

University of Alberta

**Multiple ARX Model Based Identification for
Switching/Nonlinear Systems with EM Algorithm**

by

Xing Jin

A thesis submitted to the Faculty of Graduate Studies and Research in partial
fulfillment of the requirements for the degree of

Master of Science

in

Process Control

Department of Chemical and Materials Engineering

©Xing Jin
Spring 2010
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Examining Committee

Prof. Biao Huang, Chemical and Materials Engineering

Prof. Stevan Dubljevic, Chemical and Materials Engineering

Prof. Qing Zhao, Electrical and Computer Engineering

Abstract

Benefits brought by automatic control systems through increasing the production efficiency, reducing the production cost and environmental footprint have already been seen and experienced by process industry in the past few decades. However, given the rapid increase of the complexity in process itself as well as its interactions with the outside world, it is getting more common to observe that the controlled process exhibits time-varying/switching behaviors due to, for instance, the change of its operating conditions such as grade change in polymer plants or the feedstock change for chemical reactors. In this thesis, systems showing gradually varying dynamics or abrupt changing behaviors will be referred to altogether as switched systems. In real production practice, these switching behaviors may greatly compromise the performance of the most of current control systems owing to the fact that they are not initially designed for process with switching behaviors. Due to the critical role the control systems play in ensuring the safety as well as the profitability of the plant operation, it is desirable to enable the control systems to achieve satisfactory performance for the switched process. Therefore, as a prerequisite for any model-based optimal controllers, modeling of the switched systems is of necessity and it directly determines the performance of the designed controller.

Two different types of switching mechanism are considered in this thesis, one is featured with abrupt/sudden switching while the other one shows gradual changing behavior in its dynamics. The Expectation-Maximization (EM) algorithm is employed throughout the thesis in identifying the switched systems. Identification methods with/without considering the modeling of the switching dynamics are proposed and they are tested on various numerical simulation examples as well as a

pilot scale tank system. For the identification method without considering the modeling of switching dynamics, its robustness to the data set polluted with outliers is achieved by assuming a contaminated Gaussian distribution as the distribution of noise. It is shown that, through the comparison of the identification results from the proposed method and a benchmark method, the proposed robust identification method can achieve better performance when dealing with the data set mixed with outliers. For the identification method in which the modeling of discrete switching dynamics is considered, the hidden Markov model is employed in describing the evolution of the discrete switching variable. By simultaneously estimating parameters of the discrete dynamics (hidden Markov model) and continuous dynamics (local ARX model), it is found that the performance of the identification method can be effectively increased compared with the methods without considering the switching dynamics.

In the process industry, process may gradually switch over several local sub-systems. To model the switched systems exhibiting gradual or smooth transition among different local models, in addition to estimating the local sub-systems parameters, a smooth validity (an exponential function) function is introduced to combine all the local models so that throughout the working range of the gradual switched system, the dynamics of the nonlinear process can be appropriately approximated. Scheduling variable(s) is/are defined to represent the conditions under which the process is operated and it is assumed to be measurable. The EM algorithm is applied in estimating the local model parameters as well as the key parameters for the validity functions for each local model. Verification results on a simulated numerical example and an CSTR process confirm the effectiveness of the proposed Linear Parameter Varying (LPV) identification algorithm.

Acknowledgements

Standing at the moment of wrapping up all the research work that has been done during the course of last two years and putting it into this thesis, I am really grateful to all the people who have helped me, guided me and encouraged me in my way of pursuing my MSc degree. It was, undoubtedly, a challenging (or even painful) process through which I have "grown up" into a more mature and knowledgeable engineer. All those days and nights spent in the labs become pieces of joyful moment which will definitely serve to remind me the time that I spent as a member of Computer Process Control Group.

Out of all the people that I would like to express my gratitude, Dr. Biao Huang, who supervised my research work over last two years, is the first and foremost person that my special thanks goes to. His understanding and support grant me friendly and relatively easy environment in pursuing my own goals. It was through his dedication to excellence and rigorousness in solving research problems making me understand the meaning of being an academic researcher. I would not be able to finish this thesis without his patient guidance and encouragement. The other person that deserves my special thanks is Dr. David Shook. "Sharing" an office with him grants me opportunities to talk and discuss all sorts of questions that bump into my mind which, most of time, can be quite inspirational and educative.

As a fact of being widely accepted, doing research is really a long march in which you may encounter all the difficulties which are missed out in our daily life. It is friendship that grants me joy, peace and courage to wade through all the obstacles and reach the destination. I sincerely thank all my friends: Arjun V Shenoy, Anuj Narang, Samidh Pareek, Fan Yang, Fei Qi, Jaganath Ramarathnam, Xinguang Shao, Chunyu Liu, Hector Siegler, Kartik Surisetty, Zhen Shi, Yang Lin who are always standing on my side and unselfishly offer their help whenever I need it. I wish you all the best in your pursuit for your goals.

Last but not least, I would like to express all my undescrivable gratitude to my parents, my elder brothers for their understanding and unconditional support over my whole life. Words can not describe how much I am indebted to you for all the love that you have given to me.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Identification of Switched ARX Systems	2
1.3	Brief Introduction to the EM Algorithm	5
1.3.1	EM Algorithm Revisit	5
1.4	Thesis Overview	6
1.4.1	Thesis Contributions	6
1.4.2	Thesis Outline	6
2	Robust Identification of Piecewise/Switching Autoregressive eX-	
	ogenous Process	8
2.1	Introduction	9
2.2	Problem Statement	12
2.3	EM Algorithm	13
2.3.1	Formulation of The PWARX System Identification Problem Based On The EM Algorithm	13
2.4	Initialization of EM Algorithm and Refinement of Data Classification	18
2.4.1	Initialization of EM Algorithm	18
2.4.2	Refinement of Data Classification Result	19
2.5	Robust Parameter Re-estimation of Local Models and Region Partition	21
2.5.1	Robust Parameter Re-estimation of Each Local ARX Model .	21
2.5.2	Region Partition	23
2.6	A More Complex Simulation Example	25
2.7	Application Example: Continuous Fermentation Reactor	27
2.8	Identification of An Experimental Switched Process Control System .	33
2.8.1	Case 1: Experimental Data Without Outliers	35
2.8.2	Case 2: Experimental Data With Outliers	38
2.9	Conclusion	42

3	Identification of Switched Markov Autoregressive eXogenous Systems With Hidden Switching State	43
3.1	Introduction	43
3.2	Introduction to SMARX Systems and The EM Algorithm	47
3.2.1	An Illustrative Bimodal SMARX System	47
3.2.2	EM Algorithm	48
3.3	Mathematical Solution for The SMARX System Identification	48
3.4	Verification by A Numerical Simulation Example	52
3.5	A Simulated Continuous Fermentation Reactor	54
3.5.1	Process Description and Problem Statement	54
3.5.2	Results	56
3.6	Conclusion	59
4	Identification of Switched Markov Autoregressive eXogenous Systems With Hidden Switching State	62
4.1	Introduction	62
4.2	Mathematical Formulation of Nonlinear Modeling Using Multiple Local Models	66
4.2.1	Discussion on one practical implementation issue	69
4.3	Simulations	69
4.3.1	A Numerical Simulation Example	69
4.3.2	Continuous Stirred Tank Reactor	72
4.4	Experimental Verification & Discussion	75
4.4.1	Process Description and Identification Results	75
4.4.2	Comparative Study With An Existing LPV Modeling Method	80
4.5	Discussion & Conclusion	83
5	Summary and Conclusions	85
5.1	Summary of This Thesis	85
5.2	Directions for Future Work	86

List of Tables

2.1	Comparison of identified results with different starting points	18
2.2	True bimodal PWARX system parameters and estimated parameters	24
2.3	Comparison of estimated parameters using different methods with outliers in the data set	27
2.4	Identified models under different feed substrate concentrations	30
2.5	Comparison of identification results, SD stands for standard deviation	39
3.1	True system parameters and estimated parameters	53
3.2	Comparison of identification results from different identification methods. The benchmark method is proposed by Nakada et al. (2005) and the SARX method is given by Jin & Huang (2009b). SV, CV 1, CV 2 are short for self validation and cross validation results for model 1 and model 2. N\A represents can not be calculated	60
4.1	True system parameters and estimated parameters	72
4.2	CSTR model parameters and their steady state values	73
4.3	Designed experiment parameters for the heated tank process	77
4.4	Comparison of estimated parameters using different methods with outliers in the data set	83

List of Figures

1.1	Schematic diagram of the general switched system	3
2.1	Data set generated by the PWARX system expressed in 2.3	13
2.2	An explanation figure for ‘un-decidable’ data points generated by the bimodal PWARX system in Example 1	20
2.3	Comparison of data classification results before and after refinement procedure for Example 1	21
2.4	Framework of robust PWARX system identification using EM algorithm	24
2.5	Hyperplane of the PWARX system	26
2.6	Comparison of data classification result before and after refinement procedure (Cross points denote the true mode of each data point, circle points represent classification results before refinement and square ones are the results after classification refinement)	26
2.7	Input-Output data of the continuous fermenter. Dash line represents the step response when $u_1 = 20$, solid line denotes the step response when $u_1 = 26.5$	29
2.8	Input-Output data of the continuous fermenter	29
2.9	Comparison between model prediction and the true data, self validation MSE=0.0155. Dotted line represents prediction from identified models, Solid line denotes the true data from the fermenter.	30
2.10	Comparison between model prediction and the true data when $u_2 = 26.5kg/m^3$, cross validation MSE=0.008. Dashed line represents prediction from identified models, Solid line denotes the true data from the fermenter.	31
2.11	Comparison between model prediction and the true data when $u_2 = 20kg/m^3$, cross validation MSE=0.00366. Dashed line represents prediction from identified models, Solid line denotes the true data from the fermenter.	32
2.12	Schematic diagram of the tank system with switched controller	33
2.13	(a) Step response of controller 1 (b) step response of controller 2	34
2.14	Block diagram of the switched control system	35

2.15	Input-output data of the switched control system	36
2.16	Subplot (a) denotes the estimated mode sequence of the switching evolution along the time for training data . Subplot (b) self-validation of the identified multi-ARX modes (solid line: measured water level data, dashed line: prediction from the identified hybrid model)	36
2.17	Subplot (a) denotes the estimated mode sequence of the switching evolution along the time for training data . Subplot (b) cross-validation of the identified multi-ARX modes (solid line: measured water level data, dashed line: prediction from the identified hybrid model)	37
2.18	Input-output data of the switched control system with manually added outliers	38
2.19	Comparison of the step response of the identified models for mode 1. (a) Step response of the regular EM algorithm identified model. Solid line: step response of the reference model, bold dotted line: mean step response of the identified models, dotted line: standard deviation bound of the step response from the identified models.(b)Step response of the comparative method identified model. Solid line: step response of the reference model, bold dash-dotted line: mean step response of the identified models, dash-dotted line: standard deviation bound of the step response from the identified models. (c) Step response of the robust EM algorithm identified model. Solid line: step response of the reference model, bold dash-dotted line: mean step response of the identified models, dash-dotted line: standard deviation bound of the step response from the identified models.	40
2.20	Comparison of the step response of the identified models for mode 2. (a) Step response of the regular EM algorithm identified model. Solid line: step response of the reference model, bold dotted line: mean step response of the identified models, dotted line: standard deviation bound of the step response from the identified models.(b)Step response of the comparative method identified model. Solid line: step response of the reference model, bold dash-dotted line: mean step response of the identified models, dash-dotted line: standard deviation bound of the step response from the identified models. (c) Step response of the robust EM algorithm identified model. Solid line: step response of the reference model, bold dash-dotted line: mean step response of the identified models, dash-dotted line: standard deviation bound of the step response from the identified models.	41

3.1	Graphical representation of SMARX system with hidden switching variable	46
3.2	Comparison between estimated and actual state sequence	53
3.3	Schematic diagram of the fermentation reactor	55
3.4	Input-Output data of the continuous fermenter. Dash line represents the step response when $u_1 = 20$, solid line denotes the step response when $u_1 = 26.5$	55
3.5	Input-Output data of the continuous fermenter	57
3.6	Mode sequence of the fermentation reactor	57
3.7	Self validation of the identified model, MSE=0.001423. Red solid line represents the true fermener data, blue dashed line denotes the prediction of the simulated model	58
3.8	Comparison of the true modes with the identified mode sequence, 92.1% of time is right. Red cross is the estimated mode sequence, blue solid line denotes the true mode of the process through out the simulation time period.	58
3.9	Cross validation of the identified model under mode 1, MSE=0.0.000479. Red solid line represents the true fermener data, blue dashed line denotes the prediction of the simulated model	59
3.10	Self validation of the identified model under mode 2, MSE=0.0553. Red solid line represents the true fermener data, blue dashed line denotes the prediction of the simulated model	60
4.1	Comparison of the step response between identified and true models under different T values (Red solid line: step response of the true model, black dashed line: identified LPV model using the proposed method, blue dotted line: identified LPV model using the interpolation method given in Xu et al. (2009)) (A): system step response when $T = 1.5$ (B): system step response when $T = 1.9$ (C): system step response when $T = 3.4$ (D): system step response when $T = 4$	71
4.2	Weight of each local model at different operating points	71
4.3	Input-output excitation data, (a): outlet reagent A concentration (b): cooling water flow rate	73
4.4	Self validation of the identified CSTR LPV model, MSE= 9.7535×10^{-5} . Red solid line si the real process output and the blue dashed line is the simulated output from the identified LPV model	74
4.5	Cross validation of the identified CSTR LPV model, MSE= 4.4838×10^{-5} . Red solid line si the real process output and the blue dashed line is the simulated output from the identified LPV model	74

- 4.6 Validity function of each local model under different working conditions (coolant flow rates) 75
- 4.7 Simplified heated tank system piping and instrumentation diagram . 76
- 4.8 Heated tank process input-output data (a) water temperature in the tank, process output; (b) steam flow rate, process input 77
- 4.9 Heated tank process scheduling variable 78
- 4.10 Self-validation result, the fitting percentage equals 93.6%. Red solid line is the collected process data , blue dashed line is the simulated output of the identified LPV model 78
- 4.11 Cross-validation input-output data, (a) tank temperature, process output, (b) steam flow rate, process input 79
- 4.12 Cross-validation result, the fitting percentage equals 97%. Red solid line is the collected process data , blue dashed line is the simulated output of the identified LPV model 80
- 4.13 Comparison among the predictions from different local models and the collected real process data, pink cross solid line: predictions from the model identified at level = 0.15 meters, black dash dotted line: predictions from the model identified at level = 0.25 meters, red solid line: real process response at level = 0.3 meters, blue dashed line: predictions from the model identified at level = 0.39 meters, blue plus line: predictions from the model identified at level = 0.35 meters 81
- 4.14 Noisy tank level measurement, measurement noise with variance of 0.001 is added to the original level measurement 82

List of Symbols

y_k	Switched system output(s) at sampling time instant k
u_k	Switched system input(s) at sampling time instant k
θ_i	Parameters of the i th local sub-system
α_i	The probability the i th sub-model takes effect
σ^2	Switched system noise variance
Θ	Switched system overall parameters, including $\theta_i, i = 1, 2 \dots M$, $\sigma^2, \alpha_i, i = 1, 2 \dots M$
x_k	The regressor which is composed of the process input u and the output y at sampling instant k
\bar{x}_k	The regressor with intercept $\begin{bmatrix} x_k \\ 1 \end{bmatrix}$
\hat{y}_k	Predicted value of system output(s) at sampling instant k
e_k	Switched system noise at sampling instant k
Z_k	Switched system observed data $\{y_k, x_k\}$ at sampling instant k
C_{obs}	Switched system observed data, $\{Z_1, Z_2 \dots Z_N\}$
C_{mis}	Switched system unobserved data
$C = \{C_{obs}, C_{mis}\}$	Switched system complete data
N	number of data points being collected
T	scheduling variable(s)
I	Switched system hidden switching variable(s)
α_{ij}	Transition probability from sub-model i to sub-model j
π_i	The probability of being i th sub-model initially

$L()$	Likelihood calculation function
Q	Expectation of the complete data C
W	Weighting matrix $W = \begin{bmatrix} P_{1,i} & 0 & \dots & 0 \\ 0 & P_{2,i} & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & P_{N,i} \end{bmatrix}_{(N,N)}$, $P_{k,i}, k =$ $1, 2 \dots N$ represents the probability of k th, $k = 1, 2 \dots N$ data point coming from i th, $i = 1, 2 \dots M$ sub-model
Y	$Y = [y_1 \quad y_2 \quad \dots \quad y_N]_{(N,1)}$
\bar{X}	$\bar{X} = \begin{bmatrix} x_1 & \dots & x_N \\ 1 & \dots & 1 \end{bmatrix}_{(n+1,N)}$
w_{ki}	The likelihood of k th data coming from i th local model

List of Abbreviations

<i>ARX</i>	Autoregressive eXogenous
<i>ARMAX</i>	Auto Regressive Moving Average with eXogenous Input
<i>BJ</i>	Box-Jenkins
<i>MPC</i>	Model Predictive Control
<i>SISO</i>	Single Input Single output
<i>EM</i>	Expectation-Maximization
<i>SWARX</i>	Switched Autoregressive eXogenous
<i>SMARX</i>	Switched Markov Autoregressive eXogenous
<i>AIC</i>	Akaike Information Criterion
<i>PWARX</i>	Piecewise Autoregressive eXogenous
<i>HMM</i>	Hidden Markov Model
<i>LPV</i>	Linear Parameter Varying
<i>PWA</i>	Piecewise Affine
<i>MSE</i>	Mean Square Error

Chapter 1

Introduction

1.1 Motivation

In the past few decades, automatic control systems have been widely installed in manufactory industry owing to the significant amount of benefits these systems have brought in increasing the production efficiency and safety while, on the other hand, reducing the environmental footprint caused by the production process. For instance, model predictive control, which is also known as MPC, has already seen an great success in controlling multivariate interactive process after being first introduced around 1960s (Qin & Badgwell (2003)). However, in the mean time, factors such as fierce market competition, increase of process complexity, interactions with digital control systems cause unexpected variability in process dynamics under control. Taking a polymer plant operation for example, fierce market competition drives the plant operating under different conditions from time to time so as to produce polymer with various grades. Such frequent changes in plant operation conditions can greatly compromise the performance of the control systems as most of these automatic controllers are not designed for time-varying process. How to design controllers which can achieve satisfactory or even optimal control performance under time-varying plant dynamics has become an area of interest for both academia researchers and industrial practitioners. As the prerequisite for any advanced controller design and implementation, process modeling plays a key role in determining the performance of the designed controller.

Switched systems are dynamic systems in which both continuous and discrete valued dynamics exist simultaneously. Due to its capability in describing process with time-varying dynamics, it is gaining increasing attention and considerable amount of work has already been done in switched system modeling (Barton & Pantelides (1994);Ferrari-Trecate et al. (2003);Bempoard et al. (2005);Vidal et al. (2003);Nakada et al. (2005);Ragot et al. (2003);Jin & Huang (2009b);Juloski et al. (2005)). This thesis is concerned with the identification of the switched systems.

According to the switching behavior, the switched systems are categorized into two general groups, one is featured with the abrupt change/switch of the process dynamics while the other one is characterized by the smooth/slow transition among different process dynamics. In chemical process industry, both of these two types of switching are commonly experienced. For example, abrupt stock change for the feed of chemical reactors can introduce immediate change in the reaction kinetics which ultimately causes the switching of the process dynamics. Grade change in polymer plant requires the process being transitioned gradually from one local operating condition to the other, the process dynamics does not switch immediately due to the existence of the transition period. The following sections will present the general switched system identification problem in a mathematical form and give a brief introduction on the Expectation-Maximization algorithm (EM) through which the identification algorithms are built in this thesis. The main contributions of the thesis as well as its organization are given in the end of this chapter.

1.2 Identification of Switched ARX Systems

Switched systems provide a general framework in describing nonlinear/time-variant process in which switching/transition among sub-systems with various process dynamics occurs. In this thesis, a special but representative form of the switched systems named switched Autoregressive eXogenous system (SWARX) is investigated and relevant identification methods for different types of the SWARX systems are proposed. The mathematical formulation along with the schematic diagram of a general switched ARX system are given in Equation 1.1 and Figure 1.1:

$$y_k = \theta_k^T \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, k = 1, 2 \dots N, \theta_k \in \theta_1 \dots \theta_M \quad (1.1)$$

where y_k and x_k in Equation 1.1 represent the output and the regressor of the SWARX system respectively. e_k is the Gaussian distributed noise. M denotes the number of sub-models and N represents the number of data points that have been collected. As can be seen from the equation, at each sampling time instant, the value of system parameters θ_k may change due to the switching among sub-systems. If the switching variable is observable or can be inferred from the other measurement indirectly, in other words, if the identity of the sub-system is known at each sampling instant, then the SWARX system identification problem would be trivial as linear identification methods such as ordinary least squares can be directly applied to each cluster of the data set and the local ARX model parameters can be estimated. On the other side, if the parameters of each local model as well as the variance of the process noise are already known in advance, the data points could be directly classified based on the likelihood and the value of the switching variable

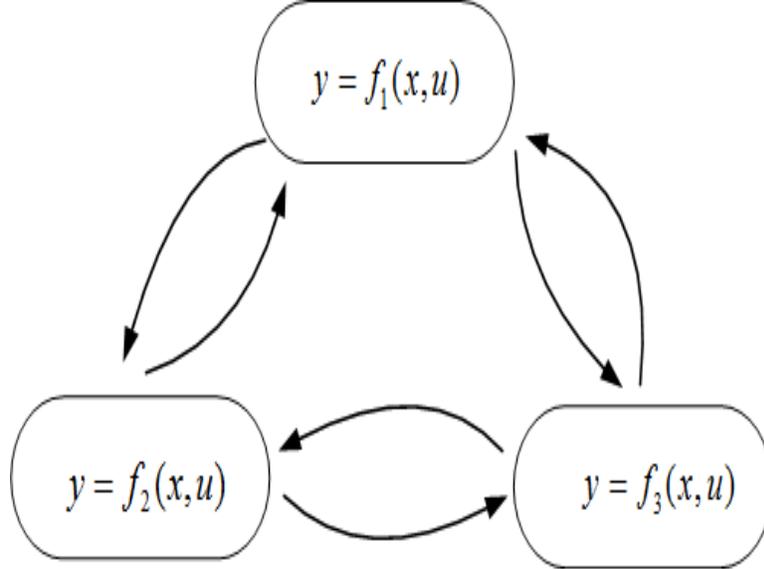


Figure 1.1: Schematic diagram of the general switched system

can be inferred from the cluster identity of each classified data point. For nonlinear systems, even if the switching variable is known or measured, the global LPV model does not depend on the local model only. The global model is weighted combination of all local models.

In practice, it is common that neither the local ARX model parameters nor the value of switching variable are known *a priori* owing to the limitation of the measurement as well as the process knowledge. Therefore, it is desirable that the proposed identification methods do not require any information regarding the system parameters. In accordance with the types of switching behavior that is considered, the identification problems that have been investigated and solved in this thesis can be categorized into two classes. For the problem that will be discussed in Chapter 2 and Chapter 3, it can be stated as:

Problem 1: *Identifying the SWARX system local ARX model parameters as well as the value of the hidden switching variable based on the collected system input and output data*

In Chapter 4, the discussion is directed to nonlinear process modeling problem in which an LPV model is identified to approximate the process dynamics over the whole operating range. It can be stated as:

Problem 2: *Identifying LPV model with multiple local ARX model structure for nonlinear systems based on the collected input and output data*

Some common assumptions are made all over the different chapters of the thesis to simplify the problems while still preserving the applicability of the proposed algorithms. The first assumption requires that the number of sub-models M an SWARX system has is known *a priori*. Although under certain circumstances, M may not be able to be inferred or observed, the number of the sub-models can still be estimated. For instance, in the experimental example illustrated by Juloski et al. (2005), based on the analysis of the working mechanism of the pick-and-place machine, 4 modes are distinguished, which equally means that the value for M equals 4. The second assumption that has been made is that the local ARX model structures including the model order, time delay are known in advance. In system identification, how to select appropriate model structures has been a twin issue with model parameter estimation as the selected model structures have direct impact on the parameter estimation accuracy. Different selection criteria such as AIC has been proposed in other literatures and relevant estimation methods can be found in Ferrari-Trecate et al. (2003); Nakada et al. (2005); Ljung (1987); Akaike (1974); Bindlish (2003). Moreover, as discussed in Jelali & Huang (2009), for most of industrial process, first or second order plus time delay model would be sufficient in capturing the process dynamics. Hence, in case the local ARX model structures are not able to be obtained from the process knowledge, the existing algorithms can be applied to search for the appropriate model structures (Zhu (2001); Kay (2001)).

The switched ARX systems that will be discussed in this thesis can be categorized into three groups based on their switching patterns and switching dynamics. As mentioned earlier in this chapter, two types of switching patterns are considered, abrupt switching and smooth switching/transition in nonlinear processes. Different identification strategies are proposed in an effort to estimate the system parameters. Moreover, for systems that switch in an abrupt fashion, further categorization is made based on the switching dynamics of the system. It will be shown that, for some switched ARX systems that exhibit complex switching behavior, instead of treating the switching variable as completely random, the Markov model can be utilized to model the discrete switching dynamics and it is found that the accuracy of the estimated parameters is higher. The switched ARX systems with switching dynamics being modeled as a Markov model are named as switched Markov ARX systems (SMARX) in the thesis.

1.3 Brief Introduction to the EM Algorithm

1.3.1 EM Algorithm Revisit

The Expectation-Maximization (EM) algorithm, after being first introduced by Dempster et al. (1977), has found extensive applications in various areas including machine learning, computer vision, speech recognition, bioinformatics, psychometrics for finding maximum likelihood estimates of parameters in probabilistic models. Assume that a complete data set C consists of two parts: $\{C_{obs}, C_{mis}\}$, C_{obs} is the data collected from the process and it is called incomplete data set. C_{mis} needs to be estimated from C_{obs} and is called missing data set. The EM algorithm consists of two consecutive steps, the first step calculates the Expectation of the complete data C with respect to the missing data while the second step searches for the parameters to increase the Expectation of the complete data. Putting the objective of the EM algorithm in a mathematical form, it can be written as (Dempster et al. (1977)):

$$L(C_{obs}, \Theta) = \int f(C | \Theta) dC_{mis} \quad (1.2)$$

where Θ in Equation 1.2 represents the system parameters while $f(\cdot)$ is the probability distribution function. In the E-Step, parameter estimation results from the previous iteration are used to compute the expectation of the complete data likelihood (Dempster et al. (1977))

$$Q(\Theta | \Theta^{old}) = E_{C_{mis} | (\Theta^{old}, C_{obs})} \{ \log L(C, \Theta) \} \quad (1.3)$$

where the conditional expectation is defined as $E_{A|B} \{g(A)\} \triangleq \int g(A) f(A | B) dA$ if A is a continuous random variable; if A is a discrete random variable the integration is replaced by summation. M-step maximizes the expectation shown in Equation 1.3 with respect to Θ so as to ensure that the newly found Θ^{New} makes the log likelihood of the complete data set C non-decreasing, which equally means that

$$Q(\Theta^{New} | \Theta^{old}) \geq Q(\Theta | \Theta^{old}), \forall \Theta \quad (1.4)$$

Hence, starting with some initial values of the parameters, the EM algorithm can ultimately converge to some stationary points after finite steps of iteration. While EM algorithm has received significant attention in many areas, it is a relatively new data processing/optimization technique in control engineering. So far, the main applications of the EM algorithm in process control literature have been mainly on parameter estimation for nonlinear/switched systems (Chitralkha et al. (2009); Goodwin & Aguero (2005, 2008)).

1.4 Thesis Overview

1.4.1 Thesis Contributions

Relevant work on switched systems identification has already been seen and different methods targeting on the identification of various types of the switched systems have been proposed (Logothetis & Krishnamurthy (1999); Doucet et al. (2001); Zhu & Xu (2008); Jin & Huang (2009b); Lee & Poolla (1996); Banerjee et al. (1997); Xu et al. (2009); Ferrari-Trecate et al. (2003); Nakada et al. (2005)). Paoletti et al. (2007) gave a nice review on the status of the research on the switched system identification along with some main methods that have been proposed. As the characteristics that distinguish the research being performed in this thesis from the other existing work, the main contributions of this thesis are listed below:

1. Formulate and solve the PWARX system identification problem under the EM algorithm framework
2. Robustness of the EM algorithm to the outliers is considered and a robust strategy is proposed for the EM algorithm. In the literature, some EM algorithms with different robustness strategy have already been suggested to handel the outliers (Saldju & Landgrebe (2000); Kalyani & k. Giridhar (2007); Saint-Jean et al. (2000)). However, the proposed strategy is based on the rigorous contaminated Gaussian distribution to describe the outliers, leading to an explicit weighted least square solution.
3. A computationally cost-efficient method is developed for the classification of un-decidable data points
4. Evaluation of the proposed identification method is performed on a simulated continuous fermenter as well as a pilot-scale switched control system
5. For SMARX systems identification, the continuous ARX model for process dynamics and the discrete-valued HMM for switching dynamics are identified simultaneously, thus improving both the clustering and identification performance.
6. Apply the EM algorithm to the identification of linear parameter varying systems and test its performance on nonlinear process modeling.

1.4.2 Thesis Outline

The rest of the thesis is organized as follows: Chapter 2 describes the identification of Piecewise/Switching ARX process and presents the validation results of the proposed identification method based on the estimation results from the simulated examples as well as the experiment performed on a pilot-scale setup. Chapter 3 explains the concept of the SMARX system and renders the identification formulation of the SMARX systems within the EM algorithm framework. Identification

results from the proposed SMARX identification method are compared against the other comparative method in terms of the parameter estimation accuracy. Chapter 4 provides the detailed procedures on formulating the LPV system identification problem within the EM algorithm framework and gives the algorithm verification results on simulated numerical and chemical engineering examples. Chapter 5 draws the conclusion based on the work that has been done in this thesis and provides some perspectives for future research on the switched system identification.

Chapter 2

Robust Identification of Piecewise/Switching Autoregressive eXogenous Process

¹ A robust identification approach for a class of switching processes named PWARX (piecewise AutoRegressive eXogenous) processes is developed in this chapter. It is proposed that the identification problem can be formulated and solved within the EM (Expectation-maximization) algorithm framework. However, unlike the regular EM algorithm in which the objective function of the maximization step is built upon the assumption that the noise comes from a single distribution, contaminated Gaussian distribution is utilized in the process of constructing the objective function which, effectively makes the revised EM algorithm robust to the latent outliers. Issues associated with the EM algorithm in the PWARX system identification such as sensitivity to its starting point as well as inability to accurately classify ‘undecidable’ data points are examined and a solution strategy is proposed. Data sets with/without outliers are both considered and the performance is compared between the robust EM algorithm and regular EM algorithm in terms of their parameter estimation performance. Finally, a modified version of MRLP (multi-category robust linear programming) region partition method is proposed by assigning different weights to different data points. In this way, negative influence caused by outliers could be minimized in region partitioning of PWARX systems. Simulation as well as application on a pilot-scale switched process control system are employed to verify the efficiency of the proposed identification algorithm.

