A Mean-Squared-Error-Based Methodology for Parameter Ranking and Selection to Obtain Accurate Model Predictions at Key Operating Conditions

by

Zahra Eghtesadi

A thesis submitted to the Department of Chemical Engineering

In conformity with the requirements for the

Degree of Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

December, 2014

To Whom I Believe

Abstract

In this thesis, a new mean-squared-error (MSE)-based criterion, r_{CCW}, is proposed to select the optimal number of parameters to estimate from the ranked list of parameters considering the operating region where accurate model predictions are desired. This new approach accounts for the trade-off between bias and variance as additional parameters from this ranked list are estimated. Next, a new forward-selection methodology based on r_{CCW} MSE-based criterion is devised to simultaneously rank and select the parameters while considering the conditions in the desired operating region. The information about the desired operating conditions is considered during both ranking and selection step in this parameter subset selection technique. This approach is valuable to modelers who want to make predictions about new products or grades or at new operating conditions, using models that are fitted only by utilizing data which were obtained at prior operating conditions. The forward selection MSE-based r_{CCW} criterion is then extended to the cases in which the Fisher information matrix (FIM) is not invertible. A singular FIM has been reported in many industrial chemical engineering models, many high-dimensional parameter estimation problems, and over-parameterized models. A singular FIM leads to a significant complication in the analysis of parameter estimation problems. In this thesis, two different approaches for parameter ranking and selection are undertaken and compared when the FIM is not invertible. The method that uses a reduced invertible FIM is shown to be superior to the alternative method that uses a pseudo inverse, using a linear regression case study. The methodology is extended for use in fundamental nonlinear dynamic models and illustrated using a film casting example.

Co-Authorship

The research presented in this thesis is conducted by me under the supervision of Professor Kim B. McAuley of the Department of Chemical Engineering at Queen's University.

Research in Chapter 2 and Chapter 3 were published in Industrial and Engineering Chemistry Research, and research in Chapter 4 will also be submitted for publication. These journal papers were drafted by me and edited and revised by Professor Kim B. McAuley, who is the co-author of the journal papers. Dr. Shaohua Wu from Honneywell Canada is the co-author of the journal paper presented in Chapter 2.

Material presented in Appendix A1 is a summary of my internship at DuPont™ Canada during December 2012 to October 2013. My research at DuPont™ Canada was supervised by Professor Kim B. McAuley from Queen's University and Mr. John Schlottke from E.I. du Pont Canada Company.

Acknowledgments

There are so many people and reasons that have made my journey through PhD a memorable one and my stay in Kingston and at Queen's a pleasant one.

I would like to express my greatest gratitude and appreciation to my supervisor Dr. Kim McAuley because of her endless support, understanding, patience and inspiration. I would like to immensely thank Kim not only because of her guidance on academic level, but also because of her thoughtfulness and honest care on family and personal levels.

I would like to thank Mr. John Schlottke and Mr. Goksin Yilmaz from DuPont Canada for their mentorship and advice during my internship at DuPont.

I would like to thank all the professors and staff in the Department of Chemical Engineering at Queen's University for their knowledge and assistance.

I would like to thank my parents and my husband for their unconditional love and constant support and encouragement.

I would like to thank my friends, fellow officemates and all the people whose company have made this episode of my life enjoyable and unforgettable.

Table of Contents

List of Figures

List of Tables

Nomenclature for Chapter 2

$A_1 = \left(X_1^T X_1\right)^{-1} X_1^T X_2$ expression used in Equation (2.13) and (2.17)

c number of measurements for each response variable

$\mathrm{Cov}(\cdot)$ variance-covariance matrix

d number of response variables

e stochastic component combined with any model mismatch

E subscript for extended model

$E(\cdot)$ expected value

i index for response variables

J objective function in parameter estimation

k scalar value

l index for measurements

m index for experimental runs

M mapping matrix from X to W

$\tilde{M}$ matrix defined in Equation(A2.12)

MSE mean-squared error prediction

n total number of measurements

p total number of parameters in the EM

$P = X(X^T X)^{-1} X^T$ expression used in Equation(2.33)

p_1 number of parameters in the SM

$P_1 = X_1(X_1^T X_1)^{-1} X_1^T$ expression used in Equation(2.5)

r number of experimental runs

R ideal gas constant

r_C likelihood ratio statistic, estimator for R_C

R_C true Wu's critical ratio

r_{CC} estimator for R_{CC}

R_{CC} true Wu's corrected critical ratio

r_{CW} estimator for R_{CW}

R_{CW} true proposed critical ratio

r_{CCW} estimator for R_{CCW}

R_{CCW} true proposed corrected critical ratio

$s_{\theta_{j0}}$ uncertainties in the initial value of parameter θ_j

s_{y_i} uncertainties in the i^{th} measured response value

S subscript for simplified model

S_E^2 unbiased estimators σ^2 in the EM

S_{ij0} corresponding element in i^{th} row and j^{th} column of the sensitivity matrix calculated at $\theta=\theta_0$

SSE sum of square residuals
T batch reactor temperature in K
T_0 reference temperature in K
$\mathrm{Tr}(\cdot)$ trace of a matrix
w total number of predictions to be made
W desired operating condition
W_1, W_2 sub matrices of W
X matrix of independent variable settings
X_1, X_2 sub matrices of X
X_i input settings correspond to the i^{th} explanatory variables
y_1, y_2, y_3, y_4 response variables (concentration) in nonlinear batch reactor
y_1^0 initial concentration of component 1
Y vector of the measured response variables
Y_i measured response corresponding to the i^{th} run condition
Y_{True} noise-free response of the process
$\hat{Y}$ vector of the model prediction
$\hat{Y}_E$ vector of the extended model prediction
$\hat{Y}_S$ vector of the simplified model prediction
Z scaled sensitivity matrix
Z_1 and Z_2 sub matrices of Z
$\left[Q^+\right]$ initial catalyst concentration
β vector of all true model parameters
β_1, β_2 sub vectors of β
$\hat{\beta}_{2E}$ unbiased estimators of β_2 in the EM
γ correlation factor
$\delta\varphi_j$ scaled parameter deviation ($\delta\theta_j$)
$\delta\theta_j$ parameter deviation
ε stochastic component
ε_i stochastic component corresponding to the i^{th} run condition
θ p-dimensional vector of unknown parameters in nonlinear batch reactor model
θ_0 p-dimensional vector of nominal parameter values in nonlinear batch reactor model
λ non-centrality parameter in a non-central F distribution
ξ_i scaled residual corresponding to the i^{th} run condition
ρ_C quadratic norm of vector R_C
ρ_{CW} quadratic norm of vector R_{CW}
σ^2 noise variance
σ_y^2 noise variance of measured response variables

$\Omega = \left(X_2^T \left(I_n - P_1 \right) X_2 \right)^{-1}$ expression used in Equation(2.22)

Nomenclature for Chapter 3

$A_1 = \left(X_1^T X_1\right)^{-1} X_1^T X_2$ expression used in Equations (3.7) and (3.11)

$\mathrm{Cov}(\cdot)$ variance-covariance matrix

e stochastic component combined with any model mismatch

E subscript for extended model

$E(\cdot)$ expected value

i index for response variables

J objective function in parameter estimation

k iteration counter

l index for measurements

m index for experimental runs

M mapping matrix from X to W

MSE mean-squared error prediction

n total number of measurements

p total number of parameters in the EM

$P = X(X^T X)^{-1} X^T$ expression used in Equation (3.11)

p_1 number of parameters in the SM

$P_1 = X_1(X_1^T X_1)^{-1} X_1^T$ expression used in Equations (3.9) and (3.11)

R ideal gas constant

R_k residual matrix defined in Equation (T3.1.2)

r_C likelihood ratio statistic, estimator for R_C

R_C true Wu's critical ratio

r_{cKub} truncated Kubokawa estimator

r_{CC} estimator for R_{CC}

R_{CC} true Wu's corrected critical ratio

r_{CW} estimator for R_{CW}

R_{CW} true proposed critical ratio

r_{CCW} estimator for R_{CCW}

R_{CCW} true proposed corrected critical ratio

$s_{\theta_{j0}}$ uncertainties in the initial value of parameter θ_j

s_{y_i} uncertainties in the i^{th} measured response value

S subscript for simplified model

S_E^2 unbiased estimators σ^2 in the EM

S_{ij0} corresponding element in i^{th} row and j^{th} column of the sensitivity matrix calculated at $\theta=\theta_0$

SSE sum of squared residuals

T batch reactor temperature in K

T_0 reference temperature in K

$\mathrm{Tr}(\cdot)$ trace of a matrix

w total number of predictions to be made

W desired operating condition

W_1, W_2 sub matrices of W

X matrix of independent variable settings

X_1, X_2 sub matrices of X

X_i input settings correspond to the i^{th} explanatory variables

y_1, y_2, y_3, y_4 response variables (concentration) in nonlinear batch reactor

y_1^0 initial concentration of component 1

Y vector of the measured response variables

Y_i measured response corresponding to the i^{th} run condition

Y_{True} noise-free response of the process

$\hat{Y}$ vector of the model prediction

$\hat{Y}_E$ vector of the extended model prediction

$\hat{Y}_S$ vector of the simplified model prediction

Z scaled sensitivity matrix

Z_1 and Z_2 sub matrices of Z

$\left[Q^+\right]$ initial catalyst concentration

β vector of all true model parameters

β_1, β_2 sub vectors of β

$\hat{\beta}_{2E}$ unbiased estimators of β_2 in the EM

γ correlation factor

$\delta\varphi_j$ scaled parameter deviation ($\delta\theta_j$)

$\delta\theta_j$ parameter deviation

ε stochastic component

ε_i stochastic component corresponding to the i^{th} run condition

θ p-dimensional vector of unknown parameters in nonlinear batch reactor model

θ_0 p-dimensional vector of nominal parameter values in nonlinear batch reactor model

ξ_i scaled residual corresponding to the i^{th} run condition

σ^2 noise variance

σ_y^2 noise variance of measured response variables

$\Omega = \left(X_2^T \left(I_n - P_1\right) X_2\right)^{-1}$ expression used in Equations (3.7) and (3.11)

Nomenclature for Chapter 4

a	regression parameter in Equation (T4.3.12) ,m^{-1}
A	expression defined in Equation (Al4.4.2)
B	expression defined in Equation (Al4.4.3)
C	parameter for evaluating the dependency of viscosity on shear rate in Equation (T4.3.6), dimensionless
C_p	polymer specific heat, kJ $kg^{-1}K^{-1}$
$C_{p,a}$	air specific heat, kJ $kg^{-1}K^{-1}$
D	parameter to calculate crystallization kinetic in Equation (T4.3.8), K(or ° C)
DR	draw ratio, dimensionless
E_a	activation energy of Arrhenius law in Equation (T4.3.6), J mol^{-1}
f	parameter accounting for the effect of crystallinity on polymer viscosity in Equation (T4.3.6), dimensionless
F	draw force (the x-component of the force acting on polymer), N
g	gravity acceleration, 9.81 m s^{-2}
h	parameter accounting for the effect of crystallinity on polymer viscosity in Equation (T4.3.6), dimensionless
h_{forc}	forced convection heat transfer coefficient, kJ s^{-1} K^{-1} m^{-2}
h_{irr}	irradiation heat transfer coefficient, kJ s^{-1} K^{-1} m^{-2}
h_{nat}	natural convection heat transfer coefficient, kJ s^{-1} K^{-1} m^{-2}
h_{tot}	total convection heat transfer coefficient, kJ s^{-1} K^{-1} m^{-2}
j	parameter for calculating the natural convection heat transfer coefficient in Equation (T4.3.11), dimensionless
J	objective function in parameter estimation in Equation (Al4.2.1)
K_0	parameter for calculating crystallization kinetic in Equation (T4.3.8), s^{-1}
$K(T)$	Nakamura's constant defined in Equation (T4.3.8), s^{-1}
L_0	initial film width, m
$L(x)$	film width distribution along the draw direction, m
m	parameter accounting for the effect of crystallinity on polymer viscosity in Equation (T4.3.6), dimensionless
$\dot{m}$	mass flow rate of polymer, kg s^{-1}
M	mapping matrix from X to W defined
MSE	mean-squared error prediction
n	total number of measurements
n_A	Avrami's exponent in Equation (T4.3.4), dimensionless
n_C	parameter for evaluating the dependency of viscosity on shear rate in Equation (T4.3.6), dimensionless
p	total number of parameters in the EM
p_1	number of parameters in the SM
P	expression defined in Equations (Al4.4.5) and (Al4.4.9)

P_1 expression defined in Equations (Al4.4.6) and (Al4.4.10)

r_C likelihood ratio statistic, estimator for Wu's critical ratio

r_{cKub} truncated Kubokawa estimator

r_{CC} estimator for Wu's corrected critical ratio

r_{CCW} estimator for Eghtesadi's corrected critical ratio

R universal gas constant, 8.314 J mol^{-1} K^{-1}

R_k residual matrix defined in Equation (Al4.1.1)

$s_{\theta_{j0}}$ uncertainties in the initial value of parameter θ_j

s_{y_i} uncertainties in the i^{th} measured response value

$S(x)$ film thickness distribution along the draw direction, m

S_0 initial film thickness, m

S_{ij0} corresponding element in i^{th} row and j^{th} column of the sensitivity matrix calculated at $\theta=\theta_0$

SSE sum of square residuals

t time, s

T extrusion temperature, K (or ° C)

T_a ambient temperature, K (or ° C)

$T_{\max}$ parameter for calculating crystallization kinetic in Equation (T4.3.8), K (or ° C)

$Tr(\cdot)$ trace of a matrix

v_x x-component of velocity, m s^{-1}

v_{x0} initial value for x-component of velocity, m s^{-1}

v_{xX} final value (at $x=X_t$) for x-component of velocity, m s^{-1}

w total number of predictions to be made

W desired operating condition

x coordinate in draw direction, m

x_{FL} freezing line location, m

X matrix of independent variable settings

X_C volumetric degree of crystallinity, dimensionless

X_{eq} equilibrium degree of crystallinity, dimensionless

X_i input settings correspond to the i^{th} explanatory variables

X_t take-up distance, m

Y vector of the measured response variables

Y_i measured response corresponding to the i^{th} run condition

Z scaled sensitivity matrix

i index for response variables

k index for iteration counter

l index for measurements

m index for experimental runs

β	vector of all true model parameters
β_a	air volumetric expansion coefficient, K^{-1} (or $° C^{-1}$)
B	parameter for calculating the natural convection heat transfer coefficient in Equation (T4.3.11), Pa^{-1}
γ	correlation factor
$\dot{\gamma}$	rate of shear in Equation (T4.3.6), s^{-1}
ΔH	latent heat of crystallization for iPP, kJ kg^{-1}
ε	polymer emissivity, dimensionless
ε_i	stochastic component corresponding to the i^{th} run condition
θ	p-dimensional vector of unknown parameters in nonlinear batch reactor model
θ_0	p-dimensional vector of nominal parameter values in nonlinear batch reactor model
κ_a	air thermal conductivity, kJ s^{-1} K^{-1} m^{-1}
μ	material's viscosity, Pa s
μ_a	air dynamic viscosity, Pa s
$\mu_{0,r}$	reference viscosity in Equation (T4.3.6), Pa s
ν_a	air kinematic viscosity, m^2 s^{-1}
ξ_i	scaled residual corresponding to the i^{th} run condition
ρ	material's density, kg m^{-3}
ρ_a	air density, kg m^{-3}
σ	Stefan Boltzmanns constant, 5.67×10^{-8} J s^{-1} m^{-2} K^{-4}
σ^2	noise variance
Σ	variance-covariance matrix

Chapter 1 Introduction

1.1. Background

Fundamental mathematical models have a significant part in design, simulation, optimization and control of a vast range of chemical and biochemical processes, and improving the accuracy of the model predictions has been the centre of many research studies for the past few decades[1–9]. To obtain model predictions, it is often required to estimate unknown model parameters. However, it is sometimes not possible to estimate all of the parameters from the available data[10–12]. Noisy and scarce experimental data, or data from correlated experimental designs, impedes the parameter estimation step[13,14]. Besides, not all the model parameters affect the model predictions equally and independently. Some parameters have important influence on predictions of the data and others do not and, as a consequence, are more difficult to estimate. These parameter estimability issues can be addressed by either conducting additional experimental runs or measuring additional responses[15–17], or by changing the structure of the model equations via removing less influential parameters, lumping several parameters together into one single parameter, or fixing some of the model parameters at their nominal values[14,18–22]. Parameter subset selection techniques, therefore, are often used to decide which parameters to fix and which to estimate[7,18,23–25]. These techniques are important because estimating too few parameters may lead to significant bias in the predictions, while estimating too many parameters decreases model's predictive ability and leads to increased variances in the model predictions[7,26]. Simplified models (SMs) which consider fewer parameters for estimation than the corresponding extended model (EM) that contains all of the parameters are sometimes developed and used. In the research in this thesis, it is assumed that the EM is a model with a sufficiently detailed structure to return reliable model predictions over the entire range of operating region if reliable values of the model parameters are available. Estimating the subset of the model parameters that

belong to a SM, rather than all of the parameters in the corresponding EM, simplifies the estimation procedure and can lead to more accurate model predictions by avoiding over-fitting[18,23,25]. By admitting fewer parameters in a SM, the prediction variance decreases and the prediction bias increases[4,7,27]. Due to this trade-off between the bias and the variance, a SM gives better predictions than a corresponding EM when the variance reduction outweighs the bias increase[13].

Parameter subset selection techniques have been described in recent review articles[14,24,28]. Collinearity index methods[29], relative gain array methods[30], Hankel singular value methods[31], orthogonalization-based methods[20,32,33], and optimization-based methods (e.g., using *A*-, *D*- or *E*- optimality)[34] are examples of parameter subset selection techniques. Many of the parameter subset selection techniques produce a ranked list of parameters. Parameters that appear near the top of the ranked list are the most important ones and should be estimated from the available data to obtain good model predictions, while parameters near the bottom of the list are the less important ones which can be fixed either at zero or at their nominal values[20,35]. Some of parameter subset selection techniques suffer from arbitrary cut-off values[24,36], while other more reliable techniques such as cross-validation are computationally intensive[20]. The cut-off values are used to decide how many parameters to estimate and how many to fix. Optimization-based techniques, due to their combinatorial nature, are sometimes difficult or impossible to use as the number of parameters increases[14]. Therefore, less computationaly intensive methods such as forward selection and backward elimination[28,37], stochastic searches such as genetic algorithms[38], and heuristic reduction techniques such as parameter clustering[39] have been developed and used to find the sub-optimal solution of the parameter subset selection problems.

Orthogonalization-based techniques are among the most popular techniques for parameter ranking and selection, and have been used extensively in the model development phase of complex chemical processes such as polymerization reactors, bioreactors, fuel cells[14,36,40–45].

The orthogonalization method[32] and generalized orthogonalization method[46] may lead to suboptimal parameter subset selection because selection of a parameter at the current step is conditional on a selected parameter subset on the parameter ranked list. Orthogonalization techniques utilize the Fisher information matrix (FIM) to rank the model parameters. The FIM contains information about the influence of model parameters on the model predictions and measurement uncertainties and is used to compute the parameter variance-covariance matrix. Orthogonalization-based techniques select a parameter subset in a way to avoid rank deficiency of the FIM[18] so that numerical conditioning problems can be avoided during parameter estimation.

Obtaining accurate model predictions while avoiding parameter estimability problems motivated our research group to develop and test mean-squared-error (MSE)-based criteria for parameter subset selection. Wu et al.[24] proposed a MSE-based criterion, r_{CC}, to determine the number of parameters to estimate from the ranked list of parameters[13,25,47]. The MSE of model predictions is equivalent to the sum of the squared bias and the variance; thus, Wu's criterion accounts for the trade-off between bias and variance as additional parameters from this ranked list are estimated. Experimental data and least-squares parameter estimation are required when using Wu's MSE-based technique to choose the optimum number of parameters to estimate. Wu's method has recently been used to choose parameter subsets for estimation in several different polymerization models[48–51]. Wu's approach does not use any arbitrary cut-off value, rather, it uses a statistically motivated critical ratio to decide how many parameters to estimate and how many to fix[24,36].

Furthermore, his method is not as computationally expensive as other reliable methods such as cross-validation[20]. Wu's technique assumes that an EM (with relatively a large number of parameters) has already been derived from the modeler's prior knowledge, and the structure of this EM is perfect. However, if after estimating an appropriate parameter subset, poor model predictions have been obtained, the model structure needs to be revisited or extra experimental runs may be required. Initially, Wu et al. used their r_{CC} criterion on the ranked list of parameters, which was obtained by using orthogonalization technique, to determine the appropriate number of parameters for estimation. Later, McLean et al.[22] proposed a MSE-based forward selection technique established on the r_{CC} criterion to simultaneously rank the parameters and select the optimum number of parameters to estimate. McLean et al. used a variety of the initial parameters guesses to reduce the sensitivity of the FIM to the initial parameters guesses which can be problematic for models that are nonlinear in the parameters. McLean et al. also compared their MSE-based forward selection techniques to another MSE-based forward selection technique proposed by Chu et al.[18]. McLean et al showed that both of these procedures require considerably less effort than the leave-one-out cross-validation and are more reliable and computationally intensive than using orthogonalization combined with the Wu's MSE-based method[22].

1.2. Research Motivations

One drawback of the recently developed MSE-based techniques is that they minimize the MSE of predictions made at the experimental settings corresponding to the available data set to select the optimal parameter subset to estimate. Model users are often interested in acquiring accurate model predictions at operating conditions which are different from the settings where data are already available. This problem motivated the research in this thesis. The main objective of the

research is to propose a new MSE-based criterion, r_{CCW}, that considers the operating region where accurate model predictions are desired[52]. The second objective of this thesis is to test this new methodology to rank and select parameters for estimation in linear and nonlinear models. This approach is valuable to modelers who want to make predictions about new products or grades or at new operating conditions, using models that are fitted only by utilizing data which were obtained at prior operating conditions. Initially, we applied the new criterion on the ranked list of parameters which was derived from the orthogonalization-based parameter ranking method. The results of initial work are shown in Chapter 2. The issue with this approach is that the knowledge about the desired operating conditions is not considered during the ranking step in parameter subset selection. Therefore, a new forward-selection methodology is developed in Chapter 3 based on using the r_{CCW} MSE-based criterion to simultaneously rank and select the parameters while considering the conditions in the desired operating region[53]. This technique ranks the parameters and selects the optimal number of parameters to include in the SM to minimize the anticipated MSE of the model predictions at the desired operating conditions specified by the model users. To utilize the r_{CCW} criterion to rank and select the parameters for estimation, the FIM must be invertible. Therefore, the r_{CCW} criterion needed to be extended to the cases where the FIM is not invertible. A singular FIM has been reported in many industrial chemical engineering models[20,48,49,54,55], where it is impossible to rank all of the model parameters using orthogonalization-based techniques. A singular FIM occurs in many high-dimensional parameter estimation problems[56,57] and in over-parameterized models[58]. A singular FIM also leads to a significant complication in the analysis of the parameter estimation problems[59]. The problem with a singular FIM is often handled by the use of a pseudoinverse of the FIM[56,60]. In this thesis, two different approaches for parameter ranking and selection are

proposed in Chapter 4 and compared when the FIM is not invertible. The first approach uses the orthogonalization algorithm to find and fix (at nominal values) the least important parameters in order to avoid rank deficiency (i.e., ill conditioning) of the FIM. The second approach uses the Moore-Penrose pseudoinverse of the FIM instead of the inverse of the FIM in the r_{CCW} criterion calculation procedure.

1.3. Research Outline

In Chapter 2 of this thesis, first, some relevant mathematical notation and definitions are presented. Later, Wu's MSE-based model selection criterion, which selects the best SM to obtain accurate predictions at available experimental setting, is summarized. Next, the methodology is extended to take the choice of the desired operating conditions into account when selecting the best SM. The theory and mathematical derivation of the r_{CCW} criterion is illustrated using a linear regression model and then extended for the use in the nonlinear cases. A linear regression example and a nonlinear batch reactor example are used to demonstrate and test the r_{CCW} criterion. In Chapter 3, a MSE-based forward selection methodology established on the r_{CCW} criterion is proposed for simultaneous parameter ranking and selection. This methodology is then described and demonstrated using the same linear and the nonlinear case studies used in Chapter 2. In Chapter 4, the problem of noninvertible FIMs is investigated. Two different approaches for parameter ranking and selection are proposed when the FIM is not invertible. Using the known true parameter values in a linear regression example, the true values of the MSE model predictions are calculated, and the efficiency of these two approaches are compared. Next, the better of the two approaches is extended for the use in nonlinear models and is illustrated using a nonlinear polymer film casting model. Lastly, in Chapter 5, conclusions of the thesis and the recommendations for the future work are presented.

1.4. References

(1) Maria, G. A Review of Algorithms and Trends in Kinetic Model Identification for Chemical and Biochemical Systems. **2004**, *18*, 195–222.

(2) Foss, B. a.; Lohmann, B.; Marquardt, W. A field study of the industrial modeling process. *Model. Identif. Control A Nor. Res. Bull.* **1998**, *19*, 153–174.

(3) Shmueli, G.; Koppius, O. R. Predictive Analytics in Information Systems Research. *Mis Q.* **2011**, *35*, 553–572.

(4) Shmueli, G. To Explain or to Predict? *Stat. Sci.* **2010**, *25*, 289–310.

(5) McLean, K. A. P. Obtaining the Best Model Predictions and Parameter Estimates Using Limited Data, 2011.

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

(7) Hocking, R. R. The Analysis and Selection of Variables in Linear Regression. *Biometrics* **1976**, *32*, pp. 1–49.

(8) Ehrenberg, A.; Bound, J. Predictability and prediction. *J. Roy. Stat. Soc. Ser. A* **1993**, *156*, 167–206.

(9) Geisser, S. The predictive sample reuse method with applications. *J. Amer. Stat. Assoc.* **1975**, *70*, 320–328.

(10) Perregaard, J. Model simplification and reduction for simulation and optimization of chemical processes. *Comput. Chem. Eng.* **1993**, *17*, 465–483.

(11) Wang, F. Y.; Zhu, Z. H.; Massarotto, P.; Rudolph, V. A simplified dynamic model for accelerated methane residual recovery from coals. *Chem. Eng. Sci.* **2007**, *62*, 3268–3275.

(12) Weijers, S. R., & Vanrolleghem, P. A. A procedure for selecting best identifiable parameters in calibrating activated sludge model no 1 to full-scale plant data. *Water Sci. Technol.* **1997**, *36*, 69–79.

(13) Wu, S.; Harris, T. J.; McAuley, K. B. The use of simplified or misspecified models: Linear case. *Can. J. Chem. Eng.* **2007**, *85*, 386–398.

(14) Kravaris, C.; Hahn, J.; Chu, Y. Advances and selected recent developments in state and parameter estimation. *Comput. Chem. Eng.* **2013**, *51*, 111–123.

(15) Dochain, D.; Vanrolleghem, P. A. *Dynamical Modelling and Estimation inWastewater Treatment Processes*; IWA Publishing: London, UK, 2001.

(16) Benyahia, B.; Latifi, M. A.; Fonteix, C.; Pla, F. *Modeling of a batch emulsion copolymerization reactor in the presence of a chain transfer agent : estimability analysis , parameters identification and experimental validation* .; Elsevier B.V., 2011; Vol. 29, pp. 126–130.

(17) Ben-Zvi, A.; McAuley, K.; McLellan, J. Identifiability study of a liquid-liquid phase-transfer catalyzed reaction system. *AIChE J.* **2004**, *50*, 2493–2501.

(18) Chu, Y.; Huang, Z.; Hahn, J. Improving prediction capabilities of complex dynamic models via parameter selection and estimation. *Chem. Eng. Sci.* **2009**, *64*, 4178–4185.

(19) Quaiser, T.; Monnigmann, M. Systematic Identifiability Testing for Unambiguous Mechanistic Modeling—Application to JAK-STAT, MAP Kinase, and NF-kB Signaling Pathway Models. *BMC Syst. Biol* **2009**, *3*.

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

(21) Lund, B. F.; Foss, B. a. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. *Automatica* **2008**, *44*, 278–281.

(22) McLean, K. A. P.; Wu, S.; McAuley, K. B. Mean-Squared-Error Methods for Selecting Optimal Parameter Subsets for Estimation. *Industrial & Engineering Chemistry Research*, 2012, *51*, 6105.

(23) Moore, C.; Doherty, J. Role of the calibration process in reducing model predictive error. *Water Resour. Res.* **2005**, *41*, W05020.

(24) McLean, K. A. P.; McAuley, K. B. Mathematical modelling of chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Can. J. Chem. Eng.* **2011**, *90*, 351–366.

(25) Wu, S.; McAuley, K. B.; Harris, T. J. Selection of simplified models: II. Development of a model selection criterion based on mean squared error. *Can. J. Chem. Eng.* **2011**, *89*, 325–336.

(26) Hawkins, D. M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12.

(27) Rao, P. Some Notes on Misspecification in Multiple Regressions. *Am. Stat.* **1971**, *25*, 37–39.

(28) Miller, A. J. *Subset Selection in Regression*; Chapman and Hall: New York, 1990.

(29) Brun, R., Reichert, P., & Kunsch, H. R. Practical identifiability analysis of large environmental simulation models. *Water Resour. Res.* **2001**, *37*, 1015–1030.

(30) Sandink, C. a.; McAuley, K. B.; McLellan, P. J. Selection of Parameters for Updating in On-line Models. *Ind. Eng. Chem. Res.* **2001**, *40*, 3936–3950.

(31) Sun, C. L., & Hahn, J. Parameter reduction for stable dynamical systems based on Hankel singular values and sensitivity analysis. *Chem. Eng. Sci.* **2006**, *61*, 5393–5403.

(32) Yao, K. Z.; Shaw, B. M.; Kou, B.; McAuley, K. B.; Bacon, D. W. Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimatibility analysis and experimental design. *Polym. React. Eng.* **2003**, *11*, 563–588.

(33) Lund, B. F.; Foss, B. a. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. *Automatica* **2008**, *44*, 278–281.

(34) Walter, E., & Pronzato, L. *Identification of parametric models from experimental data*; Springer, 1997.

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

(36) Jayasankar, B. R.; Ben-Zvi, A.; Huang, B. Identifiability and estimability study for a dynamic solid oxide fuel cell model RID F-2736-2010. *Comput. Chem. Eng.* **2009**, *33*, 484–492.

(37) Blanchet, F. G., Legendre, P., & Borcard, D. Forward selection of explanatory variables. *Ecology* **2008**, *89*, 2623–2632.

(38) Chu, Y.; Hahn, J.; Chu, Y. F., & Hahn, J. Parameter set selection for estimation of nonlinear dynamic systems. *AIChE J.* **2007**, *53*, 2858–2870.

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

(40) Lopez, D. C. C.; Barz, T.; Penuela, M.; Villegas, A.; Ochoa, S.; Wozny, G. Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production. *Biotechnol. Prog.* **2013**, *29*, 1064–1082.

(41) Ngo, V. V; Latifi, M. A.; Simonnot, M.-O. Estimability Analysis and Optimisation of Soil Hydraulic Parameters from Field Lysimeter Data. *Transp. POROUS MEDIA* **2013**, *98*, 485–504.

(42) Meshram, M.; Naderi, S.; McConkey, B.; Ingalls, B.; Scharer, J.; Budman, H. Modeling the coupled extracellular and intracellular environments in mammalian cell culture. *Metab. Eng.* **2013**, *19*, 57–68.

(43) Balakrishnan, N. P.; Samavedham, L.; Rangaiah, G. P. Personalized mechanistic models for exercise, meal and insulin interventions in children and adolescents with type 1 diabetes. *J. Theor. Biol.* **2014**, *357*, 62–73.

(44) Ngo, V. V.; Gerke, H. H.; Badorreck, A. Estimability Analysis for Optimization of Hysteretic Soil Hydraulic Parameters Using Data of a Field Irrigation Experiment. *Transp. Porous Media* **2014**, *103*, 535–562.

(45) SJ, Y.; JH, K.; JM, L. Dynamic modelling of mixotrophic microalgal photobioreactor systems with time-varying yield coefficient for the lipid consumption. *Bioresour. Technol.* **2014**, *162*, 228–235.

(46) Chu, Y.; Hahn, J. Generalization of a Parameter Set Selection Procedure Based on Orthogonal Projections and the D-Optimality Criterion. *AIChE J.* **2012**, *58*, 2085–2096.

(47) Wu, S.; McAuley, K. B.; Harris, T. J. Selection of Simplified Models: I. Analysis of Model-Selection Criteria using Mean-Squared Error. *Can. J. Chem. Eng.* **2011**, *89*, 148–158.

(48) Woloszyn, J. D.; McAuley, K. B. Application of Parameter Selection and Estimation Techniques in a Thermal Styrene Polymerisation Model. *Macromol. React. Eng.* **2011**, *5*, 453–466.

(49) Karimi, H.; Schaffer, M. A.; McAuley, K. B. Kinetic Moded for Non-Oxidative Thermal Degradation of Nylon 66. *Macromol. React. Eng.* **2012**, *6*, 93–109.

(50) Cui, W. J.; McAuley, K. B.; Whitney, R. A.; Spence, R. E.; Xie, T. Mathematical Model of Polyether Production From 1,3-Propanediol. *Macromol. React. Eng.* **2013**, *7*, 237–253.

(51) Zhao, Y. R.; McAuley, K. B.; Puskas, J. E.; Dos Santos, L. M.; Alvarez, A. Mathematical Modeling of Arborescent Polyisobutylene Production in Batch Reactors. *Macromol. THEORY SIMULATIONS* **2013**, *22*, 155–173.

(52) Eghtesadi, Z.; Wu, S.; McAuley, K. B. Development of a Model Selection Criterion for Accurate Model Predictions at Desired Operating Conditions. *Ind. Eng. Chem. Res.* **2013**, *52*, 12297–12308.

(53) Eghtesadi, Z.; McAuley, K. B. Mean Square Error Based Method for Parameter Ranking and Selection To Obtain Accurate Predictions at Specified Operating Conditions. *Ind. Eng. Chem. Res.* **2014**, *38*.

(54) Kou, B.; McAuley, K. B.; Hsu, C. C.; Bacon, D. W. Mathematical model and parameter estimation for gas-phase ethylene/hexene copolymerization with supported metallocene catalyst. *Macromol. Mater. Eng.* **2005**, *290*, 537–555.

(55) Cui, W. J.; McAuley, K. B.; Whitney, R. A.; Spence, R. E.; Xie, T. Mathematical Model of Polyether Production From 1,3-Propanediol. *Macromol. React. Eng.* **2013**, *7*, 237–253.

(56) Hero, A. O.; Fessler, J. A.; Usman, M. Exploring estimator bias- variance tradeoffs using the uniform CR bound. *IEEE Trans. Signal Process.* **1996**, *44*, 2026–2041.

(57) Sjöberg, J.; Zhang, Q.; Ljung, L.; Benveniste, A.; Delyon, B.; Rennec, P.-Y. G.-; Hjalmarsson, H.; Judisky, A. Nonlinear black-box modeling in system identification: A unified overview. *Automatica* **1995**, *31*, 1691–1724.

(58) Stoica, P.; Söderström, T. On non-singular information matrices and local identifiability. *Int. J. Control* **1982**, *36*, 323–329.

(59) Stoica, P.; Marzetta, T. L. Parameter Estimation Problems with Singular Information Matrices. *IEEE Trans. Signal Process.* **2001**, *49*, 87–90.

(60) Rao, C. R. *Linear Statistical Inference and Its Applications*; 2nd Editio.; Wiley: New York, 1973.

Chapter 2 Development of a Model Selection Criterion for Accurate Model Predictions at Desired Operating Conditions

2.1. Summary

A methodology is proposed for selecting parameters to estimate when data are too limited to estimate all kinetic, thermodynamic and mass-transfer parameters in complex models of chemical processes. When data are sparse, noisy or correlated, it is often better to obtain predictions from a simplified model (SM) where a few parameters have been removed via simplifying assumptions or some parameters are fixed at nominal values based on prior knowledge. Reducing the number of estimated parameters leads to bias in model predictions, but also lowers prediction variance. Tradeoff between bias and variance is assessed using the mean-squared error (MSE) of the model predictions. The proposed model selection criterion is an advance over previous criteria in the literature because arbitrary tuning parameters are not required, computations are relatively simple, and the user can specify key operating conditions where accurate predictions are desired. Important benefits are that over-fitting of noisy data is prevented and standard least-squares parameter estimation can be used without numerical difficulties. Monte Carlo simulations are used to assess the effectiveness of the proposed methodology for parameter selection in linear and nonlinear models. This approach will be valuable for industrial modelers who want to make accurate predictions about new product specifications or grades.

This chapter was published in Industrial and Engineering Chemistry Research in vol. 52 pp. 12297-12308 with S Wu and K B McAuley as coauthors.

2.2. Introduction

Fundamental models play a significant role in simulation, design, control and optimization of industrial processes. A comprehensive model, which will be called an extended model (EM) in this chapter, is a model with a sufficiently detailed structure so that the model equations are assumed capable of providing reliable predictions over the range of operating conditions of interest, if appropriate values of the model parameters are available[1]. In many cases, EMs contain a large number of parameters that need to be estimated from experimental measurements; however, it is often too expensive or time-consuming to obtain sufficient data to estimate all of the parameters reliably[2, 3]. Accurate estimation of the large number of parameters in an EM is hampered by data sets with noisy measurements, only a few measurements, or with correlated design matrices. Large uncertainties in parameter estimates in an EM will cause large variances in model predictions. Therefore simplified models (SMs), which have a smaller number of parameters, are sometimes developed and used. Adopting a SM, rather than the EM, simplifies the estimation procedure and can lead to more accurate model predictions, while avoiding over-fitting[4-6].

One way to derive a SM is by fixing some of the parameters in the corresponding EM, either at zero or at other nominal values, and then estimating the remaining parameters[5]. Deciding which parameters to fix and which to estimate is referred as parameter subset selection[4-8]. Methods in the literature for selecting appropriate SMs or appropriate parameter subsets to estimate have been described in recent review articles[7, 9]. Several methods produce a ranked list of parameters, with the parameters that appear near the top being parameters that require estimation so that good predictions can be obtained, and parameters near the bottom being parameters that are less important[4, 10, 11]. Some of these methods suffer from arbitrary cut-off values that are used to

determine how many parameters to estimate and how many to fix[7, 12], whereas other more reliable methods such as cross-validation are computationally expensive, since they involve evaluating the predictive power of a model by repeatedly selecting small subsets of the data as evaluation data and then re-estimating the parameters based on the remaining data[10]. This problem motivated Wu et al. to propose a criterion based on mean square error (MSE) to determine the number of parameters to estimate from the ranked list of parameters[1, 5, 13]. MSE is the sum of squared bias and variance and therefore accounts for the trade-off between bias and variance. As more parameters are estimated from the ranked list, the bias decreases and variance increases. A SM gives better predictions than the corresponding EM when the variance reduction outweighs the bias. Wu's approach, like many other ranking techniques, requires experimental data and least-squares parameter estimation to choose the optimum number of parameters to estimate. This methodology has been used recently[14, 15] to choose subsets parameters to estimate in two polymerization reactor models. One problem with Wu's approach is that it uses the MSE at the data points used for parameter estimation to select the optimal number of parameters to estimate. Users are often interested in obtaining accurate predictions at operating conditions that are different from the points at which data are already available. In this chapter, an enhanced selection methodology is proposed, based on MSE at the operating conditions of interest to the model user. The focus is on obtaining accurate model predictions rather than accurate parameter estimates[16, 17].

The ability of a model to predict new observations that were not used for model building is called predictive power or predictive accuracy[17]. Improving predictive power often calls for permitting some bias to achieve lower variance[9]. Predictive power has been assessed using out-of-sample prediction accuracy methods such as data mining, cross-validation, bootstrapping and

external validation techniques[10, 18-21]. Lower predictive power can result from over-fitting, i.e., estimating too many parameters using the available data[19]. Wu's MSE-based criterion and the new criterion proposed in this chapter are different from traditional statistical approaches to model building, wherein models with a sufficient number of terms are identified so that the models can pass statistical adequacy tests without over-fitting the data[22]. In these traditional approaches, the analyst uses experimental data to determine the appropriate model structure. However, the current methodology assumes that the modeler has already derived a model (with a large number of parameters) using prior chemical and physical knowledge, and that the structure of this model is perfect. The problem faced by the modeler is then to decide which parameters should be estimated to obtain the best possible predictions, knowing that both bias and variance will be important. After the appropriate parameters have been estimated, the modeler should then compare the model predictions and the data to determine whether the predictions are good enough that the model can be used for its intended purpose. If not, the model structure may need to be revised or additional experiments may be required.

