power-Law Biochemical System Models," J. Biotechnol. **149**(3), 132-140 (2010).

Thompson, D. E., K. B. McAuley and P. J. McLellan, "Design of Optimal Sequential Experiments to Improve Model Predictions from a Polyethylene Molecular Weight Distribution Model," Macromol. React. Eng. **4**(1), 73-85 (2010).

Thompson, D. E., K. B. McAuley and P. J. McLellan, "Parameter Estimation in a Simplified MWD Model for HDPE Produced by a Ziegler-Natta Catalyst," Macromol. React. Eng. **3**(4), 160-177 (2009).

Vajda, S. and H. Rabitz, "State Isomorphism Approach to Global Identifiability of Nonlinear Systems," IEEE Trans. Autom. Control. **34**(2), 220-223 (1989).

Vajda, S., K. R. Godfrey and H. Rabitz, "Similarity Transformation Approach to Identifiability Analysis of Nonlinear Compartmental-Models," Math. Biosci. **93**(2), 217-248 (1989a).

Vajda, S., H. Rabitz, E. Walter and Y. Lecourtier, "Qualitative and Quantitative Identifiability Analysis of Nonlinear Chemical Kinetic-Models," Chem. Eng. Commun. **83**(1), 191-219 (1989b).

Vanrolleghem, P. A., M. Van Daele and D. Dochain, "Practical Identifiability of a Biokinetic Model of Activated Sludge Respiration," Water Res. **29**(11), 2561-2570 (1995).

Walter, E., "Identifiability of State Space Models," Springer-Verlag, Berlin, Germany (1982), pp. 41-95.

Walter, E. and L. Pronzato, "On the Identifiability and Distinguishability of Nonlinear Parametric Models," Math. Comput. Simul. **42**(2-3), 125-134 (1996).

Walter, E. and Y. Lecourtier, "Global Approaches to Identifiability Testing for Linear and Nonlinear State Space Models," Math. Comput. Simul. **24**(6), 472-482 (1982).

Weijers, S. R. and P. A. Vanrolleghem, "A Procedure for Selecting Best Identifiable Parameters in Calibrating Activated Sludge Model No.1 to Full-Scale Plant Data," Water Sci. Technol. **36**(5), 69-79 (1997).

Weijers, S. R., J. J. Kok, H. A. Preisig, A. Buunen and T. W. M. Wouda, "Parameter Identifiability in the IAWQ Model No. 1 for Modelling Activated Sludge Plants for Enhanced Nitrogen Removal," Comput. Chem. Eng. **20**, S1455-S1460 (1996).

Woloszyn, J. D. and K. B. McAuley, "Application of Parameter Selection and Estimation Techniques in a Thermal Styrene Polymerization Model," Macromol. React. Eng. (2011). doi: 10.1002/mren.201100021

Wu, H., H. Zhu, H. Miao and A. S. Perelson, "Parameter Identifiability and Estimation of HIV/AIDS Dynamic Models," Bull. Math. Biol. **70**(3), 785-799 (2008).

Wu, S., T. J. Harris and K. B. McAuley, "The Use of Simplified or Misspecified Models: Linear Case," Can. J. Chem. Eng. **85**(4), 386-398 (2007).

Wu, S., K. B. McAuley and T. J. Harris, "Selection of Simplified Models: II. Development of a Model Selection Criterion Based on Mean Squared Error," Can. J. Chem. Eng. **89**(2), 325-336 (2011a).

Wu, S., K. A. P. McLean, T. J. Harris and K. B. McAuley, "Selection of Optimal Parameter Set Using Estimability Analysis and MSE-Based Model-Selection Criterion," Int. J. Advanced Mechatronic Systems, **3**(3), 188-197 (2011b).

Xia, X. and C. H. Moog, "Identifiability of Nonlinear Systems with Application to HIV/AIDS Models," Automatic Control, IEEE Transactions on. **48**, 330-336 (2003).

Yao, K. Z., B. M. Shaw, B. Kou, K. B. McAuley and D. W. Bacon, "Modeling Ethylene/Butene Copolymerization with Multi-Site Catalysts: Parameter Estimability and Experimental Design," Polym. React. Eng. **11**(3), 563-588 (2003).

Yue, H., M. Brown, J. Knowles, H. Wang, D. S. Broomhead and D. B. Kell, "Insights into the Behaviour of Systems Biology Models from Dynamic Sensitivty and Identifiability Analysis: a Case Study of an NF-κB Signalling Pathway," Mol. Biosyst. **2**(12), 640-649 (2006).

Zavala, V. M., C. D. Laird and L. T. Biegler, "Interior-Point Decomposition Approaches for Parallel Solution of Large-Scale Nonlinear Parameter Estimation Problems," Chem. Eng. Sci. **63**(19), 4834-4845 (2008).

Zhang, S. and P.M. Huck, "Parameter Estimation for Biofilm Processes in Biological Water Treatment," Water Res. **30**(2), 456-464 (1996).

Zhang, T., D. Zhang, Z. Li and Q. Cai, "Evaluating the Structural Identifiability of the Parameters of the EBPR Sub-Model in ASM2d by the Differential Algebra Method," Water Res. **44**(9), 2815-2822 (2010).

Chapter 3

Robust Mean-Squared Error-based Methods for Selecting Optimal Parameter Subsets for Estimation[2]

3.1 Summary

Engineers who develop fundamental models for chemical processes are often unable to estimate all of the model parameters, especially when the available data are limited or noisy. In these situations, modellers may decide to select only a subset of the model parameters for estimation. An orthogonalization algorithm combined with a recent mean squared error (MSE)-based criterion has been used successfully to rank parameters from most to least estimable and to determine the parameter subset that should be estimated to obtain the best predictions. A new ranking and selection technique is developed based on this MSE criterion. Unfortunately, results from these parameter selection techniques are sensitive to the initial parameter values and the uncertainty factors used to calculate sensitivity coefficients. A robustness test is proposed and applied to a batch reactor model to assess the sensitivity of the selected parameter subset to the initial parameter guesses. The selected parameter subsets are compared with those selected using another MSE-based method proposed by Chu et al. (2009). The two methods based on the MSE criterion selected fewer parameters for estimation, but require longer computation times due to the parameter estimation steps.

[2] The work in this Chapter is in preparation for submission to a journal article. K.A.P. McLean, S. Wu., and K.B. McAuley are co-authors of this journal article. Note that this thesis has been prepared using a manuscript format, so some nomenclature used is not consistent throughout the entire thesis. Please refer to Section 3.6 for the nomenclature used in this Chapter.

3.2 Introduction

Chemical engineers often build fundamental dynamic models to represent the behaviour of chemical phenomena in industrial processes. These nonlinear models frequently contain many parameters and when the values of some parameters are not well known, they must be estimated using nonlinear optimization algorithms (Maria, 2004). It may be impossible for the modeller to estimate all of the model parameters due to limited or noisy data (e.g., Vajda et al., 1989; Raue et al., 2009; McLean and McAuley, 2011). In addition, the effect of some parameters on model predictions may be negligible, or the effect of some parameters may be correlated with the effect of others. Consequently, a modeller may instead estimate only a subset of the model parameters (e.g., Kou et al., 2005a, b; Chu et al., 2009; Thompson et al., 2009; Wu et al., 2011a, b). In this situation, the modeller needs effective tools to aid in determining the appropriate parameter subset to estimate.

Methods for selecting parameter subsets for estimation have been reviewed recently (McLean and McAuley, 2011) and include: methods based on scalar measures of the Fisher Information Matrix (Weijers and Vanrolleghem, 1997; Freni et al., 2009; Machado et al., 2009), correlation and collinearity index methods (Brun et al., 2001, 2002; Chu and Hahn, 2009), methods based on principal component analysis (PCA) and eigenvalues (Vajda et al., 1989; Degenring et al., 2004; Schittkowski, 2007; Quaiser and Mönnigmann, 2009), and orthogonalization methods (Yao et al., 2003; Lund and Foss, 2008; Thompson et al., 2009). Many researchers have successfully used the orthogonalization method to rank parameters from most to least estimable and various techniques have been employed to select the number of parameters to estimate from the ranked list (e.g. using an arbitrary threshold value (Yao et al., 2003; Kou et al., 2005a, b; Jayasankar et al., 2009), judgment based on objective function (Littlejohns, 2010), or cross-validation (Thompson et al.,

2009)). A statistical method proposed by Wu et al. (2007; 2011a, b) relies on mean squared error (MSE), which takes into account both the bias and variance in the parameter estimates and model predictions. This bias term increases when a modeller chooses to use a simplified model (SM) wherein only a subset of parameters is estimated. However, a SM with fewer parameters also results in a decrease in the variance for the parameter estimates and model predictions. When the variance reduction is greater than the increase in bias, the SM gives better predictions, in terms of MSE, than the original model with all parameters estimated (Hocking, 1976; Wu et al., 2007). Wu's MSE-based method has been applied to select optimal parameter subsets in a styrene polymerization model (Woloszyn and McAuley, 2011) and a nylon 66 degradation model (Karimi et al., 2011).

One limitation of Wu's MSE-based method and other methods that are based on a local sensitivity matrix is their dependence on initial parameter values. Poor initial guesses will lead to poor derivative information contained in the sensitivity coefficients. Using different parameter values or uncertainty factors can result in a different parameter ranking and a different number of parameters selected. A method for assessing the sensitivity of the selected parameters to the initial guesses is proposed in this Chapter, wherein Wu's MSE-based method is repeated using different initial guesses. Another MSE-based technique proposed by Chu et al. (2009) employs a Monte Carlo technique that considers a large number of potential parameter values and, as such, should be more robust to poor initial guesses. However, the Monte Carlo approach is computationally expensive and is prohibitive for models with a large number of parameters. The results obtained and the computational effort required for implementing these two robust parameter-selection techniques will be compared. The method of Wu et al. relies on an orthogonalization algorithm to first rank the parameters (Yao et al., 2003; Thompson et al., 2009).

In a second step, the appropriate number of parameters is determined using a MSE-based model selection criterion (Wu et al., 2007; 2011a, b). A new selection technique will also be proposed, which uses Wu's MSE-based criterion to rank the parameters as well as select the optimal number of parameters to estimate.

