

DATA-DRIVEN MODELLING FOR PROCESS IDENTIFICATION WITH FLAT-TOPPED  
GAUSSIAN UNCERTAINTY

by

**Ruomu Tan**

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science

in

Process Control

Department of Chemical and Materials Engineering

University of Alberta

©Ruomu Tan, 2015

# Abstract

In data-driven modelling, model accuracy relies heavily on the data set collected from target process. However, various types of measurement noise exist extensively in industrial processes and the data obtained are usually contaminated. If the influence of measurement noise is neglected, both the quality of models trained from data and performance of further operations, such as control and optimization of objective variables, will be affected significantly. A good output noise model is essential in data-driven modelling if one wishes to attain a process model with satisfactory performance.

Instead of the regular Gaussian distribution assumption for the noise, a novel type of the noise distribution is proposed and corresponding solutions to the process identification problems are established accordingly in this thesis. Specifically, a flat-topped Gaussian distribution, which combines the Gaussian and uniform distribution, is formulated to model a class of disturbances that often occur in practice. Moment fitting strategy is proposed as a general approach to approximating the distribution function of summed random variables with different distributions. The Flat-topped Gaussian distribution is then applied for identification of linear processes.

As for more complicated nonlinear models, Flat-topped Gaussian distribution is considered for Gaussian Process modelling. Gibbs Sampling is incorporated and combined with the posterior distribution obtained from Gaussian Process in order to reconstruct the original output. Mixture Gaussian approximation is also used as an alternative of approximating the Flat-topped Gaussian distribution.

The proposed algorithms are validated by numerical simulations and industrial applications. Soft sensors for estimation of emulsion flow rate in Steam Assisted Gravity Drainage (SAGD) process based on data-driven modelling are developed and relevant practical issues are discussed.

# Acknowledgements

First and foremost, it gives me great pleasure in acknowledging the patience, support and guidance of my supervisors, Professor Biao Huang and Professor Zukui Li, all the way through my graduate study. I would like to thank Dr. Huang for his continuing advice which always shredded lights on my works in this thesis, as well as his patience in explaining and clarifying the concept, philosophy and principle behind the work which I had trouble understanding myself. Also, had it not been for the support of Dr. Huang, it could never be possible for me to get involved in various academical and industrial activities, which helped me broaden my horizon as well as develop interpersonal skills. I would like to thank Dr. Li for his support in helping me understand the concepts and approaches in optimization which turned out to be really beneficial for my research even though I was not familiar with them at the beginning.

It has been an honour for me to be part of the Computer Process Control (CPC) group since there are so many extraordinary members to whom I am gratefully indebted. I would like to thank Dr. Swanand Khare and Dr. Jiusun Zeng in particular, for they have brought up and explained the idea which became the very foundation of my research later on. I would like to thank all my present and previous colleagues for generously sharing their precious thoughts and ideas with me, listed but not limited to as follows: Rishik Ranjan, Yanjun Ma, Chao Shang, Shakher Sharma, Dr. Alireza Fatehi, Ouyang Wu, Yujia Zhao, Yaojie Lu, Ming Ma, Dr. Nima Sammaknejad, Anahita Sadeghian, Elham Naghoosi, Mohammad Rashed, Dr. Ruben Gonzalez, Dr. Fadi Ibrahim, Wenhan Shen and Rahul Raveendran. I really enjoyed discussing with all of them and I was always fascinated by the novel ideas and ingenious solutions which they brought up. I would also express my gratefulness to my friends for their selfless encouragement, inspiration and companionship: Jing Zhang, Su Liu, Tianbo Liu, Shuning Li, Shunyi Zhao, Kangkang Zhang, Ye Shen, Ruoxia Li, Xiaodong Xu, Yuan Yuan, Liu Liu and many else.

I would also like to thank Eliyya Shukeir and Ramesh Kadeli from Suncor Energy for their kindly help during the progress of industrial soft sensor project which makes an

important part of this thesis and I really appreciate the opportunity of working jointly with engineers towards industrial applications.

Financial support from NSERC and Alberta Innovative Technology Futures (AITF) is sincerely acknowledged.

It was such a luck for me to share the past two years here with my cousin sister, Ruohua Liang, who has just got her Bachelor of Arts degree at University of Alberta this fall. Last but not least, this thesis would not have been possible without the consistent love, trust and support from my dearest parents.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Thesis Contributions . . . . .	2
1.3	Thesis Outline . . . . .	3
<b>2</b>	<b>Linear Model Identification with Flat-topped Gaussian Uncertainty</b>	<b>4</b>
2.1	Introduction . . . . .	4
2.2	Revisit of Noise Distributions & Flat Gaussian Profile . . . . .	8
2.2.1	Noise Distributions . . . . .	8
2.2.2	Flat Gaussian Profile . . . . .	10
2.3	Flat-topped Gaussian Distribution . . . . .	11
2.3.1	Problem Statement . . . . .	11
2.3.2	Moment Fitting Method . . . . .	13
2.3.3	Parameter Estimation . . . . .	15
2.3.4	Mixture Gaussian approximation . . . . .	19
2.4	Flat-topped $t$ Distribution . . . . .	21
2.4.1	Robustness Issue in Flat-topped Gaussian Distribution . . . . .	21
2.4.2	Formulation of Flat-topped $t$ Distribution . . . . .	22
2.4.3	Potential Advantages of Flat-topped $t$ Distribution . . . . .	24
2.5	Estimation of Noise Parameters . . . . .	25
2.5.1	Noise Estimation for Flat-topped Gaussian Distribution . . . . .	25
2.5.2	Noise Estimation for Flat-topped $t$ Distribution . . . . .	26
2.5.3	Iterative Procedure of Noise Parameter Estimation . . . . .	27
2.6	Simulations & Results . . . . .	27
2.6.1	Parameter Estimation of Flat-topped Gaussian Distribution . . . . .	27
2.6.2	Simulation with Quantization Error . . . . .	30

2.7	Conclusion and Future Work . . . . .	33
<b>3</b>	<b>Nonlinear Identification of Gaussian Process with Flat-topped Gaussian Uncertainty</b>	<b>35</b>
3.1	Introduction . . . . .	35
3.2	Review of Methodologies . . . . .	37
3.2.1	Gaussian Process Modelling . . . . .	37
3.2.2	Gibbs Sampling . . . . .	41
3.3	Problem Formulation . . . . .	42
3.4	Data Reconstruction: Gibbs Sampling Approach . . . . .	45
3.4.1	Mixture Gaussian Approximation . . . . .	45
3.4.2	Posterior Distribution . . . . .	47
3.4.3	Summary of Algorithm . . . . .	50
3.5	Simulation & Results . . . . .	51
3.6	Discussion & Future Work . . . . .	55
<b>4</b>	<b>Design of Emulsion Flow Soft Sensor for SAGD Process</b>	<b>56</b>
4.1	Introduction . . . . .	56
4.2	Process Description . . . . .	58
4.3	Model Identification . . . . .	59
4.3.1	Data Preprocessing . . . . .	59
4.3.2	Training Parameters . . . . .	62
4.3.3	OLS Performance Evaluation . . . . .	66
4.4	Robust Layer Development . . . . .	70
4.4.1	Sensor Robustness . . . . .	70
4.4.2	Back-up Model Development . . . . .	70
4.4.3	Performance Evaluation . . . . .	73
4.5	Data Reconciliation in Bias Correction . . . . .	74
4.5.1	Data Reconciliation & Bias Correction Approaches in Soft Sensor Development . . . . .	74
4.5.2	Problem Formulation . . . . .	75
4.5.3	Performance Evaluation . . . . .	78
4.6	Discussion & Future Work . . . . .	79
4.7	Conclusion . . . . .	81

<b>5</b>	<b>Conclusions</b>	<b>85</b>
5.1	Summary of This Thesis . . . . .	85
5.2	Directions for Future Work . . . . .	86
	<b>Bibliography</b>	<b>88</b>

# List of Tables

2.1	Model Parameters Trained with Different $\Delta$ . . . . .	33
3.1	Examples of Mean & Covariance Functions in Gaussian Process . . . . .	38
3.2	Prediction Performance of Gaussian Process Trained by Quantized and Re-constructed Data . . . . .	54
4.1	Process Variables in Model Building . . . . .	62
4.2	Data Description in Model Building . . . . .	64
4.3	Quantitative Model Performance . . . . .	65
4.4	Quantitative Gaussian Process Performance . . . . .	66
4.5	Data Description in Fast-rate Off-line Test . . . . .	67
4.6	Quantitative Model Performance in Fast-rate Off-line Test . . . . .	69
4.7	Data Description in Pad 2 . . . . .	69
4.8	Quantitative Model Performance of Pad 2 . . . . .	69
4.9	Data Description in Back-up Model Testing . . . . .	73
4.10	Sensor Performance Evaluation Before/After Bias Correction . . . . .	80

# List of Figures

1.1	Layout of Problems and Proposed Solutions . . . . .	3
2.1	Comparison of Regressions w/ Normal & Uniform Uncertainties . . . . .	5
2.2	Continuous and Quantized Signals . . . . .	7
2.3	Process Diagram with Instrumentation Noise and Quantization Error . . . . .	7
2.4	Quantization Error vs Summation of Quantization Error and White Noise . . . . .	8
2.5	Examples of Approximated Flat Gaussian Profile . . . . .	11
2.6	Histogram of Flat-topped Gaussian Random Variable with Different Uniform Width . . . . .	12
2.7	Flow Chart of Moment Fitting Strategy . . . . .	15
2.8	Performance of Mixture Gaussian Approximation with Different $m$ . . . . .	21
2.9	Comparison of Gaussian and Flat-topped Gaussian Loss Functions . . . . .	22
2.10	Comparison of True Histogram and Approximated Function . . . . .	24
2.11	Comparison of Gaussian, Flat-topped Gaussian and Flat-topped $t$ Loss Functions . . . . .	25
2.12	Comparison of Approximated PDF and Sample Histogram . . . . .	29
2.13	Comparison of Approximated PDF and Sample Histogram . . . . .	30
2.14	ARX Model with Quantization Module . . . . .	31
2.15	Influences on Different Quantization Resolutions . . . . .	32
2.16	Comparison of Relative Error by Two Approaches . . . . .	32
2.17	Comparison of Infinite Step Prediction by Two Approaches when $\Delta = 6$ . . . . .	34
3.1	Illustration of Gibbs Sampling Procedure . . . . .	42
3.2	Process Diagram under Gaussian Process Framework . . . . .	42
3.3	Descending Trend of Correlation between $x_i$ and $x_j$ . . . . .	44
3.4	Illustration of Solving Gaussian Process Model with Additional Noise . . . . .	45
3.5	Comparison of $v \sim U[-20, 20]$ and Corresponding Gaussian Mixture with $\sigma^2 = 4, m = 20$ . . . . .	46