1. This chapter has been published in "Robust Identification of Piecewise/Switching Autoregressive eXogenous Process, AIChE Journal, DOI 10.1002/aic.12112"

2.1 Introduction

Hybrid systems are gaining increasing attention in process control research (Banejee & Arkun (1998); P. Mhaskar and N.H. El-Farra and P. D. Christofides (2005); Morari (2007)). Hybrid systems are dynamic systems in which both continuous and discrete valued dynamics exist simultaneously. A number of examples on hybrid systems can be found in fields such as process control, embedded system, electrical circuits, biological process and so on (Juloski (2004)). In chemical process industry, due to the inborn complexity of the chemical process as well as wide application of automated control systems, discrete behavior of the chemical process control system is commonly experienced (El-Farra & Christofides (2003); El-Farra et al. (2005)). The intricate interactions between continuous behavior (driven by the underlying physical laws such as mass and energy conversation) and the discrete events pose great challenges for both academic researchers and industrial practitioners. Motivated by the economic, safety and environmental considerations, relevant researches on hybrid chemical process modeling (Barton & Pantelides (1994)), optimization (Barton et al. (2000); Barton & Lee (2004)) and control have been conducted (El-Farra & Christofides (2003); El-Farra et al. (2005)).

As an important subclass of hybrid systems, PWA (Piecewise affine) system is receiving a growing interest owing to its capability of describing a large number of processes by switching among different affine subsystems when state/input comes to a different region. Furthermore, equivalence of PWA systems to other forms of hybrid system models, such as mixed logical dynamical system, linear complementary systems and max-min-plus-scaling systems has been proved under mild conditions (Bempoar & Ferrari-Trecate (2000); Heemels et al. (2001)), which further stimulates the interest in the research of PWA systems. Provided that each affine subsystem is a linear ARX model, the PWA system can be considered as a PWARX system.

In the past few years, a number of methods for PWARX systems identification have been put forward. A data clustering based identification algorithm (Ferrari-Trecate et al. (2003)) has been proposed in which data clustering, linear identification and region partition are performed together to identify PWA subsystems as well as valid region for each subsystem from input-output data. Juloski et al. (2005) employ a particle filter method to estimate local ARX model parameters by sequentially processing the input-output data. A bounded error method is developed to solve the identification problem (Bempoard et al. (2005)). After roughly classifying the data, a further refinement step is taken to get rid of ‘un-decidable’ and infeasible data points. Nakada et al. (2005) apply a statistical clustering strategy in order to classify each data point to its relevant regions. An algebraic geometric approach is introduced for the identification of switched linear hybrid systems (Vidal et al.

(2003)). In this identification algorithm, the data classification procedure is set to be independent of the sub-model parameter estimation procedure by using hybrid decoupling constraint. Ragot et al. (2003) suggest an adaptive weighting method for iteratively classifying the data points and estimation of local ARX sub-models. An optimization technique is introduced by Roll et al. (2004). By considering hybrid system identification as a prediction error minimization problem, they use mixed integer programming to search for globally optimal identification solution, which becomes computationally infeasible when facing a large data set.

It can be seen from the analysis above that for the PWARX system identification, the main issue involved is how to identify each local ARX model when partition of regression space is controlled by unobservable variable and no relevant *priori* knowledge is at hand. It needs to be pointed out that Nakada et al. (2005) have used the regular EM algorithm to cluster the data points by iteratively calculating the center of different clusters which, to some extent, is similar to the data classification process presented in Ferrari-Trecate et al.(2003) by using ‘K-means’ like algorithm. The distinguished features of the method developed in this chapter are 1) By treating the unobservable data point identity as ‘missing variable’ and formulating the PWARX system identification within the EM algorithm framework, not only are the data points from different sub-models classified, but also the local ARX model parameters are estimated in the same time. This enables us to make full use of the convergence property of the EM algorithm within finite iteration steps. 2) Robustness of the EM algorithm is considered and a novel strategy for achieving this is proposed. It is not rare to find outliers in practical applications and the performance of the regular EM algorithm could deteriorate significantly in the presence of the statistical outliers (Saldju & Landgrebe (2000)). Therefore, being robust to outliers is desirable as well as necessary for the EM algorithm.

In real applications, contamination of measurements by noise may lead to misclassification of data points which lie far away from region of the intersection area. For such misclassified data points, they can be viewed as outliers to the misattributed sub-models and accuracy of parameter estimation can be greatly deteriorated. For other data points which are classified to the correct sub-model, once they are polluted with abnormal measurement or process noise, they still need to be removed from data sets for better parameter estimation. A robust parameter estimation method is applied by weighting every data point so that in local parameter estimation, smaller or even zero weights are given to outliers while normal data points are granted with much higher weight.

After finishing the data classification and sub-model identification, a robust linear programming for multi-category discrimination of polyhedral regions (Bennett & Mangasarian (1994)) is applied. This polyhedral region discrimination method

has already been employed in Juloski et al (2005) and different data points are weighted based on their likelihood to be undecidable. However, only considering those undecidable data points may not be sufficient as we have found that misclassification could also happen for data points which lie far away from intersection area. To reduce the effect of the outliers that may be present in data clusters, weights obtained in robust parameter estimation are used for each data point in the region partition stage.

When PWARX systems are considered, we assume that the switching of the system is triggered by different operating regions of the input and output although we do not know in advance how the region of the input and output is partitioned. However, if the switching mechanism of the system can not be represented by regressor space partition, the switched system would be regarded as switching along the time. This provides a more general way of treating the switched system. In this chapter, we also apply the proposed PWARX system identification algorithm to a simulated continuous fermenter as well as an experimental switched control system upon which two controllers with different characteristics operate alternatively. The identification results verify the validity of the proposed identification algorithm and show the potential of its usage in identification of various kinds of switched linear system.

The main contribution of this chapter are: 1) Formulate and solve the PWARX system identification problem under the EM algorithm framework. 2) A robust strategy is proposed for the EM algorithm. In the literature, some EM algorithms with different robustness strategy have already been suggested to handel the outliers (Saldju & Landgrebe (2000); Kalyani & k. Giridhar (2007); Saint-Jean et al. (2000)). However, the proposed strategy is based on the rigorous contaminated Gaussian distribution to describe the outliers, leading to an explicit weighted least square solution. The effectiveness of the robust procedure in resisting the abnormal data/outliers is demonstrated through comparison with the existing bench-marking method. 3) A computationally cost-efficient method is developed for the classification of un-decidable data points. 4) Evaluation of the proposed identification method is performed on a simulated continuous fermenter as well as a pilot-scale switched control system. The capability of the proposed PWARX system identification algorithm in handling switched linear systems is demonstrated.

The remainder of the chapter is organized as follows: In section 2, the PWARX system is formulated and several fundamental issues regarding the identification are explained through an example. Section 3 gives an overall introduction to the EM algorithm and formalizations for the identification of PWARX systems are obtained based on an improved version of the EM algorithm. An approach to initialize the EM algorithm together with a data classification refinement procedure is given in

section 4. Section 5 focuses on the derivation of robust parameter estimation for each sub-model parameters with reclassified data points. Effectiveness of the proposed algorithm is demonstrated in section 6 through an illustrative simulation example. Section 7 illustrates a potential application of the proposed algorithm in chemical process operation monitoring and process modeling. Section 8 explains how a process control experiment is performed and the way the proposed PWARX system identification method is employed in the analysis of the experimental data. Section 9 draws the conclusion.

2.2 Problem Statement

As an important subclass of PWA systems, a PWARX system is formulated as Ferrari-Trecate et al. (2003); Juloski et al. (2005); Roll et al. (2004):

$$y_k = \begin{cases} \theta_1^T \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, & x_k \in \chi_1 \\ \vdots \\ \theta_M^T \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, & x_k \in \chi_M \end{cases}, k = 1, 2 \dots N \quad (2.1)$$

where N , M represent number of data points collected and number of sub-models respectively, $y_k \in R$ is the output, $x_k \in R^n$ is the regressor which consists of past input and output,

$$x_k = [y_{k-1} \quad y_{k-2} \cdots y_{k-na} \quad u_{k-1}^T \quad u_{k-2}^T \cdots u_{k-nb}^T]^T \quad (2.2)$$

where na and nb are orders of the output and input, $u \in R^m$ is the input and $n = na + m \cdot nb$. $e_k \in R$ is Gaussian distributed noise with zero mean and variance σ^2 . $\theta_i \in R^{n+1}$ is the parameter vector of the i th sub-model and $\{\chi_i\}_{i=1}^M$ is the polyhedral region of the input-output space. Given the number of sub-model M , the PWARX system identification problem can be stated as:

Problem: *Assigning each data point in data set $(x_k, y_k), k = 1, 2 \dots N$ to one of M sub ARX models and identifying the parameter vector $\{\theta_i\}_{i=1}^M$ along with polyhedral region $\{\chi_i\}_{i=1}^M$ for each local ARX model.*

It is assumed that the number of sub-models M and the order of each sub-model (ARX) are given *a priori*. In the case that the number of sub-models is unknown, there also exist methods to estimate it (Ferrari-Trecate et al. (2003); Nakada et al. (2005)). On the other hand, if the orders of the sub-ARX models are unknown, fixed high-order ARX models can be used since a sufficiently high-order ARX

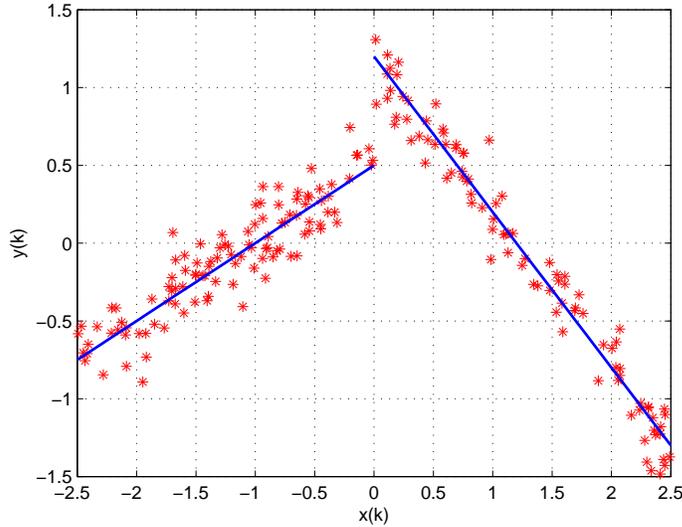


Figure 2.1: Data set generated by the PWARX system expressed in 2.3

model can approximate any linear dynamic system (Ljung (1987)). This chapter will however focus on the robust identification problem for the PWARX models given the assumptions stated above.

Example 1. Consider the following bimodal PWARX example which is used in Juloski et al (2005),

$$y_k = \begin{cases} [0.5 & 0.5] \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, & x_k \in [-2.5 \ 0] \\ [-1 & 2] \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, & x_k \in [0 \ 2.5] \end{cases}, k = 1, 2 \dots N \quad (2.3)$$

Let $N = 200$ and the data set $(x_k, y_k), k = 1, 2 \dots 200$ is generated by the bimodal PWARX system expressed by Equation 2.3. $e_k \sim N(0, 0.025)$ and input x_k follows a uniform distribution between $[-2.5 \ 2.5]$. Figure 2.1 shows the data set.

From Figure 2.1, it can be seen that the polyhedral boundary that separates two regions is the y-axis given by $\{x = 0\}$ and the regressor x is a scalar in this specific case. We will use this simple example to illustrate the PWARX identification problem.

2.3 EM Algorithm

2.3.1 Formulation of The PWARX System Identification Problem Based On The EM Algorithm

A brief introduction of the EM algorithm has been given in Chapter 1. It can be seen that the calculation of the expectation of the complete data C plays a key role

as it provides the objective function for the maximization step of the EM algorithm. The formulation for the PWARX system identification under regular and robust EM algorithms is given in the following two sections.

Regular EM Algorithm

Define $Z_k = \{x_k, y_k\}$, $k = 1, 2 \dots N$ as the observed data set generated from a PWARX system. Therefore, for data Z_k , its conditional probability equals (Mclachlan & Krishnan (1996))

$$P(Z_k | Z_{k-1} \dots Z_1) = \sum_{i=1}^M \alpha_i P(Z_k | \theta_i, Z_{k-1} \dots Z_1) \quad (2.4)$$

where α_i is the probability that i th sub-model takes effect. As C_{obs} in the PWARX system is the observed dataset $Z_k = \{x_k, y_k\}$, $k = 1, 2 \dots N$, the maximum likelihood equation for system parameter estimation is

$$\begin{aligned} \max_{\Theta} L(C_{obs}, \Theta) &= \max_{\Theta} P(C_{obs} | \Theta) \\ &= \max_{\theta_i, i=1 \dots M} \prod_{k=1}^N \sum_{i=1}^M \alpha_i P(Z_k | \theta_i, Z_{k-1} \dots Z_1) \end{aligned} \quad (2.5)$$

To simplify the problem, rather than maximizing the likelihood function directly, one usually maximizes the log likelihood function,

$$\begin{aligned} \max_{\Theta} \log L(C_{obs}, \Theta) &= \max_{\Theta} \log P(C_{obs} | \Theta) \\ &= \max_{\theta_i, i=1 \dots M} \sum_{k=1}^N \log \sum_{i=1}^M \alpha_i P(Z_k | \theta_i, Z_{k-1} \dots Z_1) \end{aligned} \quad (2.6)$$

The parameters may be estimated from Equation 2.6 by brute force maximization, but this optimization is still difficult.

To make the problem tractable and solve the maximum likelihood estimation problem, we introduce $I = \{I_1, I_2, \dots, I_N\}$ as a ‘missing variable’ to denote the sub-model identity of each data point. Following Equation 1.3, the expression for expectation of complete data $C = \{C_{obs}, I\}$ is:

$$\begin{aligned} Q(\Theta | \Theta_{old}) &= E_{I | (\Theta_{old}, C_{obs})} \{ \log P(C_{obs}, I | \Theta) \} \\ &= E_{I | (\Theta_{old}, C_{obs})} \{ \log P(Z_N, Z_{N-1} \dots Z_1, I_N \dots I_1 | \Theta) \} \\ &= E_{I | (\Theta_{old}, C_{obs})} \{ \log \prod_{k=1}^N P(Z_k, I_k | Z_{k-1}, \dots, Z_1, I_{k-1}, \dots, I_1, \Theta) \} \\ &= E_{I | (\Theta_{old}, C_{obs})} \{ \log \prod_{k=1}^N P(Z_k | Z_{k-1}, \dots, Z_1, I_k, \dots, I_1, \Theta) P(I_k) \} \\ &= E_{I | (\Theta_{old}, C_{obs})} \left\{ \sum_{k=1}^N \log [\alpha_{I_k} P(Z_k | \theta_{I_k}, Z_{k-1} \dots Z_1)] \right\} \end{aligned} \quad (2.7)$$

where α_{I_k} represents the probability that Z_k comes from the I_k th sub-model. Here, $I_k \in \{1, 2, \dots, M\}$ represents the true sub-model that Z_k comes from. In deriving Equation 2.7, we have used the fact that $P(Z_k | Z_{k-1}, \dots, Z_1, I_k, \dots, I_1, \Theta) = P(Z_k | Z_{k-1}, \dots, Z_1, I_k, \Theta)$ since I_k completely determines which sub-model that Z_k belongs to. We have also used the equation $P(I_k | Z_{k-1}, \dots, Z_1, I_{k-1}, \dots, I_1, \Theta) = P(I_k)$; namely switching between the sub-models is completely random and does not depend on which sub-model the system takes in previous instants.

By moving the Expectation operator inside the summation, Equation 2.7 becomes

$$\begin{aligned}
Q(\Theta | \Theta^{old}) &= \sum_{k=1}^N E_{I_k | (\Theta^{old}, C_{obs})} \{ \log \alpha_{I_k} + \log P(Z_k | \theta_{I_k}, Z_{k-1} \dots Z_1) \} \\
&= \sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log \alpha_i \\
&\quad + \sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log P(Z_k | \theta_i, Z_{k-1} \dots Z_1)
\end{aligned} \tag{2.8}$$

where $P(I_k = i | \Theta^{old}, C_{obs})$ in Equation 2.8 denotes the probability that the k th data point comes from the i th sub-model, and can be derived following the Bayes rule as

$$\begin{aligned}
P(I_k = i | \Theta^{old}, C_{obs}) &= P(I_k = i | \Theta^{old}, Z_k, Z_{k-1}, \dots, Z_1) \\
&= \frac{P(Z_k | I_k = i, \Theta^{old}, Z_{k-1}, \dots, Z_1) P(I_k = i | \Theta^{old}, Z_{k-1}, \dots, Z_1)}{\sum_{i=1}^M P(Z_k | I_k = i, \Theta^{old}, Z_{k-1}, \dots, Z_1) P(I_k = i | \Theta^{old}, Z_{k-1}, \dots, Z_1)} \\
&= \frac{P(Z_k | \theta_i^{old}, Z_{k-1}, \dots, Z_1) P(I_k = i)}{\sum_{i=1}^M P(Z_k | \theta_i^{old}, Z_{k-1}, \dots, Z_1) P(I_k = i)} \\
&= \frac{\alpha_i P(Z_k | \theta_i^{old}, Z_{k-1}, \dots, Z_1)}{\sum_{i=1}^M P(Z_k | \theta_i^{old}, Z_{k-1} \dots Z_1) \alpha_i}
\end{aligned} \tag{2.9}$$

For notational simplicity, we will use $P_{k,i}$, namely the probability that the k th data point comes from the i th sub-model, to denote $P(I_k = i | \Theta^{old}, C_{obs})$ in the remainder of the chapter.

Let $\bar{x}_k = \begin{bmatrix} x_k \\ 1 \end{bmatrix}$, then for the PWARX system, $\log P(Z_k | \theta_i, Z_{k-1} \dots Z_1)$ in Equation 2.8 equals

$$\begin{aligned}
\log P(Z_k | \theta_i, Z_{k-1} \dots Z_1) &= \log \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{1}{2\sigma^2} (y_k - \theta_i^T \bar{x})^T (y_k - \theta_i^T \bar{x})} \\
&= -\log \sqrt{2\pi}\sigma - \frac{1}{2\sigma^2} (y_k - \theta_i^T \bar{x})^T (y_k - \theta_i^T \bar{x})
\end{aligned} \tag{2.10}$$

Therefore, substituting Equation 2.10 of $\log P(Z_k | \theta_i, Z_{k-1} \dots Z_1)$ in the second term on the right hand side of Equation 2.8, after computing Equation 2.9 by using

previous estimation of θ_i^{old} and α_i^{old} , and then taking it into the right hand side of Equation 2.8, we get the expectation of complete data $Q(\Theta | \Theta^{old})$.

For the M-step of the EM algorithm, derivatives are taken with respect to α_i , θ_i and noise variance σ^2 in an effort to maximize the likelihood of the parameters given the observed data. After several steps of algebraic manipulations, expressions for θ_i , α_i and σ^2 at each iteration can be derived as:

$$\theta_i^{New} = \frac{\sum_{k=1}^N P_{k,i} \bar{x}_k y_k}{\sum_{k=1}^N P_{k,i} \bar{x}_k \bar{x}_k^T} \quad (2.11)$$

$$\alpha_i^{New} = \frac{\sum_{k=1}^N P_{k,i}}{N} \quad (2.12)$$

$$(\sigma^{New})^2 = \frac{\sum_{k=1}^N \sum_{i=1}^M P_{k,i} (y_k - (\theta_i^{New})^T \bar{x}_k)^T (y_k - (\theta_i^{New})^T \bar{x}_k)}{\sum_{k=1}^N \sum_{i=1}^M P_{k,i}} \quad (2.13)$$

$P_{k,i}$ is updated using value of $(\theta_i)^{New}$, $(\alpha_i)^{New}$ and $(\sigma^{New})^2$, and then the updated $P_{k,i}$ can be used for the calculation of E-step in the next iteration.

If we take a look back at Equation 2.11, it can be found that this is a typical quadratic minimization problem and the method of weighted least squares has been used unconsciously to find the parameters $(\theta_i)^{New}$ for each local ARX model. As a matter of fact, Equation 2.11 could be transferred into classic results in weighted least squares,

$$(\theta_i)^{New} = (\bar{X}^T W \bar{X})^{-1} \bar{X}^T W Y \quad (2.14)$$

where $\bar{X} = \begin{bmatrix} x_1 & \dots & x_N \\ 1 & \dots & 1 \end{bmatrix}_{(n+1,N)}$, $Y = [y_1 \ y_2 \ \dots \ y_N]_{(N,1)}$,

$W = \begin{bmatrix} P_{1,i} & 0 & \dots & 0 \\ 0 & P_{2,i} & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & P_{N,i} \end{bmatrix}_{(N,N)}$ is weighting matrix for each data point with di-

mension (N,N). Therefore, for PWARX systems, the EM algorithm works as a combination of weight updating (which is E-step) and weighted least squares (which is M-step).

Robust EM Algorithm

It is noticed that in the maximization step of the regular EM algorithm for PWARX systems identification, an assumption that the residual errors follow the single normal distribution is made. The resulted maximization step of the regular EM algorithm is essentially an ordinary least squares procedure as shown in Equation 2.14. However, if the data set contains outliers, which is not uncommon in real applications, ordinary least squares could fail and parameter estimation results obtained from it can be misleading (Tjoa & Biegler (1991)).

To offset the negative influence brought by outliers, the maximum likelihood objective function for parameter estimation is built based on a mixture distribution function (Tjoa & Biegler (1991); Albuquerque & Biegler (1996); Ragot et al. (2005)) instead of single normal distribution as being used in the maximization step of the regular EM algorithm. This mixture distribution function is also known as a contaminated Gaussian distribution.

As a result, $P(Z_k | \theta_i, Z_{k-1} \dots Z_1)$ in Equation 2.10 can be written as:

$$\begin{aligned} P((x_k, y_k) | \theta_i, (x_{k-1}, y_{k-1}) \dots (x_1, y_1)) &= P(e_k) \\ &= mP(e_k^{regular}) + (1 - m)P(e_k^{outlier}) \end{aligned} \quad (2.15)$$

In Equation 2.15, error e consists of two parts: error introduced by regular noise $e_{regular}$ and error caused by irregular noise or outlier $e_{outlier}$, and m is the probability that the noise is the regular noise.

Setting that the ratio of noise variance between $e_{outlier}$ and $e_{regular}$ equals d^2 and $d \gg 1$, Equation 2.15 can be further written as:

$$P(e_k) = m \frac{1}{\sqrt{2\pi}\sigma} \exp^{-0.5 \frac{e_k^T e_k}{\sigma^2}} + \frac{1 - m}{d} \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-0.5 e_k^T e_k}{d^2 \sigma^2} \quad (2.16)$$

Take the logarithm of Equation 2.16 and substitute it into the right hand side of Equation 2.8. Again, for the M-step of the robust EM algorithm, derivatives are taken over θ_i , α_i and noise variance σ^2 so as to ensure that Equation 1.4 is always satisfied. After several steps of mathematical manipulation, the equation for calculating new θ_i is:

$$\theta_i = \frac{\sum_{k=1}^N P_{k,i} w_k \begin{bmatrix} x_k \\ 1 \end{bmatrix} y_k}{\sum_{k=1}^N w_k \begin{bmatrix} x_k \\ 1 \end{bmatrix} \begin{bmatrix} x_k \\ 1 \end{bmatrix}^T} = (X^T W X)^{-1} X^T W Y \quad (2.17)$$

where in Equation 2.17 $w_k = \frac{m \frac{P(e_k^{regular})}{\sigma^2} + (1-m) \frac{P(e_k^{outlier})}{d^2 \sigma^2}}{m P(e_k^{regular}) + (1-m) P(e_k^{outlier})}$,

$$W = \begin{bmatrix} P_{1,i} w_1 & 0 & \dots & 0 \\ 0 & P_{2,i} w_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & P_{N,i} w_N \end{bmatrix}_{(N,N)}, \quad X = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix}_{N,2} \quad \text{and} \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}_{N,1}.$$

From Equation 2.17, it can be seen that for different data points, different weights are given based on their measurement noise as well as the probability of coming from the i th sub-model. Starting from an initial guess of the system parameters $\theta_i^{initial}$, $i = 1, 2 \dots M$, the robust EM algorithm iterates and usually converges within finite iterations.

Table 2.1: Comparison of identified results with different starting points

	Starting Point	Identified Parameters	
$\theta_{ini1} =$	$\begin{bmatrix} 0.800 & -0.839 \\ -0.116 & -0.156 \end{bmatrix}$	$\theta_{iden1} =$	$\begin{bmatrix} -0.998 & 1.999 \\ 0.517 & 0.521 \end{bmatrix}$
$\theta_{ini2} =$	$\begin{bmatrix} 0.385 & 0.922 \\ -0.919 & 0.059 \end{bmatrix}$	$\theta_{iden2} =$	$\begin{bmatrix} 0.770 & 1.013 \\ 0.117 & 0.032 \end{bmatrix}$

The *priori* knowledge of m , d and the covariance of data set σ^2 is not required since m and d can normally be set to between 0.7 - 0.9 and 10 - 15 respectively (Albuquerque & Biegler (1996)) and σ^2 has already been estimated in the EM algorithm using Equation 2.13. Furthermore, Farris et al (1979) pointed out that the robust parameter estimation procedure is not very sensitive to the values of m and d .

2.4 Initialization of EM Algorithm and Refinement of Data Classification

2.4.1 Initialization of EM Algorithm

Even though it is guaranteed that EM algorithm can converge to a stationary point after steps of iteration, there is no guarantee that this stationary point is a global or even local maxima. It may turn out to be a saddle point, and starting point plays an important role in determining what the convergence point of the EM algorithm would be (McLachlan & Krishnan (1996); Biernacki et al. (2003)).

To demonstrate the importance of the starting point for the EM algorithm, by using equations obtained in section 3, Example 1 is identified using the EM algorithm with different starting points.

The true parameters of the PWARX system are: $\theta_{true} = \begin{bmatrix} \theta_1^T \\ \theta_2^T \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 0.5 & 0.5 \end{bmatrix}$

It is shown in Table 2.1 that the EM algorithm could converge to totally different results under different starting points. This uncertainty in the performance of the EM algorithm introduced by the starting point has to be resolved as dependency on the initial conditions is undesirable for an identification algorithm.

Among various initializing strategies, pre-running of the EM algorithm for several times with randomly selected starting points before entering the main EM step is most widely used owing to its simplicity and efficiency. After taking h trials for pre-running, the one with the largest likelihood would be chosen as the initial condition for the main EM algorithm. To balance the computation time spent on the initialization step and goodness of the initial condition obtained, the stopping

condition has to be appropriately defined (Biernacki et al. (2003)):

$$\frac{L(C, \Theta)^m - L(C, \Theta)^{m-1}}{L(C, \Theta)^m - L(C, \Theta)^1} \leq l \quad (2.18)$$

where $L(C, \Theta)$ denotes the complete data likelihood and l is a stopping value which can be tuned.

2.4.2 Refinement of Data Classification Result

A classification refinement procedure was proposed in order to deal with ‘un-decidable’ data points by Bempoard et al(2005). It uses bounded error along with spatial location information of ‘un-decidable’ data points to achieve data reclassification. However, given tens of thousands of data points, we found that it is computationally formidable by searching through every data point instead of focusing on only those ‘un-decidable’ data points. Hence, in this chapter, ‘certainty’ of each data point is defined based on the probability value $P_{k,i}$, $k = 1, 2 \dots N$, $i = 1, 2 \dots M$ obtained in the EM algorithm and ‘un-decidable’ data points are determined in terms of their ‘certainty’ level.

Data clustering in the EM algorithm is realized by comparing $P_{k,i}$, $i = 1, 2 \dots M$ for the k th data and finding the largest $P_{k,i}$ among all possible modes.

$$mode_k = arg \max_i P_{k,i}, k = 1, 2, \dots N, i = 1, 2, \dots M \quad (2.19)$$

Through Equation 2.19, the mode that each data point belongs to can be determined.

However, the data point such as those displayed in Figure 2.2, which lie in the greyed zone, may be classified to either of the two sub-models, and most of time, the classification result can be quite random as it is greatly influenced by the noise level. Hence for the data points lying in the interaction area among several sub-models, they are normally ‘un-decidable’ and misclassification can happen with a great chance.

Therefore, a refinement is needed to deal with those ‘un-decidable’ data points. Here, we propose to use ‘certainty’ level of each data point to denote the probability that a data point is ‘un-decidable’. The lower the ‘certainty’ is, the more likely the data point is ‘un-decidable’. The certainty of the k th data point is defined as,

$$Certainty(Z_k) = \left| \log \frac{P_{k,i_{min}}}{P_{k,j_{min}}} \right| \quad (2.20)$$

$$(i_{min}, j_{min}) = arg \min_{i,j} \left| \log \frac{P_{k,i}}{P_{k,j}} \right|, i = 1, 2 \dots M - 1 \quad j = i + 1, \dots M \quad (2.21)$$

In Equation 2.21, the probability of a data point belonging to each sub-model is compared and a pair of sub-models $\{i_{min}, j_{min}\}$ with closest probabilities are picked.

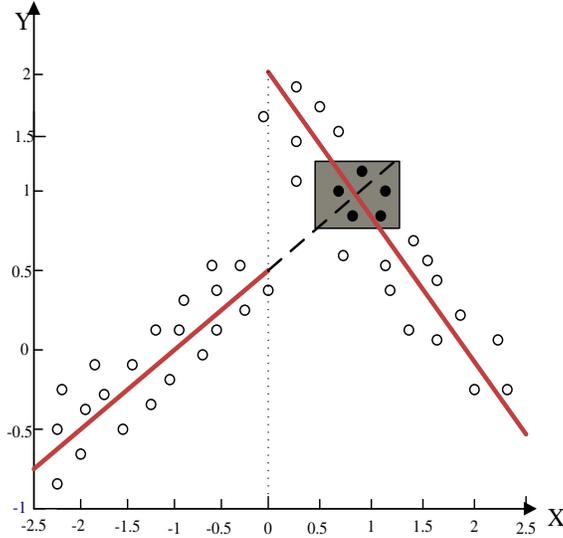


Figure 2.2: An explanation figure for ‘un-decidable’ data points generated by the bimodal PWARX system in Example 1

When probability between this couple of sub-models is equal, which implies that the data can be attributed to either sub-model, the ‘certainty’ value obtained through Equation (Equation 2.20) will be zero. Hence, in this way, ‘un-decidable’ data points can be sorted out effectively using this ‘certainty’ metric so that in the following classification refinement procedure, only those ‘un-decidable’ data with low ‘certainty’ levels are considered. The procedure for the refinement of ‘un-decidable’ data points classification result is:

Step1: Check ‘certainty’ level of the k th data point, if it is lower than some specified level lb , then go to step 2; otherwise check $(k + 1)$ th data until it reaches N .

Step2: Collect p data points that are closest to the k th data point in terms of the Euclidean distance.

Step3: For these $(p + 1)$ data points, label them as $q = 1, 2 \dots p + 1$, the probabilities $P_{q,i}$ of those data points classified to the same i th sub-model are added up, and the sub-model with the highest sum is taken as the new sub-model identity for this k th data point. This is based on the similarity principle. Namely, a data point that has low certainty value is classified to the same sub-model as its closest decidable neighbors tend to belong.

Step4: If $k < N$, go back to step 1; otherwise stop and exit the refinement procedure.

