"We create order out of chaos, beauty and meaning out of ugly randomness."

Rick Riordan
The Throne of Fire

"Winter is coming."

George R. R. Martin
A Song of Ice and Fire

University of Alberta

Economic optimization of steam plant operation

by

Tianbo Liu

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

Fall 2013
Edmonton, Alberta

Abstract

The utility system plays an important role in efficient plant operations of chemical processes. In this thesis, economic optimization of steam utility system is investigated in detail. The objective is: 1) to calculate the optimal generation amount of steam and electricity under uncertainty in process and electricity market; 2) to distribute the generated steam in a most efficient way throughout the steam network.

In this work, steam distribution system is represented as a network with dynamic process equipment models. Operating constraints and uncertain process disturbances are included to accurately represent plant operations.

A cost-benefit analysis reveals that electricity price plays an important role in optimal plant operations. Thus, to maximize the economic profit of a steam plant in the long term, a high quality electricity price prediction model is developed based on a robust switched system identification algorithm. The algorithm is formulated using Expectation-Maximization (EM) algorithm to estimate parameters in prediction model, noise distribution and switching dynamics.

Dynamic process models and electricity price prediction models are integrated into a linear programming problem that uses plant profit as the performance objective. Random process variables are included to represent process uncertainty. The optimization effect is evaluated by comparing the plant profit from routine operations and from optimized operations. The distribution of optimized plant profit is obtained by solving the distribution problem of stochastic linear programming (SLP). A metric based on Earth Mover's Distance (EMD) is introduced to measure the difference between plant profit distributions.

Based on the validation results of developed models and proposed performance evaluation method, the optimized steam plant operations show significant advantage over the routine ones when electricity prices vary considerably.

Acknowledgements

Graduate study and research is a tough but sweet process: it is tough when you try to push the boundary of knowledge; it is sweet because you are not lonely.

I would like to give my first and foremost gratitude to Dr. Biao Huang and Dr. J. Fraser Forbes, who are my dear supervisors. Without their continuous support, guidance and encouragement, I cannot finish this adventurous long journey. It is their encouragement when I was lost in directions that finally led me to the destination. My special thanks will also go to Mr. Changbin Yu, who is the industrial supervisor for the project I participated. Without Changbin's dedicated help, I may not have the chance to visit the steam plant several times in person to collect the first-hand data for this research.

I would like to thank William Weng for his foundation work in the steam plant optimization project. I would give my sincere thanks to my friends and colleagues who had helped me rule out the distracting ideas but keep the true essence, they are: Marziyeh Keshavarz, Da Zheng, Yuri Shardt, Lei Chen, Chen Li, Aditya Tulsyan, Nima Sammaknejad, Shima Khatibisepehr, Yaojie Lu, Yafeng Guo, Yu Zhao and many others from Computer Process Control group. I will always be proud to be a member of the CPC group. I would also like to thank my friends for their encouragement and accompanying, and most importantly, the joyful time we spent together: Mulang Chen, Zhankun Xi, Kangkang Zhang, Ming Ma, Liang Chen, Xing Jin, Yijia Zhu, Lei Sun, Xiongtan Yang, Swanand Khare, , Sridhar Dasani, Elliot Cameron, Ruben Gonzalez, Jinfeng Liu, Jing Zhang, Xi Huang, You Fu and so many else.

Last but not least, I would like to express my deepest love and gratitude to my parents and girl friend Shuning for their unconditional support and love over so many years.

Contents

List of Tables

List of Figures

List of Abbreviations

SMARX	Switched Markov Autoregressive eXogenous
EM	Expectation-Maximization
GM	Gaussian Mixture
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
LP	Linear Programming
SLP	Stochastic Linear Programming
DLP	Dynamic Linear Programming
RHS	Right Hand Side
EMD	Earth Mover's Distance
RBS	Random Binary Sequence
LQI	Linear Quadratic Integral
MPC	Model Predictive Control
LTI	Linear Time Invariant
NG	Natural Gas
BFW	Boiler Feed Water
PLS	Pressure Let-down Station
P	Pressure
OE	Output Error
AESO	Alberta Electric System Operator
SMP	System Marginal Price
CSTR	Continuous Stirred Tank Reactor
RMSE	Root-Mean-Square-Error

Chapter 1

Introduction

In process industries, utility systems are crucial components that supply energy for plant operation, usually in the form of steam and electricity. The steam plant has high degrees of freedom to allow significant opportunity for cost reduction through optimization [4]. Functionally, the most important task of steam plant is to meet the fluctuating process steam demand. Economically, the cost of steam plant comes from the energy and capital consumption in steam generation, and the profit comes from the sales of generated electricity to electricity grid. Thus, for efficient operation and maximal profit of steam plant, the following aspects are essential to consider: modelling the steam plant and process equipment accurately, responding to electricity market changes in operation, meeting steam user demand and generating electricity in an optimal way.

1.1 Thesis objective and scope

Steam plant operation problems are widely formulated as optimization problems with operating constraints. The objectives are to minimize energy consumption, minimize operating cost and maximize economic profit [5], [2], [6], [7], [8]. This requires deep understanding of the steam network connection and input-output models for the process equipment, which is the focus of Chapter 2.

For the optimization of steam plant operation, if only current operating conditions are taken into account, long term optimality cannot be guaranteed [9]. Thus, a predictive optimization framework is required, in which the influence of current operation on future profit will be included. This naturally requires knowledge about the future electricity market, future user demands and process equipment dynamics. A high quality electricity price prediction model will be discussed in Chapter 3 and the predictive optimization framework containing future electricity market and process information will be discussed in Chapter 4.

After solving the steam plant operation problem, a method to evaluate the optimization effect is developed (i.e., how much better the optimal operation is compared to the routine operation). This method, based on the distribution problem of stochastic linear programming, is discussed in Chapter 4. It is to quantify the added value of the optimal operation over routine ones and to introduce a bounded performance index.

The flow chart for the steam plant optimization problem in this thesis is shown in Figure 1.1.

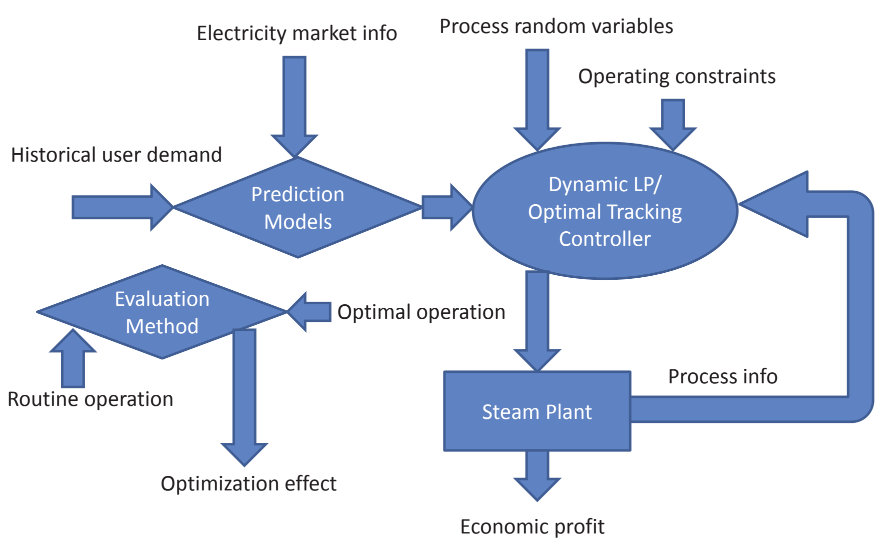


Figure 1.1: Steam plant optimization flow chart

1.2 Robust switched system identification

In this thesis, for accurate electricity price prediction, a robust switched system identification algorithm will be developed.

Linear time-invariant (LTI) models are widely used in many fields of science and engineering, such as spectroscopy, signal processing and electrical circuits, due to their structural simplicity and well developed supporting theories. In the traditional chemical process industries, LTI models are also the most used; however, due to inherent complexity of chemical processes, in many cases, it is difficult to capture complete dynamics of chemical processes and control them via a single LTI model [10]. Reasons for this include inherent physicochemical discontinuities (e.g., phase changes, flow reversals, shocks and transitions), the use of discrete actuators and sensors (e.g., on/off valves, pumps and binary sensors), and the use of logic-based switching for safety control tasks [11]. Hybrid modelling becomes a natural choice [12] in cases where the process switches among discrete operating states (usually unknown), and a different local continuous model describes the system in each state. In the literature, various approaches to hybrid system modelling have been developed. In [12], four methods are discussed for identifying switched affine and piecewise affine systems, namely: the algebraic, Bayesian, cluster-based and bounded error approaches. In [13], Jin and Huang (2010) proposed a method to identify a class of piecewise ARX model (PWARX), where the switching among various operating points is assumed to be random. To further model patterns in switching dynamics, the authors [14] proposed an extended method based on Expectation-Maximization (EM) algorithm, to identify switched ARX models where switching dynamics is assumed to be a first order hidden Markov model. As a powerful approach for maximum likelihood estimation, the EM algorithm has been widely used in different disciplines after its advent in 1970s [15]; however, in the process modelling and control literature, application of EM algorithm is relatively new. Goodwin and Aguero (2008) applied the EM algorithm to nonlinear and switched system parameter estimation [16]. As discussed earlier, Jin and Huang (2010, 2011) applied the EM algorithm to switched system identification by assuming first-order Markov switching dynamics [13], [14], [10]. Deng and Huang

(2012) further extended the previous work to identification of nonlinear parameter varying systems with missing output data [17]. The EM algorithm based method in [14] provided a good framework to model hybrid system dynamics using switched Markov ARX (SMARX) model, and it provided an estimation approach for both discrete and continuous dynamics at the same time. In addition, the closed form analytical update equation also makes this method attractive.

In the above EM based switched system identification methods, noise has been assumed to be single Gaussian distributed. This assumption has limited the flexibility when handling data with non-Gaussian distribution noise, or more commonly, low quality data with outliers. The authors in [13] provided a relatively robust way to handle outliers by dividing noise into a regular portion and irregular portion; however, in this method, prior knowledge about outliers is required, and the outlier pattern cannot be more than one. In cases where non-Gaussian distributions or multiple outlier patterns are present, the algorithm [13] cannot properly handle them. In general, due to existence of outliers and non-Gaussian distributed noise in industrial data, a single Gaussian distribution is insufficient to describe various noise characteristics. Instead, the Gaussian mixture model (GMM) can be used in modelling non-Gaussian noise because of its ability to approximate a rich class of distributions, analytical tractability, and conciseness in form [18]. The Gaussian mixture noise distribution is introduced in switched system identification in Chapter 3, which can naturally handle outliers and approximate a rich group of non-Gaussian noise distributions such as uniformly distributed noise [19].

In the literature, outlying observations (i.e., outliers) are measured values that deviate seriously from normal range of measured data, and existence of outliers in data can degrade performance of data driven models significantly [20]. It can be expected that by assuming Gaussian mixture distributed noise, outliers will naturally be modelled as components whose mixing probabilities are relatively small, and mean or variance values deviate from normal noise range significantly.

Another merit of Gaussian mixture noise assumption is its ability to represent a rich class of noise distributions. GMM can approximate arbitrarily well any density function, and the resulting class of density functions is abundant for engineering use

[19],[21]. This would increase the flexibility of proposed switched system identification algorithm.

In addition to noise distribution, a higher-order hidden Markov model may have potential in SMARX modelling for better performance in capturing more complex switching dynamics. Compared with the first-order HMM, a higher-order HMM captures more information contained in historical data and thus, is a more general and descriptive approach to model complex switching dynamics like periodicity. In [22], a comparative study has shown that a second-order Markov chain is more descriptive than its first-order counterpart in wind speed modelling. In the speech recognition literature, the higher-order hidden Markov model is also investigated [23], [24], [25].

1.3 Distribution problem for stochastic linear programming

Linear programming (LP) is a popular tool that is used for decision making, scheduling and planning in various industries, such as transportation, manufacturing and energy dispatch. Due to the stochastic nature of complex industrial operating environment, the coefficients in constraints and objective function of LP problems are usually random rather than constant. Stochastic linear programming (SLP) is the tool to deal with LP problem with random variables in formulation. Hansoia [26] states that chance-constrained LP and LP under uncertainty are two main fields where randomness is introduced into LP [27]. In this thesis, the third main research topic of SLP is considered, namely the distribution problem. The distribution problem of SLP was first introduced by Babbar [28], Tintner [29] and Wagner [30]. The aim of distribution problem is to solve for the distribution of objective function and decision variables in SLP, so that perfect information about the objective function and decision variables under uncertainty can be uncovered.

In the optimization of utility steam systems, where steam and electricity are generated to meet steam user demand or sell to the electricity grid, LP is a popular technique [7], [9]. There are unknown process parameters, such as missing or faulty measurements in the steam network, and thus the operation constraints in LP generally have unknowns in right hand side (RHS). The RHS of these constraints is always random, and is at most known in distribution estimated from historical mass and energy balances. The solution to the distribution problem is important as: 1) it provides an estimation of the added value of optimal operation under process uncertainty; 2) by knowing the distribution of plant profit and decision variables, it is possible to compare and classify different plant operating modes given appropriate probability distribution metric; 3) knowing the distribution for decision variables simplifies the sensitivity analysis for SLP and therefore simplifies the determination of bottleneck in steam plant.

As stated in [27], the solution for the distribution of objective function in a closed form is generally not possible. Note that even in relatively simple case depicted in (4.16), no easy solution can be derived. As an analytical solution to distribution problem is generally difficult or impossible to derive, the author in [31] suggests three alternative methods, namely: the simulation method, the discretization method and the incomplete description method. In this thesis, a Monte Carlo simulation approach is used to solve for the distribution of objective function and decision variables. Bracken and Soland [32], and Sarper [27] used the Monte Carlo simulation methods to solve the distribution problem of SLP. One of the drawbacks of Monte Carlo simulation based methods is the heavy computational load in estimating the distributions based on histograms of objective function. Since in most cases the resulting distribution family is unknown, only a non-parametric method can be used.

For the steam plant optimization problem, it is meaningful to compare the optimized operation with the routine ones. This comparison can be made by measuring the difference between the distribution of optimized plant profit and that of routine operations. This calls for a metric between probability distributions. Among different probability distribution metrics, the Earth Mover's Distance, indicating the minimal cost to transform from one distribution to another, is widely used. Its merits can be summarized as: 1) it accounts for perceptual similarity better than other metrics; 2) it can be calculated effectively using an LP solver; and 3) it is a true metric under mild restrictions, which is important for optimization purpose [33]. Defining the EMD between two distributions requires a defined 'distance' between the basic features that form distribution, and this distance is generally referred to as ground distance in the literature [33]. In this thesis, the Hellinger distance is used as the ground distance for EMD as: 1) the ground distance can be derived in a closed form using the Hellinger distance to avoid high computational load; 2) the Hellinger distance is a true metric bounded between 0 and 1, which makes the upper layer EMD a true metric between 0 and 1; and 3) in this thesis, the Hellinger distance can be modified so that it is adjustable according to different user requirements.

1.4 Thesis contributions and outline

An optimization framework is proposed in this thesis to maximize the operating profit in steam utility systems under process and market uncertainty. The proposed framework integrates: 1) LP problem using profit as the objective, and incorporating process/market information and operating restrictions into constraints; 2) high-quality, hour-ahead electricity price prediction model which allows the plant operation to hedge risk in peak price hours; and 3) probability density estimation of the missing process measurements from historical data using mass balances. Under this framework, the potential profit in utility systems can be assessed quantitatively, and can be maximized based on random missing measurement distribution and electricity price prediction. Cross validation shows good prediction accuracy in electricity price, and benefit analysis by industrial partner shows promising results of proposed optimization approach.

The main contributions of this thesis are:

1. Models of the steam network connection and dynamic process equipment under process uncertainty.

2. Introduction to Alberta's electricity market, and proposal of a robust switched electricity price prediction model. The characteristics of the proposed method are: 1) introducing Gaussian mixture noise distribution to model outliers and to approximate non-Gaussian noise distributions; 2) solving the formulated prob-

lem by EM algorithm and deriving a closed form solution; 3) extending the switching dynamics to second-order hidden Markov model; and 4) proposing a novel initialization strategy for EM algorithm to include prior process knowledge.

3. Formulation of a predictive optimization framework integrating the dynamic process equipment models and electricity price prediction models into an LP problem.

4. Design of controllers to track the optimal boiler load trajectory calculated by LP and development of an optimization performance index.

The remainder of this thesis is organized as follows:

Chapter 2 describes the complete steam network connection and process equipment models, which will be used in the predictive optimization framework.

Chapter 3 introduces Alberta's electricity market, and develops a robust switched electricity price prediction model to capture the electricity market dynamics for optimization use.

Chapter 4 integrates the process equipment models and electricity price prediction to form a unified predictive optimization framework and proposes a novel approach to evaluate the optimization performance.

Chapter 5 concludes the thesis and provides perspectives for future work.

Chapter 2

Steam Plant Model

In this chapter, the process equipment models that will be used for plant optimization are developed in detail. Section 2.1 discusses the mass balances in each steam common header, the steam distribution constraints, and models for random process variables. Section 2.2 describes the turbine generator model which links the process operation with electricity generation. Section 2.3 provides the drum boiler dynamic model, which is used in predictive optimization framework in Chapter 4.

2.1 Plant-wide steam mass balances

In this section, the steam plant layout and mass balances will be discussed. The steam suppliers and users connected to each common header will be introduced. The plant-wide operational constraints, as well as the models for random process variables are developed.

2.1.1 Steam network connection

The process flow diagram for steam plant considered in this thesis is given in Figure 2.1.

In the steam plant, there are a total of four steam headers connected together: 1) #900 steam common header; 2) #450 steam common header; 3) #160 steam common

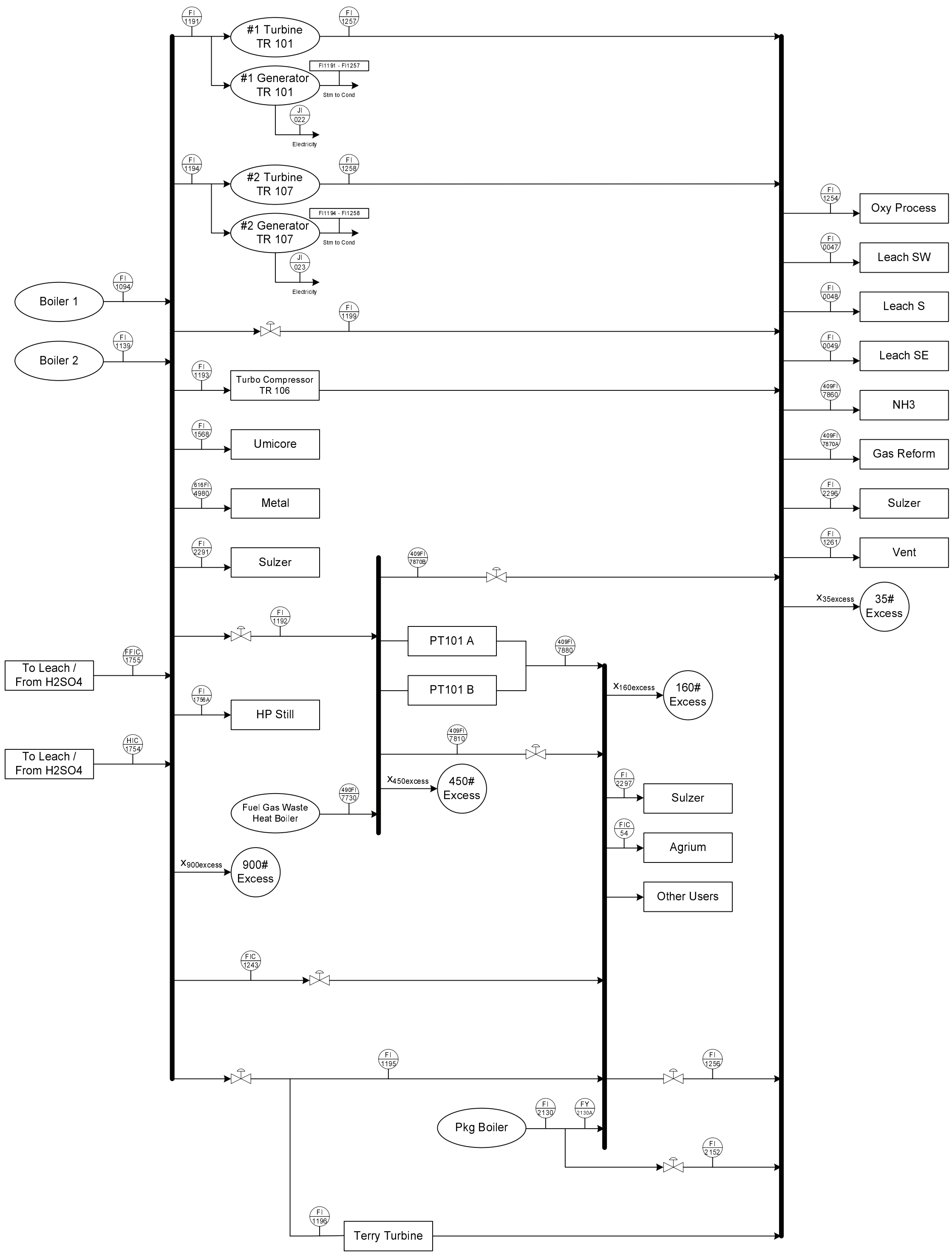


Figure 2.1: Steam plant process flow diagram

header; and 4) #35 steam common header. Two steam boilers are connected to #900 header. They consume boiler feed water (BFW) and natural gas (NG) to produce 900 psi steam. Another source of 900 psi steam is the sulfuric acid production units. Changes in user demand, steam leaks and other plant operations will influence the pressure in the common headers, and the automatic control system will manipulate the BFW and NG flow rates to compensate for this change. In the steam supply system, there are equipment that transform steam from high pressure to low pressure, such as pressure let-down stations (PLS), turbine generators, and process turbines.

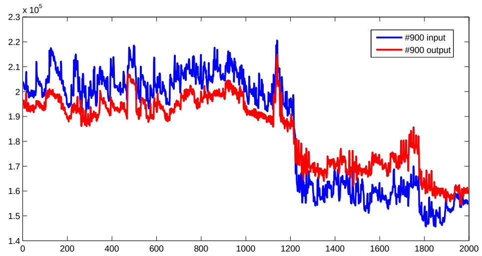
The optimal steam plant operation requires that the amount of steam generated at each operating instant satisfies the user demand, maximizes the profit via electricity generation, and distributes the generated steam in a plant-wide optimal fashion. The cost of the steam production results from: the energy and material consumed to generate steam, (i.e., water and natural gas), the labour cost and the operation, maintenance and depreciation cost of process equipment. The profit of the steam plant is from the following aspects: the value ascribed to a user demand for the steam and the generated electricity sold to the electricity grid. Thus, the key for the steam plant optimal operation are the fluctuation of electricity price and the steam user demand, which will be modelled in detail in Chapters 3 and 4, respectively.

2.1.2 Steam distribution constraints

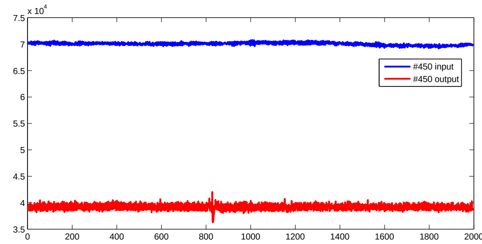
To optimize the steam plant operation, the mass balance constraints need to be respected for each common header. In practice, the amount of steam flowing into each common header should be slightly more than or equal to the total user demand from that header. They are not equal when there are leaks along the pipelines. The actual measured input and output flows for each common header are shown in Figure 2.2 from process data.

From Figure 2.2, it can be seen that the input and output steam amount for each common header do not balance each other. The missing measurements could be leaks along the pipeline or unmeasured steam flows. To be specific:

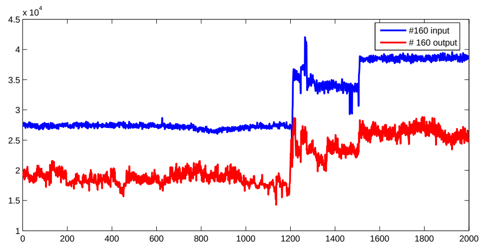
1. In the #900 common header, the input is sometime more and sometime less



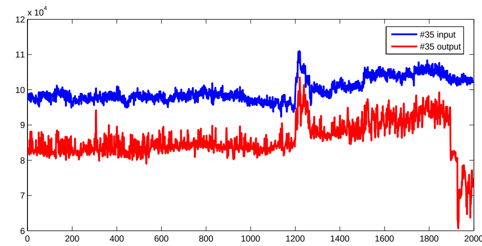
(a) #900 common header



(b) #450 common header



(c) #160 common header



(d) #35 common header

Figure 2.2: The input-output mismatch in common headers

than the output, and the average absolute difference is around $10^4 lb/hr$. Since no abnormality in the flow meters or large amount leakage is reported, there may be unmeasured steam flow, or missing measurements on both supply and the demand side of #900 common header. The imbalance amount changes from time to time, which can be modelled by a random variable from a certain distribution.

2. In the #450 common header, the input always exceeds the output, and the amount is relatively constant, which can be explained as the missing measurement on the demand side of #450 steam, and the missing demand amount can be approximately taken as a constant offset in modelling.

3. In the #160 common header, the input always exceeds the output, and the amount changes slightly from time to time, which can be explained as the missing measurement on the demand side of #160 steam, and the missing demand can be modelled by a random variable following a certain distribution.

4. In the #35 common header, the input always exceeds the output, and the

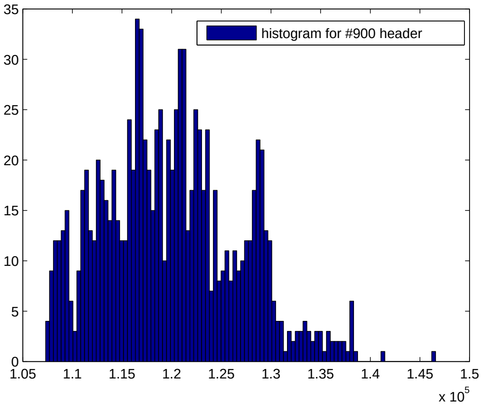
amount changes significantly from time to time, which can be explained as the missing measurement on the demand side of #35 steam, and the missing demand can be modelled by a random variable following a certain distribution.

The missing measurements in each common header could be from some hidden steam users or suppliers. Thus, if they are neglected in the steam plant optimization, the calculated optimal steam input amount for each common header might be insufficient or excess, which in both cases deviates from the optimization objective. The missing measurements will also challenge the optimization performance assessment process, which will be discussed in Chapter 4. Thus, accurate models for missing measurements and fluctuating user demands are essential to the success of the overall steam plant optimization.

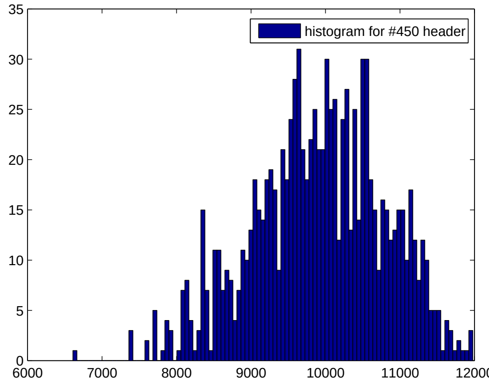
2.1.3 Random process variables

In previous subsection, there are missing measurements in all of the common headers, and the future missing amount of steam is not known directly, which yet will be useful in the predictive optimization framework. Nevertheless, using the mass balance relationship, the historical values of missing measurements can be calculated as the difference between the input and output amount of steam in each header. Therefore, taking the missing measurements as random variables, the distribution of each missing element can be estimated from the historical input-output difference. Another source of uncertainty in the steam plant is the fluctuating user demands. Unlike the missing measurements, the user demands are directly known in the optimization; however, from a long term point of view, the user demand data are random, and for the purpose of predictive optimization and performance assessment afterwards, they need to be properly modelled.