3.6	Illustration of Sampling from Distributions of $Y_1^{(i)}$ . . . . .	50
3.7	Flowchart of Gibbs Sampling Procedure . . . . .	51
3.8	Trend Plot of Nonlinear Model Used . . . . .	52
3.9	Reconstruction Performance in Training Set ( $\Delta = 2$ ) . . . . .	54
3.10	Reconstruction Performance in Training Set ( $\Delta = 0.5$ ) . . . . .	54
4.1	Flowchart of Soft Sensor Development Procedure . . . . .	58
4.2	Well Pair in SAGD Process. Source: JAPEX . . . . .	59
4.3	Example of Noisy Input Data with Outliers . . . . .	60
4.4	Example of Noisy Input Data with Outlier Replaced . . . . .	61
4.5	Comparison of Pressure Measurement Before/After Smoothing . . . . .	62
4.6	Emulsion Flow Soft Sensor Diagram . . . . .	63
4.7	Model Performances on Well Pair 1 . . . . .	64
4.8	Model Performances on Well Pair 2 . . . . .	64
4.9	Performance of Gaussian Process on Well Pair 1 . . . . .	66
4.10	Cross-validation Performance of Pad 1 Well Pair 1 . . . . .	68
4.11	Cross-validation Performance of Pad 1 Well Pair 2 . . . . .	68
4.12	Input and Output Trends of Well Pair 7 . . . . .	71
4.13	Input and Output Trends of Well Pair 7 . . . . .	71
4.14	Example 1 of Limitation of Back-up Model . . . . .	72
4.15	Example 2 of Limitation of Back-up Model . . . . .	72
4.16	Performance Comparison of Original Model and Original Model w/ Back-up . . . . .	74
4.17	Existence of Gross Error between Overall Reference and Prediction . . . . .	79
4.18	Overall Sensor Performance . . . . .	81
4.19	Individual Sensor Performance Before/After Bias Update: WP1-6 . . . . .	82
4.20	Individual Sensor Performance Before/After Bias Update: WP7-12 . . . . .	83
4.21	Individual Sensor Performance Before/After Bias Update: WP13-14 . . . . .	84

# Chapter 1

## Introduction

### 1.1 Motivation

Data-driven modelling approach has enjoyed a vigorous development in recent years as the statistical modelling and machine learning methods advance. Commonly in this procedure, the measured data are contaminated by measurement noises of various kinds. To formulate the data-driven modelling problem from statistical perspective, it is necessary to acquire the noise distribution a priori.

It is essential to obtain a good measurement noise distribution model in order to improve the accuracy of data-driven modelling. The most widely used noise assumption is white Gaussian noise which yields Ordinary Least Square based methods. However, when the regular Gaussian noise assumption mismatches with true measurement noise, it can be anticipated that model identified accordingly will deviate from the real one and prediction performance based on the inaccurate model will also be poor. In this case, other more appropriate distributions, such as Student- $t$  distribution, Mixture Gaussian distribution and Mixture  $t$  distribution, are incorporated as noise distributions to deal with more complicated measurement noise issues such as outliers and/or multi-modal behaviours [1, 2, 3].

Uniformly distributed noise, as a certain type of uncertainty, exists extensively in different measurement approaches. For instance, measurements obtained from lab analysis usually possess an error interval indicating its reliability. Due to the randomness in instrumentation and operations, it is not necessary to assign the highest weight of trust to the measured value as the Gaussian noise assumption does. Instead, all the values within the error interval are considered to be equally probable which actually corresponds to a uniform noise assumption. Quantization error, which is caused by the analog-to-digital conversion, is another example of the existence of uniform uncertainty. Unlike other noise distributions from the exponential family, uniform noise is simple in expression but will result in a

truncated likelihood function that is difficult to handle in mathematics. Therefore, the formulation of noise distribution and corresponding identification algorithm deserves a special consideration in presence of uniform noise.

This work focuses on dealing with uniformly distributed noise combined with the traditional Gaussian noise, known as Flat-topped Gaussian distribution. For a parametric model, a novel type of flat-topped distributions is proposed as an alternative to regular Gaussian noise. For nonlinear Gaussian Process model with Flat-topped Gaussian noise, an iterative de-noising algorithm is developed based on Gibbs Sampling method. In addition, the proposed algorithms are validated in soft sensor development for emulsion flow estimation in SAGD process.

## 1.2 Thesis Contributions

This thesis contributes mainly on the identification of processes with Flat-topped Gaussian distributed noises. Under different assumptions of model structure, corresponding solutions are proposed to improve identification performance. Detailed contributions of this work are listed as follows:

1. Established the Flat-topped Gaussian distribution, which is an approximation of the distribution function of the summation of a uniform and a Gaussian distributed random variable, as an alternative to Gaussian distribution for process identification. Derived its functional structure and parameter estimation.
2. Proposed moment fitting strategy for the approximation of flat-topped distributions. This strategy is capable of approximating the combinations of random variables with arbitrary distributions.
3. Developed an iterative data de-noising algorithm for Gaussian Process models with Flat-topped Gaussian noise based on Gibbs Sampling method and Mixture Gaussian approximation.
4. Designed a soft sensor for emulsion flow estimation in SAGD process. An extra robust layer and bias correction with data reconciliation were included for better on-line implementation. Evaluated the integrated sensor performance on industrial data sets obtained from real SAGD process.

## 1.3 Thesis Outline

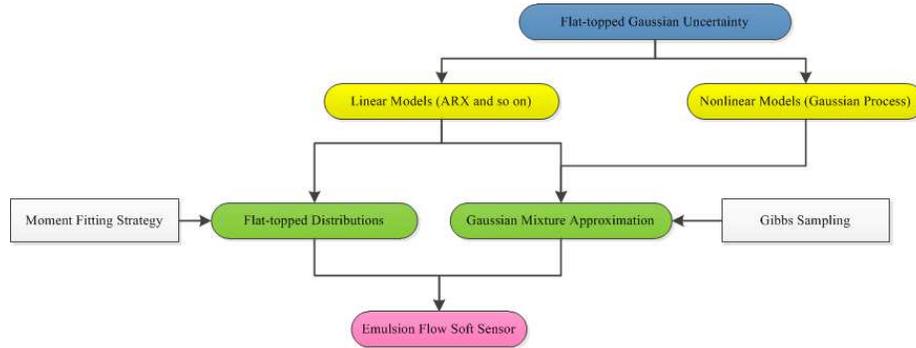


Figure 1.1: Layout of Problems and Proposed Solutions

Fig 1.1 outlines the problems and corresponding solutions of the thesis. Contributions of this work will be presented in three chapters and the layout of this thesis is as follows:

Chapter 2 proposes a novel type of flat-topped distributions to model a class of measurement noises. To analytically approximate noise distribution function, moment fitting strategy is established in order to be applied to approximate the distribution function of summed random variables.

In Chapter 3, similar output noise model under Gaussian Process framework is considered. In order to achieve a more accurate hyperparameter estimation and better prediction performance accordingly, Gibbs Sampling approach is incorporated to reconstruct the original measurements from contaminated ones.

Chapter 4 presents a procedure of industrial soft sensor development for emulsion flow in SAGD process. In off-line modelling part, efficiency of proposed Flat-topped Gaussian distribution and Gaussian Process approach is tested. Several issues in on-line implementation are discussed.

Chapter 5 draws a conclusion of the thesis.

## Chapter 2

# Linear Model Identification with Flat-topped Gaussian Uncertainty

In this chapter, a novel category of flat-topped distributions is proposed as a noise distribution in process identification as an alternative to Gaussian noise distribution. To begin with, Flat-topped Gaussian distribution which is capable of handling convoluted Gaussian and uniform noise in measurements is introduced and investigated. Moment fitting strategy, as a by-product, is formulated for estimating parameters in Flat-topped Gaussian distribution functions. Afterwards the Flat-topped  $t$  distribution is proposed by analogy. In addition to handling uniform noise, algorithm robustness is enhanced by incorporating  $t$  distribution. Two numerical examples are presented to demonstrate the feasibility of the proposed moment fitting strategy in estimating parameters of Flat-topped Gaussian distributions. Furthermore, a numerical example is studied to demonstrate the advantage of Flat-topped Gaussian distribution as noise assumption in identification of linear processes in presence of additional uniform measurement uncertainty such as quantization error.

### 2.1 Introduction

Measurement uncertainty is common in industrial practice. In reality, uncertainty is originated from different sources and varies in characteristics. In data-driven system identification, it is necessary to reduce measurement uncertainty in order to achieve good identification results and a specified noise assumption is required in identification algorithms. Two types of possible uncertainty and corresponding regression results are illustrated in Fig 2.1. The measurement uncertainty is mostly assumed to be normally distributed with its mode at the measured value as shown in Fig 2.1(a); nevertheless, when it comes to the situation where the measurement is inaccurate or less trustworthy, it is not necessary to assign the

largest weight to the measured value. Instead, all values within a certain error band near the measurement can be equally weighted as shown in Fig 2.1(b). Corresponding regression results are also compared. It can be seen that under normal uncertainty, the regression line is forced to capture all measure values as much as it can while it is more flexible when uniform assumption is adopted.

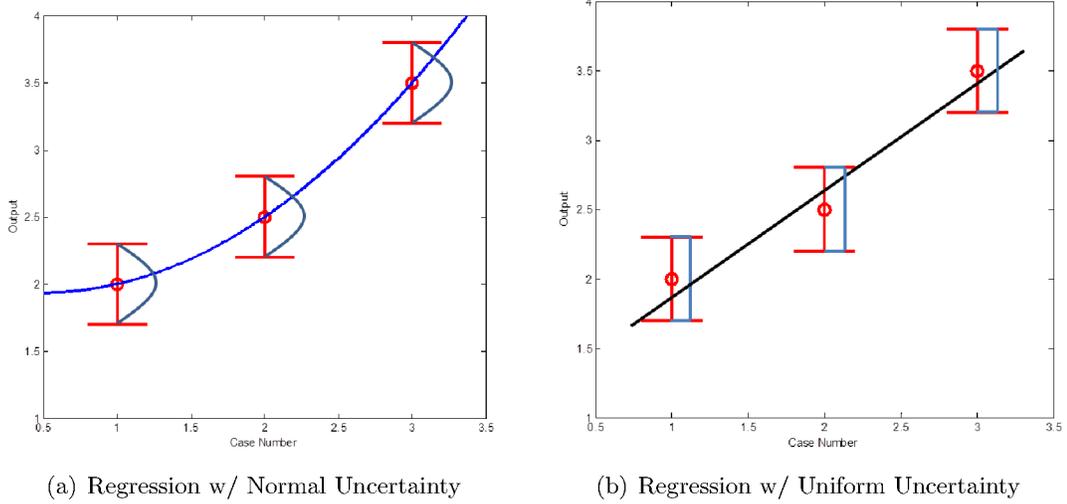


Figure 2.1: Comparison of Regressions w/ Normal & Uniform Uncertainties

There are two ways to view the measurement uncertainty in process identification: 1) loss function perspective; 2) noise distribution perspective. The common philosophy of algorithms using loss function is to define a certain form of loss function (usually monotonically increasing with respect to the absolute value of prediction error) and minimize it over model parameters. The best estimation of parameters can be obtained by minimizing the defined loss function. In this class, Ordinary Least Square (OLS) method and others of its kin are most frequently used. Ordinary Least Square method, established first in [4], is the most common regression strategy. Loss function of OLS is defined as Eqn 2.1:

$$J(\theta) = \sum_{i=1}^n (y_i - f(x_i; \theta))^2 \quad (2.1)$$

Corresponding optimal estimation is:

$$\hat{\theta} = \arg \min J(\theta) \quad (2.2)$$

where  $X = \{x_1, x_2, \dots, x_n\}$  is input data and  $Y = \{y_1, y_2, \dots, y_n\}$  is noisy output measurement.