Section 3.3 presents the proposed robust MSE-based parameter selection procedures that account for uncertainty in the initial parameter guesses. These techniques are compared with the recent MSE-based Monte Carlo approach by Chu et al. (2009). In Section 3.4 a multi-response batch reactor model and literature data (Biegler et al., 1986) are used to demonstrate the three different parameter selection techniques and the results and efficiencies are compared.

3.3 Methods

Assume that the behaviour of the process of interest can be described by a multivariate nonlinear dynamic model of the form:

$$\frac{dx}{dt} = f(x,u,\theta), \quad x(0) = x_0 \tag{3.1}$$

$$y = g(x,u,\theta) + \varepsilon \tag{3.2}$$

where x is the vector of model states with the vector of initial values x_0, u is the vector of input trajectories, θ is vector of model parameters, y is the vector of measured response variables, and ε is the vector of stochastic measurement errors. Note that θ can appear explicitly in the function g and implicitly in g through $x(t)$. With several different response variables and experimental runs performed, there may be $N=dnr$ total data points available for estimation, where d is the number of response variables, and n is the number of measurements taken for each response variable for each of the r experimental runs.

Information about influences of parameter values on model predictions is contained in a sensitivity matrix, *Z*, which is important for many parameter selection techniques. The sensitivity matrix has dimensions $N \times p$, where p is the total number of model parameters. The elements of this matrix, which are scaled for dimensional consistency, are partial derivatives of each response variable prediction g_{ilm} with respect to each parameter θ_j and are calculated as follows:

$$\frac{\partial g_{ilm}}{\partial \theta_j} \frac{s_{\theta j 0}}{s_{yi}} \tag{3.3}$$

These scaled sensitivity coefficients are calculated for each of the i=1…d response variables, each of the l=1…n measurement times, and each of the m=1…r experimental runs. The scaling factors $s_{\theta j0}$ and s_{yi} account for uncertainties in the initial guess for parameter θ_j and the ith measured response y_i (Thompson et al., 2009). The initial parameter guesses and their corresponding uncertainty levels can be set by the modeller based on his or her prior knowledge (e.g., information from literature, earlier data or parameter estimation analyses, or knowledge of physically realistic parameter values (Bates and Watts, 1988)) and the different s_{yi} values account for inaccuracies of different measurements and may be determined from replicate experiments or information from sensor suppliers. The elements of this scaled sensitivity matrix, *Z*, can be calculated by difference approximations with perturbed parameter values (Saltelli et al., 2000) or by solving sensitivity equations (Leis and Kramer, 1988).

The widely-used orthogonalization algorithm, which is outlined in Table 3.1, relies on this scaled sensitivity matrix to rank parameters from most to least estimable. The rank of each model parameter is determined based the influence of the parameter on model predictions and on correlations with other parameters. A complete description of the algorithm is provided by Yao et al. (2003) and Thompson et al. (2009). Wu et al. (2007; 2011a, b) subsequently developed a

MSE-based technique to determine the optimal number of parameters to estimate from the ranked list obtained using the orthogonalization algorithm. This technique, shown in Table 3.2, relies on a critical ratio, r_C. This statistic estimates the ratio of the squared bias to the variance that results from estimating only k model parameters, rather than estimating all p parameters. r_C is computed using optimal values of a weighted least-squares objective function J:

$$J = \sum_{i=1}^{d} \sum_{l=1}^{n} \sum_{m=1}^{r} \left(\frac{y_{ilm} - g_{ilm}}{s_{yi}} \right)^2 \tag{3.4}$$

where y_{ilm} is the measured value of the ith response variable at the lth measurement time, for the mth experimental run, and g_{ilm} is the corresponding predicted value. In Table 3.2, J_k and J_p are values of the objective function when k and p parameters are estimated, respectively. In cases where it is impossible to estimate all p parameters simultaneously due to numerical difficulties, J_p can be approximated by estimating a sufficiently large number of parameters. A corrected critical ratio, r_{CC}, is calculated in step four of the algorithm to allow for comparisons among simplified models with different numbers of parameters. An $r_{CC,k}$ value that is less than zero indicates that the SM with k parameters estimated is expected to provide better predictions, in terms of MSE, than the model with all parameters estimated. The value of k corresponding to the lowest $r_{CC,k}$ value is the number of parameters that should be selected for estimation.

Table 3.1: Orthogonalization Algorithm (Wu et al., 2011b)

1. Calculate the magnitude (i.e., the Euclidean norm) of each column in the scaled sensitivity matrix Z. The most estimable parameter corresponds to the column in Z with the largest magnitude. Set k=1.
2. Put the k columns from Z that correspond to parameters that have been ranked into matrix X_k.
3. Use X_k to predict columns in Z using ordinary least-squares: $$\hat{Z}_k = X_k \left(X_k^T X_k \right)^{-1} X_k^T Z \quad (3.5)$$ and calculate the residual matrix $$R_k = Z - \hat{Z}_k \quad (3.6)$$
4. Calculate the magnitude of each column in R_k. The $(k+1)^{th}$-most estimable parameter corresponds to the column in R_k with the largest magnitude.
5. Increase k by 1, and put the columns corresponding to the k+1 parameters that have been ranked in matrix X_k
6. Advance the iteration counter (subscripts for X and R) and repeat Steps 3 to 5, until all parameters are ranked or until it is impossible to perform the least-squares prediction of Z in Step 3 due to matrix singularity.

Table 3.2: Wu's MSE-based approach for selecting the number of parameters to estimate (Wu et al., 2011b)

1. Start with a ranked parameter list (i.e., use the orthogonalization algorithm)
2. Estimate the top-ranked parameter using weighted least-squares regression, with all other parameters fixed at initial guesses. Next, estimate the top two parameters, followed by the top three parameters and so on, until all of the ranked parameters have been estimated. Denote the value of the objective function with the top k parameters estimated and the remaining p-k parameters held fixed as $J_{k..}$
3. Compute the critical ratio $$r_{C,k} = \left(J_k - J_p \right) / \left(p - k \right) \quad (3.7)$$ for k=1...p-1.
4. For each value of k, compute the corrected critical ratio $$r_{CC,k} = \frac{p-k}{N} \left(r_{CKub,k} - 1 \right) \quad (3.8)$$ where $$r_{CKub,k} = \max\left(r_{C,k} - 1, \frac{2}{p-k+2} r_{C,k} \right) \quad (3.9)$$
5. Select the value of k corresponding to the lowest value of $r_{CC,k}$ as the appropriate number of parameters to estimate.

Note that the initial parameter ranking using the orthogonalization algorithm and the subsequent subset selection algorithm in Table 3.2 depend on the initial parameter guesses and the scaling factors used to construct the *Z* matrix. The robustness of the selected parameter subset to these assumptions can be analyzed by repeating the ranking and parameter selection using different realistic initial parameter guesses or scaling factors. Later in this Chapter, a modeling example is analyzed using 100 sets of different initial parameter guesses, chosen randomly from reasonable uncertainty ranges, which are used to build 100 different *Z* matrices. The orthogonalization algorithm for parameter ranking is repeated for each set of initial guesses. For each of the 100 resulting ranked parameter lists, Wu's MSE-based method is applied to select the appropriate number of parameters for estimation. This technique is used to determine if the same subset of parameters is frequently selected for estimation or whether the initial parameter guesses and scaling factors led to anomalous results.

A second robust MSE-based technique proposed by Chu et al. (2009) relies on Monte Carlo simulations to simultaneously rank parameters and select the optimal number to estimate. The algorithm for this technique is shown below in Table 3.3. Note that unlike the previous method, this technique does not involve parameter estimation and does not require experimental data. In this method, many different sets of potential true parameter values are considered, and sensitivity matrices are formulated for each candidate set. The potential true parameter values are selected from reasonable uncertainty ranges, for each parameter, which must be specified by the user in step one of the algorithm. By using many different sets of candidate parameter values to calculate an average bias, the selected parameter subset should be less sensitive to poor initial guesses than the method of Wu et al (2007; 2011a, b). Unscaled sensitivity matrices, *S*, with elements $\partial g_{ilm}/\partial \theta_j$, are calculated 10^4 times and each *S* matrix is partitioned into two parts: *T* and *W*. *T* is the matrix

of sensitivity vectors corresponding to parameters that may be estimated and *W* is the matrix of sensitivity vectors for parameters that will remain fixed at nominal values. Steps four and seven involve calculating the average bias introduced by considering only a subset of parameters for estimation. $\bar{\theta}_U$ in the expression for bias denotes the vector of nominal values for the unselected parameters and θ_U^* denotes the vector of possible true values for the unselected parameters.

Table 3.3: MSE-based approach developed by Chu et al. (2009) to rank and select the number of parameters to estimate

1.	Specify initial guesses (nominal values) and parameter ranges and distributions for all *p* parameters, as well as the variance σ^2 for the response variable.
2.	Randomly select 10^4 possible sets of true parameter values from the specified parameter ranges. For each set of candidate parameter values, evaluate and record the sensitivity matrix *S*.
3.	Starting with no parameters in the ranked list, formulate *p* candidate parameter subsets, each with 1 parameter that will be estimated, so that *p*-1 parameters will be left out in each case. For each selected parameter (i.e., for *j* = 1, ... , *p*), place its corresponding column from *S* into *T* and place the remaining columns in *W*
4.	For *j* = 1, ..., *p*, evaluate the bias for each of the 10^4 potential sets of true values, caused by selecting parameter *j*, but leaving the others at their nominal values: $$\left(\bar{\theta}_U - \theta_U^*\right)^T \left[W^T W - W^T T \left(T^T T\right)^{-1} T^T W\right] \left(\bar{\theta}_U - \theta_U^*\right)$$ For each candidate parameter subset, compute the average of these 10^4 bias values.
5.	The parameter with the lowest mean value for the bias is ranked most estimable.
6.	With *k-1* parameters ranked, formulate *p-k-1* candidate parameter subsets, each with *k* parameters. Each of these subsets will include the *k*-1 previously selected parameters and one additional parameter. For each candidate subset, include the corresponding columns in *T* and put columns for the unselected parameters in *W*.
7.	For each candidate subset, evaluate the bias for the 10^4 potential sets of true values, caused by selecting *k* parameters. This bias is obtained using the expression shown in step 4 and the average of the 10^4 bias values is computed for each candidate parameter subset.
8.	The selected parameter that gives the lowest mean value for the bias (in combination with the parameters that were previously ranked) is ranked next.
9.	Repeat steps 6-8 until the reduction in average bias resulting from the addition of the new parameter is less than the increase in the total variance caused by including the parameter, which is equal to σ^2.