In this chapter, we summarize the MSE-based model selection criterion developed by Wu et al., which selects the best SM to obtain good predictions at experimental setting used for parameter estimation. Next, the theory is extended to permit selection of the best SM that will provide the best predictions at key operating conditions specified by the modeler. These conditions may be very different from conditions available for model fitting. This new approach will be valuable to industrial modelers who want to make predictions about new product specifications or grades, using models that are fitted using data obtained at previous operating conditions. The method will also help modelers who want to tune parameters to obtain accurate predictions from a SM that will be used for predicting process behavior over a limited range of conditions, rather than

over the whole range of settings available in the training data set. Finally, two examples, one linear and one nonlinear, are used to test and demonstrate the proposed methodology[5, 23].

2.3. Summary of Wu's MSE-based model selection criterion

Wu's MSE-based model selection criterion was developed by considering the following single-response extended model (EM) which is linear in the parameters[1]:

$$\begin{aligned} Y &= X\beta + \varepsilon \\ &= X_1\beta_1 + X_2\beta_2 + \varepsilon \end{aligned} \tag{2.1}$$

where $Y \in R^n$ denotes a vector of measured responses, $X \in R^{n\times p}$ is a matrix of independent-variable settings, $X_1 \in R^{n\times p_1}$, $X_2 \in R^{n\times(p-p_1)}$ are submatrices that contain the first p_1 and remaining p-p_1 columns of X, respectively. $\beta \in R^p$ is a vector of all unknown model parameters, $\beta_1 \in R^{p_1}$, $\beta_2 \in R^{p-p_1}$ are sub vectors of β, and $\varepsilon \in R^n$ is a vector of additive random noise terms, where n is the total number of measurements, p is total number of parameters in the EM, p_1 is the number of parameters in the SM and $p - p_1$ is the number of parameters excluded from the SM. Therefore, the true or noise-free response is $Y_{True} = X\beta$. Wu et al. made the following least-squares assumptions[1]: X_1 and X_2 are deterministic and the stochastic component ε is a mean-zero, independent random sequence with constant variance σ^2. If σ^2 is not constant for the original response variable, then a transformation can be used to determine a new response variable with constant σ^2 [24]. A particular simplified model (SM) is of the form:

$$Y = X_1\beta_1 + e \tag{2.2}$$

where $e = X_2\beta_2 + \varepsilon$ is the stochastic component combined with any model mismatch. Note that

SM is nested within the EM.

The MSE is defined as the expected squared difference between the model prediction ($\hat{Y}$) and the noise-free response of the process (Y_{True})[25]:

$$\begin{aligned} MSE(\hat{Y}) &= E\left((\hat{Y}-Y_{True})^T(\hat{Y}-Y_{True})\right) \\ &= \left(E(\hat{Y})-Y_{True}\right)^T\left(E(\hat{Y})-Y_{True}\right)+Tr\left(Cov(\hat{Y})\right) \end{aligned} \tag{2.3}$$

where $E(\cdot)$, $Cov(\cdot)$ and $Tr(.)$ denote the expected value, variance-covariance matrix and trace, respectively. The squared bias is $\left(E(\hat{Y})-Y_{True}\right)^T\left(E(\hat{Y})-Y_{True}\right)$ and the total prediction variance is $Tr\left(Cov(\hat{Y})\right)$.

Wu et al.[1, 5, 13] proposed a strategy based on a critical ratio R_C to find whether a particular SM or the EM gives better predictions (lower MSE). This critical ratio is defined as:

$$R_C = \frac{\beta_2^T X_2^T \left(I_n - P_1\right) X_2 \beta_2}{(p-p_1)\sigma^2} \tag{2.4}$$

where

$$P_1 = X_1(X_1^T X_1)^{-1} X_1^T \tag{2.5}$$

and σ^2 is the noise variance. R_C is the ratio of squared bias introduced by removing β_2 from the EM over the variance reduction when the SM is used rather than the EM to make predictions at the experimental settings X [1]. If $R_C < 1$, the SM will give better predictions than the EM. Later, Wu et al. proposed a corrected critical ratio, R_{CC}, to compare several different SMs:

$$R_{CC} = \frac{\left(MSE(\hat{Y}_S) - MSE(\hat{Y}_E)\right)/n}{\sigma^2} = \frac{p - p_1}{n}\left(R_C - 1\right) \tag{2.6}$$

where R_C is defined in Equation (2.4) and MSE is the mean-squared error, and subscript "E" and "S" indicate the extended model and simplified model, respectively. The SM from the set of candidate SMs that gives the lowest MSE predictions correspond to the smallest value of R_{CC} [5]. R_{CC} is useful for comparing several SMs with different number of parameters. The model corresponding to the lowest value of R_{CC} is expected to give the best predictions. An estimate for R_{CC} can be obtained using the following estimator for R_C [1, 26]:

$$r_C = \frac{\hat{\beta}_{2E}^T X_2^T \left(I_n - P_1\right) X_2 \hat{\beta}_{2E}}{(p - p_1) S_E^2} = \left(J_{p_1} - J_p\right)/\left(p - p_1\right) \tag{2.7}$$

where $\hat{\beta}_{2E}$ and S_E^2 are parameter estimates and the variance estimate, respectively, which are unbiased estimates of β_2 and σ^2. J_{p_1} and J_p are values of the least-squares objective function when p_1 and p parameters are estimated, respectively. In Equation (2.7) the structure of the EM must be known to calculate r_C. For a multi-response model, the optimal value of the weighted least-squares objective function *J*, which is a function of the parameter vector (θ), is:

$$J(\theta) = \sum_{i=1}^{d} \sum_{l=1}^{c} \sum_{m=1}^{r} \left(\frac{Y_{ilm} - \hat{Y}_{ilm}(\theta)}{s_{y_i}} \right)^2 \tag{2.8}$$

where Y_{ilm} is the measured value of the i^{th} response variable at the l^{th} measurement time, for the m^{th} experimental run and $\hat{Y}_{ilm}$ is the corresponding predicted value. s_{y_i} is the uncertainty associated with the i^{th} measured response variable[26].

Note that Wu et al. used the notation $\hat{R}_C$ instead of r_C for the estimator in their earlier work[5], and then changed the notation to r_C[27]. When the noise variance is unknown, r_C in Equation (2.7) is a likelihood ratio statistic with a non-central F distribution with $(p-p_1)$ and $(n-p)$ degrees of freedom and non-centrality parameter $\lambda=(p-p_1)R_C$, given the additional assumption that ε is normally distributed. Alternatively, in situations where the noise variance, σ^2, is assumed to be known, the estimator for r_C can be obtained from:

$$r_C=\frac{\hat{\beta}_{2E}^T X_2^T\left(I_n-P_1\right)X_2\hat{\beta}_{2E}}{(p-p_1)\sigma^2} \tag{2.9}$$

where $(p-p_1)r_C$ follows a noncentral χ^2 distribution with $(p-p_1)$ degrees of freedom and noncentrality parameter $\lambda=(p-p_1)R_C$.To Account for the number of parameters in different SMs, Wu et al. suggested that modelers should use the corrected critical ratio r_{CC} to select the best model:

$$r_{CC}=\frac{p-p_1}{n}(r_C-1) \tag{2.10}$$

The best SM for making predictions at the input settings in X will have the lowest value of r_{CC}.

2.4. Proposed MSE-based model selection criterion for selecting the best model for predictions at user-specified operating conditions

The model user is often interested at making accurate predictions at different input settings than the setting in X. If the modeler wants accurate predictions corresponding to a different input matrix $W \in R^{w \times p}$, then the SM selected using r_{CC} may not be the best choice. Here we develop new model selection criterion aimed at selecting the best SM to give predictions for the settings in W. W can be partitioned in the same way as X into $W_1 \in R^{w \times p_1}$ and $W_2 \in R^{w \times (p-p_1)}$, where w is the total number of predictions to be made. The corresponding simplified and extended model predictions are:

$$\hat{Y}_S = W_1 \hat{\beta}_{1S} \tag{2.11}$$

$$\hat{Y}_E = W \hat{\beta}_E = W_1 \hat{\beta}_{1E} + W_2 \hat{\beta}_{2E} \tag{2.12}$$

To derive the new MSE-based model selection criterion, we follow a similar approach as that described in the previous section. Given the trade-off between bias and variance, we desire a criterion to select the best SM with the lowest total MSE. When $W \neq X$, the expected values and covariance matrices of predictions for the SM and EM are[1]:

$$E\left(\hat{Y}_S\right) = E\left(W_1 \hat{\beta}_S\right) = W_1 \beta_1 + W_1 A_1 \beta_2 \tag{2.13}$$

$$E\left(\hat{Y}_E\right) = E\left(W \hat{\beta}_E\right) = W_1 \beta_1 + W_2 \beta_2 \tag{2.14}$$

where

$$A_1 = \left(X_1^T X_1\right)^{-1} X_1^T X_2 \tag{2.15}$$

$$Cov\left(\hat{Y}_S\right) = \sigma^2 W_1 \left(X_1^T X_1\right)^{-1} W_1^T \tag{2.16}$$

assuming that the measurement variance will be σ^2 for any new measurement that could be made at settings corresponding to *W*.

$$Cov\left(\hat{Y}_E\right) = \sigma^2 W_1 \left(X_1^T X_1\right)^{-1} W_1^T + \sigma^2 \left(W_1 A_1 - W_2\right) \Omega \left(W_1 A_1 - W_2\right)^T \tag{2.17}$$

The mean-squared error matrices for the SM and EM are:

$$MSEM(\hat{Y}_S) = \sigma^2 W_1 (X_1^T X_1)^{-1} W_1^T + (W_1 A_1 - W_2)\beta_2 \beta_2^T (W_1 A_1 - W_2)^T \tag{2.18}$$

$$MSEM(\hat{Y}_E) = \sigma^2 W_1 (X_1^T X_1)^{-1} W_1^T + \sigma^2 (W_1 A_1 - W_2)\Omega (W_1 A_1 - W_2)^T \tag{2.19}$$

which lead to the following total mean-squared errors:

$$\begin{aligned} MSE(\hat{Y}_S) &= \sigma^2 Tr(W_1 (X_1^T X_1)^{-1} W_1^T) + Tr((W_1 A_1 - W_2)\beta_2 \beta_2^T (W_1 A_1 - W_2)^T) \\ &= \sigma^2 Tr(W_1 (X_1^T X_1)^{-1} W_1^T) + Tr(\beta_2^T (W_1 A_1 - W_2)^T (W_1 A_1 - W_2)\beta_2) \\ &= \sigma^2 Tr(W_1 (X_1^T X_1)^{-1} W_1^T) + (\beta_2^T (W_1 A_1 - W_2)^T (W_1 A_1 - W_2)\beta_2) \end{aligned} \tag{2.20}$$

$$MSE(\hat{Y}_E) = \sigma^2 Tr(W_1 (X_1^T X_1)^{-1} W_1^T) + \sigma^2 Tr((W_1 A_1 - W_2)\Omega (W_1 A_1 - W_2)^T) \tag{2.21}$$

where

$$\Omega = \left(X_2^T \left(I_n - P_1\right) X_2\right)^{-1} \tag{2.22}$$

When the SM is used for making predictions, the squared bias introduced is:

$$\begin{aligned} \text{bias}^2 \text{ increase} &= (W_1 A_1 \beta_2 - W_2 \beta_2)^T (W_1 A_1 \beta_2 - W_2 \beta_2) \\ &= \beta_2^T \left(W_1 A_1 - W_2\right)^T \left(W_1 A_1 - W_2\right) \beta_2 \end{aligned} \tag{2.23}$$

The corresponding decrease in variance due to the smaller number of parameters in the SM is:

$$\text{var decrease} = \sigma^2 Tr\left(\left(W_1A_1 - W_2\right)\Omega\left(W_1A_1 - W_2\right)^T\right) \tag{2.24}$$

The following critical ratio, which is the squared bias over the variance decrease, is analogous to the expression for R_C in Equation (2.4):

$$R_{CW} = \frac{\text{bias}^2 \text{ increase}}{\text{var decrease}} = \frac{\beta_2^T \left(W_1A_1 - W_2\right)^T \left(W_1A_1 - W_2\right)\beta_2}{\sigma^2 Tr\left(\left(W_1A_1 - W_2\right)\Omega\left(W_1A_1 - W_2\right)^T\right)} \tag{2.25}$$

When $W = X$, Equation (2.25) reduces to Equation (2.4)(See Appendix A1 for the proof) so that $R_{CW} = R_C$. Other special cases in which $R_{CW} = R_C$ are when $W = kX$ and k is any scalar and when $p - p_1 = 1$(See Appendix A1). Based on Equation(2.25), if

$$R_{CW} < 1 \tag{2.26}$$

the benefit of variance reduction outweighs the disadvantage of bias increase, so that, the SM will give a smaller value of total mean-squared prediction error (MSPE) than the EM, for predictions made using the settings in W. Inequality (2.26) is a necessary and sufficient condition for the SM being better than the EM in the sense of MSPE.

Unfortunately, the expressions in Equations (2.4) and (2.25) are based on the true values of the parameters and the true noise variance, which are often not available; therefore these true values must be estimated from the available data, assuming that the EM is correctly structured.

2.5. Estimation of R_{CW}

The true value of R_{CW} depends on the true values of the parameters, β_2, and the true noise variance, σ^2, according to Equation (2.25), which unfortunately, are usually not known. However, assuming the EM is the true model, these values can be estimated from the available data set. Replacing the true parameter values and noise variance in Equation (2.25) with estimates from the EM, obtained using X, we have:

$$r_{CW} = \frac{\hat{\beta}_{2E}^T \left(W_1 A_1 - W_2\right)^T \left(W_1 A_1 - W_2\right) \hat{\beta}_{2E}}{s_E^2 Tr\left(\left(W_1 A_1 - W_2\right) \Omega \left(W_1 A_1 - W_2\right)^T\right)} \tag{2.27}$$

where

$$\hat{\beta}_{2E} = \Omega X_2^T \left(I_n - P_1\right) Y \tag{2.28}$$

In situations where the noise variance, σ^2, is assumed to be known, the estimate for R_{CW} can be obtained from:

$$r_{CW} = \frac{\hat{\beta}_{2E}^T \left(W_1 A_1 - W_2\right)^T \left(W_1 A_1 - W_2\right) \hat{\beta}_{2E}}{\sigma^2 Tr\left(\left(W_1 A_1 - W_2\right) \Omega \left(W_1 A_1 - W_2\right)^T\right)} \tag{2.29}$$

By substituting Equation (2.28) into Equation (2.27) we obtain an alternative expression for r_{CW} in terms of the data (see Appendix A3):

$$r_{CW} = \frac{Y^T AY / \left(p - p_1\right)}{Y^T BY / \left(n - p\right)} \tag{2.30}$$

where

$$A = \frac{(p - p_1)(P - P_1)M^T M(P - P_1)}{Tr\left((P - P_1)M^T M(P - P_1)\right)} \tag{2.31}$$

$$B = I_n - P \tag{2.32}$$

$$P = X(X^T X)^{-1} X^T \tag{2.33}$$

and $M_{w \times n}$ is the mapping matrix from X to W:

$$M = W\left(X^T X\right)^{-1} X^T \tag{2.34}$$

Readers can refer to Appendix A2 for more explanation about the mapping matrix. As shown in Equations (2.31) and (2.32), matrices A and B do not depend on any parameter values, but rather they depend only on the W and X matrices. Modelers may prefer to use Equation (2.30) rather than Equation (2.27) to compute r_{CW} because Equation (2.30) is expressed in terms of the experimental measurements Y rather than estimates of the excluded parameters $\hat{\beta}_2$.

The estimator r_{CW} shown in Equations(2.27), (2.29) and (2.30) is useful because it can assist a modeler in determining whether a SM should give better predictions than an EM. If $r_{CW} < 1$then the SM should give better predictions than the corresponding EM in the sense of MSPE at the key operating conditions, W.

2.6. Corrected Critical Ratio R_{CCW} for comparing SMs

Following a similar idea used to derive Wu's corrected critical ratio, R_{CC}, here we derive a similar MSE-based criterion for choosing the best SM among a set of candidate SMs with different numbers of parameters. Similar to Equation (2.6) for R_{CC}, R_{CCW} is defined as:

$$R_{CCW} = \frac{\left(MSE(\hat{Y}_S) - MSE(\hat{Y}_E)\right)/w}{\sigma^2} = \frac{Tr\left((W_1A_1 - W_2)\Omega(W_1A_1 - W_2)^T\right)}{w}(R_{CW} - 1) \tag{2.35}$$

By substituting for W_1 and W_2 from Equation(A2.7), we have:

$$R_{CCW} = \frac{Tr\left(M^T M\left(I - P_1\right)X_2\Omega X_2^T\left(I - P_1\right)^T\right)}{w}(R_{CW} - 1) = \frac{Tr\left(M^T M\left(P - P_1\right)\right)}{w}(R_{CW} - 1) \tag{2.36}$$

In cases where $W = X$ Equation(2.36) reduces to Equation(2.6), because $M = P$ according to Equation(2.34). When comparing several SMs that are simplifications of the EM, the model with the lowest value of R_{CCW} is expected to give the best predictions at key operating points in *W*.

2.7. Linear Example

Consider a linear model with five parameters in the EM where the true values of the parameters are:

$$\beta = \left(\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5\right)^T = \left(1\ \frac{1}{2}\ \frac{1}{3}\ \frac{1}{4}\ \frac{1}{5}\right)^T \tag{2.37}$$

Input variable settings for experiments that are available for estimating parameters are:

$$X = \left(X_1\ \ X_2\ \ X_3\ \ \gamma X_1 + (1-\gamma)X_4\ \ \gamma X_2 + (1-\gamma)X_5\right) \tag{2.38}$$

where the first three columns are orthogonal and the final two columns may be correlated, depending on the value of γ. When $\gamma = 0$, all columns of X are orthogonal; as $\gamma \to 1$, columns

4 and 5 become highly correlated with columns 1 and 2, respectively. $X_i (i = 1, 2, ..., 5)$ are the following orthogonal column vectors:

$$\begin{aligned}
X_1^T &= (\ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1) \\
X_2^T &= (1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1) \\
X_3^T &= (\text{-}1\ \text{-}1\ \ 1\ \ 1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \text{-}1\ \text{-}1\ \ 1\ \ 1) \\
X_4^T &= (\text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \ 1\ \ 1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \ 1\ \ 1) \\
X_5^T &= (\text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1)
\end{aligned} \tag{2.39}$$

The setings in X matrix represent a 2^4 experimental design when γ=0. Assume that the additive noise, ε, in the EM follows a normal distribution with zero mean and constant variance and is independently and identically distributed.

To find the best SM, which gives the best predictions at the input variable settings, X, R_{CC} should be used as the model selection criterion (MSC). To select the best SM to provide predictions at desired settings W, R_{CCW} should be used. Consider several scenarios for W: 1) $W = kX$ where k is a scalar, 2) predictions are required at settings that are a subset of the rows in X $(W \subset X)$, and 3) a general case where W is any matrix that the modeler may be interested in. For case 1, it can be shown that $R_{CW} = R_C$ (and $r_C = r_{CW}$) so that the same SM will be chosen using either criterion(See Appendix A1). For cases 2 and 3, however, the criteria sometimes select different SMs. Below, we investigate the influence of the correlation factor γ, the noise variance σ^2 and the rows that appear in W to see how these factors affect the selection of a SM. First, true values of the criteria R_{CCW} and R_{CC} are used to investigate which SMs should be selected. Next, multiple sets of simulated experimental data (Monte-Carlo simulations) are used

to test the effectiveness of the estimators r_{CC} and r_{CCW} when the true parameter values are unknown.

In this example there are 2^5-2=30 possible SMs containing four or fewer parameters. To illustrate the proposed technique we have selected seven SMs arbitrarily (see Table 2.1) for comparison, along with the EM, rather than comparing the results for all of 31 models. For example, simplified model M_1 contains only the parameter β_1, simplified model M_5 contains parameters β_1 , β_3 and β_5 and the extended model M_8 contains all five parameters.

Table 2.1 A set of candidate models

Model	Parameters in candidate models				
	β1	**β2**	**β3**	**β4**	**β5**
M_1	•				
M_2				•	
M_3				•	•
M_4	•	•	•		
M_5	•		•		•
M_6		•	•	•	
M_7	•	•	•	•	
M_8	•	•	•	•	•

Table 2.2 shows model selection results for low and high levels of noise variance ($\sigma^2 = 0.10$ and $\sigma^2 = 10$, respectively) when the settings in X have low correlation ($\gamma = 0.10$) and when W is a particular subset of the rows in X (containing only the 2nd, 6th, 10th and 14th rows of X).

Table 2.3 shows analogous results for a fixed low value of the noise variance ($\sigma^2 = 0.10$) and two different values of the correlation factor ($\gamma = 0.10$ and $\gamma = 0.90$). Similar results can be obtained for other choices of W, as will be explained later in this chapter. Using different values of σ^2 and γ, the criteria R_{CCW} and R_{CC} often agree in the selection of the best SM, but in some situations they do not.

Table 2.2 Effect of noise variance at moderate correlation factor

	$\gamma = 0.10$							
Model Index	$\sigma^2 = 0.10$				$\sigma^2 = 10$			
	R_{CCW}	R_{CC}	$\%r_{CCW}$	$\%r_{CC}$	R_{CCW}	R_{CC}	$\%r_{CCW}$	$\%r_{CC}$
M_1	7.862	4.395	0	0	-0.169	-0.204	18.83	24.12
M_2	0.375	13.767	4.76	0	-0.244	-0.110	32.71	9.39
M_3	0.221	13.271	12.91	0	-0.183	-0.053	19.87	6.78
M_4	0.705	0.705	0.32	0.53	-0.117	-0.117	6.22	15.77
M_5	2.851	2.851	0	0	-0.095	-0.095	6.76	13.4
M_6	10.077	10.077	0	0	-0.023	-0.023	6.68	8.29
M_7	0.262	0.262	6.74	9.74	-0.059	-0.059	3.33	9.78
M_8	0.000	0.000	75.36	89.73	0.000	0.000	5.6	12.47

As shown in Table 2.2, when the noise variance is low (data are more reliable), both criteria tend to select complex models with more parameters. For example, when $\sigma^2 = 0.10$, the true values of the critical ratios, R_{CCW} and R_{CC} for the EM (M_8) are the lowest values among all candidate models. As a result, the EM is selected as the best model using both criteria. Model selection results from the 10000 Monte-Carlo simulations are consistent with the theoretical results. The EM is selected 75.36% and 89.73% of the time using r_{CCW} and r_{CC}, respectively. As expected, when the noise variance is higher both criteria tend to select less complicated models with a

smaller number of parameters; however, each criterion tends to select a different SM. For example, when $\sigma^2 = 10$, M_2 with only one parameter should be selected according to R_{CCW} (based on the value -0.244) and a different one-parameter model, M_1, should be selected using R_{CC}. The Monte-Carlo simulation results agree with these theoretical calculations.

As shown in Table 2.3, when the correlation factor is low, both criteria tend to select complex models with more parameters. For example, when $\gamma = 0.10$, the EM (M_8)(with five parameters) should be selected according to both R_{CCW} and R_{CC}. Monte-Carlo simulation results agree with these theoretical calculations. For higher values of the correlation factor, both criteria tend to select less complicated models with fewer parameters. For example at $\gamma = 0.90$, M_4 (with three parameters) should be selected according to both R_{CCW} (-0.115) and R_{CC} (-0.115). M_4 is the most popular model using both criteria, selected 31.01% of the time using r_{CCW} and r_{CC}. The second best model (M_5, with three parameters) was selected 26.83% of the time using r_{CCW} and r_{CC}.

Table 2.3 Effect of correlation at moderate noise variance

Model Index	$\sigma^2 = 0.10$							
	$\gamma = 0.10$				$\gamma = 0.90$			
	R_{CCW}	R_{CC}	$\%r_{CCW}$	$\%r_{CC}$	R_{CCW}	R_{CC}	$\%r_{CCW}$	$\%r_{CC}$
M_1	7.862	4.395	0	0	10.029	5.495	0	0
M_2	0.375	13.767	4.76	0	9.897	5.611	0	0
M_3	0.221	13.271	12.91	0	1.034	1.076	0.06	0.06
M_4	0.705	0.705	0.32	0.53	-0.115	-0.115	31.01	31.01
M_5	2.851	2.851	0	0	-0.088	-0.088	26.83	26.83
M_6	10.077	10.077	0	0	0.001	0.001	19.41	19.41
M_7	0.262	0.262	6.74	9.74	-0.059	-0.059	7.66	7.66
M_8	0.000	0.000	75.36	89.73	0.000	0.000	15.03	15.03

There are $\binom{16}{4} = 1820$ different possible subsets of the rows of the X matrix with 4 rows, and the results shown in Table 2.2 and Table 2.3 correspond to one of these possibilities where the modeler might desire good predictions. We investigated which models are selected using each of these 1820 possible W matrices to find out how often the R_{CCW} and R_{CC} criteria agreed in the selection of the same model considering all possible SMs with five or fewer parameters (2^5-1=31). For low noise variance and correlation ($\sigma^2 = 0.10$ and $\gamma = 0.10$), R_{CCW} and R_{CC} agreed that the extended model (M_8) should be selected regardless of the choice of W (i.e., the EM should be selected in 1820 out of the 1820 possible choices of W considered). For a high value of the correlation factor ($\gamma = 0.90$) with low noise variance ($\sigma^2 = 0.10$), R_{CCW} and R_{CC} tend to pick simpler models, as expected, and they agree on the choice of model in only 3.85% of the

cases. However, both criteria choose SMs with the same number of parameters (i.e., three-parameter SMs are chosen in 1820 out of 1820 cases). The SM selected most often by R_{CCW} is M_5 (1750 out of 1820 cases), whereas R_{CC} always selects M_4 because it does not depend on W. Similarly, for a high value of the noise variance ($\sigma^2 = 10$) with high correlation factor ($\gamma = 0.90$), corresponding to very poor data, R_{CCW} and R_{CC} tend to pick even simpler models and they agree on the choice of SM in 28.46% of the cases. Both criteria choose a SM with only one-parameter in all 1820 cases. For a high value of noise variance ($\sigma^2 = 10$) and low correlation ($\gamma = 0.10$), R_{CCW} and R_{CC} never agree on the choice of model, but they always pick one-parameter models. In summary, the two criteria R_{CCW} and R_{CC} tend to agree on the level of model complexity, but they often disagree on which SM is preferable, particularly when the data set available for parameter estimation is poor.

We also investigated the effects of increasing the number of rows in the W matrix, when we the modeler is interested in making predictions corresponding to a subset of the rows of X. For example when W contains 8 rows from the X matrix, there are $\binom{16}{8} = 12870$ different possibilities for the choice of W. In this situation, W is more similar to X than when a four-row subset is chosen. As a result, both criteria tend to agree on the best model more often.

Many of the results described above are based on the theoretical values of R_{CCW} and R_{CC}, rather than on the values of the statistics r_{CCW} and r_{CC} calculated from data. We generated 10 simulated data sets using the X matrix corresponding to $\gamma = 0.90$, with $\sigma^2 = 10$. Using this simulated data and all 1820 subset matrices W with 4 rows, we determined that r_{CCW} and r_{CC}

agreed on the best model 29.18% of the time, and on the number of parameters to estimate 99.56% of the time. These levels of agreement are similar to those for R_{CCW} and R_{CC}.

We also investigated the case where W contains the same number of rows as X, but predictions are desired at the corner points (i.e., W is determined using $\gamma = 0$, but the data are collected using a correlated design). For the values of the noise variance and correlation factor studied ($\gamma = 0.10, 0.90$ and $\sigma^2 = 0.10, 10$), both R_{CCW} and R_{CC} tend to agree on the best model. Finally, the case where $W = kX$ for $k = 2$ (i.e., when extrapolation is desired) was investigated. As expected from the proof in Appendix A1, R_{CW} and R_C always agree on the best model.

In summary, a simple five-parameter linear regression model was used to investigate the effects of noise variance, correlated experiments, and different choices for W on the models selected by R_{CCW} and R_{CC}. For high correlation and noise variance, when data are not very informative, both criteria tend to choose a simple model with fewer parameters. When data are more informative (low variance and correlation) both criteria tend to choose more complex models, regardless of the choice of the W matrix where good predictions are required. The two criteria, R_{CCW} and R_{CC}, and their statistics r_{CCW} and r_{CC} do not always select the same SM in different situations, but they often agree on the number of parameters in the selected SM. Note that the best models determined using R_{CCW} and R_{CC} are sometimes different than the models selected using the statistics r_{CCW} and r_{CC} that are calculated from the noisy data. Fortunately, r_{CCW} and r_{CC} tend to select models that are ranked as the best or nearly the best by R_{CCW} and R_{CC}. The new methodology is comparable to Wu's methodology in terms of computational effort. Wu's critical ratio has been recently used to select parameters in two polymerization reactor models with ~50

parameters[14, 15]. As a result, the proposed methodology should be applicable to larger scale systems without encountering computational difficulties. The only additional computations required are for computing the matrices A, B and M in Equations (2.31), (2.32) and (2.34).

2.8. Nonlinear Batch Reactor Example

The batch reactor model formulated by Dow Chemical Company consists of the differential and algebraic equations (DAEs) shown in Table 2.4. x_i, i=1,… ,10 is the model predictions for concentration of different species.

Table 2.4 Dynamic batch reactor model

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

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

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

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

$$\frac{dx_5}{dt} = -k_1 x_2 x_6 - 0.5 k_{-1} x_{10}$$

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

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

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

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

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

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

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

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

In this model, T is the reactor temperature in K, $T_0 = 340.15$ K is a reference temperature, $\left[Q^+ \right] = 0.0131 \text{ mol kg}^{-1}$ is the initial catalyst concentration and $R = 1.986 \text{ cal mol}^{-1}\text{K}^{-1}$ is the ideal gas constant. The response variables are three measured concentrations and a computed concentration as follows:

$$\begin{aligned} y_1 &= x_1 + \varepsilon_1 \\ y_2 &= x_2 + \varepsilon_2 \\ y_3 &= x_3 + \varepsilon_3 \\ y_4 &= y_1^0 - y_1 - y_3 \end{aligned} \tag{2.40}$$

where y_1^0 is the initial value of the concentration of component 1. Biegler et al. provide initial conditions and reported data at three different temperatures (40 °C, 67 °C, and 100 °C)[23].

This model was used by Wu et al.[27] and McLean et al.[26] to demonstrate the effectiveness of the model selection criterion r_{CC}. In their calculations, Wu and McLean assumed that measurements of type y_1, y_2 and y_3 all have the same noise variances. Table 2.5 shows the initial parameter values and associated uncertainties of the initial guesses that they used to rank the model parameters from most estimable to least estimable using the orthogonalization algorithm developed by our research group[10, 28]. As shown in Table 2.5, parameter E_2 was selected as the most estimable parameter, followed by parameter K_1 and so on. The parameters that rank near the top of the list should be estimated from the data, because they are influential parameters that are initially poorly known. Parameters that rank near the bottom of the list are not as important to estimate because they are less influential or have low initial uncertainties, meaning that their

values are already well known a priori. Parameters E_2 and K_1 appear at the top of the ranked list because of their large initial uncertainties assigned by Wu et al. (shown in the third column of Table 2.5) and their influence on predictions of the data supplied by Biegler et al.[23]. Parameter K_3 ranks at the bottom of the list due to its small initial uncertainty, its small influence on model predictions, and its correlation with other parameters that rank higher in the list.

Wu et al. formulated a series of nested SMs using the ranked list in Table 2.5[27]. The first SM is a one-parameter model, with the top-ranked parameter (E_2) estimated and all other parameters fixed at their initial values. The second SM contains the top two parameters (E_2 and K_1) to estimate, and finally the eighth SM includes the top eight parameters (with K_3 fixed). The resulting nonlinear least squares parameter estimation problems were solved using MATLAB™ (i.e., solver ode15s used to solve the stiff DAEs and subroutine lsqnonlin for parameter optimization). Detailed information concerning solver and optimizer settings is provided by McLean[29]. The corresponding values of r_{CC} are shown in Table 2.6, as are values of the proposed criterion r_{CCW} for several different choices of *W*.

Table 2.5 Initial parameters guesses, uncertainty in parameter and ranking of the parameters according to estimibility analysis

Parameters	Initial Guess	Uncertainty (s_{θ_0})	Rank
$E_2\left(cal\,mol^{-1}\right)$	1×10^{4}	1×10^{4}	**1**
$K_1\left(mol\,kg^{-1}\right)$	3×10^{-16}	2×10^{-16}	**2**
$E_{-1}\left(cal\,mol^{-1}\right)$	1×10^{4}	4×10^{3}	**3**
$E_1\left(cal\,mol^{-1}\right)$	1×10^{4}	2×10^{3}	**4**
$k_{20}\left(kg\,mol^{-1}h^{-1}\right)$	2.0	0.3	**5**
$k_{10}\left(kg\,mol^{-1}h^{-1}\right)$	1.0	0.2	**6**
$k_{-10}\left(h^{-1}\right)$	2×10^{3}	5×10^{2}	**7**
$K_2\left(mol\,kg^{-1}\right)$	5×10^{-14}	5×10^{-15}	**8**
$K_3\left(mol\,kg^{-1}\right)$	2×10^{-16}	1×10^{-17}	**9**

Table 2.6[1] Corrected critical ratios in nonlinear batch reactor example for various SMs at different desired operation conditions

		$W = W^{1\,2}$	$W = W^{2\,3}$	$W = W^{3\,4}$	$W = W^{4\,5}$
	r_{CC}	r_{CCW}	r_{CCW}	r_{CCW}	r_{CCW}
SM_1	27.344	0.128	0.128	-1.261E4	0.098
SM_2	19.727	0.123	0.123	-1.261E4	0.079
SM_3	19.401	0.123	0.123	-1.261E4	0.042
SM_4	17.978	0.123	0.123	-1.261E4	0.042
SM_5	3.708	0.041	0.041	-1.261E4	0.013
SM_6	0.074	0.046	0.046	-0.040E4	0.013
SM_7	-0.001	-0.005	-0.005	-0.049E4	-0.01
SM_8	-0.001	-0.002	-0.002	-0.006E4	-0.008
EM	0	0	0	0	0

Since the criterion r_{CCW} was developed earlier using a linear regression framework, several extensions are required so that it can be used to compare simplified nonlinear models of the form:

$$Y_i = f(X_i, \theta) + \varepsilon_i \tag{2.41}$$

[1] All shaded entries are the smallest value for the column and any other values that are smaller than the lowest value plus one percent of the highest value for the particular column

[2] Predictions at the middle temperature only ($67\ ^{\circ}C$)

[3] Predictions that extrapolate to low temperature behavior ($20\ ^{\circ}C$)

[4] Predictions at short times

[5] Predictions at long times

where Y_i is the measured response corresponding to the i^{th} run condition, X_i is the set of corresponding settings for the explanatory variables, θ is the p-dimensional vector of parameters and ε_i is the random noise.

The model can be linearized with respect to the parameters around a nominal parameter vector (θ_0):

$$\begin{aligned} f(X_i,\theta_j) &\approx f(X_i,\theta_{j0}) + S_{ij0}(\theta_j - \theta_{j0}) \\ &= f(X_i,\theta_{j0}) + S_{ij0}\delta\theta_j \end{aligned} \tag{2.42}$$

where $S_{ij0} = \frac{\partial f(X_i,\theta)}{\partial \theta_j^T}|_{\theta_0}$ is the corresponding element in i^{th} row and j^{th} column of the sensitivity matrix calculated at $\theta = \theta_0$. Thompson et al. suggested using the scaled sensitivity matrix (Z_0) obtained by dividing each element of the sensitivity matrix (S_{ij0}) by scaling factors to ensure dimensional consistency[10]:

$$Z_{ij0} = S_{ij0}\frac{s_{\theta_{j0}}}{s_{y_i}} \tag{2.43}$$

where $s_{\theta_{j0}}$ and s_{y_i} are uncertainties in the initial value of parameter θ_j and i^{th} measured value, respectively (supplied by the model user). By substituting Equations (2.42) and (2.43) into Equation(2.41), the scaled residual corresponding to the i^{th} run condition is

$$\xi_i = \frac{\left(Y_i - f(X_i,\theta_{j0})\right)}{s_{y_i}} \approx Z_{ij0}\frac{(\theta_j - \theta_{j0})}{s_{\theta_{j0}}} + \varepsilon_i \tag{2.44}$$

so that

$$\xi_i \approx Z_{ij0}\delta\varphi_j + \varepsilon_i \tag{2.45}$$

where $\delta\varphi_j$ is the scaled value of $\delta\theta_j$. In the mutiresponse situations, the measured values for the different responses stack below each other in the Y_i matrix. Similar stacking procedure is performed for the scaled sensitivity matrix and the scaled residuals corresponding to different responses.

Comparing the linear regression problem formulated in Equation (2.1) with the nonlinear regression problem corresponding to Equation(2.45), the following equations can be derived for the critical ratios for each nonlinear SM, which are analogous to the critical ratios defined in Equations (2.27) and (2.30):

$$r_{CW} = \frac{\left(\delta\hat{\varphi}_{2E}^T\right)\left(W_1A_1 - W_2\right)^T\left(W_1A_1 - W_2\right)\left(\delta\hat{\varphi}_{2E}\right)}{s_E^2 Tr\left(\left(W_1A_1 - W_2\right)\Omega\left(W_1A_1 - W_2\right)^T\right)} \tag{2.46}$$

where

$$\begin{aligned} A_1 &= \left(Z_1^T Z_1\right)^{-1} Z_1^T Z_2 \\ \Omega &= \left(Z_2^T\left(I_n - P_1\right)Z_2\right)^{-1} \\ P_1 &= Z_1(Z_1^T Z_1)^{-1} Z_1^T \end{aligned} \tag{2.47}$$

Equation (2.46) can also be written as:

$$r_{CW} = \frac{\left(\xi^T\right)A(\xi)/\left(p - p_1\right)}{\left(\xi^T\right)B(\xi)/\left(n - p\right)} \tag{2.48}$$

where

$$A = \frac{(p - p_1)(P - P_1) M^T M (P - P_1)}{Tr\left((P - P_1) M^T M (P - P_1)\right)} \tag{2.49}$$

$$B = I_n - P \tag{2.50}$$

$$M = W\left(Z^T Z\right)^{-1} Z^T \tag{2.51}$$

$$P = Z(Z^T Z)^{-1} Z^T \tag{2.52}$$

In Equations (2.46) to (2.51) the scaled residuals (ξ) replace the response variables (Y) in Equation (2.30), the scaled sensitivity matrix (Z) replaces the independent variable settings (X), and the scaled parameter deviations from their initial guesses ($\delta\varphi$) replace the parameter values (β), respectively.

For any choice of input settings, values of the model selection criterion, r_{CCW}, can be calculated for each SM of interest. Next, r_{CW} can be computed using Equation (2.48) and r_{CCW} can be calculated from Equation (2.36). The SM with the lowest value of r_{CCW} should give the best model predictions based on MSE at the desired operating points.

In this nonlinear batch reactor example, several choices of the input settings could be of interest to a modeler who wants to make good predictions at particular sets of experimental conditions. The following practical situations that might be of interest to a modeler were considered when computing W: 1) predictions at the middle temperature only ($67\ ^\circ\mathrm{C}$), corresponding to W^1 in Table 2.6; 2) predictions that require extrapolation to low temperature behavior (i.e., at $T_{low} = 20\ ^\circ\mathrm{C}$ and times corresponding to the $40\ ^\circ\mathrm{C}$ run, corresponding to W^2); 3) predictions at

short times, corresponding to the first time and half the first time where data are reported for the 40, 67 and 100 °C runs, respectively, resulting in W^3; 4) predictions at long times corresponding to the final time and twice the final time for the 40, 67 and 100 °C runs, resulting in W^4. Since estimating parameters in all possible SMs with eight or fewer parameters (i.e., 2^9-2=510 SMs) would be computationally expensive, a set of candidate SMs with one or more parameters was formulated based on the ranked list of parameters in Table 2.5, so that SM_3 contains parameters E_2, K_1 and E_{-1}, and SM_8 contains all of the parameters except K_3. Values of the model selection criteria r_{CC} and r_{CCW} for these models are shown in Table 2.6, for the four cases (W^1 to W^4) described above. For each model, the value of r_{CC} remains fixed because r_{CC} does not depend on the choice of W. The shaded entries in Table 2.6 are the smallest value in each column along with any other similar small values (i.e., any values that are smaller than the lowest value plus one percent of the highest value for the particular column). As shown in Table 2.7 SM_7 is selected when predictions are important at 67 °C , 20 °C , and long time. However, for prediction at short time, r_{CCW} selects SM_2 as the best model.