Specifically in linear model assumption, it is:

$$J(\theta) = \sum_{i=1}^n (y_i - x_i\theta)^2 \quad (2.3)$$

The well-known explicit form of optimal solution is shown in Eqn 2.4:

$$\hat{\theta} = (X^T X)^{-1} X^T Y \quad (2.4)$$

Other identification algorithms of Least Square kin evolve by incorporation of different techniques. Weighted Least Square method is developed for varying variance of noise in which the homoskedasticity no longer holds [5]. Ridge regression can be applied to deal with the case that input data matrix is ill-conditioned or nonorthogonal [6]. For identification problem with constraints, Lagrange Multiplier is adopted to formulate objective function [7]. By employing multivariate statistics, Principle Component Regression (PCR) [8] and Partial Least Square (PLS) [9] methods are developed for dimension reduction when the number of input variables is large. Recently PCR and PLS are extensively adopted in soft sensor development which considerably facilitates the prediction and control of important process variables [10, 11, 12, 13]. To summarize, by different definitions of prediction error and assignment of corresponding penalty, the loss function approach is an intuitive way to attain the estimation of model parameters. However, Least Square based methods are all aimed at solving the normally distributed uncertainty problem since the loss function always reaches its minimum at zero prediction error hence the tight tracking of measurements is preferred.

However, in practice, there exist noise distributions that are different from the normal distribution. As an example of uniformly distributed uncertainty, quantization procedure, which maps a continuous measurement into a smaller discrete-value set [14], exists widely in industrial practice due to essential digitalization of control and measurement signals. Intuitively, aforementioned quantization procedure will introduce error to the quantized output in addition to measurement noise brought by instrumentation. It has been established in [15] that quantization error can be modelled as an additive noise to the continuous signal. Furthermore, it is zero-mean and uniformly distributed under ideal A/D converter circumstance. A comparison of continuous signal and its quantized result by an ideal mid-riser A/D converter is presented in Fig 2.2.

Accordingly, Fig 2.3 describes a typical measurement and quantization procedure. Analogue instrument is adopted to measure the objective variable and results in additional white Gaussian measurement noise. Quantization is conducted successively and based on the ad-

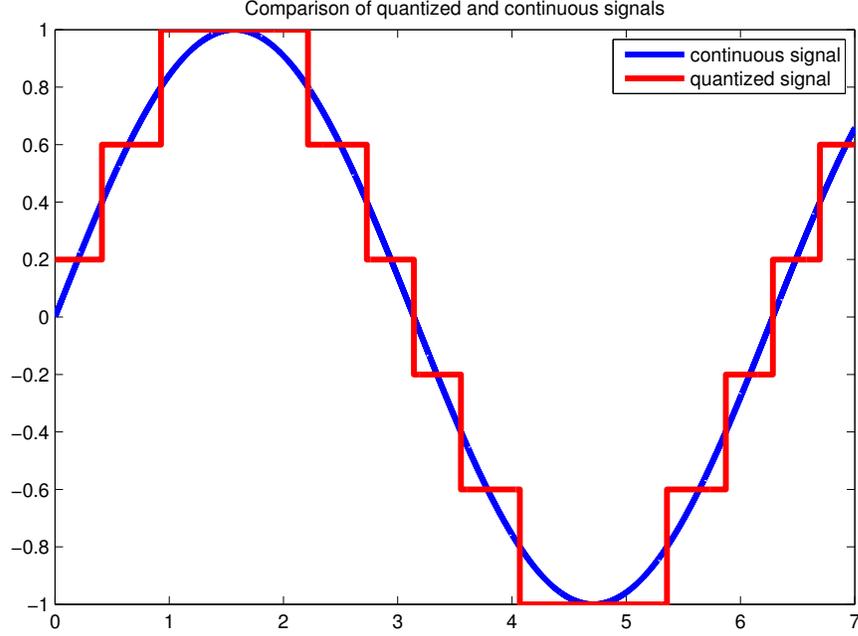


Figure 2.2: Continuous and Quantized Signals

ditive and uniform error property mentioned above, independent uniform quantization error is added to the original measurement.

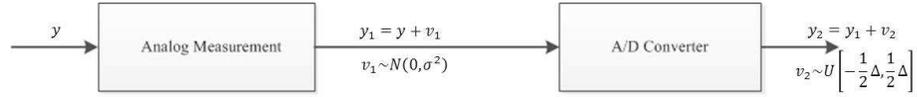


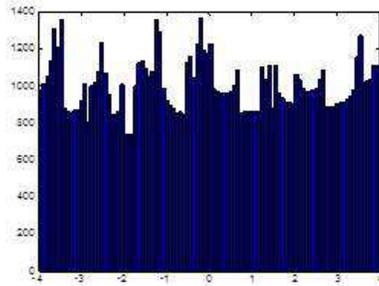
Figure 2.3: Process Diagram with Instrumentation Noise and Quantization Error

Hence entire output noise model can be formulated in Eqn 2.5. Fig 2.4 show histograms of simulated quantization error and summation of quantization error and white Gaussian noise.

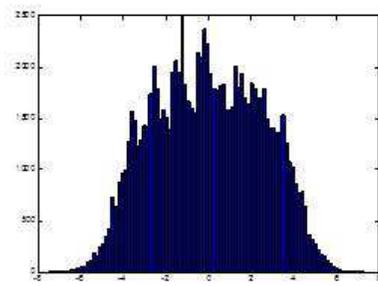
$$y_{measure} = f(x; \theta) + v_1 + v_2 \quad (2.5)$$

where  $v_1 \sim N(0, \sigma^2)$  is white Gaussian noise and  $v_2 \sim U[-\frac{\Delta}{2}, \frac{\Delta}{2}]$  is uniform quantization error.

In order to handle uniform uncertainty, various methods have been developed previously. In presence of bounded error, Support Vector Regression (SVR) [16, 17], Least Square SVR (LSSVR) [18] and related machine learning methods have been developed and applied. Similar to Least Square and other loss function methods, conceptually the objective of SVR is to minimize prediction error. However any error within the range  $[-\epsilon, \epsilon]$  is neglected



(a) Quantization



(b) Quantization Error and White Noise

Figure 2.4: Quantization Error vs Summation of Quantization Error and White Noise

and exempt from penalty. Advanced “kernel trick” [19] allows aforementioned PCA, PLS and SVM algorithms to be extended to nonlinear functional spaces. On the other hand, from noise distribution aspect, Wübbeler, Krystek and Elster proposes a numerical way of calculating the Probability Density Function (PDF) of uniform uncertainty by Monte Carlo method [20]. As an extension, this study mathematically extends the calculation of PDF and focuses on the development of analytical approximation of the noise distribution and its parameter estimation with uniform component. A flat-topped Gaussian distribution is proposed and utilized to facilitate the identification of process with uniform measurement uncertainty in this work.

The rest of this chapter is organized as follow. A revisit on existing noise distributions and Flat Gaussian profile is given in Section 2.2. Section 2.3 discusses the formulation of Flat-topped Gaussian distribution and its parameter estimation, in which moment fitting is proposed as a general strategy for density function approximation. As a robust extension of the proposed Flat-topped Gaussian distribution, Section 2.4 introduces some findings on Flat-topped  $t$  distribution. Moreover, considering practical issues, Section 2.5 introduces an iterative procedure to isolate uniform error and estimate the variance of Gaussian noise. To manifest the efficiency of proposed distribution in identification of linear processes, numerical examples are presented in Section 2.6. Key findings and of this chapter and potential future works are summarized in Section 2.7.

## 2.2 Revisit of Noise Distributions & Flat Gaussian Profile

### 2.2.1 Noise Distributions

From probability distribution point of view, measurement uncertainty is considered as an contaminating random variable added to the deterministic process output. With certain

noise distribution, objective functions and corresponding optimization problems for parameter estimation are formulated under Maximum Likelihood framework [21, 22]. Therefore, adoption of different noise distributions will lead to different objective functions and change the performance of identification algorithms. Hence intermediate objective of identification procedure is to specify the noise distribution and obtain corresponding Maximum Likelihood Estimation (MLE) in order to eliminate the influence of measurement noise on “true” output statistically.

Under noise contamination assumption, model structure is defined as Eqn 2.6:

$$y_{measure} = f(x; \theta) + v \quad (2.6)$$

where  $y_{measure}$  is measured output and  $f(x; \theta)$  is deterministic process output. In identification procedure,  $x$  is defined as regressor and  $\theta$  is model parameter to be identified.  $v$  is the additional noise term with unspecified distribution.

The objective function of the parameter estimation problem is the likelihood function of output  $Y$  given input  $X$  and parameters  $\theta$ . Assume noise distribution  $P_V(v)$  is known and independent at each sample point, the likelihood function can be expressed as:

$$\begin{aligned} L(\theta) &= P(Y|X, \theta) = \prod_{i=1}^n P(y_i|x_i, \theta) \\ &= \prod_{i=1}^n P_V(y_i - f(x_i; \theta)) \end{aligned} \quad (2.7)$$

For instance, white Gaussian noise is the most widely used assumption that yields an objective function similar to the aforementioned Ordinary Least Square solution. Following derivation illustrates the equivalence of the MLE estimation under white Gaussian noise assumption with variance  $\sigma^2$  and OLS solution of model parameters.

$$\begin{aligned} P_V(v) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{v^2}{2\sigma^2}\right) \\ \hat{\theta} &= \arg \max_{\theta} [L(\theta)] \\ &= \arg \max_{\theta} \prod_{i=1}^n P_V(y_i - f(x_i; \theta)) \\ &= \arg \max_{\theta} \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - f(x_i; \theta))^2}{2\sigma^2}\right) \\ &= \arg \min_{\theta} \sum_{i=1}^n (y_i - f(x_i; \theta))^2 \end{aligned} \quad (2.8)$$

Other noise assumptions, such as Mixture Gaussian distribution [2], student- $t$  distribution and its mixture [3], are also adopted to develop algorithms which are capable of handling specific noise issues arising in certain model identification procedure. As the noise assumption gets more sophisticated, more advanced probabilistic and statistical tools like Bayesian Inference method are applied to facilitate the Maximum Likelihood Estimation procedure. For example, a process identification approach is established by incorporating Mixture Student  $t$  distribution and Expectation Maximization (EM) algorithm in [23] in order to deal with outlier issue and parameter varying property simultaneously. In addition to the Flat-topped Gaussian distribution which is going to be considered, Mixture Gaussian approximation is also utilized in order to deal with the uniform measurement uncertainty in this chapter.

### 2.2.2 Flat Gaussian Profile

The Flat Gaussian profile was originated from optical research for description of featured light beams. Shealy and Hoffnagle [24] introduce several possible approximated expressions of this profile. Li [25] proposes a sufficient condition that approximated expression should satisfy so as to generate a flat-topped Gaussian profile with certain degree of flatness.

A variety of approximating expressions and corresponding profiles [24] with different parameter selections are shown as follow:

- Super-Gaussian Function:

$$SG(v) = \exp(-|v|^\gamma)$$

where  $\gamma \geq 2$  is even integer;

- Super-Lorentzian Profile:

$$SL(v) = \frac{1}{1 + |v|^M}$$

where  $M$  is positive integer;

- Flat-Gaussian Function:

$$FG(v) = \exp\left(-\sum_{i=1}^m \alpha_i v^{2i}\right)$$

where  $m$  is positive integer. For example, when  $m = 2$ ,  $FG(v) = \exp(-av^2 - bv^4)$ .