As an illustration, set $lb = 2$ and $p = 0.1 * N$, the refinement procedure is applied to Example 1 to show its effectiveness in correcting the misclassified ‘un-decidable’ data.

As we can see from Figure 2.3, some misclassifications have occurred in the in-

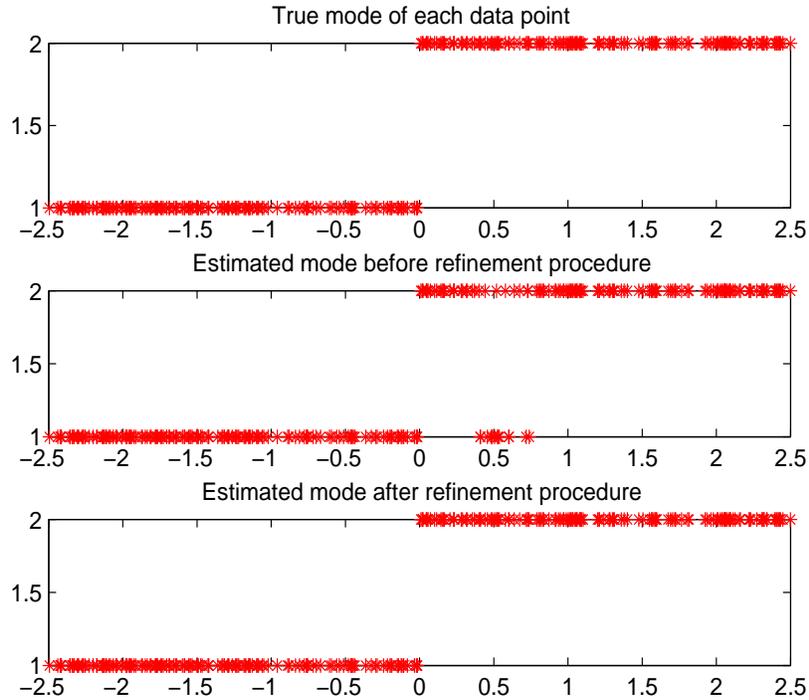


Figure 2.3: Comparison of data classification results before and after refinement procedure for Example 1

tersection area between two sub-models, and the refinement procedure can reclassify those misclassified data to the correct mode by exploring their spatial information.

2.5 Robust Parameter Re-estimation of Local Models and Region Partition

2.5.1 Robust Parameter Re-estimation of Each Local ARX Model

The reclassification procedure introduced above can effectively reduce the chance of misclassification for the 'un-decidable' data points located in the intersection areas of different sub-models by exploring the spatial information of those data points. This provides us an opportunity that by applying linear regression techniques such as ordinary least squares to each reclassified cluster, more accurate parameter estimation results could be obtained. However, if the data set of a local ARX model contains outliers, as pointed out earlier in the robust EM algorithm section, ordinary least squares can suffer. These outliers mainly come from two sources:

1. Measurements of process variables may be corrupted by occasional outliers such as noise spikes that come along with normal noise. These data may be clustered in

a correct group but they deviate significantly from the rest of data within the same group.

2. Occurrence of misclassification which mistakenly assigns data points to sub-models other than the one they truly belong to.

Therefore, it is necessary to employ another robust parameter estimation method so as to immunize the parameter estimation procedure from the outliers in each clustered data set. Here, once again, contaminated Gaussian distribution is utilized as the assumption for the residual errors and it is expected that by granting small weights to the abnormal data points, the identified local ARX models can better represent the dynamics of the system.

Let $Z_{n,i} = \{x_{n,i}, y_{n,i}\}$, $i = 1, 2 \dots M$, $n = 1, 2 \dots N_i$, where N_i represents the number of data points assigned to the i th sub-model. Assume that the sampling instant of $Z_{n,i}$ is m_n , following the same approach as being introduced in section 2, the likelihood function would be:

$$\begin{aligned} J_i &= \prod_{n=1}^{N_i} P(Z_{m_n} | \theta_i, Z_{m_n-1}, Z_{m_n-2} \dots Z_1) \\ &= \prod_{n=1}^{N_i} P(x_{m_n}, y_{m_n} | \theta_i, (x_{m_n-1}, y_{m_n-1}) \dots (x_1, y_1)) \end{aligned} \quad (2.22)$$

For an ARX model, we have

$$y_{n,i} = \theta_i^T \begin{bmatrix} x_{n,i} \\ 1 \end{bmatrix} + e_{n,i}, \quad i = 1, 2 \dots M, n = 1, 2 \dots N_i \quad (2.23)$$

Considering that the noise $e_{n,i}$ follows the contaminated Gaussian distribution which consists of regular noise $e_{regular}$ and irregular noise $e_{outlier}$, following the same way as having been done in the M-step of the robust EM algorithm, the expression for θ_i is:

$$\theta_i = \frac{\sum_{n=1}^{N_i} w_{n,i} \begin{bmatrix} x_{n,i} \\ 1 \end{bmatrix} y_{n,i}}{\sum_{n=1}^{N_i} w_{n,i} \begin{bmatrix} x_{n,i} \\ 1 \end{bmatrix} \begin{bmatrix} x_{n,i} \\ 1 \end{bmatrix}^T} = (X_i^T W_i X_i)^{-1} X_i^T W_i Y_i \quad (2.24)$$

where $w_{n,i} = \frac{m \frac{P(e_{n,i}^{regular})}{\sigma^2} + (1-m) \frac{P(e_{n,i}^{outlier})}{d^2 \sigma^2}}{mP(e_{n,i}^{regular}) + (1-m)P(e_{n,i}^{outlier})}$, $X_i = \begin{bmatrix} x_{1,i} & 1 \\ \vdots & \vdots \\ x_{N_i,1} & 1 \end{bmatrix}_{N_i,2}$, $Y_i = \begin{bmatrix} y_{1,i} \\ \vdots \\ y_{N_i,i} \end{bmatrix}_{N_i,1}$, $W_i =$

$\begin{bmatrix} w_{1,i} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{N_i,i} \end{bmatrix}_{N_i, N_i}$. If error $e_{n,i}$ satisfies Tjoa & Biegler (1991)

$$mP(e_{n,i}^{regular}) > (1-m)P(e_{n,i}^{outlier}) \quad (2.25)$$

$Z_{n,i}$ can be treated as a normal data point and a relatively high weight $w_{n,i}$

2.5.2 Region Partition

In addition to data clustering and local ARX model parameter estimation, polyhedral region $\{\chi\}_{i=1}^M$ in which each sub-model takes effect should also be estimated in order to complete the identification of a PWARX system. Here, we use the MRLP algorithm which is introduced by Bennett et al (1994). The rationale behind this method is that for the data point $Z_{n,i} = \{x_{n,i}, y_{n,i}\}$, $n = 1, 2, \dots, N_i$ that belongs to region i , the following Equation always holds (Bennett & Mangasarian (1994)):

$$x_{n,i}(\omega_i - \omega_j) > \gamma_i - \gamma_j, i, j = 1, 2 \dots M, i \neq j \quad (2.26)$$

or equivalently

$$x_{n,i}(\omega_i - \omega_j) \geq \gamma_i - \gamma_j + 1, i, j = 1, 2 \dots M, i \neq j \quad (2.27)$$

where $\omega_i, \omega_j, \gamma_i$ and γ_j are parameters of hyperplane that separates region χ_i and χ_j . The hyperplane that separates χ_i and χ_j satisfies $x_{n,i}(\omega_i - \omega_j) = \gamma_i - \gamma_j$. Define

$$G_n^{i,j} = \max(-x_{n,i}(\omega_i - \omega_j) + (\gamma_i - \gamma_j) + 1, 0) \quad (2.28)$$

Then a linear objective function can be defined as:

$$J_{rp} = \min_{\omega_i, \gamma_i} \sum_{i=1}^M \sum_{j=1, i \neq j}^M \sum_{n=1}^{N_i} G_n^{i,j} \quad (2.29)$$

where N_i denotes the number of data points classified to the i th sub-model. Normally, $G_n^{i,j}$ should always be zero when Equation 2.29 holds. However, if misclassification of data occurs, Equation 2.28 or 2.29 will be violated, which means that $G_n^{i,j}$ is larger than zero. As a result, the accuracy of region partition results from linear minimization Equation 2.29 could be reduced because of those misclassified data points.

As discussed in the robust parameter estimation section, weights for data points that are suspicious of being outliers are small or even equal to zero. These weights can be used to weight each data point in Equation 2.29. Therefore, a new objective equation would be:

$$J_{rp} = \min_{\omega_i, \gamma_i} \sum_{i=1}^M \sum_{j=1, i \neq j}^M \sum_{n=1}^{N_i} w_{n,i} G_n^{i,j} \quad (2.30)$$

To summarize, the complete framework of the algorithm introduced in this chapter is shown as Figure 2.4,

Returning to Example 1, the initial variance σ_{ini}^2 is arbitrarily chosen as 0.04. Based on the recommendation given in Biernacki et al.(2003), the stopping condition l and pre-running times h are set to be 0.01 and 10 respectively. For lb and p , we

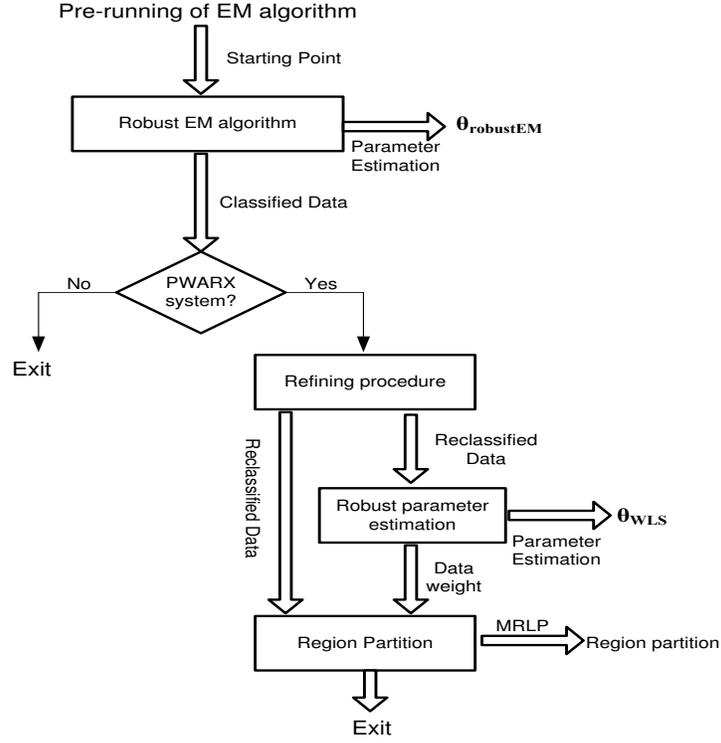


Figure 2.4: Framework of robust PWARX system identification using EM algorithm

Table 2.2: True bimodal PWARX system parameters and estimated parameters

θ_1	$\theta_{1regularEM}$	$\theta_{1robustEM}$	θ_{1WLS}	θ_2	$\theta_{2regularEM}$	$\theta_{2robustEM}$	θ_{2WLS}
-1	-0.9845	-0.9886	-0.9886	0.5	0.5047	0.5115	0.5115
2	2.0418	2.0392	2.0392	0.5	0.5435	0.5541	0.5540

find that, in general, the un-decidable data points can be effectively detected and classified with the choice of $lb = 2$ and $p = 0.1 * N$. Applying the proposed algorithm to Example 1, we obtain the results shown in Table 2.2.

$\theta_{regularEM}$ and $\theta_{robustEM}$ in Table 2.2 are the parameter estimation results from the regular EM algorithm, robust EM algorithm respectively. θ_{WLS} is obtained by applying the robust parameter estimation procedure to the reclassified data clusters. The final estimation of variance $\hat{\sigma}^2 = 0.0278$ while true data variance equals $\sigma^2 = 0.025$. As no outliers are added in the data set, the estimation result from the regular EM algorithm is similar to its robust counterpart as well as robust parameter estimation procedure. The region partition result is $x = -0.00258$.

2.6 A More Complex Simulation Example

Consider the following PWARX system with 3 sub-models (Bempeard et al. (2005); Nakada et al. (2005)):

$$y_k = \begin{cases} [-0.4 & 1 & 1.5] \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, & x_k \in \chi_1 \\ [0.5 & -1 & -0.5] \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, & x_k \in \chi_2 \\ [-0.3 & 0.5 & -1.7] \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, & x_k \in \chi_3 \end{cases}, k = 1, 2 \dots N \quad (2.31)$$

where the regressor $x_k = \begin{bmatrix} y_{(k-1)} \\ u_{(k-1)} \end{bmatrix}$. Region partition is given by

$$\begin{aligned} \chi_1 &= \{[4 \quad -1]x + 10 < 0\} \\ \chi_2 &= \left\{ \begin{array}{l} [-4 \quad 1]x - 10 \leq 0 \\ [5 \quad -1]x - 6 \leq 0 \end{array} \right\} \\ \chi_3 &= \{[-5 \quad -1]x + 6 \leq 0\} \end{aligned}$$

Hence, the hyperplane of the PWARX system is shown in Figure 2.5.

In order to test the capability of the algorithm in handling outliers, certain percentage of outliers are added to a data set generated by the PWARX system. Set total number of data points $N = 300$ in which around 10 percents are outliers. Regular Normal distributed noise $e_k \sim N(0, 0.05)$ is chosen while initial guess of noise variance σ_{ini}^2 is arbitrarily set 0.03. The input u_k follows a uniform distribution between $[-5 \quad 5]$. Let $lb = 2$, $p = 0.1 * N$, $l = 0.01$, $h = 10$, $m = 0.9$ and $d^2 = 60$, then apply the algorithm to the data set and data classification result is shown in Figure 2.6.

In Figure 2.6, we can see that before the refinement procedure, some misclassification occurs at the intersection of different sub-models (those ‘floating’ circle

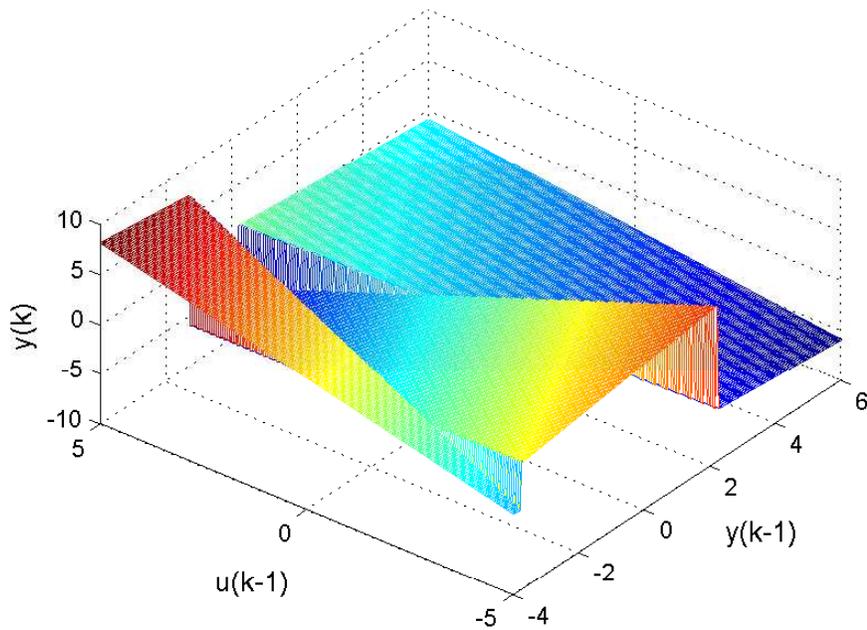


Figure 2.5: Hyperplane of the PWARX system

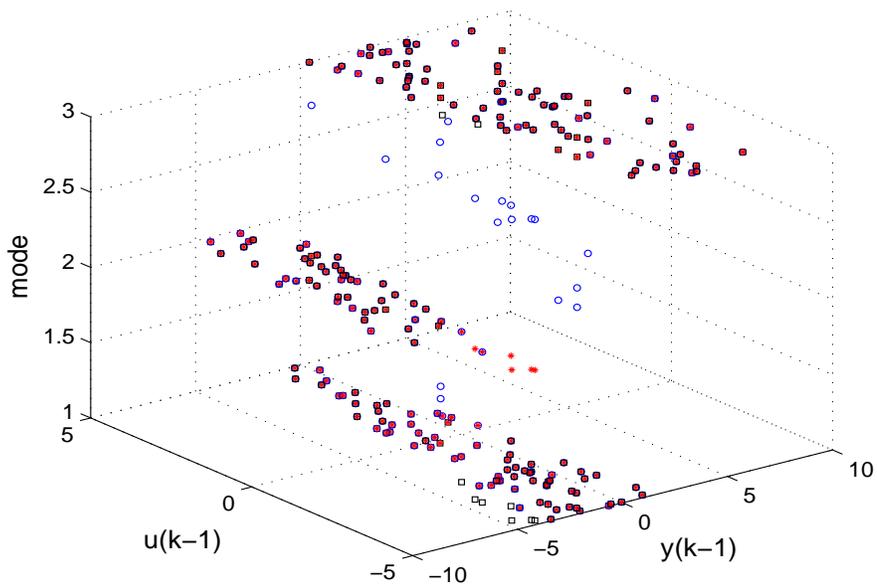


Figure 2.6: Comparison of data classification result before and after refinement procedure (Cross points denote the true mode of each data point, circle points represent classification results before refinement and square ones are the results after classification refinement)

Table 2.3: Comparison of estimated parameters using different methods with outliers in the data set

θ_1	$\theta_{1regularEM}$	$\theta_{1robustEM}$	θ_{1WLS}	θ_2	$\theta_{2regularEM}$	$\theta_{2robustEM}$	θ_{2WLS}
-0.4	-0.538	-0.516	-0.405	0.5	0.3	0.442	0.454
1	1.028	1.021	1.017	-1	-0.988	-1.0	-0.997
1.5	0.904	0.972	1.397	-0.5	-0.511	-0.524	-0.514
<hr/>							
	θ_3	$\theta_{3regularEM}$	$\theta_{3robustEM}$	θ_{3WLS}			
	-0.3	-0.229	-0.302	-0.296			
	0.5	0.508	0.486	0.498			
	-1.7	-2.049	-1.657	-1.687			

points) and it is partially eliminated after the refinement procedure. Parameter estimation results from different estimation methods are listed in Table 2.3.

where $\theta_i, i = 1, 2, 3$ are the real parameters of its corresponding sub-model respectively. The subscript regularEM stands for the parameters estimated from the regular EM algorithm; subscript robustEM represents the parameter estimation results obtained from the robust EM algorithm; θ_{WLS} denotes the identification results after applying the robust parameter estimation procedure to the reclassified data points. As can be seen in Table 2.3, the estimation from the regular EM algorithm for some parameters is greatly influenced by outliers while, on the other hand, the robust EM algorithm $\theta_{robustEM}$ and the robust parameter estimation θ_{WLS} exhibit sufficient robustness to resist the negative influence brought by outliers. However, from the estimation results of θ_1 , in which robust parameter estimation procedure outperforms the robust EM algorithm in terms of parameter estimation accuracy, it can be seen that the classification refinement procedure can effectively improve the performance of the parameter estimation.

Finally, transfer the weights obtained in robust parameter estimation to the region estimation procedure, the regressor partition results are obtained as:

The estimated hyperplane between χ_1 and χ_2 is $\hat{\lambda}_{12}x + 1 = 0$, $\hat{\lambda}_{12} = [0.33 - 0.1087]$ while χ_2 and χ_3 are separated by the hyperplane $\hat{\lambda}_{23}x + 1 = 0$, $\hat{\lambda}_{23} = [-0.823 - 0.152]$. It is noticed from Equation 2.31 that the true hyperplane that separates χ_1 and χ_2 is $\lambda_{12}x + 1 = 0$, $\lambda_{12} = [0.4 - 0.1]$, and χ_2, χ_3 are separated by $\lambda_{23}x + 1 = 0$, $\lambda_{23} = [-0.83 - 0.167]$.

2.7 Application Example: Continuous Fermentation Reactor

Discontinuous behavior of the chemical process may originate from its inherent physicochemical discontinuities (such as phase changes, flow reversals, and shocks) or some interventions/disturbances from outside world (such as human or controller

manipulation, changes in raw materials). These discrete events drive the system traveling among various operating regimes featured by different dynamics. In optimal control of plant grade transition, the governing process dynamics during the plant transition period can vary dramatically and as a result, the performance of the model based controller may deteriorate greatly if only one single local model is utilized. Hence, being able to detect the changes of process dynamics and identify the relevant local models under different operating regimes could be critical in some applications.

To illustrate potential applications of the proposed identification method for chemical process, a continuous fermentation reactor is given in this section (Henson & Seborg (1992a); Gugaliya et al. (2005); Venkat et al. (2003)). The fermenter consists of two inputs and three outputs, namely feed substrate concentration (u_1), dilution rate (u_2) and biomass concentration (y_1), substrate concentration (y_2), production concentration (y_3). Following the same parameter setting and normal operating conditions given by Gugaliya et al. (2005), a single input, single output (SISO) model between dilution rate (u_2) and biomass concentration (y_1) is identified. To ensure the local linearity of the identified model, a random binary signal with appropriate amplitude (from $0.1636h^{-1}$ to $0.19h^{-1}$) is designed for the input u_2 . Assume that the feed substrate concentration u_1 fluctuates between its nominal value $20kg/m^3$ and $26.5kg/m^3$ and cannot be tracked in a timely manner, the process dynamics under different u_1 values are compared and the step responses are shown in Figure 3.4

As can be seen from Figure 3.4, the dynamics of the model under different feed substrate concentration values exhibit different characteristics, including process gain, time constant as well as steady state values. This indicates the necessity of detecting and identifying the switching process model in order to achieve satisfactory control performance. Let the feed substrate concentration change between $20kg/m^3$ and $26.5kg/m^3$ randomly, adding white noise which is about 1% of the noise-free output in power. The obtained output and input data are shown in Figure 2.8

In this case, the switching frequency of the input signal is determined based on the minimum time constant of all the sub-models. Distinct dynamics coexist in the output data set shown in Figure 2.8 owing to the change of feed substrate concentration. Passing the data set through the proposed algorithm, the identified models as well as self-validation results are shown in Figure 2.4 and Figure 3.7.

Due to the discrepancy of process behavior under different feed substrate concentration values ($20kg/m^3$ and $26.5kg/m^3$ in this case), we are able to classify the data point at each instance to its clusters featured with different u_2 values. As a result, we can detect the grade change of the raw materials by analyzing and clustering the data. For the data set shown in Figure 2.8, 80% of the data points have been

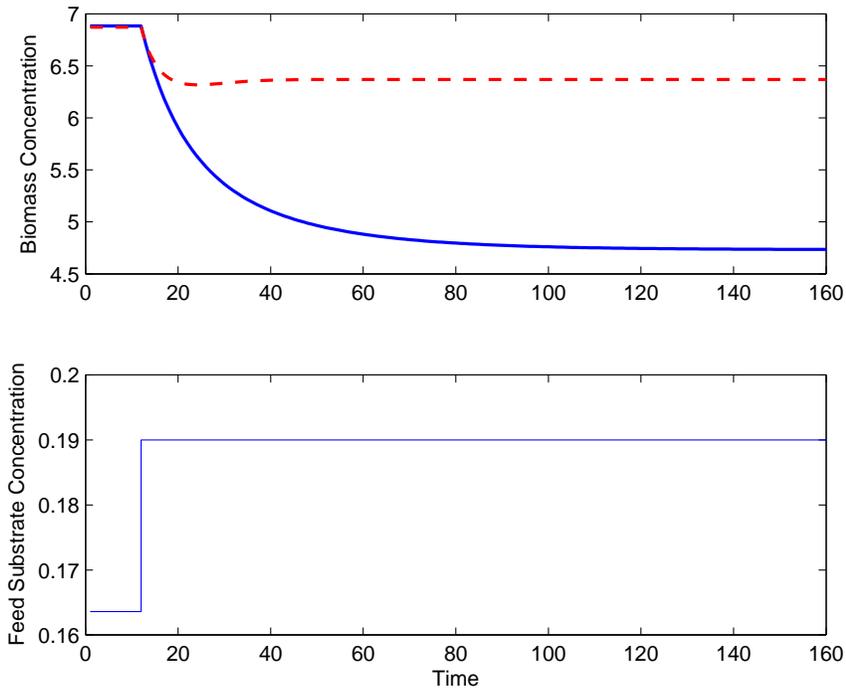


Figure 2.7: Input-Output data of the continuous fermenter. Dash line represents the step response when $u_1 = 20$, solid line denotes the step response when $u_1 = 26.5$

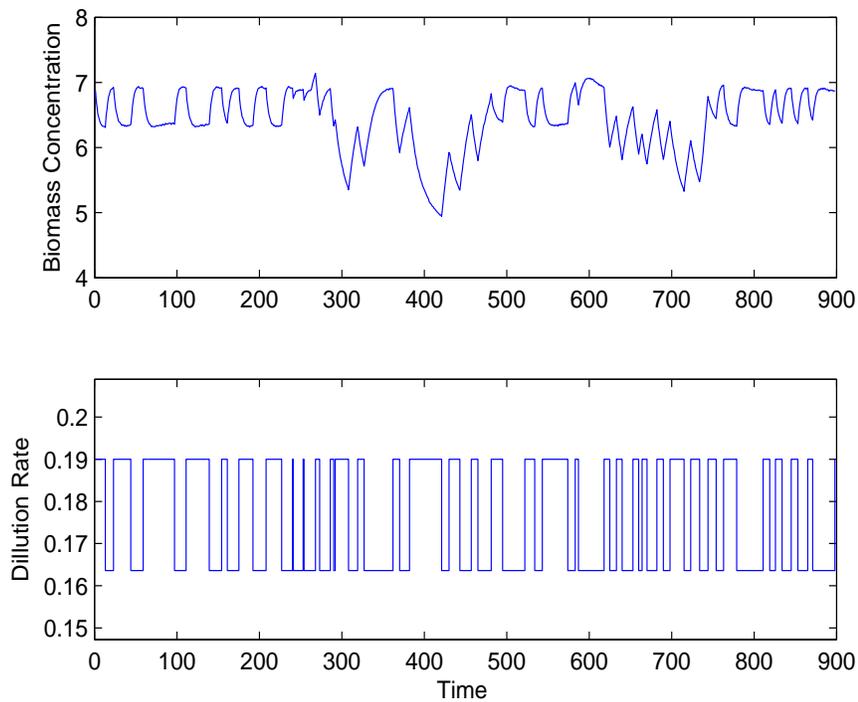


Figure 2.8: Input-Output data of the continuous fermenter

Table 2.4: Identified models under different feed substrate concentrations	
<i>Feed substrate concentration</i>	<i>Identified model from u_2 to y_1</i>
$20\text{kg}/\text{m}^3$	$y_k = 1.6550y_{k-1} - 0.6842y_{k-2} - 6.4345u_{k-1} + 6.0230u_{k-2} + 0.2676$
$26.5\text{kg}/\text{m}^3$	$y_k = 1.9010y_{k-1} - 0.9034y_{k-2} - 5.7619u_{k-1} + 5.6314u_{k-2} + 0.0374$

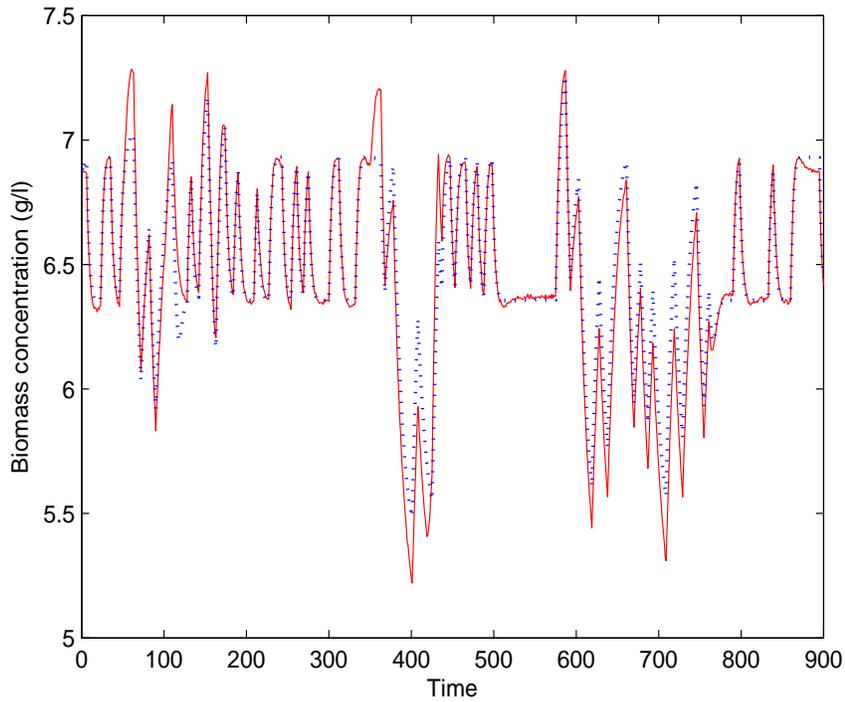


Figure 2.9: Comparison between model prediction and the true data, self validation MSE=0.0155. Dotted line represents prediction from identified models, Solid line denotes the true data from the fermenter.

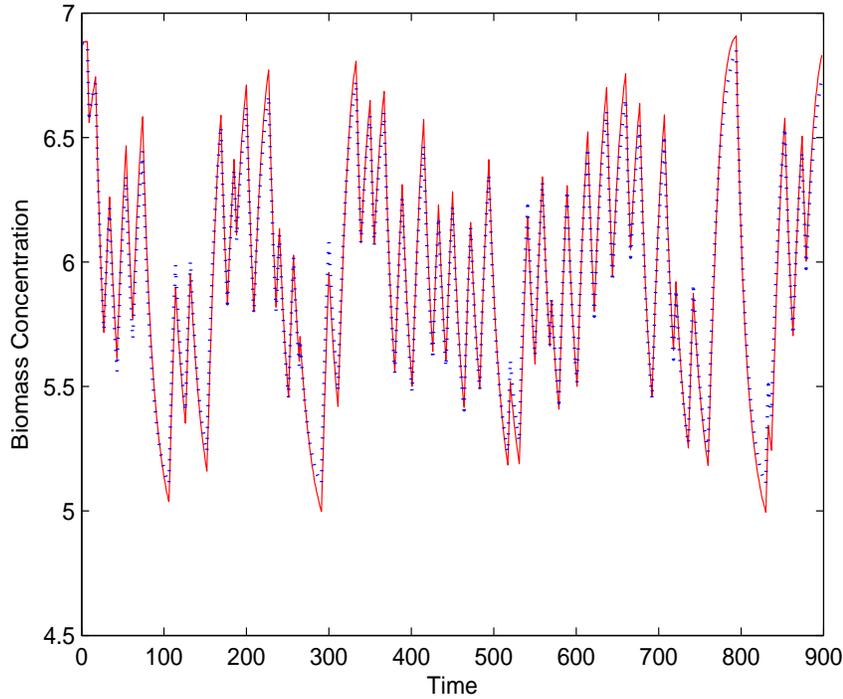


Figure 2.10: Comparison between model prediction and the true data when $u_2 = 26.5 \text{ kg/m}^3$, cross validation MSE=0.008. Dashed line represents prediction from identified models, Solid line denotes the true data from the fermenter.

classified to the right mode, which equally means that we may be able to indirectly infer the value of feed substrate concentration correctly during 80% of the operating time.