Similar to the method of Wu et al. shown in Table 3.1 and Table 3.2, Chu's method is a forward selection technique (i.e., parameters are ranked one at a time from most estimable to least estimable) that considers the trade-off between bias and variance as additional parameters are included in the estimation. However, with Wu's method, the ranking of parameters and the selection of a parameter subset are performed in separate procedures, whereas in Chu's method, the ranking and selection of parameters are combined. For typical chemical process models, it is not clear which method will give better results, and which will be more computationally expensive.

An additional new parameter selection technique is proposed in this Chapter, wherein Wu's MSE-based method is modified to rank parameters by minimizing MSE at each ranking step, similar to the method of Chu et al. However, unlike the previous two methods, this new method does not involve sensitivity matrices. Rather, this method involves parameter estimation at each ranking step to compute r_{CC}. The next parameter selected at each step corresponds to the parameter that gives the lowest value of $r_{CC,k}$. To rank all the parameters from most estimable to least estimable using this technique requires $p(p+1)/2-1$ parameter estimations. The algorithm for this procedure is outlined in Table 3.4.

Table 3.4: New method to rank and select the number of parameters to estimate

1. Starting with no parameters in the ranked list, formulate p candidate parameter subsets, each with 1 parameter that will be estimated, so that p-1 parameters will be held at their initial values.
2. For each selected parameter (i.e., for j = 1, ..., p) fit the candidate model using the weighted nonlinear least squares criterion in Equation (3.4) to compute an optimal objective function value, which is then used to calculate p different $r_{C,1}$ and $r_{CC,1}$ values using Equations (3.7) – (3.9) from Table 3.2.
3. The parameter corresponding to the lowest $r_{CC,1}$ value is ranked as the most estimable parameter.
4. With k-1 parameters ranked, formulate p-k-1 candidate parameter subsets. Each subset will contain the k-1 parameters that were previously ranked and one additional parameter.
5. For each candidate subset, fit the candidate model using weighted nonlinear least squares to compute an optimal objective function value, which is then used to calculate $r_{C,k}$ and $r_{CC,k}$ values using Equations (3.7) – (3.9) from Table 3.2.
6. The additional parameter that gives the lowest $r_{CC,k}$ value (in combination with the parameters that were previously ranked) is ranked next.
7. Repeat steps 4-6 until all the parameters are ranked, or until estimation fails due to numerical conditioning problems. Select the parameter subset corresponding to the lowest overall value of $r_{CC,k}$ as the subset that should be included in the final estimation.

Like Chu's method, this new forward selection technique combines parameter ranking and subset selection by considering the trade-off between bias and variance at each step. Unlike Chu's method, the new technique uses parameter estimation and data. Chu's method and the orthogonalization algorithm rank parameters based on sensitivity matrices and thus require only the settings at which the experimental data will be collected (e.g., measurement times, reactor temperature, and other input conditions). Like Wu's original MSE-based algorithm, this new algorithm determines the appropriate number of parameters to estimate based on the lowest value of $r_{CC,k}$. Unlike Wu's method, the new technique does not rely on sensitivity coefficients for parameter ranking; the user specifies information about the uncertainty in parameter values by enforcing reasonable bounds in the parameter estimation step.

For chemical process models, it is not clear whether the three methods outlined above will give similar results and which method will be best. In the next section, the three methods are applied to select parameter subsets for estimation in a batch reactor model.

3.4 Example: Parameter Selection and Estimation for the Dow Chemical Model

The Dow Chemical batch reactor model described by Biegler et al. (1986) consists of the differential and algebraic equations shown in Table 3.5. In these equations, $[Q^+]$ is the initial catalyst concentration and T_0 is a reference temperature (340.15 K). The four response variables in the model include three measured concentrations and a fourth computed concentration:

$$\begin{aligned} y_1 &= x_1 + \varepsilon_1 \\ y_2 &= x_2 + \varepsilon_2 \\ y_3 &= x_3 + \varepsilon_3 \\ y_4 &= y_1^0 - y_1 - y_3 \end{aligned} \tag{3.10}$$

where y_1^0 is the initial concentration of y_1. Data and initial conditions for these response variables are provided at 40, 67, and 100 °C, with some y_1 measurements missing for the experimental run at 40 °C (Biegler et al., 1986). The uncertainty associated for each measured concentration is assumed to be s_{yi} = 0.02 mol/kg.

Table 3.5: Batch reactor model equations (Biegler et al., 1986)

$$\frac{dx_1}{dt} = -k_2 x_2 x_8$$

$$\frac{dx_2}{dt} = -k_1 x_2 x_6 + k_{-1} x_{10} - k_2 x_2 x_8$$

$$\frac{dx_3}{dt} = k_2 x_2 x_8 + k_1 x_4 x_6 - 0.5 k_{-1} x_9$$

$$\frac{dx_4}{dt} = -k_1 x_4 x_6 + 0.5 k_{-1} x_9$$

$$\frac{dx_5}{dt} = k_1 x_2 x_6 - k_{-1} x_{10}$$

$$\frac{dx_6}{dt} = -k_1 x_2 x_6 + k_{-1} x_{10} - k_1 x_4 x_6 + 0.5 k_{-1} x_9$$

$$x_7 = -[Q^+] + x_6 + x_8 + x_9 + x_{10}$$

$$x_8 = \frac{K_2 x_1}{K_2 + x_7}$$

$$x_9 = \frac{K_3 x_3}{K_3 + x_7}$$

$$x_{10} = \frac{K_1 x_5}{K_1 + x_7}$$

$$k_1 = k_{10} \exp\left(\frac{-E_1}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right)$$

$$k_2 = k_{20} \exp\left(\frac{-E_2}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right)$$

$$k_{-1} = k_{-10} \exp\left(\frac{-E_{-1}}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right)$$

Numerical difficulties have been frequently encountered in previous studies of this parameter estimation problem (Biegler et al., 1986). As a result, some researchers simplified the model by reducing the number of differential equations and the number of parameters (Tjoa and Biegler, 1991; Guay and McLean, 1995; Sulieman et al., 2009). The complete model is used in the current study. Wu et al. (2011b) applied the orthogonalization algorithm and the MSE-based criterion in Tables 3.1 and 3.2 to select a parameter subset for estimation in this model, using only data at

40 °C for parameter estimation. The authors demonstrated that 5 of the 9 model parameters should be selected for estimation to obtain the best expected predictions using this limited data set. In the current chapter, the parameter selection study is extended to include the data at 67 and 100 °C to study the effect of including additional data on the number of estimable parameters. The orthogonalization technique and Wu's MSE criterion are first applied to rank and select parameters (O+r_{CC}) in Section 3.4.1. These results are compared with leave-one-out cross-validation in Section 3.4.2. In Sections 3.4.3 – 3.4.5, the three different MSE-based parameter selection methods are compared when estimating parameters from the 67 °C data alone and when using all of the data for parameter estimation.

3.4.1 Ranking and Parameter Selection using Orthogonalization and Wu's MSE Criterion (O+r_{CC})

The experimental data are divided into two cases for analysis: 1) only the mid temperature (67 °C) data are available for estimation, and 2) all the data (40, 67, and 100 °C) are available for estimation. With the mid temperature data, the three activation energies are eliminated from the model, because the reference temperature, T_0 = 340.15 K, is the same as the reactor temperature, T (see expressions for k_1, k_2, and k_{-1} in Table 3.5). Table 3.6 shows the initial parameter guesses and associated uncertainties used for parameter ranking. These initial guesses and uncertainties are the same that were used by Wu et al. (2011b). The parameter rankings obtained using the orthogonalization algorithm from Table 3.1 are also shown in Table 3.6. When only the mid temperature data are used for parameter estimation, K_1 ranks first. When all of the data are used for parameter estimation, E_2 and K_1 rank at the top of the list due to their large influence on model predictions and large uncertainties in their initial guesses. Parameter K_3 ranks at the bottom of the

list for both cases because of its small influence, precise initial value, and correlation with parameters that appear higher in the list.

Table 3.6: Initial parameter guesses, uncertainty factors and rankings obtained using the orthogonalization algorithm

Parameter	Initial Guess	Uncertainty $s_{\theta 0}$	Rank (67°C data)	Rank (All data)
k_{10} (kg mol^{-1} h^{-1})	1.0	0.2	3	6
E_1 (cal mol^{-1})	1 x 10^4	2 x 10^3	-	4
k_{20} (kg mol^{-1} h^{-1})	2.0	0.3	2	5
E_2 (cal mol^{-1})	1 x 10^4	1 x 10^4	-	1
k_{-10} (h^{-1})	2 x 10^3	5 x 10^2	4	7
E_{-1} (cal mol^{-1})	1 x 10^4	4 x 10^3	-	3
K_1 (mol kg^{-1})	3 x 10^{-16}	2 x 10^{-16}	1	2
K_2 (mol kg^{-1})	5 x 10^{-14}	5 x 10^{-15}	5	8
K_3 (mol kg^{-1})	2 x 10^{-16}	1 x 10^{-17}	6	9

Based on each ranked list, a series of SMs were formulated. First, only the top-ranked parameter was estimated and the other parameters were left at their initial guesses. Next, the top two parameters were estimated, followed by the top three parameters, and so on until all model parameters were estimated. For each parameter estimation step, the Matlab™ program "lsqnonlin" was used to estimate parameters by minimizing the objective function shown in Equation (3.4). The solver "ode15s" was used to solve the system of model equations. The tolerances settings for the ode solver and optimizer were carefully selected to avoid numerical problems and ensure accurate results (Appendix A). Problems with local optima were sometimes encountered and parameter estimation was restarted with updated parameter estimates obtained from the previous SM. As expected, estimating additional parameters resulted in an improved fit of the data, as shown by the decreasing trends in the objective functions shown in Figures 3.1 and 3.2. An improvement in the objective function is observed for up to 3 parameters estimated using

the 67 °C data and up to 6 parameters estimated using all the data. When additional parameters are estimated, there is negligible improvement in the fit.