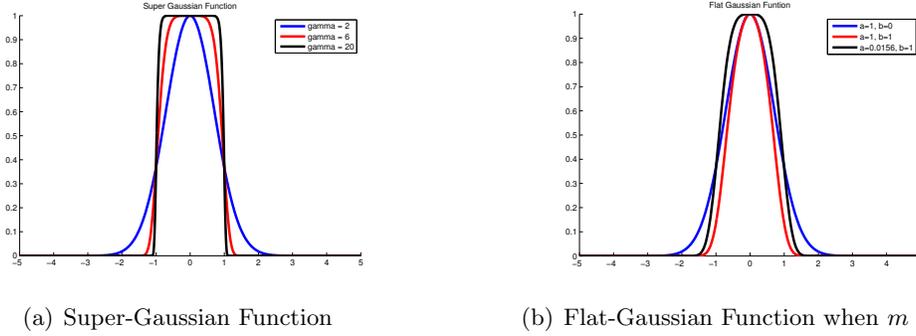


Figure 2.5: Examples of Approximated Flat Gaussian Profile

Afterwards explanation from statistical perspective is entitled to this Flat Gaussian profile [26]. It indicates that Flat Gaussian profile can be interpreted as the PDF of summation of a Gaussian distributed random variable and a uniform distributed random variable which are independent of each other. Coverage factor and other properties of this innovative distribution function are studied sequentially. As an extension, detailed reasoning and derivation of parameter estimation for the Flat-topped Gaussian distribution and its application as noise distribution will be presented later on.

## 2.3 Flat-topped Gaussian Distribution

Based on the Guide to the Expression of Uncertainty in Measurement [27], the probability density function of a summation of two random variables can be approximated by moment fitting strategy. This section combines the moment fitting strategy with approximated Flat Gaussian profile and proposes the functional structure of Flat-topped Gaussian distribution function as well as approaches to estimate its parameters.

### 2.3.1 Problem Statement

Flat Gaussian profile is used to describe a random variable that is formulated as the summation of a Gaussian random variable and a uniform distributed random variable independent of each other.

$$V = V_1 + V_2, V_1 \sim N(0, \sigma^2), V_2 \sim U[-r, r]. \quad (2.9)$$

For simplicity, assume  $E(V) = E(V_1) = E(V_2) = 0$ .

Different selections of Gaussian and uniform distribution parameters will result in diverse distribution profiles. The histograms shown in Fig 2.6 of the new random variable

illustrate the influence of parameter selection. PDFs under different parameters all have a top with different degrees of flatness in centre area while the tail demonstrates a Gaussian profile, which leads to a designation as “Flat-topped Gaussian” distribution. In order to describe properties of this new random variable and formulate objective function in Maximum Likelihood Estimation procedure correspondingly, it is necessary to obtain an expression of PDF of this random variable.

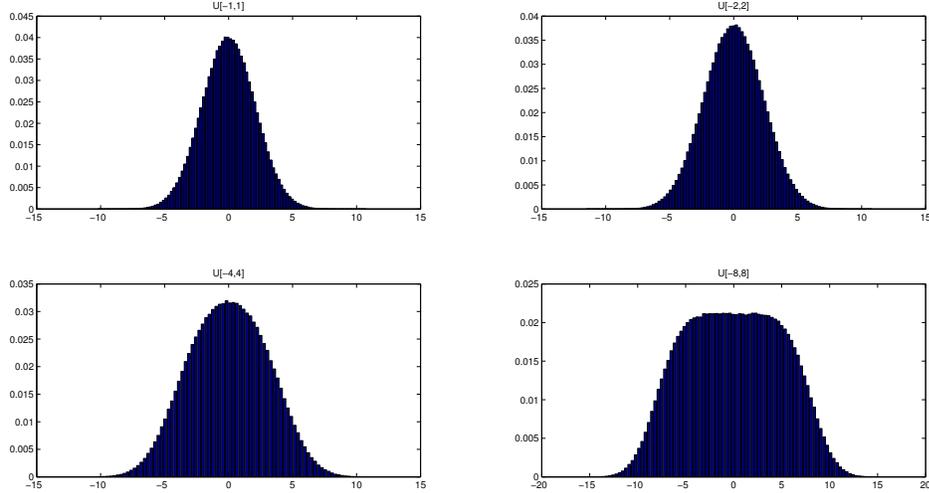


Figure 2.6: Histogram of Flat-topped Gaussian Random Variable with Different Uniform Width

Traditional approach to calculate PDF of summation of two variables such as 2.9 is convolution over two PDFs which can be justified by marginalization of joint distribution:

$$\begin{aligned}
 P_V(v) &= \int_{-\infty}^{\infty} P(v_2, v) dv_2 \\
 &= \int_{-\infty}^{\infty} P_V(v|v_2)P_{V_2}(v_2)dv_2 \\
 &= \int_{-\infty}^{\infty} P_{V_1}(v - v_2)P_{V_2}(v_2)dv_2 \\
 &= P_{V_1}(v_1) * P_{V_2}(v_2)
 \end{aligned} \tag{2.10}$$

However, it is not always tractable to calculate the convolution of two arbitrarily selected PDFs; hence approximation approach is of necessity in order to obtain the PDF of summed random variables.

### 2.3.2 Moment Fitting Method

One widely used approach for approximation of composed PDF is the moment fitting strategy [26]. With all information of the two component random variables available, this strategy will come up with an approximation function of the composed PDF with maximum entropy among all. There are two basic steps in moment fitting method:

1. Uncertainty propagation: with all parameters of component distributions known, extract different orders of moment information of the composed variable;
2. Maximum entropy distribution [28]: find the Maximum Entropy Distribution subject to moment constraints obtained in Step 1.

#### Uncertainty Propagation

**Lemma 1.** *If this relationship  $V = f(V_1, V_2)$  exists, where  $V_1$  and  $V_2$  are two independent random variables with second order moment  $M_2(V_1)$  and  $M_2(V_2)$ , then the variance of  $V$  can be calculated as:*

$$M_2(V) = \frac{\partial^2 f}{\partial V_1^2} M_2(V_1) + \frac{\partial^2 f}{\partial V_2^2} M_2(V_2) \quad (2.11)$$

Hence if  $V = V_1 + V_2$ , the second order moment of  $V$  can be calculated as:

$$M_2(V) = M_2(V_1) + M_2(V_2) \quad (2.12)$$

Lemma 1 points out that moment information of the composed random variable is able to be extracted from moment information of component random variables. Moreover, higher order moments can be calculated with respect to moments of component variables when the relationship between  $V$  and component variables  $\{V_1, V_2, \dots, V_n\}$  is linear, that is to say, if the following relationship holds:

$$V = A \cdot [V_1, V_2, \dots, V_n]^T$$

then arbitrary  $p$ -th order of moment can be calculated by Eqn 2.13 [27]:

$$M_p(V) = A^{(p)} \cdot [M_p(V_1), M_p(V_2), \dots, M_p(V_n)] \quad (2.13)$$

where  $A^{(p)}$  is defined as the  $p$ -fold Kronecker product of  $A$ :

$$A^{(p)} = \underbrace{A \otimes A \otimes \dots \otimes A}_{p \text{ times}} \quad (2.14)$$

Moment generating function can be used as an alternative to aforementioned uncertainty propagation approach for moment information extraction. In practice, application of uncertainty propagation is not limited to the summation of two random variables. However, its computational efficiency suffers dramatically from increasing number of component random variables as well as moment orders due to the Kronecker product calculations. Incorporating moment generating function makes it possible to avoid explosion in matrix dimension and therefore sets the stage for moment extraction for a number of random variables.

Moment generating function of a random variable  $V$  is defined as [29]:

$$M_V(t) = \mathbb{E}[e^{tV}], t \in \mathbb{R} \quad (2.15)$$

When this function is available,  $p$ -th order of central moment of  $V$  can be obtained:

$$\mu_p = \frac{d^{(p)}M_V}{dt^p}(0). \quad (2.16)$$

**Lemma 2.** *If  $V_1, V_2, \dots, V_k$  are a set of independent random variables and  $V = \sum_{i=1}^k a_i V_i$  is defined as their summation, then moment generating function of  $V$  is:*

$$M_V(t) = M_{V_1}(a_1 t) M_{V_2}(a_2 t) \dots M_{V_k}(a_k t) \quad (2.17)$$

However, the ascendancy between two moment extraction methods is not definitive since uncertainty propagation is also applicable in nonlinear combination of random variables while Lemma 2 is valid for linear combination only.

## Maximum Entropy Distribution

The entropy of a distribution  $P_X(x)$  of a random variable  $X$ , denoted as  $H$ , is a measurement of information contained in its distribution function [30].

$$H = E[\ln P_X(x)] = - \int P_X(x) \ln P_X(x) dx. \quad (2.18)$$

Based on first to  $m$ -th order of moment informations of  $P_Y(y)$  obtained previously by uncertainty propagation, the following constrained optimization problem can be formulated and solved to obtain an approximation for the original distribution.

$$\begin{aligned}
P_X(x) &= \max_P H \\
s.t. \int x^i P_X(x) dx &= \mu_i; i = 1, \dots, m \\
\int P_X(x) dx &= 1.
\end{aligned} \tag{2.19}$$

Fig 2.7 shows the general flow chart of moment fitting strategy. Moment information based on component variables is extracted through uncertainty propagation or moment generating function. Maximum Entropy Distribution (MED) is obtained by maximizing distribution entropy under moment constraints. It is obvious that the optimization problem mentioned above is a functional optimization problem, which cannot be solved solely by ordinary optimization approaches. The analytical representation of the optimal function should be determined antecedently to parameter estimation. The approximated expression of Flat Gaussian profile [26] is the MED under second and fourth order moment constraints [31]. Hence this moment fitting procedure is applicable for attaining an approximation of the Flat-topped Gaussian distribution.



Figure 2.7: Flow Chart of Moment Fitting Strategy

### 2.3.3 Parameter Estimation

The Flat Gaussian profile can be approximated by the following function [26]:

$$P_V(v) = A \exp(-av^2 - bv^4) \tag{2.20}$$

where  $A$ ,  $a$  and  $b$  are parameters to be determined.

To estimate parameters of this approximated distribution with known parameters of the component random variables, aforementioned moment fitting strategy provides a formulation of an optimization problem.