Table 2.7[6] Corrected critical ratios in nonlinear batch reactor example for various seven-parameter SMs at different desired operation conditions

		$W = W^1$	$W = W^2$	$W = W^3$	$W = W^4$
	r_{CC}	r_{CCW}	r_{CCW}	r_{CCW}	r_{CCW}
$SM_7^{1\,7}$	-0.0011	-0.0048	-0.0048	-0.0491E4	-0.0098
$SM_7^{2\,8}$	-0.0015	-0.0045	-0.0045	-0.0476E4	-0.0095
$SM_7^{3\,9}$	0.2674	-0.0075	-0.0075	-1.2615E4	-0.0022
$SM_7^{4\,10}$	0.0008	-0.0066	-0.0066	-0.1284E4	-0.0076

Figure 2.1 shows the total squared bias, total variance and total MSE for the model predictions as a function of number of parameters included in the SM. Part a) is for predictions of all the data of Biegler et al.. Part b) is for predictions at long reaction times, corresponding to W^4 in Table 2.6. These results were computed using nonlinear analogies to Equations (2.16), (2.17), (2.20), (2.21) and (2.23).

As expected, the squared bias for the model predictions decreases as the number of parameters increases, and the variance of the model predictions increases as the number of parameters increases. In both Figure 2.1 a) and Figure 2.1 b), the MSE for the predictions has its lowest value for SM_7, which matches the results for r_{CC} and r_{CCW} shown in Table 2.6, confirming the

[6] All shaded entries are the smallest value for the column and any other values that are smaller than the lowest value plus one percent of the highest value for the particular column

[7] Contains parameters 1, 2, 3, 4, 5, 6 and 7 in the ranked list of Table 2.5

8 Contains parameters 1, 2, 3, 4, 5, 6 and 9 in the ranked list of Table 2.5

9 Contains parameters 1, 2, 3, 5, 6, 7 and 9 in the ranked list of Table 2.5

10 Contains parameters 2, 3, 4, 5, 6, 7 and 9 in the ranked list of Table 2.5

selection of SM_7 as the best model (among the nine candidates considered) for making predictions at the data values or at the desired long-time operating conditions.

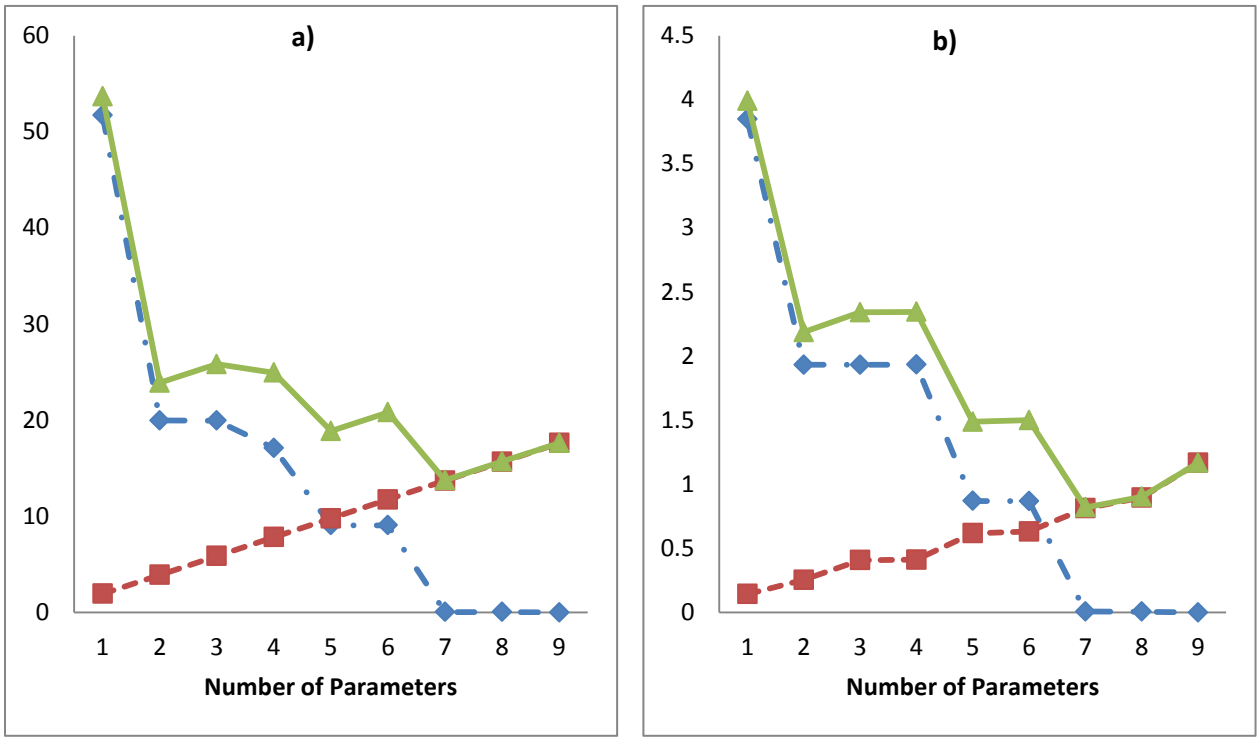

Figure 2.1: Estimated total variance -■-, squared bias -♦- and mean-squared error -▲- for a) predictions at the experimental settings used by Biegler et al. to estimate the model parameters and b) predictions at long reaction times

Because SM_7 was chosen most often in Table 2.6, we also considered three other seven-parameter SMs in addition to SM_7 to investigate the effect of different choices of operating conditions on the models selected by the corrected critical ratios (r_{CCW} and r_{CC}). We did this because the ranking of the parameters in Table 2.5 did not take into account the conditions where good model predictions are desired. Perhaps other seven-parameter models will give better predictions than the seven-parameter model from the ranked list. The results are shown in Table 2.7, where four different seven-parameter SMs are compared. As indicated in the footnote for Table 2.7, SM_7^1 includes the top seven parameters E_2, K_1, E_{-1}, E_1, k_{20}, k_{10} and k_{-10}. Models

SM_7^2, SM_7^3 and SM_7^4 are similar, except that they include the 9th ranked parameter K_3 in place of one of the higher ranked parameters. The shaded entries in Table 2.7 are the smallest value in each column along with any other similar small values (i.e., any values that are smaller than the lowest value plus one percent of the highest value for the particular column). The results in Table 2.7 indicate that r_{CC} and r_{CCW} often agree on the choice of the best SM, but sometimes they do not, so that SM_7^1 (which is the same as SM_7 in Table 2.6) is not necessarily the best SM among the set of candidate seven-parameter SMs when different operating conditions are important to the modeler.

In summary, a nonlinear approximation of the proposed corrected critical ratio (r_{CCW}) was developed to select the best SM among the set of candidate nonlinear SMs. A nine-parameter nonlinear batch reactor DAE model was used to investigate the effect of different choices of operating conditions where predictions are important on the model selected. In most of the situations studied, the SM selected by r_{CCW} is the same as that selected using r_{CC}, but in some cases, the criteria select different models.

2.9. Conclusion

A simple methodology is proposed to improve parameter subset selection, based on a new MSE-based model selection criterion r_{CCW}. This new criterion is an improvement over the previous MSE-based criterion, r_{CC}, developed by Wu et al., because it accounts for user-specified operating conditions where good model predictions are desired. These two model-selection techniques are valuable because the simplified models that are selected are expected to give better predictions (in terms of mean-squared prediction error) than the corresponding complex models, especially when data are limited.

A simple five-parameter linear regression model is used to investigate effects of noise variance, correlated experimental designs and different choices of operating conditions on model selection using the new criterion r_{CCW}. Results are compared with those for Wu's critical ratio r_{CC}. For experiments with high correlation and noise variance, both criteria tend to choose simple models with fewer parameters because the data are not very informative. When data are more informative (low noise variance and correlation) both criteria tend to choose more complex models, regardless of the choice of input conditions where good predictions are required. Simulation results show that r_{CCW} and r_{CC} often agree on the choice of the best simplified model to use, especially when there are sufficient data to estimate all of the parameters. In cases where they select different simplified models, they often agree on the number of parameters that should be included in the model.

A nonlinear approximation for the criterion r_{CCW} is developed and used to compare simplified nonlinear models for a dynamic batch reactor example[23]. This nine-parameter model, consisting of six ODEs and four algebraic equations, is used to investigate the effect of different settings where predictions are desired on the parameters that are selected for estimation. In this example, Wu's r_{CC} criterion selected seven parameters for estimation. The new criterion selected only two parameters for estimation when predictions are required only at long times. Seven parameters were selected for making predictions at the other conditions studied. The proposed technique will be valuable to modelers with limited data who want to make the best possible predictions at specific operating conditions.

2.10. References

(1) Wu, S.; Harris, T. J.; McAuley, K. B. The use of simplified or misspecified models: Linear case. *Can. J. Chem. Eng.* **2007**, *85*, 386-398.

(2) Perregaard, J. Model simplification and reduction for simulation and optimization of chemical processes. *Comput. Chem. Eng.* **1993**, *17*, 465-483.

(3) Wang, F.; Zhu, Z.; Massarotto, P.; Rudolph, V. A simplified dynamic model for accelerated methane residual recovery from coals. *Chem. Eng. Sci.* **2007**, *62*, 3268-3275.

(4) Chu, Y.; Huang, Z.; Hahn, J. Improving prediction capabilities of complex dynamic models via parameter selection and estimation. *Chem. Eng. Sci.* **2009**, *64*, 4178-4185.

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

(6) Moore, C.; Doherty, J. Role of the calibration process in reducing model predictive error. *Water Resour. Res.* **2005**, *41*, W05020.

(7) McLean, K. A. P.; McAuley, K. B. Mathematical modelling of chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Can. J. Chem. Eng.* **2011**, *90*, 351–366.

(8) Hocking, R. R. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. *Biometrics* **1976**, *32*, pp. 1-49.

(9) Miller, A. J. In *Subset selection in regression;* Chapman and Hall: London England ; New York, 1990; .

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

(11) Guo, Y.; Wang, N.; Bradshaw, R. D.; Brinson, L. C. Modeling Mechanical Aging Shift Factors in Glassy Polymers During Nonisothermal Physical Aging. I. Experiments and KAHR-a(te) Model Prediction RID B-6678-2009. *J. Polym. Sci. Part B Polym. Phys.* **2009**, *47*, 340-352.

(12) Jayasankar, B. R.; Ben-Zvi, A.; Huang, B. Identifiability and estimability study for a dynamic solid oxide fuel cell model RID F-2736-2010. *Comput. Chem. Eng.* **2009**, *33*, 484-492.

(13) Wu, S.; McAuley, K. B.; Harris, T. J. Selection of Simplified Models: I. Analysis of Model-Selection Criteria using Mean-Squared Error. *Can. J. Chem. Eng.* **2011**, *89*, 148-158.

(14) Woloszyn, J. D.; McAuley, K. B. Application of Parameter Selection and Estimation Techniques in a Thermal Styrene Polymerisation Model. *Macromol. React. Eng.* **2011**, *5*, 453–466.

(15) Karimi, H.; Schaffer, M. A.; McAuley, K. B. Kinetic Moded for Non-Oxidative Thermal Degradation of Nylon 66. *Macromol. React. Eng.* **2012**, *6*, 93-109.

(16) Shmueli, G. To Explain or to Predict? *Stat. Sci.* **2010**, *25*, 289-310.

(17) Shmueli, G.; Koppius, O. R. Predictive Analytics in Information Systems Research. *MIS Quart.* **2011**, *35*, 553-572.

(18) Hastie, T.; Tibshirani, R.; Friedman, J. H. In *The elements of statistical learning: Data Mining, Inference, and Prediction;* Springer: New York, 2009; .

(19) Hawkins, D. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1-12.

(20) Consonni, V.; Ballabio, D.; Todeschini, R. Evaluation of model predictive ability by external validation techniques. *J Chemometr* **2010**, *24*, 194-201.

(21) Konishi, S.; Kitagawa, G. In *Information criteria and statistical modeling;* Springer: New York, 2008; .

(22) Montgomery, D. C.; Peck, E. A.; Vining, G. G. In *Introduction to Linear Regression Analysis;* John Wiley & Sons: NY, 2001; .

(23) Biegler, L. T.; Damiano, J. J.; Blau, G. E. Nonlinear parameter estimation: A case study comparison. *AIChE J.* **1986**, *32*, 29-45.

(24) Huang, Y.; Draper, N. Transformations, regression geometry and R-2. *Comput. Stat. Data Anal.* **2003**, *42*, 647-664.

(25) Rice, J. A. In *Mathematical statistics and data analysis [multimedia];* Duxbury Press: Belmont, CA, 1995; .

(26) McLean, K. A. P.; Wu, S.; McAuley, K. B. Mean-Squared-Error Methods for Selecting Optimal Parameter Subsets for Estimation. *Ind. Eng. Chem. Res.* **2012**, *51*, 6105.

(27) Wu, S.; McLean, K. A. P.; Harris, T. H.; McAuley, K. B. Selection of optimal parameter set using estimability analysis and MSE-based model-selection criterion. *IJAMechS* **2011**, *3*, 188-197.

(28) Yao, K. Z.; Shaw, B. M.; McAuley, K. B.; Bacon, D. W. Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimatibility analysis and experimental design. *Polym. React. Eng.* **2003**, *11*, 563–588.

(29) McLean, K. A. P. Obtaining the Best Model Predictions and Parameter Estimates Using Limited Data, Queen's University 2011-09-27 10:52:31.588, 2011.

Chapter 3 Mean-Squared-Error-Based Method for Parameter Ranking and Selection To Obtain Accurate Predictions at Specified Operating Conditions

3.1. Summary

A mean-squared-error-based forward selection methodology is proposed for simultaneous parameter ranking and selection based on the critical ratio r_{CCW}[1]. This new technique employs information in the available data set and the operating region of interest to determine the best model with the lowest mean square prediction error. This technique involves relatively simple computations and avoids over-fitting of noisy data. This new approach is valuable when data available for parameter estimation arise from correlated experimental designs that make accurate estimation of all the parameters difficult. It is particularly beneficial when the predictions are desired in an operating region that is different from where the data are already available. Monte Carlo simulations of a linear regression example and a nonlinear case study are used to illustrate the effectiveness of the proposed method, demonstrating that results from simulated data agree with theoretical results.

This chapter was published in Industrial and Engineering Chemistry Research in vol. 53 pp. 6033 −6046 with K B McAuley as coauthor.

3.2. Introduction

Fundamental models are used for simulation, design, control and optimization of a wide variety of chemical and biochemical processes[2,3]. Therefore, improving the accuracy of model predictions has been an important subject for many researchers[4–10]. Obtaining accurate model predictions often requires estimation of unknown model parameters and making decisions about which parameters should be estimated and which should be fixed at prior values or removed from the model via simplification[11–13].

One major difficulty in parameter estimation in complex models is that it is often not possible to estimate all of the parameters from the available data[14–16]. Moreover, experimental data for parameter estimation are unavoidably noisy, and are often obtained using a limited number of runs, or using correlated experimental designs[11,17]. In addition, some parameters may have only a small influence on model predictions and the effects of some parameters may be correlated with the effects of others. These types of problem can sometimes be addressed by conducting additional experiments or by measuring additional responses to improve the conditioning of the Fisher Information Matrix[18–20]. However, the modeler may sometimes decide to change the structure of the model equations by removing some parameters which have little influence on model predictions, by lumping several parameters together into a single parameter, or by fixing some of the model parameters at their nominal values[21–23]. Estimating too many parameters, i.e. over-fitting, decreases a model's predictive ability and leads to large uncertainties in the estimated parameter values and large variances in the model predictions[8,17,24,25]. Therefore simplified models (SMs) that have a smaller number of parameters are sometimes developed and used. Using a SM, rather than the corresponding complex model, simplifies the estimation procedure and can lead to more accurate model predictions, while avoiding over-fitting[21,26,27].

When fewer parameters are included in a simplified model, the prediction variance decreases while the prediction bias increases[5,8,25]. Because of this trade-off between bias and variance, a SM gives better predictions than a corresponding complex model with more parameters when the variance reduction outweighs the bias increase[17].

One common way to derive a SM is by fixing some of the parameters in the complex model either at zero or at other nominal values based on the modeler's prior knowledge, and then estimating the remaining parameters[26]. Statistical significant tests such as *t* tests or ridge selection analysis have been used to decide which parameters to eliminate from complex models[28,29]. Deciding which parameters to fix and which to estimate is referred to as parameter subset selection[8,21,26,27,30]. Parameter subset selection techniques have been described in recent review articles[11,30,31], which describe a variety of techniques including the collinearity index[32], the relative gain array[33], the Hankel singular value[34], orthogonalization methods[23,35,36], and optimization-based methods (e.g., using *A*-, *D*- or *E*- optimality) that rely on the Fisher information matrix[37]. The primary difficulty in using these optimization-based methods is the combinatorial nature of the optimization problem[11]. As the number of parameters increases, the optimal parameter subset is difficult or even impossible to find. For example finding the best 30 parameters to estimate in a complex model with 50 parameters involves $\binom{50}{30} \cong 4.7e13$ possible outcomes. Therefore, alternative procedures are often used to find sub-optimal solutions with less computational effort. These methods include forward selection and backward selection[38](which is also called backward elimination[31]), stochastic searches such as genetic algorithms[39], and heuristic reduction techniques such as parameter clustering[40]. Although the commonly used *D* optimality criterion (e.g. [32,39]) can be used to find which 30-parameter SM is the best among all candidate 30-parameter SMs (in terms of the smallest volume of the joint

confidence region for the parameters that are estimated), it cannot be used to decide whether the best 30-parameter SM will be better than the best 31-parameter SM.

Orthogonalization techniques (see Table 3.1) are the most popular techniques for parameter ranking and selection, and have been used extensively to aid parameter estimation in models for complex processes including polymerization reactors, bioreactors and fuel cells (e.g., [11,41–44]). In the orthogonalization method, model parameters are ranked using the Fisher information matrix, Z^TZ, which contains information about the influence of model parameters on the model predictions and correlations among model parameters. The forward-selection orthogonalization method outlined in Table 3.1 may lead to suboptimal parameter subset selection because selection of the k^{th} parameter is conditional on the k-1 parameters that

Table 3.1 Orthogonalization Algorithm[35]

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 $Z_{1,k}$.
3	Use $Z_{1,k}$ to predict columns in Z using ordinary least-squares $\hat{Z}_k = Z_{1,k}\left(Z_{1,k}^T Z_{1,k}\right)^{-1} Z_{1,k}^T Z$ (T3.1.1) and calculate the residual matrix $R_k = Z - \hat{Z}_k$ (T3.1.2)
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 $Z_{1,k}$.

6	Advance the iteration counter *k* 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.

already appear on the ranked list. Chu et al. recently proposed a generalized orthogonalization method for parameter subset selection[45] that partially addresses this problem by selecting several parameters at each iteration. This method involves more computations than the traditional orthogonalization method in Table 3.1, but should provide results that are closer to the solution of the full combinatorial optimization problem. Another problem associated with the orthogonalization procedure in Table 3.1 is that the scaled local sensitivity matrix Z , whose elements are defined in Equation (3.15), depends on the initial parameter guesses, which may be very uncertain[21]. Poor initial values of some parameters may yield a misleading list of ranked parameters. McLean et al.[13] addressed this problem by ranking the model parameters 100 times using a variety of initial guesses within the uncertainty ranges for the individual parameters. In the example that they studied, seven parameters were selected for estimation about 80% of time, with six or eight parameters being selected less often. Kou et al.[46] re-ranked the parameters several times using a revised sensitivity matrix that was updated using non-linear regression. Recently, Alberton et al. used a binary search method to reduce the number of times that parameter re-estimation via non-linear regression is required when selecting parameter subsets[47].

Orthogonalization-based techniques focus on selecting a parameter subset that avoids rank deficiency of the Fisher information matrix (FIM) (i.e. ill-conditioning)[21]. Modelers are often interested in obtaining accurate model predictions as well as avoiding parameter estimation problems. Even though the two concepts, accuracy of model predictions and reliable parameter

estimates, are related, augmenting the prediction accuracy of a model is not always the same as improving parameter estimability[30]. Parameter estimation challenges and the desire for accurate model predictions motivated Wu et al. to propose a criterion based on mean-squared error (MSE) to determine the number of parameters to estimate from the ranked list of parameters obtained by orthogonalization (refer to Table 3.2)[17,26,48]. Wu's MSE-based method uses a statistically motivated critical ratio rather than an arbitrary cut-off value[30,44] to determine how many parameters to estimate and how many to fix. Fortunately, Wu's method is not as computationally expensive as other reliable methods such as cross-validation[23].

Table 3.2 Using Wu`s MSE-Based Model Selection Criterion to Determine the Optimum Number of Parameters for Estimation[49]

1 Rank model parameters from most estimable to least estimable using the orthogonalization algorithm in Table 3.1.

2 Use weighted least-squares regression to estimate the first parameter from the list, 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. Weighting factors used in parameter estimation should be consistent with measurement uncertainties s_{y_i} used for scaling during parameter ranking.

3 Compute the critical ratio

$$r_{C,k} = \left(J_k - J_p\right) / \left(p - k\right) \qquad \text{` (T3.2.1)}$$

for $k = 1 \ldots p - 1$.

Weighted least-squares objective function J is defined as:

$$J = \sum_{i=1}^{d} \sum_{l=1}^{c} \sum_{m=1}^{r} \left(\frac{Y_{ilm} - \hat{Y}_{ilm}}{s_{y_i}} \right)^2 \qquad \text{(T3.2.2)}$$

where Y_{ilm} is the measured value of the i^{th} response variable at the l^{th} measurement time, for the m^{th} experimental run and $\hat{Y}_{ilm}$ is the corresponding predicted value.

4 For each value of k, compute the corrected critical ratio

$$r_{CC,k} = \frac{p-k}{N}\left(r_{CKub,k} - 1\right) \qquad \text{(T3.2.3)}$$

where

$$r_{CKub,k} = \max\left(r_{C,k} - 1, \frac{2}{p-k+2} r_{C,k} \right) \qquad \text{(T3.2.4)}$$

5 Select the value of k corresponding to the lowest value of $r_{CC,k}$ as the appropriate number of parameters to estimate.

MSE for model predictions is equal to the sum of the squared bias and the variance; therefore, Wu's criterion accounts for the trade-off between bias and variance as additional parameters

from the ranked list are estimated. Like many other subset selection techniques[31], Wu's approach requires experimental data and least-squares parameter estimation to choose the optimum number of parameters to estimate. The parameter ranking method in Table 3.1 has been combined with Wu's MSE selection method in Table 3.2 to choose parameter subsets to estimate in four polymerization models[50–53]. Chu et al. proposed a more computationally intensive MSE-based forward selection technique, which uses Monte Carlo simulations and considers a large number of possible parameter values to calculate the average bias[21]. Chu's methodology simultaneously ranks the parameters and selects the optimal number to estimate. Because of the wide variety of initial parameter guesses that are used, it should be robust to poor initial parameter guesses[21]. Recently, Mclean et al. used Wu's criterion to simultaneously rank and select the parameters for estimation (see Table 3.3) in a nonlinear batch reactor model[13]. They repeated this analysis using the MSE-based method of Chu et al.[13]. It is not clear which of the two procedures leads to better results and which is more computationally intensive. Both procedures require substantially less effort than leave-one-out cross-validation and both are more reliable and more computationally intensive than using orthogonalization followed by Wu's method (i.e., the procedures in Table 3.1 and Table 3.2)[13].

One problem with the MSE-based techniques of Wu et al. and Chu et al. is that they use the MSE of predictions made at experimental settings corresponding to the available data points for

Table 3.3 r_{CC} Method to Rank and Select Parameters to Estimate[13]

1	Starting with no parameters in the ranked list, formulate p candidate parameter subsets, each with one 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(T3.2.2) in Table 3.2 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 (T3.2.1) to (T3.2.4) in 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 (T3.2.1) to (T3.2.4) in 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.
	parameter,

selecting the optimal parameter subset. Users are often interested in obtaining accurate predictions at operating conditions that are different from the settings where data are already available. This problem motivated Eghtesadi et al. to propose an MSE-based criterion that considers the operating region where accurate model predictions are desired (refer to Table 3.4)[1].

Table 3.4 Using Orthogonalization and the r_{CCW} Criterion to Select Parameters for Estimation[1]

1 Rank model parameters from most estimable to least estimable using the orthogonalization algorithm in Table 3.1.

2 For a linear model, use the experimental data to calculate $r_{CW,k}$ for each k (i.e. for k = 1... p – 1) from:

$$r_{CW,k} = \frac{Y^T AY / (p-k)}{Y^T BY / (n-p)} \qquad \text{(T3.4.1)}$$

where

$$A = \frac{(p-k)\left(P - P_{1,k}\right) M^T M \left(P - P_{1,k}\right)}{Tr\left(\left(P - P_{1,k}\right) M^T M \left(P - P_{1,k}\right)\right)} \qquad \text{(T3.4.2)}$$

$$B = I_n - P \qquad \text{(T3.4.3)}$$

$$M = W\left(X^T X\right)^{-1} X^T \qquad \text{(T3.4.4)}$$

$$P = X(X^T X)^{-1} X^T \qquad \text{(T3.4.5)}$$

$$P_{1,k} = X_{1,k}(X_{1,k}^T X_{1,k})^{-1} X_{1,k}^T \qquad \text{(T3.4.6)}$$

where $X_{1,k}$ is the sub-matrix of X corresponding to the parameters that are included in the SM for the kth iteration.

For nonlinear model use nonlinear weighted least-squares regression to estimate all of the ranked parameters in the ranked parameter list from step 1. Use experimental data to calculate $r_{CW,k}$ for each k from the residuals ξ defined in Equation (3.16) and A and B from Equations (T3.4.2) and (T3.4.3)

$$r_{CW,k} = \frac{\left(\xi^T\right) A(\xi) \big/ (p-k)}{\left(\xi^T\right) B(\xi) \big/ (n-p)} \qquad \text{(T3.4.7)}$$

with

$$M = W\left(Z^T Z\right)^{-1} Z^T \qquad \text{(T3.4.8)}$$

$$P = Z(Z^T Z)^{-1} Z^T \qquad \text{(T3.4.9)}$$

$$P_{1,k} = Z_{1,k}(Z_{1,k}^T Z_{1,k})^{-1} Z_{1,k}^T \qquad \text{(T3.4.10)}$$

where $Z_{1,k}$ is the sub-matrix of the scaled local sensitivity matrix Z corresponding to the parameters that are included in the SM for the kth iteration.

3 Calculate corrected critical ratio $r_{CCW,k}$ for each k from:

$$r_{CCW,k} = \frac{Tr\left(M^T M\left(P - P_{1,k}\right)\right)}{w}(r_{CW,k} - 1) \qquad \text{(T3.4.11)}$$

4 Select the value of k corresponding to the lowest value of $r_{CCW,k}$ as the appropriate number of parameters to estimate.

For example, an engineer may want to use a model to make accurate predictions over a limited range of temperatures where a chemical process will operate, and the data set available for parameter estimation may contain a wider range of temperature settings. Alternatively, the engineer may want to make accurate predictions at a new temperature where no data are available. When Eghtesadi et al. tested their new criterion, they relied on the orthogonalization-based parameter ranking method in Table 3.1. The shortcoming of this approach is that knowledge about the specific operating conditions of interest is not considered during the ranking stage of parameter subset selection. In the current chapter, a new forward-selection

methodology is developed based on r_{CCW} MSE-based criterion to rank and select the parameters simultaneously while considering the operating region of interest. The proposed technique ranks the parameters to minimize the anticipated MSE of the model predictions at operating conditions specified by the user.

First, we introduce some notation and summarize the MSE-based model-selection criterion of Eghtesadi et al.. Next, the new algorithm for simultaneously ranking and selecting parameters is described and two examples, a linear regression problem and a nonlinear batch reactor model[26,54], are used to test and demonstrate the proposed methodology. This new approach may be valuable to industrial modelers who want to make predictions about new product specifications or grades, using models that are fitted from data obtained at previous operating conditions.

3.3. Background Information

Consider the following single-response extended model (EM) which is linear in the parameters:

$$\begin{aligned} Y_E &= X\beta + \varepsilon \\ &= X_1\beta_1 + X_2\beta_2 + \varepsilon \end{aligned} \tag{3.1}$$

This EM has a sufficiently detailed structure so that the model is capable of providing reliable predictions over the range of operating conditions of interest, provided that appropriate values of the model parameters are available. In Equation (3.1) $Y \in R^n$ is a vector of measured responses and the subscript "E" indicates the extended model, rather than a SM. $X \in R^{n\times p}$ is a matrix of independent-variable settings. $X_1 \in R^{n\times p_1}$ and $X_2 \in R^{n\times(p-p_1)}$ are submatrices that contain the first

p_1 and remaining p-p_1 columns of X, respectively. $\beta \in R^p$ is a vector of the unknown model parameters. $\beta_1 \in R^{p_1}$ and $\beta_2 \in R^{p-p_1}$ are sub vectors of β, and $\varepsilon \in R^n$ is a vector of additive random noise terms. n is the total number of measurements, p is the total number of parameters in the EM and p_1 is the number of parameters in the SM:

$$Y_S = X_1\beta_1 + e \tag{3.2}$$

The following least-squares assumptions are made: X_1 and X_2 are perfectly known and the stochastic component ε is a mean-zero, uncorrelated random sequence with constant variance σ^2[29]. In Equation (3.2) $e = X_2\beta_2 + \varepsilon$ is the stochastic component combined with any model mismatch, and the subscript "S" indicates a simplified model. Note that SM is nested within the EM because it can be obtained by setting the parameters in β_2 to zero.

MSE is defined as the expected squared difference between the model prediction ($\hat{Y}$) and the noise-free response of the process (Y_{True}) where[55]:

$$Y_{True} = X\beta \tag{3.3}$$

so that:

$$\begin{aligned} MSE(\hat{Y}) &= E\left((\hat{Y} - Y_{True})^T(\hat{Y} - Y_{True})\right) \\ &= \left(E(\hat{Y}) - Y_{True}\right)^T\left(E(\hat{Y}) - Y_{True}\right) + Tr\left(Cov(\hat{Y})\right) \end{aligned} \tag{3.4}$$

where $E(\cdot)$, $Cov(\cdot)$ and $Tr(.)$ denote the expected value, variance-covariance matrix and trace, respectively. The squared bias is $\left(E(\hat{Y})-Y_{True}\right)^T\left(E(\hat{Y})-Y_{True}\right)$ and the total variance is $Tr\left(Cov(\hat{Y})\right)$.

Wu et al. proposed a strategy (see Table 3.2) based on a critical ratio R_C to find whether a particular SM or the corresponding EM gives better predictions at the points where data are already available (i.e., at the settings in *X*). This critical ratio is the ratio of the squared bias over the variance when p_1 parameters are estimated rather than all *p* parameters. They also proposed a corrected critical ratio, R_{CC}, to compare several SMs with different numbers of parameters[17,26,48] and derived an estimator, r_{CC}, for this corrected critical ratio(see Equations (T3.2.1) to (T3.2.4) in Table 3.2). The SM that is expected to provide the best predictions (lowest MSE) at the input settings in *X* corresponds to the lowest value of r_{CC}. Note that Wu et al. used the notation $\hat{R}_{CC}$ instead of r_{CC} for the estimator in their earlier work[26], and then simplified the notation to r_{CC} later[49].

The main problem associated with using r_{CC} to select a SM is that the modeler often desires accurate predictions at alternative input settings that are different from the settings in *X* where data are already available. If the modeler wants accurate predictions corresponding to a different input matrix $W \in R^{w\times p}$, then the SM selected using r_{CC} may not be the best choice. To address this problem, the revised model selection criterion r_{CCW} shown in Table 3.4 was developed[1].

The matrix of important input settings W can be partitioned in the same way as X into $W_1 \in R^{w \times p_1}$ and , $W_2 \in R^{w \times (p-p_1)}$ where w is the total number of important predictions to be made so that:

$$\hat{Y}_E = W\hat{\beta}_E = W_1\hat{\beta}_{1E} + W_2\hat{\beta}_{2E} \tag{3.5}$$

$$\hat{Y}_S = W_1\hat{\beta}_{1S} \tag{3.6}$$

Using the trade-off between bias and variance, the R_{CW} criterion aims to select the best SM with the lowest total MSE at specified important operating conditions in *W*:

$$R_{CW} = \frac{\text{bias}^2 \text{ increase}}{\text{var decrease}} = \frac{\beta_2^T \left(W_1A_1 - W_2\right)^T \left(W_1A_1 - W_2\right)\beta_2}{\sigma^2 Tr\left(\left(W_1A_1 - W_2\right)\Omega\left(W_1A_1 - W_2\right)^T\right)} \tag{3.7}$$

where

$$A_1 = \left(X_1^T X_1\right)^{-1} X_1^T X_2 \tag{3.8}$$

$$\Omega = \left(X_2^T \left(I_n - P_1\right) X_2\right)^{-1} \tag{3.9}$$

$$P_1 = X_1(X_1^T X_1)^{-1} X_1^T \tag{3.10}$$

The corrected critical ratio R_{CCW} has been derived to make comparisons among a set of candidate SMs with different numbers of parameters:

$$\begin{aligned} R_{CCW} &= \frac{Tr\left(\left(W_1A_1 - W_2\right)\Omega\left(W_1A_1 - W_2\right)^T\right)}{w}(R_{CW} - 1) \\ &= \frac{Tr\left(M^T M\left(P - P_1\right)\right)}{w}(R_{CW} - 1) \end{aligned} \tag{3.11}$$

where

$$P = X(X^T X)^{-1} X^T \tag{3.12}$$

$$M = W\left(X^T X\right)^{-1} X^T \tag{3.13}$$

The model with the lowest value of R_{CCW} is expected to give the best predictions at key operating points in *W*. The estimators for this corrected critical ratio r_{CCW} can be obtained by using Equations (T3.4.1) to (T3.4.11) in Table 3.4. A model with the lowest value of r_{CCW} is expected to give the best predictions (lowest MSE) at key operating points corresponding to W^1. Eghtesadi et al. extended the r_{CCW} criterion that was developed using a linear regression framework, for use with nonlinear EMs of the form:

$$Y_{iE} = f(X_i, \theta) + \varepsilon_i \tag{3.14}$$

where Y_i is the measured response corresponding to the i^{th} run condition, X_i is the set of corresponding settings for the explanatory variables, θ is the *p*-dimensional vector of parameters and ε_i is the random noise. They performed Taylor series expansion for the nonlinear model around a nominal parameter vector (θ_0) to obtain the nonlinear analogy for the r_{CW} criterion (Equations (T3.4.7) to (T3.4.10) in Table 3.4)[1]. In Equations (T3.4.8) to (T3.4.10) in Table 3.4, Z is the local scaled sensitivity matrix and can be partitioned into $Z_1 \in R^{n \times p_1}$ and $Z_2 \in R^{n \times (p-p_1)}$ in a similar way to *X*. Note that the FIM (i.e., X^TX or alternatively Z^TZ) must be invertible for the algorithm in Table 3.4 to be used. The sensitivity matrix, *Z*, contains information about the effects of parameter values on model predictions for all different experimental runs. The local

scaled sensitivity matrix around the initial parameter value $\theta_0, Z\big|_{\theta_0} = Z_0$, is obtained by multiplying each element of the local sensitivity matrix (S_{ij0}) by scaling factors to ensure dimensional consistency[23]:

$$Z_{ij0} = S_{ij0} \frac{s_{\theta_{j0}}}{s_{y_i}} \tag{3.15}$$

where $S_{ij0} = \frac{\partial f(X_i, \theta)}{\partial \theta_j^T}|_{\theta_0}$ is the corresponding element in i^{th} row and j^{th} column of the sensitivity matrix calculated at $\theta = \theta_0$, $s_{\theta_{j0}}$ and s_{y_i} are uncertainties in the initial value of parameter θ_j and the i^{th} measured value, respectively. The scaling factors should be chosen carefully, since scaling influences the results of parameter ranking procedure. Initial parameter guesses and their corresponding uncertainties can be determined based on the modeler's prior knowledge or information from the literature. Uncertainties associated with different measurements can be estimated from replicate experiments or can be obtained based on information from the device suppliers. To calculate the elements of the local scaled sensitivity matrix, *Z*, difference approximations can be used. These elements can also be obtained by solving sensitivity equations[56]. In Equation (T3.4.7), ξ is the scaled residual vector:

$$\xi_i = \frac{\left(Y_i - f(X_i, \theta_{j0})\right)}{s_{y_i}} \tag{3.16}$$

For any choice of input settings, values of the model selection criterion, r_{CW} can be computed using Equations (T3.4.1) to (T3.4.10), and r_{CCW} can be calculated from Equation (T3.4.11). The

SM with the lowest value of r_{CCW} should give the best model predictions based on MSE at the desired operating points.

3.4. Parameter Ranking Procedure Based on r_{CCW} Criterion

A new parameter subset selection methodology is proposed in this paper based on the MSE-based model selection criterion r_{CCW}. The proposed forward selection technique will rank parameters to minimize the MSE at the specified operating conditions. This new method, outlined in Table 3.5, uses experimental data and parameter estimation to compute $r_{CCW,k}$ at each step. The method ranks parameters based on information in the available data set and information provided by the modeler about the desired operating conditions where good predictions are required. Using the proposed technique, the appropriate number of parameters to estimate is the value k that corresponds to the lowest value of $r_{CCW,k}$. This technique requires solving $p(p+1)/2-1$ parameter estimation problems to rank all the parameters, where p is the number of parameters in the complex (extended) model. Uncertainty in the initial parameters values is accounted for by assigning physically realistic parameter bounds, based on prior knowledge, during parameter estimation. A simple linear regression model and a nonlinear batch reactor example[54] are used to test and illustrate the proposed technique. Results are compared with those from a related technique that does not account for the operating region where good predictions are required[13].

3.5. Linear Regression Example

Consider a linear five-parameter EM of the form of Equation (3.1) where the true values of the model parameters are[17]:

$$\beta = \left(\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5\right)^T = \left(1\ \frac{1}{2}\ \frac{1}{3}\ \frac{1}{4}\ \frac{1}{5}\right)^T \tag{3.17}$$

The matrix of independent variable settings *X* that was used to collect the experimental data is:

$$X = \left(X_1\ X_2\ X_3\ \gamma X_1 + (1-\gamma)X_4\ \gamma X_2 + (1-\gamma)X_5\right) \tag{3.18}$$

where the first three columns in X are orthogonal and the final two columns can be correlated, depending on the value of the correlation factor γ. For example, when $\gamma = 0$, all columns in *X* are orthogonal; as $\gamma \to 1$ columns 4 and 5 will become highly correlated with columns 1 and 2, respectively. $X_i (i = 1, 2, ..., 5)$ are orthogonal column vectors:

$$\begin{aligned}
X_1^T &= \left(\ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\right)\\
X_2^T &= \left(1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\ \ 1\ \text{-}1\right)\\
X_3^T &= \left(\text{-}1\ \text{-}1\ \ 1\ \ 1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\right)\\
X_4^T &= \left(\text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \ 1\ \ 1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \ 1\ \ 1\right)\\
X_5^T &= \left(\text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \text{-}1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\right)
\end{aligned} \tag{3.19}$$

The setings in *X* matrix represent a 2^4 experimental design when γ=0. The additive noise ε is independently and identically distributed, and follows a normal distribution with zero mean and constant variance.