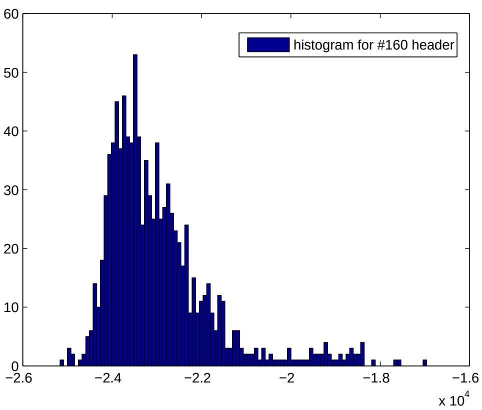
To model the random process variables appropriately, first the histograms of these process variables are plotted, and distribution families are chosen to represent the histograms. For performance assessment of the steam plant optimization under uncertainty, all of the random process variables in each header are lumped into one random variable. The histograms for each of lumped random variable can be found



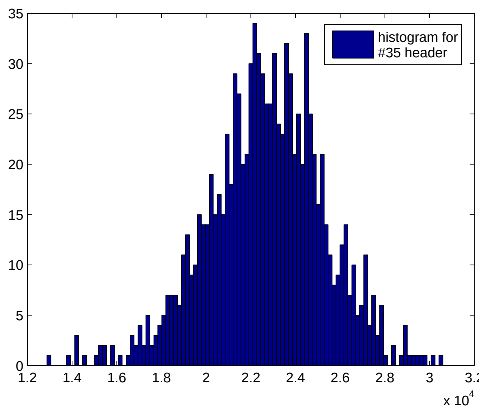
(a) #900 common header



(b) #450 common header



(c) #160 common header



(d) #35 common header

Figure 2.3: Histograms for the lumped random variables in each steam header

in Figure 2.3.

The histogram for the #900 header shows multiple modes, and it may be possible to model each of the mode by a local Gaussian distribution. The histogram for the #450 and #35 headers are similar to Gaussian. The histogram for #160 header appears asymmetric with a long tail to the right.

Based on the shape of histogram in each common header, the Gaussian mixture model is chosen to represent their distributions. In probability density estimation, the Gaussian mixture model (GMM) is widely used in modelling non-Gaussian distributions because of its parsimony, its ability to approximate a rich class of distributions, and its analytical tractability [18]. A general form of Gaussian mixture model can be expressed as follows [34]:

$$\begin{array}{l} p\left(x|\theta\right) = \sum\limits_{i=1}^{M} w_i g\left(x|\mu_i, \Sigma_i\right) \\ g\left(x|\mu_i, \Sigma_i\right) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}\left(x-\mu_i\right)\right\} \end{array} \tag{2.1}$$

where: x is D-dimensional continuous valued data; $g(x|\mu_i, \Sigma_i)$ is the local Gaussian component, with mean μ_i and covariance matrix Σ_i; and w_i is the mixing weight with $\sum_{i=1}^{M} w_i = 1$. θ represents the set of parameters $\{w_i, \mu_i, \Sigma_i\}$ to be estimated from the historical observations of lumped random variables. The Gaussian mixture can naturally handle outliers and approximate a rich group of non-Gaussian distribution features such as long tail and asymmetry [19].

Since a Gaussian mixture is chosen to approximate the distributions of the random variables in each common header, it is essential to find an effective algorithm to estimate the parameters $\{w_i, \mu_i, \Sigma_i\}$. Expectation-Maximization (EM) algorithm is one such algorithm.

The EM algorithm is a popular solution for maximum likelihood estimation problem with hidden information. An EM algorithm consists of two steps, E-step and M-step, and the algorithm updates the estimated parameters iteratively by these two steps, until convergence. In the E-step, conditional expectation of the likelihood function over hidden information (in this case, the membership of local Gaussian distribution to which each data point belongs) is calculated, and the result is normally referred to as the Q function in the literature [15], [35]. Mathematical formulation of the Q function for Gaussian mixture parameter estimation can be expressed as [36]:

$$\begin{aligned} Q\left(\theta|\theta^{old}\right) &= E_{Y|X,\theta^{old}}\left\{\ln P\left(X,Y|\theta\right)\right\} \\ &= E_Y\left\{\ln P\left(\left(x_1,y_1\right),\left(x_2,y_2\right),...,\left(x_n,y_n\right)|\theta\right)|X,\theta^{old}\right\} \\ &= E_Y\left\{\ln\prod_{i=1}^{n} P\left(\left(x_i,y_i\right)|\theta\right)|X,\theta^{old}\right\} \end{aligned} \tag{2.2}$$

where: X is the observed data collected from historical database for each common header; and Y is treated as hidden variable, denoting the Gaussian component from which each x_i comes. Thus, $y_i = 1, 2, ..., M$, assuming M local Gaussian components. The final equality in equation (2.2) follows the assumption that the historical data are independent from each other. From (2.2), the Q function can be re-written as [36]:

$$\begin{aligned} Q\left(\theta|\theta^{old}\right) &= E_Y\left\{\sum_{i=1}^{n}\ln\left(w_{y_i}g\left(x_i|\mu_{y_i},\Sigma_{y_i}\right)\right)|X,\theta^{old}\right\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{M}\left\{P\left(y_i=j|x_i,\theta^{old}\right)\left(\ln w_{y_i}+\ln g\left(x_i|\mu_{y_i},\Sigma_{y_i}\right)\right)\right\} \end{aligned} \tag{2.3}$$

where $P\left(y_i=j|x_i,\theta^{old}\right)$ is generally referred to as membership probability in the literature on Gaussian mixture parameters estimation [37], and it is the key probability that connects the E-step and M-step of the EM algorithm. By the Bayes' formula, it can be expressed as:

$$P\left(y_i=j|x_i,\theta^{old}\right)=\frac{f\left(x_i,y_i=j|\theta^{old}\right)}{f\left(x_i|\theta^{old}\right)}=\frac{w_j^{old}g\left(x_i|\mu_j^{old},\Sigma_j^{old}\right)}{\sum\limits_{j=1}^{M}w_j^{old}g\left(x_i|\mu_j^{old},\Sigma_j^{old}\right)} \tag{2.4}$$

Substituting (2.4) into (2.3), and in M-step, taking derivatives of the Q function over θ under probability measure constraints, yields the parameter update equation:

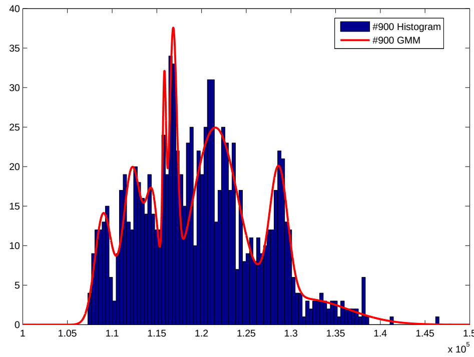
$$\begin{aligned} w_j &= \frac{\sum\limits_{i=1}^{n}P\left(y_i=j|x_i,\theta^{old}\right)}{n}, \\ \mu_j &= \frac{\sum\limits_{i=1}^{n}P\left(y_i=j|x_i,\theta^{old}\right)x_i}{\sum\limits_{i=1}^{n}P\left(y_i=j|x_i,\theta^{old}\right)}, \\ \Sigma_j &= \frac{\sum\limits_{i=1}^{n}P\left(y_i=j|x_i,\theta^{old}\right)\left(x_i-\mu_j\right)^T\left(x_i-\mu_j\right)}{\sum\limits_{i=1}^{n}P\left(y_i=j|x_i,\theta^{old}\right)} \end{aligned} \tag{2.5}$$

For detailed intermediate derivation steps and analysis of the EM algorithm in Gaussian mixture parameter estimation, see the Bilmes tutorial [36].

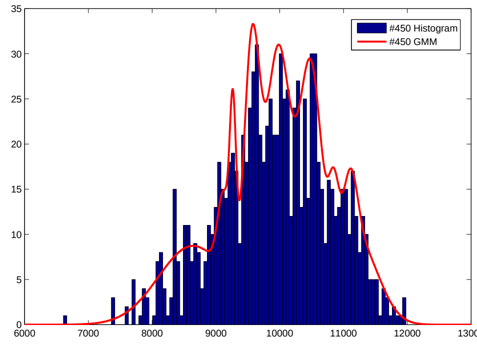
The estimated GMM distributions for the four common headers are shown in Figure 2.4. After estimation of these distributions, performance assessment of the steam plant optimization can be conducted using method proposed in Chapter 4.

2.2 Turbine generator model

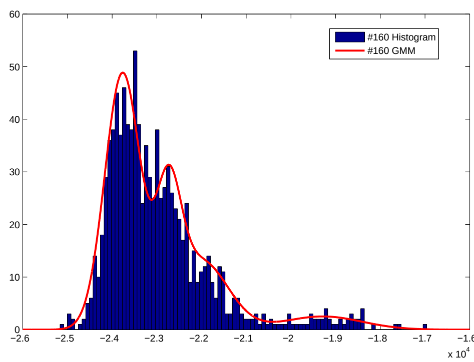
A model of the turbine generator is required as it is a potential profit maker in the steam plant due to electricity generated during peak price hours. Thus, knowing the



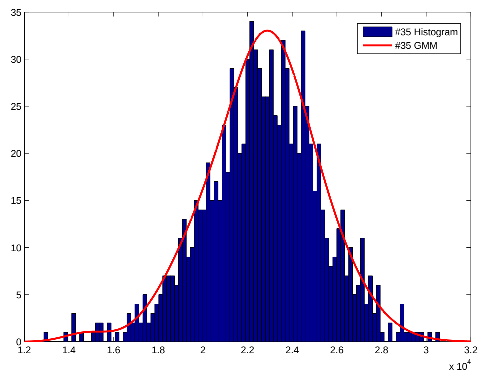
(a) #900 common header



(b) #450 common header



(c) #160 common header



(d) #35 common header

Figure 2.4: Estimated Gaussian mixture distributions for steam headers

relationship between steam passing through the turbine generators and the electricity generated is an essential element of steam plant optimization.

A simplified structure of the turbine generator and its connection to different steam common headers is shown in Figure 2.5 [2].

In [2], two approaches are introduced to model electricity generation using the amount of steam flowing through the turbine generator and the amount of #900 steam reduced to #35 steam. The first method is based on first principles or physical laws in the electricity generation process, namely the mass and energy balances, thermodynamics and so forth. Note that the turbine generator transforms the energy in the #900 steam to the electrical energy and the energy contained in #35 steam plus condensed water. The first principles model for the turbine generator can be

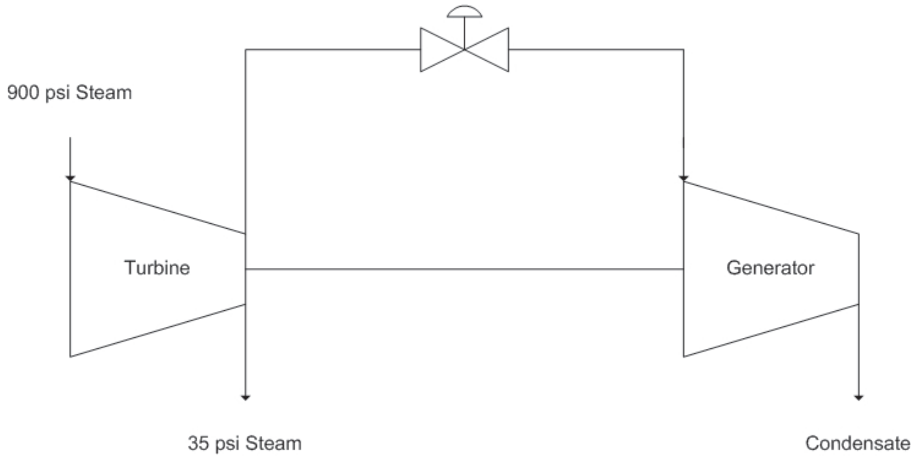


Figure 2.5: Turbine generator [2]

represented as (2.6), [2]:

$$\Delta E_{actual} = \eta_{overall} \left(H_{900} x_{tg} - H_{35} x_{tgex35} \right) \tag{2.6}$$

where: H_{900} and H_{35} are the enthalpy of corresponding steam, which are inherently non-linear functions of steam temperature and pressure; x_{tg} is the amount of #900 steam flowing into the turbine generator; x_{tgex35} is the flow-rate of steam reduced to #35 steam; and $\eta_{overall}$ is the efficiency of the turbine generator [2]. The parameters in equation (2.6) can be found in steam tables or calculated from turbine generator operation data as discussed in detail in [2].

The second approach is a data driven modelling method, based on equation (2.6). Given that the temperature and pressure of each steam common header are well controlled, the enthalpy of the steams are roughly constant, and the amount of electricity generated is approximately a linear function of x_{tg} and x_{tgex35}. After collecting the historical operating data for x_{tg}, x_{tgex35} and the amount of electricity generated, a linear regression analysis is performed to estimate the unknown parameters. Following this method, the data driven model for turbine generators can be expressed as in [2]:

$$X_{e1} \propto 0.0400 X_{tg1ex35} + 0.0905 (X_{tg1} - X_{tg1ex35}) \tag{2.7}$$

$$X_{e2} \propto 0.0435 X_{tg2ex35} + 0.1074 (X_{tg2} - X_{tg2ex35}) \tag{2.8}$$

The above turbine generator models will be used in the predictive steam plant optimization in Chapter 4.

2.3 Steam boiler model

In this section, a drum boiler dynamic model will be developed. First a model based on mass and energy balance equation is developed, and then a data based modelling procedure is used to estimate the parameters for the boiler dynamic model. The boiler dynamics will be combined with other elements models for the steam plant predictive optimization studies in Chapter 4.

2.3.1 First principles model

The motivation to include a steam boiler dynamic model in the optimization is from practical consideration. Although the optimal boiler load at each optimization step can be calculated, the steam boiler cannot adjust its input variables to achieve this load immediately due to relatively long response time of the boiler system. Thus, knowledge of the boiler dynamics can improve the effect of steam plant operating optimization.

The objective of the dynamic boiler model is to build a dynamic relationship between the steam generation and the boiler input variables: natural gas flow (NG), boiler feed water (BFW) and header pressure (P), which is a disturbance input variable. According to [38], based on the mass and energy balances, the first principles dynamic boiler model can be represented as:

$$\begin{aligned} &\frac{d}{dt}\left[\rho_s V_{st} + \rho_w V_{wt}\right] = q_f - q_s \\ &\frac{d}{dt}\left[\rho_s u_s V_{st} + \rho_w u_w V_{wt} + m_t C_p t_m\right] = Q + q_f h_f - q_s h_s \end{aligned} \tag{2.9}$$

where: V_{st} and V_{wt} are the total steam and water volumes, respectively; ρ_s and ρ_w are the densities for steam and water at operating temperature and pressure, respectively; q_f and q_s are the BFW flow-rate and the steam output, respectively; u_s and u_w are the specific internal energy of steam and water, respectively; h_f and h_s are the specific enthalpy of BFW and steam output, m_t is the mass of boiler; C_p the specific heat of boiler's metal; t_m is the temperature of the boiler metal; and Q is the external energy for heating, which is due to the burning of natural gas in the steam plant.

By further simplifying equation (2.9), the boiler dynamic model can be expressed as:

$$\begin{aligned} e_{11}\frac{dV_{wt}}{dt} + e_{12}\frac{dp}{dt} &= q_{\mathrm{f}} - q_s \\ e_{21}\frac{dV_{wt}}{dt} + e_{22}\frac{dp}{dt} &= Q + q_f h_f - q_s h_s \end{aligned} \tag{2.10}$$

where e_{ij} are parameters in the approximate linear state space model, treated as constants around steady-state operating point. Furthermore, in the above equations: q_s is taken as the system output; p is taken as the disturbance variable; NG is directly proportional to Q and is an input variable; and q_f represents the BFW flow and is another input variable. Combine equations in (2.10) to eliminate $\frac{dV_{wt}}{dt}$. After elimination of $\frac{dV_{wt}}{dt}$, equation (2.10) can be expressed as a transfer function with q_f and q_{NG} as input, p as disturbance, and q_s as output. Notice here Q is expressed as proportional to the NG flow q_{NG}:

$$\frac{e_{21}}{e_{11}}\left[q_f(s) - q_s(s) - e_{12}sp(s)\right] + e_{22}sp(s) = Cq_{NG}(s) + h_f q_f(s) - h_s q_s(s) \tag{2.11}$$

Equation (2.11) indicates that the steam generation can be represented by a dynamic model of natural gas flow(q_{NG}), boiler feed water(q_f) and pressure (P) through a transfer function. In practice, the boiler dynamics may not be as simple as in equation (2.11), because the parameters e_{21}, e_{11}, e_{12} is only constant within certain operating region and may vary if the operating condition changes dramatically. The parameter estimation of the dynamic boiler model will be discussed in next subsection.

2.3.2 Model derived from system identification

In the previous section, we discussed the dynamic boiler model in a transfer function form between the steam generation and BFW, NG, P, with P as a disturbance input.

After collecting input-output data, and pre-processing them for system identification purposes, different model structures were tested. The results of our analysis yielded an Output Error (OE) model as the desired model structure [39]:

$$y\left(t\right) = \frac{B\left(q\right)}{F\left(q\right)}u\left(t - nk\right) + e\left(t\right) \tag{2.12}$$

where $B(q)$ and $F(q)$ are polynomials of the back-shift operator to be estimated. For the multiple input, single output system in this case, $B(q)$ and $F(q)$ are vectors of polynomial coefficients, where in the steam boiler model, $B(q), F(q) \in \mathbb{R}^3$.

After taking into consideration the estimation error and the parsimony principle, the optimal orders for the polynomial coefficient vectors $B(q)$, $F(q)$ and time delay vector were determined to be: $n_b = [2, 2, 2]$, $n_f = [3, 1, 1]$, $n_k = [0, 0, 0]$. Using the MATLAB OE model estimation function, the estimated parameters are calculated as:

$$\begin{aligned}
&B_1(q) = 0.8293(\pm 0.0093) - 0.8297(\pm 0.0095)q^{-1}\\
&B_2(q) = -0.2524(\pm 0.0056) + 0.2813(\pm 0.0056)q^{-1}\\
&B_3(q) = -628.6(\pm 25.56) + 705.5(\pm 20.16)q^{-1}\\
&F_1(q) = 1 - 0.7697(\pm 0.012)q^{-1} - 0.4378(\pm 0.019)q^{-2} + 0.2338(\pm 0.012)q^{-3}\\
&F_2(q) = 1 - 0.9719(\pm 0.0028)q^{-1}\\
&F_3(q) = 1 + 0.3651(\pm 0.029)q^{-3}
\end{aligned} \tag{2.13}$$

where the $\pm$ numbers in the parentheses are the standard deviations of the estimated parameters.

The cross validation results for the identified boiler dynamic model will be presented in the next subsection.

2.3.3 Boiler model validation

Figure 2.6 shows the residual test results for the identified model. From the residual test results, the identified model passed all of the cross correlation tests between inputs and prediction error; however, it failed the autocorrelation test for prediction error. This is as expected, since the choice of OE model structure cannot ensure the prediction error to be white noise. Although OE model misses some dynamics according to the auto-correlation test, it is still an ideal model structure due to its

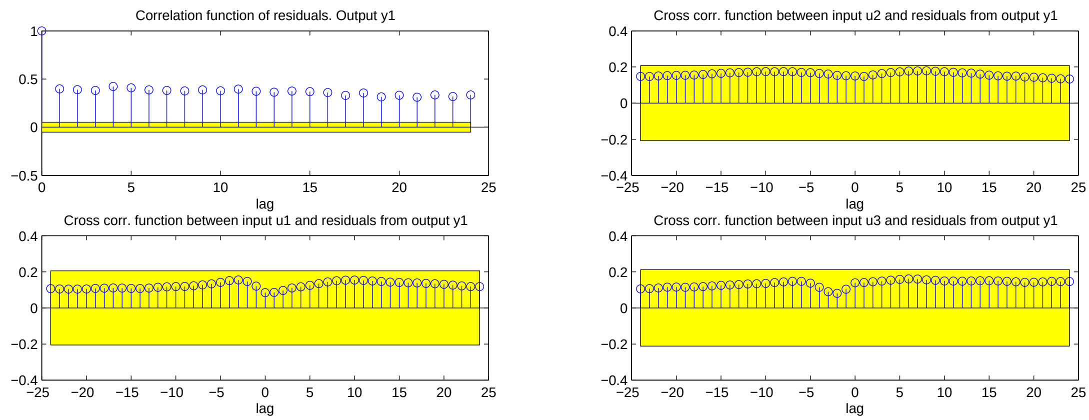


Figure 2.6: Residual test results: Top left: autocorrelation of prediction error; Top right: cross-correlation between prediction error and BFW; Bottom left: cross-correlation between prediction error and NG; Bottom right: cross-correlation between prediction error and P

parsimony and good fitting performance. The cross-validation fitting result is shown in Figure 2.7. The step test results of three inputs are shown in Figure 2.8.

Figure 2.7 shows that the cross validation fitting rate is as high as 91.15%. The step tests for NG flow, BFW flow and pressure are shown in Figure 2.8. In the NG step test result (top sub-figure in Figure 2.8), the steam generation increases at first, and then it drops to 0. As when the natural gas flow increases, the boiler load will increase first and then it drops because eventually there is no sufficient boiler feed water supply to balance the extra energy. In the BFW step test result (middle sub-figure in Figure 2.8), the steam generation increases smoothly. As the boiler drum level increases with BFW flow and when new mass balance is achieved, the steam generation will increase accordingly. In the P step test result (bottom sub-figure in Figure 2.8), steady state following oscillation can be observed.

The prediction performance and residual test results are reasonable. The cross fitting accuracy is more than 90%, the missing dynamics might be because of measurement error, model plant mismatch and other practical reasons; however, it is safe to use the dynamic boiler model at such accuracy from empirical point of view.

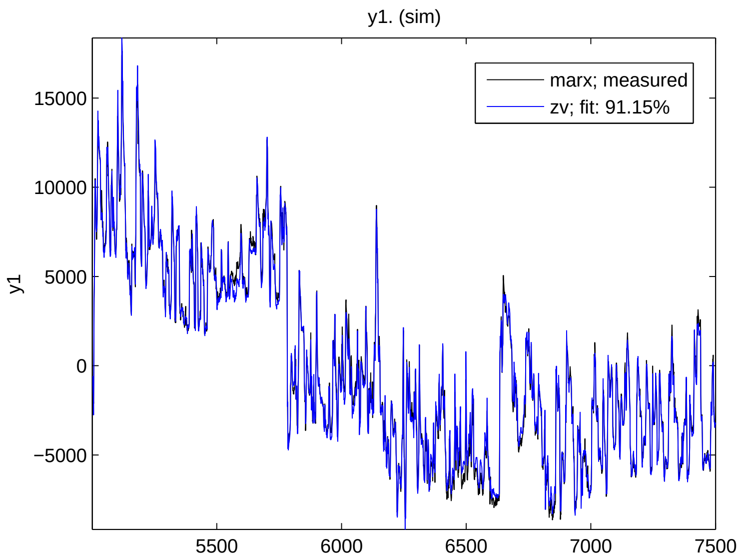


Figure 2.7: Cross fitting result of identified boiler dynamic model

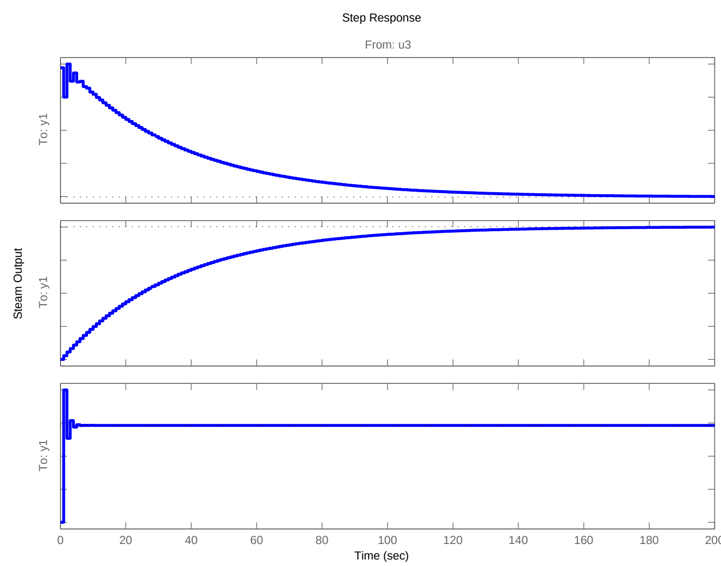


Figure 2.8: Step test result of identified boiler dynamic model. Top: step response of NG flow; Middle: step response of BFW; Bottom: step response of P

Chapter 3

Electricity price prediction model

In this chapter: first, a brief introduction to Alberta's deregulated electricity market will be made, the characteristics of the market and motivation of the multiple electricity price prediction model are explained; then, a switched Markov ARX model identification approach is proposed and illustrated by simulation and experiment results; and in the final section of this chapter, the application of proposed method to the development of a multi-model electricity price prediction method is discussed in detail.

3.1 Alberta's electricity market

Since 1996, Alberta has begun to operate its electricity market in a deregulated approach that increasingly relies on the generation and selling competition, [40]. This type of deregulated electricity market has more flexibility in electricity pricing, and thus grants more opportunities for generation units to optimize operation in response to market changes. There are some basic rules for electricity generation and selling in Alberta that are essential to understanding the predictive model construction. These contents are discussed in the following sections, as are some key facts of Alberta's electricity market.

In Alberta, all wholesale electrical energy from generation that is not consumed on site must flow through the Power Pool, which is operated by Alberta's independent system operator (ISO), named the Alberta Electric System Operator (AESO). One of the most important tasks of AESO is to operate the Power Pool such that the market operates in a fair, efficient and openly competitive manner [40]. All trading of power through the AESO is by a process of offers and bids according to a "merit order". The order is established to meet the forecast pool demand by ranking offers and bids from low to high in cost on an hourly basis. The last bid, or offer, that is dispatched every minute sets the System Marginal Price (SMP) and the time-weighted average of SMP at the end of each hour is calculated as the Pool price, which is the wholesale settlement price [41].

A wealth of information is available on the AESO website giving insight into current state of the market, such as SMP, Pool price, current supply-demand report, forecast demand, forecast Pool price and so forth. Actually the data used to build the electricity price prediction model in following sections comes from AESO website. The following characteristics of Alberta's electricity market are summarized from the literature [41], [42], [40]:

1. Obvious on-peak and off-peak electricity price patterns can be observed. According to the Alberta Energy Utilities Board, the on-peak period is from 8:00 to 21:00 Monday through Friday inclusive, except statutory holidays.

2. Pronounced periodic effects are observed. There are hourly, daily, intra-daily, and weekly repeating patterns. Within 24 hours of a day, prices increase as demand increases with a distinct hourly pattern. Electricity prices are normally higher when demand is greater. In Alberta, about 78% of demands are from industrial and commercial use, 18% from residential and 4% from farm.

3. Although there is strong correlation between demand and Pool price, interestingly, demand is not the most important driver of Pool price. Rather, it is always the unplanned (or forced) unit outages along transmission lines that drive the Pool price to high level.

4. Generators are free to make changes to their offer prices (but not offered volumes) as the market unfolds; however, two hours before the final price releases

all price changes must stop and the only allowable changes are those associated with operational issues at the units. Therefore, it is reasonable to predict the electricity price two hours ahead based on the current final offers from generation units. Note that AESO website publishes the two hour ahead Pool price prediction result, which is an important reference for the electricity price prediction model proposed later.

5. Price 'spikes'. As shocks in demand and supply are common, the electricity Pool prices are extremely volatile. It is not unusual to see the price as high as \$500/MWh or more. If the steam plants can foresee such spikes in electricity market and operate the steam generators accordingly, a considerable amount of profit can be made during 'spike hours'.