#### Uncertainty Propagation

The Flat-topped Gaussian distributed random variable  $V$  is defined as:

$$\begin{aligned}
V &= V_1 + V_2 = A \cdot [V_1 \ V_2]^T, \\
\text{where } V_1 &\sim N(0, \sigma^2) \\
V_2 &\sim U[-r, r] \\
A &= [1, 1]
\end{aligned} \tag{2.21}$$

Since the approximated function is the Maximum Entropy Distribution under moment constraints of second and fourth order, these two moments of  $V$  should be obtained by uncertainty propagation in advance:

$$\mu_2 = M_2(V) = M_2(V_1) + M_2(V_2) = \sigma^2 + \frac{r^2}{3} \tag{2.22}$$

$$\begin{aligned}
\mu_4 &= M_4(V) = (A \otimes A \otimes A \otimes A) \cdot [M_4(V_1) \ M_4(V_2)]^T \\
&= M_4(V_1) + 18M_2(V_1)M_2(V_2) + M_4(V_2) \\
&= 3\sigma^4 + 6\sigma^2 r^2 + \frac{1}{5}r^4
\end{aligned} \tag{2.23}$$

### Maximum Entropy Distribution

Denote  $\phi = \{A, a, b\}$  as the parameter set, the maximum entropy problem with moment constraints can be formulated in Eqn 2.24.

$$\begin{aligned}
\hat{\phi} &= \arg \max_{\phi} H \\
&= \arg \max_{\phi} \left[ - \int P_V(v) \ln P_V(v) dv \right] \\
&= \arg \max_{\phi} \left[ - \int A \exp(-av^2 - bv^4) \ln \{A \exp(-av^2 - bv^4)\} dv \right] \\
&= \arg \max_{\phi} \left[ -A \int (\ln A - av^2 - bv^4) \exp(-av^2 - bv^4) dv \right] \\
s.t. & \int v^2 A \exp(-av^2 - bv^4) dv = \mu_2; \\
& \int v^4 A \exp(-av^2 - bv^4) dv = \mu_4; \\
& \int A \exp(-av^2 - bv^4) dv = 1.
\end{aligned} \tag{2.24}$$

Lagrangian multiplier approach is applied to solve this constrained nonlinear optimiza-

tion problem with integrations. Dual problem can be formulated as follow:

$$\begin{aligned}
L(\phi, \lambda_1, \lambda_2, \lambda_3) = & A \int (\ln A - av^2 - bv^4) \exp(-av^2 - bv^4) dv + \\
& \lambda_1 \left( \int v^2 A \exp(-av^2 - bv^4) dv - \mu_2 \right) + \\
& \lambda_2 \left( \int v^4 A \exp(-av^2 - bv^4) dv - \mu_4 \right) + \\
& \lambda_3 \left( \int A \exp(-av^2 - bv^4) dv - 1 \right)
\end{aligned} \tag{2.25}$$

To optimize this dual function, following partial derivative equations should be satisfied:

$$\begin{aligned}
\frac{\partial L}{\partial \lambda_1} = 0; \quad \frac{\partial L}{\partial \lambda_2} = 0; \quad \frac{\partial L}{\partial \lambda_3} = 0. \\
\left\{ \begin{aligned} \int v^2 A \exp(-av^2 - bv^4) dv &= \mu_2 \\ \int v^4 A \exp(-av^2 - bv^4) dv &= \mu_4 \\ \int A \exp(-av^2 - bv^4) dv &= 1 \end{aligned} \right.
\end{aligned} \tag{2.26}$$

General solution of this exponential integration can be expressed with confluent hypergeometric function [32] shown in Eqn 2.27:

$$\begin{aligned}
\int_0^\infty v^{2k} A \exp(-av^2 - bv^4) dv \\
= \frac{1}{4} \left[ \Gamma\left(\frac{2k+1}{4}\right) {}_1F_1\left(\frac{2k+1}{4}; \frac{1}{2}; \frac{\alpha^2}{4}\right) - \alpha \Gamma\left(\frac{2k+3}{4}\right) {}_1F_1\left(\frac{2k+1}{3}; \frac{3}{2}; \frac{\alpha^2}{4}\right) \right]
\end{aligned} \tag{2.27}$$

where  $\alpha = ab^{-\frac{1}{2}}$  is the intermediate parameter,  ${}_1F_1\left(\frac{2k+1}{4}; \frac{1}{2}; \frac{\alpha^2}{4}\right)$  is the first order confluent hypergeometric function and  $\Gamma\left(\frac{2k+1}{4}\right)$  is Gamma function.

Previous system of nonlinear equations with integration can be further simplified as:

$$\left\{ \begin{aligned} A &= \frac{2b^{\frac{1}{4}}}{t_{11} - \alpha t_{33}} \\ \mu_2 &= \frac{t_{31} - \alpha t_{53}}{b^{\frac{1}{2}} t_{11} - \alpha t_{33}} \\ \mu_4 &= \frac{t_{51} - \alpha t_{73}}{b t_{11} - \alpha t_{33}} \end{aligned} \right. \tag{2.28}$$

where  $t_{ij} = \Gamma\left(\frac{i}{4}\right) {}_1F_1\left(\frac{i}{4}; \frac{j}{2}; \frac{\alpha^2}{4}\right)$ .

Due to nature of confluent hypergeometric function and Gamma function, there is no close form of solution of this system of nonlinear equations. Instead, It can be formulated in MATLAB with packaged functions for confluent hypergeometric and Gamma function, and corresponding solutions are obtained as parameter estimation by *fsolve* function in MATLAB, too. Approximated PDF function is formulated consequently.

As it is discussed in section 2.1, different parameter combinations of component PDFs will yield different degrees of flatness of the Flat-topped Gaussian distribution. If variance of uniform part is negligible, flatness of composed distribution will vanish and the PDF will degenerate to Gaussian distribution. The other extreme case is when  $\frac{a^2}{4b} \ll 1$ , indicating the uniform distribution is dominant in composed distribution, and the PDF can be simplified as Eqn 2.29:

$$P_V(v) \approx \frac{2b^{-\frac{1}{4}} \exp(-bv^4)}{\Gamma(\frac{1}{4})} \quad (2.29)$$

In this case, Eqn 2.30 suggested by [26] provides an approximated solution to  $b$ , which makes the maximum entropy procedure and consequential system of nonlinear equations 2.28 simpler.

$$b = (3\mu_2)^{-2} \quad (2.30)$$

### Maximum Likelihood Estimation

In section 2.3, an approximation of Flat-topped Gaussian noise distribution function is obtained, which makes it possible for Maximum Likelihood Estimation of model parameters under Flat-topped Gaussian noise assumption.

A typical output noise model can be formulated as:

$$y = f(x; \theta) + v \quad (2.31)$$

where noise is assumed to be  $v \sim FG(A, a, b)$ .  $FG$  denotes the aforementioned Flat-topped Gaussian distribution function with parameters  $A, a$  and  $b$ . Furthermore, independent and zero-mean conditions are applied. The PDF of noise  $v$  is:

$$P_V(v) = A \exp(-av^2 - bv^4) \quad (2.32)$$

Let  $X, Y = (x_1, y_1); (x_2, y_2); \dots; (x_n, y_n)$  denote the input-output pairs of measurement. With independent noise condition, likelihood function of observations can be written as Eqn 2.33:

$$\begin{aligned}
P(Y|X) &= \prod_{i=1}^n P(y_i|x_i) = \prod_{i=1}^n P_V(y_i - f(x_i; \theta)) \\
&= \prod_{i=1}^n A \exp(-a(y_i - f(x_i; \theta))^2 - b(y_i - f(x_i; \theta))^4)
\end{aligned} \tag{2.33}$$

To obtain maximum likelihood estimation of model parameter  $\theta$ , the optimization problem can be formulated as:

$$\begin{aligned}
\hat{\theta} &= \arg \max_{\theta} P(Y|X) \\
&= \arg \max_{\theta} \prod_{i=1}^n A \exp(-a(y_i - f(x_i; \theta))^2 - b(y_i - f(x_i; \theta))^4) \\
&= \arg \max_{\theta} A^n \exp(-a \sum_{i=1}^n (y_i - f(x_i; \theta))^2 - b \sum_{i=1}^n (y_i - f(x_i; \theta))^4) \\
&= \arg \min_{\theta} \sum_{i=1}^n (a(y_i - f(x_i; \theta))^2 + b(y_i - f(x_i; \theta))^4)
\end{aligned} \tag{2.34}$$

When comparing with the Least Square loss function:

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n (y_i - f(x_i; \theta))^2 \tag{2.35}$$

one can see that Flat-topped Gaussian noise distribution will yield a Least Square type of loss function with higher ordered error terms which can provide a certain range of tolerance for prediction error in identification. Similarly, coefficients and orders of error terms determine the flatness of the tolerance range of error in the loss function.

### 2.3.4 Mixture Gaussian approximation

In this section, we consider another approach to approximate Flat-topped Gaussian distribution. Gaussian Mixture Model is a probabilistic model which assumes that all samples are drawn from a mixture of Gaussian distributions with different parameters. Eqn 2.36 shows an example of a random variable  $X$  with Mixture Gaussian distribution. Eqn 2.37 is the corresponding PDF of  $X$ :

$$X \sim N(\mu_i, \sigma_i^2) \quad \text{with probability } \alpha_i, \quad i = 1, 2, \dots, M \tag{2.36}$$

$$P_X(x) = \sum_{i=1}^M \alpha_i N(\mu_i, \sigma_i^2) \tag{2.37}$$

Based on the nature of convolution, following derivation reveals the feasibility of Mixture Gaussian approximation for Flat-topped Gaussian distribution.

Given  $v = v_1 + v_2$ , where  $v_1 \sim N(0, \sigma^2)$  and  $v_2 \sim U[-r, r]$ ,  $P_V(v)$  can be calculated by convolution:

$$\begin{aligned}
P_V(v) &= P_{V_1}(v_1) * P_{V_2}(v_2) \\
&= \int_{-r}^r \frac{1}{2r} \times \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(v-v_2)^2}{2\sigma^2}\right] dv_2 \\
&\approx \frac{1}{2r} \sum_{i=-r}^r \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(v-v_2)^2}{2\sigma^2}\right] \\
&= \frac{1}{2r} \sum_{i=-r}^r N(i, \sigma^2)
\end{aligned} \tag{2.38}$$

Flat-topped Gaussian distribution, therefore, is essentially the average of infinite number of equally weighted Gaussian distributions with same variance and mean varying in the range  $[-r, r]$ . Intuitively, approximation accuracy of the Mixture Gaussian distributions depends on the number of component distributions. The following formulation describes the approximated Flat-topped Gaussian random variable under Mixture Gaussian distribution framework.

$$V \sim \begin{cases} N(-r, \sigma^2) & w.p. \frac{1}{m} \\ N\left(-r + \frac{2r}{m-1}, \sigma^2\right) & w.p. \frac{1}{m} \\ \dots & \\ N(r, \sigma^2) & w.p. \frac{1}{m} \end{cases} \tag{2.39}$$

Parameter  $m$  in this formulation represents the number of Gaussian mixture components. Here the original Flat-topped Gaussian distributed random variable is assumed to be drawn from  $m$  Gaussian distributions whose means are equally distributed in the range  $[-r, r]$  with the same probability  $\frac{1}{m}$  and same variance  $\sigma^2$ . This formulation in Eqn 2.39 is equivalent to the result obtained by convolution, shown in Eqn 2.38, when the number of components is set to be  $m$ .

Fig 2.8 demonstrates different degrees of accuracy of Mixture Gaussian approximation with original variables to be  $v_1 \sim N(0, 2^2)$  and  $v_2 \sim U[-4, 4]$ . As the number of component Gaussian distributions increases, corresponding approximation will become more precise.