To further validate the identified models, cross validation for both identified models is performed using newly generated data set under different u_2 values. The validation results are given in Figure 2.10 and Figure 2.11.

Satisfactory cross validation results demonstrate the capability of the proposed identification algorithm in handling the process which exhibits various distinct dynamics owing to the influences from diverse sources. Detection of the process change not only facilitates the monitoring of the process operation by informing us with happening of dramatic change in operating condition (feed substrate concentration change in this case), but also provides us with deeper perspective on the process itself during the operation (such as whether phase change happens). Moreover, multiple local models identified through the algorithm can also be used by model-based optimal controllers so that decent control performance could always be achieved regardless of the switching of process dynamics.

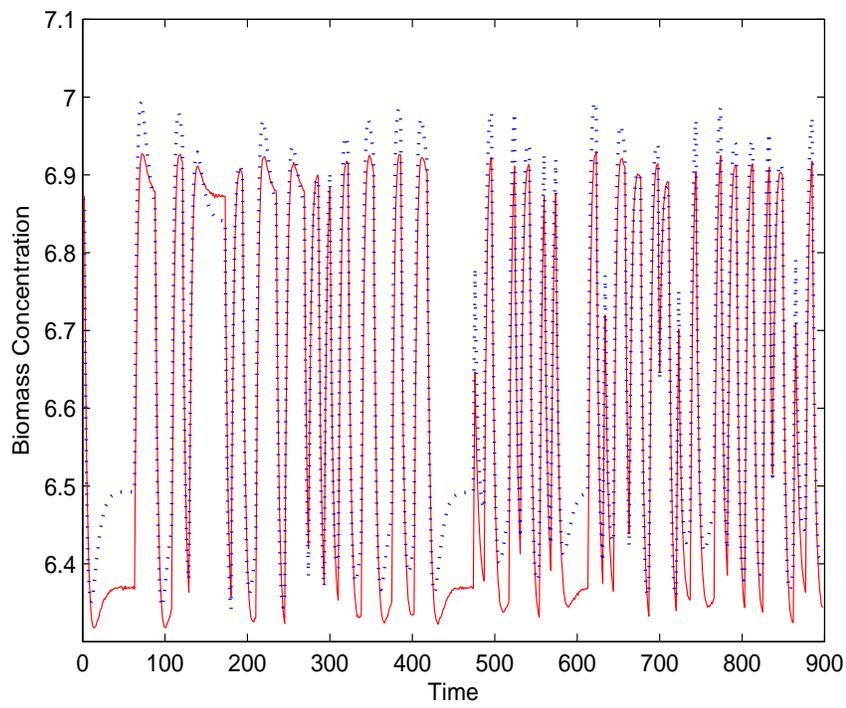


Figure 2.11: Comparison between model prediction and the true data when $u_2 = 20kg/m^3$, cross validation $MSE=0.00366$. Dashed line represents prediction from identified models, Solid line denotes the true data from the fermenter.

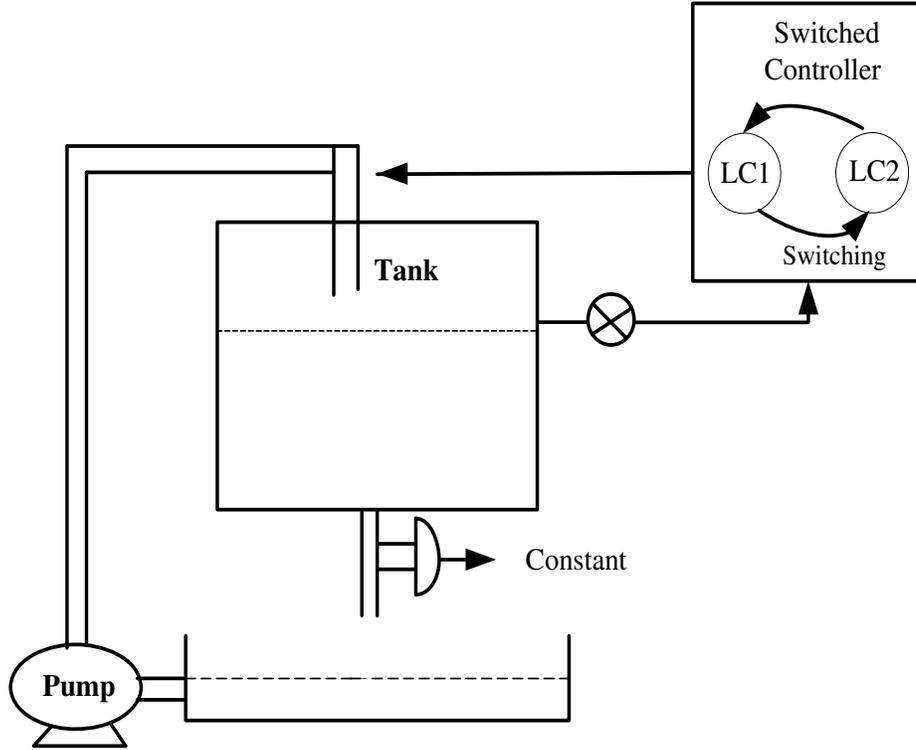


Figure 2.12: Schematic diagram of the tank system with switched controller

2.8 Identification of An Experimental Switched Process Control System

As a typical form of switched linear systems, the PWARX system switches because its input and output come to different domains as mathematically formulated in Equation 2.1. Provided that no *prior* knowledge regarding the triggering signal of the switched system is at hand, time could be considered responsible for the switching of the system. This grants us a general way of treating the switched linear system and the proposed PWARX system identification algorithm can be applied to such a switched system without need of further modification.

The tank system is a pilot scale setup on which two switching level controllers with different characteristics are installed in an effort to maintain the level of the tank. The schematic diagram is shown in Figure 2.14:

By controlling the speed of the pump, the level controller can manipulate the inlet water flow rate of the tank so as to control its level. The closed loop step response under each of the two level controllers is given in Figure 2.13.

It can be seen that both controllers do not have any offset in tracking the setpoint, which equally means that the steady state gains of the control loop under both controllers are one. However, apparent difference is observed in terms of their transient response to the setpoint change; in other words, the dynamics of the closed

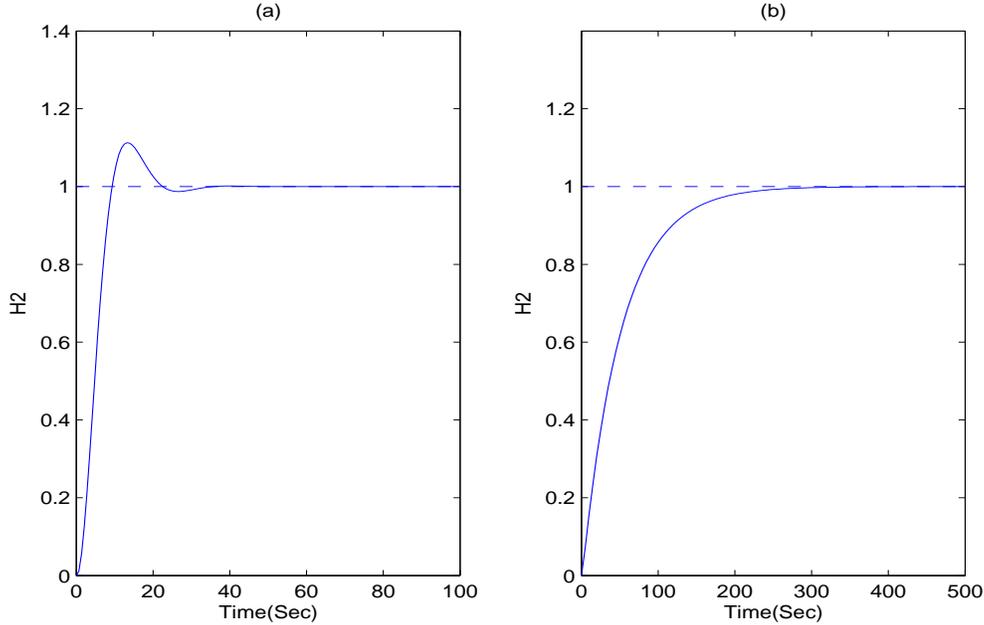


Figure 2.13: (a) Step response of controller 1 (b) step response of controller 2

loop process under different controllers are different. Hence, the control loop operates under two modes:

Mode1: The loop is running on level controller 1, which enables the controller to respond aggressively to any setpoint change or unknown disturbance with certain amount of overshoot.

Mode2: The loop is running on the sluggish level controller 2, which makes the controller act slowly in setpoint tracking or disturbance rejection. There is no overshoot under level controller 2.

For each time point, the switched controller may randomly reside on one of the two level controllers with certain probability, resulting two-mode closed-loop operations. Without knowing how the controller is switched internally, it is expected that the proposed system identification algorithm can separate the bimodal process with different dynamics and flag the identity of the controller on which the system is running at each sampling instance in the presence of the disturbances.

As shown in Figure 2.14, the setpoint of the closed loop is treated as the input of the process and the water level of the tank is considered as the output, the proposed identification method can be performed after collecting the input-output data of the closed-loop process. Cases with/without outliers in the experimental data are considered. According to the algorithm framework shown in Figure 2.4, the robust parameter estimation for each local clustered data set θ_{WLS} is calculated only when the refinement of the classification result from the robust EM algorithm is achieved. Due to the fact that the (time) switched system does not have input/output de-

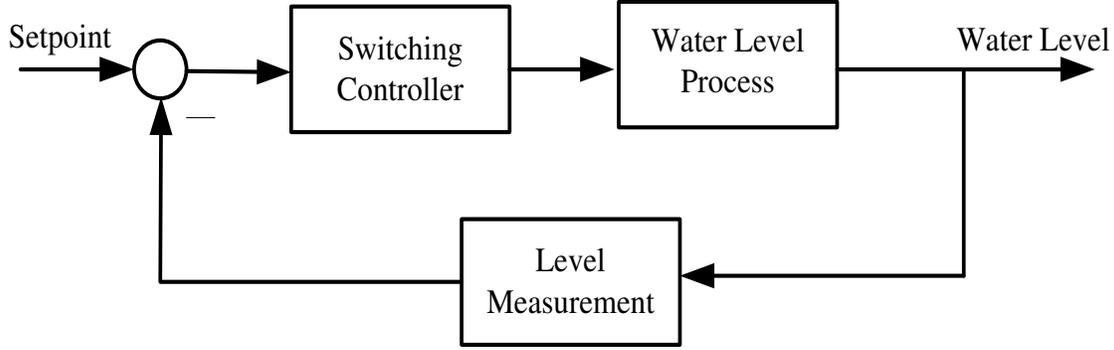


Figure 2.14: Block diagram of the switched control system

pendent partition, the region partition procedure will not be performed. Only the identified models from the regular EM algorithm $\theta_{regularEM}$ and the robust EM algorithm $\theta_{robustEM}$ are compared to demonstrate the difference between robust and regular (non-robust) algorithm, which is the focusing point of this chapter.

2.8.1 Case 1: Experimental Data Without Outliers

The model order for both sub-models of the bimodal closed-loop process can be deduced from the analysis of the process based on the physical knowledge of the system, the number of sub-models for the multi-model identification algorithm is known as 2. After running an experiment, the collected input-output data shown in Figure 2.15 are separated into two parts, one is for estimation of each local ARX model parameters while the other one is used for cross validating the identified models obtained from training data set.

The hidden variable that represents the identity of the controller mode on which the system is running at each sampling time point needs to be estimated and the accuracy of its estimation directly influences the quality of the identified models. Given the estimation of the hidden variable \hat{I}_k for the k th sampling time point, each data point can be clustered into one of the two groups.

After passing the training data set through the algorithm, it is noticed that the identified switched models from the regular EM algorithm and its robust counterpart are similar. Figure 2.16 shows the estimated hidden variable \hat{I}_k as well as the comparison between the simulated response of the identified models and the actual experimental data.

The identified closed-loop models under different level controllers are:

$$\mathbf{Mode1} : y_k = 1.1430y_{k-1} - 0.4346y_{k-2} + 0.0572u_{k-1} + 0.2415u_{k-2} \quad (2.32)$$

$$\mathbf{Mode2} : y_k = 0.9534y_{k-1} - 0.0475y_{k-2} + 0.0618u_{k-1} + 0.0336u_{k-2} \quad (2.33)$$

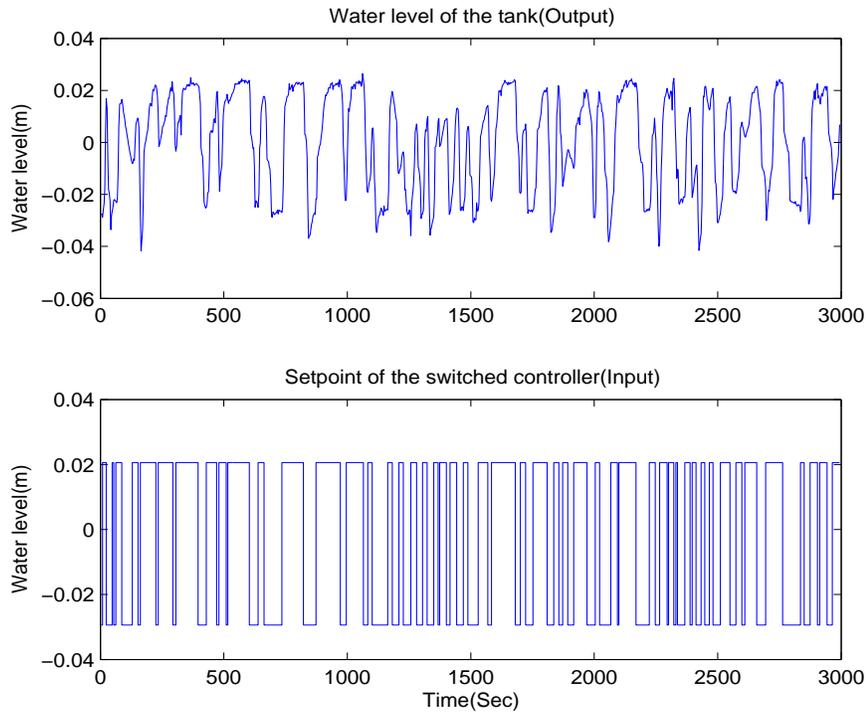


Figure 2.15: Input-output data of the switched control system

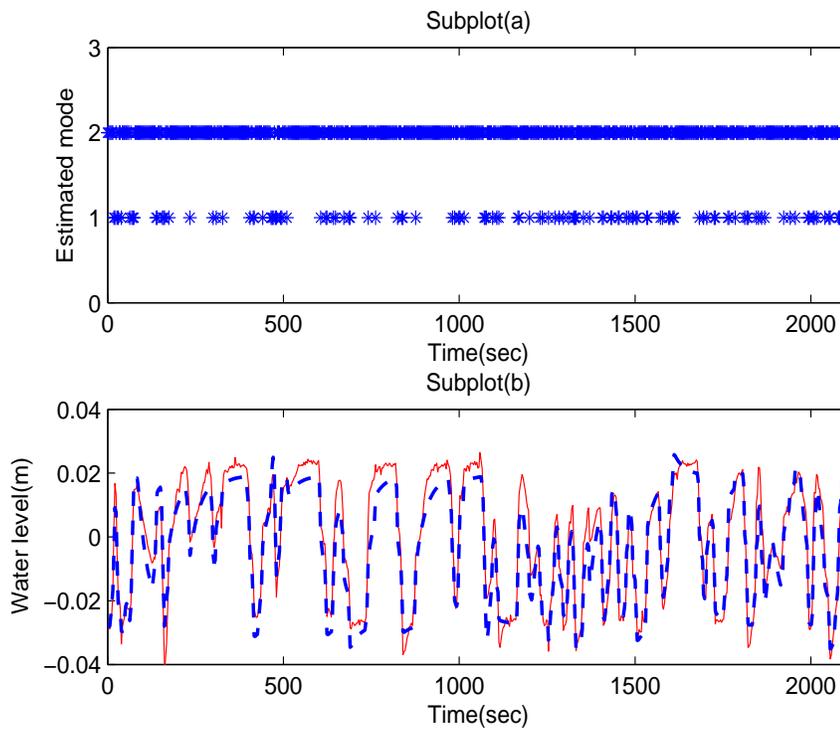


Figure 2.16: Subplot (a) denotes the estimated mode sequence of the switching evolution along the time for training data . Subplot (b) self-validation of the identified multi-ARX modes (solid line: measured water level data, dashed line: prediction from the identified hybrid model)

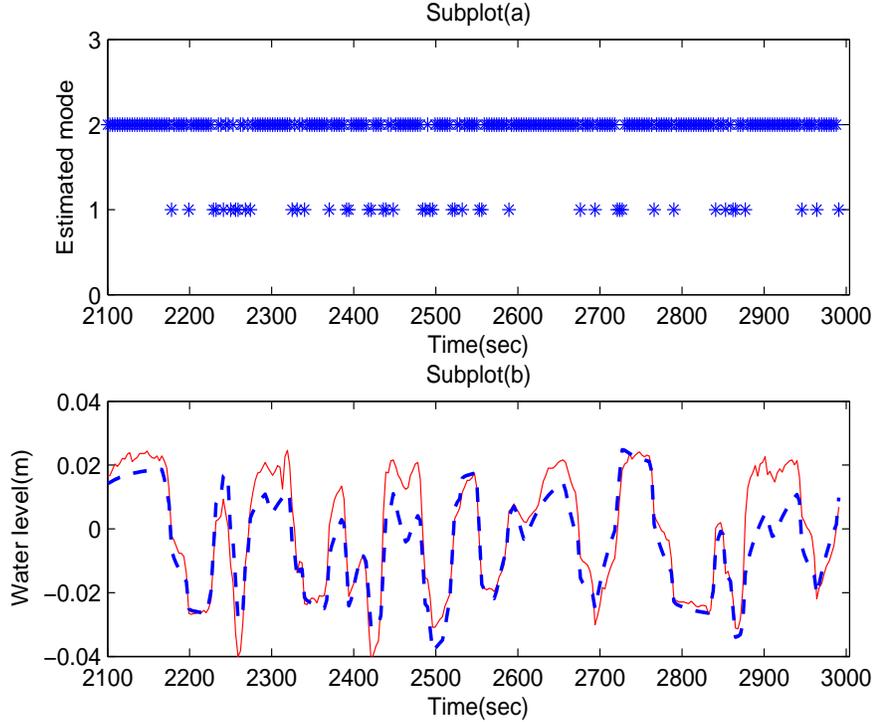


Figure 2.17: Subplot (a) denotes the estimated mode sequence of the switching evolution along the time for training data . Subplot (b) cross-validation of the identified multi-ARX modes (solid line: measured water level data, dashed line: prediction from the identified hybrid model)

As shown in Figure 2.16, even though a relatively high percentage fit is achieved in self-validation, cross-validation is still required to further test the validity of the identified models. Due to the switching time points are unknown in advance, unlike the conventional way of cross-validation, the mode identity of each data point in the cross-validation data set should be known before we are able to use the identified switched models to predict the system output. As a result, cross validation is performed through a two-step procedure. In the first step, validation data are clustered to estimate the model identity for each data point using the proposed clustering method, and then in the second step, the ARX models obtained from the identification data set are validated using the clustered data. The cross validation results of the multi-ARX models expressed in Equation 2.32 and 2.33 along with the validation data's estimated cluster identity are given in Figure 2.17.

Again, it is noticed from Figure 2.17 that the predictions from the identified models are close to the true system output, which implies that the identified sub-ARX models can effectively describe the dynamics of close-loop process under different controllers.

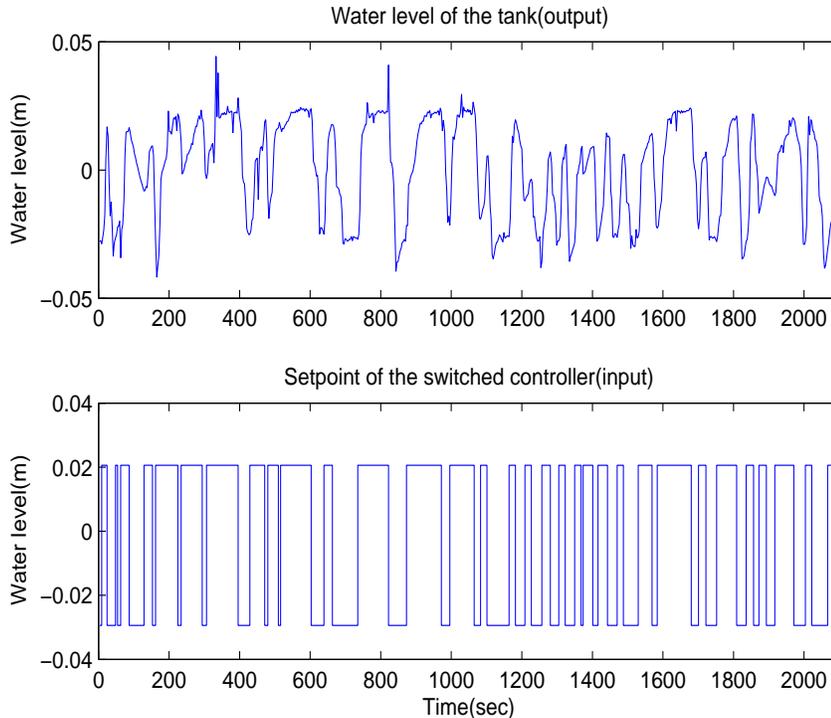


Figure 2.18: Input-output data of the switched control system with manually added outliers

2.8.2 Case 2: Experimental Data With Outliers

No apparently abnormal data points have been observed in the data set shown in Figure 2.15 and the identified models have been successfully validated. To test the robustness of the proposed identification algorithm, several outliers are randomly added to the experimental training data set and Figure 2.18 shows the new data set after outliers are added.

With the outliers it is expected that the performance of various identification procedures would degrade compared with the outlier-free case. As a result, the identified models described in Equation 2.32 and Equation 2.33 are treated as reference models against which the new models obtained from the robust EM algorithm, regular EM algorithm and the identification method introduced by Nakada et al.(2005). Here, the reasons we chose the method (Nakada et al. (2005)) as the comparative method against which the proposed robust EM algorithm is evaluated are because both of the methods use statistical theory as a way to classify the observed data. Moreover, the structures of the parameters identified from the two algorithms are quite similar which makes them more comparable. To simplify the expression, we will refer the method put forward by Nakada et al. (2005) as the comparative method hereafter.

To sufficiently investigate the performance of different identification procedures

Table 2.5: Comparison of identification results, SD stands for standard deviation

<i>Mode</i>	$\theta_{reference}$	$\theta_{robustEM}(Mean)$	$SD_{robustEM}$	$\theta_{normalEM}(Mean)$	$SD_{normalEM}$
mode 1	1.1430	1.0917	0.2627	0.5034	0.4210
	-0.4346	-0.4153	0.2015	0.0838	0.4715
	0.0572	0.0924	0.1979	0.0183	0.1834
	0.2415	0.2297	0.1972	0.4589	0.2008
mode 2	0.9534	0.9888	0.0466	0.9963	0.1442
	-0.0475	-0.0811	0.0541	-0.0999	0.1467
	0.0618	0.0645	0.0335	0.0576	0.0619
	0.0336	0.0285	0.0329	0.0461	0.0606

<i>Mode</i>	$\theta_{reference}$	$\theta_{comparison}(Mean)$	$SD_{comparison}$
mode 1	1.1430	0.4482	0.4107
	-0.4346	0.2229	0.4440
	0.0572	0.0173	0.1848
	0.2415	0.5404	0.1896
mode 2	0.9534	0.9862	0.1423
	-0.0475	-0.0894	0.1456
	0.0618	0.0559	0.0625
	0.0336	0.0477	0.0591

in the presence of outliers, Monte-Carlo simulation is performed. In each run of the simulation, the percentage of outliers is fixed as 5% and the location of those outliers are randomly determined. After 100 runs, the averaged parameters along with the standard deviation of the estimated parameters from each of the identification procedures are calculated. Table 2.5 gives the identified results:

In comparison with the reference models, it is found that the presence of the outliers greatly skews the identification results from both the comparative method and the regular EM algorithm while, on the other hand, the robust EM algorithm identification procedure renders identified models closer to the reference ones by effectively spotting the outliers and diminishing their negative influence on parameter estimation. To demonstrate the discrepancy of various identified models in a clearer way, step tests are performed respectively for the models listed in Table 2.5. It is found that 2 out of 100 sets of parameters obtained from the regular EM algorithm and the comparative method have poles outside of the unit circle, which can end up with unstable step response. These unstable models are removed before calculating the mean and standard deviation of the Monte-Carlo simulation for the regular EM algorithm and the comparative method.

The mean as well as the standard deviation of the step responses from the two sets of models identified from the EM algorithm with ($\theta_{robustEM}$) or without robust procedure ($\theta_{regularEM}$) together with the comparative method for mode 1 and 2 have been calculated. Figure 2.19 and 2.20 show the comparison results.

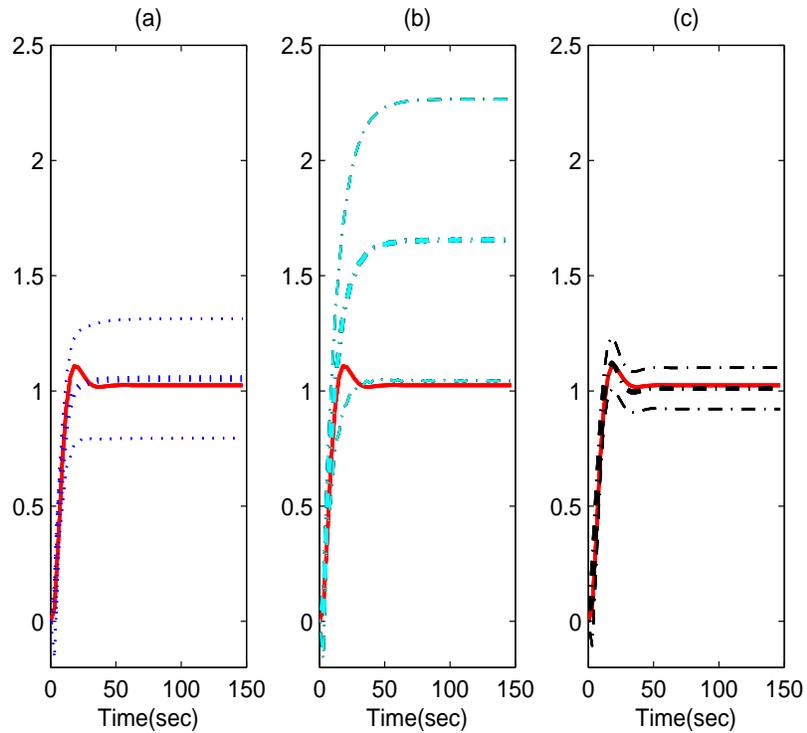


Figure 2.19: Comparison of the step response of the identified models for mode 1. (a) Step response of the regular EM algorithm identified model. Solid line: step response of the reference model, bold dotted line: mean step response of the identified models, dotted line: standard deviation bound of the step response from the identified models.(b)Step response of the comparative method identified model. Solid line: step response of the reference model, bold dash-dotted line: mean step response of the identified models, dash-dotted line: standard deviation bound of the step response from the identified models. (c) Step response of the robust EM algorithm identified model. Solid line: step response of the reference model, bold dash-dotted line: mean step response of the identified models, dash-dotted line: standard deviation bound of the step response from the identified models.

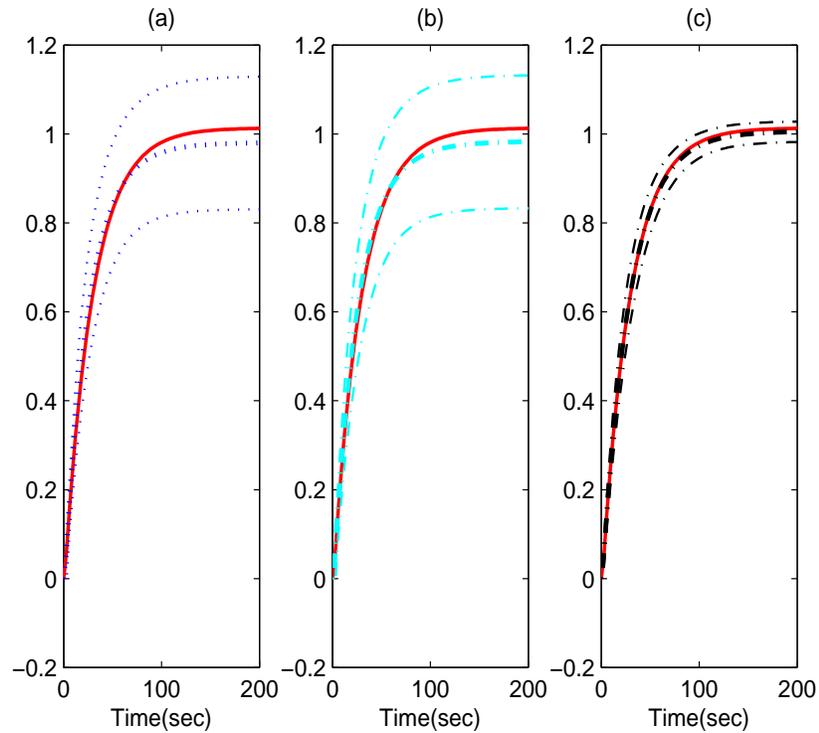


Figure 2.20: Comparison of the step response of the identified models for mode 2. (a) Step response of the regular EM algorithm identified model. Solid line: step response of the reference model, bold dotted line: mean step response of the identified models, dotted line: standard deviation bound of the step response from the identified models.(b)Step response of the comparative method identified model. Solid line: step response of the reference model, bold dash-dotted line: mean step response of the identified models, dash-dotted line: standard deviation bound of the step response from the identified models. (c) Step response of the robust EM algorithm identified model. Solid line: step response of the reference model, bold dash-dotted line: mean step response of the identified models, dash-dotted line: standard deviation bound of the step response from the identified models.

Not surprisingly, the model identified from the robust EM algorithm outperforms the models rendered by the other two methods. It gives more accurate estimation of the step response on average while, in the mean time, enjoys much smaller standard deviation bound.

2.9 Conclusion

A new approach for identification of PWARX systems is developed in this chapter. The EM algorithm is employed in data clustering as well as parameter estimation of the PWARX system. By expressing the noise distribution in a contaminated Gaussian distribution form, a robust EM algorithm is developed and its robustness to the outliers is demonstrated through simulations and pilot-scale experiments. It is shown that pre-running of the EM algorithm for several times with certain stopping condition can be an effective way to overcome its pitfall of sensitiveness to the starting point. Also, for those ‘un-decidable’ data points in the data clustering, a refinement procedure can classify them to their own clusters by using the information provided by their spatially closest data points. For PWARX systems, in case that outliers exist in the reclassified local data set, robust parameter estimation of each local ARX model is realized by employing the contaminated Gaussian distribution for residual errors. Region partition of PWARX systems is performed by using slightly modified MRLP algorithm in which data points are weighted based on their probability of being outliers. In this way, the influence of those abnormal data points could be minimized in the process of hyperplane determination. Finally, successful identification of simulated PWARX systems, a simulated continuous fermenter and a experimental switched control system is achieved by employing the proposed identification algorithm. The identification results confirm the potential capability of the proposed PWARX systems identification algorithm in handling a class of switched linear systems.