6. Price-dependent variance. Note that there is empirical evidence suggesting that the volatility of electricity prices is high when the demand is high and vice versa. Since Pool price and demand are highly correlated, the volatility of electricity prices is dependent on the Pool price as well. This implies that at different price levels, the electricity price prediction models should be different.

7. Non-negativity. There are cases where portions of electricity generation are offered at \$0/MWh to avoid being dispatched off at low Pool prices; however, there are no negative electricity price offerings in Alberta. On the other end, the upper limit of electricity price in Alberta is \$1000/MWh.

Figure 3.1 shows the hourly electricity price change within a typical month, from August 20th 2012 to September 20th 2012. From Figure 3.1, the periodic characteristics of electricity price can be observed. The on-peak hours in this month are generally from 14 to 16 in each day. The relationship between electricity price and demand can be found in Figure 3.2. From Figure 3.2, it can be seen that the correlation between the electricity demand and price is not linear. Thus, data preprocessing is necessary to capture such correlation if a linear model structure such as linear ARX model is chosen to model the electricity market dynamics.

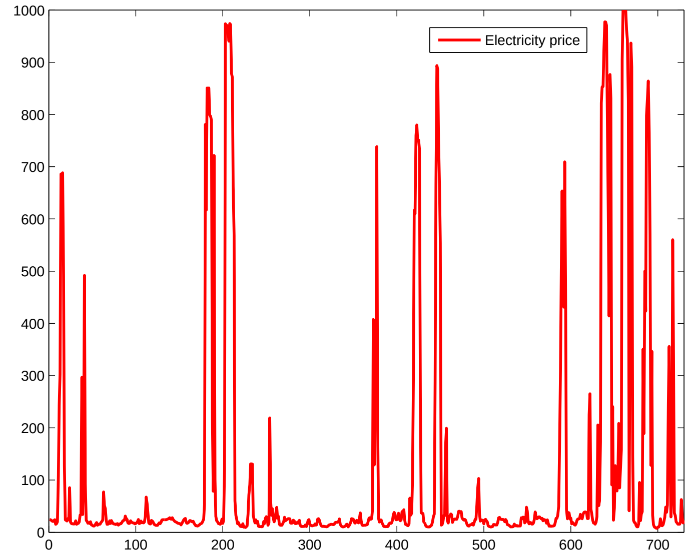


Figure 3.1: Alberta electricity price fluctuation within a month

Due to the characteristics of Alberta's electricity market discussed above, it is necessary to develop a multiple model approach to describe the electricity price during peak hours and off-peak hours, respectively. In section 3.2 derivation of the proposed method is presented and illustrated by simulation and experiment. Specific application of the proposed method to Alberta's electricity price prediction is provided in section 3.3.

3.2 Switched Markov ARX identification with Gaussian mixture noise

This section is organized as follows: 1) derivation of switched Markov ARX model with Gaussian mixture noise (SMARX-GMM) identification algorithm based on EM algorithm, extension to second-order Markov switching dynamics and a novel initialization strategy are introduced in subsection 3.2.1; 2) numerical simulation examples, CSTR simulation example and pilot-scale experiment illustration are demonstrated

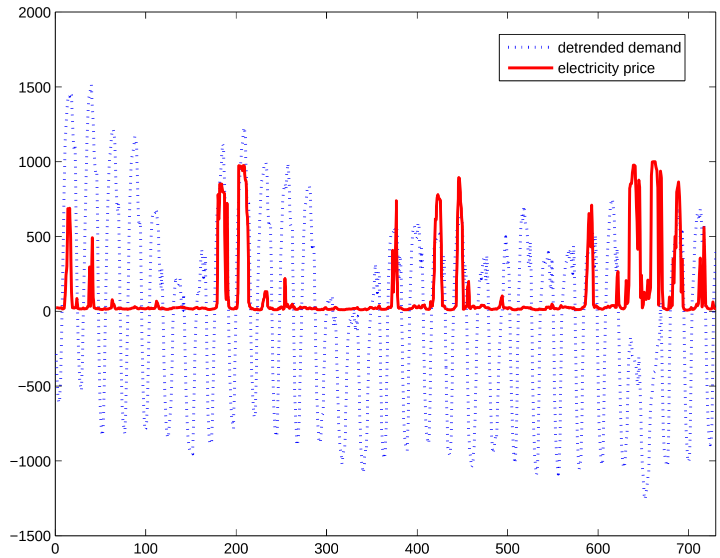


Figure 3.2: Alberta electricity price and demand correlation

and analyzed in subsection 3.2.2; and 3) comparative study and cross validation results are shown to support the proposed method.

3.2.1 Formulation and analytical solution for SMARX-GMM identification problem

In this subsection, a mathematical formulation of a switched Markov ARX model with Gaussian mixture noise is presented, and an EM algorithm based solution is developed. Both process models and noise distributions, in both discrete and continuous valued dynamics will be estimated.

The following problems will be solved in an SMARX-GMM identification framework: (1) determining which sub-model each data point is from (hidden model identities); (2) identifying parameters for each local ARX model; (3) identifying switching parameters such as the transition matrix of hidden Markov model; (4) determining which noise distribution each data point is disturbed with (hidden noise memberships); and (5) identifying parameters for each local Gaussian noise component.

A general formulation and parameter estimation for SMARX-GMM model will be presented and followed by analysis of two special cases. Application to modelling with outliers will be explained thereafter. In the last part of this subsection, an extension of SMARX identification to second-order Markov switching dynamics will be developed.

SMARX model with Gaussian mixture noise

In this part, a general SMARX-GMM model will be formulated and the parameters will be estimated. The following assumptions are made: 1) each local model is autoregressive eXogenous (ARX) model since a high-order ARX model is capable of approximating any linear dynamic model [43]; 2) the orders of local ARX model and number of local sub-models are known a priori [44], [13] and [14]; 3) switching dynamics among local models is governed by a Hidden Markov model (HMM); 4) the noise distribution is approximated by a Gaussian mixture model (GMM); 5) the number of Gaussian components is known a priori and this number can be determined by the required accuracy for the noise distribution approximation; and 6) switching among local noise distributions is random.

It is convenient to consider the SMARX identification problem assuming a GMM noise distribution, since difficulties encountered in traditional identification methods, such as handling outliers, can be solved naturally in this framework. Taking advantage of the EM algorithm allows both process and noise model parameters to both discrete and continuous switching dynamics, be estimated in a unified way. The problem formulation is as follows:

$$y_k = \begin{cases} \theta_1^T x_k \\ \vdots \\ \theta_M^T x_k \end{cases} + \begin{cases} e_{1,k} \\ \vdots \\ e_{L,k} \end{cases} , k = 1, 2, \ldots, N \tag{3.1}$$

where $y_k \in R$, $x_k \in R^n$ are the output and regressor variables of the system, respectively at kth time instant. The regressor x_k can be further expressed as (3.2):

$$x_k = \left[y_{k-1}, y_{k-2}, \ldots, y_{k-na}, u_{k-1}^T, u_{k-2}^T, \ldots, u_{k-nb}^T, 1 \right]^T \tag{3.2}$$

where: na and nb are the orders of the output and input dynamics, and are assumed to be known a priori; $u \in R^m$ is the input variable and $n = na + m \cdot nb$; $I_k = 1, 2, \ldots, M$ is defined as the hidden model identity at the time instant k, and its switching dynamics follow first-order hidden Markov model; $\theta_{I_k} \in R^{n+1}$ is the parameter vector for I_kth local model; $M_k = 1, 2, \ldots, L$ is defined as the noise distribution membership of time instant k, and its switching dynamics are random; and $e_{M_k,k} \in R$ corresponds to noise value at kth time instant from M_kth Gaussian distribution component with mean μ_{M_k} and variance $\sigma^2_{M_k}$. At any time instant, one of M local models and one of L noise distributions govern the behaviour of the switched system. To proceed, the EM algorithm needs to be revisited in the switched system identification context.

An EM algorithm consists of two steps: the E-step and the M-step. The algorithm updates the estimated parameters iteratively using these two steps until convergence. In the E-step, the conditional expectation of the likelihood function over hidden states is calculated, and the result is normally referred to as the Q function in the literature [15], [35]. The mathematical formulation of the Q function for SMARX system can be expressed as:

$$Q(\Theta|\Theta^{old}) = E_{I|(\Theta^{old}, C_{obs})}\{logP(C_{obs}, I|\Theta)\} \tag{3.3}$$

where: Θ are the parameters of the SMARX model, including both process and noise, in both discrete and continuous valued dynamic models; Θ^{old} are the parameters estimated from last iteration; I is the model identity of each time instant, which is treated as hidden variable; and C_{obs} are the observed data set, including output and regressor, which can be further denoted as $Z_N, Z_{N-1}, \ldots, Z_1$ for each time instant. By assuming that the evolution of hidden variable is governed by the following first-order Markov property:

$$P(I_k|I_{k-1}, \ldots, I_1) = P(I_k|I_{k-1}); \tag{3.4}$$

and that the conditional probability of observing Z_k is dependent on the hidden model identity at time instant k:

$$P(Z_k|Z_{k-1}, \ldots, Z_1, I_k, I_{k-1}, \ldots, I_1, \Theta) = P(Z_k|Z_{k-1}, \ldots, Z_1, I_k, \Theta); \tag{3.5}$$

Following [14], at the end of E-step, the Q function can be written as:

$$
\begin{aligned}
Q\left(\theta|\theta_{old}\right) = &\sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta^{old}, C_{obs}\right) \log P\left(Z_k|Z_{k-1}, ..., Z_1, \theta_i\right) \\
&+ \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=2}^{N} P\left(I_k = i, I_{k-1} = j|\theta^{old}, C_{obs}\right) \log \alpha_{ij} \\
&+ \sum_{i=1}^{M} P\left(I_1 = i|\theta^{old}, C_{obs}\right) \log \pi_i
\end{aligned} \tag{3.6}
$$

where: α_{ij} is the transition probability from hidden state i to hidden state j, governed by the hidden Markov model; and π_i is the probability that ith local model takes effect as initial state. Both α_{ij} and π_i are elements of Θ. In the M-step, the Q function is maximized over Θ to derive the update equation for each continuous model parameter. Estimation of discrete model parameters, such as model identity and noise membership at each time instant, will be explained shortly. The analytical update equations are calculated using optimality condition and probability measure constraints via the Karush-Kuhn-Tucker (KKT) conditions [45], [46]. The solution strategy is to separate the coupled estimation problem of process models and noise distributions: 1) update process model parameters based on GMM noise distribution using EM algorithm at each step; 2) calculate the estimation error sequence based on newly updated model parameters; 3) model the error sequence as a Gaussian mixture model, using EM algorithm to estimate the noise distribution parameters; and 4) repeat these steps until convergence. Therefore, the estimation problem will now be split into process model parameter estimation and noise distribution parameter estimation. Following the optimization procedure using KKT condition, the update equations for local process model parameters and model identity switching dynamics can be determined.

The update equation for model parameters are only relevant to the first term in equation (3.6), which is:

$$
\theta_i = \arg\left\{\max\left\{\sum_{k=1}^{N} P\left(I_k = i|\theta^{old}, C_{obs}\right) \log\left(\sum_{j=1}^{L} p_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(\frac{\left(y_k - x_k\theta_i - \mu_j\right)^2}{2\sigma_j^2}\right)\right)\right\}\right\} \tag{3.7}
$$

It is found that the direct solution for θ_i from equation (3.7) is difficult, and thus rather than maximizing the expression in (3.7) over θ_i, we try to maximize its lower bound using Jensen's inequality:

$$\begin{aligned}\sum_{k=1}^{N} P(I_k = i|\theta^{old}, C_{obs}) \log \left(\sum_{j=1}^{L} p_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(\frac{(y_k - x_k\theta_i - \mu_j)^2}{2\sigma_j^2} \right) \right) \\ \geq \sum_{k=1}^{N} P\left(I_k = i|\theta^{old}, C_{obs}\right) \sum_{j=1}^{L} p_j \log \left(\frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(\frac{(y_k - x_k\theta_i - \mu_j)^2}{2\sigma_j^2} \right) \right)\end{aligned} \tag{3.8}$$

and θ_i is thus updated by:

$$\begin{aligned}\theta_i &= \arg\left\{ \max\left\{ \sum_{k=1}^{N} P\left(I_k = i|\theta^{old}, C_{obs}\right) \sum_{j=1}^{L} p_j \log \left(\frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(\frac{(y_k - x_k\theta_i - \mu_j)^2}{2\sigma_j^2} \right) \right) \right\} \right\} \\ &= \arg\left\{ \max\left\{ \sum_{k=1}^{N} P\left(I_k = i|\theta^{old}, C_{obs}\right) \sum_{j=1}^{L} p_j \frac{(y_k - x_k\theta_i - \mu_j)^2}{2\sigma_j^2} \right\} \right\}\end{aligned} \tag{3.9}$$

After taking the derivative with respect to θ_i, θ_i is solved to be:

$$\theta_i = -\frac{A \sum_{k=1}^{N} P\left(I_k = i|\theta^{old}, C_{obs}\right) x_k - B \sum_{k=1}^{N} P\left(I_k = i|\theta^{old}, C_{obs}\right) x_k^T y_k}{B \sum_{k=1}^{N} P\left(I_k = i|\theta^{old}, C_{obs}\right) x_k^T x_k} \tag{3.10}$$

$$A = \sum_{j=1}^{L} \frac{p_j \mu_j}{\sigma_j^2}; \qquad B = \sum_{j=1}^{L} \frac{p_j}{\sigma_j^2}$$

where p_j, μ_j and σ_j can be substituted using the noise parameter estimation from last step. The transition probability and initial probability of HMM can be updated as follows:

$$\alpha_{ij}^{new} = \frac{\sum_{k=2}^{N} P\left(I_k = i, I_{k-1} = j|C_{obs}, \theta^{old}\right)}{\sum_{i=1}^{M} \sum_{k=2}^{N} P\left(I_k = i, I_{k-1} = j|C_{obs}, \theta^{old}\right)} \tag{3.11}$$

$$\pi_i^{New} = P\left(I_1 = i|C_{obs}, \theta^{old}\right) \tag{3.12}$$

where: $P(I_k = i, I_{k-1} = j|\theta_{old}, C_{obs})$ is the probability that at the kth time point, the ith local model takes effect; and at the $k-1$th time point, the jth local model takes effect. It can be further expanded as (3.13) following Bayes' rule,

$$
\begin{aligned}
P(I_k = i, \quad & I_{k-1} = j|C_{obs}, \theta^{old}) \\
&= \frac{P\left(Z_k, I_k = i, I_{k-1} = j|Z_{k-1}, ..., Z_1, \theta^{old}\right)}{P\left(Z_k|Z_{k-1}, ..., Z_1, \theta^{old}\right)} \\
&= \frac{P\left(Z_k|I_k = i, I_{k-1} = j, Z_{k-1}, ..., Z_1, \theta^{old}\right) P\left(I_k = i, I_{k-1} = j|Z_{k-1}, ..., Z_1, \theta^{old}\right)}{\sum_{m=1}^{M}\sum_{n=1}^{M} P\left(Z_k|I_k = m, I_{k-1} = n, Z_{k-1}, ..., Z_1, \theta^{old}\right) P\left(I_k = m, I_{k-1} = n|Z_{k-1}, ..., Z_1, \theta^{old}\right)} \\
&= \frac{P\left(Z_k|I_k = i, Z_{k-1}, ..., Z_1, \theta^{old}\right) P\left(I_k = i|I_{k-1} = j, \theta^{old}\right) P\left(I_{k-1} = j|Z_{k-1}, ..., Z_1, \theta^{old}\right)}{\sum_{m=1}^{M}\sum_{n=1}^{M} P\left(Z_k|I_k = m, Z_{k-1}, ..., Z_1, \theta^{old}\right) P\left(I_k = m|I_{k-1} = n, \theta^{old}\right) P\left(I_{k-1} = n|Z_{k-1}, ..., Z_1, \theta^{old}\right)} \\
&= \frac{P\left(Z_k|I_k = i, Z_{k-1}, ..., Z_1, \theta^{old}\right) {\alpha^{old}}_{ji} P_{j,k-1}}{\sum_{m=1}^{M}\sum_{n=1}^{M} P\left(Z_k|I_k = m, Z_{k-1}, ..., Z_1, \theta^{old}\right) {\alpha^{old}}_{nm} P_{n,k-1}}
\end{aligned}
\tag{3.13}
$$

where the model identity probability $P(I_k = i|\theta_{old}, C_{obs})$ is denoted as $P_{i,k}$ in (3.13) and (3.14). The first two equalities in (3.13) follows from Bayes' rule, the third equality follows from first order Markov properties (3.4) and (3.5), and the final equality uses the notation of model identity probability $P_{i,k}$ and the transition matrix element α_{ji}. Using the law of total probability $P(A) = \sum_{j=1}^{M} P(A|B)P(B)$ to marginalize the probability (3.13) to $P(I_k = i|\theta_{old}, C_{obs})$, namely $P_{i,k}$, following equation can be derived:

$$
\begin{aligned}
P(I_k = i|\theta^{old}, Z_k, ..., Z_1) &= P_{i,k} \\
&= \sum_{j=1}^{M} P\left(I_k = i, I_{k-1} = j|C_{obs,}\theta^{old}\right) \\
&= \frac{\sum_{j=1}^{M} P\left(Z_k|I_k = i, \theta^{old}, Z_{k-1}, ..., Z_1\right) \alpha_{ji}^{old} P_{j,k-1}}{\sum_{m=1}^{M}\sum_{n=1}^{M} P\left(Z_k|I_k = m, \theta^{old}, Z_{k-1}, ..., Z_1\right) \alpha_{nm}^{old} P_{n,k-1}}
\end{aligned}
\tag{3.14}
$$

The identity probability is updated inductively using (3.14), starting from $P(I_1 = i|Z_1, \theta^{old})$, which is initial state probability (3.12). This Bayes' rule based induction method, compared to updating methods based on propagation of Markov chain in the SMARX identification literature [14]:

$$P_{i,k-1} = [(\alpha_{ij})^{i-2}\pi_{old}]_{ith} \tag{3.15}$$

can include both process information and switching dynamics in the updating equation, so that better model identity clustering performance can be expected.

Following definition (2.1) and SMARX-GMM problem formulation (3.1), the GMM model can be expressed as:

$$e_k \sim \sum_{j=1}^{L} p_j N\left(\mu_j, \sigma_j^2\right) \tag{3.16}$$

where: p_j is mixing probability indicating how frequent each local Gaussian component takes effect; N represents the Gaussian density function; and μ_j, σ_j^2 are local Gaussian mean and variance, respectively. The probability of observing Z_k given the model identity I_k, historical data and parameter θ_{I_k} can be written as

$$P\left(Z_k | I_k = i, \theta_{I_k}, Z_{k-1}, ..., Z_1\right) = \sum_{j=1}^{L} p_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(e_{k,i} - \mu_j)^2}{2\sigma_j^2}\right) \tag{3.17}$$

where $e_{k,i} = y_k - \theta_i^T x_k$ is the error when the ith local model takes effect at time instant k. For estimation of noise distribution parameters, μ_j and σ_j, substitute (3.17) to (3.6) (where $I_k = i$ and θ_{I_k} are together denoted to be θ_i). First-order derivative of Q function is taken over each μ_j and σ_j, respectively, and then equating the results to zero:

$$\frac{\partial \sum\limits_{i=1}^{M} \sum\limits_{k=1}^{N} P\left(I_k = i | \theta_{old}, C_{obs}\right) \log\left(\sum\limits_{j=1}^{L} P_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(e_{k,i}-\mu_j)^2}{2\sigma_j^2}\right)\right)}{\partial \mu_j} = 0 \tag{3.18}$$

$$\frac{\partial \sum\limits_{i=1}^{M} \sum\limits_{k=1}^{N} P\left(I_k = i | \theta_{old}, C_{obs}\right) \log\left(\sum\limits_{j=1}^{L} P_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(e_{k,i}-\mu_j)^2}{2\sigma_j^2}\right)\right)}{\partial \sigma_j} = 0 \tag{3.19}$$

For notational simplicity, following probabilities are defined:

$$q\left(j, e_{k,i}\right) = \quad p_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{\left(e_{k,i} - \mu_j\right)^2}{2\sigma_j^2}\right) \tag{3.20}$$

$$p\left(j | e_{k,i}\right) = \frac{q\left(j, e_{k,i}\right)}{\sum\limits_{m=1}^{L} q\left(m, e_{k,i}\right)} \tag{3.21}$$

In the Gaussian mixture estimation literature [37], (3.21) is referred to as the membership probability, representing the possibility that jth Gaussian component takes effect at kth time instant given the estimated noise observation $e_{k,i}$. Using the notation in (3.21), and the optimality condition (3.18), (3.19), the update equation for μ_j and σ_j^2 can be derived as:

$$\mu_j{}^{new} = \frac{\sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta_{old}, C_{obs}\right)p\left(j|e_{k,i}\right)\left[y_k - (\theta_i^{New})^T x_k\right]}{\sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta_{old}, C_{obs}\right)p\left(j|e_{k,i}\right)} \tag{3.22}$$

$$\left(\sigma_j^2\right)^{new} = \frac{\sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta_{old}, C_{obs}\right)p\left(j|e_{k,i}\right)\left(e_{k,i} - \mu_j^{new}\right)^2}{\sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta_{old}, C_{obs}\right)p\left(j|e_{k,i}\right)} \tag{3.23}$$

The estimation of mixing probability p_j is relatively complicated, since there are probability measure constraints: $p_j \geq 0$ and $\sum_{j=1}^{L} p_j = 1$. Unlike in (3.11) and (3.12), optimization under these constraints is difficult, and therefore a transformation technique is applied to form an unconstrained optimization problem [37]. Transformation of p_j is:

$$p_j = \frac{e^{\gamma j}}{\sum_{m=1}^{L} e^{\gamma m}} \tag{3.24}$$

$$\frac{\partial p_j}{\partial \gamma_m} = \begin{cases} p_j - p_j^2, if j = m \\ -p_j p_m, else \end{cases}$$

where e^{γ_j} is exponential function of γ_j. Given the unconstrained variables γ_j, p_j must always satisfy the properties of probability measure, so instead of optimization over p_j, the Q function can be optimized over γ_j to get the update equation:

$$\frac{\partial \sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta_{old}, C_{obs}\right)\log\left\{\sum_{n=1}^{L}\frac{e^{\gamma n}}{\sum_{m=1}^{L} e^{\gamma m}}\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{(e_{k,i}-\mu_n)^2}{2\sigma_n^2}\right]\right\}}{\partial \gamma_j} = 0 \tag{3.25}$$

and the corresponding p_j for the maximized Q function is:

$$p_j{}^{new} = \frac{\sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta_{old}, C_{obs}\right)p\left(j|e_{k,i}\right)}{N} \tag{3.26}$$

In the update equations for μ_j,σ_j and p_j, the model identity probability, namely $P(I_k = i|\theta_{old}, C_{obs})$, has been used, and it can be updated using induction equation (3.14) and conditional probability (3.17). For estimation of the noise membership probability $p(j|e_{k,i})$, definition (3.20) and (3.21) are used. Notice that the update of process model and noise distribution parameters (3.10), (3.11), (3.12), (3.22), (3.23), (3.26) depends on identity and membership probabilities (3.14) and (3.21), and the estimation of these probabilities in turn depends on model and noise parameters. This fact results in two different EM algorithm initialization strategies.

Even though the EM algorithm is proven to converge, there is the possibility that the convergent point is only stationary rather than a global maxima. Thus, the initialization of EM algorithm is important with respect to the convergence point [35], [47]. Due to the fact that in SMARX-GMM identification, the update equations for parameters and several important probabilities are closely related to each other, one can either choose to directly initialize model and noise parameters or choose to initialize probabilities (3.14) and (3.21) instead. Traditionally, direct initialization of parameters is widely used, and strategies such as pre-running EM algorithm with random initial guesses for multiple times are very popular [47],[13]. This strategy, though simple in implementation, is an ad-hoc method and cannot make good use of prior process information to facilitate calculation.

For example, in a biological fermenter, it is known that different temperatures can result in different operating points. Although the influence of temperature on system dynamics is only fuzzily known, it is worth investigating how this information can be used in initialization of EM algorithm for better convergence performance.

In the method proposed in this section, initializing identity/membership probabilities is adopted instead of the traditional approach to initializing parameters directly, because in this way prior process and noise information can be used. The specific initialization procedures are:

1. After specifying the number of sub-process models M, the variable(s) most representative of hidden operating states is classified to different membership symbols. For example, in the biological fermenter case, temperature is chosen

as the variable to indicate operating point, and the temperature measurement sequence is discretized into HIGH, MIDDLE, and LOW, according to their relative values. Based on prior knowledge, at each operating point, the occurrence pattern of these symbols should be different.

2. A Hidden Markov model (HMM) is trained based on the discretized measurement sequence of symbols, and the corresponding transition probability matrix and emission probability matrix are estimated.

3. Process the estimated transition matrix and emission matrix, together with the sequence of symbols in (1) using an inference algorithm for HMM, such as $\alpha-\beta$ algorithm [48]. Initial values of identity probabilities are calculated as a result. The returned identity probability sequence is the most probable inference of local models at each time instant based on the indication of representative variable(s).

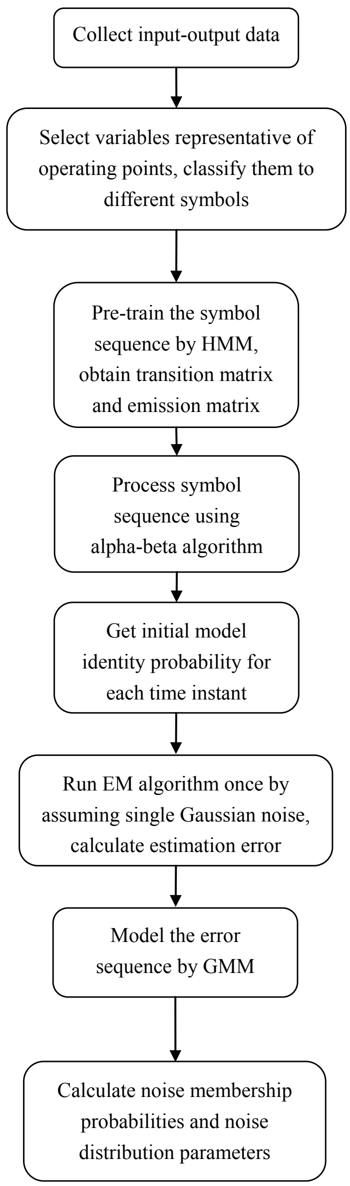
4. After initialization of the identity probability, run the EM algorithm once assuming single Gaussian distributed noise. Local model parameters and switching dynamics will be calculated. Use these parameters together with input-output data to calculate the estimation error.

5. As items in 1., 2. and 3., classify the error sequence to symbols representing different noise levels, and train the HMM using the symbol sequence to get a sequence of initial values for membership probability by $\alpha - \beta$ algorithm. Use the membership probability to initialize the EM algorithm in estimating the GMM parameters for the error sequence.