To find the best SM for making predictions at desired settings corresponding to W, parameters can be ranked and selected using the r_{CCW} criterion according to the procedure outlined in Table 3.5. In this example, the influence of the correlation factor γ and the noise variance σ^2 are

Table 3.5 Proposed Forward-Selection Method to Rank and Select Parameters Using r_{CCW}

1	Start with p candidate simplified models, each with only one parameter to estimate and the remaining p-1 parameters fixed at their initial values.
2	Set the iteration counter at $k=1$. Fit the parameters for each candidate SM using weighted nonlinear least-squares regression, and then calculate p different $r_{CW,1}$ and $r_{CCW,1}$ values using Equations(T3.4.1) to (T3.4.11) in Table 3.4 and a W matrix that corresponds to the operating conditions of interest. The lowest value of $r_{CCW,1}$ corresponds to the parameter that should be ranked first because it provides a model with the lowest MSE predictions at the operating conditions corresponding to W.
3	With k parameter ranked, consider p-k candidate models where each contains all k previously ranked parameters plus one other additional parameter. Fit the parameters for each candidate model using weighted nonlinear least-squares regression, and then calculate p-k different $r_{CW,k}$ and $r_{CCW,k}$ values using Equations(T3.4.1) to (T3.4.11) in Table 3.4 for the given choice of the operating conditions. The added parameter leading to the lowest value of $r_{CCW,k}$ is ranked next.
4	Steps 3 should be repeated until all p parameters are ranked. The parameter subset that should be included in the best model (in terms of MSE at the chosen operating conditions) contains the k parameters that correspond to the lowest value of $r_{CCW,k}$ over all of the different values of k.

investigated to understand how these factors affect the selection of a particular SM. Also, several different choices of the conditions specified in W are investigated. First, values of the criterion R_{CCW}, computed using the true parameter values in Equation(3.17) and the definition in Equation(3.11), are used to investigate the parameter ranking that should be obtained and the SM that should be selected. Next, 10000 sets of simulated experimental data (Monte-Carlo

simulations) are used to examine the effectiveness of the estimator r_{CCW} in the more practical situation where true values of the model parameters are unavailable. Table 3.6 shows parameter ranking and model selection results for low and high levels of the noise variance ($\sigma^2 = 0.1$ and $\sigma^2 = 10$, respectively) for a low level of correlation between columns of X ($\gamma = 0.1$) when W is a subset of the rows in X (i.e., the 2nd, 6th, 10th and 14th rows of the X matrix which were chosen arbitrarily). Table 3.7 shows similar results for low and high levels of the correlation factor ($\gamma = 0.1$ and $\gamma = 0.9$) at a fixed low value of the noise variance ($\sigma^2 = 0.1$). As shown in Table 3.6, at low correlation, when the noise variance is low ($\sigma^2 = 0.1$) β_4 is the top-ranked parameter with R_{CCW}=0.3753, followed by β_5, β_3, β_1 and β_2. The same ranking is obtained when the noise variance increases to $\sigma^2 = 10$ and the data are less reliable. When $\sigma^2 = 0.1$, Table 3.6 indicates that the EM, which contains all the five parameters, should be used because the lowest value of R_{CCW} corresponds to the EM. Parameter ranking and model selection results from 10000 Monte-Carlo simulations when $\sigma^2 = 0.10$ and $\gamma = 0.1$ are shown in Figure 3.1 and Figure 3.2, respectively. These simulation results are consistent with the theoretical results in Table 3.6.

Table 3.6 Effect of Noise Variance at Low Correlation on the Optimal Parameters to Estimate Using the Proposed Algorithm in Table 3.5 and the Five- Parameter Linear Example, $W \subset X$

	γ=0.1			
# Param	σ^2=0.1		σ^2=10	
	Rank	R_{CCW}	Rank	R_{CCW}
1	β_4	0.3753	β_4	-0.2437
2	β_5	0.2209	β_5	-0.1834
3	β_3	2.4650	β_3	-0.0991
4	β_1	2.4070	β_1	-0.0378
5	β_2	0	β_2	0

Table 3.7 Effect of Correlation at Low Noise Variance on the Optimal Parameters to Estimate Using the Proposed Algorithm in Table 3.5 and the Five- Parameter Linear Example, $W \subset X$

	σ^2=0.1			
# Param	γ=0.1		γ=0.9	
	Rank	R_{CCW}	Rank	R_{CCW}
1	β_4	0.3753	β_3	2.7305
2	β_5	0.2209	β_4	4.3966
3	β_3	2.4650	β_2	0.0010
4	β_1	2.4070	β_1	-0.0585
5	β_2	0	β_5	0

According to Figure 3.1, β_4 is ranked first 73.05% of the time. $\beta_5, \beta_3, \beta_1$ and β_2 are ranked 2nd, 3rd, 4th and 5th most of the time. As shown in Figure 3.2, when $\sigma^2 = 0.10$ and $\gamma = 0.1$, the 5-parameter model (i.e. EM) is selected 81% of the time using r_{CCW}, followed by the two- and one-parameter models (11% and 8% of the time, respectively). It is somewhat unexpected that a three- or four-parameter model was never selected. Figure 3.3 and Figure 3.4 show the parameter ranking and model selection results from Monte-Carlo simulations when $\sigma^2 = 10$ and $\gamma = 0.1$. As expected, using this higher noise variance r_{CCW} tends to select less-complicated models with fewer parameters. As shown in Figure 3.3, the frequency of the parameter ranking from the Monte-Carlo simulations agrees, more or less, with the theoretical results in Table 3.6.

According to Figure 3.4, when $\sigma^2 = 10$, an SM that only contains β_4 is selected as the best model 59% of the time using r_{CCW}.

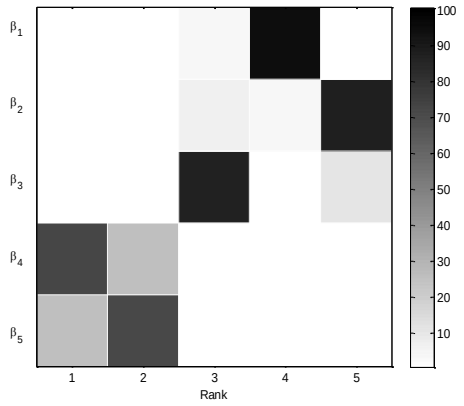

Figure 3.1 Parameter Ranking Frequency in the Linear Five-Parameter Model Using the Proposed MSE-Based Ranking Criterion Outlined in Table 3.5, σ^2=0.1, γ=0.1, $W \subset X$

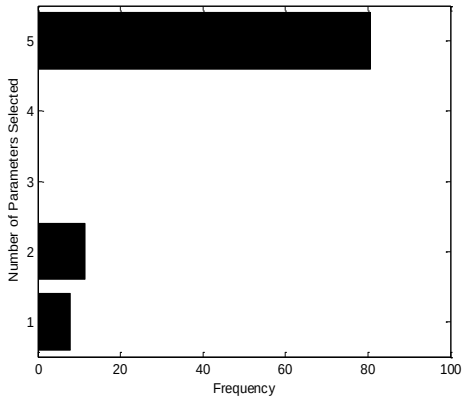

Figure 3.2 Frequency of the Number of Selected Parameters in the Linear Five-Parameter Model Using the Proposed MSE-Based Ranking Criterion Outlined in Table 3.5, σ^2=0.1, γ=0.1, $W \subset X$

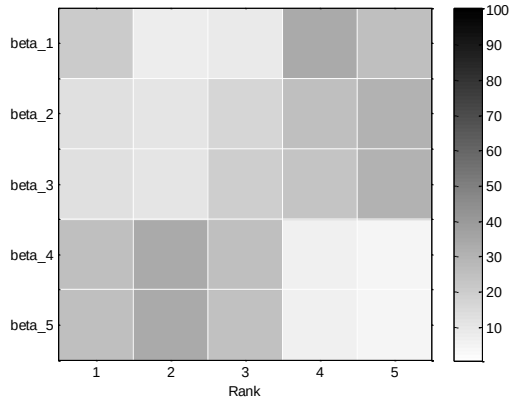

Figure 3.3 Parameter Ranking Frequency in the Linear Five-Parameter Model Using the New MSE-Based Ranking Criterion Outlined in Table 3.5, σ^2=10, γ=0.1, $W \subset X$

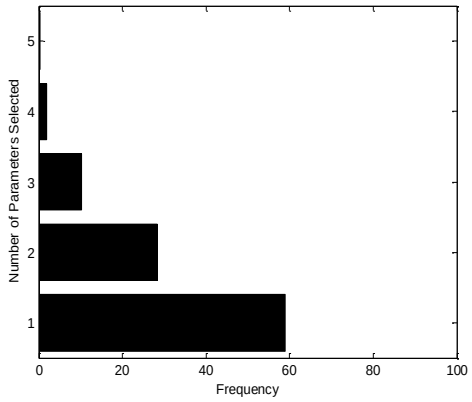

Figure 3.4 Frequency of the Number of Selected Parameters in the Linear Five-Parameter Model Using the Proposed MSE-Based Ranking Criterion Outlined in Table 3.5, σ^2=10, γ=0.1, $W \subset X$

According to Table 3.7, when ($\gamma = 0.9$), β_3 is selected as the top-ranked parameter using $R_{CCW} = 2.7305$, followed by β_4, β_2, β_1 and β_5. The SM which includes the four top-ranked parameters (i.e., $\beta_3, \beta_4, \beta_2$ and β_1) is selected to provide the best predictions for the settings in W.

Monte-Carlo simulation results when $\sigma^2 = 0.1$ and $\gamma = 0.9$ using r_{CCW} are consistent with the theoretical results obtained using R_{CCW} and are summarized elsewhere[1]. Table 3.8 shows the results of various parameter ranking and model selection techniques using the simulated data sets generated with $\sigma^2 = 10$ and $\gamma = 0.1$ for the same W (i.e., the 2nd, 6th, 10th and 14th rows of the X matrix) used to obtain the results above. Results in the first column were

Table 3.8 Comparison of Parameter Ranking and Selection Using Different Methods Outlined in Table 3.2 to Table 3.5 and the Five-Parameter Linear Model, σ^2=10, γ=0.1, $W \subset X$

	Rank			
Parameter	**Orthogonalization+ r_{CC} method in Table 3.2**	**McLean's MSE-based method in Table 3.3**	**Orthogonalization+ r_{CCW} method in Table 3.4**	**Proposed MSE-based method in Table 3.5**
β_1	1*	1*	1*	4
β_2	2	2	2	5
β_3	3	3	3	3
β_4	4	4	4	1*
β_5	5	5	5	2

obtained by ranking the parameters using the orthogonalization procedure in Table 3.1 and then selecting the optimal number of parameters to estimate using the r_{CC} criterion as described in Table 3.2. The asterisk indicates that estimating only β_1 should provide the best predictions, on average, for responses that correspond to the settings in X. The results in the second column, which are the same as those in the first column, were obtained using r_{CC} to rank the parameters and determine the optimal number to estimate, according to the method describe in Table 3.3. The results in the third column, which are the same as those in the first two columns, were obtained by first ranking the parameters using orthogonalization (Table 3.1) and then selecting the model with the smallest value of r_{CCW}. Results in the final column, which were obtained

using the proposed methodology from Table 3.5, indicated that a different one-parameter model, which contains only β_4 should provide the best predictions, on average, for the settings in W.

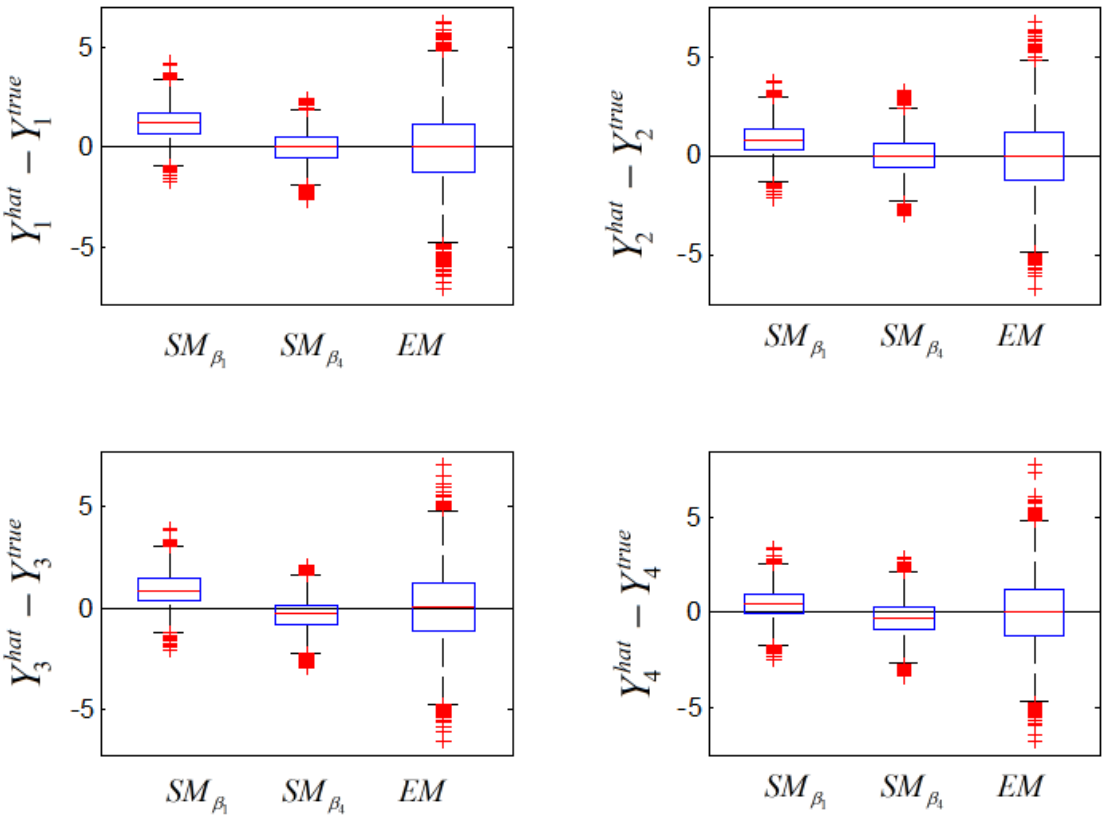

Figure 3.5 Comparison of Model Predictions at Setting in W ($W \subset X$) for the Five-Parameter Linear Model when σ^2=10, γ=0.1. SM_{β_1} Contains only β_1, SM_{β_4} Contains only β_4. SM_{β_1} Was Selected Using the Procedures in Table 3.2, Table 3.3 and Table 3.4. SM_{β_4} Was Selected Using the Proposed Procedure in Table 3.5

Comparison of the model predictions obtained using parameters estimated from the 10000 simulated data sets for the different selected models in Table 3.8 are shown in Figure 3.5 using the four different settings in W. Results in this figure are boxplots of predictions made using settings in the first row of W. The label $SM_{\beta 1}$ indicates predictions obtained using the model with only β_1 estimated, which was selected in the first three columns of Table 3.8 and the label $SM_{\beta 4}$ indicates predictions obtained using the model with only β_4 estimated, which was selected

using the proposed methodology. As shown in Figure 3.5 the $SM_{\beta 4}$ model gives better predictions, on average, for the 10000 data sets generated than the $SM_{\beta 1}$ model gives. Both of these SMs tend to give better predictions than when all of the parameters in the EM are estimated. These results confirm that the proposed methodology is useful for deciding which parameters should be estimated when accurate predictions are desired at a specified set of operating conditions, for the particular model studied and the particular choice that was made for *W*. Analogous results were obtained for a different choice of *W*

$$W = \begin{bmatrix} 1 & -2 & -2 & -2 & -2 \\ 1 & -2 & -2 & 2 & -2 \\ 1 & -2 & -2 & -2 & 2 \\ 1 & -2 & -2 & 2 & 2 \end{bmatrix} \tag{3.20}$$

involving extrapolation to independent variable conditions that are beyond the range from -1 to +1 used to obtain the experimental data. Table 3.9 shows that the proposed methodology indicates that β_3 is the only parameter that should be estimated in this situation using data obtained with σ^2=0.1 and γ=0.9. Boxplots in Figure 3.6 confirm that the predictions from this model are superior, on average, to those obtained using the EM and also $SM_{\beta 1, \beta 2, \beta 3}$, which was selected using the other methods.

Table 3.9 Comparison of Parameter Ranking and Selection Using Different Methods Outlined in Table 3.2 to Table 3.5 and the Five-Parameter Linear Model, σ^2=0.1, γ=0.9, $W \not\subset X$

	Rank			
Parameter	**Orthogonalization+ r_{CC} method in Table 3.2**	**McLean's MSE-based method in Table 3.3**	**Orthogonalization+ r_{CCW} method in Table 3.4**	**Proposed MSE-based method in Table 3.5**
β_1	1	1	1	3
β_2	2	2	2	4
β_3	3*	3*	3*	1*
β_4	4	4	4	5
β_5	5	5	5	2

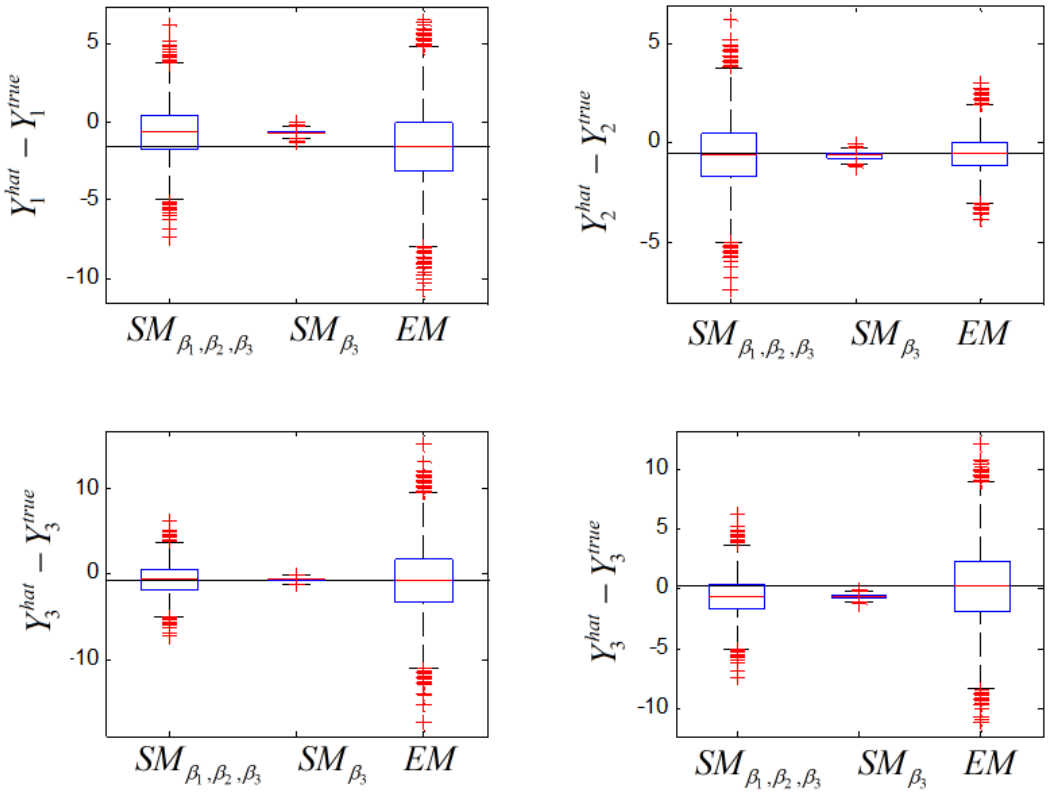

Figure 3.6 Comparison of Model Predictions at Setting in W ($W \not\subset X$) for the Five-Parameter Linear Model when σ^2=0.1, γ=0.9. $SM_{\beta_1,\beta_2,\beta_3}$ Contains Three Parameters β_1, β_2 and β_3. SM_{β_3} Contains only β_3 . $SM_{\beta_1,\beta_2,\beta_3}$ Was Selected Using the Procedures in Table 3.2, Table 3.3 and Table 3.4. SM_{β_3} Was Selected Using the Proposed Procedure in Table 3.5.

Note that Wu et al. constructed confidence intervals for the critical ratio R_{CC}[17], which is appropriate for comparing SMs with the complex model when W=X. They demonstrated that these confidence intervals tend to be wide when the number of parameters included in the SM is small compared to the total number of parameters in the full model. This situation happens when the data are not very informative about the model parameters, resulting in statistical tests that cannot easily discriminate between different models. As a result, we anticipate that confidence intervals for R_{CCW} would also be wide when selecting SMs rather than the full model, and will therefore not be useful for determining which SM is better at the 95% confidence level. Discrimination at lower confidence levels would be possible.

In summary, a simple five-parameter linear regression model was used to study the influence of noise variance, correlation, and different choices of W on the parameters selected for estimation using the proposed forward selection MSE-based technique outlined in Table 3.5. When data are very informative (i.e. low noise variance and low correlation in the experimental design), a large number of parameters are selected for estimation using r_{CCW} and the other techniques used for comparison. However, when data are less informative, either due to high noise variance or high correlation, a smaller number of model parameters tends to be selected for estimation. Estimating the parameters selected via the proposed methodology leads to better predictions, on average, than estimating the parameters selected by the competing techniques.

3.6. Nonlinear Regression Example

A batch reactor model from the Dow Chemical Company, consisting of several differential and algebraic equations (DAEs), is shown in Table 3.10[54]. In this model, T is the reactor temperature in K, $T_0 = 340.15\ \mathrm{K}$ is a reference temperature, $\left[Q^+\right] = 0.0131\ \mathrm{mol\,kg^{-1}}$ is the initial catalyst

concentration and $R = 1.986$ cal mol^{-1}K^{-1} is the ideal gas constant. Four response variables consisting of three measured concentrations and a computed concentration are:

$$\begin{aligned} y_1 &= x_1 + \varepsilon_1 \\ y_2 &= x_2 + \varepsilon_2 \\ y_3 &= x_3 + \varepsilon_3 \\ y_4 &= y_1^0 - y_1 - y_3 \end{aligned} \tag{3.21}$$

where y_1^0 is the initial concentration of component 1. Initial conditions and measurements at three different temperatures, 40 °C, 67 °C, and 100 °C were reported by Biegler et al.[54]. The uncertainties associated with the response variables are all assumed to be $s_{y_i} = 0.02$ mol/kg. The initial parameter values and their corresponding lower and upper bounds are shown in Table 3.11.

Table 3.10 Dow Chemical Batch Reactor Model

$$\frac{dx_1}{dt} = -k_2 x_2 x_8$$
$$\frac{dx_2}{dt} = -k_1 x_2 x_6 + k_{-1} x_{10} - k_2 x_2 x_8$$
$$\frac{dx_3}{dt} = -k_2 x_2 x_8 + k_1 x_4 x_6 - 0.5 k_{-1} x_9$$
$$\frac{dx_4}{dt} = -k_1 x_4 x_6 - 0.5 k_{-1} x_9$$
$$\frac{dx_5}{dt} = -k_1 x_2 x_6 - 0.5 k_{-1} x_{10}$$
$$\frac{dx_6}{dt} = -k_1 x_2 x_6 + k_{-1} x_{10} - k_1 x_4 x_6 + 0.5 k_{-1} x_9$$
$$x_7 = -\left[Q^+\right] + x_6 + x_8 + x_9 + x_{10}$$
$$x_8 = \frac{K_2 x_1}{K_2 + x_7}$$
$$x_9 = \frac{K_3 x_3}{K_3 + x_7}$$

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

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

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

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

Several parameter-estimation studies using this Dow Chemical model have reported numerical difficulties[13,54]. In some cases the model was simplified by reducing the number of differential equations and the number of parameters[57–59]. However, in the current study, the complete model which includes nine parameters is considered. In doing so, it was necessary to select tolerances for the DAE solver (ode15s) and the optimizer (lsqnonlin) in Matlab, carefully.

Table 3.11 Initial Parameter Guesses with Corresponding Lower and Upper Bounds on Plausible Parameter Values

Parameters	Initial Guess	LB	UB
$k_{10}\left(kg\,mol^{-1}h^{-1}\right)$	1.0	0.8	2
$E_1\left(cal\,mol^{-1}\right)$	1×10^4	0.8×10^4	1.2×10^4
$k_{20}\left(kg\,mol^{-1}h^{-1}\right)$	2.0	1.7	3
$E_2\left(cal\,mol^{-1}\right)$	1×10^4	0	2×10^4
$k_{-10}\left(h^{-1}\right)$	2×10^3	1.5×10^3	2.6×10^3
$E_{-1}\left(cal\,mol^{-1}\right)$	1×10^4	0.6×10^4	1.4×10^4
$K_1\left(mol\,kg^{-1}\right)$	3×10^{-16}	1×10^{-16}	5×10^{-16}
$K_2\left(mol\,kg^{-1}\right)$	5×10^{-14}	4.5×10^{-14}	5.5×10^{-14}
$K_3\left(mol\,kg^{-1}\right)$	2×10^{-16}	1.9×10^{-16}	2.1×10^{-16}

Wu et al.[49] used the orthogonalization algorithm in Table 3.1 and their r_{CC} criterion in Table 3.2 to rank and select the parameters for estimation in this model. The results, shown in the first column of Table 3.12, show that a 7-parameter SM (with parameters K_2 and K_3 held constant) was selected as the best SM for making predictions at the settings in X. Later, McLean et al.[13] used the methodology in Table 3.3 to simultaneously rank and select parameters for estimation based on r_{CC}. The results, shown in the second column of Table 3.12, indicate that the same 7-parameter SM was selected, which makes sense because both methods made use of the r_{CC} criterion. However, note that the parameters in the second column are ranked in a different order than in the first column (e.g., parameter E_2 is the top-ranked parameter using the orthogonalization approach and k_{20} is the top-ranked parameter using r_{CC}). Columns three to six in Table 3.12 contain results obtained using the proposed technique based on r_{CCW}, which is outlined in Table 3.5. To study the performance of this new technique, several choices of input settings where good predictions might be desired by the modeler are considered. The following four practical regions of interest are considered: 1) predictions only at the middle temperature ($67\ ^{\circ}\mathrm{C}$), corresponding to $W = W^1$ in third column of Table 3.12; 2) extrapolated predictions for low temperature behavior (i.e., at $T_{low} = 20\ ^{\circ}\mathrm{C}$ and times associated with Biegler's $67\ ^{\circ}\mathrm{C}$ run) corresponding to $W = W^2$ in the fourth column of Table 3.12; 3) predictions at short times, corresponding to the first time and half the first time where data were reported by Biegler et al. for run conditions at all three temperatures (40, 67 and $100\ ^{\circ}\mathrm{C}$ runs), resulting in $W = W^3$, used to generate the fifth column in Table 3.12; 4) predictions at long times corresponding to the final times and twice the final times in the data reported by Biegler et al. at all three temperatures, resulting in $W = W^4$ to produce the sixth column of Table 3.12. Computing the elements of W for

these four different cases is straightforward. For example, W^1 is a matrix with 63 rows (corresponding to the three predicted responses at 21 different times at which data were reported for the run conducted at $67^{\circ}\mathrm{C}$) and 9 columns (corresponding to the 9 model parameters). The elements of W^1 are scaled sensitivity coefficients, similar to those in Equation (3.15), determined by using difference approximations (i.e., by perturbing of each of the nine parameters, one at a time, by 5%). The difference between the Z matrix and W^1 is that only the sensitivity coefficients corresponding to the middle temperate (67 °C) are included in W^1. When computing the corresponding scaled sensitivity matrices, uncertainties in the parameters were set at half the distance between the upper and lower bounds in Table 3.11. Uncertainties in the measurements were all set at $s_{y_i} = 0.02$ mol/kg. Similarly, W^2 is a matrix with 63 rows corresponding to the three predicted responses at 21 different times where predictions for 20 °C operation are desired (i.e., the same times used for the run at 67 °C[54]), and 9 columns corresponding to the 9 model parameters. The elements of W^2 were calculated using simulations at 20 °C where each of the nine parameters was perturbed one at a time by 5%.

Table 3.12 Comparison of Parameter Ranking and Selection Using Different Techniques Outlined in Table 3.2, Table 3.3 and Table 3.5 for the Biegler Nonlinear Batch Reactor Example with Different Choices of *W*.

			Proposed MSE-based method (Table 3.5)			
# Param	**Orthogonalization +r_{CC} (Table 3.2)**	**McLean's MSE-based method (Table 3.3)**	$W = W^1$ [11]	$W = W^2$ [12]	$W = W^3$ [13]	$W = W^4$ [14]
1	E_2	k_{20}	K_1	K_1	k_{10}	K_1
2	K_1	k_{10}	K_2	E_2	K_3	k_{20}
3	E_{-1}	E_{-1}	k_{20}	E_{-1}	k_{-10}	E_1
4	E_1	E_1	E_{-1}	K_2	k_{20}	k_{-10}
5	k_{20}	k_{-10}	k_{-10}	k_{20}	E_2	k_{10}
6	k_{10}	K_1	E_1	k_{-10}	K_2	E_2
7	k_{-10}	E_2	k_{10}	K_3	E_{-1}	E_{-1}
8	K_2	K_2	E_2	k_{10}	K_1	K_2
9	K_3	K_3	K_3	E_1	E_1	K_3

As shown by the shaded area in Table 3.12, a 5-parameter SM, which includes parameters K_1, K_2, k_{20}, E_{-1} and k_{-10}, is selected when W^1 is specified. A 7-parameter SM, with k_{10} and E_1 held constant, is selected when W^2 is used. Similarly, 6- and 4-parameter SMs are selected using W^3 and W^4, respectively. In all four cases, different parameters were selected for estimation using

[11] Predictions at the middle temperature only ($67\ ^\circ C$)

[12] Predictions that extrapolate to low temperature behavior ($20\ ^\circ C$)

[13] Predictions at short times

[14] Predictions at long times

the proposed methodology than the seven parameters that were selected using the r_{CC}-based methods in the first two columns of Table 3.12, which do not account for the operating regions of interest.

In case of W^1 (i.e., middle temperature case), the activation energy parameters E_1 and E_2 are excluded from the selected SM, and E_{-1} is ranked lower than in the first two columns of Table 3.12. This result makes sense because the activation energy parameters are relatively unimportant for making predictions at the reference temperature of 67 °C. Parameter K_3 continues to rank at the bottom of the list because of its small initial uncertainty and high correlation with other parameters that ranked higher in the list of ranked parameters. In case of W^2 (i.e., extrapolation to low-temperature behaviour), however, the activation energy parameters are important. Parameters E_2 and E_{-1} rank near the top of the list. However, E_1 is left out of the selected SM due to correlation with other selected parameters. In case of W^3 (i.e., extrapolation to short-time behaviour), parameter k_{10} ranks at the top of the list. Parameter k_{10} is particularly influential at short times, because it is the rate constant for a reaction between two reactants that are present at the start of the batch. However, in case of W^4 (i.e., extrapolation to long-time behaviour), parameter k_{10} ranks 5th and is not selected for estimation due to its relative unimportance for making accurate predictions at long reaction times.

Figure 3.7 shows model predictions obtained using the initial parameter values (dotted lines) and predictions obtained using estimated parameters (solid lines) for the SM selected using McLean's MSE-based method (third column of Table 3.12). The predictions using this seven-parameter SM match the data obtained at 40 °C very well over the entire duration of the experiment because the selection criterion seeks to obtain good predictions for all of the data points.

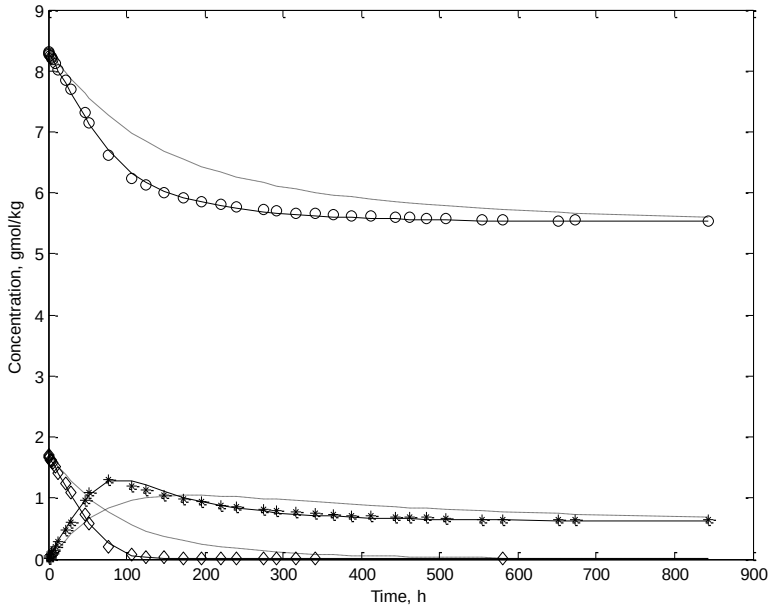

Figure 3.7 Model Predictions Obtained by Using the Seven-Parameter SM Arising from McLean's MSE-Based Method (Solid Lines) and Predictions Made Using Initial Parameter Values (Dotted Lines) for Reactor Operation at 40 °C. Corresponding Data Values Are Shown Using Symbols ◊ y_1; o y_2; * y_3.

Figure 3.8 shows model predictions obtained using parameter estimates from the SM selected using the proposed methodology with W^4 to ensure accurate predictions at long times (final column of Table 3.12). As expected, the resulting four-parameter SM predicts the data very well at the final time (t=843 h). However, predictions are not as good as predictions in Figure 3.7 at shorter times because the choice of W^4 indicated that predictions at these times are not important. Good predictions (not shown) are obtained at short times when parameter estimates corresponding to the six-parameter SM in the sixth column are used.

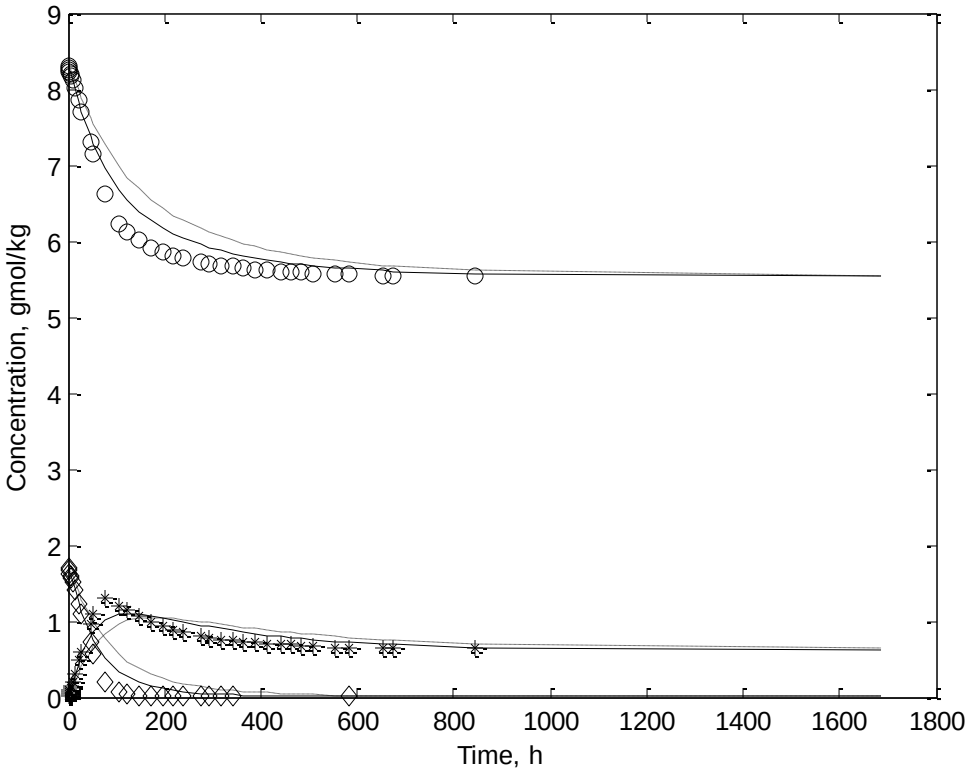

Figure 3.8 Model Predictions Obtained by Using the Four-Parameter SM Arising from the Proposed Method with W4 for Good Predictions at Long Times (Solid Lines) and Predictions Made Using Initial Parameter Values (Dotted Lines) for Reactor Operation at 40 °C. Corresponding Data Values Are Shown Using Symbols ◊ y1; o y2; * y3.

In summary, the proposed methodology for simultaneous parameter ranking and selection was used to rank and select parameters in a nonlinear nine-parameter batch reactor model to find which SM gives the best model predictions (in terms of MSE) for four desired operating regions of interest. In the cases studied here, the SMs selected by the proposed methodology had equal or fewer parameters compared to the SMs selected by the orthogonalization algorithm and the methodology of McLean et al. because some of the parameters are relatively unimportant for achieving good model predictions at the different operating conditions of interest.

3.7. Conclusion

A MSE-based forward selection technique is proposed for simultaneous parameter ranking and selection based on our previously derived critical ratio, r_{CCW}[1]. This new technique uses the available data set, along with information about the particular operating region of interest, to rank the parameters and select the best model, with the lowest anticipated mean-squared prediction error.

A simple five-parameter linear regression model is used to investigate the effects of noise variance, correlation in experimental design, and the operating region of interest on parameter ranking and selection using the proposed MSE-based technique. Monte-Carlo simulations using the statistic r_{CCW} show that results obtained using simulated data agree well with theoretical results. Parameter ranking and model selection results are compared with those from an orthogonalization algorithm that uses the statistics r_{CC} or r_{CCW} to select the optimal number of parameters to estimate. When data are reliable (low noise variance and low correlation) all techniques tend to choose relatively complex models with a large number of estimated parameters. When data are less informative (i.e., there is high correlation or high noise variance), all of the techniques tend to choose simple models with fewer parameters. Note that proposed method often produces different parameter rankings and selects different parameters for estimation compared with orthogonalization-based methods, because the proposed method incorporates information about the operating region of interest.

A nine-parameter nonlinear dynamic batch-reactor model, consisting of six ODEs and four algebraic equations, is used to illustrate the proposed methodology and to compare it with previous methods[13,49]. Four different choices for the operating region of interest are considered to illustrate the proposed methodology. The parameters selected for estimation are those that are

important for predicting reactor behaviour in the four different situations. The proposed parameter ranking and selection technique should be helpful to modelers who want to obtain the best possible predictions at desired operating conditions, particularly when data are limited. In future, it will be important to extend the algorithm so that it can be used in situations where the FIM for the available experimental data is not invertible.

3.8. References

(1) Eghtesadi, Z.; Wu, S.; McAuley, K. B. Development of a Model Selection Criterion for Accurate Model Predictions at Desired Operating Conditions. *Ind. Eng. Chem. Res.* **2013**, *52*, 12297–12308.

(2) Maria, G. A Review of Algorithms and Trends in Kinetic Model Identification for Chemical and Biochemical Systems. **2004**, *18*, 195–222.

(3) Foss, B. a.; Lohmann, B.; Marquardt, W. A field study of the industrial modeling process. *Model. Identif. Control A Nor. Res. Bull.* **1998**, *19*, 153–174.

(4) Shmueli, G.; Koppius, O. R. Predictive Analytics in Information Systems Research. *Mis Q.* **2011**, *35*, 553–572.

(5) Shmueli, G. To Explain or to Predict? *Stat. Sci.* **2010**, *25*, 289–310.

(6) McLean, K. A. P. Obtaining the Best Model Predictions and Parameter Estimates Using Limited Data, 2011.

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

(8) Hocking, R. R. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. *Biometrics* **1976**, *32*, pp. 1–49.

(9) Ehrenberg, A.; Bound, J. Predictability and prediction. *J. Roy. Stat. Soc. Ser. A* **1993**, *156*, 167–206.

(10) Geisser, S. The predictive sample reuse method with applications. *J. Amer. Stat. Assoc.* **1975**, *70*, 320–328.

(11) Kravaris, C.; Hahn, J.; Chu, Y. Advances and selected recent developments in state and parameter estimation. *Comput. Chem. Eng.* **2013**, *51*, 111–123.

(12) Lund, B. F.; Foss, B. a. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. *Automatica* **2008**, *44*, 278–281.

(13) McLean, K. A. P.; Wu, S.; McAuley, K. B. Mean-Squared-Error Methods for Selecting Optimal Parameter Subsets for Estimation. *Industrial & Engineering Chemistry Research*, 2012, *51*, 6105.

(14) Perregaard, J. Model simplification and reduction for simulation and optimization of chemical processes. *Comput. Chem. Eng.* **1993**, *17*, 465–483.

(15) Wang, F. Y.; Zhu, Z. H.; Massarotto, P.; Rudolph, V. A simplified dynamic model for accelerated methane residual recovery from coals. *Chem. Eng. Sci.* **2007**, *62*, 3268–3275.

(16) Weijers, S. R., & Vanrolleghem, P. A. A procedure for selecting best identifiable parameters in calibrating activated sludge model no 1 to full-scale plant data. *Water Sci. Technol.* **1997**, *36*, 69–79.

(17) Wu, S.; Harris, T. J.; McAuley, K. B. The use of simplified or misspecified models: Linear case. *Can. J. Chem. Eng.* **2007**, *85*, 386–398.

(18) Dochain, D.; Vanrolleghem, P. A. *Dynamical Modelling and Estimation inWastewater Treatment Processes*; IWA Publishing: London, UK, 2001.