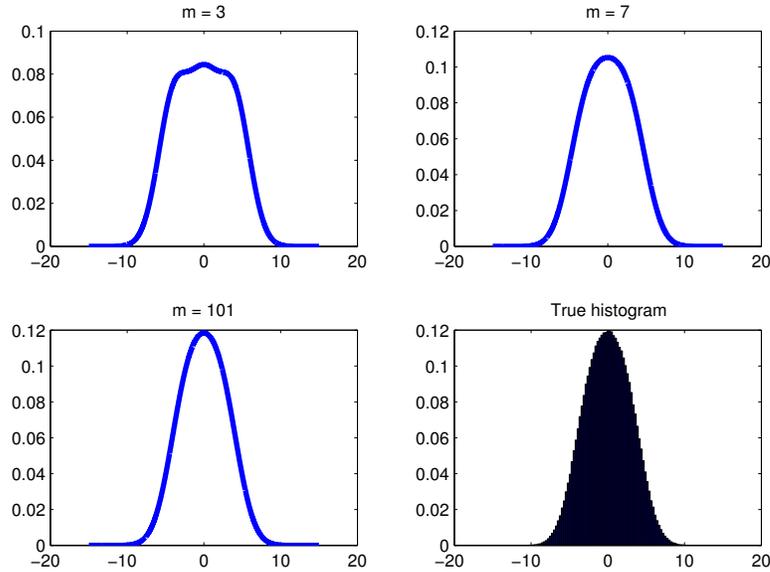


Figure 2.8: Performance of Mixture Gaussian Approximation with Differnet  $m$

## 2.4 Flat-topped $t$ Distribution

### 2.4.1 Robustness Issue in Flat-topped Gaussian Distribution

Recall the comparison of loss functions formulated by Flat-topped Gaussian and ordinary Gaussian noise distributions in Eqn 2.34 and 2.35. Flat-topped Gaussian distribution will yield higher order loss functions and therefore reduce algorithm robustness to outliers. Fig 2.9 is an illustration of the sensitivity to prediction error of different objective functions. For an error value  $v_1$  which lies in the error band  $[-\delta, \delta]$ , Flat-topped Gaussian loss function has a better tolerance and punishes less comparing with the original Gaussian one. However, for a potential outlier  $v_2$  outside the error band, Flat-topped Gaussian loss function increases rapidly and penalize significantly on the outliers.

Hence in order to enhance robustness,  $t$  distribution is considered as an alternative to the Gaussian distribution component since the longer tail in its distribution function provides better tolerance for outliers. Flat-topped  $t$  distribution, formed by summation of two random variables that follows  $t$  distribution and uniform distribution respectively, is proposed and will be studied in this section.

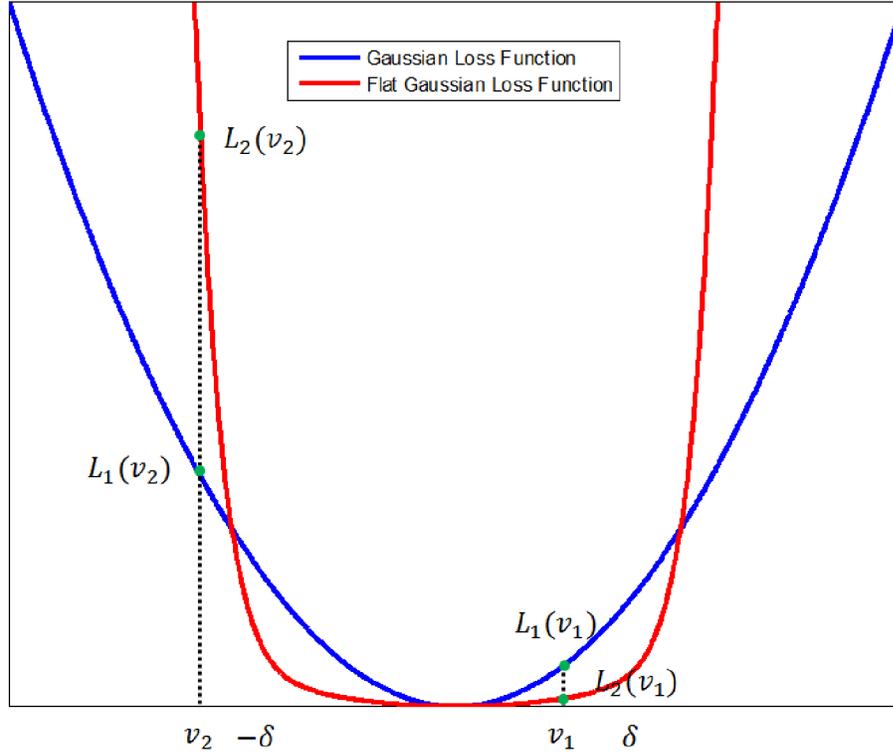


Figure 2.9: Comparison of Gaussian and Flat-topped Gaussian Loss Functions

## 2.4.2 Formulation of Flat-topped $t$ Distribution

### $t$ Distribution

The standardized Student- $t$  distribution shown below is first derived by Helmet and Lüroth [33, 34]. Gosset indicated under the pseudonym Student that the mean of a small sample set drawn from a Gaussian distributed population with unknown standard deviation follows this distribution [35] and Fisher [36] first named it as the “Student- $t$ ” distribution.

$$P_T(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\phi}\Gamma(\frac{\nu}{2})\sigma} \left( 1 + \frac{1}{\nu} \left( \frac{t-\mu}{\sigma} \right)^2 \right)^{-\frac{\nu+1}{2}} \quad (2.40)$$

where  $\mu$  and  $\sigma$  are the mean and variance of random variable  $t$  respectively.  $\nu$  is degree of freedom of this distribution which is an extra parameter comparing to Gaussian distribution that defines the profile of distribution function [37]. It holds that when  $\nu \rightarrow \infty$ ,  $t$  distribution approaches Gaussian distribution. Likelihood and corresponding loss functions under  $t$  distributed noise are derived in Eqn 2.41:

$$\begin{aligned}
\hat{\theta} &= \arg \max_{\theta} P(Y|X) \\
&= \arg \max_{\theta} \prod_{i=1}^n \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\phi}\Gamma(\frac{\nu}{2})\sigma} \left( 1 + \frac{1}{\nu} \left( \frac{y_i - f(x_i; \theta)}{\sigma} \right)^2 \right)^{-\frac{\nu+1}{2}} \\
&= \arg \max_{\theta} n \log \left( \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\phi}\Gamma(\frac{\nu}{2})\sigma} \right) - \frac{\nu+1}{2} \sum_{i=1}^n \log \left( 1 + \frac{1}{\nu} \left( \frac{y_i - f(x_i; \theta)}{\sigma} \right)^2 \right) \\
&= \arg \min_{\theta} \sum_{i=1}^n \log \left( 1 + \frac{1}{\nu} \left( \frac{y_i - f(x_i; \theta)}{\sigma} \right)^2 \right)
\end{aligned} \tag{2.41}$$

Since the natural logarithm function is monotonically increasing with respect to its argument, this loss function is magnified as squared prediction error increases. As it is proposed in Section 2.3, higher ordered error terms will result in a relatively “flat” top in neighbourhood of mean. Therefore an analogy is made to increase the order of error term in  $t$  distribution function to provide a new distribution with a tolerance range of prediction error.

### Flat-topped $t$ Distributed Random Variable

By analogy to the formulation of Flat-topped Gaussian distribution, the summation of a  $t$  distributed random variable and a uniform random variable can be defined as Flat-topped  $t$  distributed random variable:

$$V = V_1 + V_2, V_1 \sim t_p(0, \sigma^2, \nu), V_2 \sim U[-r, r]. \tag{2.42}$$

For simplicity, assume  $E(V) = E(V_1) = E(V_2) = 0$ .

Following histograms of this new random variable suggest that similar to Flat-topped Gaussian distribution, it also has a flat top and  $t$  distributed tail. Feasibility of moment fitting strategy for approximating Flat-topped  $t$  distribution function is open to discussion, since maximum entropy constraints and corresponding MED structure are yet to be established for proposed Flat-topped  $t$  random variable. However, compromised structure can be generated by analogy to Flat-topped Gaussian distribution. Similar to formulation of Flat-topped Gaussian distribution, the squared error term is substituted by a summation of second and fourth power error terms in order to yield a flat top in the area around zero; coefficients of two terms can be adjusted to accommodate different degrees of flatness.

$$P_V(v) = A \left( 1 + \frac{a}{\nu} \left( \frac{v}{\sigma} \right)^2 + \frac{b}{\nu} \left( \frac{v}{\sigma} \right)^4 \right)^{-\frac{\nu+1}{2}} \tag{2.43}$$

According to the comparison result of histogram of  $V = V_1 + V_2$  generated above, it is reasonable to claim that when coefficients are selected properly, this structure is capable of approximating the Flat  $t$  profile.

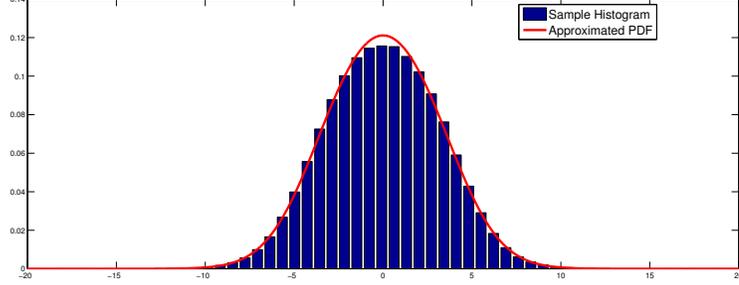


Figure 2.10: Comparison of True Histogram and Approximated Function

In Fig 2.10, the sample histogram is generated from a random variable  $V = V_1 + V_2$ , where  $V_1 \sim t_p(0, 2^2, 10)$  and  $V_2 \sim U[-4, 4]$ . Red curve is the graph of following function:

$$P(v) = 0.1211 \left( 1 + \left( \frac{v}{11.4960} \right)^2 + \left( \frac{v}{10.2490} \right)^4 \right)^{-\frac{10+\frac{1}{2}}{2}} \quad (2.44)$$

This simulation case demonstrates that if function parameters are properly selected, revised structure of  $t$  distribution with higher order terms is eligible for approximating Flat-topped  $t$  distributed random variables. However, parameter selection still remains for further investigation.

### 2.4.3 Potential Advantages of Flat-topped $t$ Distribution

Fig 2.9 demonstrates that Flat-topped Gaussian noise distribution suffers severely from outliers in samples due to increased order or error term in its loss function at the cost of enhancing its tolerance of errors within the error band. Meanwhile,  $t$  distribution possesses good robustness comparing with original Gaussian distribution. Therefore Flat-topped  $t$  distribution will effectively improve robustness of the loss function comparing with Flat-topped Gaussian distribution without loss of tolerance of errors within the error band. Fig 2.11 demonstrates the potential advantage of the loss function generated based on Flat-topped  $t$  distribution in robustness in comparison to the regular Gaussian and Flat-topped Gaussian loss functions.

The Flat-topped  $t$  distribution may also outperform piecewise linear objective functions generated by algorithms in SVR class since objective function is continuous under Flat-topped  $t$  assumption, which can facilitate subsequent optimization procedure.