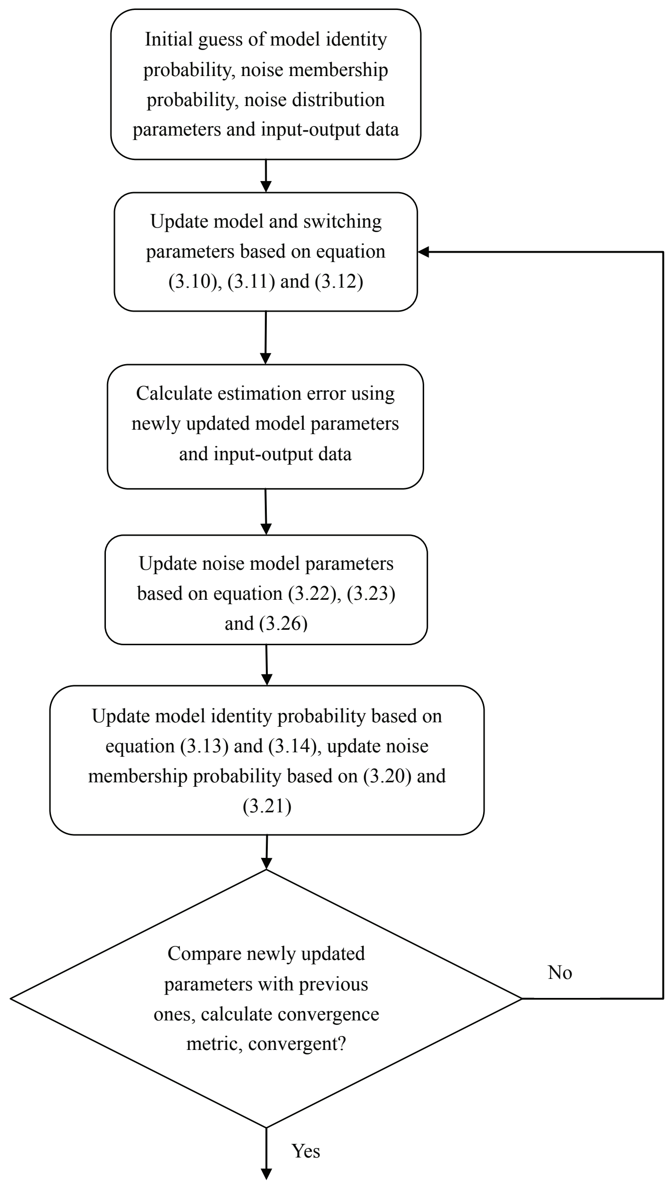
6. Run the EM algorithm until convergence with initialization of identity probability, membership probability and the noise parameters estimated in item 5.

The initialization strategy and SMARX-GMM identification algorithm flowcharts are shown in Figure 3.3.

For the clustering of model identities and noise memberships at each time instant, the calculation is based on the model identity and noise membership probabilities at



(a) Initialization strategy



(b) SMARX-GMM identification

Figure 3.3: SMARX-GMM initilization strategy and algorithm flowchart

the final iteration of algorithm. The candidate for model identity/noise membership with the highest calculated probability based on equation (3.14) and (3.21) is inferred as the true model identity/noise membership at each time instant.

Two special cases in SMARX-GMM identification

In this part, the proposed method is approved to two special cases of the SMARX-GMM model.

First of all, a single Gaussian distributed noise case is considered. In this case, $j = 1$, $p_j = 1$ and thus the parameters to be estimated reduce to:

$$\mu^{new} = \frac{\sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta_{old}, C_{obs}\right)[y_k - (\theta_i^{New})^T x_k]}{N} \tag{3.27}$$

$$\left(\sigma^2\right)^{new} = \frac{\sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta_{old}, C_{obs}\right)(e_{k,i} - \mu^{new})^2}{N} \tag{3.28}$$

and accordingly, the conditional probability of observing Z_k reduces to:

$$P\left(Z_k|I_k = i, \theta^{old}, Z_{k-1}, ..., Z_1\right) = \frac{1}{\sqrt{2\pi(\sigma^2)^{new}}}\exp^{-\frac{1}{2(\sigma^2)^{new}}\left(e_{k,i}-\mu^{new}\right)^2} \tag{3.29}$$

Substituting these reduced results to (3.14), (3.13), and then to (3.10), (3.11), (3.12) (notice that in this simple case, the parameter A in (3.10) reduces to μ and B reduces to 1). The resulting update equations are the EM solution to the single Gaussian noise SMARX identification problem. As expected, this result is consistent with traditional single Gaussian noise SMARX identification in the literature [14].

The second special case of SMARX-GMM model is the SMARX model with operating-point-dependent noise. In industrial operations, valves, meters or sensors have limited operating range, and out-of-range operations will lead to inaccurate results. These factors can introduce different noise distributions around different operating point. As a result, it is reasonable to assume that noise distributions will change with different process operating points. In this case, it is assumed that process model and noise distribution switching occurs at the same time (i.e., each process model has a corresponding noise distribution). To be specific, the problem is:

$$y_k = \begin{cases} \theta_1^T x_k + e_{1,k}, & \text{if } I_k = 1; \\ \vdots & \qquad\qquad k = 1, 2, \ldots, N \\ \theta_M^T x_k + e_{M,k}, & \text{if } I_k = M; \end{cases} \tag{3.30}$$

where the notations is the same as that introduced in subsection 3.2.1, and $e_{I_k,k}$ is assumed to be white noise with zero mean and variance $\sigma_{I_k}^2$. Applying the EM algorithm to this formulation, the E-step result is the same as in Equation (3.6). So the update equations for process parameters (3.10), (3.11), (3.12) still hold (in (3.10), A reduces to 0 and B reduces to 1 in this case). In this formulation, the probability of observing Z_k, given the model identity I_k, historical data and θ^{old} can be written as:

$$P\left(Z_k | I_k = i, \theta^{old}, Z_{k-1}, ..., Z_1\right) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp^{-\frac{1}{2\sigma_i{}^2}\left(y_k - \theta_i^T x_k\right)^2} \tag{3.31}$$

Notice that the subscript of model parameter θ and noise parameter σ are the same. Following similar procedures as in subsection 3.2.1, the estimation of σ_i requires:

$$\frac{\partial \sum_{i=1}^{M} \sum_{k=1}^{N} P\left(I_k = i | \theta_{old}, C_{obs}\right) \log\left[\frac{1}{\sqrt{2\pi\sigma_i}} \exp\left(-\frac{e_{k,i}^2}{2\sigma_i^2}\right)\right]}{\partial \sigma_i} = 0 \tag{3.32}$$

and:

$$(\sigma_i^{new})^2 = \frac{\sum_{k=1}^{N} P\left(I_k = i | C_{obs}, \theta^{old}\right) \left[y_k - (\theta_i^{New})^T x_k\right]^T \left[y_k - (\theta_i^{New})^T x_k\right]}{\sum_{k=1}^{N} P\left(I_k = i | C_{obs}, \theta^{old}\right)} \tag{3.33}$$

As the noise mean is assumed to be zero, σ_i is the only noise parameter to be estimated. Therefore, at this point, SMARX-GMM identification of the second special formulation is completely solved. After taking the initialization strategy discussed previously, the EM algorithm will run iteratively and parameters are updated at each iteration until convergence.

High-order SMARX model

In previous subsections, a first-order hidden Markov model is assumed to govern the evolution of process sub-models. This assumption, widely used in modelling, control and optimization in chemical process research fields [49],[50], though concise in formulation, has its limitation in describing complex patterns of switching dynamics [23], [24], [25]. [22] .

In this part, a second-order hidden Markov model is adopted to govern the switching dynamics. In this case, both the E-step and the M-step in the EM algorithm need to be re-derived due to the structural difference in problem formulation. White noise is assumed as the noise distribution. The corresponding high-order SMARX-GMM model identification results can be derived in a similar way by combining results of this part and subsection 3.2.1.

Some properties of the second-order SMARX model are given in following equations:

$$
\begin{aligned}
&P\left(I_k|Z_{k-1},...,Z_1,I_{k-1},...,I_1,\theta\right)=P\left(I_k|I_{k-1},I_{k-2}\right)\\
&P\left(Z_k|Z_{k-1},...,Z_1,I_k,I_{k-1},...,I_1,\theta\right)=P\left(Z_k|Z_{k-1},...,Z_1,I_k,\theta\right)
\end{aligned}
\tag{3.34}
$$

The first equality in (3.34) follows second-order Markov property and the second equality implies that an observation at time instant k is solely dependent on the I_kth local model and historical data. Using these equations, the Q function is:

$$
\begin{aligned}
Q\left(\theta|\theta_{old}\right) &= E_{I|(\theta_{old},C_{obs})}\left\{\log P\left(C_{obs},I|\theta\right)\right\}\\
&= E_{I|(\theta_{old},C_{obs})}\left\{\log P\left(Z_N,Z_{N-1},...,Z_1,I_N,...,I_1|\theta\right)\right\}\\
&= E_{I|(\theta_{old},C_{obs})}\left\{\log\prod_{k=1}^{N}P\left(Z_K,I_K|Z_{K-1},...,Z_1,I_{K-1},...,I_1,\theta\right)\right\}
\end{aligned}
\tag{3.35}
$$

Equation (3.35) is derived from the definition of Q function, and use of the probability chain rule. Then, the Q function can be further written as:

$$
\begin{aligned}
Q\left(\theta|\theta_{old}\right) &= E_{I|(\theta_{old},C_{obs})}\left\{\log\prod_{k=1}^{N}P\left(Z_K|Z_{K-1},...,Z_1,I_K,I_{K-1},...,I_1,\theta\right)P\left(I_K|I_{K-1},I_{K-2}\right)\right\}\\
&= E_{I|(\theta_{old},C_{obs})}\left\{\begin{array}{l}\sum_{k=3}^{N}\left[\log P\left(Z_K|Z_{K-1},...,Z_1,I_K,\theta\right)+\log P\left(I_K|I_{K-1},I_{K-2}\right)\right]\\ +\log P\left(Z_2|Z_1,I_2,\theta\right)+\log P\left(I_2|I_1\right)\\ +\log P\left(Z_1|I_1,\theta\right)+\log P\left(I_1\right)\end{array}\right\}
\end{aligned}
\tag{3.36}
$$

Equation (3.36) expands the final equality in (3.35) using the conditional probability formula and then uses second-order SMARX model properties (3.34) to simplify the expression. The final equality in (3.36) separates the first two terms from the summation for convenience in defining initial conditions. The Q function can then be derived as:

$$
\begin{aligned}
Q\left(\theta|\theta_{old}\right) &= \sum_{I} P\left(I|\theta^{old}, C_{obs}\right) \\
&\left(\begin{array}{l} \sum\limits_{k=1}^{N} \log P\left(Z_K|Z_{K-1},...,Z_1,I_K,\theta\right) + \sum_{k=3}^{N}\log P\left(I_K|I_{K-1},I_{K-2}\right) \\ + \log P\left(I_2|I_1\right) + \log P\left(I_1\right) \end{array}\right) \\
&= \sum_{I_1}...\sum_{I_N} P\left(I_1,...,I_N|\theta^{old}, C_{obs}\right) \\
&\left(\begin{array}{l} \sum\limits_{k=1}^{N} \log P\left(Z_K|Z_{K-1},...,Z_1,I_K,\theta\right) + \sum_{k=3}^{N}\log P\left(I_K|I_{K-1},I_{K-2}\right) \\ + \log P\left(I_2|I_1\right) + \log P\left(I_1\right) \end{array}\right) \\
&= \sum_{i=1}^{M}\sum_{k=1}^{N} P\left(I_k = i|\theta^{old}, C_{obs}\right) \log P\left(Z_K|Z_{K-1},...,Z_1,\theta_i\right) \\
&+ \sum_{l=1}^{M}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=3}^{N} P\left(I_k = i, I_{k-1} = j, I_{k-2} = l|\theta^{old}, C_{obs}\right) \log \alpha_{lj,i} \\
&+ \sum_{i=1}^{M}\sum_{j=1}^{M} P\left(I_1 = i, I_2 = j|\theta^{old}, C_{obs}\right) \log \pi_{ji} + \sum_{i=1}^{M} P\left(I_1 = i|\theta^{old}, C_{obs}\right) \log \pi_i
\end{aligned}
\tag{3.37}
$$

For the second-order hidden Markov model, the transition probability is defined as the probability that the hidden variable jumps from two consecutive states to a new one, denoted as $\alpha_{lj,i}$. Thus, the transition matrix is an $M^2 \cdot M$ dimensional matrix with row elements summing up to 1. Equation (3.37) follows from the definition of statistical expectation and terms irrelevant in the summation indices are marginalized so that $P\left(I_1,...,I_N|\theta^{old}, C_{obs}\right)$ reduces to probabilities such as $P\left(I_k = i|\theta^{old}, C_{obs}\right)$, where only I_kth term is relevant in the summation, and $P\left(I_k = i, I_{k-1} = j, I_{k-2} = l|\theta^{old}, C_{obs}\right)$, where I_k, I_{k-1}, I_{k-2} are relevant to the summation. Also notice that in second-order HMM, the model identity probabilities for several consecutive time instants are introduced in the Q function, and this includes $P\left(I_k = i, I_{k-1} = j, I_{k-2} = l|\theta^{old}, C_{obs}\right)$, $P\left(I_1 = i, I_2 = j|\theta^{old}, C_{obs}\right)$, and also $P\left(I_k = i|\theta^{old}, C_{obs}\right)$ as in previous subsections. Two initial conditions π_{ji} and π_i are as expected due to higher switching dynamics order (i.e., first-order requires one initial condition, and second-order requires two initial conditions). In the M-step, the Q function in (3.37) along with optimality conditions and probability measure constraints are used together to derive the closed form analytical update equations for the process model, noise and switching dynamics parameters:

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

$$\alpha_{lj,i}^{new} = \frac{\sum_{k=3}^{N} P\left(I_k = i, I_{k-1} = j, I_{k-2} = l | C_{obs}, \theta^{old}\right)}{\sum_{i=1}^{M} \sum_{k=3}^{N} P\left(I_k = i, I_{k-1} = j, I_{k-2} = l | C_{obs}, \theta^{old}\right)}$$

$$\pi_i^{New} = P\left(I_1 = i | C_{obs}, \theta^{old}\right) \tag{3.38}$$

$$\pi_{ji}^{New} = P\left(I_1 = i, I_2 = j | C_{obs}, \theta^{old}\right)$$

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

As mentioned earlier, model identity probabilities for consecutive time instants are needed for the parameter estimation in (4.35). These probabilities can be derived using Bayesian induction in a similar fashion to that in (3.13) and (3.14). Details of the derivation are shown in the following equations:

$$\begin{aligned}
P\quad &\left(I_k = i, I_{k-1} = j, I_{k-2} = l | C_{obs}, \theta^{old}\right) \\
&= \frac{P\left(Z_k, I_k=i, I_{k-1}=j, I_{k-2}=l | C_{obs}, \theta^{old}\right)}{P\left(Z_k | C_{obs}, \theta^{old}\right)} \\
&= \frac{P\left(Z_k | I_k=i, I_{k-1}=j, I_{k-2}=l, \theta^{old}, Z_{k-1}, \ldots, Z_1\right) P\left(I_k=i | I_{k-1}=j, I_{k-2}=l, \theta^{old}\right) P\left(I_{k-1}=j, I_{k-2}=l | C_{obs}, \theta^{old}\right)}{\sum\limits_{m=1}^{M} \sum\limits_{n=1}^{M} \sum\limits_{s=1}^{M} P\left(Z_k, I_k=m, I_{k-1}=n, I_{k-2}=s | C_{obs}, \theta^{old}\right)} \\
&= \frac{P\left(Z_k | I_k=i, \theta^{old}, Z_{k-1}, \ldots, Z_1\right) P\left(I_k=i | I_{k-1}=j, I_{k-2}=l, \theta^{old}\right) P\left(I_{k-1}=j, I_{k-2}=l | C_{obs}, \theta^{old}\right)}{\sum\limits_{m=1}^{M} \sum\limits_{n=1}^{M} \sum\limits_{s=1}^{M} P\left(Z_k | I_k=m, \theta^{old}, Z_{k-1}, \ldots, Z_1\right) P\left(I_k=m | I_{k-1}=n, I_{k-2}=s, \theta^{old}\right) P\left(I_{k-1}=n, I_{k-2}=s | C_{obs}, \theta^{old}\right)} \\
&= \frac{P\left(Z_k | I_k=i, \theta^{old}, Z_{k-1}, \ldots, Z_1\right) \alpha_{lj,i}^{old} P\left(I_{k-1}=j, I_{k-2}=l | C_{obs}, \theta^{old}\right)}{\sum\limits_{m=1}^{M} \sum\limits_{n=1}^{M} \sum\limits_{s=1}^{M} P\left(Z_k | I_k=m, \theta^{old}, Z_{k-1}, \ldots, Z_1\right) \alpha_{sn,m}^{old} P\left(I_{k-1}=n, I_{k-2}=s | C_{obs}, \theta^{old}\right)}
\end{aligned} \tag{3.39}$$

By marginalizing (3.39), identity probabilities for the reduced model identity indices are:

$$\begin{aligned}
P(I_k = i, \quad & I_{k-1} = j|C_{obs}, \theta^{old}) \\
&= \sum_{l=1}^{M} P\left(I_k = i, I_{k-1} = j, I_{k-2} = l|C_{obs}, \theta^{old}\right) \\
&= \frac{\sum\limits_{l=1}^{M} P\left(Z_k|I_k=i,\theta^{old},Z_{k-1},\ldots,Z_1\right)\alpha_{lj,i}^{old} P\left(I_{k-1}=j,I_{k-2}=l|C_{obs},\theta^{old}\right)}{\sum\limits_{m=1}^{M}\sum\limits_{n=1}^{M}\sum\limits_{s=1}^{M} P\left(Z_k|I_k=m,\theta^{old},Z_{k-1},\ldots,Z_1\right)\alpha_{sn,m}^{old} P\left(I_{k-1}=n,I_{k-2}=s|C_{obs},\theta^{old}\right)}
\end{aligned} \tag{3.40}$$

$$\begin{aligned}
P(I_k = i \quad & |C_{obs}, \theta^{old}) \\
&= \sum_{j=1}^{M}\sum_{l=1}^{M} P\left(I_k = i, I_{k-1} = j, I_{k-2} = l|C_{obs}, \theta^{old}\right) \\
&= \frac{\sum\limits_{j=1}^{M}\sum\limits_{l=1}^{M} P\left(Z_k|I_k=i,\theta^{old},Z_{k-1},\ldots,Z_1\right)\alpha_{lj,i}^{old} P\left(I_{k-1}=j,I_{k-2}=l|C_{obs},\theta^{old}\right)}{\sum\limits_{m=1}^{M}\sum\limits_{n=1}^{M}\sum\limits_{s=1}^{M} P\left(Z_k|I_k=m,\theta^{old},Z_{k-1},\ldots,Z_1\right)\alpha_{sn,m}^{old} P\left(I_{k-1}=n,I_{k-2}=s|C_{obs},\theta^{old}\right)}
\end{aligned}$$

Notice that in equation (3.39) and (3.40), identity probabilities can be expressed in a unified way as functions of $P\left(I_{k-1} = n, I_{k-2} = s|C_{obs}, \theta^{old}\right)$. Therefore, following the same Bayesian induction update strategy as in Section 3.2.1, a sequence of $P\left(I_k, I_{k-1}|C_{obs}, \theta^{old}\right)$ can be updated first using induction. $P\left(I_k, I_{k-1}, I_{k-2}|C_{obs}, \theta^{old}\right)$ and $P\left(I_k|C_{obs}, \theta^{old}\right)$ can then be calculated based on (3.39), (3.40). The proposed initialization strategy introduced in Section 3.2.1 can also apply to second-order SMARX identification algorithm, by merging each two consecutive operating-point-representative symbols to form a new sequence, and then applying $\alpha - \beta$ algorithm to calculate initial values of $P\left(I_k, I_{k-1}|C_{obs}, \theta^{old}\right)$.

3.2.2 Simulation and experimental verification

In this subsection, several examples will illustrate the proposed method. One example for general SMARX-GMM identification and one for data with outliers will be presented to show the robustness of proposed method. In addition, one example of a CSTR identification and one pilot-scale physical experiment validation will be presented to explore the proposed method with a benchmark problem and a physical system.

Numerical simulation

To illustrate the proposed algorithm, an SMARX model with two sub-system models and two Gaussian component noise distributions:

$$y_k = \begin{cases} [1, y_{k-1}, y_{k-2}, u_{k-1}, u_{k-2}] \begin{bmatrix} 10 \\ 0.5 \\ 0.3 \\ -0.2 \\ 0.5 \end{bmatrix}, I_k = 1 \\ \\ [1, y_{k-1}, y_{k-2}, u_{k-1}, u_{k-2}] \begin{bmatrix} 100 \\ -0.5 \\ -0.3 \\ 0.2 \\ -0.5 \end{bmatrix}, I_k = 2 \end{cases} + e_k, k = 1, 2, \ldots, N \tag{3.41}$$

where the noise distribution is:

$$e_k \sim 0.7N(3, 0.25) + 0.3N(-7, 1) \tag{3.42}$$

The hidden model identity I_k can take values of 1 or 2, and transition matrix for these two states are $\begin{bmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{bmatrix}$. The input $u(t)$ is generated from a Gaussian distribution: $N(0, 100)$, and 1500 data points are generated for identification use. Part of the simulated SMARX model output with Gaussian mixture noise is shown in Figure 3.4.

From Figure 3.4, it can be seen that SMARX outputs are noisy and thus the identities of the underlying hidden models are difficult to observe directly from the data sequence. Applying the proposed identification algorithm, the estimation results are shown in Table 3.1. The clustering result for model identity and noise membership

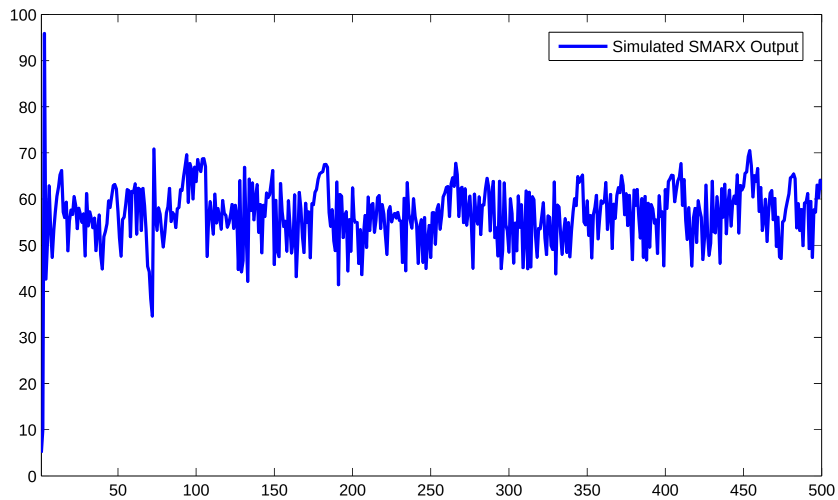


Figure 3.4: SMARX system output from simulation

are shown in Figure 3.5 and Figure 3.6. In Table 3.1, symbols without a *hat* denote the true model/noise parameters, symbols with a *hat* denote the estimated parameters using the proposed method. It can be seen from the results that the proposed method gives accurate estimation. From the clustering results in Figure 3.5 and Figure 3.6, both model identities and noise memberships are clustered accurately compared to their true counterparts. To be specific, the rate of successful clustering is 97.64% for model identity, and 99.07% for noise membership.

Now consider application of the proposed algorithm to switched system identification when data contain outliers. With the model given in (3.41), the histogram of noise in this case is shown in Figure 3.7. It can be seen that in most cases, the noise follows a Gaussian distribution centered near zero but occasionally, large value outliers appear. The influence of outliers on the output data can be seen in Figure 3.8.

It can be observed from Figure 3.8 that the output data are severely contaminated by outliers and this can significantly affect traditional identification algorithms [14], as shown in comparison results in Table 3.2. In Table 3.2, the symbols without a *hat* denote the true model/noise parameters, symbols with a *hat* and p in the parentheses

Table 3.1: EM algorithm identification results

θ_1	$\hat{\theta}_1$	θ_2	$\hat{\theta}_2$	p	$\hat{p}$	μ	$\hat{\mu}$
10	9.41	100	97.2	0.7	0.71	3	2.87
0.5	0.45	-0.5	-0.45	0.3	0.29	-7	-6.94
0.3	0.38	-0.3	-0.30	σ	$\hat{\sigma}$	α_{ij}	$\hat{\alpha_{ij}}$
-0.2	-0.31	0.2	0.16	0.5	0.59	0.95 0.05	0.95 0.05
0.5	0.43	-0.5	-0.45	1	1.09	0.1 0.9	0.12 0.88

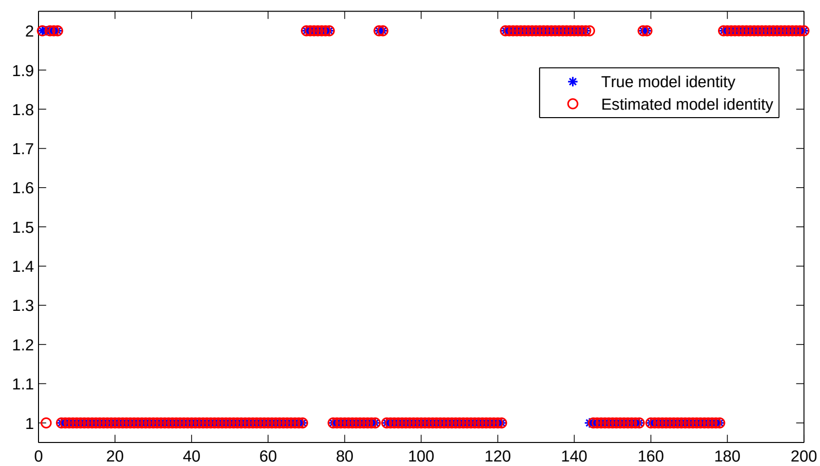


Figure 3.5: Model identity clustering result

denote the estimated parameters using the proposed method, and symbols with a *hat* and t in the parentheses denote the estimated parameters using a traditional method when the noise is assumed to be of single Gaussian distribution. It can be seen from the comparison results that for data with 5% outliers, the traditional method gives unreliable parameters estimation in both local model dynamics and switching dynamics; however, the proposed method gives accurate estimation. For instance, for the estimation of transition matrix elements α_{ij}, the traditional method results in a matrix with a dominant first column, indicating that the system tends to operate more in the first sub-model, while the proposed method gives a diagonally dominant matrix

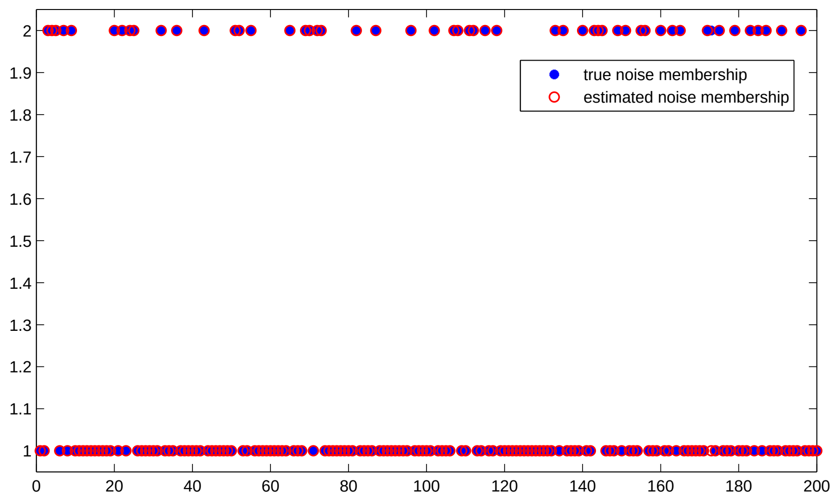


Figure 3.6: Noise membership clustering result

Table 3.2: SMARX-GMM v.s. traditional method in modelling data with outliers

θ_1	$\hat{\theta}_1(p)$	$\hat{\theta}_1(t)$	θ_2	$\hat{\theta}_2(p)$	$\hat{\theta}_2(t)$	p	$\hat{p}(p)$	$\hat{p}(t)$	μ	$\hat{\mu}(p)$	$\hat{\mu}(t)$
10	10.68	15.16	100	102.72	52.54	0.95	0.94	1	1	1.15	0
0.5	0.50	0.50	-0.5	-0.51	-1.13	0.05	0.06	NA	-19	-18.70	NA
0.3	0.28	0.24	-0.3	-0.33	0.92	σ	$\hat{\sigma}(p)$	$\hat{\sigma}(t)$	α_{ij}	$\hat{\alpha_{ij}}(p)$	$\hat{\alpha_{ij}}(t)$
-0.2	-0.12	-0.08	0.2	0.24	-0.12	0.5	0.63	3.28	0.95 0.05	0.96 0.04	0.94 0.06
0.5	0.46	0.05	-0.5	-0.65	-0.06	1	1.06	NA	0.10 0.90	0.11 0.89	0.77 0.23

which is consistent with the actual transition dynamics. For clustering accuracy, the proposed method gives a 97.85% successful clustering rate for model identity and 99.86% for noise membership, while traditional method gives only 67.1% clustering accuracy for model identity.