(19) Benyahia, B.; Latifi, M. A.; Fonteix, C.; Pla, F. *Modeling of a batch emulsion copolymerization reactor in the presence of a chain transfer agent : estimability analysis , parameters identification and experimental validation* .; Elsevier B.V., 2011; Vol. 29, pp. 126–130.

(20) Ben-Zvi, A.; McAuley, K.; McLellan, J. Identifiability study of a liquid-liquid phase-transfer catalyzed reaction system. *AIChE J.* **2004**, *50*, 2493–2501.

(21) Chu, Y.; Huang, Z.; Hahn, J. Improving prediction capabilities of complex dynamic models via parameter selection and estimation. *Chem. Eng. Sci.* **2009**, *64*, 4178–4185.

(22) Quaiser, T.; Monnigmann, M. Systematic Identifiability Testing for Unambiguous Mechanistic Modeling—Application to JAK-STAT, MAP Kinase, and NF-kB Signaling Pathway Models. *BMC Syst. Biol* **2009**, *3*.

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

(24) Hawkins, D. M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12.

(25) Rao, P. Some Notes on Misspecification in Multiple Regressions. *Am. Stat.* **1971**, *25*, pp. 37–39.

(26) Wu, S.; McAuley, K. B.; Harris, T. J. Selection of simplified models: II. Development of a model selection criterion based on mean squared error. *Can. J. Chem. Eng.* **2011**, *89*, 325–336.

(27) Moore, C.; Doherty, J. Role of the calibration process in reducing model predictive error. *Water Resour. Res.* **2005**, *41*, W05020.

(28) Maria, G.; Rippin, D. Note concerning two techniques for complex kinetic pathway analysis. *Chem. Eng. Sci.* **1993**, *48*, 3855–3864.

(29) Montgomery, D. C.; Peck, E. A.; Vining, G. G. *Introduction to Linear Regression Analysis*; 3rd ed.; John Wiley & Sons: NY, 2001.

(30) McLean, K. A. P.; McAuley, K. B. Mathematical modelling of chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Can. J. Chem. Eng.* **2011**, *90*, 351–366.

(31) Miller, A. J. *Subset selection in regression*; Chapman and Hall: London England ; New York, 1990.

(32) Brun, R., Reichert, P., & Kunsch, H. R. Practical identifiability analysis of large environmental simulation models. *Water Resour. Res.* **2001**, *37*, 1015–1030.

(33) Sandink, C. a.; McAuley, K. B.; McLellan, P. J. Selection of Parameters for Updating in On-line Models. *Ind. Eng. Chem. Res.* **2001**, *40*, 3936–3950.

(34) Sun, C. L., & Hahn, J. Parameter reduction for stable dynamical systems based on Hankel singular values and sensitivity analysis. *Chem. Eng. Sci.* **2006**, *61*, 5393–5403.

(35) Yao, K. Z.; Shaw, B. M.; Kou, B.; McAuley, K. B.; Bacon, D. W. Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimatibility analysis and experimental design. *Polym. React. Eng.* **2003**, *11*, 563–588.

(36) Lund, B. F.; Foss, B. a. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. *Automatica* **2008**, *44*, 278–281.

(37) Walter, E., & Pronzato, L. *Identification of parametric models from experimental data*; Springer, 1997.

(38) Blanchet, F. G., Legendre, P., & Borcard, D. Forward selection of explanatory variables. *Ecology* **2008**, *89*, 2623–2632.

(39) Chu, Y.; Hahn, J.; Chu, Y. F., & Hahn, J. Parameter set selection for estimation of nonlinear dynamic systems. *AIChE J.* **2007**, *53*, 2858–2870.

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

(41) Lopez, D. C. C.; Barz, T.; Penuela, M.; Villegas, A.; Ochoa, S.; Wozny, G. Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production. *Biotechnol. Prog.* **2013**, *29*, 1064–1082.

(42) Ngo, V. V; Latifi, M. A.; Simonnot, M.-O. Estimability Analysis and Optimisation of Soil Hydraulic Parameters from Field Lysimeter Data. *Transp. POROUS MEDIA* **2013**, *98*, 485–504.

(43) Meshram, M.; Naderi, S.; McConkey, B.; Ingalls, B.; Scharer, J.; Budman, H. Modeling the coupled extracellular and intracellular environments in mammalian cell culture. *Metab. Eng.* **2013**, *19*, 57–68.

(44) Jayasankar, B. R.; Ben-Zvi, A.; Huang, B. Identifiability and estimability study for a dynamic solid oxide fuel cell model RID F-2736-2010. *Comput. Chem. Eng.* **2009**, *33*, 484–492.

(45) Chu, Y.; Hahn, J. Generalization of a Parameter Set Selection Procedure Based on Orthogonal Projections and the D-Optimality Criterion. *AIChE J.* **2012**, *58*, 2085–2096.

(46) Kou, B.; McAuley, K. B.; Hsu, C. C.; Bacon, D. W. Mathematical model and parameter estimation for gas-phase ethylene/hexene copolymerization with supported metallocene catalyst. *Macromol. Mater. Eng.* **2005**, *290*, 537–555.

(47) Alberton, K. P. F.; Alberton, A. L.; Di Maggio, J. A.; Díaz, M. S.; Secchi, A. R. Accelerating the parameters identifiability procedure: Set by set selection. *Comput. Chem. Eng.* **2013**, *55*, 181–197.

(48) Wu, S.; McAuley, K. B.; Harris, T. J. Selection of Simplified Models: I. Analysis of Model-Selection Criteria using Mean-Squared Error. *Can. J. Chem. Eng.* **2011**, *89*, 148–158.

(49) Wu, S.; McLean, K. A. P.; Harris, T. J. H.; McAuley, K. B. Selection of optimal parameter set using estimability analysis and MSE-based model-selection criterion. *Int. J. Adv. Mechatron. Syst.* **2011**, *3*, 188–197.

(50) Woloszyn, J. D.; McAuley, K. B. Application of Parameter Selection and Estimation Techniques in a Thermal Styrene Polymerisation Model. *Macromol. React. Eng.* **2011**, *5*, 453–466.

(51) Karimi, H.; Schaffer, M. A.; McAuley, K. B. Kinetic Moded for Non-Oxidative Thermal Degradation of Nylon 66. *Macromol. React. Eng.* **2012**, *6*, 93–109.

(52) Cui, W. J.; McAuley, K. B.; Whitney, R. A.; Spence, R. E.; Xie, T. Mathematical Model of Polyether Production From 1,3-Propanediol. *Macromol. React. Eng.* **2013**, *7*, 237–253.

(53) Zhao, Y. R.; McAuley, K. B.; Puskas, J. E.; Dos Santos, L. M.; Alvarez, A. Mathematical Modeling of Arborescent Polyisobutylene Production in Batch Reactors. *Macromol. THEORY SIMULATIONS* **2013**, *22*, 155–173.

(54) Biegler, L. T.; Damiano, J. J.; Blau, G. E. Nonlinear parameter estimation: A case study comparison. *AIChE J.* **1986**, *32*, 29–45.

(55) Rice, J. A. *Mathematical statistics and data analysis [multimedia]*; Duxbury Press: Belmont, CA, 1995.

(56) Leis, J. R.; Kramer, M. A. The Simultaneous Solution and Sensitivity Analysis of Systems Described by Ordinary Differential Equations. *ACM Trans. Math. Softw.* **1988**, *14*, 34.

(57) Tjoa, I. B.; Biegler, L. T. Simultaneous Solution and Optimization Strategies for Parameter Estimation of Differential- Algebraic Equation Systems. *Ind. Eng. Chem. Res.* **1991**, *30*, 376.

(58) Sulieman, H.; Kucuk, I.; McLellan, P. J. Parametric Sensitivity: a Case Study Comparison. *Comput. Stat. Data Anal.* **2009**, *53*, 2640.

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

Chapter 4 Mean-Squared-Error-Based Method for Parameter Ranking and Selection with Non-Invertible Fisher Information Matrix

4.1. Summary

Two different approaches, using the mean-squared-error-based corrected critical ratio r_{CCW}, are developed and investigated to simultaneously rank and select model parameters to obtain accurate model predictions at key operating conditions when the Fisher information matrix (FIM) is non-invertible. These approaches are evaluated using randomly generated data sets in a linear regression example. Results are compared with those from other techniques which do not consider the operating region of interest to the modeler. The influence of several factors including the quality of the initial parameter guesses, uncertainty ranges for the initial parameter values, noise variance and operating regions of interest are investigated. It is shown that using a reduced FIM with full rank led to more reliable model predictions for a variety of cases than the alternative approach using the pseudo-inverse of the FIM. The proposed reduced-FIM methodology also provided better predictions than a technique that does not consider the operating region where reliable predictions are required. The proposed methodology is illustrated using a nonlinear differential equation model of a polymer film casting process. The proposed methodology is beneficial when it is impossible to accurately estimate all of model parameters from the available data and when the modeler requires accurate predictions in an important operating region.

This chapter has been prepared for submission to AIChE Journal with K B McAuley as coauthor.

4.2. Introduction

Improving accuracy of model predictions has been an important area of research for many years[1–7]. To obtain accurate model predictions, it is necessary to determine appropriate values of the model parameters. In models with large numbers of parameters and limited data it is important to decide which parameters should be estimated and which parameters should be either fixed at nominal values or removed from the model via simplification[8–10]. Furthermore, it is often not possible to estimate all of the parameters in complicated models from the available data[11–13]. Noisy and sparse experimental data, correlated parameter influences, and correlated experimental designs can all cause difficulties in parameter estimation problems[8,14]. On the one hand, estimating too few parameters leads to significant bias in the predictions. On the other, over-fitting or estimating too many parameters decreases the model's predictive ability and leads to large variances in the model predictions[5,15]. Therefore, simplified models (SMs) with fewer parameters are sometimes used instead of complex models[16–18]. With fewer parameters in a SM, the prediction variance decreases and the prediction bias increases compared with the predictions of corresponding complex model that is assumed to be perfectly structured[2,5,19]. When the variance reduction is greater than the bias increase, a SM gives better predictions than the complex model, due to the trade-off between bias and variance[14]. Several parameter subset selection techniques have been developed and used to decide which parameters should be estimated and which parameters should be fixed at their nominal values[5,8,16–18,20–30]. Orthogonalization-based techniques (Algorithm 4.1), which use the FIM to rank model parameters from the most estimable to the least estimable are the most popular methods for parameter ranking and selection, and have been widely used to estimate parameters in complex models[8,31–37]. The FIM summarizes information about the effects of the model parameters on the

model predictions and on uncertainties in the measurements that will be used for parameter estimation. The desire to obtain accurate model predictions motivated researchers to propose criteria based on mean-squared error (MSE) predictions[14]. Several MSE-based techniques (Algorithm 4.2 and Algorithm 4.3) use statistically-derived critical ratios to determine which and how many parameters to estimate[10,14,16,17,20,34,38,39]. MSE-based methods are attractive because they account for the trade-off between bias and variance as additional parameters from the ranked list are estimated[14,39,40]. One of the shortcomings of these techniques is that they usually use the MSE of predictions made at experimental settings corresponding to the available data points when selecting the parameter subset. Recently, an MSE-based criterion has been proposed that considers the operating region where accurate model predictions are desired (Algorithm 4.4)[39]. A forward-selection methodology was developed based on this MSE-based criterion to rank and select the parameters simultaneously, while considering the operating region of interest to the model user (Algorithm 4.5)[41]. Unfortunately, the FIM must be invertible for this algorithm to be used. The main purpose of this paper is to extend our recent methodology for parameter ranking and selection to cases where the FIM is not invertible.

A singular FIM occurs, for instance, in many high-dimensional parameter estimation problems[42,43] and in system identification with over-parameterized models[44]. In many practical chemical engineering models, it is impossible to rank all of the model parameters using orthogonalization-based techniques, due to a non-invertible FIM[25,35,45–47]. A singular FIM also leads to a significant complication in the analysis of parameter estimation problems and for the theory of the Cramér–Rao lower bound (CRB), which is a popular means to bound the variances of unbiased estimators[48].

In the statistics and system identification literature, a singular FIM is often handled by the use of a pseudoinverse of the FIM[42,49]. Li et al. showed that the Moore-Penrose generalized inverse of a singular FIM can be interpreted as the CRB corresponding to the minimum variance among all choices of minimum constraint functions[50]. Stoica et al. derived the CRB as a solution to an unconstrained quadratic maximization problem[48]. They proved that there is no unbiased estimator for parameters with finite variance, except under "unusual conditions"[48]. Rothenberg proved that the singular FIM is equivalent to local non identifiability of the parameters to be estimated[51]. When a FIM is singular, it is common to use biased estimators for the parameters[48]. Ram et al. illustrated how a priori information about some of the model parameters can be used to eliminate the singularity of the FIM[52]. Ben-Haim et al. demonstrated how to construct estimators by introducing constraints or by specifying an appropriate bias function, when the FIM is singular[53].

In the current chapter, first, we introduce some notation and summarize background information about the MSE-based model selection criterion and the algorithm for simultaneous parameter ranking and selection proposed earlier[39,41]. Next, we propose and compare two different approaches for parameter ranking and selection when the FIM is not invertible. A linear regression problem is used to demonstrate these approaches and their effectiveness to choose the best SM for accurate model predictions. These approaches are extended for use in nonlinear models and illustrated using a case study of a polymer film-casting process which contains 28 model parameters.

4.3. Background Information

A MSE-based methodology was developed by Wu et al., based on the corrected critical ratio, R_{CC}, to compare the accuracy of model predictions obtained using different SMs with different

numbers of parameters (Algorithm 4.2). This methodology selects the SM that gives the best predictions at conditions specified by $X \in R^{n \times p}$ corresponding to the data available for parameter estimation[14,17,38]. Note that n is the total number of measurements and p is the total number of parameters in the extended model (EM), which contains all of the model parameters and is believed to be correctly structured so that it could give accurate predictions if appropriate parameter values were available. This critical ratio is the ratio of the squared bias over the variance reduction when p_1 parameters (i.e., the number of parameters included in the SM) are estimated rather than all p parameters contained in the EM. Wu et al. derived an estimator, r_{CC}, for this corrected critical ratio and showed the lowest value of r_{CC} corresponds to the SM that is expected to provide the best predictions (lowest MSE) at the input settings in X. Later, McLean et al. used Wu's r_{CC} criterion to simultaneously rank and select parameters for estimation in a nonlinear model (Algorithm 4.3)[10]. In case of non-invertible FIM, the r_{CC} criterion can be readily used to rank or select parameters[35,45]. The only assumption required is that the weighted least-squares objective function (i.e., Equation (Al4.2.1) in Algorithm 4.2) would not become any lower if additional (highly correlated) parameters that cause the FIM to become non-invertible were estimated in addition to the higher ranked parameters. Using this assumption, parameters that cause a rank-deficient FIM can be held at their nominal values during the ranking and selection procedure and SMs that contain parameters that lead to rank-deficiency can be ignored because they would not lead to the lowest value of r_{CC}. The main disadvantage of using r_{CC} is that the modeler often desires accurate predictions at alternative input settings that are different from the settings in the already available data set. To address this problem, we proposed the revised model selection criterion r_{CCW} (Algorithm 4.4)[39]. Using the trade-off between bias and variance, this revised criterion selects the best SM with the lowest total MSE at desired operating

conditions corresponding to the matrix of settings, $W \in R^{w \times p}$, where w is the total number of important predictions to be made. The model that corresponds to the lowest value of r_{CCW} is expected to give the best predictions (lowest MSE) at the desired operating conditions[39]. The r_{CCW} criterion was developed using a linear regression framework and then extended for use with nonlinear models of the form:

$$Y_i = f(X_i, \theta) + \varepsilon_i \tag{4.1}$$

where Y_i is the measured response corresponding to the i^{th} run condition, X_i is the set of corresponding settings for the explanatory variables, θ is the p-dimensional vector of parameters and ε_i is the random noise. Later, a forward parameter-subset-selection methodology, based on the r_{CCW} criterion, was used to rank parameters to minimize the anticipated MSE at the specified operating conditions(Algorithm 4.5)[41]. To use this technique the FIM, which is $X^{\mathrm{T}}\Sigma^{-1}X$ for linear models and $Z^{\mathrm{T}}Z$ for nonlinear models must be invertible. The scaled sensitivity matrix, Z, contains information about the effects of parameters on model predictions for all experimental runs and responses. The elements of this scaled sensitivity matrix, can be obtained by multiplying local parameter sensitivities S_{ij0} (obtained by linearizing around the nominal parameter values, θ_0) by scaling factors to ensure dimensional consistency[25]:

$$Z_{ij0} = S_{ij0} \frac{s_{\theta_{j0}}}{s_{y_i}} \tag{4.2}$$

where $S_{ij0} = \frac{\partial f(X_i, \theta)}{\partial \theta_j^T}\Big|_{\theta_0}$ is the corresponding element in i^{th} row and j^{th} column of the sensitivity matrix calculated at $\theta=\theta_0$, $s_{\theta j0}$ and s_{yi} are uncertainties in the initial value of parameter θ_j and the

i^{th} measured response, respectively. The scaling factors affect the results of the parameter ranking procedure and therefore should be chosen carefully to reflect the modelers' prior beliefs about plausible parameter values and measurement variances. Measurement uncertainties can be estimated from replicate experiments or can be obtained based on information from measurement device suppliers. Difference approximations can be used to calculate the elements of the local scaled sensitivity matrix. These elements can also be obtained by solving sensitivity equations[54]. Detailed information about the r_{CCW} criterion and the simultaneous parameter ranking and selection procedure based on this criterion are provided elsewhere[39,41] and summarized in Algorithm 4.4 and Algorithm 4.5 When the FIM is not invertible, values for the r_{CCW} criterion cannot be calculated. In this paper, two possible solutions to the problem of non invertibility of the FIM are proposed. The first approach uses the orthogonalization algorithm, summarized in Algorithm 4.1, to rank all parameters, after which only the top k parameters are considered for estimation, where k=rank(FIM). The remaining p-k parameters are fixed at their nominal values. Then the r_{CCW} criterion can be used to simultaneously rank and select from among the k parameters to find the best SM for making predictions at the desired operating points (corresponding to columns labelled r_{CCW}^1 in Table 4.1 and Table 4.2). The second approach uses the Moore-Penrose pseudoinverse of the FIM in place of the inverse of the FIM when computing r_{CCW} (corresponds to columns labelled r_{CCW}^2 in Table 4.1 and Table 4.2).

Algorithm 4.1 Orthogonalization algorithm[26]

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 Z_k.

3 Use Z_k to predict columns in Z using ordinary least-squares

$$\hat{Z}_k = Z_k \left(Z_k^T Z_k \right)^{-1} Z_k^T Z \qquad \text{(Al4.1.1)}$$

and calculate the residual matrix

$$R_k = Z - \hat{Z}_k \qquad \text{(Al4.1.2)}$$

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 Z_k.

6 Advance the iteration counter k 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.

Algorithm 4.2 Wu's Criterion to find the Optimum Number of Parameters for Estimation[55]

1 Rank model parameters from most estimable to least estimable using the orthogonalization algorithm (i.e., Algorithm 4.1).

2 Use weighted least-squares regression to estimate the first parameter from the list, 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. Weighting factors used in the objective function for parameter estimation should be consistent with measurement uncertainties s_{y_i} used for scaling during parameter ranking. For example, for dynamic experiments the weighted-least squares objective function could be:

$$J = \sum_{i=1}^{d}\sum_{l=1}^{c}\sum_{m=1}^{r}\left(\frac{Y_{ilm} - \widehat{Y}_{ilm}}{s_{y_i}}\right)^2 \qquad \text{(Al4.2.1)}$$

where Y_{ilm} is the measured value of the i^{th} response variable at the l^{th} measurement time, for the m^{th} experimental run and $\widehat{Y}_{ilm}$ is the corresponding predicted value.

3 Compute the critical ratio

$$r_{C,k} = \left(J_k - J_p\right)/\left(p-k\right) \qquad \text{` (Al4.2.2)}$$

for $k = 1... p – 1$.

where J_k is the objective function when the top k parameters have been estimated.:

4 For each value of k, compute the corrected critical ratio

$$r_{CC,k} = \frac{p-k}{N}\left(r_{CKub,k} - 1\right) \qquad \text{(Al4.2.3)}$$

where

$$r_{CKub,k} = \max\left(r_{C,k} - 1, \frac{2}{p-k+2} r_{C,k}\right) \qquad \text{(Al4.2.4)}$$

5 Select the value of k corresponding to the lowest value of $r_{CC,k}$ as the appropriate number of parameters to estimate.

Algorithm 4.3 r_{CC} Criterion to Rank and Select Parameter for Estimation[10]

1. Starting with no parameters in the ranked list, formulate p candidate parameter subsets, each with one 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 (Al4.2.1) in Algorithm 4.2 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 (Al4.2.2) to (Al4.2.4) in Algorithm 4.2.
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 (Al4.2.2) to (Al4.2.4) in Algorithm 4.2.
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.

Algorithm 4.4 Using Orthogonalization and the r_{CCW} Criterion to Select Parameter for Estimation[39]

1 Rank model parameters from most estimable to least estimable using the orthogonalization algorithm in Algorithm 4.1.

2 For a linear model, use the experimental data to calculate $r_{CW,k}$ for each k (i.e. for k = 1... p-1) from:

$$r_{CW,k} = \frac{Y^T AY / (p-k)}{Y^T BY / (n-p)} \qquad \text{(Al4.4.1)}$$

where

$$A = \frac{(p-k)(P - P_{1,k}) M^T M (P - P_{1,k})}{Tr\left((P - P_{1,k}) M^T M (P - P_{1,k})\right)} \qquad \text{(Al4.4.2)}$$

$$B = I_n - P \qquad \text{(Al4.4.3)}$$

$$M = W\left(X^T X\right)^{-1} X^T \qquad \text{(Al4.4.4)}$$

$$P = X(X^T X)^{-1} X^T \qquad \text{(Al4.4.5)}$$

$$P_{1,k} = X_{1,k}(X_{1,k}^T X_{1,k})^{-1} X_{1,k}^T \qquad \text{(Al4.4.6)}$$

where $X_{1,k}$ is the sub-matrix of X corresponding to the parameters that are included in the SM for the kth iteration.

For nonlinear model use nonlinear weighted least-squares regression to estimate all of the ranked parameters in the ranked parameter list from step 1. Use experimental data to calculate $r_{CW,k}$ for each k from the residuals ξ defined in Equation (2.44) and A and B from Equations (Al4.4.2) and (Al4.4.3).

$$r_{CW,k} = \frac{(\xi^T) A(\xi) / (p-k)}{(\xi^T) B(\xi) / (n-p)} \qquad \text{(Al4.4.7)}$$

with

$$M = W\left(Z^T Z\right)^{-1} Z^T \qquad \text{(Al4.4.8)}$$

$$P = Z(Z^T Z)^{-1} Z^T \qquad \text{(Al4.4.9)}$$

$$P_{1,k} = Z_{1,k}(Z_{1,k}^T Z_{1,k})^{-1} Z_{1,k}^T \qquad \text{(Al4.4.10)}$$

where $Z_{1,k}$ is the sub-matrix of the scaled local sensitivity matrix Z corresponding to the parameters that are included in the SM for the kth iteration.

3 Calculate corrected critical ratio $r_{CCW,k}$ for each k from:

$$r_{CCW,k} = \frac{Tr\left(M^T M\left(P - P_{1,k}\right)\right)}{w}(r_{CW,k} - 1) \qquad \text{(Al4.4.11)}$$

4 Select the value of k corresponding to the lowest value of $r_{CCW,k}$ as the appropriate number of parameters to estimate.

Algorithm 4.5 r_{CCW} Forward-Selection Method to Rank and Select Parameters[41]

1 Start with p candidate simplified models, each with only one parameter to estimate and the remaining p-1 parameters fixed at their initial values.

2 Set the iteration counter at $k = 1$. Fit the parameters for each candidate SM using weighted nonlinear least-squares regression, and then calculate p different $r_{CW,1}$ and $r_{CCW,1}$ values using Equations (Al4.4.1) to (Al4.4.11) in Algorithm 4.4 and a W matrix that corresponds to the operating conditions of interest. The lowest value of $r_{CCW,1}$ corresponds to the parameter that should be ranked first because it provides a model with the lowest MSE predictions at the operating conditions corresponding to W.

3 With k parameter ranked, consider p-k candidate models where each contains all k previously ranked parameters plus one other additional parameter. Fit the parameters for each candidate model using weighted nonlinear least-squares regression, and then calculate p-k different $r_{CW,k}$ and $r_{CCW,k}$ values using Equations (Al4.4.1) to (Al4.4.11) in Algorithm 4.4 for the given choice of the operating conditions. The added parameter leading to the lowest value of $r_{CCW,k}$ is ranked next.

4 Steps 3 should be repeated until all p parameters are ranked. The parameter subset that should be included in the best model (in terms of MSE at the chosen operating conditions) contains the k parameters that correspond to the lowest value of $r_{CCW,k}$ over all of the different values of k.

4.4. Linear Regression Example

4.4.1. Linear model

Consider a linear seven-parameter EM of the form of $Y_E = X\beta + \varepsilon$ where the true values of the model parameters are:

$$\beta = (\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5\ \beta_6\ \beta_7)^T = \left(1\ \frac{1}{2}\ \frac{1}{3}\ \frac{1}{4}\ \frac{1}{5}\ \frac{1}{6}\ \frac{1}{7}\right)^T \tag{4.3}$$

The additive noise ε is independently and identically distributed, and follows a normal distribution with zero mean and constant variance. The matrix of independent variable settings X is:

$$X = \left(X_1\ \ X_2\ \ X_3\ \ \gamma X_1 + (1-\gamma)X_4\ \ \gamma X_2 + (1-\gamma)X_5\ \ \mathrm{X}_1\ \ \mathrm{X}_2\right) \tag{4.4}$$

where the first three columns in X are orthogonal, the 4th and the 5th columns can be correlated, depending on the value of the correlation factor γ, and the 6th and the 7th columns are the same as first two columns. For example, when $\gamma = 0$, the first five columns in X are orthogonal; as $\gamma \to 1$ columns 4 and 5 will become highly correlated with columns 1 and 2, respectively. $X_i (i = 1, 2, ..., 5)$ are orthogonal column vectors:

$$\begin{aligned}
X_1^T &= (\ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1) \\
X_2^T &= (1\ \text{-1}\ \ 1\ \text{-1}\ \ 1\ \text{-1}\ \ 1\ \text{-1}\ \ 1\ \text{-1}\ \ 1\ \text{-1}\ \ 1\ \text{-1}\ \ 1\ \text{-1}) \\
X_3^T &= (\text{-1}\ \text{-1}\ \ 1\ \ 1\ \text{-1}\ \text{-1}\ \ 1\ \ 1\ \text{-1}\ \text{-1}\ \ 1\ \ 1\ \text{-1}\ \text{-1}\ \ 1\ \ 1) \\
X_4^T &= (\text{-1}\ \text{-1}\ \text{-1}\ \text{-1}\ \ 1\ \ 1\ \ 1\ \ 1\ \text{-1}\ \text{-1}\ \text{-1}\ \text{-1}\ \ 1\ \ 1\ \ 1\ \ 1) \\
X_5^T &= (\text{-1}\ \text{-1}\ \text{-1}\ \text{-1}\ \text{-1}\ \text{-1}\ \text{-1}\ \text{-1}\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1)
\end{aligned} \tag{4.5}$$

The setings in X matrix represent a 2^4 experimental design when γ=0. Notice, in this example that the FIM=$(X^TX)\sigma^{-2}$ is singular because rank(FIM)=5 rather than 7. Formulating the EM in deviations around the assumed parameter initial guesses, β_0, gives:

$$\left(Y_E - X\beta_0\right) = X\left(\beta - \beta_0\right) + \varepsilon \tag{4.6}$$

This model form is useful for understanding the influence of good or bad initial guesses (nominal values) in parameter ranking and selection procedure. In this example, the influence of the noise variance σ^2, correlation factor γ, initial parameter guesses β_0, and uncertainties in the initial parameter guesses $s_{\beta_0}$are investigated to understand the effects of these factors on the selection of a particular SM. Two different choices of the conditions specified in W are also investigated (i.e., W=W^{sub} contains a subset of the rows in X) or involves extrapolation away from conditions in X (i.e., W=W^{ext})). W^{sub} contains the 2nd, 6th, 10th and 14th rows of the X matrix which were chosen arbitrarily, and W^{ext} is:

$$W^{ext} = \begin{bmatrix} 1 & -2 & -2 & -2 & -2 & 1 & -2 \\ 1 & -2 & -2 & 2 & -2 & 1 & -2 \\ 1 & -2 & -2 & -2 & 2 & 1 & -2 \\ 1 & -2 & -2 & 2 & 2 & 1 & -2 \end{bmatrix} \tag{4.7}$$

4.4.2. r_{CC} and r_{CCW} algorithm results

The effectiveness of the two different approaches for parameter ranking and selection when the FIM is non-invertible is investigated. 10000 sets of simulated experimental data are generated and used to compute r_{CCW}, so that the best SM is selected 10000 times.

In Table 4.1 and Table 4.2, results are shown for $W=W^{sub}$ and $W=W^{ext}$, respectively. Three different scenarios for the initial parameter guesses and their uncertainties (i.e., Case 1, Case 2 and Case 3) are considered. In Case 1, the initial parameter guesses are quite accurate (10% higher than the true values for all parameters) and in Cases 2 and 3, the initial parameter guesses are worse and the corresponding uncertainties are higher, as indicated in the tables captions. In Table 4.1 and Table 4.2, two levels for the correlation factor γ (i.e., γ=0.1 and γ=0.9), and multiple levels for the noise variance σ^2 (e.g., σ^2=0.01 and σ^2=1) are considered. Also, two approaches for parameter ranking and selection (i.e., r_{CCW}^1 (using orthogonalization and fixing of parameters to achieve an invertible FIM) and r_{CCW}^2 (using a pseudoinverse)) are investigated. These results are compared against the typical strategy[35,47,55] that uses orthogonolization and r_{CC} (without considering W) to study the importance of considering W when ranking and selecting parameters. Note that when the r_{CCW}^1 approach is used, only the first five columns corresponding to the first five parameters are included in W^{ext} because the 6th and 7th parameters are automatically removed during the orthogonalization/ranking step. The most common parameter rankings among the 10000 data sets are shown in Table 4.1 and Table 4.2. For example in Table 4.1, Case1, when γ=0.1 and σ^2=0.001 both approaches tend to select β_4 as the top ranked parameter, followed by β_5 and β_3. The * indicates that the five-parameter model was chosen most often. Shaded boxes indicate the lowest average MSE for predictions from the 10000 generated data sets. These shaded boxes indicate that, in almost all comparisons shown in Table 4.1 and Table 4.2, the r_{CCW}^1 methodology performs better than r_{CCW}^2 and r_{CC}. Results obtained using the pseudoinverse approach are less reliable. In both Table 4.1 and Table 4.2, in Case3, additional calculations have been performed to include higher values for the noise variance. Similar trends are obtained for other initial parameter guesses that are different than those in Cases 1, 2 and 3.

Boxplots in Figure 4.1 confirm that the SM1 predictions based on the r_{CCW}^{1} approach are better, on average, compared to SM2 predictions obtained based on the r_{CCW}^{2} approach, and compared to SM0 predictions obtained based on McLean's r_{CC} approach and compared to the EM predictions. More or less similar trends are observed for other model responses and conditions as they are shown in Figure A4.1 and Figure A4.2 in Appendix 4.1. According to Figure 4.1, there is no bias in the EM model predictions as it is expected; however, there are higher variances in the EM model predictions than there are in the SMs model predictions such as SM0, SM1 and SM2 predictions.

Based on the results presented in this section, it seems that the first approach for parameter ranking and selection performs better than the second. After using the orthogonalization algorithm to decide which parameters cannot be ranked, the reduced FIM obtains full rank. The r_{CCW} criterion can then be used to simultaneously rank and select parameters to find the best SM for making predictions at the specified operating conditions of interest.

Table 4.1 Parameter ranking and selection results corresponding to W^{sub} and two different approaches when FIM is non-invertible. Case 1: $\beta_0 = 1.1*\beta$, $s_{\beta 0} = 0.1*\beta_0$; Case 2: $\beta_0 = 1.5*\beta$, $s_{\beta 0} = \beta_0$; Case 3: $\beta_0 = 10*\beta$, $s_{\beta 0} = 0.1*\beta_0$. $r_{CCW}^{\ 1}$ corresponds to the first approach and $r_{CCW}^{\ 2}$ corresponds to the second approach for parameter ranking and selection when FIM is not invertible.

	Case 1											
	σ^2=.001						**σ^2=.1**					
	γ=0.1			**γ=0.9**			**γ=0.1**			**γ=0.9**		
	r_{CC}	**$r_{ccw}^{\ 1}$**	**$r_{ccw}^{\ 2}$**	**r_{CC}**	**$r_{ccw}^{\ 1}$**	**$r_{ccw}^{\ 2}$**	**r_{CC}**	**$r_{ccw}^{\ 1}$**	**$r_{ccw}^{\ 2}$**	**r_{CC}**	**$r_{ccw}^{\ 1}$**	**$r_{ccw}^{\ 2}$**
Rank β_1	1	4	6	1	4*	5*	1*	4	7	1*	5	7
Rank β_2	2	5*	7	2	3	3	2	5	5	7	4	6
Rank β_3	3	3	3	3*	1	1	3	3	3	2	1*	1*
Rank β_4	4	1	1	4	2	2	4	1*	1*	4	2	2
Rank β_5	5*	2	2	5	5	7	5	2	2	3	3	3
Rank β_6	6		4	6		4	6		4	6		4
Rank β_7	7		5*	7		6	7		6	5		5
$MSEY_{Ave}$	.006	.0015	.0016	.006	.0011	.0012	.6	.0501	.0545	.6	.0667	.0738
	Case 2											
	σ^2=.001						**σ^2=.1**					
	γ=0.1			**γ=0.9**			**γ=0.1**			**γ=0.9**		
	r_{CC}	**$r_{ccw}^{\ 1}$**	**$r_{ccw}^{\ 2}$**	**r_{CC}**	**$r_{ccw}^{\ 1}$**	**$r_{ccw}^{\ 2}$**	**r_{CC}**	**$r_{ccw}^{\ 1}$**	**$r_{ccw}^{\ 2}$**	**r_{CC}**	**$r_{ccw}^{\ 1}$**	**$r_{ccw}^{\ 2}$**
Rank β_1	1	4	6	1	4	6	1	4	6	1	4	5
Rank β_2	2	5*	7	2	3	3	2	5	7	2	3*	6
Rank β_3	3	3	3	3	1	1	3	3	3	3*	1	1
Rank β_4	4	1	1	4*	2	2	4*	1*	1*	4	2	2
Rank β_5	5*	2	2	5	5*	5*	5	2	2	5	5	3*
Rank β_6	6		4	6		4	6		4	6		4
Rank β_7	7		5*	7		7	7		5	7		7
$MSEY_{Ave}$	.006	.0013	.0013	.006	.0013	.0013	.6	.1260	.1352	.6	.1111	.1158

	Case 3											
	σ^2=.001						σ^2=.1					
	γ=0.1			γ=0.9			γ=0.1			γ=0.9		
	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2
Rank β_1	1	4	6	1	4	6	1	4	6	1	4	6
Rank β_2	2	5*	7	2	3	3	2	5*	7	2	3	3
Rank β_3	3	3	3	3	1	1	3	3	3	3	1	1
Rank β_4	4	1	1	4	2	2	4	1	1	4	2	2
Rank β_5	5*	2	2	5*	5*	5*	5*	2	2	5*	5*	5*
Rank β_6	6		4	6		4	6		4	6		4
Rank β_7	7		5*	7		7	7		5*	7		7
$MSEY_{Ave}$	.006	.0013	.0013	.006	.0013	.0013	.6	.1250	.1250	.6	.1384	.1384

	Case 3											
	σ^2=.1						σ^2=10					
	γ=0.1			γ=0.9			γ=0.1			γ=0.9		
	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2
Rank β_1	1	4	6	1	4	6	1	4	6	1	4*	6
Rank β_2	2	5*	7	2	3	3	2	5*	7	2	3	3*
Rank β_3	3	3	3	3	1	1	3	3	3	3*	1	1
Rank β_4	4	1	1	4	2	2	4	1	1	4	2	2
Rank β_5	5*	2	2	5*	5*	5*	5*	2	2	5	5	7
Rank β_6	6		4	6		4	6		4	6		4
Rank β_7	7		5*	7		7	7		5*	7		5
$MSEY_{Ave}$	.6	.1250	.1250	.6	.1384	.1384	60	15.0834	16.0759	60	10.9268	11.6517

	Case 3											
	σ^2=10						σ^2=100					
	γ=0.1			γ=0.9			γ=0.1			γ=0.9		
	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2
Rank β_1	1	4	6	1	4*	6	1	4	6	1	4	7
Rank β_2	2	5*	7	2	3	3*	2	5	7	2	3*	6
Rank β_3	3	3	3	3*	1	1	3*	3	3	3*	1	1
Rank β_4	4	1	1	4	2	2	4	1*	1*	5	2	2
Rank β_5	5*	2	2	5	5	7	5	2	2	4	5	3*
Rank β_6	6		4	6		4	6		4	6		4
Rank β_7	7		5*	7		5	7		5	7		5
$MSEY_{Ave}$	60	15.0834	16.0759	60	10.9268	11.6517	600	97.4803	109.9907	600	112.6950	121.4489

Table 4.2 Parameter ranking and selection results corresponding to W^{ext} and two different approaches when FIM is non-invertible. Case 1: $\beta_0 = 1.1*\beta$, $s_{\beta 0} = 0.1*\beta_0$; Case 2: $\beta_0 = 1.5*\beta$, $s_{\beta 0} = \beta_0$; Case 3: $\beta_0 = 10*\beta$, $s_{\beta 0} = 0.1*\beta_0$. r_{CCW}^1 corresponds to the first approach and r_{CCW}^2 corresponds to the second approach for parameter ranking and selection when FIM is not invertible.