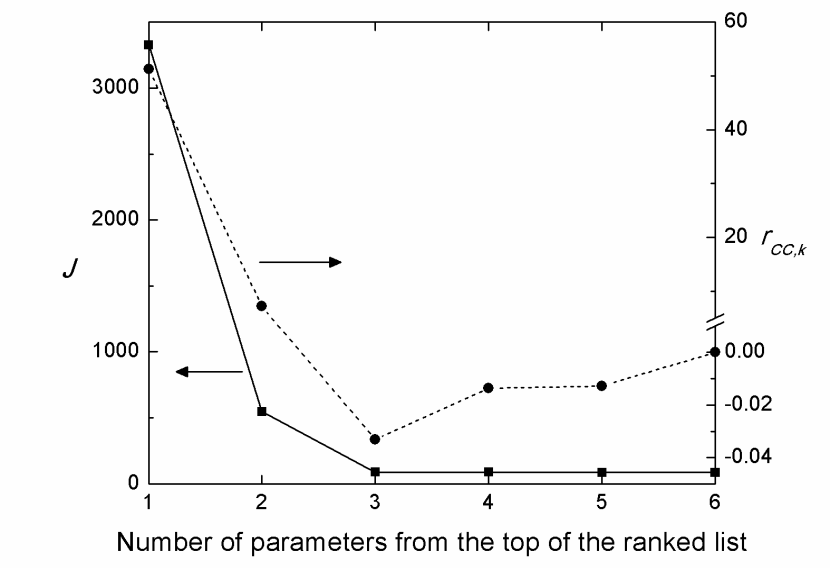

Figure 3.1: Effect of the number of parameters estimated on the objective function *J* and $r_{CC,k}$ values using only the 67 °C data. Note the break and the different scales used on the vertical axis for $r_{CC,k}$.

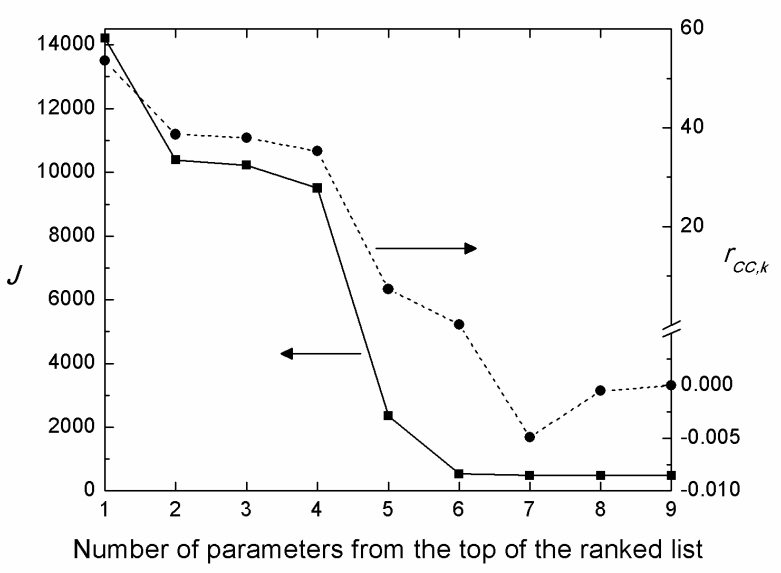

Figure 3.2: Effect of the number of parameters estimated on the objective function *J* and $r_{CC,k}$ values using the data at 40, 67, and 100 °C. Note the break and the different scales used on the vertical axis for $r_{CC,k}$.

The MSE-based model selection criterion was then used to determine the appropriate number of parameters from the ranked lists. The corresponding $r_{CC,k}$ values, shown in Figures 3.1 and 3.2, were calculated from the objective function values using the algorithm in Table 3.2. With only the 67°C data, the lowest value of the criterion occurs when the top three parameters are estimated. When all the data are used for parameter estimation, $r_{CC,k}$ was smallest at k=7. Thus, the estimating three and seven parameters, respectively, should give the best predictions with the lowest MSE. That is, with all the data available, parameters E_2, K_1, E_{-1}, E_1, k_{20}, k_{10}, and k_{-10} should be estimated; if only 67 °C data are available, parameters K_1, k_{20}, and k_{10} should be estimated. As expected, including additional data at different temperatures enabled the estimation of additional parameters. These results agree with the work of Schittkowski (2007) who selected seven parameters for estimation using a backward selection eigenvalue method. Schittkowski used

different assumptions about the initial parameter guesses and scaling factors, but six of the seven selected parameters were the same as the parameters selected in this work. Note that Schittkowski's eigenvalue method relies on an arbitrary cut-off to categorize parameters as estimable or inestimable.

Model predictions at 40 °C, resulting from estimation of the top seven parameters (with K_2 and K_3 held fixed) are shown in Figure 3.3 along with the experimental data. The model predictions match the data very well. Similar good agreement was obtained between predictions and data at 67 and 100 °C (not shown).

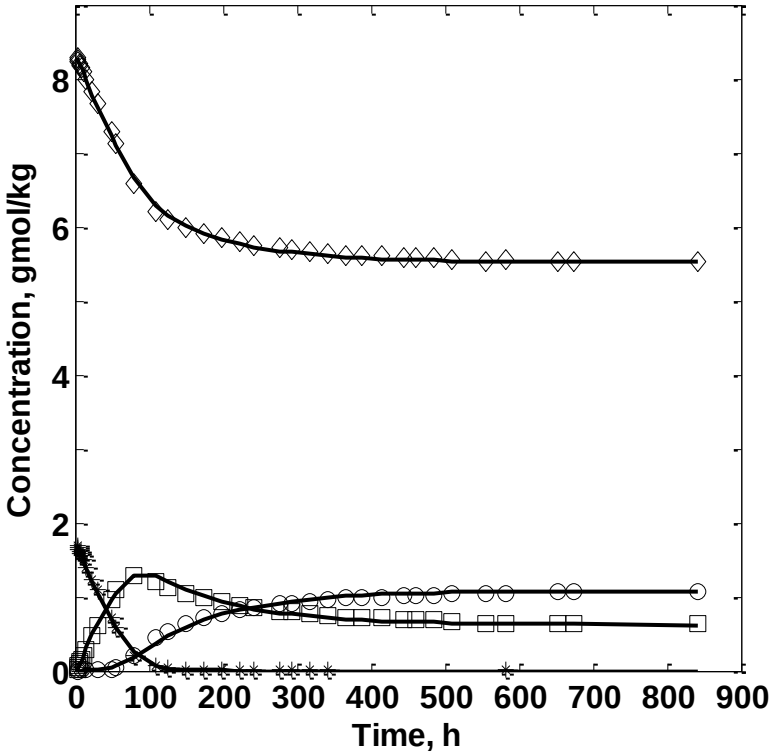

Figure 3.3: Comparison of model predictions obtained by estimating the seven top-ranked parameters using all the data, and measured data obtained at 40°C. The remaining two parameters were fixed at nominal values. *, ◊ and □ indicate measured values of y_1, y_2 and y_3, respectively, and ○ is the calculated value of y_4. Solid lines represent model predictions.

3.4.2 Leave-one-out Cross-validation

The results obtained using the O+r_{CC} technique were verified using cross-validation, which is a computationally-intensive technique used frequently in model selection and validation (Stone, 1974). Prior to the development of Wu's MSE-based technique, cross-validation was used by Thompson et al. (2009) to select the appropriate number of parameters to estimate from a ranked list of parameters in a polyethylene model. The method used by Thompson et al. involved leaving out one data point at a time, and using the remaining N-1 data points for parameter estimation.

The adjusted parameter values were then used to estimate the value of the withheld data point. This procedure was repeated for all the data points to compute the cross-validation objective function:

$$CV_k = \sum_{q=1}^{N} \left(\frac{y_q - \hat{y}_{qk}}{s_{yq}} \right)^2 \tag{3.11}$$

where y_q is the value of the withheld data point, and $\hat{y}_{qk}$ is the corresponding estimated value. The lowest CV_k value corresponds to the k-parameter SM that is expected to give the best predictions. In the current study, this procedure was applied using the 63 experimental points in the 67 °C data set and the 256 data points from all three data sets. The resulting cross-validation objective function values obtained using all of the data are shown in Figure 3.4. The cross-validation technique confirms that seven parameters should be selected for estimation, which is the same result that was obtained using the O+r_{CC} technique. CV_k decreases up to k=7 as the predictive ability of the model improves initially with additional parameters included in the estimation. However, CV_k increases beyond seven parameters, indicating a decline in the predictive ability of the model (i.e., with more than seven parameters estimated, the model fits the noise rather than the data). With only the 67 °C data, cross-validation confirmed that three parameters should be selected for estimation (not shown).

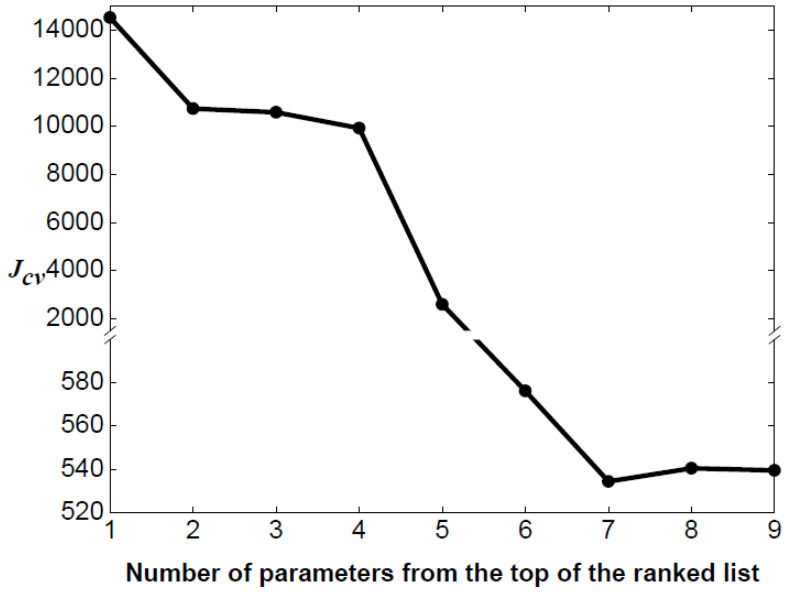

Figure 3.4: Effect of the number of parameters estimated on the cross-validation objective function values using the data of Biegler et al. (1986) collected at 40, 67, and 100 °C. Note the break and the different scales used on the vertical axis.