In summary, based on the numerical simulation results for the proposed method, the clustering performances for both model identity and noise membership have been satisfactory. Moreover, compared with a traditional SMARX identification method, the proposed algorithm gives more reliable parameter estimation results for model/noise parameters and switching dynamics, especially when data are of poor quality and contain outliers.

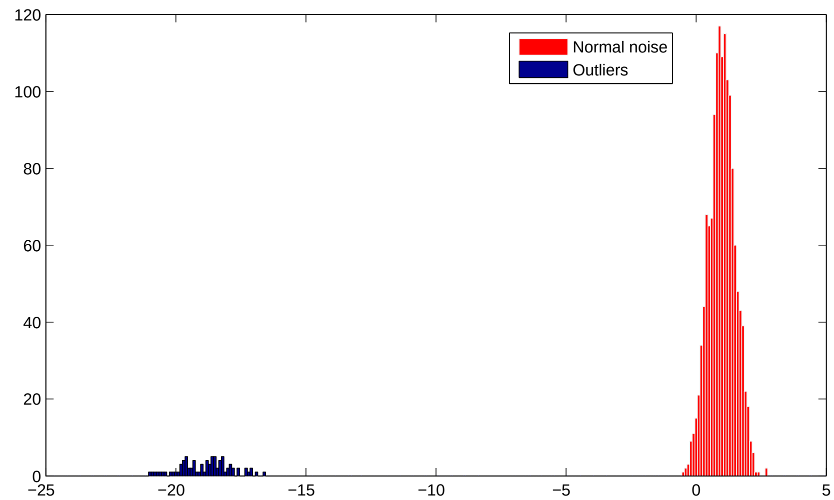


Figure 3.7: Noise distribution with outliers

CSTR simulation example

In this part, identification of a nonlinear CSTR process is considered at several representative operating points. Gaussian mixture noise is added to the output data. The proposed SMARX-GMM algorithm is used to identify the approximate local linear models.

In the CSTR, an exothermic irreversible reaction $A \rightarrow B$ takes place in a constant volume reactor with single coolant taking effect. The first principles model [1] can be expressed as:

$$\begin{aligned} \frac{dC_A(t)}{dt} &= \frac{q(t)}{V}\left(C_{A0}(t) - C_A(t)\right) - k_0 C_A(t) \exp\left(\frac{-E}{RT(t)}\right) \\ \frac{dT(t)}{dt} &= \frac{q(t)}{V}\left(T_0(t) - T(t)\right) - \frac{(-\Delta H) k_0 C_A(t)}{\rho C_p} \exp\left(\frac{-E}{RT(t)}\right) \\ &\quad + \frac{\rho_c C_{pc}}{\rho C_p V} q_c(t) \left\{1 - \exp\left(\frac{-hA}{q_c(t)\rho C_p}\right)\right\} \left(T_{c0}(t) - T(t)\right) \end{aligned} \tag{3.43}$$

The variables and parameters used above can be found in Table 3.3 [1].

The output is product concentration C_A, the input is the coolant flow rate q_c, and T is the CSTR temperature, an intermediate state. q_c also represents operating point, as discussed in [1], and different operating points show different step responses and thus indicate different process dynamics. In this simulation example, three different

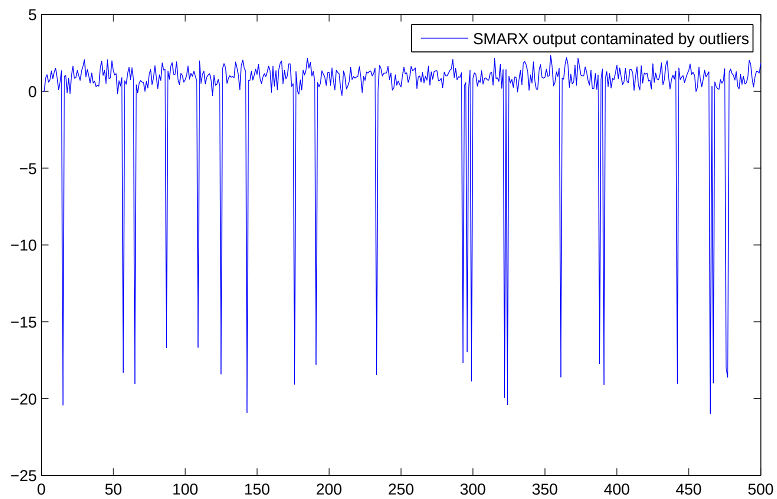


Figure 3.8: SMARX output with outliers

operating points are chosen, namely $q_c = 97, 103$ and $109L/min$. To apply the proposed algorithm, an input random binary sequence (RBS) is designed based on local model dynamics from a step response test, and potential model switching is designed to happen every 300 sampling units. Gaussian mixture noise $e_k \sim 0.7N(1.5\times 10^{-3}, (1 \times 10^{-3})^2) + 0.3N(-3.5 \times 10^{-3}, (1 \times 10^{-3})^2))$ is added to output data. Part of the input-output data is shown in Figure 3.9. The cross validation results are shown in Figure 3.10. It can be seen from the cross validation result, that the identified model, using the proposed algorithm, captures local model dynamics and model switching dynamics well.

Pilot scale experimental verification

In this part, a pilot scale experiment is conducted to illustrate the performance of the proposed algorithm. The schematic of the process is shown in Figure 3.11 [3], where three tanks are connected in series and bottom reservoir is connected to an adjustable DC pump. Notice that the cross section of the three tanks are different: the top tank has constant cross section; the middle one is conical; and bottom tank is spherical. The variability in cross section area leads to non-linearity of the system.

Table 3.3: CSTR model parameters, [1]

Process flow rate (q)	100	L/min
Feed concentration (C_{A0})	1	mol/L
Feed temperature (T_0)	350	K
Inlet coolant temperature (T_{c0})	350	K
CSTR volume (V)	100	L
Heat transfer term (hA)	7×10^5	L
Reaction rate constant (K_0)	7.2×10^{10}	min^-1
Activation energy term (E/R)	1×10^4	K
Heat of reaction ($-\delta H$)	-2×10^5	cal/mol
Liquid density ($\rho, \rho c$)	1×10^3	g/L
Specific heats (C_p, C_{pc})	1	cal/gK

During the experiment, certain operating points are chosen so that linear models well approximate the actual local non-linear system. Three tanks are connected via manipulable valves and by changing valve opening, different operating modes can be achieved.

The first principles model of this three-tank non-linear system is [3]:

$$\begin{aligned} \frac{dH_1}{dt} &= \frac{1}{\beta_1\left(H_1\right)}q - \frac{1}{\beta_1\left(H_1\right)}C_1 H_1^{\alpha 1} \\ \frac{dH_2}{dt} &= \frac{1}{\beta_2\left(H_2\right)}C_1 H_1^{\alpha 1} - \frac{1}{\beta_2\left(H_2\right)}C_2 H_2^{\alpha 2} \\ \frac{dH_3}{dt} &= \frac{1}{\beta_3\left(H_3\right)}C_2 H_2^{\alpha 2} - \frac{1}{\beta_3\left(H_3\right)}C_3 H_3^{\alpha 3} \end{aligned} \tag{3.44}$$

where C_1,C_2,C_3 are resistance coefficient of output orifice for the three tanks, respectively; H_1,H_2,H_3 are fluid levels in each tank; q is the inflow from the DC pump into top tank; β_1,β_2,β_3 are cross section areas of tanks, which are function of water levels in the lower two tanks; and α_1,α_2,α_3 are flow coefficients for the different tanks. By assuming that the inlet flow for each tank is small enough, the outlet flow from each tank can be considered to be laminar, and α_1,α_2,α_3 takes the value of 0.5.

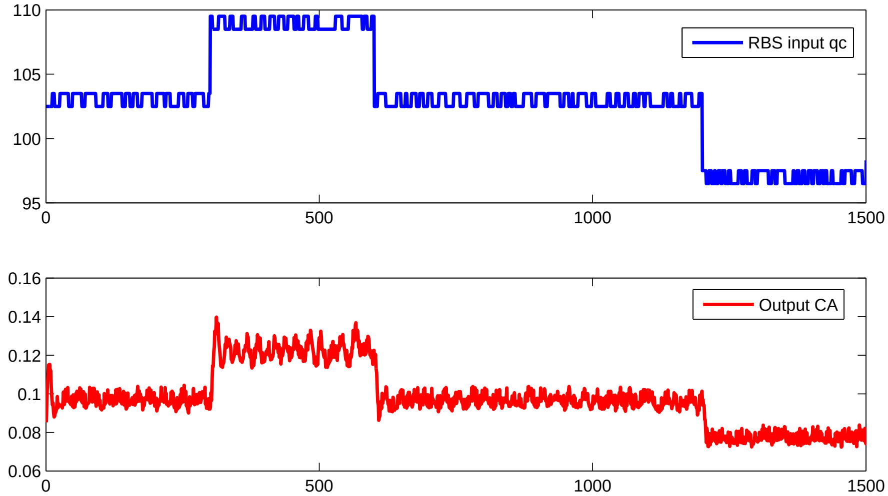


Figure 3.9: CSTR input-output data

In this experiment, q is taken as input and H_2 is taken as output, by changing opening of C_1 and C_2, two different operating modes can be achieved. A first-order Markov chain is designed to govern the switching dynamics of C_1 and C_2, with transition probability matrix $\begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}$.

A random binary sequence is generated as the input signal to excite the system based on step test results given in Figure 3.12. All positive responses are in the operating mode when C_1 is more open (80%), and C_2 is less open (80%), and negative responses are in the operating mode when C_1 is less open (60%), and C_2 is more open (100%).

From Figure 3.12, the steady state water levels for these two sub-models are different and calculations based on step test results show that the time constant for the faster system is approximately 40 seconds and for the slower one is approximately 120 seconds. The input-output data are collected as shown in Figure 3.13.

First, direct identification using experimental data is conducted by the proposed algorithm, and through various trials, it was found that a single Gaussian noise model can describe the noise structure quite well. The cross validation results of the identi-

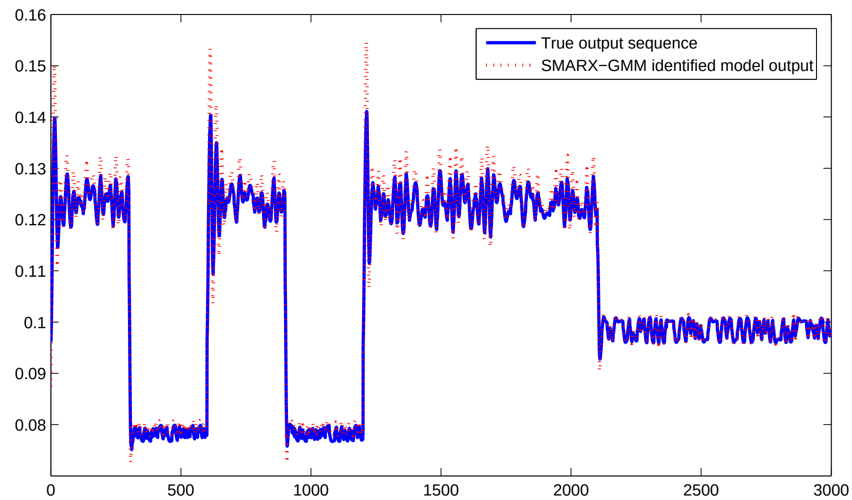


Figure 3.10: CSTR identification using SMARX-GMM algorithm, cross fitting result

fied model are shown in Figure 3.14 and Figure 3.15. It was determined that 83.12% of data are correctly clustered. Although the model identity clustering accuracy is not as good as in the simulation examples, the cross validation result looks good in Figure 3.14. Also, the estimation of transition matrix is $\begin{bmatrix} 0.75 & 0.25 \\ 0.08 & 0.92 \end{bmatrix}$, and this shows the ability of proposed algorithm in identifying the switching dynamics among sub-models.

To further test the proposed algorithm when data are contaminated by outliers, Gaussian mixture noise is added to the output measurement as:

$$e_k \sim 0.7N(-0.6, (0.2)^2) + 0.3N(1.4, (0.5)^2) \tag{3.45}$$

Using the proposed algorithm, the identification results are shown in Table 3.4, in comparison with results using a traditional single noise distribution method. The cross validation and clustering results are shown in Figure 3.16 and Figure 3.17, respectively. It can be seen that with outliers in the data, the clustering performance is significantly influenced. This is because the sub-models' dynamics are similar to each other, and the outliers make the model identities even harder to cluster; however, the proposed method still gives good cross validation results and reliable parameter

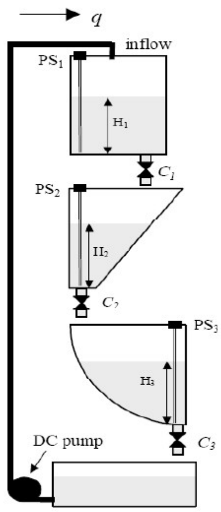


Figure 3.11: Multi tank system schematic, [3]

estimation such as transition probability matrix and noise model parameters. The difference in noise variance estimation is the added noise superimposed by the measurement noise in the raw data. On the other hand, the traditional identification algorithm suffers and gives unreliable parameters estimation. For example, the estimated transition matrix using the traditional method is a matrix dominated by second column, indicating that the system always tends to operate in state 2, while the true transition matrix is diagonally dominant, indicating that the system tends to stay in one state rather than to jump elsewhere.

Table 3.4: SMARX-GMM identification results for three-tank system

α_{ij}	$\hat{\alpha}_{ij}(p)$	$\hat{\alpha}_{ij}(t)$	p	$\hat{p}$	μ	$\hat{\mu}$	σ	$\hat{\sigma}$
0.8 0.2	0.67 0.33	0.17 0.83	0.7	0.69	-0.6	-0.62	0.2	0.59
0.1 0.9	0.09 0.91	0.29 0.71	0.3	0.31	1.4	1.40	0.5	0.76

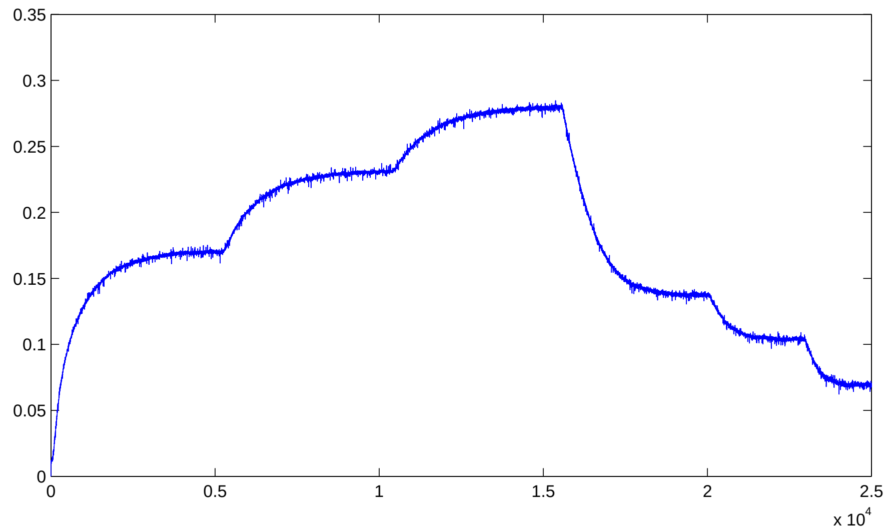


Figure 3.12: Step response of middle tank level in two operating modes

3.3 Electricity price prediction model

In this section, the method proposed in section 3.2 will be applied to the prediction of electricity price in Alberta. Based on the characteristics summarized in section 3.1, first the input variables selection and feature extraction are conducted and then, the model developed using proposed method is validated in subsection 3.3.3.

3.3.1 Input variables selection and data pre-processing

From section 3.1, the following facts are known: 1) the electricity price in Alberta has a strong periodic pattern; 2) the forecast Pool price will reflect future electricity price; 3) the actual demand correlates with the Pool price; and 4) the day ahead forecast demand will influence the bidding prices from generation units which is related to the Pool price.

Therefore, we choose the forecast Pool price, day ahead forecast system demand, actual system demand and time itself as input variables to predict the real time Pool price. In section 3.2, an approach to model multiple switched Markov ARX model is proposed. To use this method for Pool price prediction, there are two key issues to

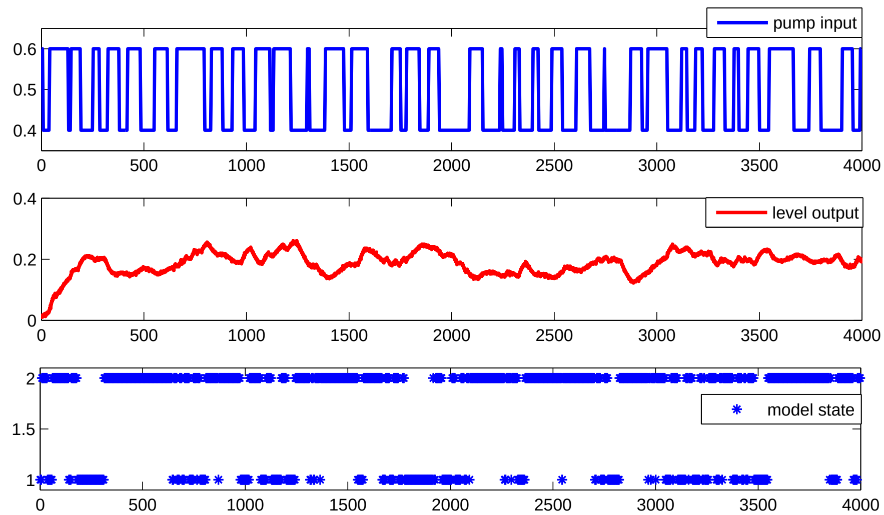


Figure 3.13: Top figure shows the pump input RBS signal; middle figure is level output; bottom figure is model identity at each time instant

consider:

1. From Figure 3.1 and 3.2, the relation between Pool price and system demand, time sequence are obviously non-linear. Thus, to use switched linear ARX models to predict Pool price, a preprocessing procedure is necessary to transform the input variables so that the transformed variables have linear correlation with the Pool price.

2. Since multiple models will be used in the prediction of Pool price, the data for system identification need to be segmented for initialization of the proposed algorithm as discussed in subsection 3.2.1. Therefore, the features of different Pool price levels need to be extracted to segment the input-output data.

The issue in item 1. will be discussed in this subsection, and item 2. in next subsection.

First the time sequence is pre-processed to construct a linear correlation with the Pool price. Due to the periodic pattern in the Pool price, we need first to transform the time sequence to be periodic by making it to time o'clock form, and different

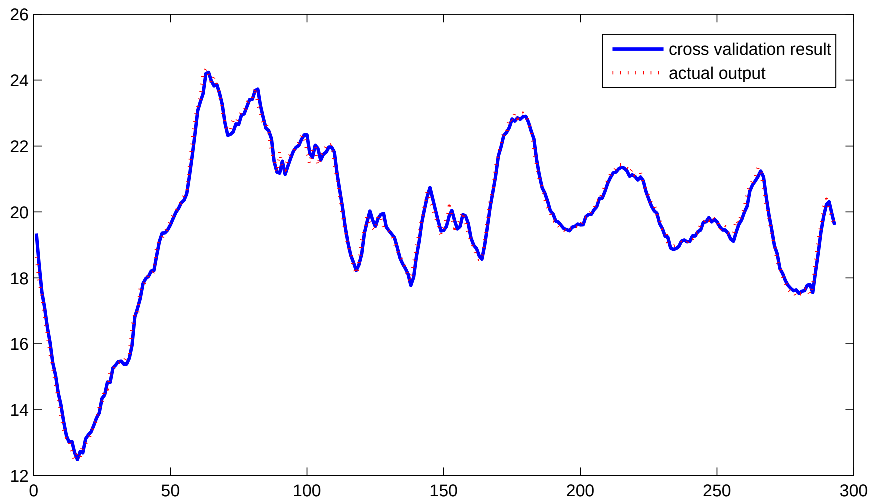


Figure 3.14: Cross fitting result using SMARX-GMM algorithm with single noise distribution

weights need to be put on peak and off-peak hours. The designed weighting formula are:

$$
\begin{aligned}
&F(t) = k(t,N)\, e^{-\frac{(t-t_p)^2}{2\sigma^2}} \\
&tp \sim PMF \\
&k(t,N) = f\left(P\left(S(t) = N|Obs, Trans, Emis\right)\right)
\end{aligned}
\tag{3.46}
$$

where: $F(t)$ is the preprocessed time sequence, weighted by a Gaussian function $e^{-\frac{(t-t_p)^2}{2\sigma^2}}$; t_p is the highest-price hour in each day, and is a random variable with probability mass function PMF (PMF can be learnt from price data); σ is a tuning parameter, determining the duration of the peak-price hours; and $k(t,N)$ is the peak-price magnitude, determined by the probability $P(S(t) = N|Obs, Trans, Emis)$ that at t o'clock, how possible the price is governed by peak or off-peak model given electricity price observation Obs and estimated HMM parameters: transition matrix $Trans$ and emission matrix $Emis$.

For the preprocessing of the demand sequence, it is found that in most cases, only when the system demand is over 9000MW, a high level Pool price will be triggered. Thus, the demand sequence should be preprocessed in a way that the parts over 9000

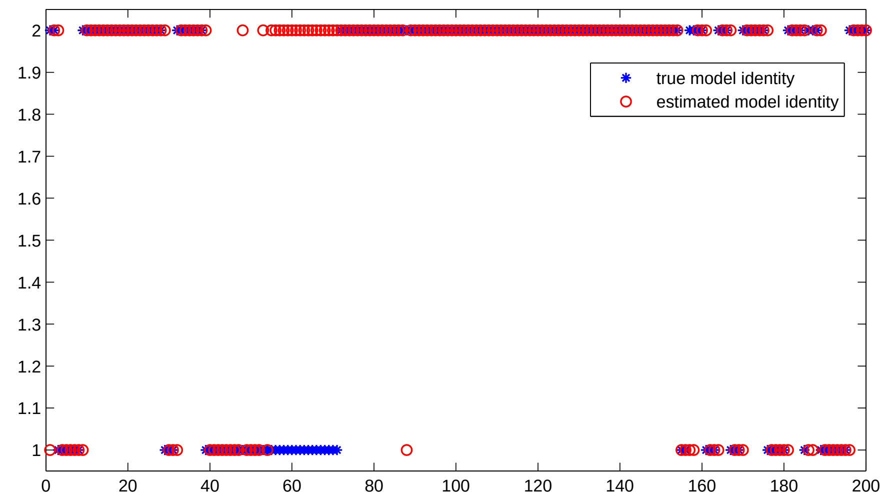


Figure 3.15: Model identity clustering result

are emphasized while the low demand intervals are flattened. After preprocessing, the correlation between Pool price and time sequence, system demand are shown in Figure 3.18.

3.3.2 Feature extraction

In last subsection, the problem of preprocessing the input sequence is discussed to form linear correlations between Pool price and input variables. In this subsection, the feature extraction of Pool price sequence is discussed to segment data for the initialization of EM algorithm as discussed in subsection 3.2.1.

Based on the Pool price sequence, first the data are divided into three types: peak-up, peak-down and off-peak (this is classified simply based on the absolute value and tendency of the sequence). Then Hidden Markov model is trained based on the symbol sequence resulted from this division to cluster the input-output data for the initialization of EM algorithm as introduced in 3.2.1.

To be specific, the data segmentation are conducted with the following rules:

1. First divide the Pool prices into five groups based on their absolute value;

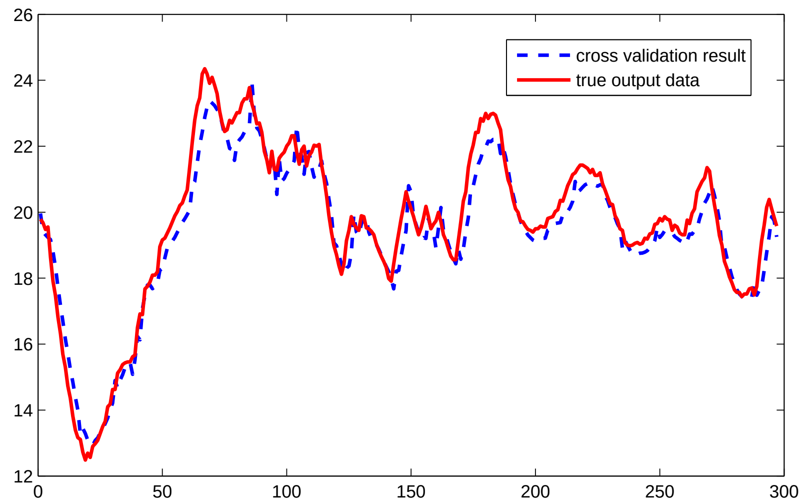


Figure 3.16: Cross fitting result of SMARX-GMM identification

group indices from low to high represent: group #1: less than \$30/MWh, group #2: from \$30/MWh to \$100/MWh, group #3: from \$100/MWh to \$300/MWh, group #4: from \$300/MWh to \$500/MWh, and group #5: more than \$500/MWh.

2. Calculate the group index differences of each two neighbouring Pool prices. If the difference is greater than 2, then the first Pool price is classified to be the peak-up data; if it's less than -2, the first Pool price is classified to be the peak-down data; else if the group index itself is greater or equal to 4, it will be classified to be either peak-up or peak-down data with probability; else it will be classified to be off-peak data.

3. Even in the peak hours, low prices and the off-peak characteristics may occur; on the other hand, in the off-peak hours, sometimes extremely high prices can show up. Therefore hidden Markov model is trained given the classified Pool price sequence in item 2. to estimate the transition probability and emission probability among peak-up, peak-down and off-peak prices.