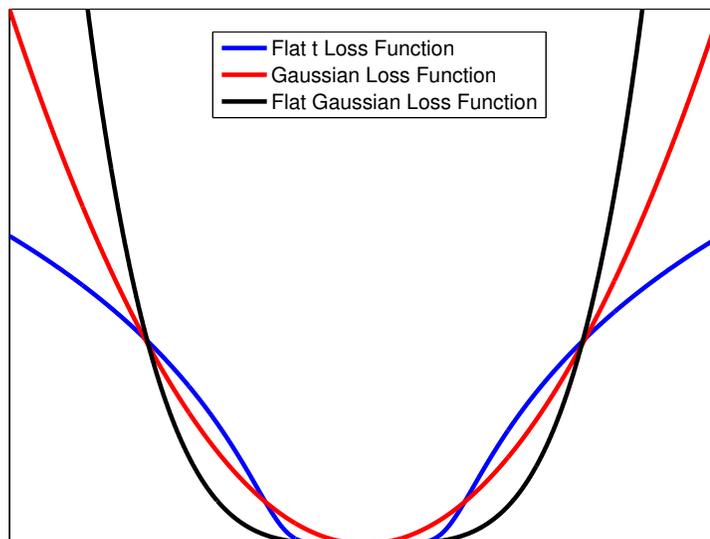


Figure 2.11: Comparison of Gaussian, Flat-topped Gaussian and Flat-topped  $t$  Loss Functions

## 2.5 Estimation of Noise Parameters

An additional demand of the proposed moment fitting strategy for PDF approximation is that it requires all parameters of component variables to be known a priori, while in practice noise level is not easy to quantify. Hence it is necessary to estimate noise level from data beforehand. To start, we would need to have the range of the uniform component. This information is normally available such as the lab analysis accuracy range is often known. In quantization procedure the resolution of quantizer is normally assumed to be fixed and known in advance as well, which can be utilized to solve this problem. The following section will give a brief discussion on estimation of measurement noise with known range of uniform distribution.

### 2.5.1 Noise Estimation for Flat-topped Gaussian Distribution

The fundamental idea of this error separation procedure is to use sample moments to approximate population moments. If instrumentation noise follows Gaussian distribution,

then the following relationship in Eqn 2.45 holds:

$$\begin{aligned} M_2(V) &= M_2(V_1) + M_2(V_2) = \sigma^2 + \frac{r^2}{3} \\ &= \frac{1}{n-1} \sum_{i=1}^n (y_i - f(x_i; \theta))^2 \end{aligned} \quad (2.45)$$

where  $2r$  is the the range of uniform distribution.  $M_2(V)$  is the variance of output uncertainty and estimated by a consistent estimator using sample variance.

Solve this equation and estimation of instrumentation noise is obtained:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - f(x_i; \theta))^2 - \frac{r^2}{3} \quad (2.46)$$

### 2.5.2 Noise Estimation for Flat-topped $t$ Distribution

For Flat-topped  $t$  distributed uncertainty, the degree of freedom of  $t$  distribution is to be estimated simultaneously with its variance. Since the number of unknown variables increases, fourth order sample moment  $M_4(V)$  is involved to establish a system of nonlinear equations with  $\sigma$  and  $\nu$  as variables:

$$\left\{ \begin{aligned} M_2(V) &= M_2(V_1) + M_2(V_2) = \frac{\nu}{\nu-2} \sigma^2 + \frac{r^2}{3} \\ &= \frac{1}{n-1} \sum_{i=1}^n (y_i - f(x_i; \theta))^2 \\ M_4(V) &= \frac{3\nu^2}{\nu^2 - 6\nu + 8} \sigma^4 + 6 \frac{\nu}{\nu-2} \sigma^2 r^2 + \frac{1}{5} r^4 \\ &= \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i; \theta))^4 \end{aligned} \right. \quad (2.47)$$

Solution of this system of equations shown in Eqn 2.48 is the parameter estimation of  $t$  distribution:

$$\left\{ \begin{aligned} \hat{\nu} &= \frac{4M_4(V) - 4r^2 M_2(V) - 6(M_2(V))^2 + \frac{6}{5} r^4}{M_4(V) - 3r^2 M_2(V) + \frac{6}{15} r^4} \\ \hat{\sigma}^2 &= \left( M_2(V) - \frac{r^2}{3} \right) \frac{\nu-2}{\nu} \end{aligned} \right. \quad (2.48)$$

It must be pointed out that the sample moments adopted here are different from moments of output measurement  $Y$ . In fact it is estimated from model residuals and highly dependent on model parameter  $\theta$ . Hence  $\theta$  should be obtained beforehand in order to calculate sample moments. Dilemma occurs that the objective of noise estimation is to

reach a better approximation of output uncertainty and better model parameter estimation  $\hat{\theta}$  afterwards, however the noise estimation procedure itself requires model parameter  $\theta$  to generate prediction error. An iterative approach which optimizes noise estimation and parameter estimation alternately is therefore needed for solving this problem.

### 2.5.3 Iterative Procedure of Noise Parameter Estimation

Given training data set  $D = \{X, Y\}$  where  $Y$  is contaminated by additive instrumentation noise and uniform error respectively and they are independent, instrumentation noise level can be separated from additional uniform error and estimated through following iterative procedure:

1. Initialization of model parameter  $\theta$ . Least Square solution  $\theta^{(0)} = \theta_{LS}$  may be applied;
2. Given  $i$ -th estimation  $\theta^{(i)}$ , obtain the error vector  $e^{(i)} = Y - f(X; \theta^{(i)})$ ;
3. Based on error vector  $e^{(i)}$ , calculate standard derivation  $\sigma^{(i)}$  of Gaussian measurement noise; for  $t$  distributed noise case, calculate  $\sigma^{(i)}$  and degree of freedom  $\nu^{(i)}$  simultaneously;
4. Apply moment fitting strategy to attain updated approximation of PDF of measurement noise and additional uniform error;
5. Adopt the approximated PDF obtained in Step 4 to build updated objective function  $J^{(i)}(\theta)$  and find  $\theta^{(i+1)} = \arg \min_{\theta} J^{(i)}(\theta)$ ;
6. Repeat Step 2 to 5 till converges.

## 2.6 Simulations & Results

In this section, qualified performance of proposed moment fitting strategy in approximation of PDF of Flat-topped Gaussian distribution will be verified by simulations. The results from white Gaussian noise assumption and Flat-topped Gaussian noise distribution are compared through estimation of a linear and an ARX model with quantization error in their output.

### 2.6.1 Parameter Estimation of Flat-topped Gaussian Distribution

As it is discussed in section 2.3, parameters of Flat-topped Gaussian distribution can be estimated by moment fitting strategy. Hence in this section, two numerical examples are

presented to qualitatively demonstrate the efficiency of proposed approximation of Flat-topped Gaussian distribution function by comparing the sample distribution histogram and PDF plot with parameters solved through moment fitting strategy.

**Component variables:**  $V_1 \sim N(0, 2^2)$ ;  $V_2 \sim U[-2\sqrt{2}, 2\sqrt{2}]$

In this case,

$$M_2(V_1) = 4; \quad M_2(V_2) = \frac{8}{3}.$$

which indicates that variances of  $V_1$  and  $V_2$  are similar in magnitude. Therefore the approximation PDF of  $V = V_1 + V_2$  maintains the form:

$$P_V(v) = A \exp(-av^2 - bv^4)$$

Parameters of approximated PDF can be estimated by solving the nonlinear system of equations, namely Eqn 2.28, proposed in Section 2.3. Following the moment fitting procedure, parameters of approximated  $P_V(v)$  can be solved as follow:

1. Uncertainty propagation: given component distribution as  $V_1 \sim N(0, 2^2)$  and  $V_2 \sim U[-2\sqrt{2}, 2\sqrt{2}]$ , corresponding moment information is extracted as Eqn 2.49.

$$\begin{aligned} M_2(V) &= M_2(V_1) + M_2(V_2) = 2^2 + \frac{(2\sqrt{2})^2}{3} = \frac{20}{3} \\ M_4(V) &= M_4(V_1) + 18M_2(V_1)M_2(V_2) + M_4(V_2) \\ &= 2^4 + 6 * 2^2(2\sqrt{2})^2 + \frac{(2\sqrt{2})^4}{5} = 124.8 \end{aligned} \quad (2.49)$$

2. Maximum entropy distribution: solve the system of nonlinear equations in Eqn 2.28 to get the maximum entropy parameter estimation.

$$\begin{cases} A = \frac{2b^{\frac{1}{4}}}{t_{11} - \alpha t_{33}} \\ \frac{20}{3} = \frac{t_{31} - \alpha t_{53}}{b^{\frac{1}{2}} t_{11} - \alpha t_{33}} \\ 124.8 = \frac{t_{51} - \alpha t_{73}}{t_{11} - \alpha t_{33}} \end{cases} \quad (2.50)$$

where  $\alpha = ab^{-\frac{1}{2}}$  is the intermediate variable. The solution to Eqn 2.50 is:

$$\begin{cases} A = 0.15012 \\ a = 0.06608 \\ b = 2.38 \times 10^{-3} \end{cases} \quad (2.51)$$

And corresponding approximated PDF of  $V$  is:

$$P_V(v) = 0.15012 \exp(-0.06608v^2 - 2.38 \times 10^{-3}v^4)$$

To qualitatively evaluate the degree of precision of the approximated PDF,  $2 \times 10^6$  random samples of  $V_1$  and  $V_2$  are generated respectively and use the histogram of their summation as sample distribution of  $V$ .

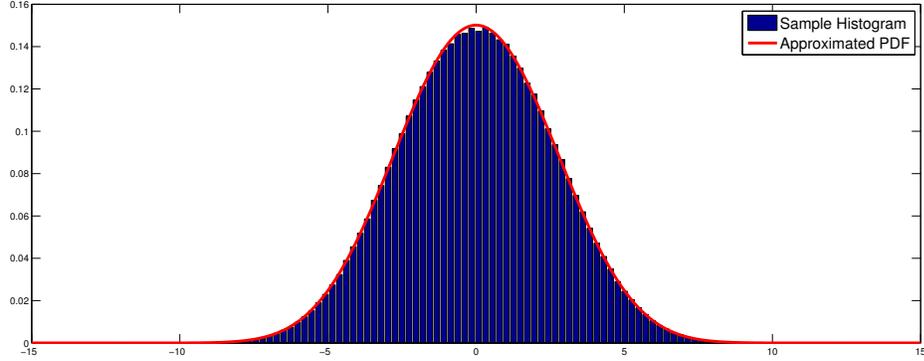


Figure 2.12: Comparison of Approximated PDF and Sample Histogram

Fig 2.12 demonstrates the comparison of sample histogram and approximated  $P_V(v)$ . It can be concluded that parameters of  $P_V(v)$  solved by moment fitting strategy provides a relatively good approximation of its real distribution function.

**Component variables:**  $V_1 \sim N(0, 2^2)$ ;  $V_2 \sim U[-6, 6]$

In this case,

$$M_2(V_1) = 4; \quad M_2(V_2) = 12. \quad (2.52)$$

which indicates that variances of  $V_2$  is dominant and approximation PDF of  $V = V_1 + V_2$  degenerates to:

$$P_V(v) = \frac{2b^{-\frac{1}{4}} \exp(-bv^4)}{\Gamma(\frac{1}{4})} \quad (2.53)$$

With the simplified form in Eqn 2.30 in effect and no need of the maximum entropy procedure, Eqn 2.54 provides the estimation of parameter  $b$ .

$$\begin{aligned} M_2(V) &= M_2(V_1) + M_2(V_2) = 2^2 + \frac{6^2}{3} = 16 \\ b &= (3M_2(V))^{-2} = 4.3403 \times 10^{-4} \end{aligned} \quad (2.54)$$

And corresponding approximated PDF of  $V$  is:

$$P_V(v) = 0.07962 \exp(-4.3403 \times 10^{-4} v^4)$$

Similarly, to qualitatively evaluate the degree of precision of the approximated PDF,  $2 \times 10^6$  random samples of  $V_1$  and  $V_2$  are generated respectively and use the histogram of their summation as sample distribution of  $V$ .