Chapter 3

Identification of Switched Markov Autoregressive eXogenous Systems With Hidden Switching State

¹ Identification of the Switched Markov Autoregressive eXogenous (ARX) systems is considered in this chapter. With a Markov chain model governing the evolution of the hidden switching state, a Switched Markov ARX System (SMARX) is formulated and a solution strategy is proposed. Expectation-Maximization (EM) algorithm is employed in the identification of the SMARX systems in which both a Hidden Markov Model (HMM) for the discrete switching dynamics and a local ARX model parameters for continuous dynamics are estimated. Through the comparison between the proposed method and previous switched ARX system identification methods on simulated examples, it is shown that by modeling both the switching and continuous dynamics, the accuracy of the identification results can, to various extent, be increased.

3.1 Introduction

With tighter environmental regulation and fiercer market competition, process industry has been aggressively searching for better solutions to reduce the environment footprint as well as increase the production quality. For instance, model predictive control (MPC), owing to its capability of handling multi-variable process in an optimal way, has gained tremendous popularity in process industry after being firstly introduced in 1970s (Qin & Badgwell (2003)). While all these model based control, optimization and monitoring techniques are extensively applied nowadays, their performance is, to large extent, dependent on the the quality of the process models

1. This chapter has been submitted to Automatica

they adopt. In view of all the importance of process models, a significant amount of work has already been done on process identification and dynamic modeling (Ljung (1987); Huang & Kadali (2008)). Although linear time-invariant modeling methods have been successful in many cases, in chemical process industry, however, the system dynamics change caused by interactions between process and its operating environment as well as intrinsic physio-chemical discontinuities is commonly experienced and this poses great challenges to traditional system identification methods (El-Farra & Christofides (2003); El-Farra et al. (2005)).

Hybrid systems, in which continuous dynamics and discrete events coexist, are considered to be a natural modeling framework for switched systems. Issues such as optimal control, process optimization and modeling of hybrid systems have been extensively investigated by both industrial practitioners and academic researchers (Barton & Pantelides (1994), Barton et al. (2000); Barton & Lee (2004); El-Farra & Christofides (2003); El-Farra et al. (2005)). As a special but representative form of hybrid systems, switched ARX systems assume that all the local continuous dynamics of hybrid systems are linear and ARX model structure is employed to represent the linear dynamics (Jin & Huang (2009b); Vidal (2008)). Relevant works on switched ARX process identification have been put forward by a considerable number of researchers (Ferrari-Trecate et al. (2003); Juloski et al. (2005); Bempoard et al. (2005); Nakada et al. (2005); Vidal et al. (2003); Vidal (2008); Ragot et al. (2003); Roll et al. (2004)). For example, Juloski et al. (2005) proposed a Bayesian theory based method for the identification of a class of switched ARX systems represented by Piecewise ARX system. By employing particle filters as a means to iteratively classify the data, the data points are clustered and the local model parameters are estimated based on the classification result. For more detailed discussion and review of the work that has been done on hybrid system modeling, readers may refer to tutorials on hybrid systems written by Simone Paoletti et al. (2007) and Luisella Balbis et al. (2007) as well as the references therein.

In our previous work (Jin & Huang (2009b)), we proposed a method for the identification of a class of switched ARX process in which, the transitions among different subsystems are triggered by traveling of regressors (PWARX system) or purely time driven (time based switched ARX system). The Expectation-Maximization (EM) algorithm is employed and the whole identification problem is formulated in such a way that the data classification results along with the estimation of sub-model parameters are obtained simultaneously. Moreover, to deal with the real data set which may be polluted by outliers, robustness of the algorithm is considered and achieved by employing contaminated Gaussian distribution to describe the noise distribution. However, modeling of the switching dynamics has not been considered in the switched ARX modeling. In practice, the switching dynamics

may follow some patterns that could be modeled by, for example, Markov chain (Costa et al. (2004)), or in other cases, the switching dynamics exhibit certain behavior and it would be beneficial to use the Markov chain to describe it (Wong & Lee (2009)). In the literature, switched systems with Markovian switching dynamics are normally referred to as Jump Markov Systems (JMS) or Jump Markov Linear systems (JMLS) if the subsystems are further specified as linear. Considerable amount of research has been done on modeling, control and optimization of JMLS owing to its potential applications in speech recognition (Rabiner (1989)), process control (Morales-Menendez et al. (2003)), signal processing (Logothetis & Krishnamurthy (1999); Doucet et al. (2001)), chemical engineering (Tamir (1998)) and other interesting areas (de Souza (2007); Costa et al. (2004)). An interesting illustrative example in which the concept of JMLS was successfully applied is the optimal control of a solar thermal receiver (Sworder & Rogers (1983)). In that case, a Markov jump controller was designed by assuming that the transition between different types of weathers (cloudy or sunny) follows a simple first order Markov chain. There are various model structures in describing the linear dynamics, such as linear state space model, linear first principle model (normally a set of differential equations in continuous case and difference equations in discrete case) and linear empirical input-output model (ARX, BOX-Jenkins, Output Error). In this chapter, we will confine our discussion to ARX models. This assumption not only greatly simplifies the computation problem so that we can focus on the main issues of JMLS modeling, but also reasonable approximation in practice given the capability of the ARX model structure in approximating any linear dynamics (Ljung (1987)). For notation simplicity, abbreviation SMARX will be used hereafter when referring to the discussed Switched Markov ARX system.

The mathematical formulation of SMARX systems is as follows:

$$y_k = \theta_{I_k}^T \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, k = 1, 2 \dots N \quad (3.1)$$

where $y_k \in R$, $x_k \in R^n$ represent the output and regressor of the system at k th time point respectively. Regressor x_k is expressed as:

$$x_k = [y_{k-1} \quad y_{k-2} \cdots y_{k-na} \quad u_{k-1}^T \quad u_{k-2}^T \cdots u_{k-nb}^T]^T \quad (3.2)$$

where na and nb are orders of the output and input, $u \in R^m$ is the input and $n = na + m \cdot nb$. $e_k \in R$ is Gaussian distributed noise with zero mean and variance σ^2 . $\theta_i \in R^{n+1}$ is the parameter vector of the I_k th sub-model.

For the SMARX system as expressed in Equation 3.1, random switching variable $I \in \{1, 2 \dots S\}$ evolves in a Markovian fashion. In other words, the state of variable I at time point k only depends on its immediate past and it has only S discrete

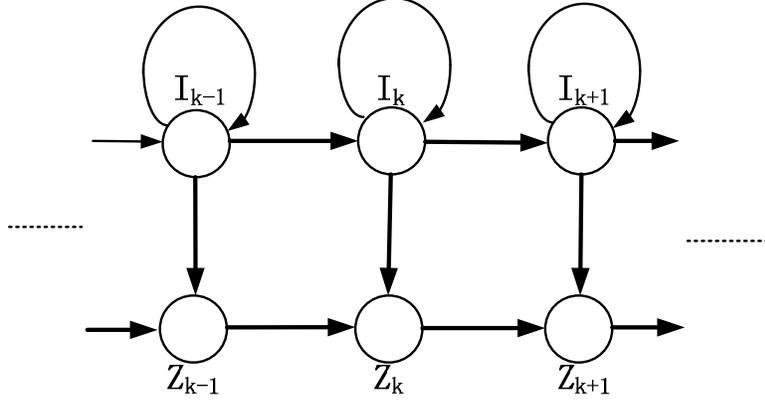


Figure 3.1: Graphical representation of SMARX system with hidden switching variable

values. Putting this in a mathematical form, we have

$$\alpha_{ij} = P(I_k = i \mid I_{k-1} = j), k = 2, 3 \dots N, 1 \leq i, j \leq S \quad (3.3)$$

$$\sum_{j=1}^S \alpha_{ij} = 1 \quad (3.4)$$

When the switching variable I is not known *a priori* and cannot be measured directly, the discrete switching dynamics described by a latent variable I is normally referred to as a Hidden Markov Model (HMM). As one of important statistical models, HMM has already been applied to various areas (Rabiner (1989); Tamir (1998)). A graphical representation of the SMARX system with hidden switching variable is shown in Figure 3.1.

Following the same assumptions regarding the fixed orders as well as the fixed number of local ARX models as being made in previous literature (Jin & Huang (2009b); Juloski et al. (2005); Ferrari-Trecate et al. (2003)), the tasks involved in the identification of SMARX systems with hidden switching variable include (1) estimating the value of the hidden switching state at each time point; (2) estimating the transition probability matrix of the Hidden Markov Model; (3) estimating the parameters of the local ARX models.

The main contribution of this chapter is to identify both the continuous ARX model for process dynamics and the discrete-valued HMM for switching dynamics, thus improving both the clustering and identification performance. Comparisons between the proposed method and the most relevant previous algorithms for the identification of piecewise/switched ARX systems in terms of their identification performance are conducted using a simulated chemical process example. It is found that the incorporation of a Markov model in the identification procedure can, to various extent, improve the relevant identification results. To simplify the notation when referring to the comparative methods, the method proposed by Jin etc. (2009)

(Jin & Huang (2009b)) will be abbreviated as SARX method and a representative algorithm put forward by Nakada et al. (2005) (Nakada et al. (2005)) is referred to as Nakada method hereafter.

The chapter is organized as follows: An illustrative numerical example of SMARX systems is given in Section 2 followed by a brief introduction of the EM algorithm. Section 3 includes all the mathematical derivations for the identification of SMARX process with hidden switching states. A numerical simulation example is put forward in Section 4 and a comparative study of the identification results from the proposed method, the SARX method and the Nakada method is conducted. Section 5 illustrates the application results for a simulated continuous fermentation reactor when the grade of the fermenter's feeding material changes in a Markovian fashion. Results from the the proposed method and two previous methods are examined and compared. Section 6 draws the conclusion and provides our perspective on potential applications of SMARX systems in process identification and control.

3.2 Introduction to SMARX Systems and The EM Algorithm

3.2.1 An Illustrative Bimodal SMARX System

To render a simple but representative example of SMARX systems, a bimodal switched system is constructed in which the transition between two sub-models is controlled by a hidden variable. The system is given as follows:

$$y_k = \begin{cases} [y_{k-1} & u_{k-1}] \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + e_k, & I_k = 1 \\ [y_{k-1} & u_{k-1}] \begin{bmatrix} -1 \\ 2 \end{bmatrix} + e_k, & I_k = 2 \end{cases}, k = 1, 2 \dots N \quad (3.5)$$

where I_k in Equation 3.5 is a hidden state which evolves in a first order Markov chain manner and the transition probability matrix of the Markov chain is set to be $A = (\alpha_{ij}) = \begin{bmatrix} 0.2 & 0.8 \\ 0.75 & 0.25 \end{bmatrix}$. The ARX model parameters a_1 and b_1 change according to the value of the hidden state I at each time instance. Meanwhile, except for the system input-output data, neither switching state at each time point $I_k, k \in 1, 2 \dots N$ nor the transition probability matrix (α_{ij}) can be directly measured or observed.

In practical words, the assumption of the switching in accordance with the Markovian transition probability implies that, for example, the probability for the process to operate in mode 1 is 20% and the probability for it to switch from mode 1 to mode 2 is 80%. Thus the Markov chain models certain tendency of process

operation. For instance, it is known that a reactor which has multiple equilibrium points tends to operate in more stable operating points and its transition among different operating points may be described by a Markov chain.

3.2.2 EM Algorithm

As being introduced in our previous work (Jin & Huang (2009b,c)), the EM algorithm consists of two consecutive steps, the first step calculates the Expectation of the complete data while the second step searches for the parameters to increase the Expectation of the complete data. Hence, starting with some initial values of the parameters, the EM algorithm can ultimately converge to some stationary points after finite steps of iteration. After first being introduced by Dempster et al. (1977), the EM algorithm has found extensive applications in various areas including machine learning, computer vision, speech recognition, bioinformatics, psychometrics and so on. While EM algorithm has received significant attention in many areas, it is a relatively new data processing/optimization technique in control engineering. So far, the main applications of the EM algorithm in process control literature have been mainly on parameter estimation for nonlinear/switched systems (Chitralkha et al. (2009); Goodwin & Aguero (2005, 2008)).

For linear systems with Gaussian distributed noise, the expectation of the complete data set can be derived in an analytical form, which is normally referred to as Q function in the literature (Dempster et al. (1977); McLachlan & Krishnan (1996)). How to derive the expression for the Q function and maximize it (regular EM algorithm) or increase it (generalized EM algorithm) is an interest of active research when employing the EM algorithm to solve certain optimization problems. The detailed mathematical derivation for the Q function and the maximization procedure for the discussed SMARX process is given in the following section.

3.3 Mathematical Solution for The SMARX System Identification

Let Θ be the parameters of the switching ARX model, it is composed of state transition probability $\alpha_{ij}, i = 1, 2 \dots M, j = 1, 2 \dots M$, the initial state distribution $\pi_i, i = 1, 2 \dots M$, the local ARX model parameters $\theta_i, i = 1, 2 \dots M$ and the process noise variance σ^2 . Given the observed data set $C_{obs} = Z_N, Z_{N-1} \dots Z_1$ and hidden state I to denote cluster identify of data points at each sampling instant, the complete data $C = (C_{obs}, I)$ is constructed. Since the state I is actually hidden, the Q function can be derived as the conditional expectation over the hidden state I given the observed data C_{obs} and the initial/previous parameter estimation results Θ_{old} ,

$$\begin{aligned}
Q(\Theta | \Theta_{old}) &= E_{I|(\Theta^{old}, C_{obs})} \{ \log P(C_{obs}, I | \Theta) \} \\
&= E_{I|(\Theta^{old}, C_{obs})} \{ \log P(Z_N, Z_{N-1} \dots Z_1, I_N \dots I_1 | \Theta) \} \\
&= E_{I|(\Theta^{old}, C_{obs})} \{ \log \prod_{k=1}^N P(Z_k, I_k | Z_{k-1}, \dots Z_1, I_{k-1}, \dots I_1, \Theta) \} \\
&= E_{I|(\Theta^{old}, C_{obs})} \{ \log \prod_{k=1}^N P(Z_k | Z_{k-1}, \dots Z_1, I_k, \dots I_1, \Theta) P(I_k | I_{k-1}) \} \\
&= E_{I|(\Theta^{old}, C_{obs})} \{ \sum_{k=2}^N [\log P(Z_k | Z_{k-1} \dots Z_1, I_k, \Theta) + \log P(I_k | I_{k-1})] \\
&\quad + \log P(Z_1 | I_1, \Theta) + \log P(I_1) \} \\
&= \sum_I P(I | \Theta^{old}, C_{obs}) (\sum_{k=1}^N \log P(Z_k | Z_{k-1} \dots Z_1, I_k, \Theta) \\
&\quad + \sum_{k=2}^N \log P(I_k | I_{k-1}) + \log P(I_1)) \\
&= \sum_{I_1} \dots \sum_{I_N} P(I_1 \dots I_N | \Theta^{old}, C_{obs}) \{ \sum_{k=1}^N \log P(Z_k | Z_{k-1} \dots Z_1, I_k, \Theta) \\
&\quad + \sum_{k=2}^N \log P(I_k | I_{k-1}) + \log P(I_1) \} \\
&= \sum_{i=1}^M \sum_{k=1}^N P(I_k = i | \Theta^{old}, C_{obs}) \log P(Z_k | Z_{k-1} \dots Z_1, \Theta_i) \\
&\quad + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=2}^N P(I_k = i, I_{k-1} = j | \Theta^{old}, C_{obs}) \log \alpha_{ij} \\
&\quad + \sum_{i=1}^M P(I_1 = i | \Theta^{old}, C_{obs}) \log \pi_i
\end{aligned} \tag{3.6}$$

In the derivation of Equation 3.6, we have used the identities such as $P(Z_k | Z_{k-1} \dots Z_1, I_k, \dots I_1, \Theta) = P(Z_k | Z_{k-1} \dots Z_1, I_k, \Theta)$ and $P(I_k | I_{k-1} \dots I_1) = P(I_k | I_{k-1})$ in accordance with the Markov chain property of I_k . Moreover, since the expectation over the hidden state I is taken with respect to the hidden state I over the time period from $k = 1$ till $k = N$, the summation over I has to be expanded to N different time instants over their possible values. To maximize the Q function over parameters Θ , derivative operation is taken with respect to each individual component of the system parameters Θ . Therefore, optimal local ARX

model parameters $\theta_i, i = 1 \dots M$ are calculated by taking the derivative with respect to the local ARX model parameters and then equating the derivatives to zero:

$$\begin{aligned} & \frac{\partial \sum_{i=1}^M \sum_{k=1}^N P(I_k = i | \Theta^{old}, C_{obs}) \log P(Z_k | Z_{k-1} \dots Z_1, \Theta_i)}{\partial \theta_i} \\ &= \frac{\partial \sum_{i=1}^M \sum_{k=1}^N P(I_k = i | \Theta^{old}, C_{obs}) \log \left(\frac{1}{\sqrt{2\pi\sigma^{old}}} \exp \frac{(y_k^T y_k - 2y_k^T x_k \theta_i + \theta_i^T x_k^T x_k \theta_i)}{2(\sigma^{old})^2} \right)}{\partial \theta_i} \\ &= 0 \end{aligned} \quad (3.7)$$

After several steps of mathematical manipulations, the relevant updating equations for $\theta_i, i = 1 \dots M$ is:

$$\theta_i^{New} = \frac{\sum_{k=1}^N P(I_k = i | C_{obs}, \Theta^{old}) x_k^T y_k}{\sum_{k=1}^N P(I_k = i | C_{obs}, \Theta^{old}) x_k^T x_k} \quad (3.8)$$

Following the similar procedure, optimal estimation of α_{ij} is obtained by taking the derivative with respect to the relevant components in the Q function. However, the optimization problem here is constrained by $\sum_{i=1}^M \alpha_{ij} = 1$ and as a result, Lagrange multiplier λ is introduced and the unconstrained optimization equation is obtained:

$$\begin{aligned} & \frac{\partial \sum_{i=1}^M \sum_{j=1}^M \sum_{k=2}^N P(I_k = i, I_{k-1} = j | \Theta^{old}, C_{obs}) \log \alpha_{ij} + \lambda (\sum_{i=1}^M \alpha_{ij} - 1)}{\partial \alpha_{ij}} \\ &= 0 \end{aligned} \quad (3.9)$$

This yields the updating equation for α_{ij} as

$$\alpha_{ij}^{New} = \frac{\sum_{k=2}^N P(I_k = i, I_{k-1} = j | C_{obs}, \Theta^{old})}{\sum_{i=1}^M \sum_{k=2}^N P(I_k = i, I_{k-1} = j | C_{obs}, \Theta^{old})} \quad (3.10)$$

The initial distribution of the hidden state $\pi_i, i = 1 \dots M$ can also be found by introducing the Lagrange multiplier λ and then taking the derivative,

$$\frac{\partial \sum_{i=1}^M P(I_1 = i | \Theta^{old}, C_{obs}) \log \pi_i + \lambda (\sum_{i=1}^M \pi_i - 1)}{\partial \pi_i} = 0 \quad (3.11)$$

As a result, we can get the updating equation for $\pi_i, i = 1, \dots M$

$$\pi_i^{New} = P(I_1 = i | \Theta^{old}, C_{obs}) \quad (3.12)$$

The updating equation for noise variance σ^2 is derived by taking the derivative of the first component of the Q function with respect to it,

$$\begin{aligned} & \frac{\partial \sum_{i=1}^M \sum_{k=1}^N P(I_k = i | \Theta^{old}, C_{obs}) \log P(Z_k | Z_{k-1} \dots Z_1, \Theta_i)}{\partial \sigma} \\ &= \frac{\partial \sum_{i=1}^M \sum_{k=1}^N P(I_k = i | \Theta^{old}, C_{obs}) (-\log \sqrt{2\pi}\sigma - \frac{1}{2\sigma^2} (y_k - x_k \theta_i^{New})^T (y_k - x_k \theta_i^{New}))}{\partial \sigma} \\ &= 0 \end{aligned} \quad (3.13)$$

Finally, the updating equation obtained from Equation 3.13 for noise variance σ^2 equals

$$(\sigma^{New})^2 = \frac{\sum_{i=1}^M \sum_{k=1}^N P(I_k = i | C_{obs}, \Theta^{old})(y_k - x_k R_i^{New})^T (y_k - x_k R_i^{New})}{\sum_{i=1}^M \sum_{k=1}^N P(I_k = i | C_{obs}, \Theta^{old})} \quad (3.14)$$

It can be seen that in the derivation of equations for local model parameters $\theta_i, i = 1, 2 \dots M$ (Equation 3.8), transition probabilities $\alpha_{ij}, i = 1, 2 \dots M, j = 1, 2 \dots M$ (Equation 3.10) and noise variance σ^2 (Equation 3.14), the intermediate terms $P(I_k = i | C_{obs}, \Theta^{old}), k = 1, 2 \dots N$ and $P(I_k = i, I_{k-1} = j | C_{obs}, \Theta^{old}), k = 2, 3 \dots N$ are required as a part of updating equations. To calculate the intermediate terms from the estimated parameters in the previous iteration, the expression for $P(I_k = i, I_{k-1} = j | \Theta^{old}, Z_k, \dots Z_1)$ can be derived as:

$$\begin{aligned} & P(I_k = i, I_{k-1} = j | \Theta^{old}, Z_k, \dots Z_1) \\ &= \frac{P(Z_k | I_k = i, I_{k-1} = j, \Theta^{old}, Z_{k-1} \dots Z_1)}{P(Z_k | \Theta^{old}, Z_{k-1} \dots Z_1)} P(I_k = i | I_{k-1} = j, \Theta^{old}, Z_{k-1} \dots Z_1) \\ &\cdot P(I_{k-1} = j | \Theta^{old}, Z_{k-1} \dots Z_1) \\ &= \frac{P(Z_k | I_k = i, \Theta^{old}, Z_{k-1} \dots Z_1) P(I_k = i | I_{k-1} = j, \Theta^{old}) P(I_{k-1} = j | \Theta^{old}, Z_{k-1} \dots Z_1)}{\sum_{m=1}^M \sum_{n=1}^M P(Z_k, I_k = m, I_{k-1} = n | \Theta^{old}, Z_{k-1} \dots Z_1)} \\ &= \frac{P(Z_k | I_k = i, \Theta^{old}, Z_k \dots Z_1) P(I_k = i | I_{k-1} = j, \Theta^{old}) P(I_{k-1} = j | \Theta^{old}, Z_{k-1} \dots Z_1)}{\sum_{m=1}^M \sum_{n=1}^M P(Z_k | I_k = m, Z_{k-1} \dots Z_1, \Theta^{old}) P(I_k = m | I_{k-1} = n, \Theta^{old}) P(I_{k-1} = n | Z_{k-1} \dots Z_1, \Theta^{old})} \end{aligned} \quad (3.15)$$

where i, j, m, n in Equation 3.15 represent cluster identity of the data points at k and $k - 1$ time instance respectively. Θ denotes the overall parameters of the Markov switched system including the local model parameters, transition probabilities and noise variance. Each component of Equation 3.15 in both the denominator and numerator can be calculated or retrieved based on the parameter estimation results from the previous iteration. For instance, $P(I_k = i | I_{k-1} = j, \Theta^{old})$ equals α_{ij}^{old} which denotes the transition probability obtained in the previous calculation, $P(Z_k | I_k = i, \Theta^{old}, Z_{k-1} \dots Z_1)$ represents the probability of observing Z_k given all the past observed data as well as the previously estimated local model parameters, $P(I_{k-1} = j | \Theta^{old}, Z_{k-1} \dots Z_1)$ is obtained through the discrete-valued state propagation of Markov chain starting from the initial estimation of $P(I_1 = j | \Theta^{old}, Z_{k-1} \dots Z_1) = (\pi_j)^{old}$. Therefore, although it appears to be more complex after expanding $P(I_k = i, I_{k-1} = j | \Theta^{old}, Z_k, \dots Z_1)$ into multiple components, however, the expansion/conversion enables us to utilize the parameter estimation results Θ^{old} from the previous step to calculate the probability $P(I_k = i, I_{k-1} = j | \Theta^{old}, Z_k, \dots Z_1)$. After calculating the probabilities as in Equation 3.15, use the results in Equation 3.10 so that the updating equation for $\alpha_{i,j}$ can be calculated.

The term $P(I_k = i | C_{obs}, \Theta^{old}), k = 1, 2 \dots N$ is needed in calculating updating equations for local model parameters as well as the noise variance, which are denoted

by Equation 3.8 and Equation 3.14 respectively. Its computation equation can be derived as

$$\begin{aligned}
P(I_k = i | \Theta^{old}, C_{obs}) &= \sum_{j=1}^M P(I_k = i, I_{k-1} = j | \Theta^{old}, Z_k, Z_{k-1}, \dots, Z_1) \\
&= \frac{\sum_{j=1}^M P(Z_k | I_k = i, \Theta^{old}, Z_{k-1} \dots Z_1) P(I_k = i | I_{k-1} = j, \Theta^{old}) P(I_{k-1} = j | \Theta^{old}, Z_{k-1} \dots Z_1)}{\sum_{m=1}^M \sum_{n=1}^M P(Z_k | I_k = m, Z_{k-1} \dots Z_1, \Theta^{old}) P(I_k = m | I_{k-1} = n, \Theta^{old}) P(I_{k-1} = n | Z_{k-1} \dots Z_1, \Theta^{old})}
\end{aligned} \tag{3.16}$$

where the identities i, j, m, n share the same meaning as given in Equation 3.15. Based on the calculation results from Equation 3.15 and Equation 3.16, the updating equations for various parameters $\theta_i, i = 1, 2 \dots M$, $\alpha_{ij}, i = 1, 2 \dots, M, j = 1, 2 \dots M$, $\pi_i, i = 1, 2 \dots M$ and σ^2 can be derived.

3.4 Verification by A Numerical Simulation Example

To validate the proposed identification algorithm, a simulated SMARX process with three sub-systems is constructed and its equations are given as follows,

$$y_k = \begin{cases} \begin{bmatrix} y(k-1) & y(k-2) & u(k-1) & u(k-2) \end{bmatrix} \begin{bmatrix} 1.143 \\ -0.4346 \\ 0.0572 \\ 0.2415 \end{bmatrix} + e_k, & I_k = 1 \\ \begin{bmatrix} y(k-1) & y(k-2) & u(k-1) & u(k-2) \end{bmatrix} \begin{bmatrix} 0.9534 \\ -0.0475 \\ 0.0618 \\ 0.0336 \end{bmatrix} + e_k, & I_k = 2 \\ \begin{bmatrix} y(k-1) & y(k-2) & u(k-1) & u(k-2) \end{bmatrix} \begin{bmatrix} 1.178 \\ -0.09 \\ 0.089 \\ 0.15 \end{bmatrix} + e_k, & I_k = 3 \end{cases}, k = 1, 2 \dots N \tag{3.17}$$

I_k in Equation 3.17 is a hidden random integer variable which evolves according to a first order Markov chain model and $I_k \in 1, 2, 3, k = 1 \dots N$. Under different realization of hidden variable I_k , corresponding sub-system with distinct parameters

will take effect. Set the transition matrix $A = (\alpha_{ij}) = \begin{bmatrix} 0.25 & 0.1 & 0.65 \\ 0.55 & 0.35 & 0.1 \\ 0.15 & 0.15 & 0.7 \end{bmatrix}$ and noise

variance $e_k \sim N(0, 0.025)$. The input u is randomly changed within $[-5, 5]$ following uniform distribution. 900 data points are generated and passed through the proposed identification algorithm, the identified sub-system parameters are shown in Table 3.1.

$\theta_i, i = 1, 2, 3$ in Table 3.1 are true system parameters while $\hat{\theta}_i, i = 1, 2, 3$ are the estimated ones. It is noticed that the estimated sub-system parameters are close

Table 3.1: True system parameters and estimated parameters

θ_1	$\hat{\theta}_1$	θ_2	$\hat{\theta}_2$	θ_3	$\hat{\theta}_3$
1.143	1.1432	0.9534	0.9536	1.178	1.1850
-0.4346	-0.4381	-0.0475	-0.0516	-0.09	-0.0941
0.0572	0.0597	0.0618	0.0644	0.089	0.0889
0.2415	0.2404	0.0336	0.0316	0.15	0.1499

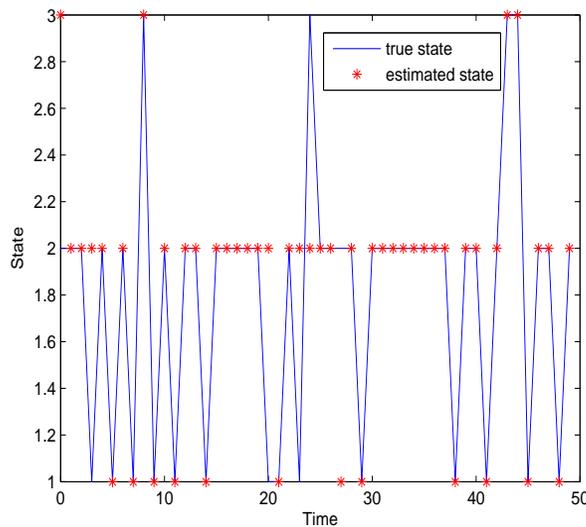


Figure 3.2: Comparison between estimated and actual state sequence

to their real counterparts. Moreover, the estimated transition probability matrix is

$$\hat{\alpha}_{ij} = \begin{bmatrix} 0.181 & 0.1242 & 0.6948 \\ 0.5349 & 0.3839 & 0.0812 \\ 0.1762 & 0.1393 & 0.6845 \end{bmatrix} \text{ and the estimated noise variance is } \hat{\sigma}^2 = 0.0250.$$

Some minor discrepancy between the estimated transition probabilities and its true ones is observed. Due to the uncertainty introduced by the noise, it is expected that the misclassification of some data points may happen during the process of identification, which leads to the deviation of the estimated transition probabilities from their true values. Hence, when dealing with process data collected from industrial or pilot-scale setup, pre-filtering of the noise may be conducted as overly high noise variance may result in poor clustering results.

Part of the estimated hidden state sequence $\hat{I}_k, k = 1 \dots N$ is shown along with its true counterpart in Fig. 3.2. Some minor disagreement between the true I_k and the estimated \hat{I}_k is spotted in the figure.