A modeller may not be confident in the values of the initial guesses and scaling factors that are required using the O+r_{CC} technique. Different values for the uncertainties and initial guesses will result in different sensitivity coefficients in the Z matrix and potentially different ranked parameter lists and subsets selected for estimation. It is possible that a different set of parameters could give a better fit or better predictions due to a poor choice of initial guesses. The sensitivity of the parameter subset selected to the initial parameter guesses is assessed in the following section.

3.4.3 Application of Proposed Robustness Test

To assess the robustness of the previous selection technique for this batch reactor model, the procedure in the O+r_{CC} technique was repeated 100 times, wherein new ranked lists and

parameter subsets were obtained using different initial guesses. These 100 new initial guesses were randomly chosen for each of the 9 parameters, assuming a uniform distribution over the uncertainty ranges given in Table 3.6.

With the different initial guesses, 100 different rankings were obtained by repeating the orthogonalization technique. The frequency of the ranking for each parameter using the data at all three temperatures is shown in the shaded box diagram in Figure 3.5 (Stortelder, 1998) and in Table B.1 in Appendix B. In this diagram, the parameters are listed according to their original rank along the y-axis and the degree of shading for each cell depends on the frequency of that rank for the parameter. For example, parameter E_2, which was originally ranked first, ranked as the top-selected parameter using 93 of the sets of initial parameter guesses. Parameters K_1, E_{-1}, K_2, and K_3 all retained their original ranking more than 50% of the time. However, there was more variation among parameters ranked four though seven.

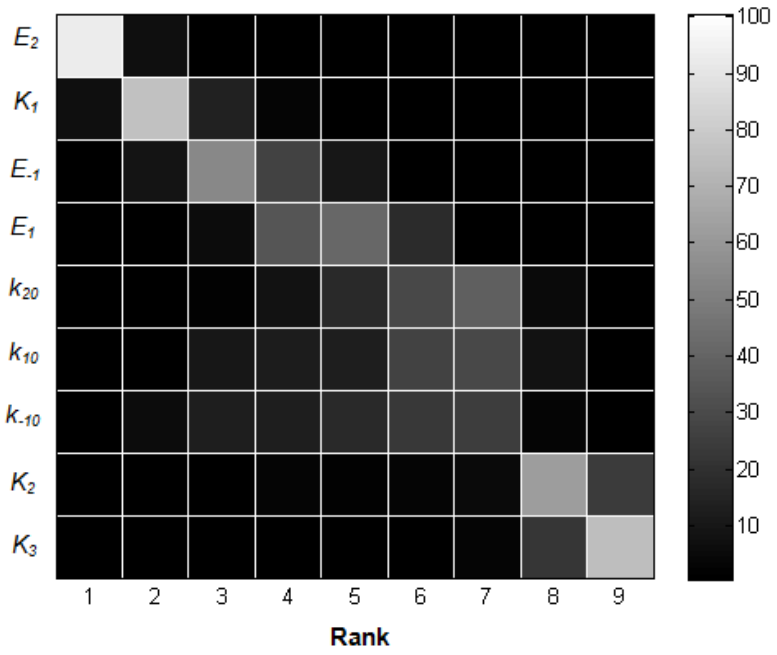

Figure 3.5: Frequency of parameter rankings using all data for 100 random initial parameter guesses

Wu's MSE-based criterion was applied to each of the 100 ranked lists to determine the optimal number of parameters to estimate for each list. The frequecies of the optimal number of parameters selected for estimation are shown in Figure 3.6. As shown in the figure, the original seven parameters were selected for estimation nearly 80% of the time. Different results are obtained using different models with different initial guesses, scaling factors, and data. For example, the same robustness procedure using 100 initial guesses was performed using only the 67 °C data for estimation (not shown). It was found that four parameters were instead selected for estimation the majority of the time, with the original three-parameter subset selected 19 out of 100 times.

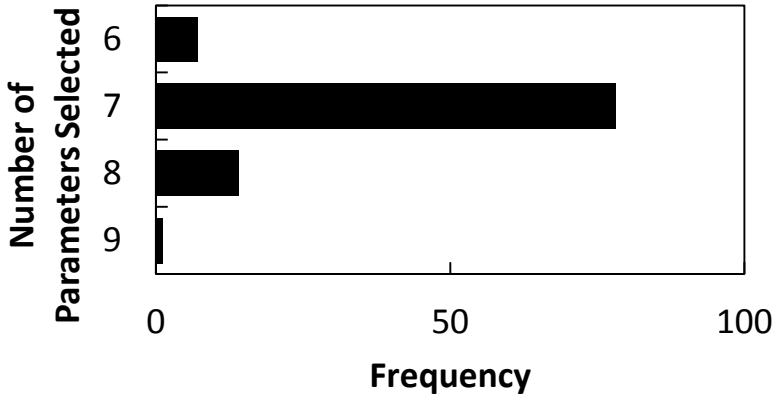

Figure 3.6: Frequency of the number of parameters selected using all data and 100 random initial parameter guesses

This robust parameter selection technique involves many parameter estimation steps. It can be computationally expensive to perform the series of parameter estimations 100 times for each ranked list, especially for complex models with many parameters. For comparison, the robust parameter selection technique by Chu et al. (2009), which does not use parameter estimation to rank and select parameters, is applied to the batch reactor model in the following section.

3.4.4 Application of Chu's MSE-based Method

The algorithm outlined in Table 3.3 was applied to the batch reactor model using the data collected at 67 °C and the data collected at all temperatures. With Chu's method, 100 sets of possible true parameter values were first selected from the uncertainty ranges given in Table 3.6, assuming a uniform distribution. This procedure was also repeated with 10000 sets of possible parameter values as recommended by Chu et al. (2009). The parameters were ranked according to the average bias terms calculated at each step. In Table 3.7, the resulting rankings obtained using 10000 sets of possible parameter values are compared with the parameter rankings obtained in Section 3.4.1. The number of parameters selected using each method is denoted by the symbol *. With the 67 °C data, the parameter rankings and selected subset were identical using 100 and 10000 sets. With all the data, parameters ranked four through six changed positions when using 100 sets compared to 10000 sets, but the same subset was selected in both cases.

Table 3.7: Comparison of parameter rankings and subsets selected using the orthogonalization technique and Wu's MSE-based technique and the method by Chu et al. (2009) using 10000 sets of possible parameter values

	O+r_{CC} technique		Chu's MSE technique	
Parameter	Rank (67°C data)	Rank (All data)	Rank (67°C data)	Rank (All data)
k_{10}	3*	6	3	5
E_1	-	4		6
k_{20}	2	5	4	7
E_2	-	1		1
k_{-10}	4	7*	2	4
E_{-1}	-	3		3
K_1	1	2	1	2
K_2	5	8	5	8
K_3	6	9	6*	9*

The rankings obtained using the O+r_{CC} technique and Chu's MSE method were similar. Using the data at 40, 67, and 100 °C, the top-three and the bottom-two parameters were identical using both methods. Only the parameters that ranked four through seven showed slight differences in rankings. The rankings for these parameters also showed the largest variability using the robust method in the previous section (see Figure 3.5). Unlike the proposed robustness method, Chu's method indicated that all parameters should be selected for estimation using either the 67 °C data alone or all the data. This discrepancy may be due to the difference in the way that each method approximates the bias term. Wu's MSE-based technique uses data and weighted-least-squares, whereas in Chu's method, the modeller assumes a uniform distribution of potential true parameter values, and calculates an average bias from a random sample. As a result Wu's MSE-based technique can make use of information from the data set, whereas Chu's method must rely on assumptions about potential true values. The trade-off between bias and variance estimated using Chu's method may be very different from the trade-off determined by Wu's method.

Chu's method combines parameter ranking and selection in a single algorithm. Wu's MSE-based parameter selection method can also be extended to simultaneously rank and select parameters, as shown in Table 3.4. This new approach is applied to the batch reactor model in the following section.

3.4.5 Application of New MSE-based Ranking Method (Sr_{CC})

A new MSE-based ranking method, which simultaneously ranks and selects parameters using $r_{CC,k}$ (Sr_{CC} technique, see Table 3.4 for algorithm), was tested using the model and data from Biegler et al. (1986). Since the model that predicts the 67 °C data set contains six parameters, each of the six parameters was first estimated individually with the remaining five held fixed. Estimation of k_{20} resulted in the lowest $r_{CC,1}$ value of 16.81 and thus k_{20} was selected

as the top-ranked parameter, as shown in Table 3.8. Next k_{20} was estimated along with each of the remaining five parameters. The parameter pair that gave the lowest value of $r_{CC,2}$=1.24 was k_{20} and k_{10}. The minimum value of $r_{CC,3}$= -0.042 was obtained when parameters k_{20}, k_{10}, and K_1 were estimated. Note that the negative value of $r_{CC,3}$ indicates that the SM with three parameters should give better predictions, in terms of MSE, than the model with all six parameters estimated. When the best four parameters were estimated, $r_{CC,4}$ was higher than $r_{CC,3}$. This indicates that estimating three parameters, with the remaining three parameters held fixed, is preferable to estimating four. This method was repeated for all the data and the seven-parameter subset containing E_2, K_1, E_{-1}, E_1, k_{20}, k_{10}, and k_{-10} was found to give the lowest $r_{CC,k}$ value, as shown in Table 3.8. Note that the subsets selected using this new MSE-based method are identical to the subsets selected using the O+r_{CC} technique, but again disagree with Chu's MSE method. This new MSE-based method and orthogonalization combined with Wu's MSE-based method use information from the data when assessing the trade-off between bias and variance, whereas Chu's method does not.