	Case 1											
	σ^2=.001						σ^2=.1					
	γ=0.1			γ=0.9			γ=0.1			γ=0.9		
	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2
Rank β_1	1	5*	6	1	2	4	1*	2	2	1*	2	5
Rank β_2	2	4	7	2	1*	6	2	4	5	7	4	7
Rank β_3	3	1	1	3*	3	3	3	3	6	2	3*	3
Rank β_4	4	2	2	4	4	5	4	1*	1*	4	1	1
Rank β_5	5*	3	3	5	5	1*	5	5	7	3	5	2*
Rank β_6	6		5*	6		2	6		3	6		4
Rank β_7	7		4	7		7	7		4	5		6
$MSEY_{Ave}$	.006	0.0054	0.0054	.006	0.1392	0.1806	.6	0.1914	0.2068	.6	7.9061	7.9533

Case 2

	σ^2=.001						σ^2=.1					
	γ=0.1			γ=0.9			γ=0.1			γ=0.9		
	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2
Rank β_1	1	5*	6	1	4	4	1	4	6	1	2*	5
Rank β_2	2	4	7	2	3	7	2	5	7	2	1	6
Rank β_3	3	1	1	3	1	1	3	1	1*	3*	3	3
Rank β_4	4	2	2	4*	5*	5	4*	3	3	4	4	7
Rank β_5	5*	3	3	5	2	2	5	2*	2	5	5	1
Rank β_6	6		5*	6		6*	6		4	6		2*
Rank β_7	7		4	7		3	7		5	7		4
$MSEY_{Ave}$	.006	0.0047	0.0047	.006	0.3787	0.3858	.6	0.5581	0.6339	.6	8.1664	9.4472

Case 3

	σ^2=.001						σ^2=.1					
	γ=0.1			γ=0.9			γ=0.1			γ=0.9		
	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2
Rank β_1	1	5*	6	1	4	4	1	5*	6	1	4	4
Rank β_2	2	4	7	2	3	7	2	4	7	2	3	7
Rank β_3	3	1	1	3	1	1	3	1	1	3	1	1
Rank β_4	4	2	2	4	5*	5	4	2	2	4	5*	5
Rank β_5	5*	3	3	5*	2	2	5*	3	3	5*	2	2
Rank β_6	6		5*	6		6*	6		5*	6		6*
Rank β_7	7		4	7		3	7		4	7		3
$MSEY_{Ave}$	.006	0.0047	0.0047	.006	0.3035	0.3035	.6	0.4735	0.4735	.6	42.6204	42.6544

	Case 3											
	σ^2=1						σ^2=10					
	γ=0.1			γ=0.9			γ=0.1			γ=0.9		
	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2	r_{CC}	r_{ccw}^1	r_{ccw}^2
Rank β_1	1	5*	6	1	3	4	1	5*	6	1	2*	4
Rank β_2	2	4	7	2	5	5	2	4	7	2	1	2
Rank β_3	3	1	1	3	1*	1*	3	1	1	3*	3	3
Rank β_4	4	2	2	4	4	6	4	2	2	4	4	5
Rank β_5	5*	3	3	5*	2	7	5*	3	3	5	5	1
Rank β_6	6		5*	6		3	6		5*	6		2*
Rank β_7	7		4	7		2	7		4	7		7
$MSEY_{Ave}$	.6	4.7346	4.7346	.6	269.4007	318.3253	60	57.3526	59.4314	60	1.2531E3	1.6075E3

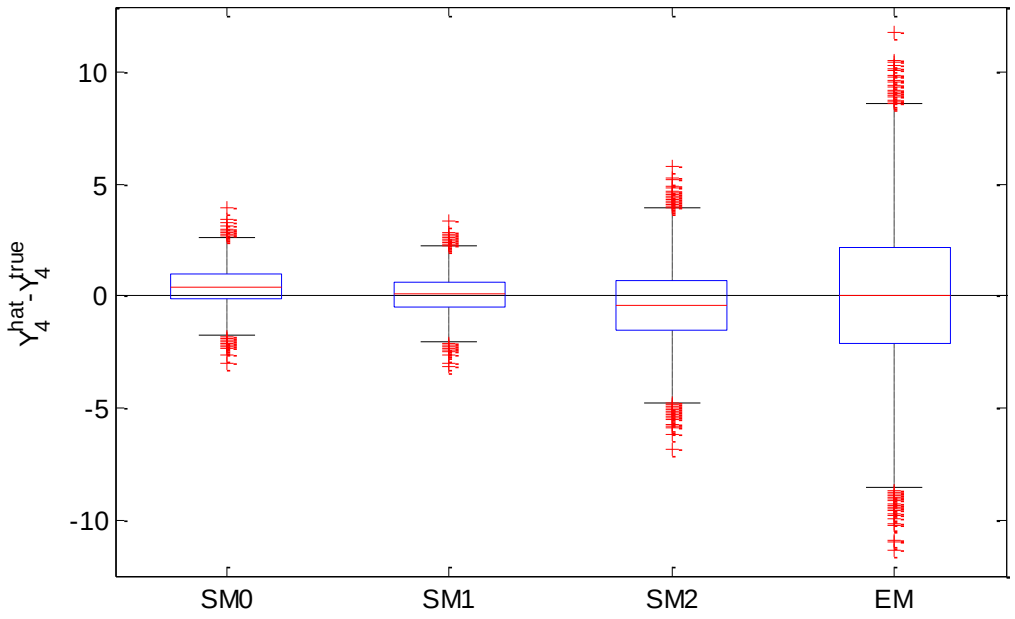

Figure 4.1 Comparison of model predictions of the forth response at settings in W^{ext} ($W \not\subset X$) for the seven-parameter linear model for Case 2 when σ^2=0.1, γ=0.9. SM0 contains three parameters β_1, β_2 and β_3 and was selected using the r_{CC} approach. SM1 contains two parameters β_1 and β_2 and was selected using the $r_{CCW}{}^1$ approach. SM2 contains two parameters β_5 and β_6 and was selected using the $r_{CCW}{}^2$ approach. The EM contains all seven parameters.

4.5. Nonlinear Regression Example

4.5.1. Polymer Film Casting Model

Lamberti et al[56,57] developed a one dimensional non-isothermal crystalline polymer film casting model (Figure 4.2) to obtain the five ODEs shown in Table 4.3, where L is film width distribution along the machine (draw) direction, x is the coordinate in the draw direction, v_x is the x-component of velocity, F is the draw force (the x-component of force acting on polymer), X_c is volumetric degree of crystallinity and T is the temperature of the polymer. The terminology is defined in the nomenclature section. Lamberti et al. used a commercial grade isotactic polypropylene(iPP) supplied by Montell, BA238G3 (polypropylene with 27% w/w of rubber C2–C4 added with 1.5% w/w of talc) in their experimental work[56,58].

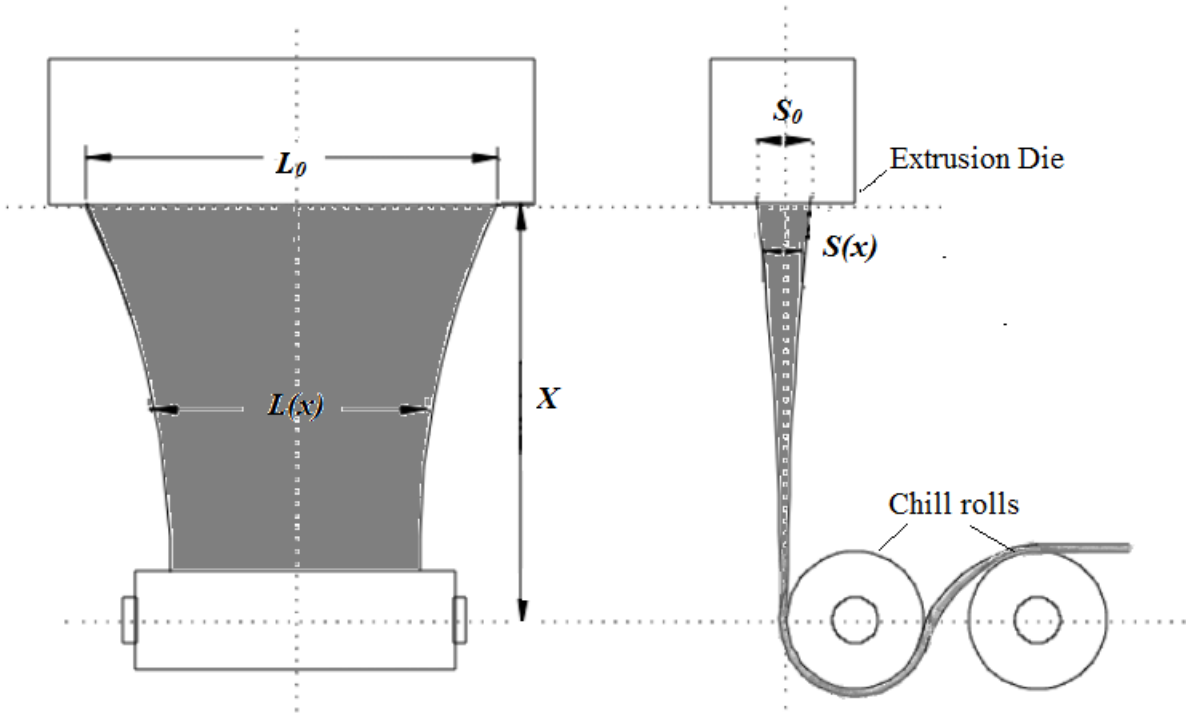

Figure 4.2 Film Casting Schematic[56]

Table 4.3 Polymer Film Casting Model[56,57]

$\frac{dL}{dx}=\frac{6\mu\dot{m}}{\rho FL}-\sqrt{\left(\frac{6\mu\dot{m}}{\rho FL}\right)^2+2}$: material balance	(T4.3.1)
$\frac{dv_x}{dx}=\frac{v_x}{4}\left(\frac{F\rho}{\mu\dot{m}}-\frac{2}{L}\frac{dL}{dx}\right)$: momentum balance	(T4.3.2)
$\frac{dF}{dx}=\dot{m}\left(\frac{dv_x}{dx}-\frac{g}{v_x}\right)$: force balance	(T4.3.3)
$\frac{dX_c}{dx}=\frac{\left[X_{eq}-X_c(x)\right]}{v_x}n_A\ln 2\left[\int_0^x K(T(x))\frac{d\xi}{v_x}\right]^{n_A-1}K(T(x))$: crystalline phase balance	(T4.3.4)

$$\frac{dT}{dx} = \frac{2h_{tot}(T_a - T)L}{C_p \dot{m}} + \frac{\Delta H}{C_p}\frac{dX_c}{dx} \text{ : energy balance} \quad \text{(T4.3.5)}$$

$$\mu(T, \dot{\gamma}, X_c) = \frac{\mu_{0,r}\exp(E_a / RT)}{\left\{1 + \left[\mu_{0,r}\exp(E_a / RT)C\dot{\gamma}\right]^{1-n_c}\right\}}\left[1 + f\exp\left(-\frac{h}{X_c^m}\right)\right] \quad \text{(T4.3.6)}$$

$$\dot{\gamma} = \frac{v_{xX} - v_{x0}}{x_{FL}} \quad \text{(T4.3.7)}^{61}$$

$$K(T) = K_0 \exp\left[-\frac{4\ln 2(T - T_{\max})^2}{D^2}\right] \quad \text{(T4.3.8)}$$

$$h_{tot} = h_{forc} + h_{nat} + h_{irr} \quad \text{(T4.3.9)}$$

$$h_{forc} = 0.66\frac{\kappa_a}{X_t}\left[\frac{v_X X_t \rho_a}{\mu_a}\right]^{0.50}\left[\frac{C_{p,a}\,\mu_a}{\kappa_a}\right]^{0.33} \quad \text{(T4.3.10)}$$

$$h_{nat} = \mathrm{B}\frac{\kappa_a}{X_t}\left[\mathrm{Gr}.\mathrm{Pr}\right]^j$$
$$\mathrm{Gr} = \frac{g\,\beta_a\,X_t^3\left(T - T_a\right)}{\upsilon_a^2} \quad \text{(T4.3.11)}$$
$$\mathrm{Pr} = \frac{C_{p,a}\,\mu_a}{\kappa_a}$$

$$h_{irr} = \varepsilon(S)\sigma\frac{T_a^4 - T^4}{T_a - T}$$
$$\varepsilon(S) = \left[1 - \exp\left(-aS\right)\right] \quad \text{(T4.3.12)}$$

Film thickness (S) can be calculated from the material balance equation:

$$\frac{d\left(v_x SL\right)}{dx} = 0 \tag{4.8}$$

Boundary conditions for the system of ODEs are given by:

$$x = 0, \quad \begin{cases} L = L_0 \\ v_x = v_{x0} \\ T = T_0 \\ X_c = 0 \end{cases} \qquad ; x = X_t,\ v_x = v_{xX} \tag{4.9}$$

The unusual structure for the boundary conditions (i.e., no boundary conditions for the draw force and two for the velocity) calls for a shooting procedure to solve the ODEs, where the initial value of the draw force is assumed and the ODEs are solved numerically. Next, the boundary condition for the velocity is compared with the predicted value and the initial draw force is adjusted until convergence is obtained. Model parameters, their nominal values and associated uncertainties are listed in Table 4.4. These uncertainty ranges were specified based on other prior parameter values found in the literature or using engineering judgment. Note that parameters *D*, B and *j* take different values in different operating regions and are treated as six parameters rather than three in the parameter selection procedure. The experimental run conditions reported by Lamberti et al. are listed in Table 4.5. In the "C" run conditions in Table 4.5, normalized width and normalized thickness measurements are available along the draw direction, but no temperature measurements were made. In the "E" run conditions in Table 4.5, normalized width and temperature measurements are available along the draw direction, but no thickness measurements are available. For runs E2 to E5 replicate runs with a few temperature measurements but no thickness and width measurements were performed.

Table 4.4 Nominal parameter values[56,62] for the nonlinear film casting model summarized in Table 4.3 and their associated uncertainties

#of param	Parameter	Value	s_β
1	$\mu_{0,r}\,(\text{Pa.s})$	0.07565	±5%
2	$E_a / R\,(\text{K})$	4966.22	±5%
3	$C\left(\text{Pa}^{-1}\right)$	1.064×10^{-4}	±10%
4	$n_C\,(\text{Dimensionless})$	0.4262	±10%
5	$f\,(\text{Dimensionless})$	1000	±30%
6	$h(\text{Dimensionless})$	0.2	±30%
7	$m(\text{Dimensionless})$	2.0	±2%
8	$K_0\left(\text{s}^{-1}\right)$	148.4	±1%
9	$T_{\max}\,(\text{K})$	331.0	±1° C
10	$n_A\,(\text{Dimensionless})$	0.35	±2%
11	$X_{eq}\,(\text{Dimensionless})$	0.58[56]	±5.5%
12	$a\left(\text{m}^{-1}\right)$	2662	±30%
13	$T_0\,(\text{K})$	473	±3° C
14	$T_a\,(\text{K})$	298	±6° C
15	$\mu_a\,(\text{Pa.s})$	1.8103×10^{-5}	±1%
16	$\upsilon_a\left(\text{m}^2.\text{s}^{-1}\right)$	1.5952×10^{-5}	±1%
17	$\beta_a\left(\text{K}^{-1}\right)$	1/298	±1%
18	$\kappa_a\left(\text{W.m}^{-1}.\text{K}^{-1}\right)$	0.0271	±1%
19	$C_{p,a}\left(\text{J.Kg}^{-1}.\text{K}^{-1}\right)$	1005.38	±1%
20	$C_p\left(\text{J.Kg}^{-1}.\text{K}^{-1}\right)$	1925.928[62]	±30%
21	$\Delta H\left(\text{J.Kg}^{-1}\right)$	87922.8[62]	±30%
22	$x_{FL}\,(\text{m})$	0.20	±30%
23	$D(\text{K})$	$\begin{cases} D_1 = 6.5 & \text{if} \quad T < T_{\max} \\ D_2 = 41.0 & \text{if} \quad T > T_{\max} \end{cases}$	±10%
24			±10%
25	$\text{B}(\text{Dimensionless})$	$\begin{cases} B_1 = 0.25 & \text{if} \quad \text{Gr.Pr} < 2\times10^9 \\ B_2 = 0.33 & \text{if} \quad \text{Gr.Pr} > 2\times10^9 \end{cases}$	±20%
26			±20%
27	$j(\text{Dimensionless})$	$\begin{cases} j_1 = 0.56 & \text{if} \quad \text{Gr.Pr} < 2\times10^9 \\ j_2 = 0.12 & \text{if} \quad \text{Gr.Pr} > 2\times10^9 \end{cases}$	±20%
28			±20%

Table 4.5 Experimental runs operating conditions[56,58]

Run #	$\dot{m}$ kg s^{-1}	v_{x0} m s^{-1}	v_{xX} m s^{-1}	DR dimensionless	L_0 M	L_X m	S_0 m	X_t m
C2	1.617×10^{-4}	2.12×10^{-3}	0.06733	31.7	0.2	0.087	0.0005	0.45
C3	2.577×10^{-4}	3.39×10^{-3}	0.10633	31.4	0.2	0.064	0.0005	0.45
C4	1.963×10^{-4}	2.58×10^{-3}	0.10433	40.4	0.2	0.069	0.0005	0.45
E2	2.01×10^{-4}	4.41×10^{-3}	102.0×10^{-3}	23.1	0.2		0.0003	0.45
E3	1.52×10^{-4}	3.32×10^{-3}	104.7×10^{-3}	31.5	0.2		0.0003	0.45
E4	1.37×10^{-4}	3.00×10^{-3}	104.7×10^{-3}	34.9	0.2		0.0003	0.45
E5	2.34×10^{-4}	5.13×10^{-3}	105.0×10^{-3}	20.5	0.2		0.0003	0.45

4.5.2. Results from Orthogonalization algorithm combined with r_{CC}

The orthogonalization algorithm (Algorithm 4.1) was used to rank all the model parameters from the most estimable to the least estimable. In this example, the scaled sensitivity matrix Z is a 178×28 matrix, where each row corresponds to the predicted responses for the run conditions at which data were reported and each column corresponds to a model parameter. The elements of the Z matrix are determined by using the difference approximation (i.e., by perturbing every parameter, one at a time, by 1%). When computing the scaled sensitivity matrix, the uncertainties in the parameters are set according to the corresponding values in Table 4.4. The thickness and width measurements uncertainties are fixed at 2% of their initial values and temperature uncertainties are calculated from the pooled variance of the duplicate runs in every experimental run conditions listed in Table 4.5. In this example, the FIM is not invertible, since rank(FIM)=25 instead of 28. The rank deficiency of the FIM means that only 25 parameters can be ranked using the orthogonalization algorithm. As such, the three left-out parameters (D_1, B_2 and j_2 (refer to Table 4.4)) were be fixed at their nominal values when using the r_{CC} criterion (Algorithm 4.2) with p=28 to decide how many parameters should be selected. Note that D_1 is the 23rd parameter listed in Table 4.4. It is the value of D that applies when $T<T_{max}$. Similarly B_2 and j_2 are values

of the 26th and 28th parameters listed in Table 4.4. These parameters are only used when Gr.Pr>2×10^9. It appears that these left-out parameters were not used in the model predictions because $T>T_{max}$ and Gr.Pr<2×10^9 during all of the runs where data are available. As shown in the second column of Table 4.6, by using orthogonalization and the r_{CC} criterion (Algorithm 4.2) to rank and select the parameters, a SM with only two parameters, j_1 and E_a / R, is selected for estimation using the 178 data values (i.e., temperature, thickness and width measurements). Figure 4.3 shows the weighted least-squares objective function with the top k parameters estimated and the remaining parameters held fixed at their nominal values (J_k, k=1, 2, …, p-1, obtained by using Equation (Al4.2.1) in Algorithm 4.2) versus the number of ranked parameters included in SM (k). According to Figure 4.3, the objective function does not get noticeably better after estimating the top two parameters. Figure 4.4 shows Wu's r_{CC} corrected critical ratio versus the number of parameters, calculated according to Algorithm 4.2. As shown in this figure, a SM with k=2 (i.e., j_1 and E_a / R) is selected as the best SM to obtain accurate predictions at the available data points. Estimating additional parameters from the ranked list in the second column of Table 4.6 would be expected to increase the total prediction variance for the 178 data values used for parameter estimation by more than the decrease in the total prediction bias caused by estimating more parameters. Figure 4.5 a) and b) shows the model predictions of the normalized thickness and the normalized thickness along the draw direction, respectively, for the C3 run conditions. In this figure, thinner lines correspond to the predictions of a model with initial parameter values, thicker lines correspond to the predictions of the two-parameter SM chosen according to the Algorithm 4.2, asterisks show the normalized width and squares show the normalized thickness corresponding to the experimental runs for the C3 condition listed in Table 4.5.

Table 4.6 Parameter ranking and selection results based on different methodologies

Param	Ortho. + r_{CC}	r_{CC} to rank & select	r_{CCW}[1]	
			$W=W^{sub}$	$W=W^{ext}$
$\mu_{0,r}$	5	8	4	14
E_a / R	2*	2	12	15
C	15	7	10	3
n_C	10	9	14	1
f	25	10	7	5
h	19		5	12
m	22		16	17
K_0	16		18	4
T_{max}	12		17	18
n	13		1	16
X_{eq}	21	6	2	22
a	14		13	7
T_0	3		8	8
T_a	8	8	11	11
μ_a	23		24	25
υ_a	17		20	19
β_a	20		25	23
k_a	18		23	21
$C_{p,a}$	24		22	24
C_p	6	3	19	20
ΔH	9	5*	21	13
x_{FL}	7	7	9	2
D_1				
D_2	4	4	15	9
B_1	11		3	10
B_2				
j_1	1	1	6*	6*
j_2				

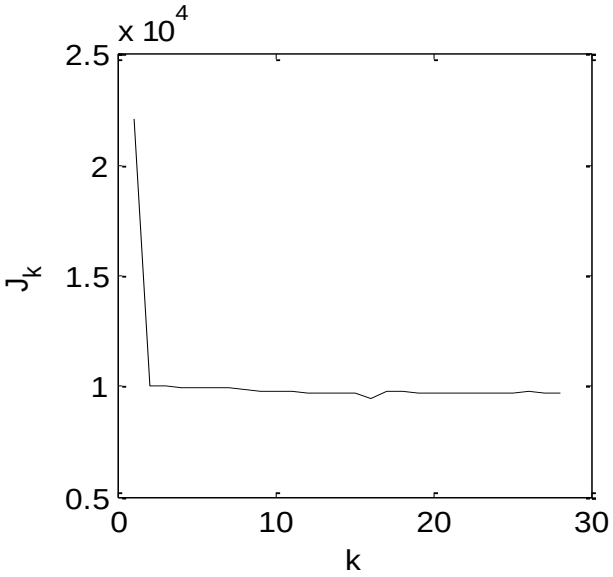

Figure 4.3 Sum of squared error predictions for a SM with *k* parameters versus the number of parameters, *k*. Parameters have been ranked according to the orthogonalization algorithm

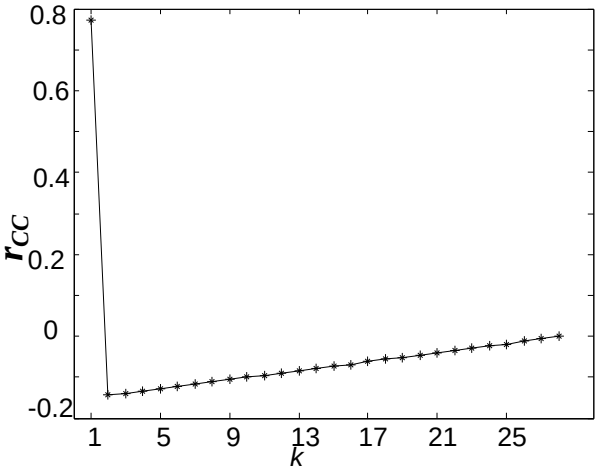

Figure 4.4 Wu's critical ration versus the number of parameters. Parameters have been ranked according to orthogonalization algorithm

The model predictions for the rest of the run conditions listed in Table 4.5 are shown in Figure A4.3 to Figure A4.7 in Appendix A4. According to Figure 4.5, the model predictions correspond to the two-parameter SM match the experimental data better that the model predictions using initial parameter guesses. Moreover, the two-parameter SM is capable of predicting the plateau region after the freezing line where film reaches the final width and thickness[56,58–60].

Figure 4.6 a) and b) shows the model predictions of temperature and normalized width along the draw direction, respectively, for the E2 run conditions specified in Table 4.5. The model predictions obtained by using the initial parameter values (a thinner line) and the model predictions obtained after estimating the two parameters in the chosen SM according to the orthogonalization and r_{CC} criterion (a thicker line) are shown in this figure. According to this figure, the model predictions correspond to the two-parameter SM match the experimental data (diamonds in Figure 4.6) better than the model predictions using initial parameter guesses.

Similar agreements between the measured and predicted values have been observed for the rest of the experimental run conditions listed in Table 4.5.

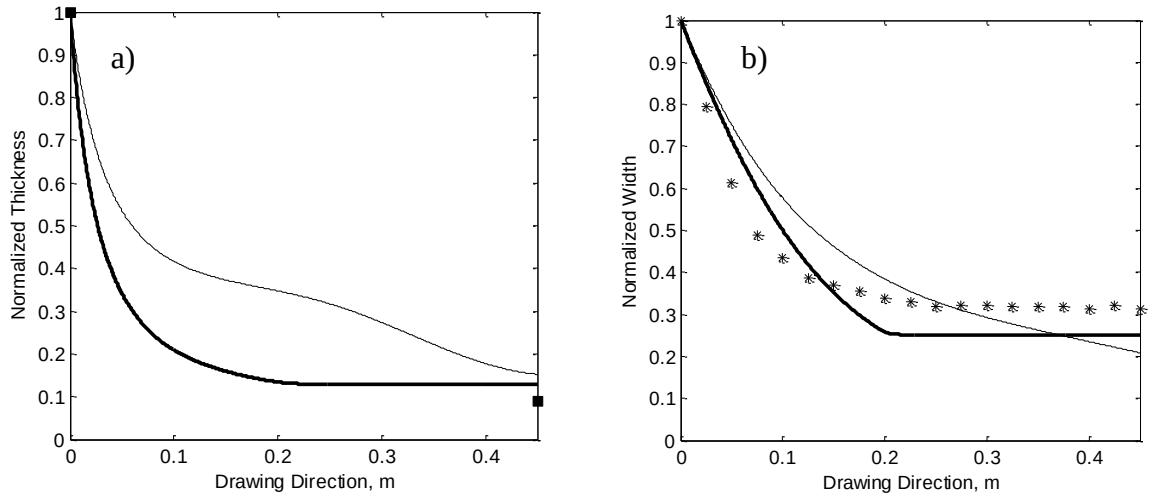

Figure 4.5 Model predictions obtained by using the two-parameter SM in Table 4.6 arising from the orthogonalization and r_{CC} method (thicker line) and predictions made using initial parameter values (thinner line) for the run condition C3 listed in Table 4.5 for a) normalized thickness model predictions. Squares correspond to a thickness measurement along the draw direction. b) normalized width model predictions. Asterisks correspond to the normalized width measurements along the draw direction.

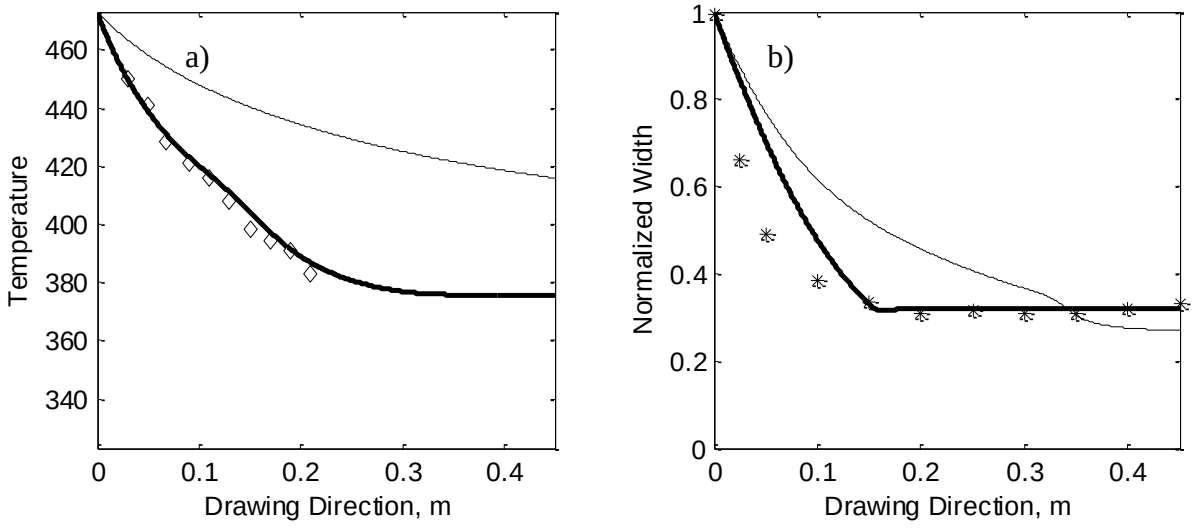

Figure 4.6 Model predictions obtained by using the two-parameter SM in Table 4.6 arising from the orthogonalization and r_{CC} method (thicker line) and predictions made using initial parameter values (thinner line) for the run condition E2 listed in Table 4.5 for a) temperature model predictions. Diamonds correspond to the temperature measurements along the draw direction. b) normalized width model predictions. Asterisks correspond to the normalized width measurements along the draw direction.

4.5.3. Results when r_{CC} is used for ranking and selection

For comparison, McLean's methodology (Algorithm 4.3) was used to rank and select the parameters for estimation in this polymer film casting model. As shown in the third column of Table 4.6, a five-parameter SM which includes $j_1, E_a / R, C_p, D_2$ and ΔH was chosen as the best SM for making predictions corresponding to the available data set. In this procedure, the three parameters (i.e., D_1, B_2 and j_2) that cause a rank-deficient FIM were first identified using the orthogonalization algorithm and were fixed at their nominal values. In this non-isothermal polymer film casting process, parameters that influence the polymer temperature and the rate of change of temperature in the casting process seem to be more important according to Algorithm 4.3. For example, j_1, which appears as an exponent for calculating the natural convection heat transfer coefficient, is selected as the most important parameter. This result can be explained because of the relatively large contribution of natural convection compared to forced convection when calculating the total heat transfer coefficient, due to the low film velocity. Natural convection is also more important than radiation, which is relatively high close to the die exit but decreases rapidly along the draw direction. Parameters E_a / R which relates to the effect of temperature on polymer viscosity and C_p which is involved in the rate of change of polymer temperature due to heat loss to the surroundings are ranked as the second and the third most important parameters, respectively. This outcome is because, in the experimental data provided by Lamberti et al., a long take-up distance (i.e., the distance between die exit and chill roll) and relatively low extrusion flow rate are used, so that there are large temperature changes during film casting. As a result, these parameters which contribute to the effects of temperature on the polymer film casting model are important. Because of the highly non-isothermal nature of the film stretching in the air, the polymer can crystalized before it contacts the chill rolls. Thus, it is

important that the chosen SM accounts also for the solidification process during the film stretching in the air. Therefore, parameters D_2 and ΔH which appear in calculations of the crystallization kinetics and heat generation due to crystallization, respectively, are selected among the set of the influential model parameters for estimation in order to obtain accurate model predictions using McLean's methodology.

4.5.4. r_{CCW} algorithm results

The r_{CCW} criterion (Algorithm 4.5) is applied on the polymer film casting model to rank and select the parameters for estimation. Because the FIM in non-invertible, the orthogonalization algorithm (Algorithm 4.1) was first used to determine the three least important parameters (i.e., D_1, B_2 and j_2) that cause invertibility problems. These parameters were then fixed at their nominal values and the corresponding columns were removed from the scaled sensitivity matrix Z. The resulting reduced FIM considers the remaining 25 parameters listed in Table 4.4. This approach was adapted because using a reduced FIM tended to give better results than the pseudoinverse approach when selecting parameters using r_{CCW} for the linear regression case study. Two different choices of the W matrices, $W=W^{sub}$(W contains a subset of the rows for the available data set) and $W=W^{ext}$(W is an extrapolation away from the available data set), are considered here. W^{sub} ($W^{sub} \subset Z$) is a subset of the rows of the available data set corresponding to the E4 experimental run conditions listed in Table 4.5. E4 run conditions are interesting because they correspond to the highest value of the draw ratio among the run conditions where the temperature measurements are available (i.e., among the E5, E4, E3 and E2 run conditions). W^{ext} ($W^{ext} \not\subset Z$) is an extrapolation away from the conditions in C4 corresponding to a 25% increase in the extrusion flow rate. These particular choices of the W matrices are arbitrary and are for the

sole purpose of demonstration of the r_{CCW} methodology. W^{sub} is a 35×28 matrix whose elements are the same as the scaled sensitivity coefficients in the Z matrix corresponding to the run E4.

The results of parameter ranking and selection for the two choices of the W matrices are shown in columns 4th and 5th in Table 4.6. As shown in column 4th, a six-parameter SM which includes parameters $n, X_{eq}, B_1, \mu_{0,r}, h$ and j_1 is chosen as the best SM for making predictions at $W=W^{sub}$. The selection of this particular subset of the aforementioned parameters makes physical sense. In the E4 run conditions, a relatively high draw ratio caused by a decrease in the extrusion flow rate at a constant take-up velocity (compare E4 and E3 conditions in Table 4.5) is used. Extrusion flow rate reduction leads to cooling of the polymer along a smaller distance or in other words to an enlargement of the freezing zone which leads to an earlier onset of the crystallization and solidification process of the film stretching in the air. Therefore, it is necessary to obtain accurate model predictions of the crystallinity and the polymer cooling rate. Therefore, parameters n and X_{eq} which influence the predicted rate change of crystallinity and eventually the rate of change of temperature (Table 4.3) are very important and are ranked at the top of the list. Parameters B_1 and j_1 which appear as a coefficient and an exponent, respectively, in the natural convection heat-transfer coefficient are important because of their contributions in the predicted rate of change of temperature. The high draw ratio caused by the relatively low extrusion flow rate leads to an increase in the final film width[56]. Therefore, parameter $\mu_{0,r}$ and h which influence the predicted polymer viscosity become important because of their contributions to the predicted rate of change of extruded polymer width. Figure 4.7 a) and b) shows the model predictions of temperature and normalized width along the draw direction, respectively, for $W=W^{sub}$.

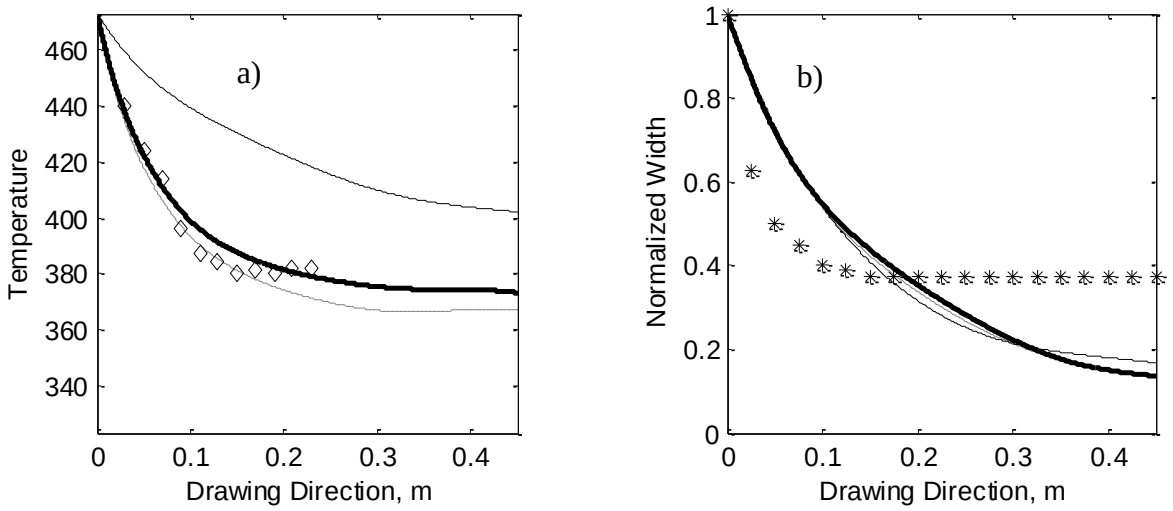

Figure 4.7 Model predictions obtained by using the six-parameter SM arising from the r_{CCW}^{1} approach (thicker line), predictions made using initial parameter values (thinner line) and predictions made using the six-parameter SM arising from the r_{CC} approach (dotted line) for $W=W^{sub}$ for a) temperature model predictions. Diamonds correspond to the temperature measurements along the draw direction. b) normalized width model predictions. Asterisks correspond to the normalized width measurements along the draw direction.

Figure 4.7 shows the model predictions that correspond to the six-parameter SM arising from the r_{CCW}^{1} approach (a thicker line), predictions made using initial parameter values (a thinner line) and predictions that correspond to the six-parameter SM arising from the r_{CC} approach (a dotted line) for $W=W^{sub}$. In this figure, experimental measurements for temperature and normalized width are shown with diamonds and asterisks, respectively. According to Figure 4.7 a), temperature predictions that correspond to the six-parameter SMs match the experimental data better than the model predictions using initial parameter guesses. Furthermore, the six-parameter SM is capable of predicting a plateau region starting at the freezing line, which has been observed in the experimental temperature data provided by Lamberti et al[58]. Using the parameter values in their paper (i.e., the initial values in Table 4.4) Lamberti et al. were not successful in predicting the temperature plateau region[58]. The model predictions arising from the r_{CC}-based

SM over-estimate the experimental measurements in the temperature plateau region more than the r_{CCW}^{1}-based SM does. The temperature plateau region, which is observed during the higher draw ratio experiments such as run E4[58], can be explained by significant increase in crystallization rate starting at the freezing zone which releases the same amount of heat as it lost by the film. Therefore, again it is important to obtain accurate estimates of parameters involved in crystallization kinetics such as n and X_{eq}. According to Figure 4.7 b), normalized width model predictions do not match the experimental data well enough. This result can be because of the assumed uncertainty in the width measurements. Due to the presense of replicate temperature runs, the calculated pooled variance for temperarure measurements seems to be more accurate than the assumed variance for width measurements.

According to the right-most column in Table 4.6, when W=W^{ext}, a different six-parameter SM which includes n_C, x_{FL}, C, K_0, f and j_1 is chosen as the best SM for making predictions at the settings in W^{ext}. W^{ext} is constructed based upon the C4 run conditions but with a 25% increase in the extrusion flow rate. W^{ext} is a 21×28 matrix whose elements were calculated by using simulations of the C4 run conditions with a 25% increase in the extrusion flow rate, where each of the 28 parameters was perturbed one at a time by 1%. An increase in the flow rate at constant draw ratio leads to a reduction in the final film width[56]. Thus, polymer viscosity and parameters which influence the polymer viscosity through the effect of rate of deformation (n_C and C) and crystallinity (f) have become important due to the effects of viscosity on film width. An increase in the flow rate at constant draw ratio and take-up velocity also leads to cooling of the polymer along a longer distance from the die exit and causes the freezing line to move away from a die exit[58]. Therefore, x_{FL} is the second most important parameter to estimate. Parameters

K_0 and j_1 were chosen because of their important influence on the rate of change of temperature due to the crystallization kinetics and the natural convection, respectively. Figure 4.8 a), b) and c) shows predicted normalized thickness, normalized width and temperature along the draw direction. In this figure, the model predictions obtained by using the six-parameter SM arising from the r_{CCW}^1 approach (thicker line), the initial parameter values (thinner line) and the six-parameter SM arising from the r_{CC} approach (dotted line) when W=W^{ext}, are shown. The information in the W matrix used by the r_{CCW} methodology does make a difference in the selected parameters and in the model predictions, as indicated by Figure 4.8. Unfortunately, when W=W^{ext}, there is no measured data available to compare with the model predictions. Figure 4.8 b) shows that the predicted final film width using both r_{CCW}^1 and r_{CC} approaches (but not the model with the initial parameter values) is less than the predicted final film width for the run C4 presented in Figure A4.6. This results make physical sense when W=W^{ext}, because the settings in the W^{ext} and the run C4 are the same, except a 25% increase in the extrusion flow rate (i.e., an increase in the flow rate at the constant draw ratio).

A polymer film casting model which contains 28 parameters with a non-invertible FIM is used as a case study to compare between several different techniques for parameter ranking and selection in order to obtain accurate model predictions at settings specified by model users. Different sets and numbers of parameters have been chosen according to different strategies and the results have been justified. Parameter j_1 has been always selected to include in the SM because of its significant influence on the rate of change of temperature of the extruded film, but the other selected parameters depended on the choice of the model selection criterion and on operating conditions of importance that were specified when r_{CCW} was used.

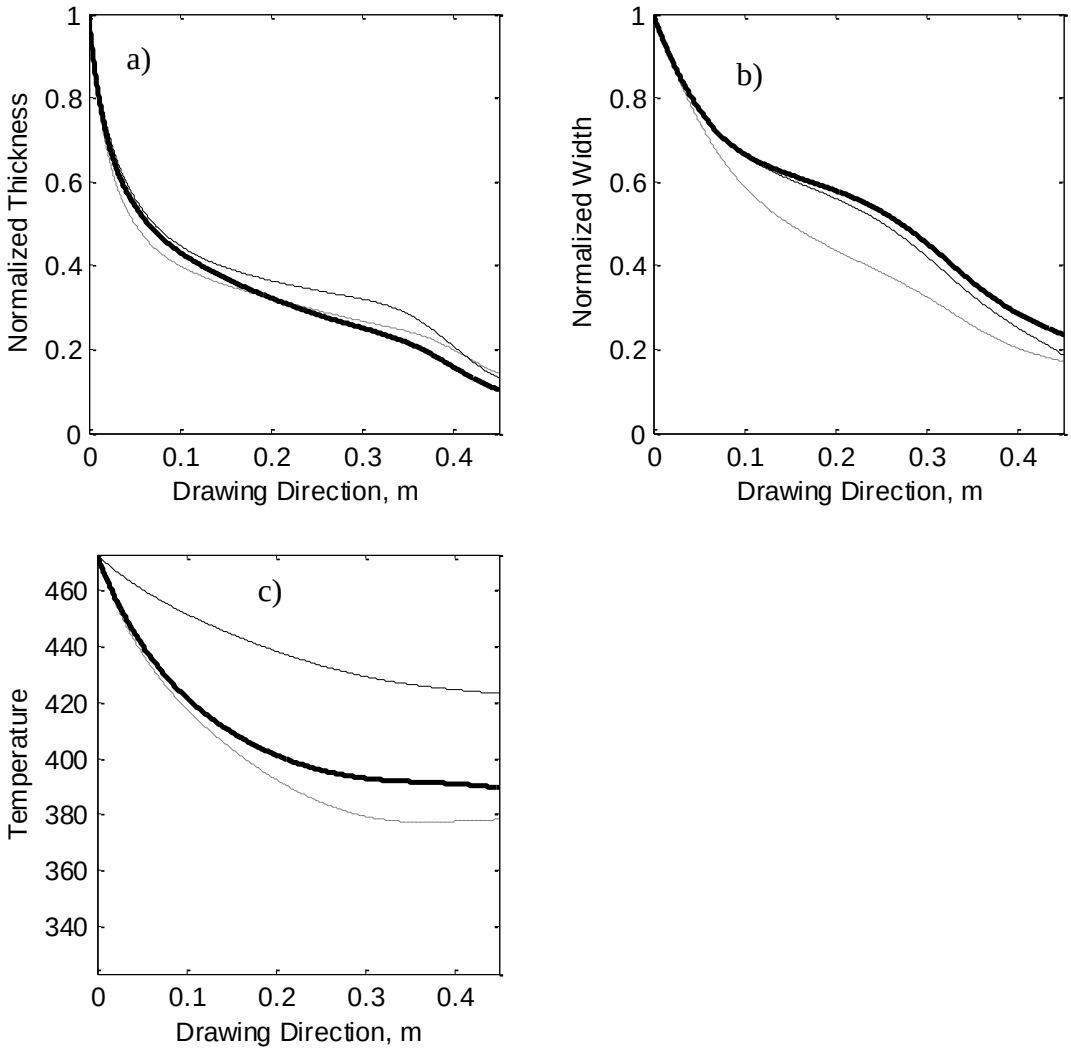

Figure 4.8 Model predictions obtained by using the six-parameter SM arising from the r_{CCW}^{1} approach (thicker line), predictions made using initial parameter values (thinner line) and predictions made using the six-parameter SM arising from the r_{CC} approach (dotted line) for $W=W^{ext}$ for a) normalized thickness model predictions b) normalized width model predictions c) temperature model predictions.