3.5 A Simulated Continuous Fermentation Reactor

3.5.1 Process Description and Problem Statement

A well mixed continuous fermentation reactor (Henson & Seborg (1992b)) is adopted here to perform the test of the algorithm on a nonlinear process. This example was adopted in our previous work (Jin & Huang (2009b)) and is used in this section as a comparison of the new proposed method with the previous results. The simplified model consists of relatively few parameters which, makes itself more appropriate for control oriented simulation and optimization purpose. The nonlinear process dynamics are described by the following equations:

$$\dot{X} = -DX + \mu X \quad (3.18)$$

$$\dot{S} = D(S_f - S) - \frac{1}{Y_{X/S}}\mu X \quad (3.19)$$

$$\dot{P} = -DP + (\alpha\mu + \beta)X \quad (3.20)$$

$$\mu = \frac{\mu_m(1 - \frac{P}{P_m})S}{K_m + S + \frac{S^2}{K_i}} \quad (3.21)$$

where X, S, P are biomass concentration, substrate concentration, and production concentration respectively. Feed substrate concentration S_f and dilution rate D are normally treated as the system inputs. Cell-mass yields $Y_{X/S}$, α and β are system parameters. Nonlinearities of the process are mainly introduced by Equation 3.21 in which μ_m (maximum specific growth rate), P_m (product saturation constant), K_m (substrate saturation constant) and K_i (substrate inhibition constant) are considered to be the model parameters. The schematic diagram of the fermentation reactor is given in Figure 3.3:

The discussed fermenter shown in Figure 3.3 has two manipulated inputs and three states. Assume that the dynamic model from the input dilution rate D to the state biomass concentration X is of interest, to accurately capture the dynamic relationship between these two variables, all the other input variables including the feed substrate concentration are maintained as constant throughout the identification period. In other words, any fluctuation in the feed substrate concentration will be considered as external noise to the system under identification.

When the process from the dilution rate to the biomass concentration is undergoing certain excitation procedure, any undetected abrupt change of the feed substrate concentration may exert undesirable influences on the identified model. For instance, if the feed substrate concentration jumps from its normal operating point $20kg/m^3$ to $26.5kg/m^3$ due to upstream feed stock change, the dynamics of the

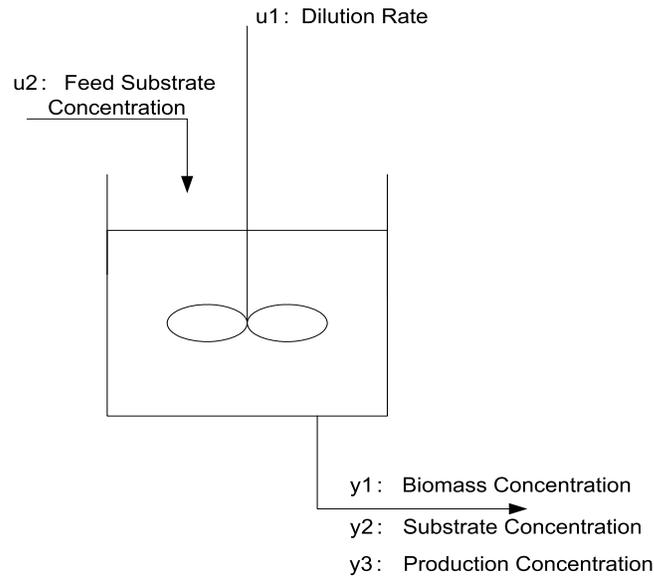


Figure 3.3: Schematic diagram of the fermentation reactor

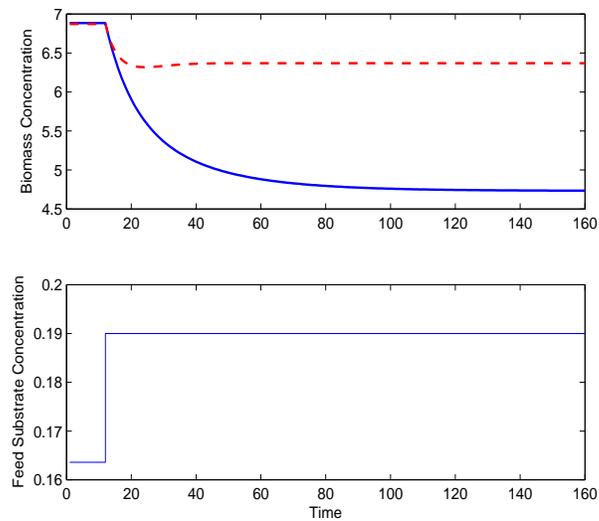


Figure 3.4: Input-Output data of the continuous fermenter. Dash line represents the step response when $u_1 = 20$, solid line denotes the step response when $u_1 = 26.5$

process under investigation will change dramatically. Figure 3.4 shows the step responses of the SISO model under different feed substrate concentration values.

Therefore, if a model is identified directly from the identification data set collected under different feed substrate concentration values, it would be expected that the identified model can hardly capture the behavior of the complete process. As a result, achievement of satisfactory performance by any model-based controller or other optimization techniques which use the identified model would be unlikely. Moreover, from the perspective of process operation monitoring, it will also be desirable that any significant change in operating conditions, such as the property of raw feed material, can be detected immediately. Hence, in the following sections, we will illustrate the capability of the proposed method in detecting the process change as well as in process modeling under different operating conditions. The switching dynamics will also be identified in the form of a Markov chain model.

3.5.2 Results

As discussed in the previous section, for chemical/bio-chemical reactors, fluctuation of the feed stream properties such as flow rate, temperature and concentrations may greatly change the reaction kinetics in the reactor. Here, we take a scenario in which the fermenter feed substrate concentration fluctuates between two operating values. Each of the feed substrate concentration values corresponds to one specific mode, say, mode 1 corresponding to $S_f = 26.5\text{kg}/\text{m}^3$ and mode 2 to the case when $S_f = 20\text{kg}/\text{m}^3$. Suppose that the mode change/switching does not occur frequently; in other words, the abrupt jump of the concentration value is infrequent during the operation. To generate a mode sequence with only a few switchings, the diagonal value of the transition matrix is set to be close to 1 while the other elements of the matrix is given a quite small number. As a result, the underlying transition probability matrix is assumed as $A = (\alpha_{ij}) = \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}$, $i = 1, 2, j = 1, 2$. Adopting the same model parameters as given by Seborg et al. (1992), the simulated responses are shown in Figure 3.5.

From Figure 3.5, we can roughly tell that the output biomass concentration dynamics exhibit distinct behaviors during the excitation period of the input signal. Passing the collected data through the proposed algorithm, the identified models as well as the estimated mode switching sequence are shown in Figure 3.7 and Figure 3.8.

Figure 3.8 shows that the estimated/identified mode sequence is in a good agreement with the true modes of the process. Moreover, self-validation results given in Figure 3.7 indicate that the identified models under each operation mode (feed substrate concentration value) are sufficient to capture the dynamics of the process.

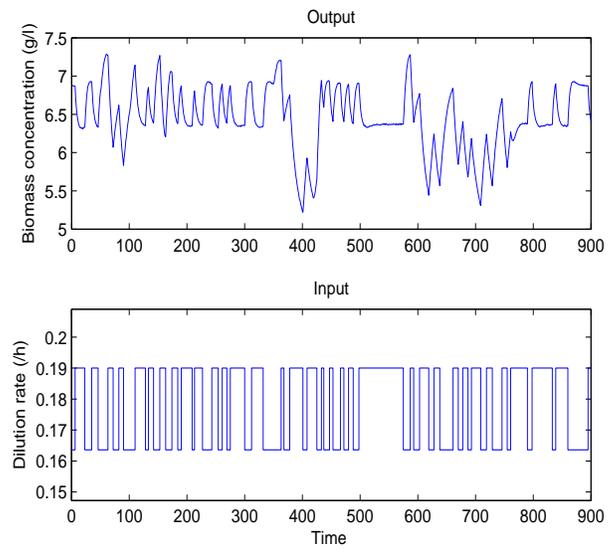


Figure 3.5: Input-Output data of the continuous fermenter

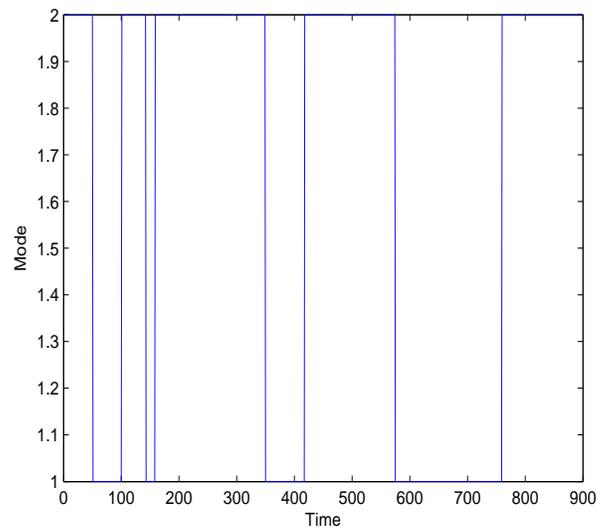


Figure 3.6: Mode sequence of the fermentation reactor

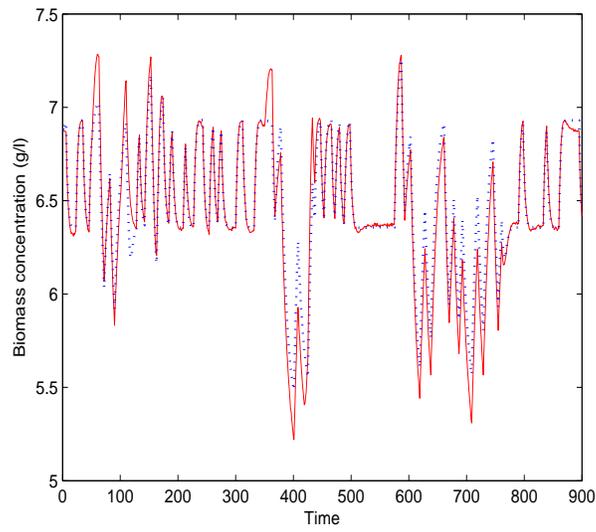


Figure 3.7: Self validation of the identified model, $MSE=0.001423$. Red solid line represents the true fermenter data, blue dashed line denotes the prediction of the simulated model

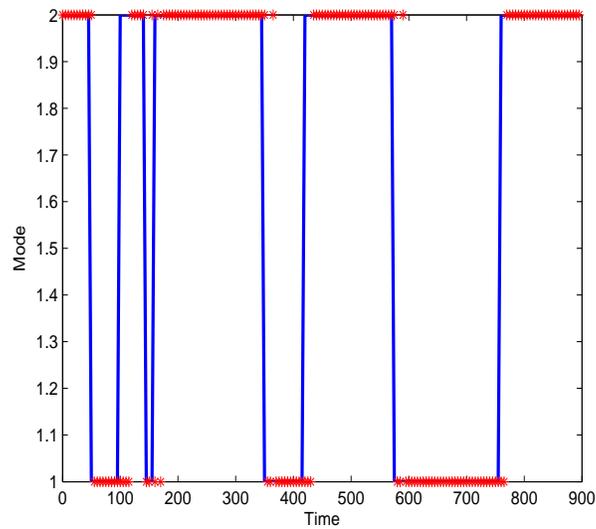


Figure 3.8: Comparison of the true modes with the identified mode sequence, 92.1% of time is right. Red cross is the estimated mode sequence, blue solid line denotes the true mode of the process through out the simulation time period.

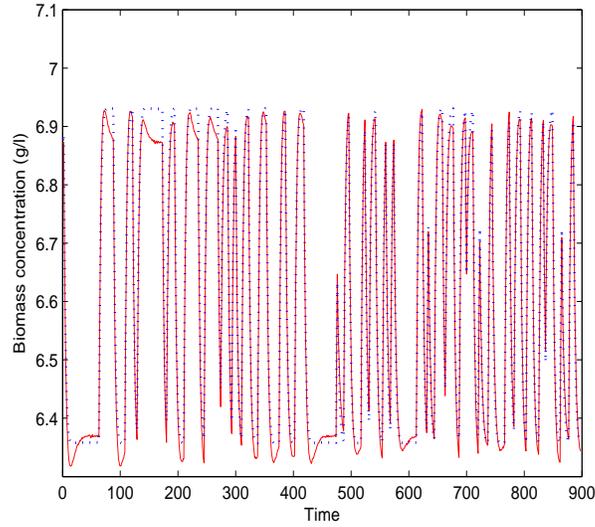


Figure 3.9: Cross validation of the identified model under mode 1, MSE=0.0.000479. Red solid line represents the true fermenter data, blue dashed line denotes the prediction of the simulated model

Cross validation is also performed here to further verify the identified models. Figure 3.9 and Figure 3.10 render the relevant cross validation results.

Small Mean Squared Errors (MSE) are achieved for each of the identified local model which effectively verify the performance of the identified models. Moreover, the estimated transition probability matrix $\hat{A} = \hat{\alpha}_{ij} = \begin{bmatrix} 0.9767 & 0.0233 \\ 0.0132 & 0.9868 \end{bmatrix}$, $i, j = 1, 2$ is identified, which is close to the real values.

To render a better view of the difference brought by modeling the switching dynamics, the proposed SMARX system identification method is compared with the SARX method as well as the method introduced by Nakada et al (2005). With the same data set from the fermentation reactor as well as the same initial parameter guess for each identification method, the identified results are listed in Table 3.2.

It is noticed from Table 3.2 that one of the models identified from both the SARX method and the Nakada method has a high MSE value in its cross validation. Table 3.2 ensures us that it is beneficial to incorporate the modeling of the switching mechanism into the identification procedure. Not only will the switching sequence be identified better, but also the accuracy of the estimated local ARX model parameters is increased.

3.6 Conclusion

Switched Markov ARX systems identification problem is investigated in this paper. The problem is formulated and solved within a probabilistic framework. The work

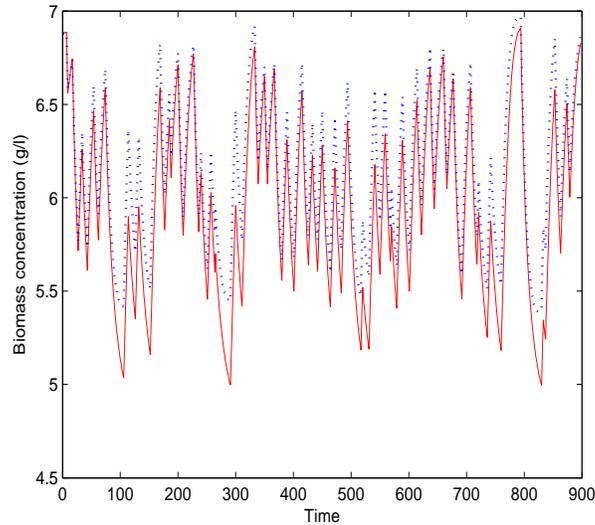


Figure 3.10: Self validation of the identified model under mode 2, MSE=0.0553. Red solid line represents the true fermenter data, blue dashed line denotes the prediction of the simulated model

Table 3.2: Comparison of identification results from different identification methods. The benchmark method is proposed by Nakada et al. (2005) and the SARX method is given by Jin & Huang (2009b). SV, CV 1, CV 2 are short for self validation and cross validation results for model 1 and model 2. N\A represents can not be calculated

Method	SV (MSE)	CV 1 (MSE)	CV 2 (MSE)
The proposed method	0.01423	0.00047	0.0553
The benchmark method	0.1883	0.00024	231.02
The SARX method	0.0226	0.00024	143.5

Method	Transition Matrix	Rate of Accuracy
The proposed method	$\begin{bmatrix} 0.9767 & 0.0233 \\ 0.0132 & 0.9868 \end{bmatrix}$	92.11%
The benchmark method	N\A	53.33%
The SARX method	N\A	53.33%

in this paper extended previous work on switched/piecewise ARX system identification by considering switching dynamics through a hidden Markov model. From the analysis of the identification results of the simulated switched system examples, it is found beneficial to incorporate the switching dynamics into the identification procedure and identify all the parameters (including continuous local model parameters as well as the discrete switching HMM model parameters).

Chapter 4

Identification of Switched Markov Autoregressive eXogenous Systems With Hidden Switching State

¹ This paper is concerned with the identification of a nonlinear process which operates over several working points with consideration of transition dynamics between the working points. Operating point changes due to economic considerations (e.g. grade change in polymer plants) or working environment changes (e.g. feed raw materials property change) are commonly experienced in process industry. These transitions among different operating conditions excite the inherent nonlinearity of the chemical process and pose significant challenges for process modeling. To circumvent the difficulties, we propose a probability-based identification method in which a linear parameter varying (LPV) model is built using process input-output data. Without knowing the local model dynamics *a priori*, only small excitation signals around each operating point are required to identify linear models of the local dynamics, and then the local models are synthesized with transition data to construct a global LPV model. Simulated numerical examples as well as an experiment performed on a pilot-scale heated tank are employed to demonstrate the effectiveness of the proposed method. The conclusion section raises some practical issues involved with the implementation of the algorithm and relevant suggestions are given.

4.1 Introduction

During the last few decades, process plants have seen an astonishing increase in their complexity which poses a great challenge for process modeling and control.

1. This chapter has been submitted to Journal of Process Control

Although linear modeling technology has been quite sophisticated after decades of development both in academic research and industrial applications (Ljung (1987)), however, due to the nonlinearity of the chemical production process, the performance of single linear model-based controllers or optimizers may be compromised or even unsatisfactory for certain highly nonlinear process (such as PH process). To overcome the limitations imposed by the nonlinearity of the process, researchers have developed different strategies for nonlinear process modeling. A black box modeling approach characterized by the usage of theoretically sound nonlinear functions such as Nonlinear Autoregressive eXogenous (NARX) models (Henson & Seborg (1992a); Proll & Karim (1994)) or artificial neural network models (S.Piche et al. (2000)) is able to identify nonlinear process models solely based on the process data. No specific process knowledge is required and the model parameters are chosen to minimize certain optimization criteria. However, as the model is built purely based on process data, to ensure the validity of the model within a large operating range, a global identification test throughout the whole operating region of the process has to be conducted which, may greatly interfere the plant operation. In addition, the search of the nonlinear model structure is also a considerable challenge. Fundamental modeling based on the first principles of the process (such as heat/mass conversation law) (Henson & Seborg (1992a)) has also gained great attention for nonlinear process modeling. It normally provides more meaningful model structure with less model parameters when compared with nonlinear models identified from black box methods. However, given the complexity of the industrial-scale process nowadays, it may become overly costly to build a model that meets the accuracy requirement due to the lack of the understanding of the process.

A Linear Parameter Varying (LPV) model, as being indicated by its name, is featured by its linear model structure and varying model parameters. After being first introduced by Shamma et al. (1991) in their study of gain scheduling control of LPV process, a considerable number of publications on LPV model identification have already been seen thanks to its capability in approximating complex nonlinear systems (Lee & Poolla (1996); Banerjee et al. (1997); Bamieh & Giarr (2002); Xu et al. (2009); Murray-Smith & Johansen (1997)). Bamieh et al. (2002) put forward a LPV identification method in which the input-output data as well as the scheduling parameters are assumed to be measured. As the LPV system model parameters vary with the change of the scheduling variable, polynomial functions with unknown coefficients are employed to describe the relationships of parameter changes. The identification problem is formulated in such a way that the least squares method (off-line estimation) or recursive least squares method (on-line estimation) can be directly applied to estimate the coefficients of the polynomial dependence functions. The method requires the input signal to be manipulated sufficiently throughout the

whole operating range so that persistent excitation of the process can be achieved. This requirement can be costly or even unrealistic in practice as too much upset caused by the input excitation signal may not be allowed. To circumvent the difficulty of designing the input signal for LPV system identification, Xu et al (2009) proposed a novel identification procedure in which only input excitation around certain chosen operating points is needed to approximate a global LPV model for the nonlinear process. To blend the local linear models identified around their relevant operating points, validity of each local model at every sampling time instance is calculated and the overall prediction of the LPV model is the weighted average of the predictions from all the local models. However, as a prerequisite for the identification method, all the local linear models need to be known *a priori* which, may deteriorate the performance of the identification method if any of the local models is not able to satisfactorily approximate the local process dynamics due to improper identification or undetected change of the process dynamics. Moreover, as Xu et al. (2009) employed cubic spline functions to represent the validity functions of each local model for the LPV model, how to search for the appropriate orders for each of those valid functions can be challenging in some cases. Therefore, developing an LPV model identification method with low test cost and less requirement on the prior knowledge of the process motivates us to conduct the research in this paper.

The model predictive controller (MPC), owing to its capability of handling multivariate process in an optimal way, has been widely accepted and implemented in process industry after its first advent in 1970s. Although linear MPC has already become a sophisticated advanced process control technology with well defined development and implementation procedures, nonlinear MPC, which is featured by the adoption of a nonlinear process model for process prediction and optimization, is still relatively less popular compared with its counterpart, linear MPC. The main reasons that prevent the nonlinear MPC from wide application lie in 1) complexity of nonlinear process modeling; 2) overly high computational burden with complex nonlinear process models for online implementation. An LPV model, owing to its simpler structure compared with first principles or nonlinear black box empirical models and its capability in approximating the nonlinear process dynamics, has already been utilized by a number of researchers in the design of nonlinear model predictive controllers (Xu et al. (2009); Baneqee & Arkun (1998); Giarre et al. (2006); Dougherty & Cooper (2003); Foss et al. (1995); Aufderheide & Bequette (2003)). In the test of these LPV model based nonlinear predictive controllers, it is shown that compared with predictive controllers with first principle models, the LPV model based controllers achieve near optimal control performance while requiring much less effort in building the nonlinear predictive model as well as in the need of computation power.

The work introduced in this paper aims at identifying the LPV models without requiring the prior knowledge of the local model parameters; namely all local models are identified simultaneously together with their transitions to form a global LPV model. Following the similar assumptions regarding the scheduling variables and the segmentation of the process as being made by Xu et al. (2009); Bamieh & Giarr (2002), we assume that the scheduling variable values can be measured or inferred and the number of local models under different working conditions is known *a priori*. Given the knowledge of the scheduling variable, the segmentation of the nonlinear process as well as the process input-output data, an LPV model is identified using the EM algorithm which by itself is an iterative optimization procedure. The EM algorithm has been employed in our previous work on switched autoregressive eXogenous systems (SMARX) identification and satisfactory identification results have been achieved (Jin & Huang (2009b,a)). For most of chemical systems, when shifting from one operating point to the other, its dynamic behavior tends to vary gradually. In other words, abrupt switching from one local linear model to another is unlikely. Therefore, we assume smooth transition of the operating points and a smooth validity function is utilized to combine multiple local models to approximate the transition dynamics. In this paper, an exponential function is employed to calculate the validity of each local model under different working conditions. As the segmentation of the process is given a priori, the centers of the exponential functions are fixed and need not be estimated. Therefore, the main tasks involved in the LPV modeling are estimation of the local model parameters as well as the validity width of each local model.

Simulated as well as experimental examples are employed to demonstrate the efficiency of the proposed LPV model identification algorithm. It is shown that, through the illustrated examples, only small excitation signals are required for LPV identification and the local model parameters as well as the validity functions are estimated simultaneously. The remainder of this paper is organized as follows: Section 2 lays out the mathematical formulation for the LPV model identification and a relevant identification procedure is provided. Section 3 shows the identification results on a simple illustrative numerical simulation example adopted from Zhu et al. (2008) as well as an CSTR benchmark process. Section 4 elaborates the design of the experiment conducted on a heated tank process and renders verification results on the identified LPV model. Section 5 draws conclusions and provides our perspective on several implementation issues associated with the algorithm.

4.2 Mathematical Formulation of Nonlinear Modeling Using Multiple Local Models

In the normal operation of a chemical plant, the process may transit among several operating points. Assume that the information regarding the operating conditions of the system can be inferred from the measurement of the *scheduling variable* T , given M different operating points, their relevant local linear models are centered around M constant values, which can be denoted as $T_i, i = 1, 2 \dots M$. Let θ_i be the i th local linear model parameters, the probability of observing Z_k given all the past observations can be calculated as:

$$P(Z_k | Z_{k-1} \dots Z_1) = \sum_{i=1}^M \alpha_{ki} P(Z_k | Z_{k-1} \dots Z_1, \theta_i) \quad (4.1)$$

where α_{ki} in Equation 4.1 represents the the normalized weight given to the probability of \hat{Z}_k being at local model i . To achieve smooth combination of the predictions from M local models, an exponential weighting function is employed and the unnormalized weighting factor for i th local model w_{ki} is calculated by the following equation:

$$w_{ki} = \exp\left(\frac{-(T_k - T_i)^2}{2(o_i)^2}\right) \quad (4.2)$$

where T_k in Equation 4.2 denotes the measurement of the *scheduling variable* at the k th sampling time instant, o_i denotes the validity width of the i th local model. Therefore, the normalized weight α_{ki} can be derived as:

$$\alpha_{ki} = \frac{w_{ki}}{\sum_{i=1}^M w_{ki}} \quad (4.3)$$

In this paper, the centers of M local linear models $T_i, i = 1, 2 \dots M$ are assumed to be known *a priori*. This is a reasonable assumption owing to the fact that the plant is normally operated under several explicitly designed conditions to produce the desired products. The scheduling variable $T_i, i = 1, 2 \dots M$ is assumed to be measured or can be inferred from the plant routine operation data. Under this assumption, the parameters that need to be estimated from the process input-output data $Z_k, k = 1, 2 \dots N$ would be $\Theta = \{\theta_1, \theta_2 \dots \theta_M, o_1 \dots o_M\}$. Using the maximum likelihood principle, the LPV model parameters Θ equals

$$\begin{aligned} \Theta &= \max_{\Theta} \prod_{k=1}^N P(Z_k | Z_{k-1} \dots Z_1) \\ &= \max_{\Theta} \prod_{k=1}^N \sum_{i=1}^M \alpha_{ki} P(Z_k | Z_{k-1} \dots Z_1, \theta_i) \end{aligned} \quad (4.4)$$

As pointed in Jin & Huang (2009b), brute force optimization over Equation 4.4 can be computationally formidable and as a result, a hidden variable I_k is introduced

to represent the identity of the sub-model that takes effect at time point k . Denoting the observed data $Z_k, T_k, k = 1, 2 \dots N$ as C_{obs} and the hidden state $I_k, k = 1, 2 \dots N$ as C_{mis} , the complete data would be $C = \{Z_k, T_k, I_k\}, k = 1, 2 \dots N$. The expectation of the complete data C in the procedure of the EM algorithm is derived as:

$$\begin{aligned}
Q(\Theta | \Theta_{old}) &= E_{I|(\Theta^{old}, C_{obs})} \{ \log P(C_{obs}, I | \Theta) \} \\
&= E_{I|(\Theta^{old}, C_{obs})} \{ \log P(Z_N, Z_{N-1} \dots Z_1, T_N, \dots T_1, I_N \dots I_1 | \Theta) \} \\
&= E_{I|(\Theta^{old}, C_{obs})} \{ \log \prod_{k=1}^N P(Z_k, T_k, I_k | Z_{k-1}, \dots Z_1, T_{k-1}, \dots T_1, I_{k-1}, \dots I_1, \Theta) \} \\
&= E_{I|(\Theta^{old}, C_{obs})} \{ \log \prod_{k=1}^N P(Z_k | Z_{k-1} \dots Z_1, T_k \dots T_1, I_k \dots I_1, \Theta) \\
&\quad \cdot P(I_k | Z_{k-1} \dots Z_1, T_k \dots T_1, I_{k-1} \dots I_1, \Theta) \\
&\quad \cdot P(T_k | Z_{k-1} \dots Z_1, T_{k-1} \dots T_1, I_{k-1} \dots I_1, \Theta) \} \\
&= E_{I|(\Theta^{old}, C_{obs})} \{ \log \prod_{k=1}^N P(Z_k | Z_{k-1}, \dots Z_1, I_k, \dots I_1, \Theta) \cdot P(I_k | o, T_k) \} \cdot P(T_k) \\
&= E_{I|(\Theta^{old}, C_{obs})} \{ \sum_{k=1}^N \log [P(I_k | o, T_k) \cdot P(Z_k | \theta_{I_k}, Z_{k-1} \dots Z_1) \cdot P(T_k)] \} \\
&= \sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log P(I_k = i | o, T_k) \\
&\quad + \sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log P(Z_k | Z_{k-1} \dots Z_1, \theta_i) \\
&\quad + \sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log P(T_k)
\end{aligned} \tag{4.5}$$

where $P(I_k = i | o, T_k)$ in Equation 4.5 equals α_{ki} and $Q(\Theta | \Theta_{old})$ is defined as the expectation over the overall system parameter Θ given the initial/previous estimation results of Θ . In Equation 4.5, $P(Z_k | Z_{k-1} \dots Z_1, T_k \dots T_1, I_{k-1} \dots I_1, \Theta)$ is considered to be equal to $P(Z_k | Z_{k-1} \dots Z_1, I_k \dots I_1, \Theta)$ based on the fact that the observed data at k th time instant Z_k is directly dependent on the data identity $I_k \dots I_1$. The value of the scheduling variable $T_k \dots T_1$ determines the identity of the collected data set. Similarly, $P(I_k | Z_{k-1} \dots Z_1, T_k \dots T_1, I_{k-1} \dots I_1, \Theta)$ is equal to $P(I_k | o, T_k)$ as according to Equation 4.2, the data identity at k th time instant is determined together by the value of T_k as well as the validity width of each local model. $P(T_k | Z_{k-1} \dots Z_1, T_{k-1} \dots T_1, I_{k-1} \dots I_1, \Theta)$ is equal to $P(T_k)$ by a typical assumption that the scheduling variable T_k is independent of the past observations and model parameters. To calculate the parameters, derivative is taken with respect to the relevant components in $Q(\Theta | \Theta_{old})$ function as being expressed in Equation 4.5. For estimation of local model parameters $\theta_i, i = 1, 2 \dots M$, derivative is taken over

the term $\sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log P(Z_k | Z_{k-1} \dots Z_1, \theta_i)$ in Equation 4.5 with respect to $\theta_i, i = 1, 2 \dots M$

$$\frac{\partial \sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log P(Z_k | Z_{k-1} \dots Z_1, \theta_i)}{\partial \theta_i} = 0 \quad (4.6)$$

As a result, the local ARX linear models parameter $\theta_i, i = 1, 2 \dots M$ is derived as:

$$\theta_i^{New} = \frac{\sum_{k=1}^N P(I_k = i | \Theta^{old}, C_{obs}) x_k^T y_k}{\sum_{k=1}^N P(I_k = i | C_{obs}, \Theta^{old}) x_k^T x_k} \quad (4.7)$$

where x_k, y_k in Equation 4.7 represent local ARX models regressor and the process output respectively.

Based on Equation 4.1, the prediction of the identified LPV model under certain working point $\hat{y}_k(T_k)$, which is denoted by the scheduling variable T_k , can be derived as the weighted average summation of the prediction from each local model,

$$\hat{y}_k(T_k) = \sum_{i=1}^M \alpha_{ki} \hat{y}_{ki} \quad (4.8)$$

where \hat{y}_{ki} in Equation 4.8 represents the prediction from i th local model at k th time instant. With each local model being formulated in ARX model form, the Equation 4.8 can be further written as

$$X_k^T \hat{\theta}_k(T_k) = \sum_{i=1}^M \alpha_{ki} X_k^T \hat{\theta}_i \quad (4.9)$$

where X_k in Equation 4.9 denotes the regressor that is constructed as per the ARX model structure. From Equation 4.9, it can be seen that the identified LPV model is able to predict the process output at any given value of the scheduling variable T .