Table 3.8: Comparison of parameter rankings and subsets selected using the orthogonalization method with Wu's MSE-based technique and the new MSE-based ranking method

	O+r_{CC} technique		Sr_{CC}	
Parameter	Rank (67°C data)	Rank (All data)	Rank (67°C data)	Rank (All data)
k_{10}	3*	6	2	2
E_1	-	4		4
k_{20}	2	5	1	1
E_2	-	1		7*
k_{-10}	4	7*	4	5
E_{-1}	-	3		3
K_1	1	2	3*	6
K_2	5	8	6	9
K_3	6	9	5	8

Table 3.8 shows that the rankings of some parameters differed between using the O+r_{CC} technique and the Sr_{CC} technique. For example, using the 67 °C data, parameter K_1 ranked at the top of the list using the O+r_{CC} technique due to its influence on model predictions and large specified uncertainty. This ranking was obtained prior to parameter estimation using the data. However, with the Sr_{CC} technique, changing K_1 from its initial guess did not have as much impact on $r_{CC,1}$ as adjusting k_{20}. As a result, the Sr_{CC} technique selected k_{20} as the top-ranked parameter. Similarly, when using all the data, the Sr_{CC} technique first selected parameter k_{20} instead of E_2.

Note that with this forward-selection method, the estimated value of the selected parameter in the 1-parameter subset depends on the initial values of the other parameters. Thus, this method is also prone to problems arising from poor initial guesses. To assess the robustness of this new MSE-based method, the procedure was repeated using 20 sets of initial guesses selected randomly from the uncertainty ranges given in Table 3.6, using a uniform distribution. As shown in Figures 3.7 and 3.8, there was more variation among parameters ranked two through six, but the same seven parameters were frequently selected for estimation. The frequency of the parameter rankings are also given in Table B.2 in Appendix B.

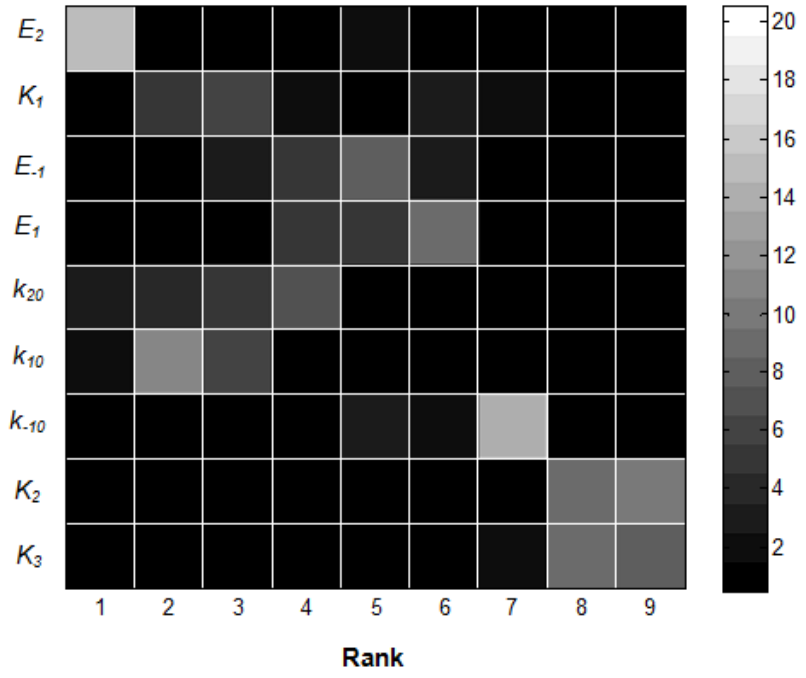

Figure 3.7: Frequency of parameter rankings using the new MSE-based ranking method with all of the data for 20 random initial parameter guesses

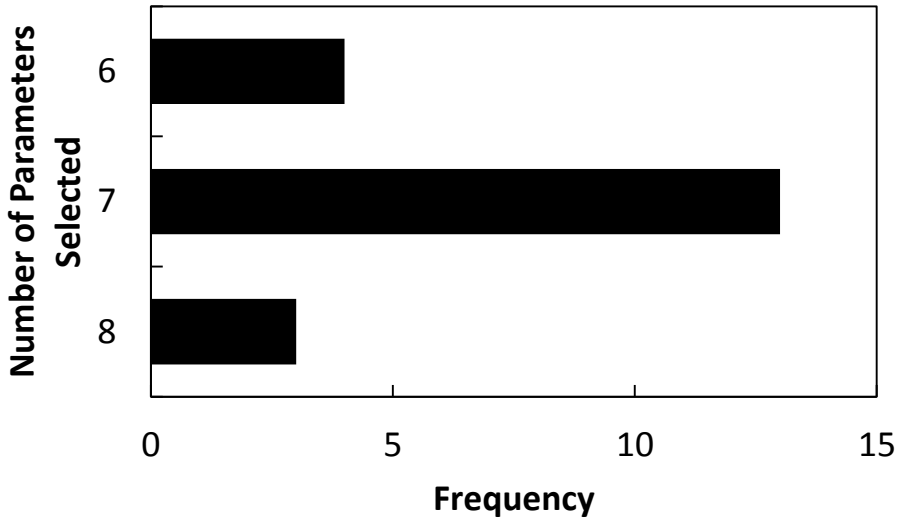

Figure 3.8: Frequency of the number of parameters selected using the new MSE-based ranking method with all of the data for 20 random initial parameter guesses

A shortcoming of this method (and the two methods above) is that it selects parameters for estimation considering the trade-off between bias and variance for predictions made at the experimental points for parameter estimation. A modeller will be interested in using the model to obtain predictions at different operating conditions, rather than at conditions where data are already available. In future, it is recommended that these robust parameter selection techniques be extended to select parameters for estimation that will provide good predictions at desired operating conditions.

This new MSE-based method, which requires the most parameter estimation steps, is the most computationally expensive of the three techniques tested. The computational requirements of the three different robust MSE-based techniques are compared in the next section.

3.4.6 Comparison of Computational Effort

The computational efforts required for the various methods are summarized in Table 3.9. The computing times are compared when using only the 67 °C data (six parameters and 63 data points) and using all of the data (9 parameters and 256 data points). These computation times, reported in seconds, were determined using the tic and toc function in MATLAB with a HP Pavilion dv6700 notebook computer. As shown in the first row of Table 3.9, the computation time required for the O+r_{CC} technique is approximately 2 minutes when using the 67 °C data, and 14 minutes using all of the data. This computation time is much shorter than the time required for leave-one-out cross-validation, which takes approximately 2.6 and 75 hours. To complete the proposed robustness test with 100 sets of initial parameter guesses requires 76140 seconds using all the data, which is approximately 21 hours of computation time. When 100 sets of possible parameter values are considered, Chu's MSE method requires a similar amount of computational effort compared to using the O+r_{CC} technique. Using 10000 sets of possible parameter values, as

suggested by Chu et al., requires 100 times the effort. Note that Chu's method does not involve parameter estimation, so obtaining parameter values will require additional computation time. The Sr_{CC} technique, which uses Wu's MSE criterion to simultaneously rank and select parameters, involves $p(p+1)/2-1$ parameter estimations for each set of initial guesses. With the original set of initial guesses, this method requires 248 seconds to select parameters using the 67 °C data set, which is about twice the effort required for the O+r_{CC} technique. With all of the data, the computation time is approximately 3.5 times greater than the time required for the O+r_{CC} technique. Approximately 12 hours of computation time are required for 20 different sets of initial parameter guesses, using all the data.

For comparison, an additional technique was performed wherein the orthogonalization algorithm was repeated with 100 different initial parameter guesses to confirm that the ranking for the initial guesses was obtained most frequently. Wu's MSE-based selection technique was performed using the initial guesses and the most frequent ranking. As shown in Table 3.9, the computation time is much lower (by factor of 50) compared to the proposed robustness test from Section 3.4.3. This final strategy is recommended to modellers to assess the sensitivity of the parameter set selected to assumptions about initial guesses, as it is able to demonstrate the variability in parameter ranking. This method does not require as many time-consuming parameter estimation steps as the proposed robustness technique from Section 3.4.3, or the new Sr_{CC} technique.

Table 3.9: Comparison of computational effort for the parameter selection techniques. Computation times are reported in seconds.

Technique	67°C data	All data
O+r_{CC}	123	840
Leave-one-out cross-validation	9345	268756
Proposed robustness test with 100 sets of initial guesses	11439	76140
Chu's MSE-based method		
100 sets of possible parameter values	164	712
10000 sets of possible parameter values	16368	71243
Sr_{CC}		
Original parameter guesses	248	2913
20 sets of initial guesses	-	43203
Orthogonalization technique with 100 sets of initial guesses and Wu's MSE-based selection technique applied to 1 ranking	279	1512

3.5 Conclusions

Chemical engineers who build fundamental models often encounter numerical difficulties when trying to estimate all of the model parameters due to the number of unknown parameters and/or the limited data available. A recent MSE-based method was developed by Wu et al. to aid the modeller in selecting the appropriate parameter subset to include in the estimation. Unfortunately, this method and other methods that rely on a sensitivity matrix may be prone to problems resulting from a poor choice of initial parameter guesses.

In this Chapter, different robust MSE-based techniques were applied to a batch reactor model and their computational requirements were compared. The different techniques produced similar parameter rankings and the techniques that involved Wu's MSE-based criterion selected the same subsets for estimation. The parameters selected using these techniques are also consistent with those selected using leave-one-out cross-validation, but much less computation time is required. The method of Chu et al., wherein the bias term was approximated without making use of

experimental data gave markedly different results, indicating that additional parameters should be estimated.

The computational requirements of the robust parameter selection techniques varied, with the MSE-based methods that involve many parameter estimations requiring more computational effort than Chu's method. For modellers who have concerns about their choice of initial guesses and/or uncertainties, it is recommended that they first repeat the ranking procedure (orthogonalization algorithm in Table 3.1) several times. If a consistent parameter ranking is obtained, then the MSE-based method of Wu et al. can be applied to several sets of initial guesses that resulted in this ranking. The number of parameters to select for estimation can then be determined from the best value of the critical ratio r_{CC}.