Figure 3.19 shows the data segmentation and feature extraction result from hidden

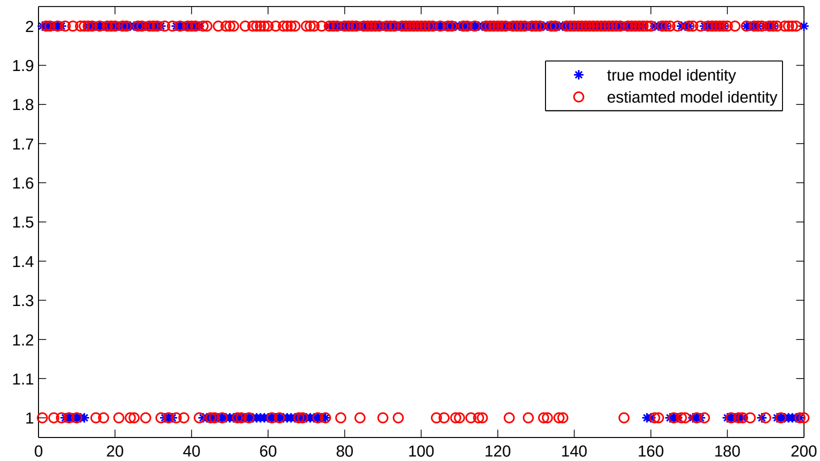


Figure 3.17: Model identity estimation result in Gaussian mixture noise case

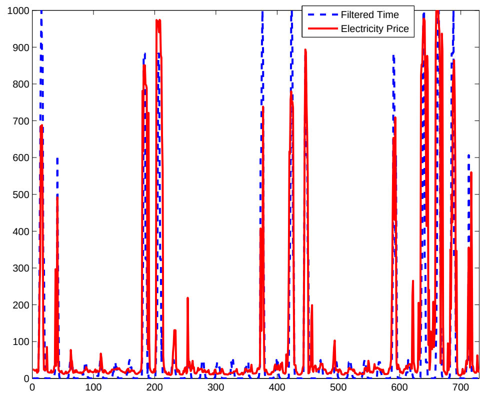
Markov model training. State 1 means peak-up Pool price, state -1 means peak-down Pool price, state 0 means off-peak price:

3.3.3 Model identification and validation

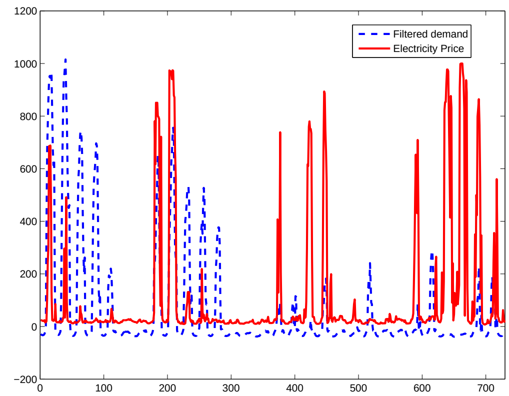
After input variables selection, feature extraction and data segmentation, the multiple Pool price SMARX model is developed using the proposed method in section 3.2. First we initialize the EM algorithm with the probabilities that each Pool price belongs to one of three sub-models from HMM training, and run the EM algorithm to estimate the model parameters for the three local Pool price models: peak-up, peak-down and off-peak. After obtaining the model parameters for each local Pool price model, the prediction formula for the expected value of next hour's Pool price are expressed as:

$$\begin{aligned} & E\left(P_e\left(i+1\right)\right) = \sum_{k=1}^{3}\sum_{j=1}^{3} P\left(S\left(i\right)=j|Obs, Trans, \Theta\right) Trans\left(jk\right) P_{ek}\left(i+1\right) \\ & Trans\left(jk\right) = P\left(S\left(i+1\right)=k|S\left(i\right)=j\right) \end{aligned} \tag{3.47}$$

where: $P\left(S\left(i\right)=j|Obs, Trans, \Theta\right)$ is the model identity probability denoting that at i o'clock, how possible Pool price is governed by the jth sub-model given the input-output data Obs, model parameters Θ and transition probability $Trans$; and $P_{ek}(i+1)$



(a) Pool price and preprocessed time sequence



(b) Pool price and preprocessed system demand

Figure 3.18: Correlation between Pool price and preprocessed inputs

is the predicted Pool price by the kth sub-model. The left hand side of equation (3.47) is the expected Pool price for the coming hour, and it is calculated as the summation of the product of each sub-model prediction times its identity probability in next hour. The identity probability in next hour is calculated as the product of current identity probability times transition probability to a potential sub-model in next hour.

The above prediction of Pool price relies mainly on the transition of hidden Markov model, and thus might be insufficient for long term prediction as first-order Markov transition cannot capture long term dynamics; however, for the prediction of hour ahead Pool price, the proposed method has very good prediction accuracy compared to the forecast Pool price released by AESO website, comparison results are shown in Figures 3.20 and 3.21.

Figure 3.20 is the fitting result based on the training data using proposed method, compared with the actual Pool price and forecast released by AESO. Figure 3.21 is the cross fitting result using Pool price data in the month after fitting data set. From the figures, it can be seen that both the AESO forecast and proposed prediction are quite accurate when electricity price is low (actually the AESO forecast data are slightly more accurate in off-peak hours from the figures); however, during peak hours, the AESO forecasts are sometimes several hundred dollars away from the actual Pool price, while the proposed method predicts accurately with much closer prediction

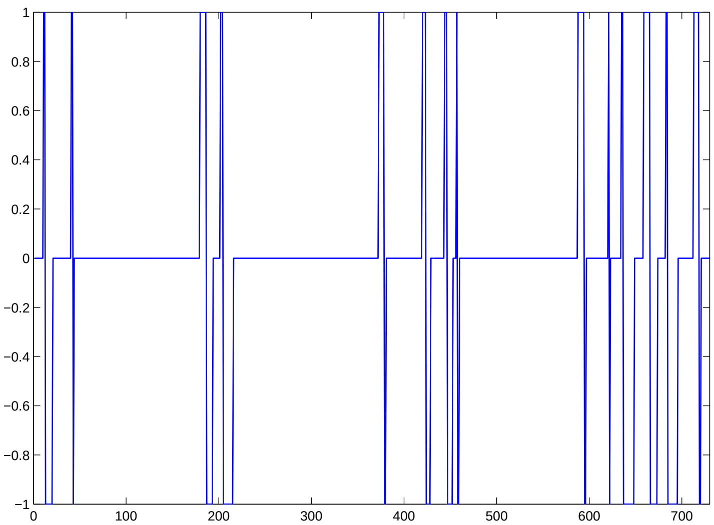


Figure 3.19: HMM feature extraction results, 1: peak-up price, -1: peak-down price, 0: off-peak price

results. The quantitative comparison for the proposed method and AESO forecast can be shown in Table 3.5: where MAE stands for mean absolute error, RMSE stands

Table 3.5: Cross validation performance comparison

Results	MAE	RMSE	Correlation	Fitting rate
AESO	28.83	83.12	92.7%	58.25%
Proposed	26.48	57.66	95.7%	71.04%

for root-mean-square error, correlation is the correlation between the predicted prices and the actual prices, fitting rate is the cross fitting percentage. From performance comparison in the table, it can be seen that in terms of the quantitative indices, the proposed method has significantly better performance especially with respect to RMSE and cross fitting rate.

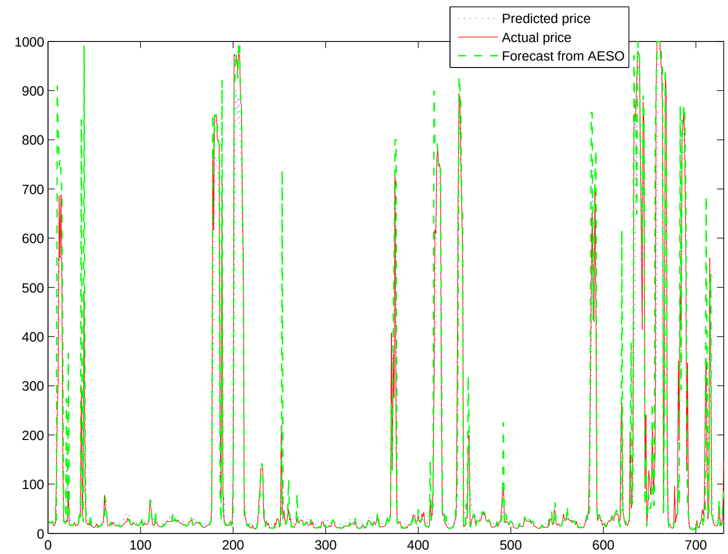


Figure 3.20: Electricity price fitting result with training data. Red: actual price; Blue: predicted price; Green: forecast from AESO

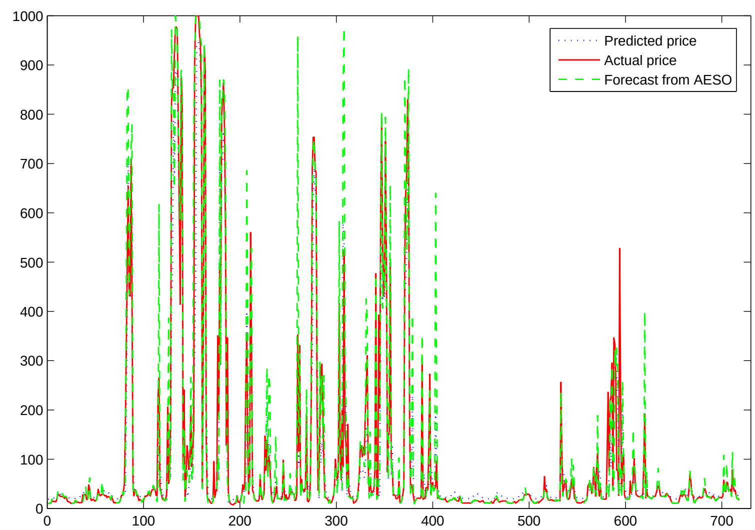


Figure 3.21: Electricity price fitting result with cross validation data. Red: actual price; Blue: predicted price; Green: forecast from AESO

Chapter 4

Predictive steam plant optimization

In this chapter, a linear optimization problem is formulated that includes the predicted Pool price and steam boiler dynamics. This method maximizes a profit function over time. Two solution approaches are proposed to solve the optimization problem: 1) the first approach is an open-loop strategy that integrates the dynamic boiler model directly into the problem and solves for boiler manipulation variables via a dynamic linear programming (DLP) problem; and 2) the second approach is a closed-loop strategy that solves the optimization to obtain the optimal boiler load trajectories and designs controllers to track them. In the final section of this chapter, a performance assessment method is proposed to evaluate the effect of the steam plant optimization under uncertainty.

4.1 Problem formulation

In this section, a linear optimization problem is formulated that includes the predicted Pool price and boiler dynamics. Objective function, constraints and the random process variables will be discussed.

4.1.1 Economic objective function

The main objective for steam plant optimization is maximizing the operating profit, given the electricity price and process steam demand. The steam plant profit arises from two sources: 1) steam users pay for the amount of steam consumed at retail price set by company's financial department; and 2) the plant produces electricity through turbine generators to support the processes and sell to the electricity grid at real time Pool price. The operating cost of the steam plant comes from the energy and materials consumed to generate the steam and equipment operation costs. To be specific, the optimization objective function for the steam plant in Figure 2.1 is expressed as:

$$C_{900}\left(X_{boi1}+X_{boi2}\right)+C_{160}X_{pkg}-\left(V_{900u}P_{900}+V_{160u}P_{160}+V_{35u}P_{35}+X_{e1}P_e+X_{e2}P_e\right) \tag{4.1}$$

where: C_{900}, C_{160} represents the cost to generate the 900 psi and 160 psi steam as provided by the financial department, in $\$/lb$; P_{900}, P_{160}, P_{35} represents the retail prices for each type of steam charged to the users as provided by the financial department, in $\$/lb$; V_{900u}, V_{160u}, V_{35u} represents the user demands in different headers, from process measurements and mass balance calculations, in lb; P_e represents the real time Pool price as provided by the AESO website and prediction model developed in Chapter 3, in $\$/kw$; X_{boi1}, X_{boi2} are boiler loads, representing the amount of steam generated by the drum boilers, in lb; X_{pkg} represents the amount of steam generated by the back-up package boiler, in lb; X_{e1} and X_{e2} are electricity generation amounts, representing the turbine generator's electricity output, in kw.

Furthermore, the electricity generation X_{e1} and X_{e2} can be expressed in terms of steam distribution operations, including the amount of 35 psi steam extracted, $X_{tg1ex35}$, $X_{tg2ex35}$, in lb and the amount of 900 psi steam passing through the turbine generators, X_{tg1} and X_{tg2}. As derived in Chapter 2, the following grey box turbine generator models are used to relate the electricity output and plant operation variables [2]:

$$
\begin{aligned}
X_{e1} &= -362.7790 + 0.0400X_{tg1ex35} + 0.00905X_{cond1} \\
X_{e2} &= -603.0405 + 0.0435X_{tg1ex35} + 0.1074X_{cond2}
\end{aligned}
\tag{4.2}
$$

where X_{cond} is the water condensed in the turbine generators. Note that in mass balance equation, the input steam to turbine generators should match the output (i.e., the condensed water plus extracted 35 psi steam). Thus, the following relationship holds:

$$
\begin{aligned}
X_{cond1} &= X_{tg1} - X_{tg1ex35} \\
X_{cond2} &= X_{tg2} - X_{tg2ex35}
\end{aligned}
\tag{4.3}
$$

In the economic objective function (4.1): all of the steam prices are treated as constants; V_{900u}, V_{160u} and V_{35u} come from process measurements; and P_e is determined by the electricity market. To include boiler dynamics, plant operation is optimized based not only on current profit, but also on potential profit in the future, which requires Pool price prediction as part of future profit function. The Pool price prediction is provided by the prediction model introduced in Chapter 3.

4.1.2 Constraints with random process variables

The operating constraints for the dynamic steam plant optimization include:

1. a mass balance constraint for each steam common header, which requires that the input steam flow-rate should be no less than the output steam flow-rate, with the possibility of leakage along steam pipelines:

$$
\begin{aligned}
X_{boi1}+ \quad & X_{boi2} + V_{H_2SO_4} \geq \\
& X_{9t4} + X_{9t1} + X_{prv9t3} + X_{tg1} + X_{tg2} + V_{tc} + V_{tt} + V_{900u} + \Delta_{900}
\end{aligned}
\tag{4.4}
$$

   where: $V_{H_2SO_4}$ represents the #900 steam supply from the sulfuric acid production unit; X_{9t4} represents the amount of steam transformed from #900 steam to #450 steam via pressure let-down station (PLS); X_{9t1} represents the amount of steam transformed from #900 steam to #160 steam via the PLS; X_{prv9t3} represents the amount of steam transformed from #900 steam to #35 steam via the PLS; V_{tc} represents the steam used to run the turbine compressor; V_{tt} represents the steam used in the Terry turbine; and Δ_{900} represents the unaccounted for steam in #900 steam common header, which is a random variable known only in its distribution during optimization.

$$\begin{aligned} v_{9t4}EX_{9t4} + V_{WasteHeat1} + V_{WasteHeat2} \geq \\ X_{4t1} + X_{4t3} + V_{4t1} + Shift + \Delta_{450} \end{aligned} \tag{4.5}$$

where: v_{9t4E} is the efficiency of the PLS to transform #900 steam to #450 steam, usually assumed to be 1 during optimization process; $V_{WasteHeat1}$ and $V_{WasteHeat2}$ are the amounts of steam supply from waste heat boilers; X_{4t1} and X_{4t3} are transformed steam via PLS; V_{4t1} is the amount of #450 steam used by the process turbines; $Shift$ is the major user of #450 steam; and Δ_{450} is the amount of unaccounted for steam in the #450 steam common header, and its density function has been modelled in Chapter 2.

$$\begin{aligned} v_{4t1}EV_{4t1} + v_{9t1}EX_{9t1} + R_{pkg}X_{pkg} + X_{4t1} \geq \\ V_{160u} + X_{1t3} + X_{pkgt3} + \Delta_{160} \end{aligned} \tag{4.6}$$

where: $v_{4t1}E$, $v_{9t1}E$ are the transformation efficiencies of PLSs, from #900 and #450 steam to #160 steam; R_{pkg} is the percentage of steam generated by the package boiler flowing into the #160 steam common header; X_{1t3} is the transformed steam via PLS from #160 steam to #35 steam; Δ_{160} represents the unaccounted for steam in the #160 steam common header, and its density function is modelled in Chapter 2.

$$\begin{aligned} X_{1t3} + (1 - R_{pkg})\, X_{pkg} + X_{4t3} + X_{prv9t3} + X_{tg1ex35} \\ + X_{tg2ex35} + E_{tt}V_{tt} + E_{tc}V_{tc} + X_{pkgt3} \geq V_{35u} + \Delta_{35} \end{aligned} \tag{4.7}$$

where: E_{tt} and E_{tc} are the transformation efficiencies of Terry turbine and turbo compressor; and Δ_{35} represents the unaccounted for steam in the #35 steam common header, and its density function is modelled in Chapter 2.

2. operational restrictions: this type of constraints are the physical limitations of industrial equipment, such as boiler capacity, turbine generator load limits and

so forth. These restrictions are collected from the equipment data sheets and process knowledge. Unlike the mass balance constraints, which include random elements, the operational constraints are deterministic. As they are largely based on the physical properties of process equipment, which are considered to be stable over time. The operational constraints are listed as follows:

$$\begin{aligned}
&X_{tg1} \leq 51000 \\
&X_{tg2} \leq 51000 \\
&0 \leq X_{tg1ex35} \leq 30000 \\
&0 \leq X_{tg2ex35} \leq 30000 \\
&10000 \leq X_{tg1} - X_{tg1ex35} \leq 19000 \\
&10000 \leq X_{tg2} - X_{tg1ex35} \leq 19000
\end{aligned} \tag{4.8}$$

In the above constraints and objective function, all the variables with capital X are the decision variables for optimization purposes. Variables with capital V and Δ are time-varying process information, which are known from process measurements or estimations. Other variables with capital E, P or C are parameters, and are derived from prior knowledge or prediction models.

4.1.3 Optimization with steam boiler dynamics

From subsection 4.1.1 and 4.1.2, the random variables in the right hand side (RHS) of constraints can be determined at any instant of time from historical data; however, to include their future values for predictive optimization purpose, we use the estimated values based on their distribution. Pool price and steam price are available from the prediction model or from statistics provided by the financial department. In real time optimization, since all estimations of the uncertain coefficients in the linear optimization problem are available, the resulting problem is a deterministic linear program, and efficient solvers are available to solve such problems.

In practice, for implementation of the optimization results, some of the decision variables are manipulated instantaneously, such as the pressure let-down station opening and the turbine generators' load change. On the other hand, for equipment like drum boilers, the system response time is relatively long due to the slow heating process. These differences in dynamics must be considered when implementing the optimization results. Thus, an optimization approach incorporating boiler dynamics is necessary to ensure that the large range of dynamics are properly reflected in the determination of optimal operations.

From Chapter 2, the boiler steam generation is dependent on the amount of natural gas flow (NG), boiler feed water (BFW) and the disturbance from common header pressure (P). According to equation (2.12), the dynamic boiler model can be represented in a general form as follows:

$$y = \frac{A\left(z^{-1}\right)}{B\left(z^{-1}\right)}u_1 + \frac{C\left(z^{-1}\right)}{D\left(z^{-1}\right)}u_2 + \frac{E\left(z^{-1}\right)}{F\left(z^{-1}\right)}u_d \tag{4.9}$$

where: u_1 is NG flow, and is a manipulated variable; u_2 is BFW flow, and is another manipulated variable; and u_d is P, and is a disturbance input. From a step test result between each input and steam output in subsection 2.3.3, the time constants range from several to tens of minutes. This means that at the beginning of each hour when the Pool price is released, the drum boiler cannot achieve the optimal loads calculated by the optimization algorithm for a considerable amount of time. Thus, a predictive optimization method is needed to adjust the boiler operation and steam distribution in advance, using the predicted Pool price and boiler dynamics. In this way, NG and BFW flow can be manipulated so that the long time dynamics of steam boilers are reasonably negated. Once the actual Pool price is released (and is not too different from the prediction), the optimal load can be achieved quickly. A moving horizon framework might be useful to compensate for the Pool price prediction error and influence of process disturbances. The specific dynamic optimization solution strategies incorporating boiler dynamics and predicted Pool price will be explained in detail in next section.

4.2 Solution strategies

The solution strategies incorporating Pool price prediction and drum boiler dynamic model are introduced in this section. The first one includes the truncated dynamic boiler model directly in the optimization process, so that the boiler manipulated variables and steam distribution operations are calculated all at once. The other strategy separates optimization from boiler control by designing tracking controller for optimal boiler load trajectories.

4.2.1 Dynamic linear programming, an open-loop approach

The first solution strategy is to integrate the boiler dynamics directly into the linear optimization problem formulation, so that the manipulated variables NG and BFW can be calculated directly via optimization in an open-loop control fashion: specifically, replacing the boiler load decision variables X_{boi1}, X_{boi2} by NG and BFW through drum boiler dynamic models, and solving the resulting linear optimization problem with respect to NG and BFW to determine the optimal boiler manipulation and steam distribution. To illustrate this solution strategy, an example problem with small number of decision variables and simple dynamics is given below:

$$\begin{aligned}
&min \quad 0.5x_0 + 0.7x_2 \\
&s.t. \\
&\qquad 3x_0 + 8x_2 \leq 100 \\
&\qquad 5x_0 + 3x_2 \leq 75 \\
&\qquad x_0, x_2 \geq 0 \\
&\qquad x_0 = \tfrac{z^{-1}}{1-0.5z^{-1}} x_1
\end{aligned} \tag{4.10}$$

where: x_0 is the decision variable (think it as boiler load) related to a manipulated variable x_1 (think it as NG or BFW) through a dynamic model. Replacing x_0 in the original problem with the manipulated variable, x_1, by expressing the dynamic model in an impulse response form:

$$x_0 = z^{-1}x_1 + 0.5z^{-2}x_1 + 0.25z^{-3}x_1 + 0.125z^{-4}x_1 ... \tag{4.11}$$

In this case, as multiple-step decisions influence optimal solution at each time instant, we minimize the summation of profit function over some time horizon to best reflect operating performance instead of minimizing the objective function at only individual time instant. By truncating equation (4.11) at an appropriate length after which the influence of previous x_1 to x_0 can be omitted, we keep the first 4 terms in (4.11) to approximate x_0:

$$\begin{array}{ll} min & \sum_{t=s}^{s+n} [\sum_{i=1}^{i=4} 0.5^i x_1(t-i) + 0.7x_2(t)] \\ s.t. & \\ & 3\sum_{i=1}^{i=4} 0.5^{i-1} x_1(t-i) + 8x_2(t) \leq 100 \\ & 5\sum_{i=1}^{i=4} 0.5^{i-1} x_1(t-i) + 3x_2(t) \leq 75 \\ & \sum_{i=1}^{i=4} 0.5^{i-1} x_1(t-i), x_2(t) \geq 0 \\ & t = s, s+1, ..., s+n \end{array} \tag{4.12}$$

where n is the optimization horizon, analogous to the prediction horizon in Model Predictive Control (MPC). Once n is set, the resulting problem is a linear program with $2n-1$ decision variables (i.e., $x_1(s), x_2(s), \cdots, x_1(s+n-1), x_2(s+n)$) and $4n$ constraints. This problem can be solved efficiently by LP solvers. The resulting LP includes the boiler dynamics within the optimization problem. Once the LP problem is solved, only the first moves of manipulated variables $x_1^*(s)$ and $x_2^*(s)$ are implemented following an MPC philosophy. At next round optimization, we shift s to $s+1$, and use the newly updated information in the RHS of constraints (such as newly updated steam user demands) and in the objective function parameters (Pool price). The truncation length of equation (4.11) can also be adjusted to serve user requirements on the accuracy of dynamic model approximation.

Figure 4.1 shows the accuracy of this solution strategy to an example problem. The time constant of the dynamic system in this example is approximately several time instants, and thus the prediction horizon n is set to be 20 so that influence of early past inputs on current output can be reasonably neglected. Figure 4.1 shows the difference between the solutions based on the truncated dynamic model and the optimal solution (i.e., using a very long truncation length) within prediction horizon.

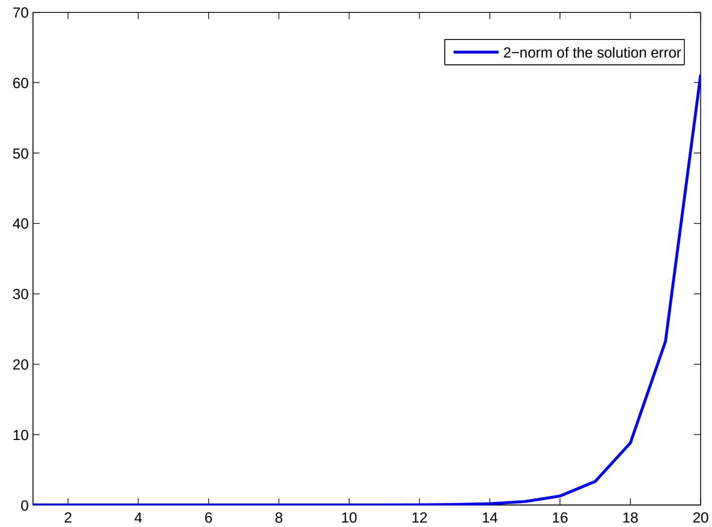


Figure 4.1: 2-norm of the error between the truncated DLP and the globally optimal solution within prediction horizon

It can be seen from Figure 4.1, when setting prediction horizon as 20, decision variables for the first 14 steps are almost the same as that of the globally optimal solution(i.e., 2-norm of the solution error is almost 0). In the last several steps, the truncated solution deviates from the globally optimal one since the truncated DLP does not have access to future information in RHS. As the DLP solution strategy follows MPC philosophy, only the decision variables for the first step is implemented. Thus, the deviation in the last steps will not influence the truncated solution's performance if appropriate prediction horizon is selected.

For the drum boiler dynamics, the time constant can be as long as half an hour. In this case, the prediction horizon must be set sufficiently large for good approximation accuracy. This may result in a relatively large scale dynamic LP problem to solve, and the computational burden might be an issue for real time optimization. On the other hand, the effect of DLP strategy depends on the quality of future predictions of constraints' RHS and objective function's parameters. In steam plant optimization, high quality predictions for steam user demand, unaccounted for steam flow and future

Pool price are required for this solution strategy. The advantage of this method is that the optimal manipulated variables are computed directly from the optimization, and this ensures the best possible performance if the knowledge is perfect about future process behaviour and electricity market.

4.2.2 MPC and LQI tracking controller

In previous subsection, the manipulated variables NG and BFW are formulated into the dynamic linear programming problem, and the algorithm calculates the optimal manipulated variables directly. In this subsection, the steam plant optimization and the control of boiler load are separate: 1) the optimization is solved first with appropriate prediction horizon, and the optimal boiler load trajectory as well as the steam distribution operation variables are calculated; 2) a controller is designed to track the optimal steam load trajectory by manipulating NG and BFW . For this approach, an optimal linear quadratic integral (LQI) control law and an MPC control law are designed to serve the tracking purpose.

For the LQI tracking controller, at each control interval, the optimal boiler load trajectory is updated based on the newly released process and market information. The NG and BFW are calculated by the LQI controller to track the boiler load trajectory with minimal error, see Figure 4.2. Controller gain K is calculated by MATLAB LQI control function *lqi*. First the dynamic boiler model is expressed as a discrete state space equation:

$$\begin{aligned} x\left(k+1\right) &= Ax\left(k\right) + Bu\left(k\right) \\ y\left(k\right) &= Cx\left(k\right) + Du\left(k\right) \end{aligned} \tag{4.13}$$

where: A, B, C and D are parameters representing the boiler dynamics; $x(k)$ are the states; $u(k)$ includes NG and BFW, while P is the disturbance variable; and $y(k)$ is the steam generation amount. The control law is $u = -Kz = -K[x; xi]$, where x_i is the augmented state representing the integrated tracking error. This control law minimizes the following objective function [51]:

$$J\left(u\right) = \sum_{k=0}^{\infty} \left[z^T\left(k\right) Qz\left(k\right) + u^T\left(k\right) Ru\left(k\right) + 2z^T\left(k\right) Nu\left(k\right)\right] \tag{4.14}$$

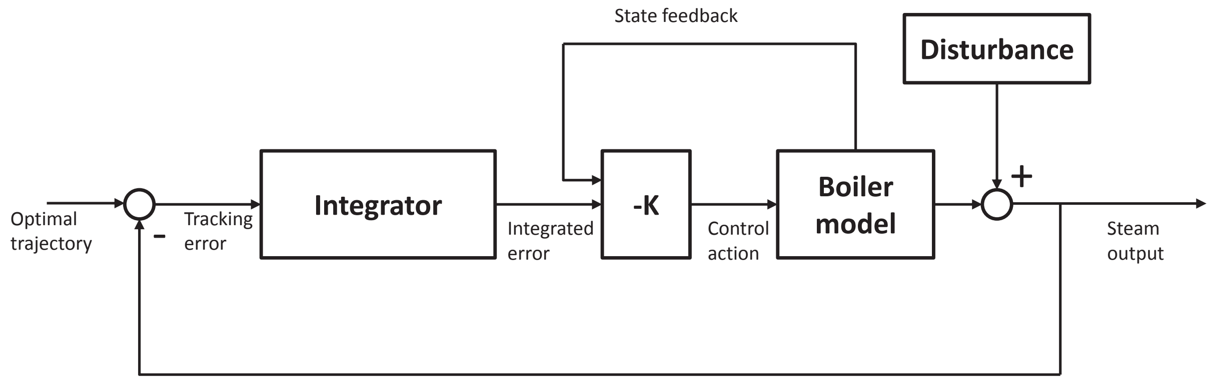


Figure 4.2: Optimal tracking controller framework

In equation (4.14), Q, R, and N are tuning parameters. Figure 4.3 shows the tracking error of designed LQI controller for a trajectory of the optimal boiler load. Notice that the boiler load's order of magnitude is $10^5 lb/hr$, and the standard deviation of the tracking error is about $1252.5 lb/hr$. The relatively small variance in tracking error illustrates the tracking performance using LQI controller.