To estimate the process noise variance σ^2 , derivative is taken over the term $\sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log P(Z_k | Z_{k-1} \dots Z_1, \theta_i)$ with respect to σ and the resulted equation for estimating σ^2 is derived as:

$$(\sigma^{New})^2 = \frac{\sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) (y_k - x_k \theta^{New})^T (y_k - x_k \theta^{New})}{\sum_{k=1}^N \sum_{i=1}^M P(I_k = i | C_{obs})} \quad (4.10)$$

As for the local model validity width $o_i, i = 1, 2 \dots M$, due to the usage of exponential function in combing the local models, it cannot be derived in an analytical form when minimizing the term $\sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log P(I_k = i | o, T_k)$ in Equation 4.5. The mathematical formulation of the optimization problem in the search for optimal $o_i, i = 1, 2 \dots M$ values can be expressed as:

$$\begin{aligned} & \max_{o_i, i=1, 2 \dots M} \sum_{k=1}^N \sum_{i=1}^M P(I_k = i | \Theta^{old}, C_{obs}) \log P(I_k = i | o, T_k) \\ & S.t. \quad o_{min} \leq o_i, i = 1, 2 \dots M \leq o_{max} \end{aligned} \quad (4.11)$$

where $P(I_k = i | \Theta^{old}, C_{obs})$ in Equation 4.11, as being explained before, represents the probability of k th data point belonging to i th sub-model. $\log P(I_k = i | o, T_k)$ is equivalent to $\log \alpha_{ki}$ which can be calculated from Equation 4.3. A nonlinear numerical optimization method is required in order to calculate the optimal $o_i, i = 1, 2 \dots M$ which is bounded by o_{min} (the lower bound for $o_i, i = 1, 2 \dots M$) and o_{max} (the upper bound for $o_i, i = 1, 2 \dots M$) for each local model. In this paper, we adopt a constrained nonlinear optimization function named 'fmincon' provided by MATLAB as to calculate the best width of each Gaussian weighting function associated with relevant local models.

4.2.1 Discussion on one practical implementation issue

Robustness VS Optimality

The LPV model identification method introduced in this paper is essentially an iterative optimization procedure in which the local model parameters $\theta_i, i = 1, 2 \dots M$ along with the local model validity width $o_i, i = 1, 2 \dots M$ are identified simultaneously at each iteration. By gradually increasing the accuracy of the estimation at each iteration, it is expected that the estimated parameters will converge to the optimal values after finite number of iterations. However, when the nonlinear optimization scheme adopted in searching for the optimal $o_i, i = 1, 2 \dots M$ slips into a local maxima, the parameter estimation results may be severely skewed and the prediction from the identified LPV model can rarely capture the nonlinear process dynamics. Therefore, before applying the proposed algorithm to the real process data, it is of benefit to narrow the searching space for parameters by incorporating additional information such as the process priori knowledge or preliminary estimation results from the other estimation methods. It was found that the validity width is the most effective parameter influencing the algorithm robustness. For example, to increase the robustness of the algorithm, the upper bound for local models validity width o_{max} should be reduced (in an extreme case when $o_{max} = 0$, the local models will be identified simply based on the corresponding local experiment data). Therefore, the upper bound o_{max} can be used as a tuning parameter through which the robustness of the algorithm can be adjusted.

4.3 Simulations

4.3.1 A Numerical Simulation Example

A first order process with varying model parameters is utilized here to demonstrate the efficiency of the proposed LPV model identification method. This simple process has been adopted by Zhu & Xu (2008) as an illustrative simulation example. The

process is described by the following equation:

$$G(s, w) = \frac{K(T)}{\tau(T)s + 1} \quad (4.12)$$

where both the process gain $K(T)$ and the process time constant $\tau(T)$ are nonlinear function of the *scheduling variable* T . The specific nonlinear relation is expressed as follows:

$$K(T) = 0.6 + T^2, T \in [1, 4] \quad (4.13)$$

$$\tau(T) = 3 + 0.5T^3, T \in [1, 4] \quad (4.14)$$

Apparently, over the whole operating range of the process, the process gain as well as the time constant change dramatically and one single linear process model identified under one specific working condition would hardly capture the dynamics of the process throughout its complete operating range. As a result, a nonlinear model or an LPV model is required here in order to sufficiently describe the behavior of the process under various operating conditions (or various T values in other words).

The proposed LPV modeling method is employed and it is assumed that the process is only to be tested around $T_1 = 1, T_2 = 2.25$ and $T_3 = 4$ with Random Binary excitation signals with small magnitude. To realize the transition from one operating point to the other, *scheduling variable* T is gradually increased by fixed incremental size and no additional excitation signal is added. In the simulation, white noise with a variance of 0.015 is added in the simulated output to account for unknown process noise.

After collecting the input-output data from the simulation, the proposed LPV model identification method is applied and the identified results are shown in Figure 4.1. Here, to better test the validity of the identified model, step responses of the identified model under randomly selected T values are calculated and they are compared with the step response of the true process model.

Comparison result displayed in Figure 4.1 shows that the identified LPV model not only can well capture the process dynamics under the training operating conditions, but also perform well in capturing the process dynamics under other operating points. This confirms the effectiveness of the identified LPV model in approximating the global process dynamics throughout the whole operating range. Figure 4.2 provides a weighting map of each local model under different T values. Based on this calculated weighting map as well as Equation 4.9, model predictions can be calculated under all the T values chosen from the operating range.

Finally, the identified local ARX model parameters are given in Table 4.1. It is noticed that in comparison with the true model parameters, the estimated ones are close to the real values.

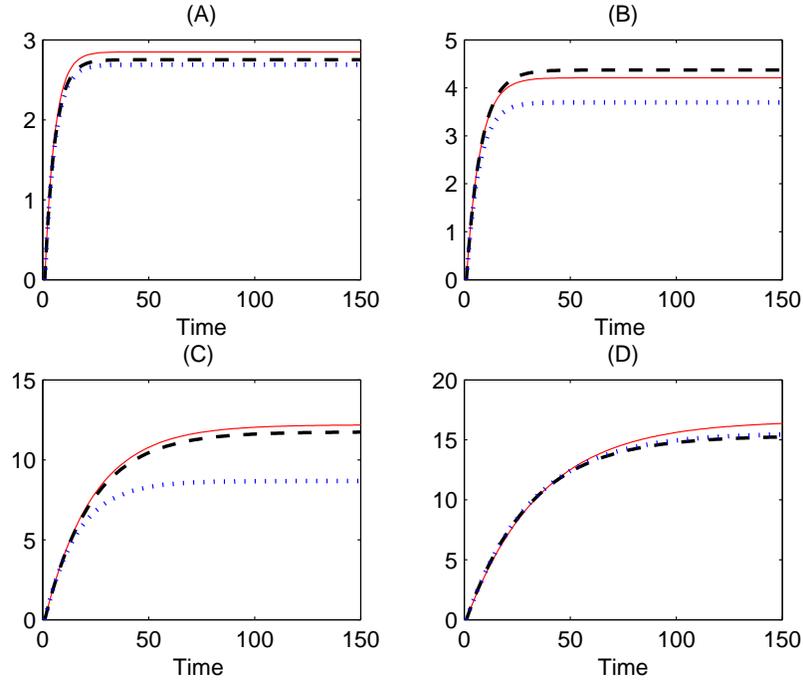


Figure 4.1: Comparison of the step response between identified and true models under different T values (Red solid line: step response of the true model, black dashed line: identified LPV model using the proposed method, blue dotted line: identified LPV model using the interpolation method given in Xu et al. (2009)) (A): system step response when $T = 1.5$ (B): system step response when $T = 1.9$ (C): system step response when $T = 3.4$ (D): system step response when $T = 4$

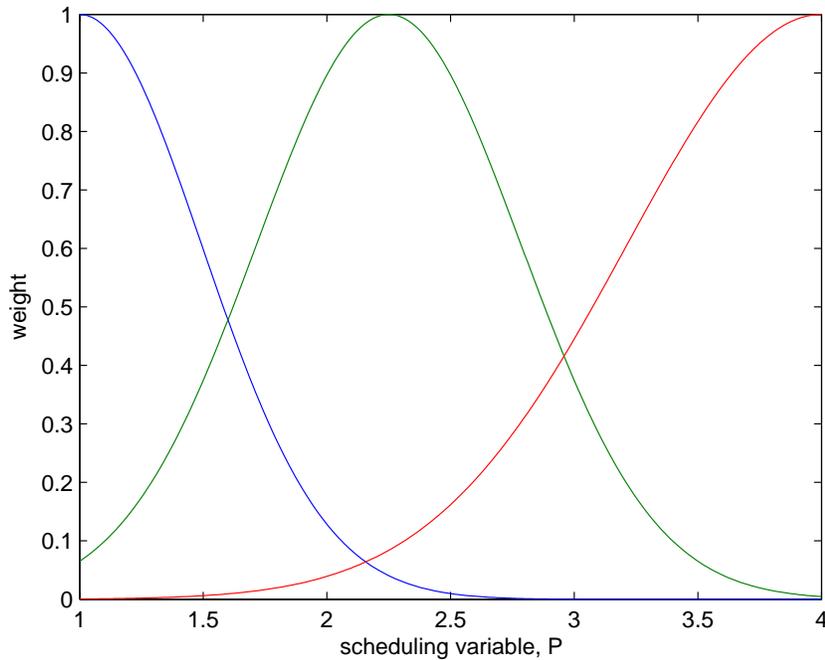


Figure 4.2: Weight of each local model at different operating points

Table 4.1: True system parameters and estimated parameters

θ_1	$\hat{\theta}_1$	θ_2	$\hat{\theta}_2$	θ_3	$\hat{\theta}_3$
0.7515	0.7649	0.8914	0.8845	0.9718	0.9687
0.3976	0.4607	0.6152	0.6149	0.4676	0.5032

4.3.2 Continuous Stirred Tank Reactor

As a commonly used production unit in chemical and petro-chemical processes, the continuous stirred tank reactor (CSTR) has been widely investigated and accepted as a benchmark process for nonlinear process modeling, optimization and control. In this section, we employ an exothermic CSTR with irreversible reaction $A \rightarrow B$. Based on the mass as well as heat balance of the process, the first principle model can be derived as (Xu et al. (2009); Senthil et al. (2006)):

$$\frac{dC_A(t)}{dt} = \frac{q(t)}{V}(C_{A0}(t) - C_A(t)) - k_0 C_A(t) \exp\left(\frac{-E}{RT(t)}\right) \quad (4.15)$$

$$\begin{aligned} \frac{dT(t)}{dt} = & \frac{q(t)}{V}(T_0(t) - T(t)) - \frac{(-\Delta H)k_0 C_A(t)}{\rho C_p} \exp\left(\frac{-E}{RT(t)}\right) + \\ & \frac{\rho_c C_{pc}}{\rho C_p V} q_c(t) \left\{1 - \exp\left(\frac{-hA}{q_c(t)\rho C_p}\right)\right\} (T_{c0}(t) - T(t)) \end{aligned} \quad (4.16)$$

where the explanations and their corresponding steady state values for the parameters of Equation 4.15 and Equation 4.16 are given in Table 4.2. As can be seen from Table 4.2, coolant flow rate q_c is defined as the process input while the concentration of component A and the reactor temperature T are treated as the outputs. By manipulating the flow rate of the coolant, the reactor temperature changes accordingly which leads to the change of the reaction kinetics. As a result, products with different concentration of component A can be produced. In this paper, we assume that the concentration of the outlet reagent A is of interest and it is controlled through manipulating the coolant flow rate. Therefore, a single input single output (SISO) model between the coolant flow rate and the reagent A concentration is built to approximate the process dynamics across the whole operating range. Measurement white noise with the magnitude of about 0.5% of the noise-free output in power is added to the simulated output concentration and the input-output data of the process are given in Figure 4.3.

As can be seen from Figure 4.3, five operating points are considered and the coolant flow rate varies from 97L/min to 109L/min. During the transition period between different operating points, the coolant flow rate increases by a fixed step size and no additional excitation signal is added. Without knowing the parameters of each local model, applying the proposed algorithm to the input-output data and the identification results are obtained as shown in Figure 4.4 and Figure 4.5.

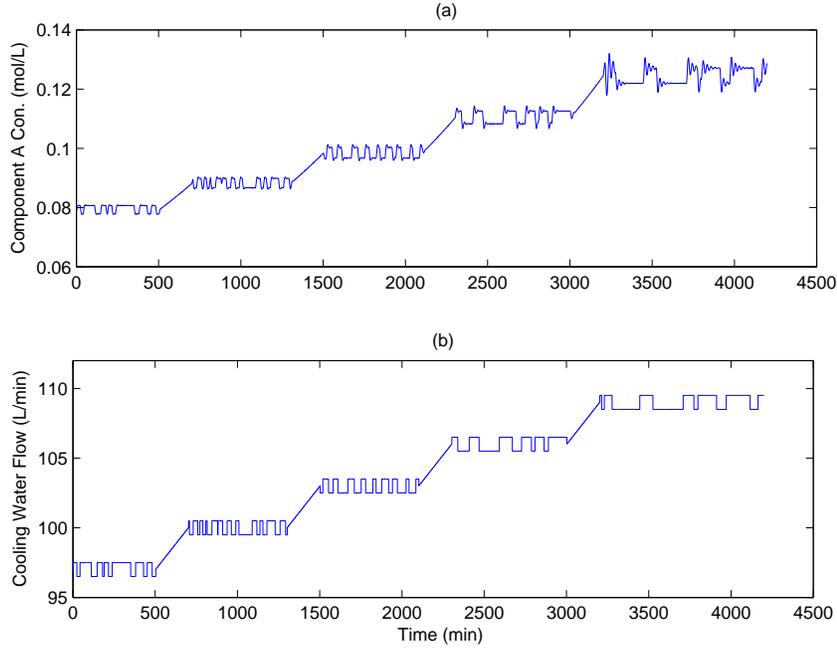


Figure 4.3: Input-output excitation data, (a): outlet reagent A concentration (b): cooling water flow rate

Table 4.2: CSTR model parameters and their steady state values

parameters	steady state value
production concentration of Component A, C_A	$output_1$
temperature of the reactor, T	$output_2$
feed Concentration of Component A, C_{A0}	1mol/L
feed temperature, T_0	350.0 K
specific heats, C_p, C_{pc}	1 cal/(g K)
liquid density, ρ, ρ_c	1×10^3 g/L
heat of reaction, $-\Delta H$	-2×10^5 cal/mol
activation energy term, E/R	1×10^4 K
reaction rate constant, k_0	$7.2 \times 10^{10} min^{-1}$
heat transfer term, hA	7×10^5 cal/(min K)
the reactor volume, V	100L
inlet coolant temperature, T_{c0}	350.0 K
process flow rate, q	100 L/min
coolant flow rate, q_c	input

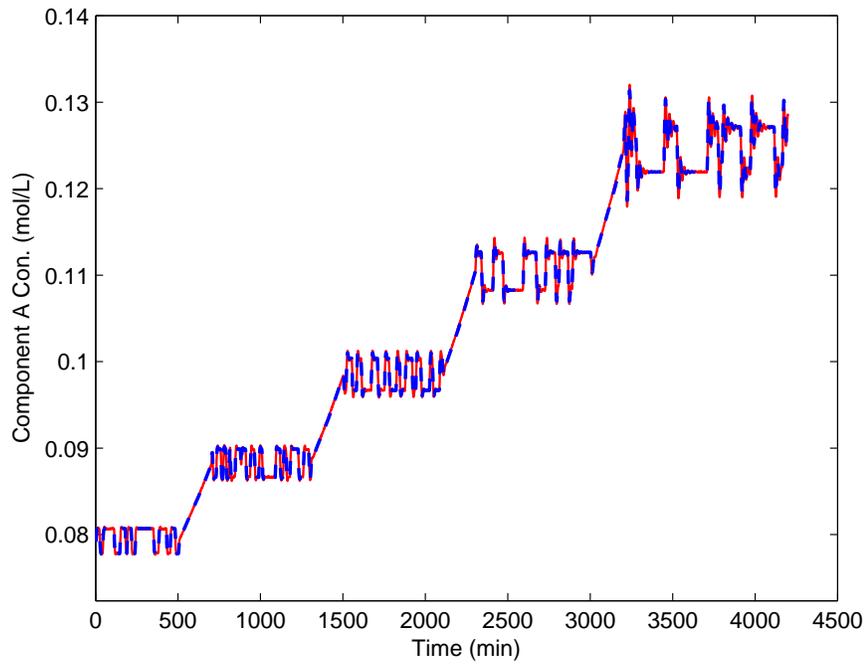


Figure 4.4: Self validation of the identified CSTR LPV model, $MSE=9.7535 \times 10^{-5}$. Red solid line is the real process output and the blue dashed line is the simulated output from the identified LPV model

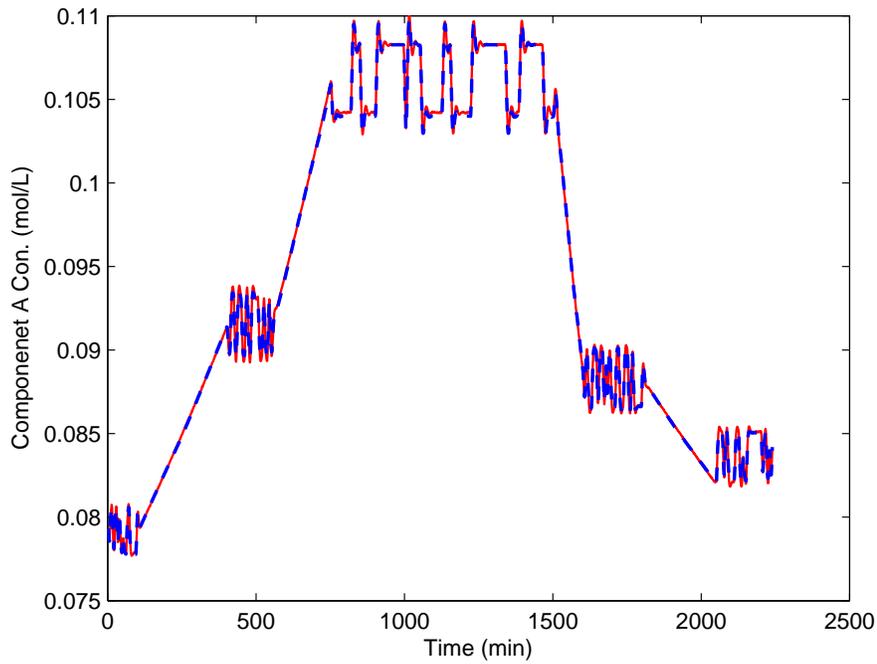


Figure 4.5: Cross validation of the identified CSTR LPV model, $MSE=4.4838 \times 10^{-5}$. Red solid line is the real process output and the blue dashed line is the simulated output from the identified LPV model

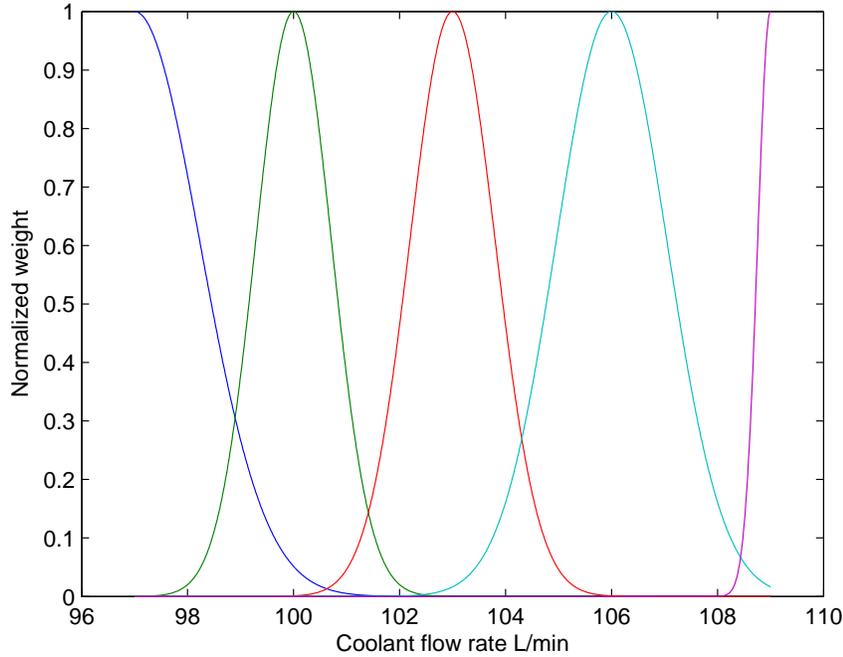


Figure 4.6: Validity function of each local model under different working conditions (coolant flow rates)

To achieve a better validation of the identified model, the data used for cross validation are generated under different coolant flow rates from that of the training data. For example, in the cross validation, we test the performance of the identified model under $q_c = 105$, which is not considered in the model training data. The validation results are satisfactory and good match between the prediction from the identified model and the real simulated data is achieved both in self validation (Figure 4.4) and cross validation (Figure 4.5). The validity/weighting values calculated for each local model with different coolant flow rates are given in Figure 4.6.

Based on the weighting map shown in Figure 4.6, the predictions from each of the five local linear models can be effectively combined and the overall prediction of the LPV model is obtained under different coolant flow rates.

4.4 Experimental Verification & Discussion

4.4.1 Process Description and Identification Results

To further verify the capability of the proposed algorithm, nonlinear process identification experiment is designed and performed on a pilot-scale heated tank system. The simplified process piping and instrumentation diagram (P & ID) is shown in Figure 4.7.

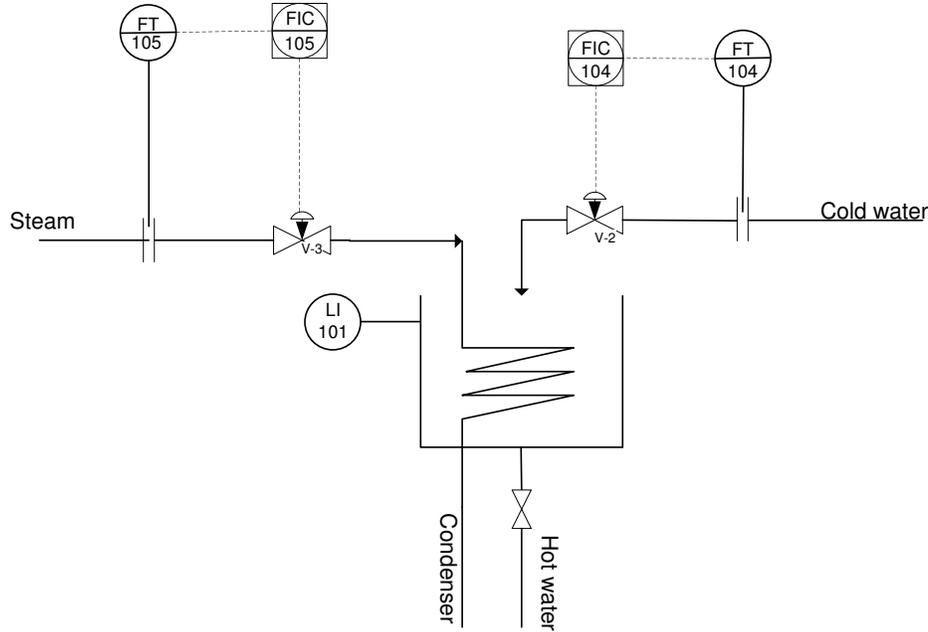


Figure 4.7: Simplified heated tank system piping and instrumentation diagram

As can be seen from Figure 4.7, the saturated steam is used to heat the water in the tank. By passing the steam through the coil, energy travels from the steam to the water through heat exchange. As a result, given a fixed volume of water in the tank, the temperature of the water is mainly influenced by the flow rate of the steam. Considering that the process dynamics from the steam flow rate (input) to the tank water temperature (output) is of interest, to identify the process model, the flow rate of the steam is manipulated as the excitation signal and the water temperature change is recorded as the process response. In the configuration as shown in Figure 4.7, the flow rate of the steam and the cold water are controlled by the controllers tagged as *FIC105* and *FIC104* and the level of the tank is measured through a field mounted level transmitter. Cascade control scheme is applied to the tank level control in which a level controller, considered as a master controller, calculates the set point for the cold water flow rate controller.

With the same system configuration, for instance, same tank exit manual valve position, etc., under different water levels, the transfer function from the steam flow rate to the water temperature could vary accordingly. In other words, the process is essentially nonlinear and relevant local linear approximation around each tank level cannot be applied to the other levels. Owing to the significant effect of the water level on the process dynamics, the tank level is denoted as the scheduling variable of the system. Four different operating points are selected and random binary excitation signal (RBS) is designed for the steam flow rate. The design parameters for the experiment are shown in Table 4.3:

Operating points, $T_k, k = 1, 2, 3, 4$	Steam flow rate excitation signal
0.15 (m)	10 kg/hr - 15 kg/hr
0.25 (m)	10 kg/hr - 15 kg/hr
0.35 (m)	10 kg/hr - 15 kg/hr
0.39 (m)	10 kg/hr - 15 kg/hr

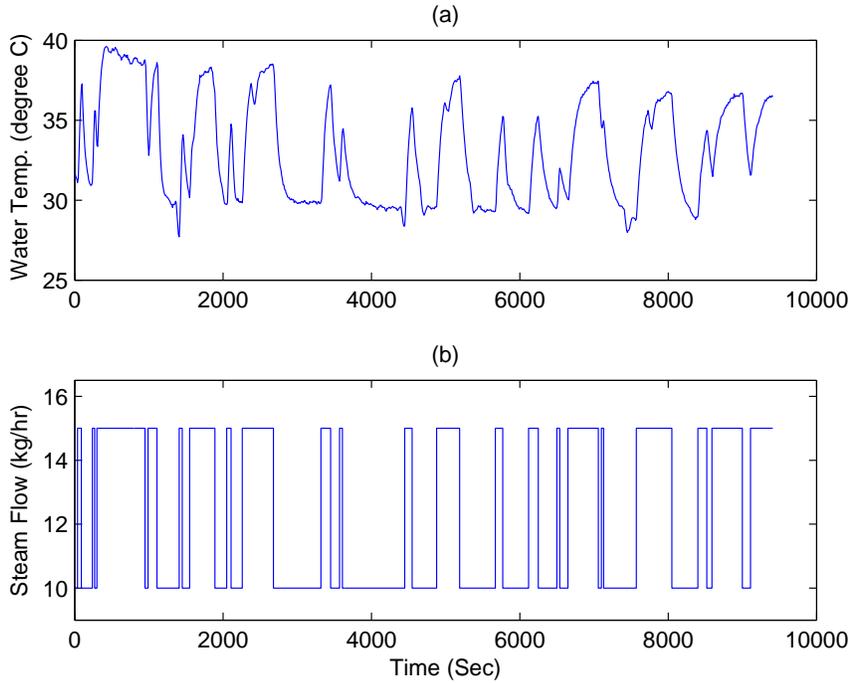


Figure 4.8: Heated tank process input-output data (a) water temperature in the tank, process output; (b) steam flow rate, process input

The transition between different operating points is operated in a way that the tank level is slightly increased by a fixed step over the transition period and it reaches the next operating level when the transition period is over. The process input - output data along with the scheduling variable (which is tank level in this case) are given in Figure 4.8 and Figure 4.9.

As pointed out in the discussion section on how to set the tuning parameter o_{max} values under different collected process data sets, when the process measurement is considered to be noisy, to increase the robustness of the algorithm, relatively small values are going to be assigned to o_{max} and as a result, it is expected that the optimality of the algorithm could be compromised. In this calculation, the o_{max} values for each of the local model are set to be $0.5^{\circ}C$ and identification results are shown in Figure 4.10.

The estimated LPV model performs well in self-validation test which indicates the efficiency of the proposed method. However, to further test the identified LPV

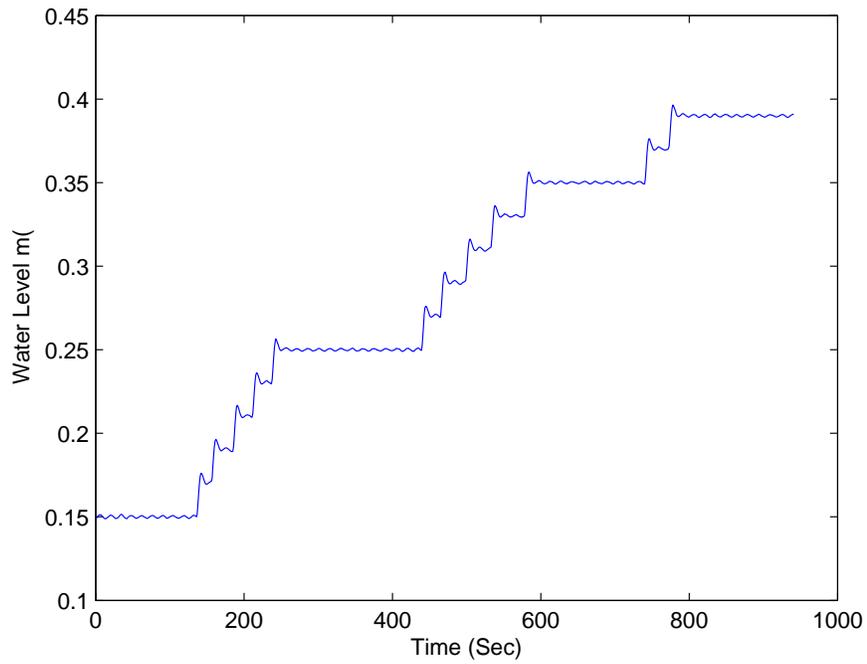


Figure 4.9: Heated tank process scheduling variable

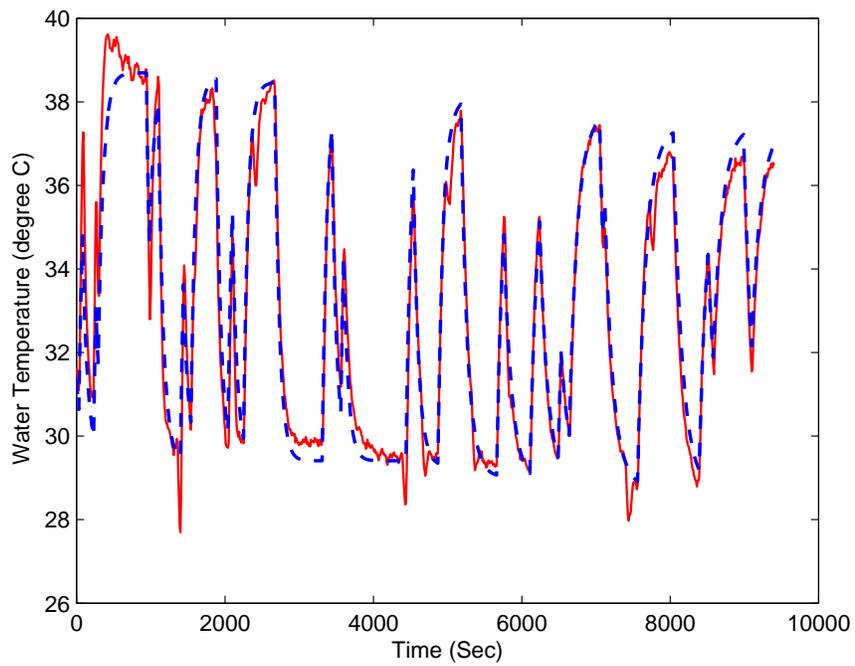


Figure 4.10: Self-validation result, the fitting percentage equals 93.6%. Red solid line is the collected process data , blue dashed line is the simulated output of the identified LPV model

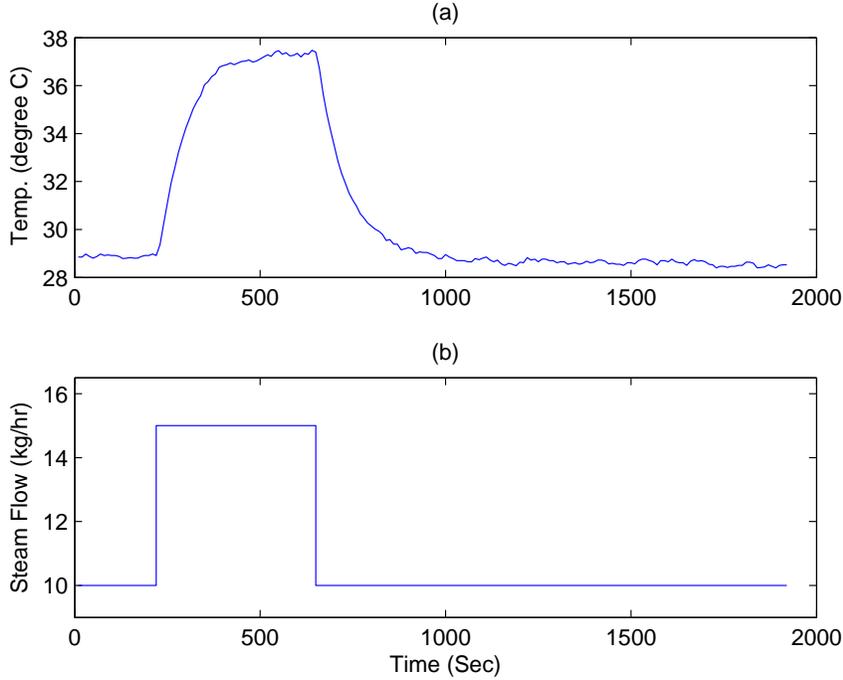


Figure 4.11: Cross-validation input-output data, (a) tank temperature, process output, (b) steam flow rate, process input

model, cross-validation is also performed in an effort to further evaluate the quality of the identified model. Here, simple step tests with steam flow rate changes between 10kg/hr and 15kg/hr are conducted under the tank level equals 0.3 meter. The process input-output data are given in Figure 4.11

Setting the same initial condition as the cross-validation data, the comparison between the prediction from the identified model against the collected process validation data is plotted in Figure 4.12

The tank level around which the step tests are performed is 0.3 meters, which has not been chosen as one of the local operating points in the training data as given in Figure 4.9. To vividly illustrate the nonlinearity of the process, predictions from the locally identified models under different levels are compared with the collected process data when the water level equals 0.3 meters. The comparison results are given in Figure 4.13. It is noticed that due to the nonlinearity brought by the change of the process operating condition (water tank level change in this case), the identified local models can not satisfactorily capture the process dynamics when at level equals 0.3 meters.