The three robust MSE-based selection techniques compared in this Chapter all determine the parameters that should be estimated to provide the best predictions at the experimental points used for parameter estimation. Usually, model users want to make predictions at operating points that are different from those where measurements are available. In future, MSE-based parameter selection techniques should be extended to select parameters that will give predictions at the operating conditions that are important to the model user.

3.6 Nomenclature

d	number of response variables
f	right-hand side of nonlinear ODE model
g	response prediction from nonlinear ODE model
i	index for response variables
k	1) index for number of candidate parameters selected 2) reaction rate constant
l	index for measurements
m	index for experimental runs
n	number of measurements for each response variable
p	number of unknown parameters
q	index for data point in cross-validation

r number of experimental runs
r_C critical ratio in MSE-based algorithm
r_{CC} corrected critical ratio in MSE-based algorithm
r_{Ckub} truncated Kubokawa estimator
s_{yi} uncertainty associated with measured response
$s_{\theta j0}$ uncertainty associated with initial parameter guess
t time
u input trajectories
x state variables
y response variables
$\hat{y}$ predicted responses using parameter estimates
E activation energy
J objective function in parameter estimation
N total number of data points
Q^+ initial catalyst concentration
R ideal gas constant
R_k residual matrix in parameter ranking algorithm
S unscaled sensitivity matrix
T 1) columns from S corresponding to sensitivity vectors for parameters that may be included in the estimation
2) temperature
T_0 reference temperature
W columns from S corresponding to sensitivity vectors for parameters that are not included in the estimation
X subset of Z matrix corresponding to selected parameters
Z scaled sensitivity matrix

Greek symbols

ε stochastic component
θ unknown parameters
$\bar{\theta}_U$ vector of nominal values for unselected parameters
θ_U^* vector of potential true values for unselected parameters
σ^2 noise variance

3.7 References

Bates, D. M. and D. G. Watts, "Nonlinear Regression Analysis and Its Applications," John Wiley & Sons, NY (1988), pp. 72-76, 147-149.

Biegler, L. T., J. J. Damiano and G. E. Blau, "Nonlinear Parameter Estimation: A Case Study Comparison," AIChE J. **32**(1), 29-45 (1986).

Brun, R., P. Reichert and H. R. Künsch, "Practical Identifiability Analysis of Large Environmental Simulation Models," Water Resour. Res. **37**(4), 1015-1030 (2001).

Brun, R., M. Kühni, H. Siegrist, W. Gujer and P. Reichert, "Practical Identifiability of ASM2d Parameters – Systematic Selection and Tuning of Parameter Subsets," Water Res. **36**(16), 4113-4127 (2002).

Chu, Y. and J. Hahn, "Parameter Set Selection via Clustering of Parameters into Pairwise Indistinguishable Groups of Parameters," Ind. Eng. Chem. Res. **48**(13), 6000-6009 (2009).

Chu, Y., Z. Huang and J. Hahn, "Improving Prediction Capabilities of Complex Dynamic Models via Parameter Selection and Estimation," Chem. Eng. Sci. **64**(19), 4178-4185 (2009).

Degenring, D., C. Froemel, G. Dikta and R. Takors, "Sensitivity Analysis for the Reduction of Complex Metabolism Models," J. Process Control. **14**(7), 729-745 (2004).

Freni, G., G. Mannina and G. Viviani, "Identifiability Analysis for Receiving Water Body Quality Modelling," Environ. Modell. Softw. **24**(1), 54-62 (2009).

Guay, M. and D. D. McLean, "Optimization and Sensitivity Analysis for Multiresponse Parameter Estimation in Systems of Ordinary Differential Equations," Comput. Chem. Eng. **19(**12), 1271-1285 (1995).

Hocking, R. R., "Analysis and Selection of Variables in Linear Regression," Biometrics **32**(1), 1-49 (1976).

Jayasankar, B. R., A. Ben-Zvi and B. Huang, "Identifiability and Estimability Study for a Dynamic Solid Oxide Fuel Cell Model," Comput. Chem. Eng. **33**(2), 484-492 (2009).

Karimi, H., M. A. Schaffer and K. B. McAuley, "A Kinetic Model for Non-Oxidative Thermal Degradation of Nylon 66," submitted to Macromolecular Reaction Engineering, April (2011).

Kou, B., K. B. McAuley, J. C. C. Hsu and D. W. Bacon, "Mathematical Model and Parameter Estimation for Gas-Phase Ethylene/Hexene Copolymerization with Metallocene Catalyst," Macromol. Mater. Eng. **290**(6), 537-557 (2005a).

Kou, B., K. B. McAuley, C. C. Hsu, D. W. Bacon and K. Z. Yao, "Mathematical Model and Parameter Estimation for Gas-Phase Ethylene Homopolymerization with Supported Metallocene Catalyst," Ind. Eng. Chem. Res. **44**(8), 2428-2442 (2005b).

Leis, J. R. and M. A. Kramer, "The Simultaneous Solution and Sensitivity Analysis of Systems Described by Ordinary Differential Equations," ACM Trans. Math. Softw. **14**(1), 45-60 (1988).

Littlejohns, J. V., K. B. McAuley and A. J. Daugulis, "Model for a Solid-Liquid Stirred Tank Two-Phase Partitioning Bioscrubber for the Treatment of BTEX," J. Hazard. Mater. **175**(1-3), 872-882 (2010).

Lund, B. F. and B .A. Foss, "Parameter Ranking by Orthogonalization – Applied to Nonlinear Mechanistic Models," Automatica. **44**(1), 278-281 (2008).

Machado, V. C., G. Tapia, D. Gabriel, J. Lafuente and J. A. Baeza, "Systematic Identifiability Study Based on the Fisher Information Matrix for Reducing the Number of Parameters Calibration of an Activated Sludge Model," Environ. Modell. Softw. **24**(11), 1274-1284 (2009).

Maria, G., "A Review of Algorithms and Trends in Kinetic Model Identification for Chemical and Biochemical Systems," Chem. Biochem. Eng. Q. **18**(3), 195-222 (2004).

McLean, K. A. P. and K. B. McAuley, "Mathematical Modelling of Chemical Processes – Obtaining the Best Model Predictions and Parameter Estimates using Identifiability and Estimability Procedures," Can. J. Chem. Eng. Accepted for publication (2011).

Quaiser, T. and M. Mönnigmann, "Systematic Identifiability Testing for Unambiguous Mechanistic Modeling – Application to JAK-STAT, MAP Kinase, and NF-κB Signaling Pathway Models," BMC Systems Biology. **3**(50), (2009).

Raue, A., C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller and J. Timmer, "Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood," Bioinformatics. **25**(15), 1923-1929 (2009).

Saltelli, A., K. Chan and E. M. Scott, "Sensitivity Analysis," John Wiley & Sons, NY (2000), pp. 15-50.

Schittkowski, K., "Experimental Design Tools for Ordinary and Algebraic Differential Equations," Ind. Eng. Chem. Res. **46**(26), 9137-9147 (2007).

Stone, M., "Cross-validation Choice and Assessment of Statistical Predictions," J. Roy. Stat. Soc. B. **36**(2), 111-147 (1974).

Stortelder, W., "Parameter Estimation in Nonlinear Dynamical Systems," Ph.D. Thesis, CWI Amsterdam (1998), pp. 91.

Sulieman, H., I. Kucuk, and P. J. McLellan, "Parametric Sensitivity: a Case Study Comparison," Comput. Stat. Data Anal. **53**(7), 2640-2652 (2009).

Thompson, D. E., K. B. McAuley and P. J. McLellan, "Parameter Estimation in a Simplified MWD Model for HDPE Produced by a Ziegler-Natta Catalyst," Macromol. React. Eng. **3**(4), 160-177 (2009).

Vajda, S., H. Rabitz, E. Walter and Y. Lecourtier, "Qualitative and Quantitative Identifiability Analysis of Nonlinear Chemical Kinetic-Models," Chem. Eng. Commun. **83**(1), 191-219 (1989).

Weijers, S. R. and P. A. Vanrolleghem, "A Procedure for Selecting Best Identifiable Parameters in Calibrating Activated Sludge Model No.1 to Full-Scale Plant Data," Water Sci. Technol. **36**(5), 69-79 (1997).

Woloszyn, J. D. and K. B. McAuley, "Application of Parameter Selection and Estimation Techniques in a Thermal Styrene Polymerization Model," Macromol. React. Eng. (2011). doi: 10.1002/mren.201100021

Wu, S., T. J. Harris and K. B. McAuley, "The Use of Simplified or Misspecified Models: Linear Case," Can. J. Chem. Eng. **85**(4), 386-398 (2007).

Wu, S., K. B. McAuley and T. J. Harris, "Selection of Simplified Models: II. Development of a Model Selection Criterion Based on Mean Squared Error," Can. J. Chem. Eng. **89**(2), 325-336 (2011a).

Wu, S., K. A. P. McLean, T. J. Harris and K. B. McAuley, "Selection of Optimal Parameter Set Using Estimability Analysis and MSE-Based Model-Selection Criterion," Int. J. Advanced Mechatronic Systems, **3**(3), 188-197 (2011b).

Yao, K. Z., B. M. Shaw, B. Kou, K. B. McAuley and D. W. Bacon, "Modeling Ethylene/Butene Copolymerization with Multi-Site Catalysts: Parameter Estimability and Experimental Design," Polym. React. Eng. **11**(3), 563-588 (2003).

Chapter 4

Contributions, Conclusions, and Recommendations

4.1 Contributions

Methods for selecting optimal parameter subsets for estimation were reviewed and strategies were developed to aid users with complex models and limited data. In Chapter 2, the literature concerning identifiability and estimability is summarized and advice for modellers with large models is presented. New techniques are presented in Chapter 3 that can aid modellers who are interested in selecting only a subset of parameters for estimation. Three techniques based on mean squared error (MSE) were applied to a Dow batch reactor model (Biegler et al., 1986). The orthogonalization algorithm and Wu's original MSE-based criterion were used to rank and select parameters using a portion of the data and all of the available data. Accomplishing this outcome required careful tuning of error tolerance parameters for the DAE solver and the optimizer parameter options as described in Appendix A. For the first time in the literature, Monte Carlo methods are demonstrated to be effective for assessing the robustness of the selected parameter subset to initial parameter guesses. In addition, a new MSE-based method for simultaneous parameter ranking and subset selection is proposed and evaluated. Comparisons are made with results from the robust Monte-Carlo approach described above and with a competing technique developed by Chu and Hahn (2009).