The second approach uses a model predictive controller (MPC) for the optimal boiler load tracking. The prediction horizon is set long enough so that sufficient market information and process dynamics can be included. The tracking performance of MPC is shown in Figure 4.4. In this case, the standard deviation of the MPC tracking error is $822.9 lb/hr$, which is smaller than the LQI tracking controller.

The advantage of the tracking controller method is that it can track any calculated boiler load trajectory once the controller is designed, and it can negate the disturbance in common header's pressure. The disadvantage of this method compared to the open-loop DLP solution is that the control result in this method may deviate from the optimal trajectory because of the existence of tracking error.

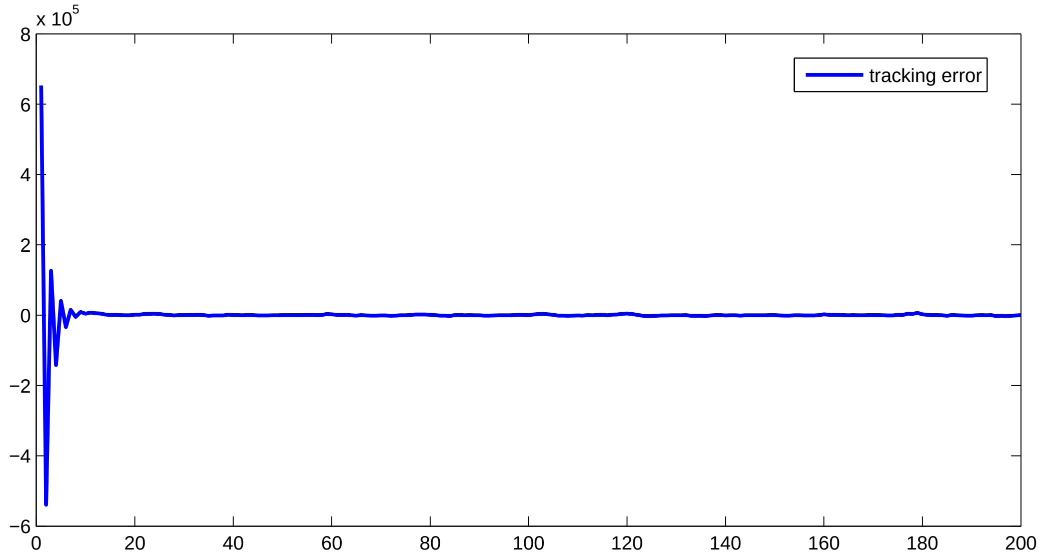


Figure 4.3: Tracking error of the LQI tracking controller

4.3 Performance assessment of the steam plant optimization

In this section, the performance assessment of the proposed optimization approach is formulated as the distribution problem of stochastic linear programming (SLP). The distribution problem of SLP is first introduced by Babbar [28], Tintner [29] and Wagner [30], and its general expression is [27]:

$$\begin{aligned} &\min \quad \mathrm{Z} = \mathrm{C}^T x = \sum_{j=1}^{n} c_j x_j \\ &s.t. \\ &\qquad A^T x = \sum_{j=1}^{n} a_{ij} x_j \geq b_i, i = 1, 2, ..., m \end{aligned} \tag{4.15}$$

where $x_j \geq 0$ and some or all of LP coefficients c_j, a_{ij} and b_j are random variables with known probability distributions [27]. The decision variables x_j^* and the optimal objective function Z^* depend on the realizations of A, b, and c through the solution of LP. Thus x_j^* and Z^* are random variables with distributions determined by the distributions of c_j, a_{ij} and b_j. The objective of distribution problem is to solve for

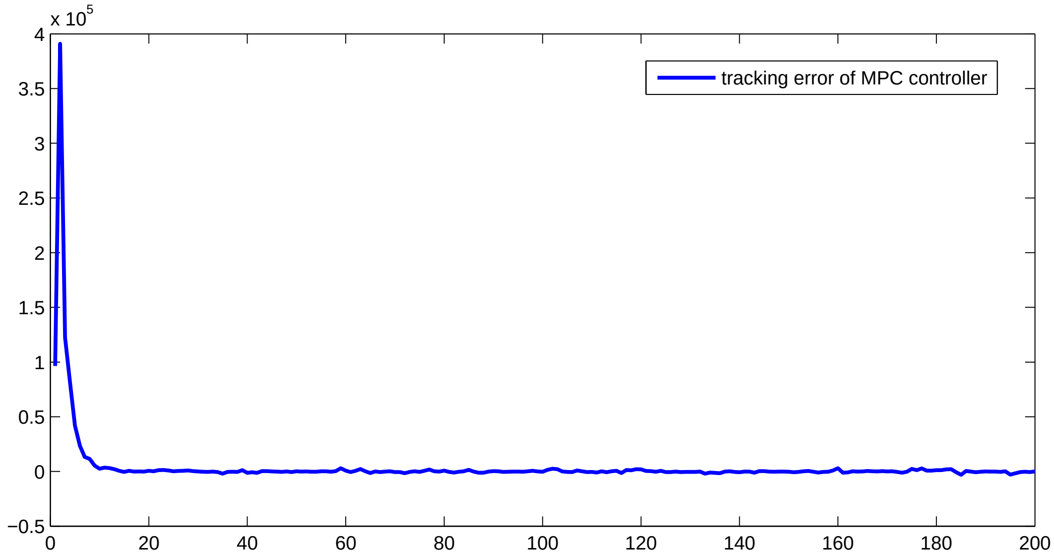


Figure 4.4: Tracking error of MPC controller

the distribution of Z^* and x^*, so that perfect knowledge about the objective function or decision variables of SLP is available. In this section, based on the formulation of steam plant optimization, the distribution problem is considered for the following type of SLP:

$$\begin{aligned} \min \quad & Z = C^T x \\ s.t. \quad & \\ & Ax \leq b\left(\xi\right) \\ & x \geq 0 \end{aligned} \tag{4.16}$$

where A and c are constants, representing process mass balances and Pool prices. The uncertain terms are the RHS of constraints, which are the random steam user demands and missing measurements as discussed in Chapter 2. The special case where $b(\xi)$ follows a Gaussian distribution is important as analytical results can be derived.

The significance of solving the distribution problem is: 1) it provides the perfect information about optimal plant operation under process uncertainty; 2) through knowledge of the distribution of steam plant profit function, it is straightforward to compare and classify different operating modes with proper distribution metric; and 3) knowledge of the distribution of the decision variables facilitates sensitivity analysis of the SLP, and thus it is possible to explore bottlenecks in the steam plant operation.

4.3.1 Distribution problem for SLP with Gaussian RHS

In this subsection, a basic theorem for the distribution problem of SLP with Gaussian RHS will be proposed and proven. Monte Carlo simulation is used to solve the distribution problem, and it requires the parameter estimation of Gaussian mixture distribution. This is done via Expectation-Maximization algorithm in this subsection. After that, a distribution metric based on earth mover's distance (EMD) with adjustable Hellinger distance is introduced to measure the difference between two probability distributions.

Gaussian mixture solution to distribution problem

Theorem 4.3.1. *In a stochastic linear programming problem with a RHS vector that follows a Gaussian distribution in the constraints (as expressed in (4.16)), the distribution for the decision variables and objective function follow the distributions of Gaussian mixture under assumptions that there are unique optimal solutions for realizations of $b(\xi)$, i.e.*

$$\begin{array}{lcl} b\left(\xi\right) & \sim & N\left(\mu,\Sigma\right) \\ x^{*} & \sim & \sum_{i=1}^{M} p_i N\left(\mu_i,\Sigma_i\right) \\ z^{*} & \sim & \sum_{j=1}^{L} p_j N\left(\mu_j,\Sigma_j\right) \end{array} \tag{4.17}$$

where the parameters for each local Gaussian component are functions of μ and Σ. Assuming $b \in R^m$, and $x \in R^n$, then the number of local Gaussian components is bounded above by:

$$\begin{array}{lcl} L & \leq & M \\ L & \leq & C^{n}_{m+n-\left\lfloor \frac{n+1}{2} \right\rfloor} + C^{n}_{m+n-\left\lfloor \frac{n+2}{2} \right\rfloor} \\ M & \leq & C^{n}_{m+n} \end{array} \tag{4.18}$$

Proof. According to the fundamental theorem of linear programming, the optimal value of z must be achieved at one of the vertices of the convex polyhedron defined by the constraints of linear programming:

$$P = \{x \in R^n | Ax \leq b, x \geq 0\} \tag{4.19}$$

In the case where $b(\xi) \sim N(\mu, \Sigma)$ is a random variable, for different realizations of $b(\xi)$, $x^{(i)}(b)$ is denoted as the vertices of polyhedron given $b(\xi)$, and all of the possible vertices for different realizations of $b(\xi)$ is denoted as a finite set

$$S = \{x^{(1)}(b), ..., x^{(r)}(b)\} \tag{4.20}$$

each with the probability $P(Ax^{(i)}(b) \leq b(\xi)), i = 1, 2, ..., r$ to exist. Denote $x^*(b)$ as the optimal decision variable given $b(\xi)$, and all the vertices $x^*(b)$ make a set S^*:

$$x^*(b) \in S^* \subseteq S \tag{4.21}$$

and the probability that vertex $x^*(b)$ is the optimal decision variable is:

$$P\{Ax^{(*)}(b) \leq b(\xi), C^T x^*(b) \leq C^T x^{(i)}(b), i = 1, 2, ..., r\} \tag{4.22}$$

As $b(\xi)$ varies, the geometrical shape of the polyhedron P will change. Following the assumption that for each realization of $b(\xi)$, the optimal solution to resulting LP is unique:

$$A_0 x^*(b) = b_0(\xi) \tag{4.23}$$

where A_0 is an invertible matrix corresponding to the active constraints, and $b_0(\xi)$ is the corresponding RHS of active constraints, also from a Gaussian distribution. Thus:

$$\begin{aligned} x^*(b) &= A_0^{-1} b_0(\xi) \\ z^*(b) &= c^T A_0^{-1} b_0(\xi) \end{aligned} \tag{4.24}$$

Since $b_0(\xi)$ follows a Gaussian distribution, $x^*(b)$ and $z^*(b)$ also follows a Gaussian distribution as they are linear combinations of $b_0(\xi)$. Due to the variation in RHS, $b(\xi)$, the active constraints set will change, namely selection of A_0 and b_0 will change; however, there are only finite many possible selections of A_0 and b_0, each with a probability in (4.22). That is, with a Gaussian RHS, $b(\xi)$, the optimal decision variables and objective function $x^*(b)$ and $z^*(b)$ follow Gaussian distribution with probability (4.22), which proves the result in (4.17). Notice that $x^* \in S^*$, and $p_i = P\{Ax^{(*)}(b) \leq b(\xi), C^T x^*(b) \leq C^T x^{(i)}(b), i = 1, 2, ..., r\}$, and p_i and p_j do not necessarily be one-one correspondence in (4.17) because linear combination of different Gaussian random variables may sum up to be the same in (4.23). Thus, $L \leq M$.

The number of Gaussian components for x^*, namely M in the Gaussian mixture distribution is the cardinality of set S^*, and M is thus bounded by the cardinality of S.

The upper bound of cardinality for S is C^n_{m+n}, which is the combination of choosing n constraints to construct a vertex from $m+n$ constraints. Note that many of the vertices constructed in this way will be infeasible, and thus not in P. Furthermore, far more vertices do not have a chance to be the optimal one (i.e., probability (4.22) is 0 for all possible realizations of $b(\xi)$). Therefore, this upper bound will be very loose.

For the upper bound of L, the dual problem of (4.16) needs to be considered

$$\begin{aligned} \min \quad & z' = b^T(\xi) y \\ s.t. \quad & \\ & A^T y \geq c \\ & y \geq 0 \end{aligned} \tag{4.25}$$

Since the optimal value will be the same for the primal problem (4.17) and the dual (4.24), for each realization of $b(\xi)$, the distribution for z^* and z'^* will be the same. The number of Gaussian components for the mixture distribution of z'^* is bounded above by the number of vertices in the convex polyhedron:

$$Q = \{y \in R^m | A^T y \leq c, y \geq 0\} \tag{4.26}$$

Therefore, the rough upper bound for L will be C^m_{m+n}, same as the value C^n_{m+n}, while from the Upper Bound Theorem of McMullen, [52], [53], the maximum number of vertices of polyhedron Q is expressed as $C^n_{m+n-\lfloor \frac{n+1}{2} \rfloor} + C^n_{m+n-\lfloor \frac{n+2}{2} \rfloor}$.

Therefore, the upper bound for the number of Gaussian components in the dual problem is $C^n_{m+n-\lfloor\frac{n+1}{2}\rfloor}+C^n_{m+n-\lfloor\frac{n+2}{2}\rfloor}$, and this is also the upper bound for the primal problem's distribution, which is the upper bound for L. $\square$

Notice that the upper bound for L and M are only dependent on m and n, which may be very loose in practice. A relatively tighter upper bound for L is the actual number of vertices in the dual problem polyhedron Q, which can be estimated by the algorithm proposed by Avis and Devroye [52].

Following Theorem 2.1, the distribution of objective function and decision variables in the SLP problem with Gaussian RHS constraints, are Gaussian mixture distributions. Since the direct analytical formula to determine the Gaussian mixture's parameters are generally not possible [32], Monte Carlo methods become a natural alternative. The specific procedures are as follows:

1. Draw samples from Gaussian distribution of $b(\xi)$ (the distribution parameters can be estimated from historical data). The necessary number of samples can be inferred from the upper bound of L and M.

2. Solve for resulting LPs by replacing random $b(\xi)$ with its realizations to obtain the optimal decision variables and objective functions, and collect data.

3. Estimate the parameters of Gaussian mixture distribution for the objective function and decision variables, where the number of Gaussian components is bounded above by M and L.

For parameter estimation, the EM algorithm is used as discussed in subsection 2.1.3.

Distribution metric between Gaussian mixtures

Once the Gaussian mixture distribution is estimated, it is reasonable to ask a question: how different are two or several such distributions? For example, in the case of steam plant optimization, it is valuable for the decision maker to know how much better the profit distribution under optimization is than that for routine operation. Comparison of the expected value or variance of profit, although simple, is an incomplete approach to evaluating the steam plant's performance. A distribution metric that can capture difference over the whole distribution is attractive, especially in the Gaussian mixture case, where each local component usually stands for a specific operation mode. If the metric can take into account the difference between local Gaussian components, it would be very helpful to the decision makers.

Earth Mover's Distance (EMD) will be introduced as the distribution metric between Gaussian mixture distributions to evaluate the performance of steam plant optimization. The definition of EMD is expressed as follows [33]:

$$\begin{aligned} \min : \quad & work\,(P,Q,F) = \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}d_{ij} \\ s.t. \quad & \\ & f_{ij} \geq 0 \\ & \sum_{j=1}^{n} f_{ij} \leq w_{pi} \\ & \sum_{i=1}^{m} f_{ij} \leq w_{qj} \\ & \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} w_{pi}, \sum_{j=1}^{n} w_{qj}\right) \end{aligned} \tag{4.27}$$

where $P = \{(p_1, w_{p_1}), ..., (p_m, w_{p_m})\}$, $Q = \{(q_1, w_{q_1}), ..., (q_n, w_{q_n})\}$ are two Gaussian mixture distributions, with different mixing probability $[w_{p_1}, ..., w_{p_m}]$ and $[w_{q_1}, ..., w_{q_n}]$, and different local Gaussian parameters $[p_1, ..., p_m]$ and $[q_1, ..., q_n]$. The calculation of EMD is based on the a transportation problem, which is formulated as an LP. The problem is to find a flow $F = [f_{ij}]$ that minimizes the overall cost with f_{ij} denoting the 'amount of earth' moved between the 'earth piles' w_{p_i} and w_{q_j}, with the 'moving cost' d_{ij}.

Using the calculated flow from LP solution, the normalized EMD can be expressed as:

$$EMD\,(P,Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}d_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}} \tag{4.28}$$

The physical interpretation of EMD in the case of Gaussian mixture distributions is the minimum cost needed to transform from one Gaussian mixture to another given the defined cost d_{ij} between local Gaussian components.

In the literature of EMD [33], d_{ij} is referred to as ground distance, denoting the distance between the basic features (in Gaussian mixture are the local Gaussian components) of distributions.

The problem here is to choose a proper ground distance between local Gaussian components so that the resulting EMD has appropriate properties for optimization, clustering and analysis.

Hellinger distance is chosen as such ground distance for EMD between Gaussian mixtures. The definition of the Hellinger distance for continuous distribution P and Q can be expressed as [54]:

$$H\left(P,Q\right)=\left[\frac{1}{2}\int\left(\sqrt{p\left(x\right)}-\sqrt{q\left(x\right)}\right)^{2}dx\right]^{1/2} \tag{4.29}$$

The Hellinger distance is chosen for the following reasons [54]:

1. The value of Hellinger distance is bounded between 0 and 1. This is convenient as performance index.

2. $H(P,Q)=0$ if and only if $p(x)=q(x)$ everywhere along the distribution, which means distribution $P=Q$; $H(P,Q)=1$ if and only if $p(x)q(x)=0$ everywhere on the distribution, which implies disjoint P and Q. This property gives Hellinger distance clear physical meaning as the dissimilarity of distributions.

3. Hellinger distance is a true metric, and satisfies symmetry, non-negativity, and the triangle inequality. This means the resulting EMD must be a true metric too [33], which is very appealing for applications based on this metric, such as optimization and clustering.

The Hellinger distance is closely related to the Bhattacharyya coefficient by equation (4.30), specifically, the Hellinger distance between two Gaussian distributions can be expressed as equation (4.31) [55].

$$\begin{aligned} H\left(p,q\right)&=\sqrt{1-BC\left(p,q\right)}\\ BC\left(p,q\right)&=\int\sqrt{p\left(x\right)q\left(x\right)}dx \end{aligned} \tag{4.30}$$

$$\begin{aligned} P &= \tfrac{\Sigma_0+\Sigma_1}{2} \\ BC\left(G\left(x;\mu_0,\Sigma_0\right),G\left(x;\mu_1,\Sigma_1\right)\right) &= \exp\left(-\tfrac{1}{8}(\mu_0-\mu_1)^T P^{-1}\left(\mu_0-\mu_1\right)\right)\sqrt{\tfrac{\sqrt{|\Sigma_0||\Sigma_1|}}{|P|}} \end{aligned} \tag{4.31}$$

Gaussian mixture is composed of local Gaussian distributions, and the EMD between two Gaussian mixtures can be calculated by taking the Hellinger distance between local Gaussian distributions as ground distance. Direct analytical expression for Hellinger distance between two Gaussian mixtures is not available; however, the Hellinger distance between Gaussian distributions can be written in a closed form as in (4.31). The upper layer EMD can be efficiently solved by LP, and lower layer Hellinger distance can be solved from analytical expression in (4.31), and thus the EMD between two Gaussian mixtures can be calculated efficiently in this way.

The adjustable Hellinger distance is developed by introducing one additional parameter α into equation (4.31) as follows:

$$\begin{aligned} BC\left(G\left(x;\mu_0,\Sigma_0\right),G\left(x;\mu_1,\Sigma_1\right)\right) &= \exp\left(-\tfrac{1}{8}\alpha(\mu_0-\mu_1)^T P^{-1}\left(\mu_0-\mu_1\right)\right)\sqrt{\tfrac{\sqrt{|\Sigma_0||\Sigma_1|}}{|P|}} \\ P &= \tfrac{\Sigma_0+\Sigma_1}{2} \\ 0 &\leq \alpha \leq \infty \end{aligned} \tag{4.32}$$

After introducing α, the adjustable Hellinger distance between two Gaussian components (4.32) can be treated as the original Hellinger distance between $N(\sqrt{\alpha}\mu_0,\Sigma_0)$ and $N(\sqrt{\alpha}\mu_1,\Sigma_1)$. Therefore all the desirable properties of Hellinger distance mentioned previously will hold. If set $\alpha \leq 1$, the user shrinks the influence of the difference in mean values between the two Gaussian distributions. In the extreme case, when $\alpha = 0$, all the emphasis is put on the difference in variance. If $\alpha \geq 1$, the user amplifies the difference in the mean value, so that even tiny difference in mean value can be reflected by the adjustable Hellinger distance.

4.3.2 Simulation example and application to steam plant optimization

In this subsection, the application of proposed performance index to an example problem and the evaluation of steam plant optimization will be discussed.

Numerical simulation for an example problem

Example: Company A decides to invest in a new production line for product α, and the sources of raw material for α are from two different suppliers, with daily supply amount denoted as x_1 and x_2. To set the purchasing plan, and to know the optimal daily net profit of this new product under constraints of daily budget, labour force limit, and market demand, a deterministic linear programming problem is formulated as:

$$
\begin{aligned}
\max \quad & z = 8.5x_1 + 6.5x_2 \quad \text{(daily net profit)} \\
s.t. \quad & \\
& 2.5x_1 + 1.5x_2 \leq 1250 \quad \text{(Daily budget)} \\
& 4x_1 + 3.5x_2 \leq 2500 \quad \text{(Labour force limit)} \\
& x_1 + x_2 \leq 750 \quad \text{(Daily market demand)} \\
& x_1, x_2 \geq 0
\end{aligned}
\tag{4.33}
$$

where the parameters in the objective function represent net profits using raw materials from different suppliers, and the parameters in the constraints represent different limitations on the two types of raw material. For example, in the daily budget constraint, 2.5 and 1.5 stands for the unit purchasing price of raw materials from suppliers.

This is a standard LP problem, where the solution is $z = \$4886.4$ and $x_1 = 227.3$, $x_2 = 454.5$. Due to the uncertainty in budget, available labour force and market fluctuation, it is impossible to estimate the net profit as a specific number. The resulting optimal objective function and decision variables will be random variables with known distribution to the best knowledge, when some of the problem parameters are unknown and randomly varying.

From previous experience running company A, and uncertainty analysis of this new investment on product α, the decision makers of the company find that from probability distribution point of view, the RHS of constraints in equation (4.33) can be approximated by Gaussian distributions centering at the nominal RHS value given in (4.33):

$$\begin{aligned}
\max \quad & z = 8.5x_1 + 6.5x_2 \quad \text{(daily net profit)} \\
s.t. \quad & \\
& 2.5x_1 + 1.5x_2 \leq 1250 + N(0, 300^2) \quad \text{(Budget uncertainty)} \\
& 4x_1 + 3.5x_2 \leq 2500 + N(0, 500^2) \quad \text{(Labour force uncertainty)} \\
& x_1 + x_2 \leq 750 + N(0, 150^2) \quad \text{(Market demand uncertainty)} \\
& x_1, x_2 \geq 0
\end{aligned} \tag{4.34}$$

Now the problem is, with Gaussian distributed RHS of constraints, what is the distribution of the optimal z^* and x_1^*, x_2^*. Knowing these allows the decision makers to assess the expected net profit and its variance with perfect information to make more reliable profit/risk analysis.

First draw 1000 samples from the Gaussian RHS, and then solve the resulting LPs to obtain optimal z^* and x_1^*, x_2^*. Follow the procedures in Chapter 2 and estimate the GM distributions of z^* and x_1^*, x_2^* using EM algorithm. By Theorem 4.3.1, the maximum possible number of Gaussian components for z^* is $C_{3+2-1}^2 + C_{3+2-2}^2 = 9$, for (x_1^*, x_2^*) is $C_{3+2}^2 = 10$. While in practice, some or most of the Gaussian components may not exist or only occur with negligible probabilities.

After the estimation procedure, the distributions for the optimal objective function and decision variables are given in the Figure (4.5) and Figure (4.6). From Figure (4.5), it can be seen that the distribution for objective function is non-Gaussian. For Figure (4.6), it seems the joint distributions are Gaussian lying at two different points on the x_1, x_2 plane; however, from the estimation result, the joint distributions in Figure (4.5) should be GMs rather than single Gaussian. This contradiction is from the non-negativity constraints of problem (3.37), which is deterministic and makes the estimated Gaussian component degenerate on the boundary. A degenerate Gaussian component has the variance almost 0 along some direction in the decision space, and thus has dominant amplitude. To make other Gaussian components in Figure (4.6) obvious for a better illustration, the non-negativity constraints are relaxed with a Gaussian RHS as $x_1, x_2 \leq 0.5 + N(0, 0.1^2)$, shown in Figure (4.7).

The expectation of the optimal objective function is calculated based on the estimated GM distribution, as $E(z^*) = \$4445.2$. The added value of knowing z^*'s distribution can be calculated as the expected value of perfect information (EVPI), which is the difference between the deterministic solution to (3.36) and the expected value of z^* for (4.34) [32].

$$EVPI = |4886.4 - 4445.2| = \$441.2 \tag{4.35}$$

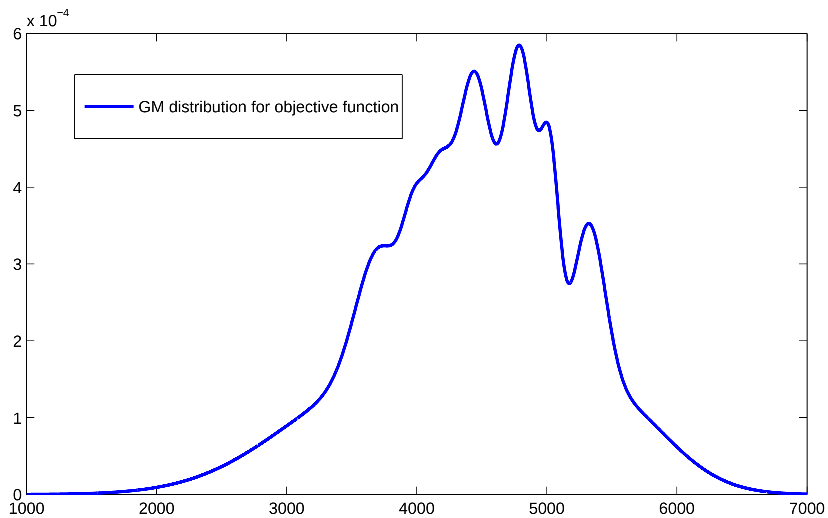


Figure 4.5: Estimated GM distribution for the optimal objective function z

Thus, the estimation based only on the mean value of RHS in (4.34) will be \$441.2 more optimistic on the potential daily net profit of the new product line, which is 9.92% more than the expected profit based on perfect information (the entire profit distribution). The variance of daily profit can be calculated as \$799.1 by the following closed form variance formula for mixture distribution:

$$\begin{aligned} Var(z) &= E[z-\mu(z)]^2 = E(z^2) - \mu(z)^2 \\ &= \sum_{j=1}^{L} p_j(\mu_j^2 + \sigma_j^2) - (\sum_{j=1}^{L} p_j\mu_j)^2 \end{aligned} \tag{4.36}$$

From the distribution of the objective function, the decision makers can also calculate different quantiles in which they are interested, when making the decision or doing profit/risk analysis.