4.6. Conclusion

In this chapter, two different methodologies have been proposed and investigated for the extension of the previously derived MSE-based r_{CCW} criterion to the cases where FIM is non-invertible. A distinct characteristic of the r_{CCW} forward selection methodology is its consideration of the desired operating conditions where accurate predictions are desired. The first approach, indicated by r_{CCW}^{1}, uses an orthogonalization algorithm to determine relatively unimportant

parameters that make the FIM non-invertible. A reduced FIM is then constructed where there problematic parameters are fixed at their nominal values. The r_{CCW} criterion in then used to simultaneously rank and select the parameters. The second approach, r_{CCW}^2, uses a pseudoinverse in place of the inverse of the FIM in r_{CCW} calculations.

The demonstration of these two approaches on a seven-parameter linear example in which the true values of the model parameters are known showed that the first approach for parameter ranking and selection tends to perform better than the second, when the FIM is non-invertible. Comparisons between the model predictions and true values showed that the selected SMs have good predictive ability at the operating regions of interest, making them better than the EM or the SMs selected by other competing methods.

The resulting r_{CCW}^1 approach was used to simultaneously rank and select the parameters for estimation in a non-isothermal polymer film extrusion model with 28 parameters and a non-invertible FIM. Two different choices for the operating regions of interest were considered. The different parameter subsets that were selected for making predictions in different regions of interest made physical sense. The proposed methodology is valuable when it is impossible to accurately estimate all the model parameters using limited data sets, so that only a subset of the model parameters should be selected for estimation. This technique is also important when the modelers want to make accurate predictions at key input settings that are different from the settings in the available data sets.

4.7. References

(1) Shmueli, G.; Koppius, O. R. Predictive Analytics in Information Systems Research. *Mis Q.* **2011**, *35*, 553–572.

(2) Shmueli, G. To Explain or to Predict? *Stat. Sci.* **2010**, *25*, 289–310.

(3) McLean, K. A. P. Obtaining the Best Model Predictions and Parameter Estimates Using Limited Data, 2011.

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

(5) Hocking, R. R. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. *Biometrics* **1976**, *32*, pp. 1–49.

(6) Ehrenberg, A.; Bound, J. Predictability and prediction. *J. Roy. Stat. Soc. Ser. A* **1993**, *156*, 167–206.

(7) Geisser, S. The predictive sample reuse method with applications. *J. Amer. Stat. Assoc.* **1975**, *70*, 320–328.

(8) Kravaris, C.; Hahn, J.; Chu, Y. Advances and selected recent developments in state and parameter estimation. *Comput. Chem. Eng.* **2013**, *51*, 111–123.

(9) Lund, B. F.; Foss, B. a. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. *Automatica* **2008**, *44*, 278–281.

(10) McLean, K. A. P.; Wu, S.; McAuley, K. B. Mean-Squared-Error Methods for Selecting Optimal Parameter Subsets for Estimation. *Industrial & Engineering Chemistry Research*, 2012, *51*, 6105.

(11) Perregaard, J. Model simplification and reduction for simulation and optimization of chemical processes. *Comput. Chem. Eng.* **1993**, *17*, 465–483.

(12) Wang, F. Y.; Zhu, Z. H.; Massarotto, P.; Rudolph, V. A simplified dynamic model for accelerated methane residual recovery from coals. *Chem. Eng. Sci.* **2007**, *62*, 3268–3275.

(13) Weijers, S. R., & Vanrolleghem, P. A. A procedure for selecting best identifiable parameters in calibrating activated sludge model no 1 to full-scale plant data. *Water Sci. Technol.* **1997**, *36*, 69–79.

(14) Wu, S.; Harris, T. J.; McAuley, K. B. The use of simplified or misspecified models: Linear case. *Can. J. Chem. Eng.* **2007**, *85*, 386–398.

(15) Hawkins, D. M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12.

(16) Chu, Y.; Huang, Z.; Hahn, J. Improving prediction capabilities of complex dynamic models via parameter selection and estimation. *Chem. Eng. Sci.* **2009**, *64*, 4178–4185.

(17) Wu, S.; McAuley, K. B.; Harris, T. J. Selection of simplified models: II. Development of a model selection criterion based on mean squared error. *Can. J. Chem. Eng.* **2011**, *89*, 325–336.

(18) Moore, C.; Doherty, J. Role of the calibration process in reducing model predictive error. *Water Resour. Res.* **2005**, *41*, W05020.

(19) Rao, P. Some Notes on Misspecification in Multiple Regressions. *Am. Stat.* **1971**, *25*, pp. 37–39.

(20) McLean, K. A. P.; McAuley, K. B. Mathematical modelling of chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Can. J. Chem. Eng.* **2011**, *90*, 351–366.

(21) Miller, A. J. *Subset selection in regression*; Chapman and Hall: London England ; New York, 1990.

(22) Brun, R., Reichert, P., & Kunsch, H. R. Practical identifiability analysis of large environmental simulation models. *Water Resour. Res.* **2001**, *37*, 1015–1030.

(23) Sandink, C. a.; McAuley, K. B.; McLellan, P. J. Selection of Parameters for Updating in On-line Models. *Ind. Eng. Chem. Res.* **2001**, *40*, 3936–3950.

(24) Sun, C. L., & Hahn, J. Parameter reduction for stable dynamical systems based on Hankel singular values and sensitivity analysis. *Chem. Eng. Sci.* **2006**, *61*, 5393–5403.

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

(26) Yao, K. Z.; Shaw, B. M.; Kou, B.; McAuley, K. B.; Bacon, D. W. Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimatibility analysis and experimental design. *Polym. React. Eng.* **2003**, *11*, 563–588.

(27) Lund, B. F.; Foss, B. a. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. *Automatica* **2008**, *44*, 278–281.

(28) Walter, E., & Pronzato, L. *Identification of parametric models from experimental data*; Springer, 1997.

(29) Blanchet, F. G., Legendre, P., & Borcard, D. Forward selection of explanatory variables. *Ecology* **2008**, *89*, 2623–2632.

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

(31) Lopez, D. C. C.; Barz, T.; Penuela, M.; Villegas, A.; Ochoa, S.; Wozny, G. Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production. *Biotechnol. Prog.* **2013**, *29*, 1064–1082.

(32) Ngo, V. V; Latifi, M. A.; Simonnot, M.-O. Estimability Analysis and Optimisation of Soil Hydraulic Parameters from Field Lysimeter Data. *Transp. POROUS MEDIA* **2013**, *98*, 485–504.

(33) Meshram, M.; Naderi, S.; McConkey, B.; Ingalls, B.; Scharer, J.; Budman, H. Modeling the coupled extracellular and intracellular environments in mammalian cell culture. *Metab. Eng.* **2013**, *19*, 57–68.

(34) Jayasankar, B. R.; Ben-Zvi, A.; Huang, B. Identifiability and estimability study for a dynamic solid oxide fuel cell model RID F-2736-2010. *Comput. Chem. Eng.* **2009**, *33*, 484–492.

(35) Karimi, H.; Schaffer, M. A.; McAuley, K. B. Kinetic Moded for Non-Oxidative Thermal Degradation of Nylon 66. *Macromol. React. Eng.* **2012**, *6*, 93–109.

(36) Cui, W. J.; McAuley, K. B.; Whitney, R. A.; Spence, R. E.; Xie, T. Mathematical Model of Polyether Production From 1,3-Propanediol. *Macromol. React. Eng.* **2013**, *7*, 237–253.

(37) Littlejohns, J.; McAuley, K.; Daugulis, A. Model for a solid-liquid stirred tank two-phase partitioning bioscrubber for the treatment of BTEX. *J. Hazard. Mater.* **2010**, *175*, 872–882.

(38) Wu, S.; McAuley, K. B.; Harris, T. J. Selection of Simplified Models: I. Analysis of Model-Selection Criteria using Mean-Squared Error. *Can. J. Chem. Eng.* **2011**, *89*, 148–158.

(39) Eghtesadi, Z.; Wu, S.; McAuley, K. B. Development of a Model Selection Criterion for Accurate Model Predictions at Desired Operating Conditions. *Ind. Eng. Chem. Res.* **2013**, *52*, 12297–12308.

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

(41) Eghtesadi, Z.; McAuley, K. B. Mean Square Error Based Method for Parameter Ranking and Selection To Obtain Accurate Predictions at Speci fi ed Operating Conditions. *Ind. Eng. Chem. Res.* **2014**, *38*.

(42) Hero, A. O.; Fessler, J. A.; Usman, M. Exploring estimator bias- variance tradeoffs using the uniform CR bound. *IEEE Trans. Signal Process.* **1996**, *44*, 2026–2041.

(43) Sjöberg, J.; Zhang, Q.; Ljung, L.; Benveniste, A.; Delyon, B.; Rennec, P.-Y. G.-; Hjalmarsson, H.; Judisky, A. Nonlinear black-box modeling in system identification: A unified overview. *Automatica* **1995**, *31*, 1691–1724.

(44) Stoica, P.; Söderström, T. On non-singular information matrices and local identifiability. *Int. J. Control* **1982**, *36*, 323–329.

(45) Cui, W. J.; McAuley, K. B.; Whitney, R. A.; Spence, R. E.; Xie, T. Mathematical Model of Polyether Production From 1,3-Propanediol. *Macromol. React. Eng.* **2013**, *7*, 237–253.

(46) Kou, B.; McAuley, K. B.; Hsu, C. C.; Bacon, D. W. Mathematical model and parameter estimation for gas-phase ethylene/hexene copolymerization with supported metallocene catalyst. *Macromol. Mater. Eng.* **2005**, *290*, 537–555.

(47) Woloszyn, J. D.; McAuley, K. B. Application of Parameter Selection and Estimation Techniques in a Thermal Styrene Polymerisation Model. *Macromol. React. Eng.* **2011**, *5*, 453–466.

(48) Stoica, P.; Marzetta, T. L. Parameter Estimation Problems with Singular Information Matrices. *IEEE Trans. Signal Process.* **2001**, *49*, 87–90.

(49) Rao, C. R. *Linear Statistical Inference and Its Applications*; 2nd Edition; Wiley: New York, 1973.

(50) Li, Y.; Yeh, P. An Interpretation of the Moore-Penrose Generalized Inverse of a Singular Fisher Information Matrix. *IEEE Trans. Signal Process.* **2012**, *60*, 5532–5536.

(51) Rothenberg, T. J. Identification in Parametric Models. *Econometrica* **1971**, *39*, 577–591.

(52) Ram, S.; Ward, E. S.; Ober, R. J. A stochastic analysis of distance estimation approaches in single molecule microscopy: quantifying the resolution limits of photon-limited imaging systems. *Multidimens. Syst. Signal Process.* **2013**, *24*, 503–542.

(53) Ben-Haim, Z.; Eldar, Y. C. On the Constrained Cramér – Rao Bound With a Singular Fisher Information Matrix. *IEEE Signal Process. Lett.* **2009**, *16*, 453–456.

(54) Leis, J. R.; Kramer, M. A. The Simultaneous Solution and Sensitivity Analysis of Systems Described by Ordinary Differential Equations. *ACM Trans. Math. Softw.* **1988**, *14*, 34.

(55) Wu, S.; McLean, K. A. P.; Harris, T. J. H.; McAuley, K. B. Selection of optimal parameter set using estimability analysis and MSE-based model-selection criterion. *Int. J. Adv. Mechatron. Syst.* **2011**, *3*, 188–197.

(56) Lamberti, G.; Titomanlio, G.; Brucato, V. Measurement and modelling of the film casting process 1 . Width distribution along draw direction. *Chem. Eng. Sci.* **2001**, *56*, 5749–5761.

(57) Lamberti, G. Process validation of the normalized rheological function behavior during polymer crystallization. *Rheol. Acta* **2011**, *51*, 259–265.

(58) Lamberti, G.; Titomanlio, G.; Brucato, V. Measurement and modelling of the film casting process 2 . Temperature distribution along draw direction. *Chem. Eng. Sci.* **2002**, *57*, 1993–1996.

(59) Lamberti, G.; Brucato, V. Real-Time Orientation and Crystallinity Measurements during the Isotactic Polypropylene Film-Casting Process. *J. Polym. Sci. Part B Polym. Phys.* **2003**, *41*, 998–1008.

(60) Titomanlio, G.; Lamberti, G. Modeling flow induced crystallization in film casting of polypropylene. *Rheol. Acta* **2004**, *43*, 146–158.

(61) Lamberti, G. Flow-Induced Crystallization During Isotactic Polypropylene Film Casting. *Polym. Eng. Sci.* **2011**, 851.

(62) Barot, G.; Rao, I. J. Modeling the film casting process using a continuum model for crystallization in polymers. *Int. J. Non. Linear. Mech.* **2005**, *40*, 939–955.

Chapter 5 Conclusions and Recommendations

A MSE-based criterion, r_{CCW}, is developed to improve the accuracy of the model predictions in important operating regions by selecting the appropriate parameters to estimate from limited data. The important operating region is specified by users selected input conditions that may be different from the points at which data are already available. The r_{CCW} criterion is first developed in Chapter 2 and tested using linear regression models. In Chapter 2 to Chapter 3 the methodology is extended for use in nonlinear regression cases via a linearization approach. The unique characteristic of the r_{CCW} criterion is the focus of the methodology on considering the operating region of interest, something that has not been considered in the previously derived model selection criteria.

Initially, in Chapter 2, an orthogonalization algorithm was used to rank the model parameters from the most estimable to the least estimable after which the r_{CCW} criterion was implemented on this ranked list of parameters to determine how many parameters to estimate to obtain accurate model predictions at the desired operating conditions.

Later, in Chapter 3 , the r_{CCW} criterion was used as part of a new forward selection methodology to aid not only in the parameter selection step but also in the parameter ranking step. In this way, information about the desired operating region is used in both ranking and selection steps to obtain the best model predictions, in terms of total mean-squared error, at the specified conditions of interest.

Finally, in Chapter 4, the r_{CCW} forward-selection methodology is extended to help modelers in parameter ranking and selection in cases where the FIM is noninvertible. Two different r_{CCW}-based approaches are proposed. The first approach uses an orthogonalization algorithm to obtain a reduced FIM that has full rank. Then the r_{CCW}-based forward selection methodology is used to

rank and select the parameters. The second approach uses the pseudoinverse of the FIM in place of the inverse of the FIM in calculations of the r_{CCW}-based forward selection methodology. It is shown that the first approach tends to produce more reliable and accurate model predictions for a variety of linear regression models and simulated data sets. As a result, this approach is recommended for situation where the FIM is non-invertible.

In all of these three aforementioned developmental stages for the r_{CCW} criterion, simple five- and seven-parameter linear regression examples are used to illustrate the methodologies and to investigate the effects of the noise variance, the correlation in experimental designs and the different choices of operating conditions on model selection. A nonlinear dynamic 9-parameter batch reactor model and a 28-parameter polymer film casting model (with a noninvertible FIM) have also been used to illustrate the proposed methodology and to investigate the effects of the different choices for the operating regions of interest. The results of the r_{CCW}-based criterion are compared with the results of the Wu's and McLean's r_{CC}-based methodologies for model selection which do not account for the operating region of interest. It is shown that the model predictions arising from the parameters selected and estimated using the r_{CCW}-based criterion are more accurate, on average, than those arising from Wu's and McLean's r_{CC} -based methodologies.

Specific conclusions arising from this research are:

1. Monte Carlo simulations using the statistic r_{CCW} show that results obtained using simulated data agree well with theoretical results based on R_{CCW}.
2. When data are more informative (i.e., low noise variance or correlation factor), both r_{CCW} and r_{CC} criteria tend to choose more complex models, with a large number of estimated parameters.

3. When data are less informative (i.e., high noise variance or correlation factor), both techniques tend to choose simple models with fewer parameters.
4. According to the linear example simulation results, r_{CCW} and r_{CC} often agree on the choice of the best SM to use, especially when there are sufficient data to estimate all of the parameters. In cases where r_{CCW} and r_{CC} select different SMs, they often agree on the number of parameters that should be included in the model.
5. r_{CCW}-based forward selection methodology often produces different parameter rankings and selects different parameters for estimation compared with orthogonalization-based methods, because the proposed method incorporates information about the operating region of interest.
6. When the FIM is non-invertible, linear example simulation results showed that the first approach for parameter ranking and selection (r_{CCW}^1) tends to perform better than the second (r_{CCW}^2). Comparisons between the model predictions and true values showed that the selected SMs have good predictive ability at the operating regions of interest, making them better than the EM or the SMs selected by other competing methods.

According to the results presented in this thesis, it is recommended that modelers should include information about the operating region of interest (using the r_{CCW}-based framework) to rank and select parameters to obtain accurate model predictions, if they know where they want to use the model. It is also recommended that when the FIM is noninvertible, it is better to use orthogonalization to identify non estimable parameters and to construct a reduced FIM than it is to use a pseudoinverse when computing r_{CCW}. The proposed methodology in this thesis for parameter subset selection may be useful in future for model based design of experiment,

particularly for V-optimal design. The optimal experimental designs are aimed at selecting the new experimental conditions that will result in the best model predictions at the desired operating conditions[1]. The suggested new sequential experimental design techniques may be particularly beneficial when some of the model parameters are unestimable using the initial data set and the FIM is noninvertible. After conducting the sequential experiments some of the initially unestimable parameters may become estimable. In addition, the proposed r_{CCW}-based techniques proposed in this thesis can benefit from more validation work with larger scale models that include a larger number of parameters.

References

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

Appendix A1 Special Cases of R_{CW}

There are numbers of special cases for R_{CW} which will be derived below:

1. When only one parameter is excluded from the EM to generate the SM

When $p - p_1 = 1$, X_2 is a column vector of length n and Ω and β_2 are scalars. Substituting Equation (A2.5) into Equation(2.25) gives:

$$R_{CW} = \frac{\beta_2^T X_2^T \left(I_n - P_1\right) M^T M \left(I_n - P_1\right) X_2 \beta_2}{\sigma^2 Tr\left(\Omega X_2^T \left(I_n - P_1\right) M^T M \left(I_n - P_1\right) X_2\right)} \tag{A1.1}$$

which simplifies to:

$$\begin{aligned} R_{CW} &= \frac{\beta_2^2 X_2^T \left(I_n - P_1\right) M^T M \left(I_n - P_1\right) X_2}{\sigma^2 \Omega X_2^T \left(I_n - P_1\right) M^T M \left(I_n - P_1\right) X_2} \\ &= \frac{\beta_2^2}{\sigma^2} X_2^T \left(I_n - P_1\right) X_2 \\ &= \frac{\beta_2^T X_2^T \left(I_n - P_1\right) X_2 \beta_2}{(p - p_1)\sigma^2} \\ &= R_C \end{aligned} \tag{A1.2}$$

Therefore R_C and R_{CW} are the same for this special case.

2. When predictions are desired at the settings where $W = kX$

When $W = kX$ and k is a scalar**,** Equation(2.34) becomes:

$$\begin{aligned} M &= W\left(X^T X\right)^{-1} X^T \\ &= kX\left(X^T X\right)^{-1} X^T \\ &= kP \end{aligned} \tag{A1.3}$$

Substituting for M in Equation(2.30) gives:

$$
\begin{aligned}
R_{CW} &= \frac{\beta_2^T X_2^T \left(I_n - P_1\right) M^T M \left(I_n - P_1\right) X_2 \beta_2}{\sigma^2 Tr\left(\Omega X_2^T \left(I_n - P_1\right) M^T M \left(I_n - P_1\right) X_2\right)} \\
&= \frac{k^2 \beta_2^T X_2^T \left(I_n - P_1\right) P \left(I_n - P_1\right) X_2 \beta_2}{k^2 (p - p_1)\sigma^2} \\
&= \frac{\beta_2^T X_2^T \left(P - P_1\right) X_2 \beta_2}{(p - p_1)\sigma^2} \\
&= \frac{\beta_2^T X_2^T \left(I_n - P_1\right) X_2 \beta_2}{(p - p_1)\sigma^2} \\
&= R_C
\end{aligned}
\tag{A1.4}
$$

Appendix A2 Relationship between R_{CW} and R_C

Equation Section (Next) R_C and R_{CW} can be formulated as the quadratic norm of vectors:

$$R_C = \rho_C^T \rho_C \tag{A2.1}$$

$$R_{CW} = \rho_{CW}^T \rho_{CW} \tag{A2.2}$$

where

$$\rho_C = \frac{(I_n - P_1) X_2 \beta_2}{\sqrt{(p - p_1)\sigma^2}} \tag{A2.3}$$

$$\rho_{CW} = \frac{(W_1 A_1 - W_2)\beta_2}{\sqrt{\sigma^2 Tr\left((Z_1 A_1 - Z_2)\Omega(Z_1 A_1 - Z_2)^T\right)}} \tag{A2.4}$$

ρ_C is a $n \times 1$ column vector and ρ_{CW} is a $w \times 1$ column vector. Assume that there is a mapping matrix from X to W where:

$$W_{w \times p} = M_{w \times n} X_{n \times p} \tag{A2.5}$$

The mapping matrix, M, is given by:

$$M = W\left(X^T X\right)^{-1} X^T \tag{A2.6}$$

From Equation(A2.5), the mapping for partitioned matrices, W_1 and W_2, is:

$$\begin{pmatrix} W_1 & W_2 \end{pmatrix} = M \begin{pmatrix} X_1 & X_2 \end{pmatrix} \quad \Rightarrow \quad \begin{matrix} W_1 = MX_1 \\ W_2 = MX_2 \end{matrix} \tag{A2.7}$$

Substituting for W_1 and W_2 from Equation(A2.7) into Equation(A2.4) gives:

$$\rho_{CW} = \frac{M(I_n - P_1)X_2\beta_2}{\sqrt{\sigma^2 Tr\left(\Omega X_2^T (I_n - P_1) M^T M (I_n - P_1) X_2\right)}} \tag{A2.8}$$

where P_1 is defined in Equation(2.5). From Equations(A2.8) and(A2.3):

$$\begin{aligned}\rho_{CW} &= \frac{M(I_n - P_1)X_2\beta_2}{\sqrt{\sigma^2 Tr\left(\Omega X_2^T (I_n - P_1) M^T M (I_n - P_1) X_2\right)}} \\ &= \frac{M\sqrt{(p-p_1)}}{\sqrt{Tr\left((I_n - P_1) X_2 \Omega X_2^T (I_n - P_1) M^T M\right)}} \frac{(I_n - P_1)X_2\beta_2}{\sqrt{(p-p_1)\sigma^2}} \\ &= M\sqrt{\frac{(p-p_1)}{Tr\left((P - P_1) M^T M\right)}}\ \rho_C\end{aligned} \tag{A2.9}$$

since

$$(I_n - P_1) X_2 \Omega X_2^T (I_n - P_1) = P - P_1 \tag{A2.10}$$

Therefore

$$\rho_{CW} = \tilde{M}\ \rho_C \tag{A2.11}$$

where

$$\tilde{M} = M\sqrt{\frac{(p-p_1)}{Tr\left((P - P_1) M^T M\right)}} \tag{A2.12}$$

and M and Ω are defined by Equation(A2.6) and (2.22) respectively. Given X and W, both M and $\tilde{M}$ are non-random quantities and do not depend on any unknown variables. Based on decomposition of R_C and R_{CW} in Equations(A2.1) and(A2.2):

$$R_C = \rho_C^T \rho_C = Tr(\rho_C \rho_C^T) \tag{A2.13}$$

Combining Equation(A2.2) and(A2.11) gives:

$$\begin{aligned} R_{CW} &= \rho_{CW}^T \rho_{CW} = \rho_C^T \tilde{M}^T \tilde{M}\ \rho_C \\ &= Tr(\tilde{M}^T \tilde{M}\ \rho_C \rho_C^T) \\ &\le Tr(\tilde{M}^T \tilde{M}) \times Tr(\rho_C \rho_C^T) = Tr(\tilde{M}^T \tilde{M}) R_C \end{aligned} \tag{A2.14}$$

From Equation(A2.14) it can be concluded that if $R_C < 1/Tr(\tilde{M}^T \tilde{M})$, then $R_{CW} < 1$, which means that the SM is better than the EM for making predictions corresponding to W.

Appendix A3 Proof of Equation (2.31)

By substituting Equation (2.28) into Equation(2.27):

$$
\begin{aligned}
r_{CW} &= \frac{\hat{\beta}_{2E}^{T} X_2^{T}\left(I_n - P_1\right) M^{T} M\left(I_n - P_1\right) X_2 \hat{\beta}_{2E}}{s_E^2 Tr\left(\Omega X_2^{T}\left(I_n - P_1\right) M^{T} M\left(I_n - P_1\right) X_2\right)} \\
&= \frac{Y^{T}\left(\left(m-p\right)\dfrac{\left(I_n - P_1\right) X_2 \Omega X_2^{T}\left(I_n - P_1\right) M^{T} M\left(I_n - P_1\right) X_2 \Omega X_2^{T}\left(I_n - P_1\right)}{Tr\left(\Omega X_2^{T}\left(I_n - P_1\right) M^{T} M\left(I_n - P_1\right) X_2\right)}\right) Y \Big/ \left(m-p\right)}{Y^{T}\left(I_n - P\right) Y \big/ \left(n-m\right)} \\
&= \frac{Y^{T} A Y \big/ \left(m-p\right)}{Y^{T} B Y \big/ \left(n-m\right)}
\end{aligned}
\tag{A3.1}
$$

where

$$
\begin{aligned}
A &= \frac{\left(m-p\right)\left(I_n - P_1\right) X_2 \Omega X_2^{T}\left(I_n - P_1\right) M^{T} M\left(I_n - P_1\right) X_2 \Omega X_2^{T}\left(I_n - P_1\right)}{Tr\left(\Omega X_2^{T}\left(I_n - P_1\right) M^{T} M\left(I_n - P_1\right) X_2\right)} \\
&= \frac{\left(m-p\right)\left(I_n - P_1\right) X_2 \Omega X_2^{T}\left(I_n - P_1\right) M^{T} M\left(I_n - P_1\right) X_2 \Omega X_2^{T}\left(I_n - P_1\right)}{Tr\left(M^{T} M\left(I_n - P_1\right) X_2 \Omega X_2^{T}\left(I_n - P_1\right)\right)} \\
&= \frac{\left(m-p\right)\left(P - P_1\right) M^{T} M\left(P - P_1\right)}{Tr\left(M^{T} M\left(P - P_1\right)\right)}
\end{aligned}
\tag{A3.2}
$$

since P and P_1 are idempotent and Equation (A2.10) is valid, Equation(A3.2) can be rewrite as:

$$
A = \frac{\left(m-p\right)\left(P - P_1\right) M^{T} M\left(P - P_1\right)}{Tr\left(\left(P - P_1\right) M^{T} M\left(P - P_1\right)\right)} \tag{A3.3}
$$

Appendix A4

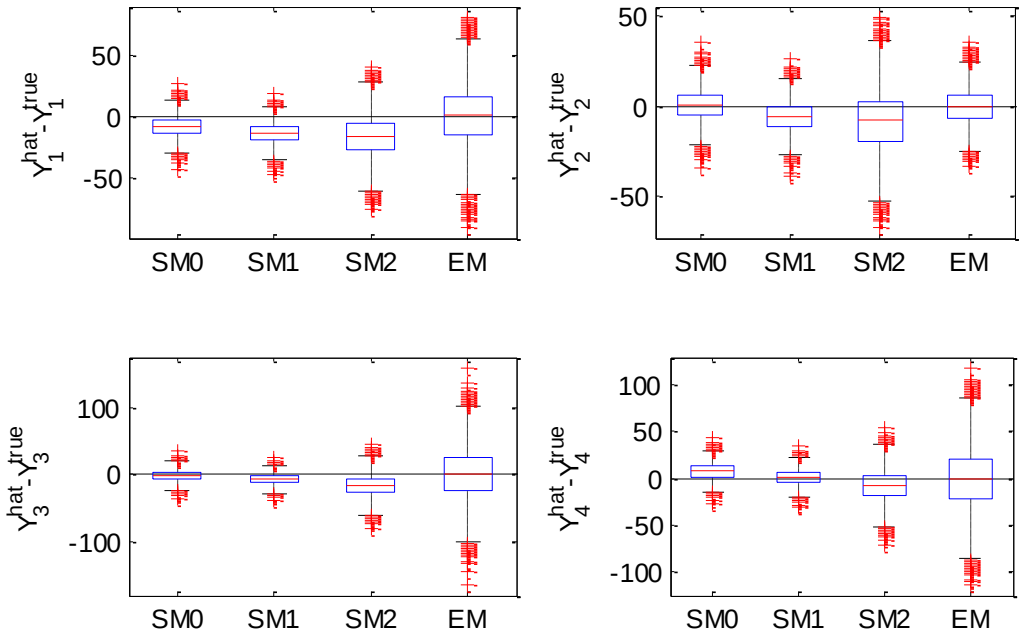

Figure A4.1 Comparison of model predictions at settings in W^{ext} ($W \subset X$) for the seven-parameter linear model for Case 3 when $\sigma^2=10$, $\gamma=0.9$. SM0 contains three parameters β_1, β_2 and β_3 and was selected using the McLean's r_{CC} approach. SM1 contains two parameters β_1 and β_2 and was selected using the $r_{CCW}^{\ 1}$ approach. SM2 contains two parameters β_5 and β_6 and was selected using the $r_{CCW}^{\ 2}$ approach. The EM contains all seven parameters.

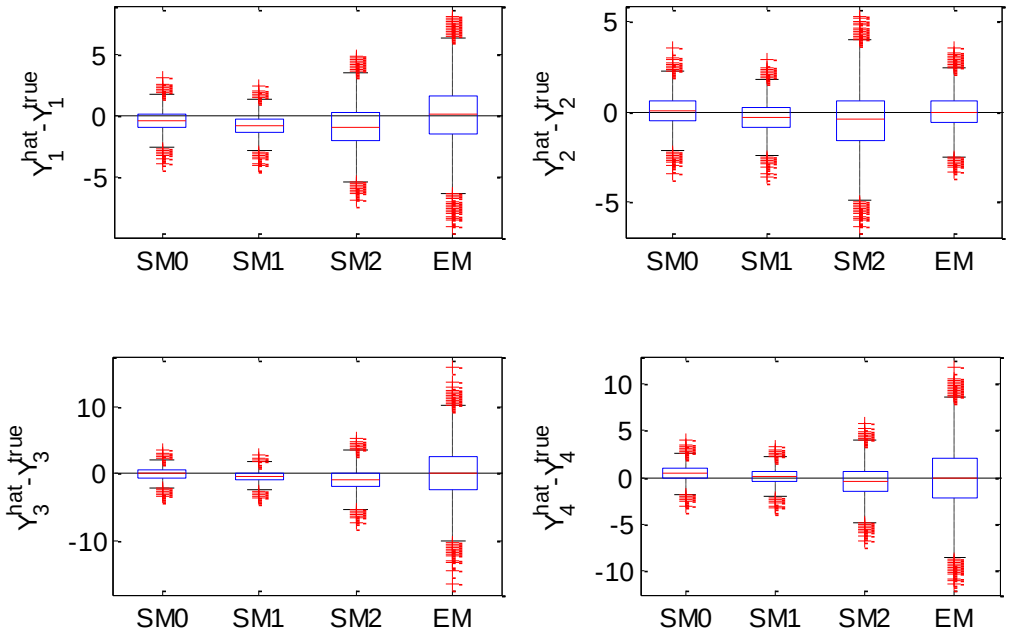

Figure A4.2 Comparison of model predictions at settings in W^{ext} ($W \subset X$) for the seven-parameter linear model for Case 2 when σ^2=0.1, γ=0.9. SM0 contains three parameters β_1, β_2 and β_3 and was selected using the McLean's r_{CC} approach. SM1 contains two parameters β_1 and β_2 and was selected using the $r_{CCW}{}^1$ approach. SM2 contains two parameters β_5 and β_6 and was selected using the $r_{CCW}{}^2$ approach. The EM contains all seven parameter

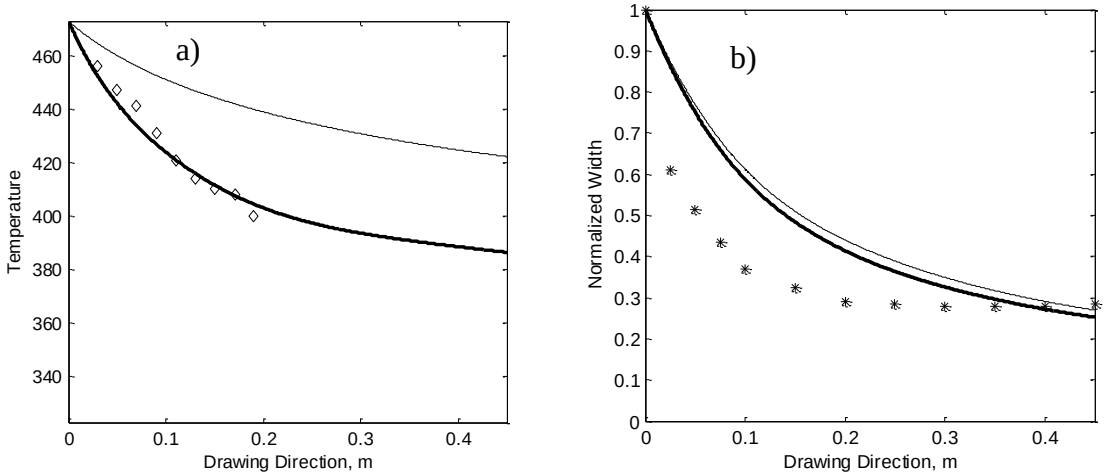

Figure A4.3 Model predictions obtained by using the two-parameter SM in Table 4.6 arising from the orthogonalization and r_{CC} method (thicker line) and predictions made using initial parameter values (thinner line) for the run condition E5 listed in Table 4.5 for a) temperature model predictions. Diamonds correspond to the temperature measurements along the draw direction. b) normalized width model predictions. Asterisks correspond to the normalized width measurements along the draw direction.

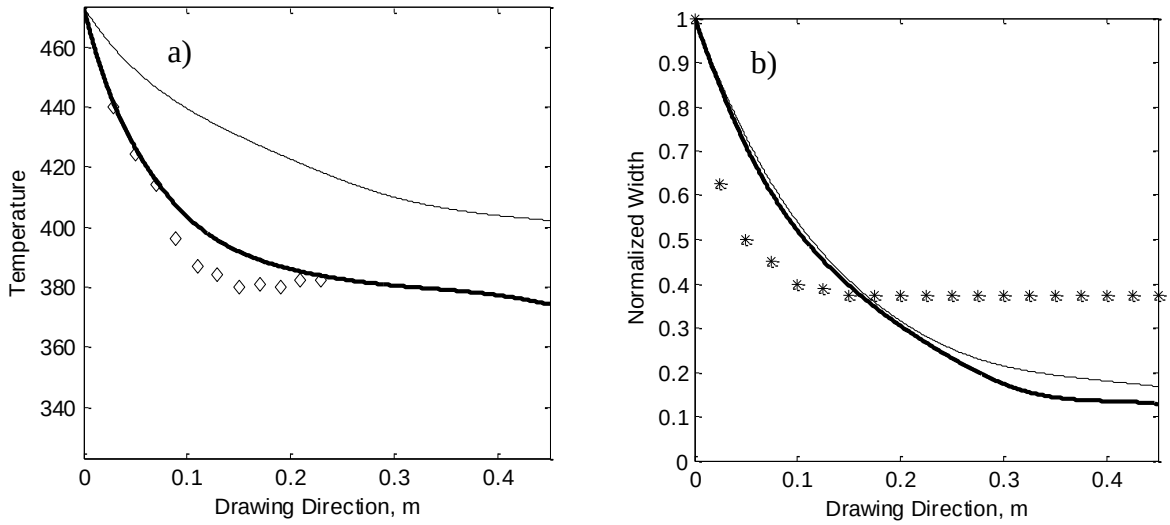

Figure A4.4 Model predictions obtained by using the two-parameter SM in Table 4.6 arising from the orthogonalization and r_{CC} method (thicker line) and predictions made using initial parameter values (thinner line) for the run condition E4 listed in Table 4.5 for a) temperature model predictions. Diamonds correspond to the temperature measurements along the draw direction. b) normalized width model predictions. Asterisks correspond to the normalized width measurements along the draw direction.

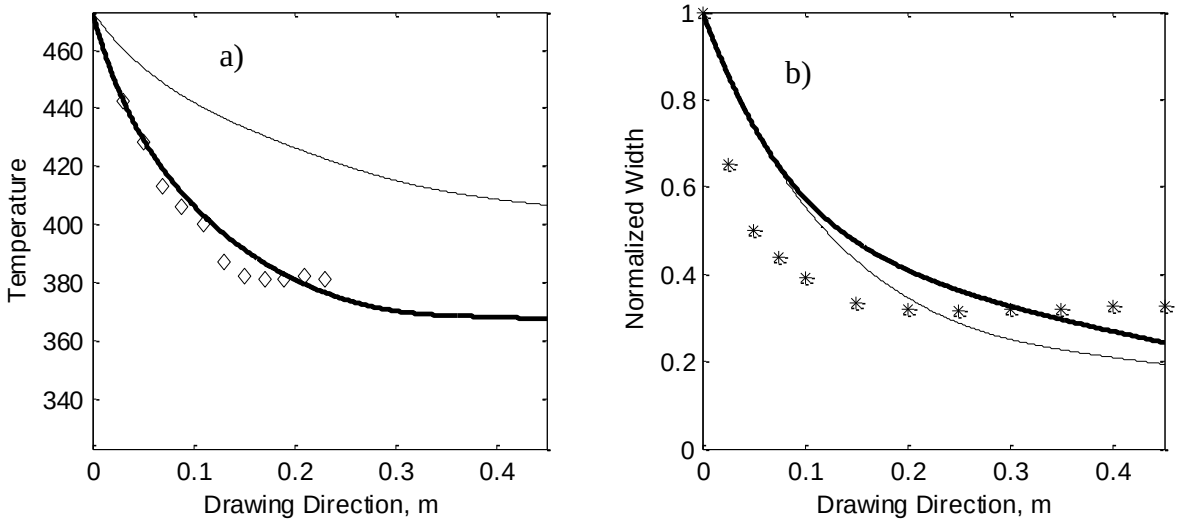

Figure A4.5 Model predictions obtained by using the two-parameter SM in Table 4.6 arising from the orthogonalization and r_{CC} method (thicker line) and predictions made using initial parameter values (thinner line) for the run condition E3 listed in Table 4.5 for a) temperature model predictions. Diamonds correspond to the temperature measurements along the draw direction. b) normalized width model predictions. Asterisks correspond to the normalized width measurements along the draw direction.