4.2 Conclusions

The conclusions that can be drawn from the research presented in Chapters 2 and 3 are:

1) Aside from using linearization-based techniques, assessing the identifiability of parameters in large nonlinear models is difficult to perform. The other identifiability analysis techniques reviewed in Chapter 2 were all applied to models with less than 15 parameters and states. Linearization-based techniques, which are the simplest to use, can confirm identifiability of a nonlinear model, but cannot provide firm conclusions about the identifiability of a nonlinear model if the corresponding linear model is not identifiable.
2) Techniques for assessing estimability are easier to perform in large nonlinear models. If model parameters are estimable, then they are also identifiable. If some parameters are not estimable, the problem may be a result of either estimability or identifiability issues.
3) The eigenvalue and orthogonalization techniques were demonstrated in the literature to be the most successful parameter ranking methods for subset selection. These backward- and forward-selection techniques may be sub-optimal compared with combinatorial approaches, but are easy to use and require less computational effort.
4) Successfully estimating the model parameters in the Dow Chemical batch reactor model required careful tuning of the error tolerance parameters for the DAE solver "ode15s" and optimizer parameter options for the "lsqnonlin" function. Estimating 8 or 9 parameters was made possible by loosening the relative error tolerance and increasing the minimum change for finite difference gradients in the optimizer. However, selection of these parameter options required some trial and error.

5) Using the orthogonalization algorithm with Wu's MSE-based criterion, additional parameters were selected for estimation when more data was available for parameter estimation (3 parameters were selected with only the 67 °C data versus 7 parameters selected using all of the data). The selected parameter subsets were the same as those obtained using cross-validation, which required much more computation time.
6) The proposed robustness technique was applied to the Dow example using 100 different sets of initial parameter values. With all of the data, the seven-parameter subset that was initially selected for estimation was frequently selected using the different sets of initial guesses. With only the 67 °C data, four parameters were frequently selected in the robustness test.
7) The selected subsets using the new MSE-based technique were the same as those selected using the orthogonalization algorithm to first rank parameters. This technique requires more parameter estimation steps and thus more computation time compared with using the orthogonalization algorithm and Wu's original MSE-based technique.
8) The method proposed by Chu et al. (2009) required the least amount of computation time compared to the techniques involving many parameter estimations, but selected additional parameters for estimation. This method does not rely on information from the available data to select parameters for estimation, but rather estimates the trade-off between bias and variance using assumptions about potential true parameter values.

4.3 Recommendations

4.3.1 Recommendations for Modellers

For modellers who are concerned that their model may have identifiability or estimability problems, it is recommended that they first attempt parameter estimation. It may be possible to obtain reasonable parameter estimates with narrow confidence intervals without having to perform any time-consuming identifiability or estimability analysis. However, if parameter estimation fails, estimability analysis is the next recommended step since it is easier to assess than identifiability in large models. If a modeller is faced with an estimability problem, he/she may: i) simplify the model, ii) design additional experiments, or iii) fix some parameter at nominal values and estimate only a subset of parameters.

Modellers who choose to simplify the model can use the identifiability and estimability techniques discussed in Chapter 2 to determine problems with the model structure and which combinations of parameters cannot be estimated together. If a modeller decides to design additional experiments, the sensitivity matrix can be augmented with the proposed experimental settings and the Fisher Information matrix with the original and new settings can be used to obtain new confidence intervals for parameters. Modellers interested in estimating only a subset of parameters and fixing others at nominal values can use the eigenvalue and orthogonalization techniques described in Chapter 2 to rank parameters from most estimable to least estimable. The number of parameters to select for estimation can then be determined using Wu's MSE-based criterion.

This parameter subset selection technique depends on initial parameter guesses and scaling factors. Modellers who are uncertain about their initial assumptions may repeat the orthogonalization algorithm using many different sets of initial parameter values. Such a Monte

Carlo approach can be used to generate different ranked parameter lists. Wu's MSE-based criterion can then be applied to several sets of initial guesses that give the most frequent parameter ranking. The best value of the critical ratio, r_{CC}, corresponds to the number of parameters that should be estimated to obtain the best predictions.

4.3.2 Recommendations for Future Work

The parameter subset selection techniques in Chapter 3 are based on the MSE at the experimental points available for parameter estimation. Modellers are generally interested in making predictions at conditions where data do not exist. As a result, the three MSE-based techniques should be extended to select the parameters for estimation that will give the best predictions at the operating conditions of interest to the modeller.

4.4 References

Biegler, L. T., J. J. Damiano and G. E. Blau, "Nonlinear Parameter Estimation: A Case Study Comparison," AIChE J. **32**(1), 29-45 (1986).

Chu, Y., Z. Huang and J. Hahn, "Improving Prediction Capabilities of Complex Dynamic Models via Parameter Selection and Estimation," Chem. Eng. Sci. **64**(19), 4178-4185 (2009).

Appendix A

Tuning Error Tolerance Parameters in the Dow Chemical Batch Reactor Example

The Dow chemical batch reactor model and data (Biegler et al., 1986) used as the example in Chapter 3 consisted of a system of stiff differential/algebraic equations (DAEs). Biegler et al. outline the various approaches taken by different researchers who attempted to solve this problem and summarize the difficulties encountered. For example, Biegler et al. (1986) state that "none of the investigators was able to solve the problem automatically in one go," and "all the solutions required careful attention by the researchers to several computational difficulties." These researchers encountered problems with ill-conditioning as well as convergence problems in the optimization scheme with certain parameter values. These convergence problems were found to occur less frequently by adjusting the error tolerances. To avoid some of difficulties encountered, most of the researchers reduced the model structure to a system of ordinary differential equations (ODEs) and eliminated some of the model parameters.

Wu et al. (2011) attempted to solve the original DAE model with the full system of equations and parameters, wherein only the Arrhenius equations were re-parameterized to reduce correlation between some parameters. Using the solver "nlscon" for estimating parameter values and the solver "ode15s" for solving the system of stiff DAEs, Wu was able obtain parameter estimates and demonstrate the parameter subset selection technique using only the 40 °C data. Wu used a relative tolerance of 1×10^{-8} with the nlscon function, and a relative tolerance of 1×10^{-8} and an absolute tolerance of 1×10^{-20} with "ode15s". When estimating six or more parameters, Wu adjusted the relative tolerance to 1×10^{-6} for the DAE solver. With the additional

data at 67 and 100 °C, Wu could not overcome the numerical difficulties that were encountered and was unable to demonstrate the parameter subset selection technique.

For estimating the model parameters in this thesis, the optimizer was changed from "nlscon" to "lsqnonlin", which is a function included in the MATLAB optimization toolbox for solving nonlinear least-squares problems. The DAE solver, "ode15s", was left unchanged, but the error tolerances needed to be carefully adjusted. Numerical difficulties were often encountered, especially when trying to estimate 8 or 9 parameters. The relative and absolute error tolerances for the DAE solver were left unchanged for the low temperature data at 1×10^{-8} and 1×10^{-20}, respectively. However, the solver failed using these tolerances with the additional data at 67 and 100 °C. To successfully estimate the parameters using all of the data, looser error tolerances were required for the DAE solver. With the mid and high temperature data, parameters were successfully estimated by implementing a relative tolerance of 1×10^{-4} and an absolute tolerance of 1×10^{-10}. When estimating 8 or 9 parameters, the relative tolerance required further adjustment. The quality of the relative tolerance was further reduced and a value of 1×10^{-2} was used. Despite the implementation of these looser error tolerances, certain combinations of parameter values caused the solver to fail, especially when repeating the simulations with different initial parameter values. Some trial and error was required wherein the relative tolerance was adjusted within the range of 1×10^{-2} to 1×10^{-4}, depending on the number of parameters to be estimated.

Problems were also encountered with local optima. The optimizer would sometimes get stuck in a local minimum and would not adjust the parameter values enough to move away from the minimum. For this reason, the option "DiffMinChange", the minimum change for finite difference gradients, in the "lsqnonlin" function was changed from its default value of 1×10^{-8} to

1×10^{-2}. Changing this option would also help avoid certain combinations of parameter values that would cause the DAE solver to fail. Again, some trial and error was required in tuning this parameter; DiffMinChange was adjusted within the range of 1×10^{-4} to 4×10^{-2} to obtain reasonable parameter estimates. To further improve the success rate of the parameter estimations, the option "TolX", the termination tolerance on the parameter values, was initially adjusted to 1×10^{-4} and further adjusted to 1×10^{-2} when estimating 8 or 9 parameters.

Appendix B

Numerical Values for the Parameter Ranking Frequencies in Figures 3.5 and 3.7

The numerical values corresponding to each of the shaded cells in Figures 3.5 and 3.7 are provided below in Tables B.1 and B.2.

Table B.1: Frequency of parameter rankings corresponding to Figure 3.5

	Rank								
	1	2	3	4	5	6	7	8	9
k_{10}	93	7	0	0	0	0	0	0	0
E_1	7	76	14	3	0	0	0	0	0
k_{20}	0	9	54	27	10	0	0	0	0
E_2	0	1	6	34	41	18	0	0	0
k_{-10}	0	0	2	8	17	29	38	5	1
E_{-1}	0	1	10	12	13	27	29	8	0
K_1	0	6	13	13	17	23	25	3	0
K_2	0	0	1	3	2	3	5	62	24
K_3	0	0	0	0	0	0	3	22	75

Table B.2: Frequency of parameter rankings corresponding to Figure 3.7

	Rank								
	1	2	3	4	5	6	7	8	9
k_{10}	15	0	0	1	2	1	1	0	0
E_1	0	5	6	2	1	3	2	1	0
k_{20}	0	0	3	5	8	3	0	0	1
E_2	0	0	0	5	5	9	0	1	0
k_{-10}	3	4	5	7	1	0	0	0	0
E_{-1}	2	11	6	0	0	1	0	0	0
K_1	0	0	0	0	3	2	14	0	1
K_2	0	0	0	0	0	0	1	9	10
K_3	0	0	0	0	0	1	2	9	8