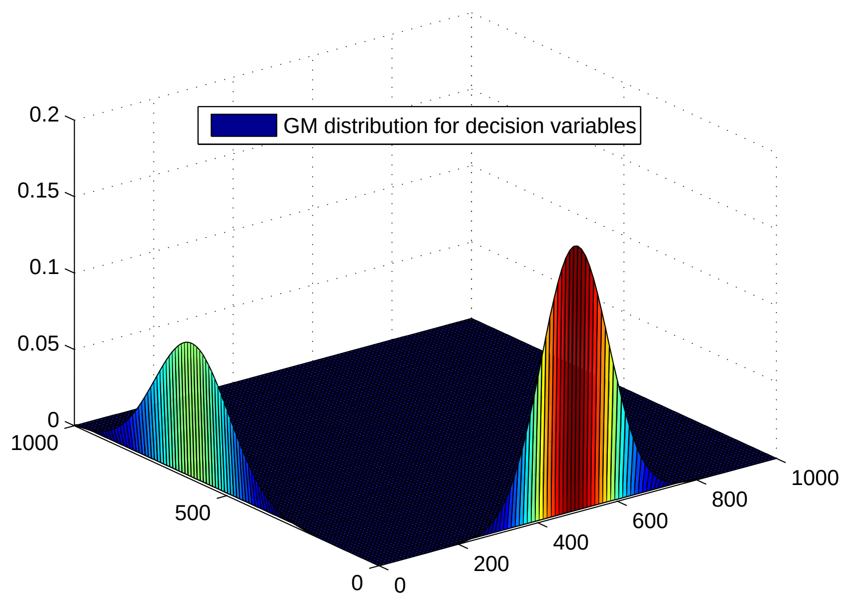


Figure 4.6: Estimated GM distribution for the optimal decision variables x_1^* and x_2^*

The next problem to consider is: if the estimated Gaussian mixture is taken as the benchmark purchasing plan, as time goes by, when the uncertainty of the constraints changes, how can the difference between the benchmark and actual purchasing activity be quantified?

EMD with Hellinger distance as the ground distance is used to measure the difference between two Gaussian mixture distributions representing the purchasing plans under uncertainty. The figures (4.8) to (4.10) show the calculated EMD between different purchasing plan pairs.

Figure (4.8) shows the decision variables' distributions under same RHS uncertainty from two rounds of Monte Carlo simulation. The estimated distribution should be very similar as shown in Figure (4.8), and actually the resulting EMD is 0.051 which is very close to 0 as expected. Figure (4.9) shows the difference between the benchmark distribution and the one with different RHS uncertainty (in this case, shifted mean value), the resulting EMD with Hellinger as ground distance is 0.558, showing the dissimilarity of the two distributions. If the adjustable parameter α is

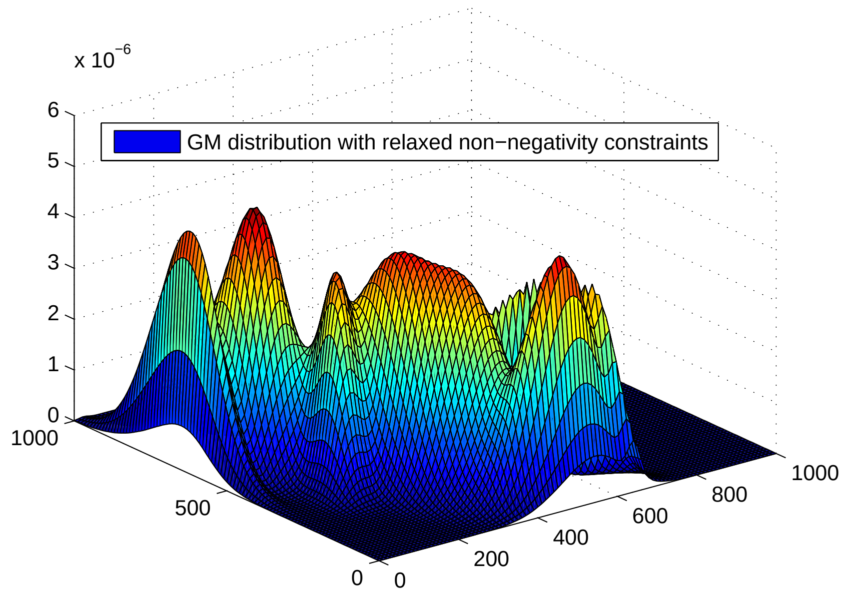


Figure 4.7: Estimated GM distribution for the optimal decision variables x_1^* and x_2^* with relaxed non-negativity constraints

set to be 0, the EMD will shrink to 0.3428; however, if $\alpha = 10$, the result will increase to 0.8528. Figure (4.10) shows the difference between the benchmark and the one with increased RHS Gaussian variance, the resulting EMD with Hellinger as ground distance is 0.411, showing the dissimilarity of the two distributions. If the adjustable parameter α is set to be 0, the EMD will shrink to 0.3513, it is not a big change since the mean values of the two distributions are relatively close after all; if $\alpha = 10$, the result will increase to 0.5667.

Application to steam plant optimization performance assessment

In this section, the application of proposed performance index to steam plant optimization will be shown. The routine operation of steam plant pays more attention to satisfy the process users' demand, while does not adjust the steam generation in response to the change in electricity price.

After solving the steam plant optimization problem using LP as explained pre-

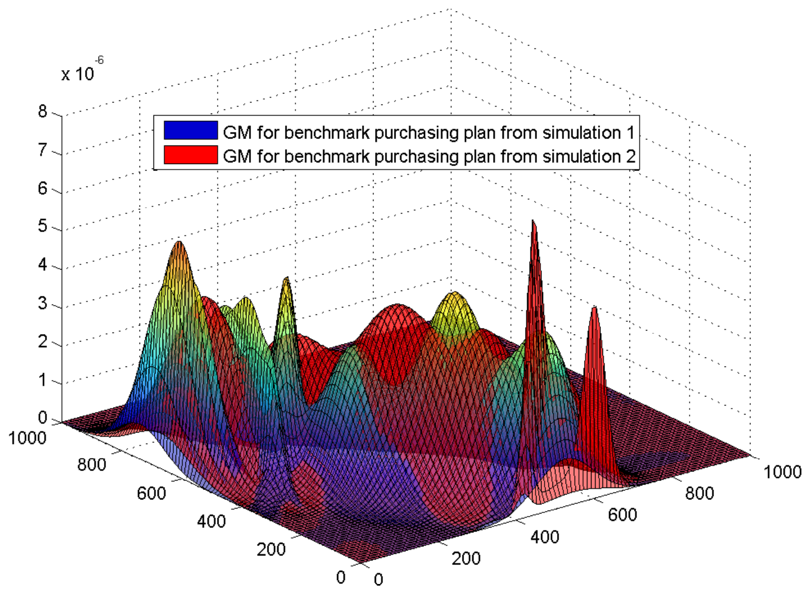


Figure 4.8: GM distribution for benchmark purchasing plan from two runs of Monte Carlo simulations, $EMD = 0.051$

ciously, the Monte Carlo simulation is run to draw samples from the RHS of constraint. The comparison of profit from routine operation and the optimized operation can be made using proposed performance index. For the deregulated electricity market in Alberta, the price can vary from several dollars per MWh to \$1000/MWh, and the higher the electricity price is, the more profit the optimization can contribute. Figure (4.11) to Figure (4.14) show the comparison of hourly profit from routine operation and optimized operation at different electricity price. At the same time, EMD is calculated to quantify the difference between the optimized operation and routine one. From Figure (4.11) to Figure (4.14), it can be seen that at each electricity price level, the optimized operation leads to less loss (if the profit turned out to be negative) or more profit. When the electricity price is relatively low, as in Figure (4.11) and Figure (4.12), the routine operation result and optimized operation result is comparable with small portion of overlapping; however, when the electricity price is high, as in Figure (4.13) and Figure (4.14), the added value of the optimized operation is significant, with thousands dollar's improvement per hour. The resulting EMD for the first two figures are already close to 1, indicating a quite improved hourly profit. For the last two figures, it is 1, indicating overwhelmingly improved hourly profit. Notice that using adjustable Hellinger distance as ground distance for EMD, these numbers can be adjusted to be smaller if more resolution is required in the performance index. A threshold EMD can be set to detect if the current operation is within the acceptable range near the optimal operation so that the steam plant operation can be monitored in a cost efficient way.

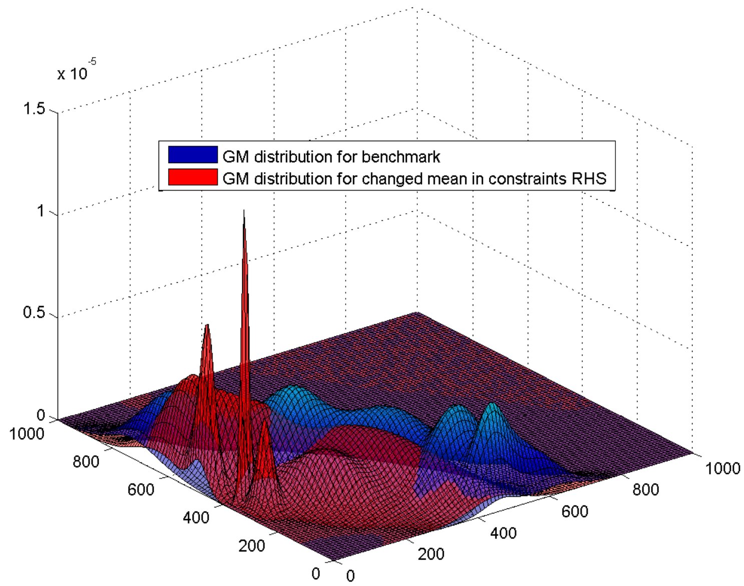


Figure 4.9: GM distributions comparison (different mean value of RHS). $EMD = 0.558$; $\alpha = 0$, EMD=0.3428; $\alpha = 10$, EMD=0.8528

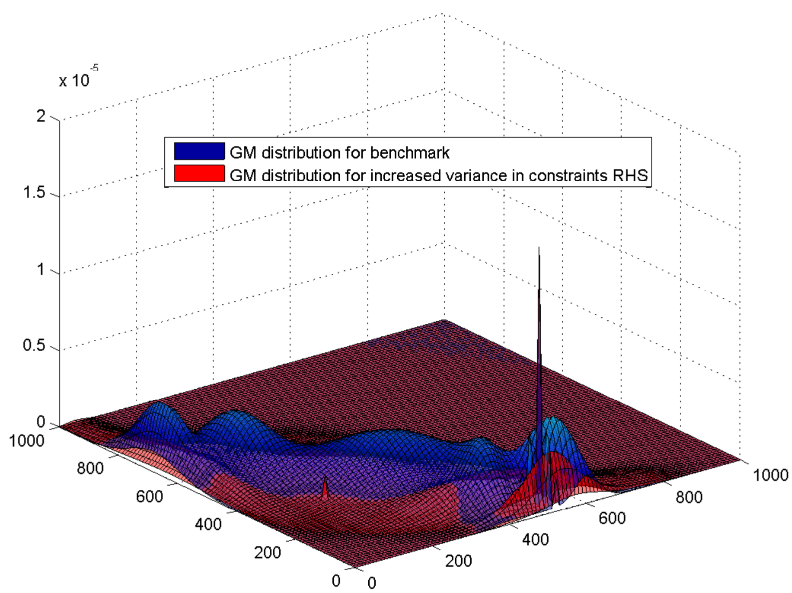


Figure 4.10: GM distributions comparison (increased variance of RHS). $EMD = 0.411$; $\alpha = 0$, EMD=0.3513; $\alpha = 10$, EMD=0.5667

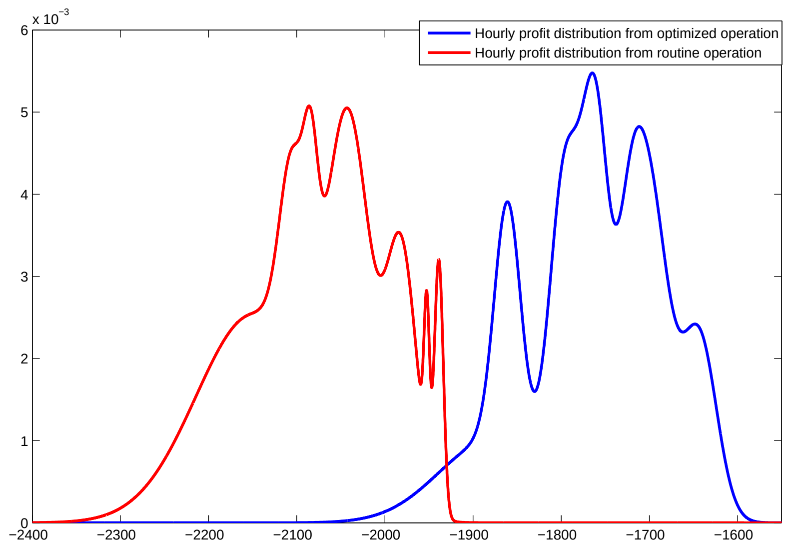


Figure 4.11: Hourly profit distribution at electricity price \$20, the expected improvement is \$312.73/h, $EMD = 0.9538$

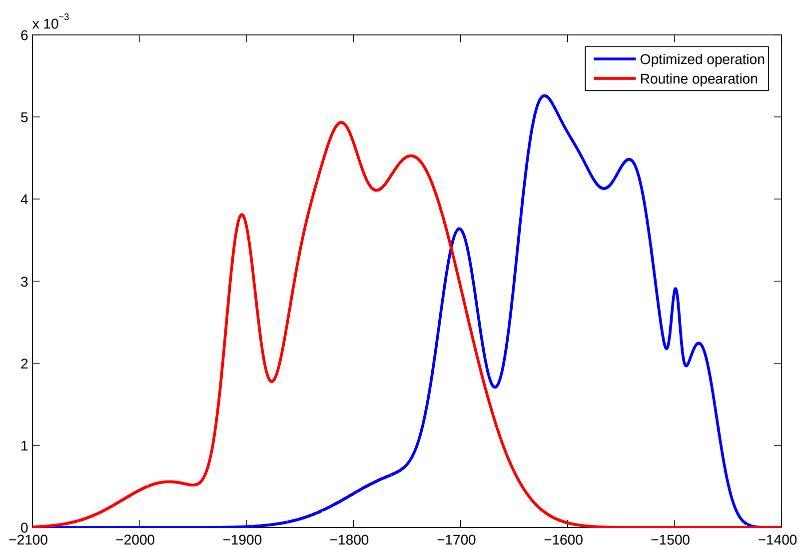


Figure 4.12: Hourly profit distribution at electricity price \$100, the expected improvement is \$195.40/h, $EMD = 0.8003$

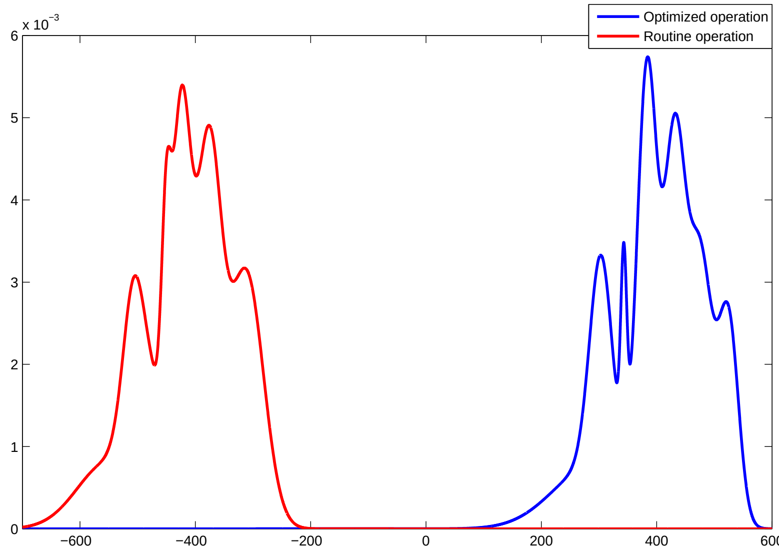


Figure 4.13: Hourly profit distribution at electricity price \$500, the expected improvement is \$813.52/h, $EMD = 1.00$

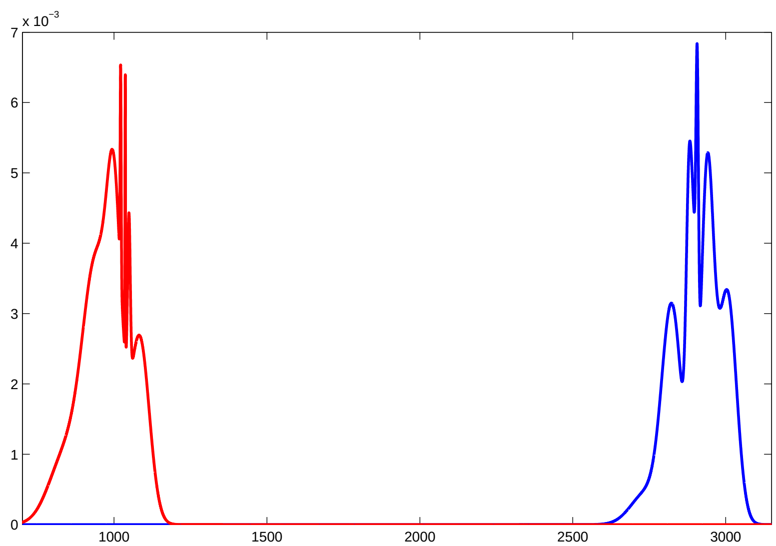


Figure 4.14: Hourly profit distribution at electricity price \$900, the expected improvement is \$1930.9/h, $EMD = 1.00$

Chapter 5

Conclusion

In this chapter, the work for this thesis is summarized, conclusions are drawn and future research directions are proposed.

5.1 Summary of this thesis

This thesis focuses on economic optimization of steam utility plant operation, in which the plant generates steam and electricity for process use as well as sales to the electricity grid.

A predictive optimization approach is proposed in this thesis to improve the steam plant profit. This approach incorporates steam plant models and electricity price prediction so that not only current but also future information can be used to maximize the long term profit.

Process models are developed in Chapter 1, such as the mass balances for each common header, random process variable models, turbine generator models and the steam boiler models. Multiple methods are used to develop the models: probability density estimation using EM algorithm, first principles modelling based on physical laws and system identification using process data. This chapter is the foundation for the predictive optimization, and the developed models are accurate in terms of fitting-rate or cross validation results.

Electricity price prediction model is developed in Chapter 2. A robust switched system identification method is proposed to build the price prediction model. This method introduces Gaussian mixture noise distribution to handle outliers robustly. The application of proposed method to electricity price prediction shows better performance than the forecast price published by AESO website, especially when electricity price is very high.

In Chapter 4, the steam plant model, random process variables, process equipment dynamic models and electricity price prediction are integrated to be a predictive optimization framework based on linear program. To implement the optimal boiler load trajectory calculated by the optimization, two control strategies are proposed: the dynamic linear programming approach is an open-loop control strategy and calculates the optimal manipulation variables directly from the steam plant optimization; and the optimal tracking controller approach is a closed-loop control strategy, which tracks the optimal boiler load trajectories via LQI or MPC control law. To evaluate the added value of the proposed optimization, a quantitative performance assessment approach is developed. It compares the optimized plant profit with the profit under routine plant operation by calculating the EMD between the distributions representing the profits. Based on the performance assessment result, the optimized plant operation is significantly better than the routine operation especially when the electricity price varies a lot.

5.2 Directions for future work

In this thesis, the steam plant optimization problem is solved under process uncertainty, including process dynamic models and the electricity price predictions. To further improve the performance of steam plant operation, following aspects can be considered in future work:

1. In this thesis, varying electricity price is taken into account on the profit side of the economic objective function, and the cost of natural gas is treated as a constant in steam generation. In reality, natural gas price does change with time and energy market situation. The prediction model for natural gas price will be a direction to improve the profit of steam plant.

2. The user demands for each steam header are assumed to be random variables. While in reality, the user demands for different units can be modelled based on the production schedule. With such models, the future user demand can be predicted in a more accurate way for better optimization effect.

3. Based on the distribution metric defined in Chapter 4, routine operation and optimal operation of steam plant can be compared using earth movers' distance. Following this way, the plant operation data can be clustered in real time such as optimal operation, normal operation, uneconomic operation, abnormal operation and so forth. An alarm/guidance system can be developed based on the clustering results for operators' and engineers' convenience.

Bibliography

[1] Yu Zhao, Biao Huang, Hongye Su, and Jian Chu. Prediction error method for identification of lpv models. *Journal of process control*, 2011.

[2] William Weng. Stochastic optimization for manufacturing processes. *University of Alberta*, 2013.

[3] Biao Huang and Rasoul Milasi. Multi tank system. *University of Alberta*, 2009.

[4] PS Varbanov, S Doyle, and R Smith. Modelling and optimization of utility systems. *Chemical Engineering Research and Design*, 82(5):561–578, 2004.

[5] Khadem Moslehi, M Khadem, R Bernal, and G Hernandez. Optimization of multiplant cogeneration system operation including electric and steam networks. *Power Systems, IEEE Transactions on*, 6(2):484–490, 1991.

[6] Wangyun Won, Kwang Soon Lee, In Seop Kim, Bongkook Lee, Seungjoo Lee, and Seokyoung Lee. Model predictive control of condensate recycle process in a cogeneration power station: Controller design and numerical application. *Korean Journal of Chemical Engineering*, 25(5):972–979, 2008.

[7] Andrés Collazos, François Maréchal, and Conrad Gähler. Predictive optimal management method for the control of polygeneration systems. *Computers & Chemical Engineering*, 33(10):1584–1592, 2009.

[8] Giancarlo Ferrari-Trecate, Eduardo Gallestey, Paolo Letizia, Matteo Spedicato, Manfred Morari, and Marc Antoine. Modeling and control of co-generation power

plants: a hybrid system approach. *Control Systems Technology, IEEE Transactions on*, 12(5):694–705, 2004.

[9] Chi-Wai Hui and Yukikazu Natori. An industrial application using mixed-integer programming technique: a multi-period utility system model. *Computers & chemical engineering*, 20:S1577–S1582, 1996.

[10] X. Jin, B. Huang, and D.S. Shook. Multiple model lpv approach to nonlinear process identification with em algorithm. *Journal of Process Control*, 21(1): 182–193, 2011.

[11] N.H. El-Farra and P.D. Christofides. Coordinating feedback and switching for control of hybrid nonlinear processes. *AIChE Journal*, 49(8):2079–2098, 2004.

[12] S. Paoletti, A.L. Juloski, G. Ferrari-Trecate, and R. Vidal. Identification of hybrid systems: A tutorial. *European Journal of Control*, 13(2-3):242–260, 2007.

[13] X. Jin and B. Huang. Robust identification of piecewise/switching autoregressive exogenous process. *AIChE Journal*, 56(7):1829–1844, 2010.

[14] X. Jin and B. Huang. Identification of switched markov autoregressive exogenous systems with hidden switching state. *Automatica*, 2011.

[15] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[16] R.B. Gopaluni. Identification of non-linear processes with known model structure under missing observations. In *Proceedings of the IFAC 17th World Congress, Seoul, Korea, July 6*, volume 11, 2008.

[17] J. Deng and B. Huang. Identification of nonlinear parameter varying systems with missing output data. *AIChE Journal*, 2012.

[18] S.J. Roberts. Novelty detection using extreme value statistics. In *Vision, Image and Signal Processing, IEE Proceedings-*, volume 146, pages 124–129, 1999.

[19] N Kostantinos and D. Hatzinakos. Gaussian mixtures and their applications to signal processing. *University of Toronto*, 2010.

[20] S. Khatibisepehr and B. Huang. Bayesian methods for process identification with outliers. In *American Control Conference (ACC), 2012*, pages 3516–3521, 2012.

[21] H.W. Sorenson and D.L. Alspach. Recursive bayesian estimation using gaussian sums. *Automatica*, 7(4):465–479, 1971.

[22] A. Shamshad, MA Bawadi, WMA Wan Hussin, TA Majid, and SAM Sanusi. First and second order markov chain models for synthetic generation of wind speed time series. *Energy*, 30(5):693–708, 2005.

[23] Scott M Thede and Mary P Harper. A second-order hidden markov model for part-of-speech tagging. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 175–182. Association for Computational Linguistics, 1999.

[24] Brett Watson and A Chung Tsoi. Second order hidden markov models for speech recognition. In *Proceedings, Fourth Australian International Conference on Speech Science and Technology*, pages 146–151, 1992.

[25] J-F Mari, J-P Haton, and Abdelaziz Kriouile. Automatic word recognition based on second-order hidden markov models. *Speech and Audio Processing, IEEE Transactions on*, 5(1):22–25, 1997.

[26] Behram J Hansotia. Stochastic linear programming with recourse: A tutorial*. *Decision Sciences*, 11(1):151–168, 1980.

[27] Hüseyin Sarper. Monte carlo simulation for analysis of the optimum value distribution in stochastic mathematical programs. *Mathematics and computers in simulation*, 35(6):469–480, 1993.

[28] MM Babbar. Distributions of solutions of a set of linear equations (with an application to linear programming). *Journal of the American Statistical Association*, 50(271):854–869, 1955.

[29] Gerhard Tintner. Stochastic linear programming with applications to agricultural economics. In *Proceedings of the Second Symposium in Linear Programming*, volume 1, pages 197–228. National Bureau of Standards: Washington, DC, 1955.

[30] Harvey M Wagner. On the distribution of solutions in linear programming problems. *Journal of the American Statistical Association*, 53(281):161–163, 1958.

[31] Ion M Stancu-Minasian and Victor Giurgiutiu. *Stochastic programming with multiple objective functions*. Editura Academiei, 1984.

[32] Jerome Bracken and Richard M Soland. Statistical decision analysis of stochastic linear programming problems. *Naval Research Logistics Quarterly*, 13(3):205–225, 1966.

[33] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40 (2):99–121, 2000.

[34] Douglas Reynolds. Gaussian mixture models. *Encyclopedia of Biometric Recognition*, 2(17.36):14–68, 2008.

[35] G.J. McLachlan and T. Krishnan. *The EM algorithm and extensions*, volume 382. 2007.

[36] Jeff A Bilmes et al. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4(510):126, 1998.

[37] C. Tomasi. Estimating gaussian mixture densities with em–a tutorial. *Duke University*, 2004.

[38] Karl Johan Åström and Rod D Bell. Drum-boiler dynamics. *Automatica*, 36(3): 363–378, 2000.

[39] Lennart Ljung. *System identification*. Wiley Online Library, 1999.

[40] Market Surveillance Administrator. Alberta wholesale electricity market. 2010.

[41] Lei Xiong. Stochastic models for electricity prices in alberta. *University of Calgary*, 2004.

[42] G. Lee and B. Jr. Stochastic financial models for electricity derivatives. *Standord University*, 1999.

[43] L. Ljung. System identification: Theory for the user. *Preniice Hall Inf and System Sciencess Series, New Jersey*, 7632, 1987.

[44] A.L. Juloski, S. Weiland, and W. Heemels. A bayesian approach to identification of hybrid systems. *Automatic Control, IEEE Transactions on*, 50(10):1520–1533, 2005.

[45] H.W. Kuhn and A.W. Tucker. Nonlinear programming. In *Second Berkeley symposium on mathematical statistics and probability*, volume 1, pages 481–492, 1951.

[46] W. Karush. Minima of functions of several variables wih inequalities as side constraints. *M.Sc. dissertation, University of Chicago*, 1939.

[47] C. Biernacki, G. Celeux, and G. Govaert. Choosing starting values for the em algorithm for getting the highest likelihood in multivariate gaussian mixture models. *Computational Statistics & Data Analysis*, 41(3):561–575, 2003.

[48] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[49] R. Morales-Menéndez, N. De Freitas, and D. Poole. Estimation and control of industrial processes with particle filters. In *American Control Conference, 2003. Proceedings of the 2003*, volume 1, pages 579–584, 2003.

[50] A. Tamir. *Applications of Markov chains in chemical engineering*. 1998.

[51] Peter Colin Young and JC Willems. An approach to the linear multivariable servomechanism problem. *International journal of control*, 15(5):961–979, 1972.

[52] David Avis and Luc Devroye. Estimating the number of vertices of a polyhedron. *Information processing letters*, 73(3):137–143, 2000.

[53] Peter M Gruber and Jörg M Wills. *Handbook of convex geometry*, volume 2. 1993.

[54] Rafał ŁOPATKA and Dominik OLSZEWSKI. Employing probabilistic dissimilarity for feature discovery in a game of chess. *Electrical Review*, 87(1):240–244, 2011.

[55] Leandro Pardo. *Statistical inference based on divergence measures*, volume 185. Chapman and Hall/CRC, 2005.