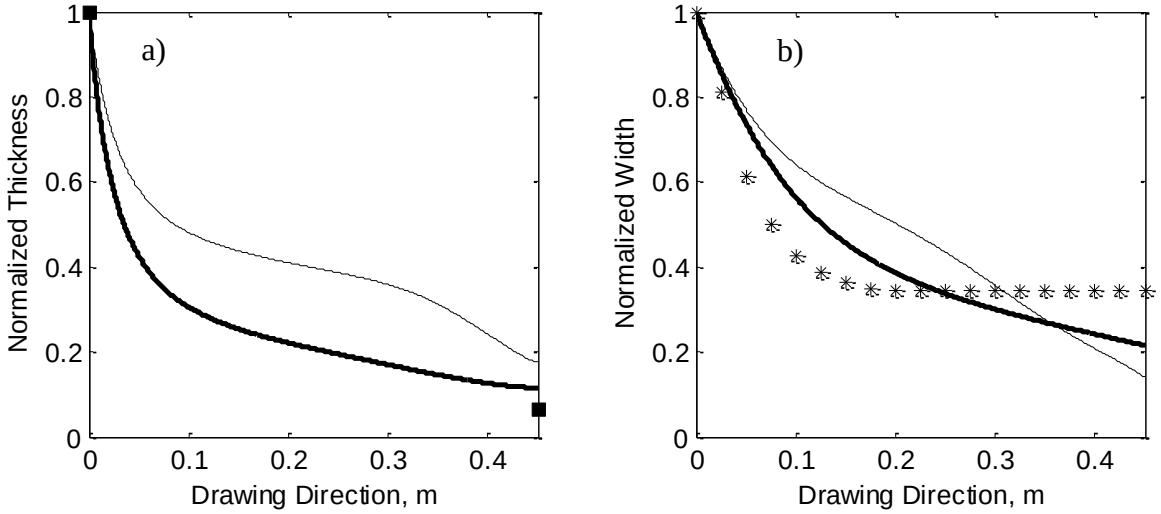

Figure A4.6 Model predictions obtained by using the two-parameter SM in Table 4.6 arising from the orthogonalization and r_{CC} method (thicker line) and predictions made using initial parameter values (thinner line) for the run condition C4 listed in Table 4.5 for a) normalized thickness model predictions. A Square corresponds to a thickness measurement along the draw direction. b) normalized width model predictions. Asterisks correspond to the normalized width measurements along the draw direction.

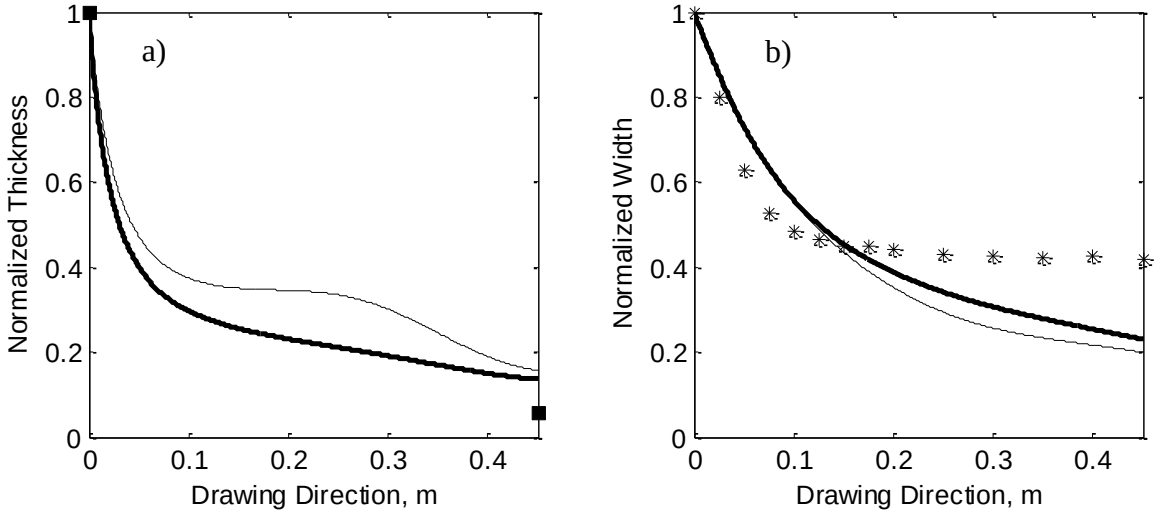

Figure A4.7 Model predictions obtained by using the two-parameter SM in Table 4.6 arising from the orthogonalization and rCC method (thicker line) and predictions made using initial parameter values (thinner line) for the run condition C2 listed in Table 4.5 for a) normalized thickness model predictions. A Square corresponds to a thickness measurement along the draw direction. b) normalized width model predictions. Asterisks correspond to the normalized width measurements along the draw direction.

Appendix A5 DuPont™ Tedlar® Film Casting Mathematical Modeling

Summary

An empirical mathematical model that predicts the thickness profile of extruded DuPont™ Tedlar® polyvinyl fluoride (PVF) film that can lead to improved process monitoring and control is developed in this appendix, particularly during product transitions and batch switches. Development of this model will require a detailed understanding of the extrusion and film casting process, as well as identification of key polymer and process parameters that influence film gauge. The model development work began by collecting pertinent steady-state data and dynamic process data during grade changes and batch switches. Important factors (e.g., polymer viscosity, extruder process variables, die geometry, sensor geometry, and rolling speed) were identified. ProMV software was used to build and validate the experimental models to predict gauge profile of the extruded Tedlar® film. The analysis performed in this study showed that there is a great value in upgrading the current feedback control scheme to a cascade control strategy in order to speed up the control action and reduce transition product waste.

A5.1 Introduction and Literature Review

Continuous sheet formation or film extrusion appears in many manufacturing processes including coating, fabric dyeing, paper manufacturing, metal rolling and polymer film extrusion[1]. Average film thickness (gauge) and spatial uniformity are important characteristics to ensure the quality of the polymer films for numerous industrial applications. Poor thickness control leads to unacceptable variability in product quality, production of off-specification products, increased energy and raw material costs, lower production rates and even machine shutdown. Tedlar® film

requires uniform thickness because it is used in surface protection of photovoltaic modules and aerospace interiors[2].

Later in this appendix, the Tedlar® casting process is described and literature models that have been used to predict thickness profiles in related processes are reviewed in Section A5.1. In Section A5.2, the data gathering step and multivariate statistical analysis to build a Projection to Latent Structure (PLS) model for the Tedlar® film casting process is described. Then, the PLS model is augmented further to include the effects of time delay and other important factors in the model. In Section A5.3, model validation is performed, followed by a conclusion and list of references.

A5.1.1 Properties and Application of Tedlar®

The commercial polyvinyl fluoride film Tedlar® was first trademarked by DuPont™ over 50 years ago. Since then, Tedlar® has become very popular because of its high outdoor durability as well as thermal stability, resistance to chemical and sunlight degradation, easy fabrication and cleaning, low-maintenance, high mechanical properties, flexibility, and high solar energy transmittance rate[2]. Tedlar® is a thermoplastic which has low thermal conductivity, high specific heat, and high melt viscosity. The desirable characteristics of Tedlar® have led to variety of applications in different industries such as photovoltaic (a backsheet in photovoltaic devices), transportation (a laminate in aircraft interior, stowage bins, galleys, and windows), construction (a laminate on architectural substrates), and other applications[2]. As such, ever increasing demand for PVF films led DuPont to expand the capacity of Tedlar® polymerization and extrusion. In August 2009, DuPont invested $120 million for expanding of the PVF capacity. In January 2010,

an additional investment of $175 million for expansion of an oriented film production line announced by DuPont[2].

A5.1.2 Film Types

DuPont™ PVF is available as transparent and pigment films. Films are made in non-oriented and oriented grades in several thicknesses, typically ranging from 0.012 mm to 0.102 mm, and various tensile modifications[2]. Oriented films, which have enhanced mechanical properties, are made by melt extrusion of PVF dispersed in a solvent, followed by biaxial orientation of the extruded film in the machine and transverse (cross) directions. Unoriented PVF film is cast onto a carrier belt (web); therefore, negligible stretching occurs. Unoriented films are more formable and compliant, have lower tensile strength and higher elongation at break than oriented films[2].

A5.1.3 Tedlar® Film Casting and Orienting

Tedlar® film casting cannot be achieved directly from the melt due to the occurrence of Tedlar® degradation prior to its melting point. PVF does not dissolve in most solvents at room temperature and pressure because of its high crystallinity and large amount of intermolecular hydrogen bonding. Therefore, PVF has to be dispersed in a polar solvent with high boiling point to form a film below its melting point. The solvent is recovered by evaporation after extrusion. Swelling of the PVF by the solvent reduces the melting point of PVF below its original value so that a single phase forms in the extruder. Pigments, plasticizer, stabilizers, and other additives can be added to the film by dispersing them with the polymer in the solvent.

At the DuPont™ Tedlar® facility located in Buffalo, several different grades of Tedlar® are produced using a common film casting area where the mix feed dispersion is coalesced into a

clear, thick, “paste-like” substance in the extruder. The extruder’s main function is to heat the mixture to a homogenous melt and to pump at a constant rate feed to casting hopper. The mixture in the extruder is heated up through a combination of mechanical shear stress from the mixing action of the screw and conductive heat transfer from the barrel. The extruder output rate is a function of screw speed, screw geometry (e.g., screw length and diameter, flight depth, and compression ratio), and melt viscosity. The flow of the extrudate is adjusted by a metering pump to control the unit weight of the film (i.e., the average thickness of the film that is produced over time). The melt is filtered through a screen pack to remove any solid impurities or contamination and then formed into a thick film by flowing through a die at the bottom of the casting hopper. The extruder’s die is located immediately above the bath. The melt web then flows through a set of heated rollers that operate at different speeds (slow rolls and fast rolls) to stretch the film in the machine direction (MD). A tenter frame drying oven is used to dry the film and stretch it in the transverse direction (TD). Various degrees of biaxial orientation can be done on extruded film to develop tensile strength in the film, and to improve mechanical, optical, and barrier film properties. Extruded film has a molecular orientation primarily in the extrusion direction (MD). Film stretching, which is performed at a temperature below the melting point of the polymer, yields partial orientation of polymer molecules in the stretch direction. The film is then wound onto rolls for further downstream processing.

A5.1.4 Film Casting Issues

The main purpose of the casting area is to provide a film with uniform thickness (i.e. gauge) in both the MD and the TD. A number of problems must be solved to produce a tough thin PVF film. When stretching is conducted at temperatures below 100 °C, “line drawing” (i.e., non-uniform thickness reduction over a short length in the MD) tends to occur between the slow and

fast rolls, making it difficult to obtain good film gauge control. Another phenomenon called "fibrillation", which happens during TD orientation, refers to physical breaks in the film along lines parallel to the MD. Fibrillation (i.e., tearing) is attributed to PVF crystallization. The stress force required for film orientation increases with the degree of crystallinity, and can go beyond the film strength, resulting in film tearing. Increasing temperature to above 100 °C to minimize line drawing problems leads to worse problems with fibrillation. Maintaining a minimum amount of solvent in the extruded film during biaxial orientation, results in fairly uniform thickness across PVF film in both MD and TD. To ensure thickness uniformity, solvent evaporation must be continued at a preset rate during TD stretch to reach to a "critical solvent content" by the end of TD orientation. The tentering oven is long enough to dry the film completely before the film exits the oven[2]. Another problem called "gauge bands" is related to the occurrence of thick spots on the extruded film. Sometimes the thick spots are thick enough to cause "raised gauge bands" leading to visible wrinkling in the final rolled sheet, and sometimes they are not. In the latter case they are called "hidden gauge bands" or "buckle lanes"[3].

A measurement system has been put in place at the Tedlar® facility in Buffalo to monitor the gauge profile across the sheet at two different locations, one in the Casting End (CE) after hopper die, and the other in the Wind-Up (WU) area, after the film has been dried significantly and before it is rolled onto a drum. In general, film-thickness sensing methods include beta-ray absorption, gamma-ray absorption, X-ray absorption, X-ray fluorescence, infrared, microwave, visible light, magnetic, electric capacitance, force distribution, and ultrasonics[1]. There are a number of actuators (i.e. die bolts) located across the hopper die which can be adjusted by stepper motors to maintain thickness uniformity. When a greater pressure is applied to the film

by adjusting a particular die bolt, the film becomes thinner at that location and at nearby locations in the TD.

A5.1.5 Mathematical Modeling and Control of Film Casting Processes

Modeling and control of polymer thickness profiles in the MD and TD is a complicated task because the thickness at each location is influenced by die bolts positions and by disturbances in the wind-up speed and polymer properties. The dimension of the control problem is large, due to the large number of actuators and sensing locations across the width of the sheet. The positions of individual die bolts are physically constrained and may be poorly known; leading to difficult constrained control problems. Also time delays between the extruder feed tanks and the sensors, and between the actuators and sensors change with production rate and thickness settings[4]. In all film gauge control problems, the main objective is to maintain the thickness of the film under tight and uniform control across the sheet when disturbances arise. Servo control, wherein the average thickness setpoint is changed may also be important when transitioning between different film grades. Various approaches to control of sheet and film thickness have been reported in the literature including linear quadratic control[5], anti-windup compensation[6] and linear model predictive control (MPC)[7–9].

One way to model the flow of molten polymer through a die to form a sheet or an extruded product is by solving fundamental material, momentum and energy balance equations. Because the properties of the polymer film change with time and location, the required balance equations are partial differential equations (PDEs)[10–14]. Formulating appropriate PDEs and boundary conditions requires simplifying assumptions (e.g., incompressible laminar flow, non-slip condition at solid surfaces)[14]. Solution of these complex theoretical models is computationally

intensive, making them inappropriate for on-line monitoring and control. As a result, easier-to-use empirical dynamic models (i.e., transfer-function and discrete state-space models) have been identified using process data and have been applied to industrial polymer extrusion and sheet-forming problems[4,15]. These empirical models can handle problems associated with noisy data, moving gauge sensors, process disturbances, actuator constraints, significant time delays between actuators and sensors, and complicated interactions between the effects of neighbouring actuators on sheet thickness at various locations along the moving sheet[1,16–18]. These interactions are caused by fluid flow in both MD and TD and by physical connections between actuators[1]. Assumptions about the nature of interactions can influence the model structure and parameter values, contributing to the success or failure of the monitoring and control system.

Multivariate techniques, which include Principal Component Analysis (PCA) and Projection to Latent Structure (PLS), have been used for modeling and control of film casting processes[19,20]. These multivariate projection techniques have the ability to decompose correlated variables into a series of latent variables and to reduce the dimensionality of the process data matrix. Multivariate statistical analysis techniques help to shed light on the importance of considering different process variables and polymer properties in the model developed in Section A5.2. For example, it may be possible to find out about the influence (or lack of influence) of individual die bolts positions on different thickness measurements across the film in the WU area (i.e., the lane identification problem).

In Tedlar® polymer extrusion process, disturbances that influence the thickness profile come from many different areas including the polymer, the additives and the process conditions in the mix tanks and casting area. Changes from one grade to the next result in large disturbances to the current film gauge control system due to the different rheological properties associated with

different grades. Disturbances associated with switching from one batch to the next (within the same grade) lead to similar but smaller problems for the current control system. DuPont engineers believe that the timing and severity of these transition disturbances can be predicted, based on knowledge about which grade the company will switch to next, and upstream measurements of the process conditions. Our multivariate statistical analysis results of Tedlar® polymer extrusion process (refer to Section A5.2) showed that there is room for improvement of the current control strategy in Tedlar® plant by switching to a cascade control system. Therefore, CE measurements, which are available sooner than WU measurements, can be implemented in a cascade control fashion to quickly adjust the die bolts positions and reduce transitional waste and time.

A5.2 Data Collection and Multivariate Statistical Analysis

As mentioned in the previous section, to achieve consistent gauge in both MD and TD of extruded Tedlar® film, a measurement system is used to monitor the gauge profile across the sheet at two different locations, one in the Casting End (CE) after the hopper die, and the other in the Wind-Up (WU) area after film has been dried significantly. Currently, the system employs a feedback control system based on information from the WU measurements to adjust the die bolt positions. All the information in the CE measurements is disregarded by the control system, since it was believed by DuPont's engineers to have high variability as a result of high moisture content in the extruded film. An investigation of potential usage of the CE measurements for developing process understanding is performed in the current study. The CE measurements, which are available one or two minutes before the film reaches the WU area, might also be useful for cascade control. Multivariate statistical analyses using PLS have been performed to find out about the correlation between the CE measurements and the WU measurements.

A5.2.1 Data Collection

Historical data which included six film transitions and twenty batch switches were collected from the Tedlar® extrusion plant between January and April 2013. Typically, data were collected over the course of about 24 hours for each batch switch or film transition to ensure that the steady state condition has been reached before and after the changes occured. CE and WU measurements, die bolts positions (DBPs) and some important upstream variables (e.g., upstream temperature, pressure, flow-rate, etc.) were collected.

A5.2.2 Multivariate Statistical Analysis

ProMV software designed by ProSensus MultiVariate was used to perform multivariate statistical analysis and develop a PLS model in this appendix. The main purpose of building this PLS model is to understand how well the WU measurements can be predicted from the CE measurements (Figure A5.1). We are also interested to know about the relation between DBPs and CE measurements and the relation between DBPs and WU measurements (Figure A5.1).

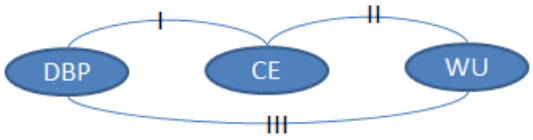

Figure A5.1 Relation between I: DBPs and CE measurements, II: CE measurements and WU measurements, and III: DBPs and WU measurements

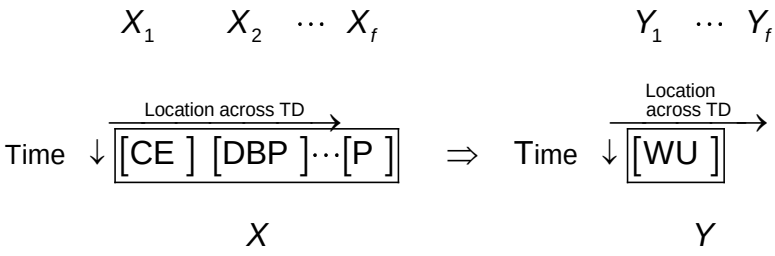

Figure A5.2 The *X* and *Y* matrices in a typical PLS model for Tedlar® plant

The *X* and *Y* matrices in a typical PLS model for Tedlar® plant are shown in Figure A5.1. In Figure A5.1 different rows inside each block (i.e., CE, DBP and P (e.g., upstream pressure of some kind)) correspond to the different data points which are measured at different time. Different columns inside each block correspond to the different data points which are measured at different locations across the film width or TD.

PCA analysis of the CE and WU measurements (i.e., *X*=[CE] or [WU]) revealed that central measurements are more similar than the edge measurements. These results can be explained because of the occurrence of the neck-in phenomena (i.e., reduction in the film thickness and width) and formation of beads (i.e., regions of greater thickness) at the film edges[21]. The streamlines in the central region of the film are parallel, indicating no or little deformation of fluid elements in the width direction (i.e., planae elongational flow); however, the streamlines

are converging near the film edges, indicating deformation of fluid elements in both width and thickness directions (i.e., uniaxial elongation flow)[21,22]. The PCA analysis results also showed that the variability in CE measurements is slightly higher than the variability in WU measurements (i.e., data points are more scattered in WU loading plot). This higher variability can be explained by higher moisture content in CE measurements.

Variable importance plots resulted from PLS analysis of DBPs and CE/WU measurements (i.e., *X*=[DBP] and *Y*=[CE] or [WU]) show that the similar DBPs are important for making CE /WU thickness predictions (i.e., DPBs #23 and #6 are the most important one in both cases). These results could be because of high correlation between CE and WU measurements due to either underlying physical phenomena or the control algorithm and operator intervention. Further PLS analysis of CE and WU measurements (i.e., *X*=[CE] and *Y*=[WU]) revealed that the central measurements are highly correlated with each other. The corresponding loading plot is shown in Figure A5.3. Model validation, which will be explained in Section A5.3, is performed by splitting all data into two groups; the training and prediction set data.

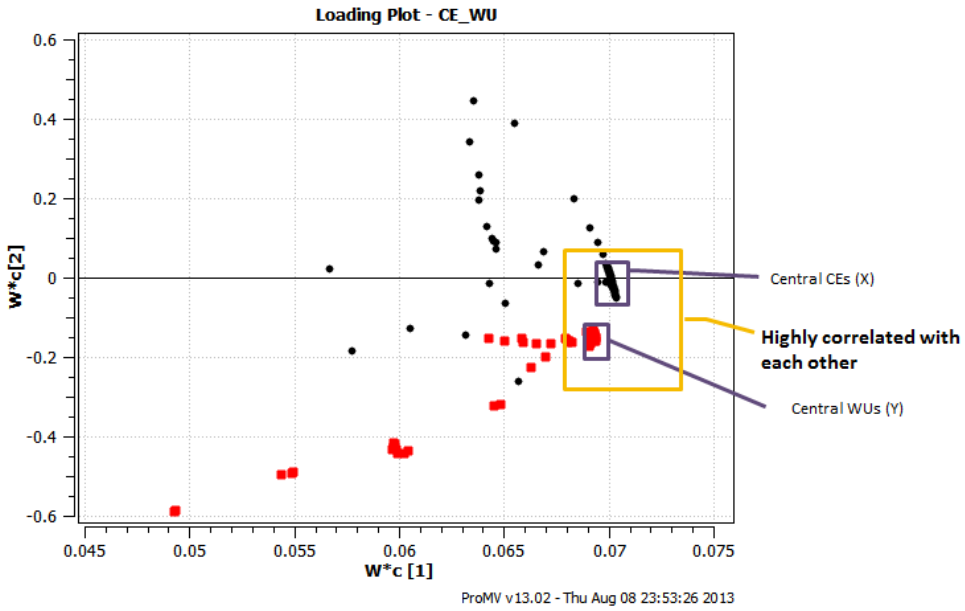

Figure A5.3 Variable loading plot for the PLS model based on the CE and WU measurements

A5.2.3 PLS Model Improvement

A5.2.3.1 Introducing Lagged Variables

Improvement of the PLS model was carried out to account for process dynamics including dead time through addition of the lagged variables into the input settings when constructing the X matrix for the PLS model[23,24]. *X* and *Y* matrices for a typical PLS model with lagged variables are shown in Figure A5.4.

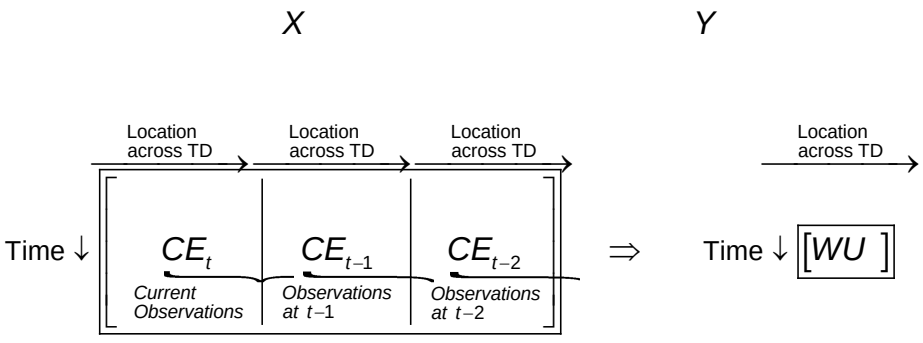

Figure A5.4 *X* and *Y* matrices for a typical PLS model with lagged variables

In this PLS model every block in *X* or *Y* matrices is expanded to include the observations at one, two or more previous time intervals. In this study, we only expanded the *X* matrix to include the lagged variables and not the *Y*, for simplicity. PLS analysis of the CE, lagged CE and WU measurements (i.e., *X*=[[CE][lagged-CE]] and *Y*=[WU]) revealed that considering only lagged variables corresponding to the previous time interval seems to be sufficient for improving the WU predictions in the prediction set data (Figure A5.6, to be explained in Section A5.3). This finding is consistent with the process dead time of about one minute since the CE measurements are available about every minute.

A5.2.3.2 Introducing Addition Upstream Variables

A PLS model for predicting WU measurements was further improved by introducing additional operating variables (e.g., DBPs) and process data (e.g., upstream temperature, pressure, flow-rate, etc.) (i.e., *X*=[[CE][lagged_CE][DBP][Upstream_Var]] and *Y*=[WU]). The upstream process variables included but were not limited to film unit weight setpoint, metering pump speed ratio, metering pump discharge pressure, static mixer outlet pressure, extruder speed, cast film width, mix water content and mix solid content. A variable importance plot of this PLS model showed that film unit weight and metering pump speed ratio are the most important variables in prediction of the WU measurements, presumably due to the fact that in Tedlar plant the metering pump speed determines the casting throughput rate and controls the film unit weight. In this PLS model the second most important variables are CE measurements after which the DBPs and some of other process variables are ranked. Therefore, information in the CE measurements seems to be more important than the information in the DBPs and some upstream variables. The process variable important plot for predicting WU measurements are shown in

Figure A5.5. When constructing a PLS model to predict WU measurements from CE measurements, DBPs and process variables, the quality of the WU fit for the training set data (training set and prediction set data will be elaborated in Section A5.3) has not been improved significantly by introducing process variables into the PLS model. However, the erratic behavior of WU predictions in the prediction set data is reduced significantly by introducing process variables (compare Figure A5.6 and Figure A5.8). In this PLS model, including lagged-CE measurements significantly improves WU predictions and reduced the root mean squared error prediction (RMSEP) by 38% (Figure A5.8).

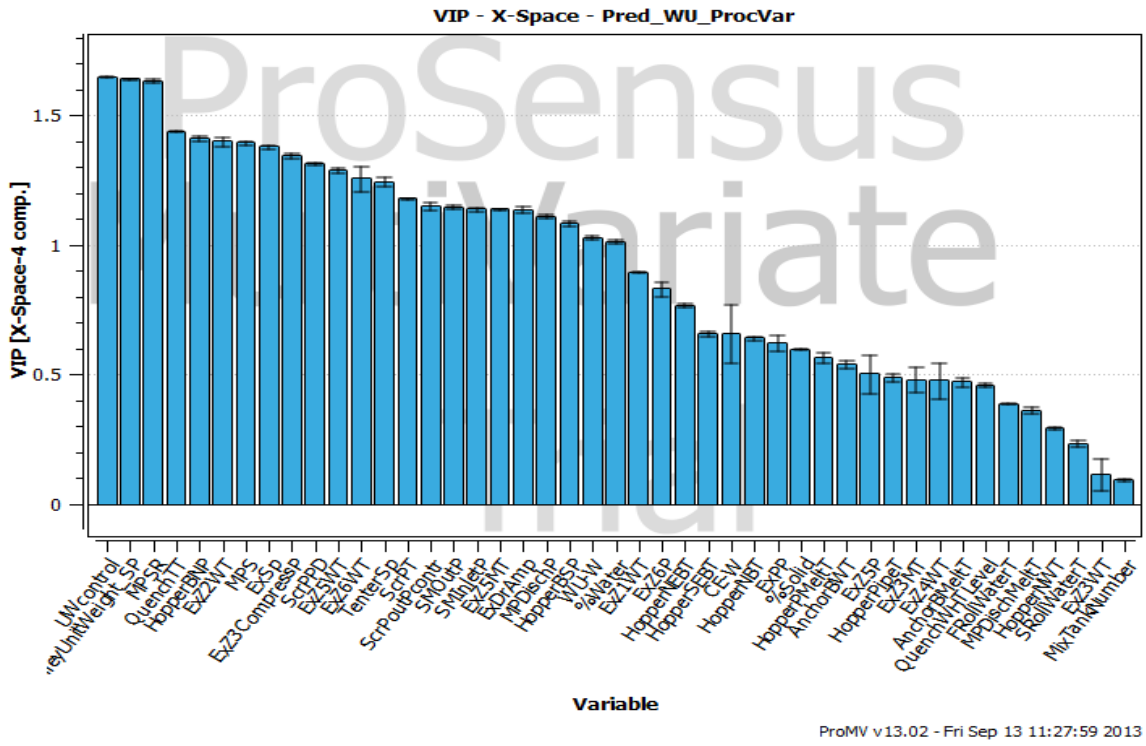

Figure A5.5 Process variable importance plot for predicting WU measurements from some upstream process variables

A5.3 PLS Model Validation

The predictive ability of the PLS model is assessed by dividing the data into two categories; the training set data and the prediction set data. In this study, the prediction set data contains

measurements for a film change and two batch switches, and the rest of the measurements are included in the training set data. A PLS model is built based on the information in the training set data (excluding the prediction set data). Next, the PLS model predictions are compared against observations for the prediction set data.

Figure A5.6 shows the comparison between the WU model predictions and the observations for the prediction set data at three different locations across the film width when the PLS model is built based on the information in the CE and WU measurements in the training set data. Figure A5.6 shows that the predictions capture some trends but they are erratic. A similar plot for the central WU position is shown in Figure A5.7 when the PLS model also includes the information in one or two lagged CE measurements. In Figure A5.7, the WU predictions corresponding to the PLS which contains one lagged CE measurements have the lowest RMSEP. By introducing the DBPs and some upstream process variables into the one-lagged PLS model, the erratic behavior and the model predictive ability improved significantly as shown in Figure A5.8.

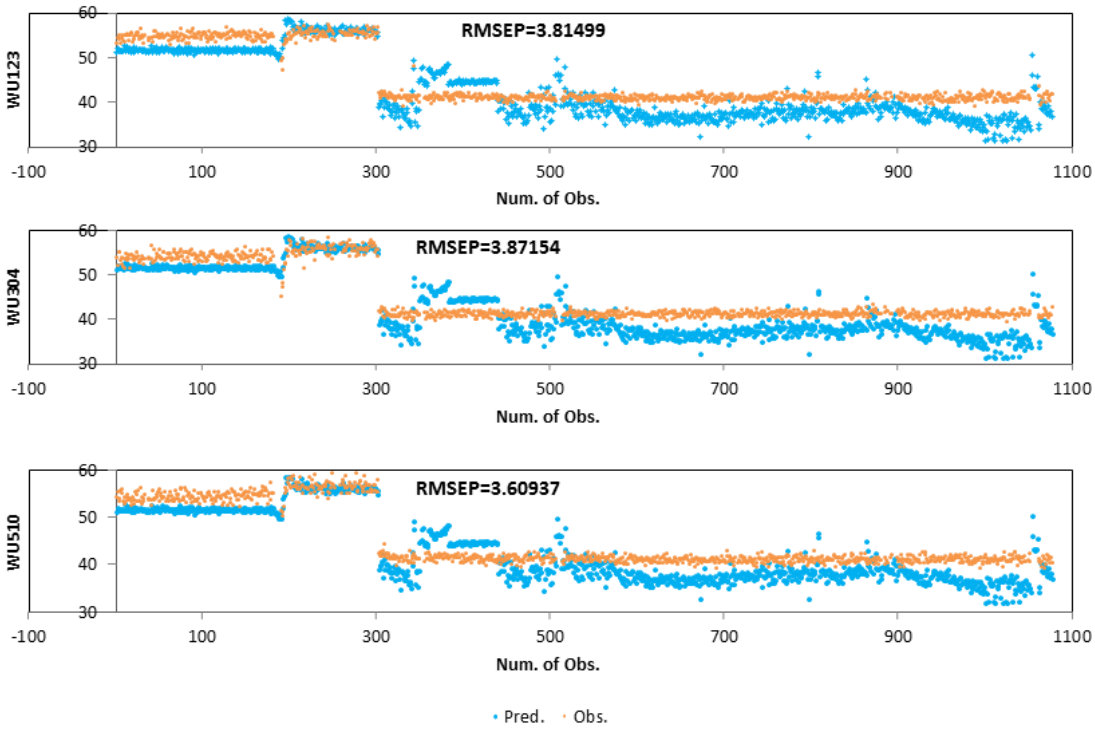

Figure A5.6 Comparison between the WU predictions and the observations for the prediction set data at three different locations across the film width when the PLS model is built based on the information in the CE and WU measurements in the training set data. Blue and orange dots correspond to the prediction and observation values, respectively.

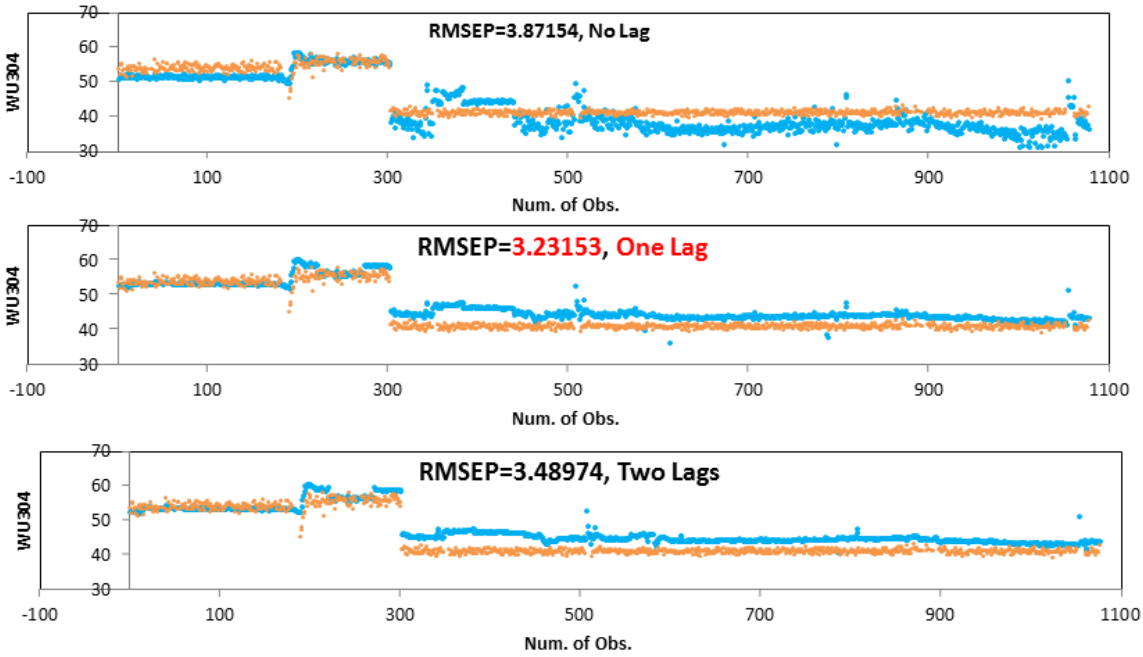

Figure A5.7 Comparison between the WU predictions and the observations for the prediction set data at the center of the film when the PLS model is built based on the information in the CE and WU measurements and one or two lagged CE measurements in the training set data. Blue and orange dots correspond to the prediction and observation values, respectively.

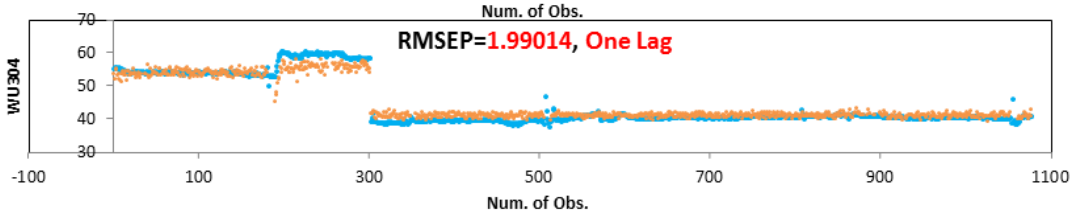

Figure A5.8 Comparison between the WU predictions and the observations for the prediction set data at the center of the film when the PLS model is built based on the information in the DBPs, some upstream process variables, CE and WU measurements and one-lagged CE measurements in the training set data. Blue and orange dots correspond to the prediction and observation values, respectively.

A5.4 Conclusions

Multivariate data analysis and experimental modeling of the DuPont™ Tedlar® PVF film casting process is performed in this appendix in order to collect information for improving the current control strategy of the plant and to thereby reduce product transition waste. Steady state and dynamic historical data during several film changes and grade switches were collected. ProMV software was used for statistical data analysis. The analysis performed here showed that the WU

measurements are highly correlated with CE measurements, DBPs and some upstream variables. Therefore, CE measurements could meaningfully be implemented in the control scheme of the extrusion process in a cascade fashion to achieve uniform film thickness especially during product transition and batch switches.

A5.5 References

(1) VanAntwerp, J. G.; Featherstone, A. P.; Braatz, R. D.; Ogunnaike, B. a. Cross-directional control of sheet and film processes. *Automatica* **2007**, *43*, 191–211.

(2) Ebnesajjad, S. *Polyvinyl Fluoride Technology and Application of PVF*; First Edit.; ELSEVIER, 2013.

(3) Lightfoot, T. No Title, 2013.

(4) Rawlings, J. B.; Chien, I.-L. Gage control of film and sheet-forming processes. *AIChE J.* **1996**, *42*, 753–766.

(5) Bergh, Luis G., MacGregor, J. F. Spatial control of sheet and film forming processes. *Can. J. Chem. Eng.* **1987**, *65*, 148.

(6) Peng, Y., Vrancic, D., & Hanus, R. Anti-windup, bumpless transfer, and conditioned transfer for PID controllers. *IEEE Control Syst.* **1996**, *16*, 48–57.

(7) Boyle, T. J. Control of cross directional variations in web forming machines.pdf. *Can. J. Chem. Eng.* **1977**, *55*, 457–461.

(8) Rao, C. V; Campbell, J. C.; Rawlings, J. B.; Wright, S. J. Efficient Implementation of Model Predictive Control for Sheet and Film Forming Processes. In *Proceedings of the American Control Conference*; 1997; pp. 2940–2944.

(9) Braatz, R. D., M. L. Tyler, M. Morar, , F. R. Pranckh, and L. S. Identification and cross-directional control of coating processes. *AIChE J.* **1992**, *38*, 1329–1339.

(10) Agassant, J. F.; Demay, Y.; Sollogoub, C.; Silagy, D. Cast film extrusion- An overview of experimental and theoretical approaches. *Int. Polym. Proc.* **2005**, *20*, 136–148.

(11) Barq, P.; Haudin, J. M.; Agassant, J. F. Isothermal and anisothermal models for cast film extrusion. *Int. Polym. Proc.* **1992**, *7*, 334–349.

(12) Zheng, H.; Yu, W.; Zhou, C. X.; Zhang, H. B. Three-dimensional simulation of the non-isothermal cast film process of polymer melts. *J. Polym. Res.* **2006**, *13*, 433–440.

(13) Wang, Y.; Tsay, H.-Y. Computer-aided geometric design of the preform dies for flat film extrusion through 3-D finite element flow analysis. *Polym. Eng. Sci.* **1994**, *34*, 1090–1100.

(14) Bird, R. B.; Armstrong, R. C.; Hassager, O. *Dynamics of Polymeric Liquids Volume 1 Fluid Mechanics*; Second Edi.; John Wiley & Sons, 1987.

(15) Liang, Y.; Kurihara, K.; Saito, K.; Murakami, H.; Kumangai, K. Profile Control of Plastic Sheet in an Industrial Polymer Processing Process. *Polym. Eng. Sci.* **1998**, *38*, 1740–1750.

(16) Corscadden, K. W.; Duncan, S. R. Multivariable disturbance modelling for web processes. *Ind. J. Syst. Sci.* **2000**, *31*, 97–106.

(17) Duncan, S.; Wellstead, P. Processing data from scanning gauges on industrial web processes. *Automatica* **2004**, *40*, 431–437.

(18) Tyler, M. L.; Morari, M. Estimation of cross-directional properties: Scanning vs. Stationary Sensors. *AIChE J.* **1995**, *41*, 846–854.

(19) Rigopoulos, A.; Arkun, Y. Principal components analysis in estimation and control of paper machines. *Comput. Chem. Eng.* **1996**, *20*, S1059–S1064.

(20) Reinig, J.; Pottmann, M.; Jackson, K. *MULTIVARIATE MODEL BASED FEEDFORWARD CONTROL OF TEDLAR FILM TRANSITIONS*; 2007.

(21) Canning, K.; Co, A. Edge Effects in Film Casting of molten Polymers. *J. Plast. Film Sheeting* **2000**, *16*, 188–203.

(22) Dobroth, T.; Erwin, L. Causes of edge beads in cast films. *Polym. Eng. Sci.* **1986**, *26*, 462–467.

(23) Shi, R.; Macgregor, J. F. Modeling of dynamic systems using latent variable and subspace methods. *J. Chemom.* **2000**, *14*, 423–439.

(24) MacGregor, J. F.; Yu, H.; García Muñoz, S.; Flores-Cerrillo, J. Data-based latent variable methods for process analysis, monitoring and control. *Comput. Chem. Eng.* **2005**, *29*, 1217–1223.