To test the capability of the identified nonlinear model in describing the dynamic behavior of the process, the data shown in Figure 4.11 are passed through the identified model and relevant validation result is given in Figure 4.12. It can be seen from Figure 4.12 that the real process data and the model prediction is in good

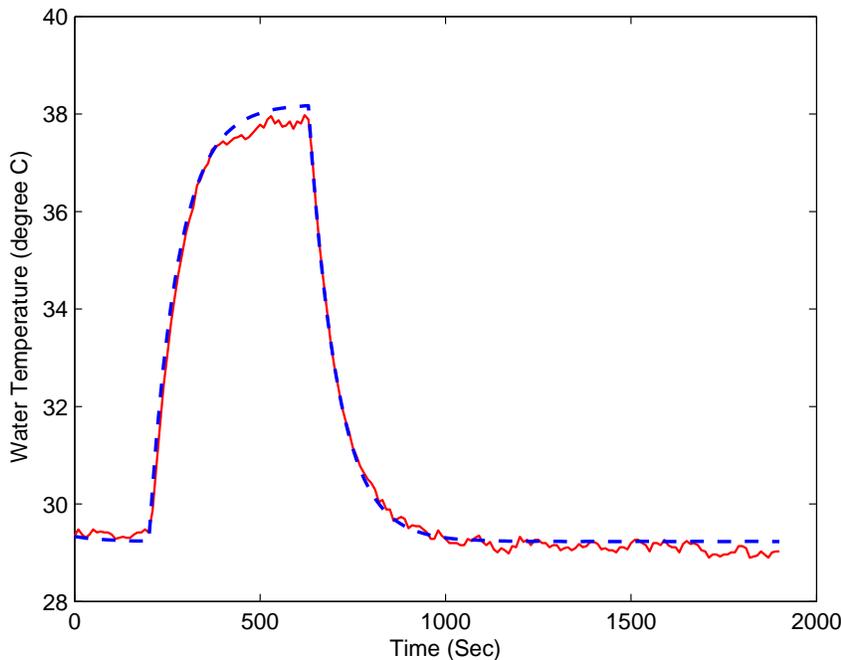


Figure 4.12: Cross-validation result, the fitting percentage equals 97%. Red solid line is the collected process data , blue dashed line is the simulated output of the identified LPV model

agreement with each other which effectively confirms the efficiency of the algorithm.

4.4.2 Comparative Study With An Existing LPV Modeling Method

Xu et al. (2009) and Zhu (2008) have proposed a nonlinear process modeling method in which an LPV model is identified based on the interpolation of each local model obtained around its relevant operating condition (Xu et al. (2009); Zhu & Xu (2008)). It has been shown that by appropriately combing all the local models using certain type of smooth validity functions, the identified LPV model is able to predict the nonlinear process behavior given inputs as well as the information on the process operating conditions. Owing to the similarity of the modeling philosophy, we perform a comparative study between the proposed method and the method that has been given by Xu et al. (2009); Zhu & Xu (2008) in this section. For notation simplicity, the LPV model identification method proposed in Xu et al. (2009); Zhu & Xu (2008) is referred to as the comparative method afterwards in this section.

The same data used in the previous section will be applied to the comparative method. It consists of the training data as shown in Figure 4.8 and the testing data as in Figure 4.11. As pointed in the introduction section, one of the main drawbacks associated with the method proposed in Zhu & Xu (2008); Xu et al. (2009)

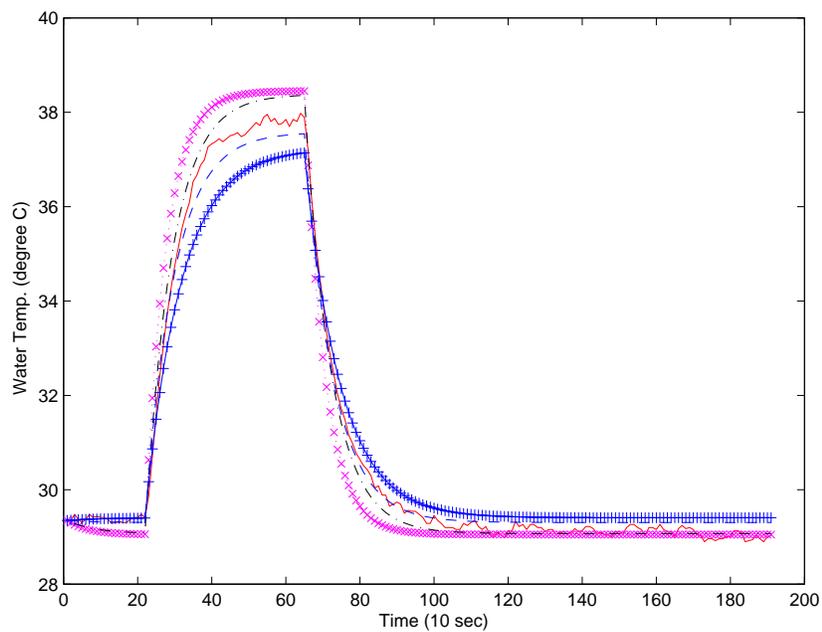


Figure 4.13: Comparison among the predictions from different local models and the collected real process data, pink cross solid line: predictions from the model identified at level = 0.15 meters, black dash dotted line: predictions from the model identified at level = 0.25 meters, red solid line: real process response at level = 0.3 meters, blue dashed line: predictions from the model identified at level = 0.39 meters, blue plus line: predictions from the model identified at level = 0.35 meters

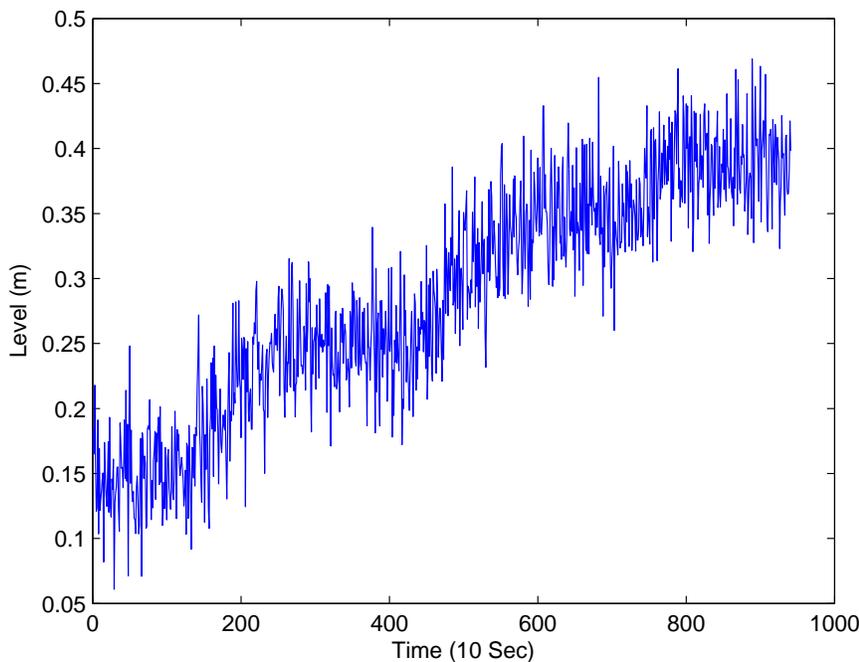


Figure 4.14: Noisy tank level measurement, measurement noise with variance of 0.001 is added to the original level measurement

is its dependence on the locally identified models. Although it is assumed that the scheduling variable/variables through which the operating condition of the process being inferred can be measured or observed in this paper, however, given noisy measurement of the scheduling variable/variables, the inference of the process operating condition can be quite difficult or even misleading under certain circumstance. In other words, the selection of the data set for local models identification may not be that easy as it is even when the scheduling variable/variables measurements are available. Taking the heated tank process for example, measurement noise with the variance of 0.001 is added to the original level measurement and the resulted level readings are shown as in Figure 4.14.

As can be seen from Figure 4.14, when the scheduling variable measurement is noisy, selection of the local data sets by hand can be challenging which makes the inappropriate segmentation of the process data more likely. To demonstrate the dependence of the comparative method on the segmentation of the process data, comparative study between the proposed method and the comparative method is conducted in this section. Two scenarios are contrived in which the first scenario provides the accurate local models for the comparative method to start with while in the second scenario, the local models are estimated with poorly segmented data (data mixed with the local data set and its neighbouring transition data) due to the noisy level measurement as shown in Figure 4.14. To ensure the fairness of the

Table 4.4: Comparison of estimated parameters using different methods with outliers in the data set

Method	SV MSE (Scenario 1)	CV MSE (Scenario 1)
The Comparative Method	0.3924	0.0349
The Proposed Method	0.4081	0.0349
Method	SV MSE (Scenario 2)	CV MSE (Scenario 2)
The Comparative Method	0.6972	0.0413
The Proposed Method	0.4342	0.037

comparison, the same global data set is used for both of the methods and the level measurement data shown in Figure 4.9 and Figure 4.14 are used as the scheduling variable measurement for Scenario 1 and Scenario 2 respectively. Under two different scenarios, the process data with different scheduling variable measurements are passed through the algorithms and the relevant validation results are summarized in Table 4.4

As can be seen from Table 4.4, under Scenario 1 in which appropriate local models are provided, both of the methods are able to achieve satisfactory identification results while, when the prerequisite on the quality of the local models is not satisfied for the comparative method due to the various reasons, e.g. poor selection of the local data set caused by noisy scheduling variable measurement in this case, it delivers lower identification performance and the proposed method outperforms it under this situation owing to the fact that the local models as well as the smooth combination functions are estimated simultaneously using the whole global data set in the proposed method.

4.5 Discussion & Conclusion

Modeling of nonlinear process with multiple operating conditions is considered in this paper. By incorporating process information such as scheduling variables, the number of operating point that the system is likely to be operated on, the model order of each local model identified around its corresponding operating point, an LPV model is identified for the nonlinear process and it has been demonstrated that within the specified working range, the identified model provides satisfactory approximation of the process dynamics. To further validate the effectiveness of the proposed identification method, an experiment is performed on a pilot-scale setup and it is demonstrated that the proposed LPV modeling method can be applied to the collected experimental data to effectively identify an LPV model for nonlinear process. Some issues may rise when it comes to real industrial process applications and we would like to provide our perspective on circumventing some of those issues

- (1) How to decide the excitation signal magnitude for each local model? We did not

quantitatively specify the magnitude of the excitation signal in the work that has been done so far; however, as identifying a local linear model around certain operating point has been a relatively sophisticated practice after decades of development, design information such as the magnitude of the input excitation signal can well be obtained from the process operation knowledge. (2) How to model the process dynamics outside the operating range of the training data? It would be difficult for the identified LPV model to predict the system behavior outside the operating range of the training data set. Therefore, we suggest collecting the input-output data from a wider operating range which covers all the potential prediction region. If overly wide operating range excitation is not allowed due to safe process operation or cost consideration, then the identified local linear model which locates closest may be chosen as the prediction model. (3) How to determine the operating points over which the local identification experiments are to be performed. In this paper, we assume that the operating points at which the local models are identified have already been specified in such a way that it is optimal for identifying the nonlinear process. However, relevant research work on optimal working points design for nonlinear process modeling using LPV models has already been conducted by researchers in the past few years such as the research performed by Khalate et al. (2009). These algorithms can be well employed in practice when process priori knowledge is not sufficient to determine appropriate operating points for the construction of LPV models.

Chapter 5

Summary and Conclusions

5.1 Summary of This Thesis

This thesis is concerned with the identification of switched systems. The popularity of the switching phenomena happening in industrial process along with the relevant challenges it poses for process control and monitoring motivate us to investigate and try to solve the modeling of different types of switched systems. The work is confined to the switched ARX systems as we believe that it greatly simplifies the switched system identification problems while, at the mean time, preserves the applicability of the proposed identification methods to the real experimental or even industrial scale processes. Different types of switched ARX systems are considered. The Piecewise ARX system, in which the validity region of each sub-system is separated by the regressor hyperplane, is considered owing to its capability of describing a large number of processes by switching among different ARX sub-systems. To immunize the identification algorithm from the negative influence of the abnormal data points that may be present in the data set, a robust EM algorithm is developed by expressing the noise distribution in a contaminated Gaussian distribution form. For switched systems with random switching behavior, the robust EM algorithm is applied directly for estimating the system parameters as well as the cluster identity of each data point. Moreover, for the PWARX system identification, a robust identification procedure which consists of robust EM algorithm block (for local ARX models parameter estimation), data clustering block and sub-system region partition block (for calculation of the boundaries among different sub-systems) is put forward and its efficiency in identifying the PWARX system parameters through the data set with/without outliers is demonstrated. Simulated numerical examples as well as experiment performed on pilot-scale setup are used to confirm the capabilities of the proposed identification methods in identifying the PWARX systems and switched ARX systems with random switching behavior.

The identification methods given in Chapter 2 do not take the switching dynam-

ics into the consideration which, under certain circumstances, may exhibit complex switching behavior or it is known that the system switches in a Markov model fashion. The identification problem of switched Markov ARX systems (SMARX) is evaluated in Chapter 3 and an identification method built upon the EM algorithm is presented in which the discrete switching dynamics along with the continuous process dynamics are estimated simultaneously. The benefits of simultaneous estimation of the continuous dynamics and discrete switching dynamics are confirmed and demonstrated from the comparison of different switched system identification methods in terms of their parameter estimation accuracy. The identification methods employed in the comparison include the proposed method in Chapter 3, the method for general switched ARX system identification given in Chapter 2, the switched system identification method proposed by Nakada et al. (2005).

Chapter 4 moves the discussion into a relatively different domain in which the investigated switched system does not have immediate switching among different local models. Instead, the switching that occurs requires certain amount of time before the transition is finished. Such kind of gradual switching has been commonly experienced in the process industry due to the necessity of operating condition change from time to time. The linear parameter varying model is employed in describing the switched process and the LPV model identification problem is formulated and solved based on the assumption that certain scheduling variable(s) through which the operating mode of the process is indicated can be measured or inferred from the measurement. Without knowing the local ARX models parameters and the validity region of each local model, the process input-output data along with the scheduling variable measurement (sometimes, the scheduling variable can be the input, output or any other process variables) are analyzed and passed through the algorithm. Simple nonlinear numeric model along with an example of nonlinear CSTR process are utilized to verify the efficiency of the proposed method. The identification results demonstrate the capability of the proposed identification method in identifying the local models parameters and their validity regions. By combining all these identified local models based on their respective validity region, a global nonlinear process model is obtained and it can be used for the prediction of the process behavior under different operating conditions.

5.2 Directions for Future Work

Up to this point of the thesis, we have been concentrating on the identification of switched ARX systems which, are worthwhile to investigate given their potential applications in nonlinear process modeling, process abrupt change detection as well as advanced process control. In this section, we would like to share our perspective

on the fields and directions that are worthy of future investigation.

1. Extending the discussion to other local linear model structures. As being pointed at the beginning, one of the basic assumptions made throughout the thesis is that all the sub-systems share the same model structure: ARX model. Although it has been proved that ARX model with sufficiently high order is capable of approximating any process dynamics (Ljung (1987)), however, being aware of the complexity that the real process can be, ARX model structure may not be the optimal choice compared with other linear model structures such as Box-Jenkins (BJ) model, Output-error (OE) model or even linear state space model. Therefore, it will be of beneficial to assign different model structures to different local models accordingly based on the *priori* information of the process.

2. Considering the switched systems with linear as well as simple nonlinear sub-systems. So far, it is assumed that all the sub-models of the switched systems are linear. However, simple nonlinear systems can estimate the local process dynamics more accurately than its linear counterpart due to the fact that all real systems are essentially nonlinear. It is expected that, owing to the introduction of simple nonlinear models for the process under certain operating conditions, the identified switched system is able to approximate the process with better accuracy while less number of local models are required.

3. Applying the general EM algorithm to nonlinear process modeling using multiple local models. In Chapter 4, the nonlinear process modeling problem is considered and it is formulated as LPV model identification problem. The proposed identification method is built upon the regular EM algorithm in which the direct derivative is taken when searching for the optimal parameters in each iteration. Although the direct derivative facilitates the convergence of the algorithm, however, it makes the whole algorithm more sensitive to the noise level. In the general EM algorithm, instead of taking the direct derivative, it only searches for the new set of parameters through which the expectation of the complete data set can be increased (not optimally increased). This is helpful in enabling the algorithm to be less sensitive to the data noise although, as the expense, it may take more time for the algorithm to converge.

4. Estimating the number of sub-models automatically from the data set. It is assumed all over the thesis that the number of sub-models of the switched system is known in advance either from the process knowledge or from the existing estimation algorithm. However, it may be desirable to incorporate the estimation block for the number of the sub-systems into the switched system identification procedure so that by iteratively estimating the parameters and the local model number, the switched system parameters can be more accurately estimated.

Bibliography

- H. Akaike (1974). ‘A new look at the statistical model identification’. *IEEE Transactions on Automatic Control* **19**(6):716–723.
- J. S. Albuquerque & L. Biegler (1996). ‘Data Reconciliation and Gross Error detection for dynamic systems’. *AIChE Journal* **42**(10):2841–2856.
- B. Aufderheide & B. W. Bequette (2003). ‘Extension of dynamic matrix control to multiple models’. *Computers & Chemical Engineering* **27**(8):1079 – 1096.
- B. Bamieh & L. Giarr (2002). ‘Identification of linear parameter varying models’. *International Journal of Robust & Nonlinear Control* **12**(9):841 – 853.
- A. Baneqee & Y. Arkun (1998). ‘Model predictive control of plant transitions using a new identification technique for interpolating nonlinear models’. *Journal of Process Control* **8**:441–457.
- A. Banerjee, et al. (1997). ‘Estimation of Nonlinear Multiple Systems Using Linear Models’. *AIChE Journal* **43**(5):1204 – 1226.
- P. I. Barton, et al. (2000). ‘Optimization of hybrid discrete/continuous dynamic systems’. *Computers & Chemical Engineering* **24**:2171–2182.
- P. I. Barton & C. K. Lee (2004). ‘Design of process operations using hybrid dynamic optimization’. *Computers & Chemical Engineering* **28**:955–969.
- P. I. Barton & C. C. Pantelides (1994). ‘Modeling of combined discrete/continuous processes’. *AIChE Journal* **40**(6):966–979.
- A. Bemporad & G. Ferrari-Trecate (2000). ‘Observability and Controllability of piecewise affine and hybrid systems’. *IEEE Transactions on Automatic Control* **45**(10):1864–1876.
- A. Bemporad, et al. (2005). ‘A bounded error approach to piecewise affine system identification’. *IEEE Transactions on Automatic Control* **50**(10):1567–1580.

- K. P. Bennett & O. L. Mangasarian (1994). ‘Multicategory discrimination via linear programming’. *Optimization Methods and Software* **3**(1-3):27–39.
- C. Biernacki, et al. (2003). ‘Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models’. *Computational Statistics and Data Analysis* **41**(3):561–575.
- R. Bindlish (2003). ‘Model order selection for process identification applied to an industrial ethylene furnace’. *Journal of Process Control* **13**(6):569–577.
- S. Chitralakha, et al. (2009). ‘Comparison of Expectation-Maximization based parameter estimation using Particle Filter, Unscented and Extended Kalman Filtering techniques’. In *15th IFAC Symposium on System Identification*, Saint-Malo, France.
- O. Costa, et al. (2004). *Discrete-Time Markov Jump Linear Systems*. Springer.
- V. M. de Souza (2007). *Markovian Jump Linear Systems Via Convex Optimization*. Ph.D. thesis, Politecnico Di Milano.
- A. P. Dempster, et al. (1977). ‘Maximum likelihood from incomplete data via the EM algorithm’. *Journal of the Royal Statistical Society, Series B* **39**(1):1–38.
- A. Doucet, et al. (2001). ‘Particle Filters for State Estimation of Jump Markov Linear Systems’. *IEEE TRANSACTIONS ON SIGNAL PROCESSING* **49**(3):613–624.
- D. Dougherty & D. Cooper (2003). ‘A practical multiple model adaptive strategy for multivariable model predictive control’. *Control Engineering Practice* **11**(6):649–664.
- N. H. El-Farra & P. D. Christofides (2003). ‘Coordinating feedback and switching control of hybrid nonlinear processes’. *AIChE Journal* **49**(8):2079–2098.
- N. H. El-Farra, et al. (2005). ‘Output feedback control of switched nonlinear systems using multiple Lyapunov functions’. *Systems & Control Letters* **54**:1163–1182.
- G. Ferrari-Trecate, et al. (2003). ‘A clustering technique for the identification of piecewise affine systems’. *Automatica* **39**(2):205–217.
- B. Foss, et al. (1995). ‘Nonlinear predictive control using local models - applied to a batch fermentation process’. *Control Engineering Practice* **3**(3):389 – 396.
- L. Giarre, et al. (2006). ‘LPV model identification for gain scheduling control: An application to rotating stall and surge control problem’. *Control Engineering Practice* **14**(4):351–361.

- G. Goodwin & J. Agüero (2005). ‘Approximate EM algorithms for parameter and state estimation in nonlinear stochastic models’. In *Proceedings of IEEE conferences on Decisions and Control*, pp. 368–373.
- G. Goodwin & J. Agüero (2008). ‘Identification of nonlinear processes with known model structure under missing observations’. In *In Proceedings of the 17th IFAC World Congress*, Seoul, South Korea.
- J. K. Gugaliya, et al. (2005). ‘Multi-model decomposition of nonlinear dynamics using a fuzzy-CART approach’. *Journal of Process Control* **15**:417–434.
- W. Heemels, et al. (2001). ‘Equivalence of hybrid dynamical models’. *Automatica* **37**(7):1085–1901.
- M. A. Henson & D. E. Seborg (1992a). ‘Nonlinear control strategies for continuous fermenters’. *Chemical Engineering Science* **47**(4):821–835.
- M. A. Henson & D. E. Seborg (1992b). ‘Nonlinear control strategies for continuous fermenters’. *Chemical Engineering Science* **47**(4):821–835.
- B. Huang & R. Kadali (2008). *Dynamic Modeling, Predictive Control and Performance Monitoring: A Data-driven Subspace Approach*. Springer.
- M. Jelali & B. Huang (2009). *Detection and Diagnosis of Stiction in Control Loops: State of the Art and Advanced Methods*. Springer.
- X. Jin & B. Huang (2009a). ‘Identification of switched Markov autoregressive eXogenous systems with hidden switching state’. *submitted to Automatica* .
- X. Jin & B. Huang (2009b). ‘Robust identificaiton of Switching/Piecewise Autoregressive eXogenous process’. *AIChE Journal*, DOI 10.1002/aic.12112 .
- X. Jin & B. Huang (2009c). ‘Robust identification of hybrid systems using EM algorithm’. In *8th World Congress of Chemical Engineering*, Montreal, Canada.
- A. Juloski (2004). *Observer design and identification methods for hybrid systems: theory and experiments*. Ph.D. thesis, Eindhoven University of Technology, Netherlands.
- A. Juloski, et al. (2005). ‘A Bayesian approach to identification of hybrid systems’. *IEEE Transactions on Automatic Control* **50**(10):1520–1533.
- S. Kalyani & k. Giridhar (2007). ‘Robust statistics based Expectation Maximization algorithm for channel tracking in OFDM systems’. In *Communications, 2007. ICC '07. IEEE International Conference on*, pp. 3051–3056.

- S. Kay (2001). ‘Conditional Model Order Estimation’. *IEEE Transactions on Signal Processing* **49**(9):1910–1917.
- A. Khalate, et al. (2009). ‘Optimal experimental design for LPV identification using a local approach’. In *15th Symposium on System Identification*, Saint Malo, France.
- L. Lee & K. Poolla (1996). ‘Identification of Linear Parameter-Varying Systems using Nonlinear Programming’. In *Proceedings of the 35th IEEE Conference on Decision and Control*, pp. 1545–1550.
- L. Ljung (1987). *System identification, Theory for the user*. Prentice Hall, New Jersey.
- A. Logothetis & V. Krishnamurthy (1999). ‘Expectation Maximization Algorithms for MAP Estimation of Jump Markov Linear Systems’. *IEEE TRANSACTIONS ON SIGNAL PROCESSING* **47**(8):2139–2156.
- G. J. McLachlan & T. Krishnan (1996). *The EM algorithm and its extensions*. John Wiley & Sons, New York.
- R. Morales-Menendez, et al. (2003). ‘Estimation and control of industrial processes with particle filters’. In *American Control Conference*, Denver, Colorado.
- M. Morari (2007). ‘Recent Development in the Control of Constrained Hybrid Systems’. In *CPC 7*, lake Louise, Canada.
- R. Murray-Smith & T. A. Johansen (1997). *Multiple Model Approaches to Modelling and Control*. Talyor & Francis.
- H. Nakada, et al. (2005). ‘Identification of piecewise affine systems based on statistical clustering technique’. *Automatica* **41**(5):905–913.
- P. Mhaskar and N.H. El-Farra and P. D. Christofides (2005). ‘Predictive Control of Switched Nonlinear Systems with Scheduled Mode Transitions’. *IEEE Transactions on Automatic Control* **50**:1670–1680.
- T. Proll & M. N. Karim (1994). ‘Model-predictive pH control using real time NARX approach’. *AIChE Journal* **2**(40):269–282.
- S. J. Qin & T. A. Badgwell (2003). ‘A survey of industrial model predictive control technology’. *Control Engineering Practice* **11**(7):733–764.
- L. R. Rabiner (1989). ‘A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition’. In *PROCEEDINGS OF THE IEEE*, vol. 77, pp. 257–286.

- J. Ragot, et al. (2005). ‘Data Reconciliation: A robust approach using contaminated distribution’. In *16th IFAC World Congress*, Prague, Czech Republic.
- J. Ragot, et al. (2003). ‘Parameter estimation of switching piecewise linear systems’. In *Proceedings of the 42th IEEE conference on Decision and Control*, pp. 5783–5788, Maui, HI.
- J. Roll, et al. (2004). ‘Identification of piecewise systems via mixed integer programming’. *Automatica* **40**(1):37–50.
- C. Saint-Jean, et al. (2000). *Advances in pattern recognition*. Springer, Berlin.
- T. Saldju & A. Landgrebe (2000). ‘Robust parameter estimation for mixture model’. *IEEE Transactions on Geoscience and Remote Sensing* **38**(1):439–445.
- R. Senthil, et al. (2006). ‘Nonlinear State Estimation Using Fuzzy Kalman Filter’. *Ind. Eng. Chem. Res.* **25**(45):8678C8688.
- S.Piche, et al. (2000). ‘Nonlinear model predictive control using neural networks’. *IEEE Control Systems Magazine* **20**(3):53–62.
- D. Sworder & R. Rogers (1983). ‘A LQG solution to a control problem with solar thermal receiver’. *IEEE Transactions on Automatic Control* **28**:971C978.
- A. Tamir (1998). *Applications of Markov Chains in Chemical Engineering*. Elsevier Science & Technology Books.
- I. B. Tjoa & L. Biegler (1991). ‘Simultaneous strategy for data reconciliation and gross error detection of nonlinear systems’. *Computers & Chemical Engineering* **15**(10):679–690.
- A. N. Venkat, et al. (2003). ‘Identification of complex nonlinear process based on fuzzy decomposition of the steady state space’. *Journal of Process Control* **13**:473–488.
- R. Vidal (2008). ‘Recursive identification of switched ARX systems’. *Automatica* **44**:2274C2287.
- R. Vidal, et al. (2003). ‘An algebraic approach to the identification of a class of linear hybrid systems’. In *Proceedings of the 42th IEEE conference on Decision and Control*, pp. 167–172, Maui, HI.
- W. C. Wong & J. H. Lee (2009). ‘Realistic disturbance modeling using Hidden Markov Models: Applications in model-based process control’. *Journal of Process Control* .

- Z. Xu, et al. (2009). ‘Nonlinear MPC using identified LPV model’. *Industrial & Engineering Chemistry Research* **6**(48):3043–3051.
- Y. Zhu (2001). *Multivariable System Identification For Process Control*. Elsevier Science.
- Y. Zhu & Z. Xu (2008). ‘A Method of LPV Model Identification for Control’. In *Proceedings of the 17th World Congress of the International Federation of Automatic Control*, pp. 5018 – 5023.