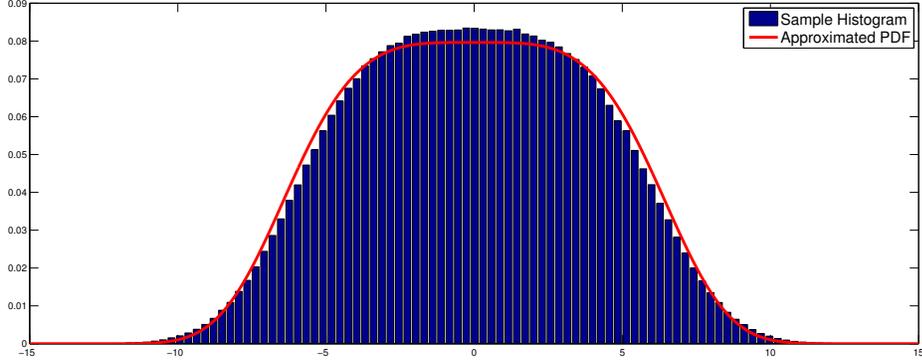


Figure 2.13: Comparison of Approximated PDF and Sample Histogram

Fig 2.13 demonstrates the comparison of sample histogram and approximated  $P_V(v)$ . It can be concluded that in addition to the situation when variance of component random variables  $V_1$  and  $V_2$  are similar in magnitude, the approximated equation is also expedient when uniform distribution is dominant, i.e. the flatness is the primary component in the composed distribution.

### 2.6.2 Simulation with Quantization Error

In this section, a numerical example on data generated by an ARX model simplified from [38] with additional quantization module is presented to validate the Maximum Likelihood Estimation based on the proposed Flat-topped Gaussian noise distribution. Eqn 2.55 is the transfer function of a first order continuous process.

$$G(s) = \frac{K}{\tau s + 1} \quad (2.55)$$

where

$$K = 1.6, \tau = 3.5. \quad (2.56)$$

Corresponding ARX model with quantization is formulated in Eqn 2.57 and Fig 2.14:

$$\begin{aligned} y(t) &= \theta_1 y(t-1) + \theta_2 u(t) + e \\ q(t) &= Q(y(t)) = \Delta \times \left( \left\lfloor \frac{y(t)}{\Delta} \right\rfloor + \frac{1}{2} \right) \end{aligned} \quad (2.57)$$

where the true value of parameters are  $\theta_1 = 0.751$  and  $\theta_2 = 0.398$ .

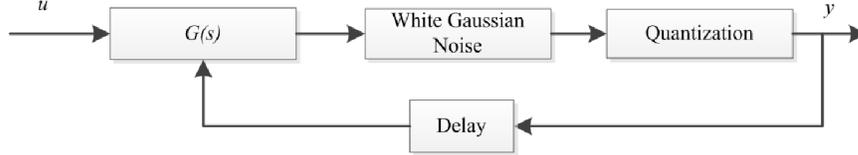


Figure 2.14: ARX Model with Quantization Module

In simulation, system input is a random binary sequence (RBS) signal generated as persistent excitation. White Gaussian noise is generated from standard Gaussian distribution  $N(0, 1)$  with zero mean and unit variance. The step size of quantization module, namely  $\Delta$  controls the result after the response is generated. As is illustrated in Fig 2.15, influence of different quantization step sizes on original signal differs significantly from each other. For example, blue curve in Fig 2.15 reflects that when the resolution is good enough, i.e. step size  $\Delta$  is small, there is barely any influence from quantization procedure on the original signal and loss of information is little. However, when the resolution is poor as the green curve, quantization procedure influences severely on the original signal and a large amount of information is lost afterwards. Quantization modules with different levels of resolution are applied to the output of this process in order to verify the proposed Flat-topped Gaussian noise in different circumstances. Relative error of parameters defined in Eqn 2.58 is adopted as an evaluation criterion.

$$RelativeError = \left| \frac{\theta_{true} - \hat{\theta}}{\theta_{true}} \right| \times 100 \quad (2.58)$$

While the variance of white Gaussian noise is fixed, a set of  $\Delta$  value is selected as  $\Delta = \{0, 2, 3, 4, 5, 6, 7, 8\}$  (here  $\Delta = 0$  stands for no quantization error) and tested in sequence to compare the performance between the proposed distribution and regular Gaussian distribution. A Monte Carlo simulation of 10 times is conducted each time when  $\Delta$  value is fixed. Table 2.1 lists all the parameter estimations and corresponding relative errors obtained simultaneously by two approaches under different  $\Delta$  values. Fig 2.16 shows the

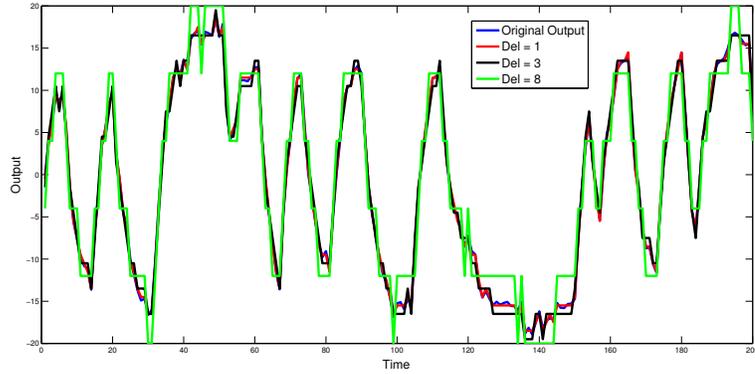


Figure 2.15: Influences on Different Quantization Resolutions

trends of relative error of model parameters trained under two noise assumptions from data as the  $\Delta$  value increases and amount of information carried by the output data diminishes. It can be seen that in the sense of learning model parameters, influence of quantization error becomes increasingly significant as step size amplifies when Gaussian noise assumption is applied. Meanwhile for the proposed Flat-topped Gaussian noise assumption, though large step size also impacts its parameter learning performance, this algorithm is relatively insensitive to quantization and consequential loss of information, and can still maintain a fair level of precision even if the quantization error is severe. On the other hand, when there is no quantization error, i.e.  $\Delta = 0$ , the proposed noise distribution is capable of resulting in an equivalent solution as the ordinary Gaussian one which is optimal in this situation. As an instance of model accuracy comparison, Fig 2.17 is the infinite step prediction comparing to the original output with no measurement noise or quantization error on cross-validation data set of the original ARX model with parameters trained when  $\Delta = 4$ .

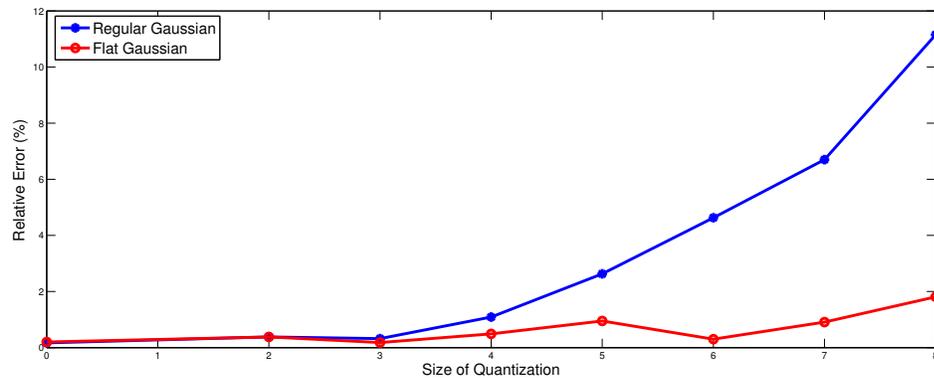


Figure 2.16: Comparison of Relative Error by Two Approaches

		Ordinary Gaussian	Flat-topped Gaussian
$\Delta = 0$	$\theta_1$	0.7492	0.7491
	$\theta_2$	0.3984	0.3986
	Relative Error (%)	0.17	0.20
$\Delta = 2$	$\theta_1$	0.7495	0.7512
	$\theta_2$	0.3958	0.3951
	Relative Error	0.38	0.38
$\Delta = 3$	$\theta_1$	0.7488	0.7516
	$\theta_2$	0.3966	0.3969
	Relative Error	0.32	0.18
$\Delta = 4$	$\theta_1$	0.7447	0.7559
	$\theta_2$	0.3927	0.3697
	Relative Error	1.09	0.49
$\Delta = 5$	$\theta_1$	0.7306	0.7536
	$\theta_2$	0.3879	0.4042
	Relative Error	2.63	0.95
$\Delta = 6$	$\theta_1$	0.7228	0.7471
	$\theta_2$	0.3761	0.3983
	Relative Error	4.63	0.30
$\Delta = 7$	$\theta_1$	0.6802	0.7632
	$\theta_2$	0.3822	0.3988
	Relative Error	6.70	0.91
$\Delta = 8$	$\theta_1$	0.6601	0.7255
	$\theta_2$	0.3574	0.3971
	Relative Error	11.15	1.81

## 2.7 Conclusion and Future Work

Key findings of this study are listed here:

- A novel Flat-topped Gaussian distribution is considered to model the output uncertainty when independent quantization error and instrumentation noise exist simultaneously.
- Moment fitting strategy is proposed as a general approach to approximate distribution functions when a close-form solution is intractable.
- The proposed approximation for Flat-topped Gaussian distribution is extended to that of  $t$ -distribution.
- An iterative procedure to extract instrumentation noise parameter is proposed in order to form a better approximation of output uncertainty and therefore improve identification accuracy when instrumentation noise level is not available in advance.

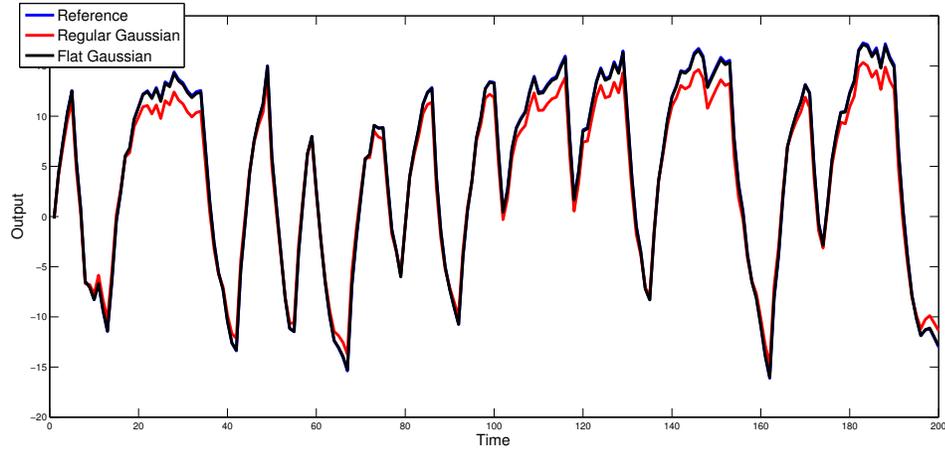


Figure 2.17: Comparison of Infinite Step Prediction by Two Approaches when  $\Delta = 6$

Several topics in this study are to be further explored. From statistical perspective, modifications of moment fitting strategy are necessary when approximating Flat-topped  $t$  distribution and generalized combination of arbitrarily distributed random variables. As for process identification, existence of quantization error or measurement inaccuracy should not be limited to output noise model. Structural change of processes, such as state space model and Gaussian Process model, brings extra difficulties to problem formulation and demands more sophisticated identification algorithms.

## Chapter 3

# Nonlinear Identification of Gaussian Process with Flat-topped Gaussian Uncertainty

In this chapter, flat-topped uncertainty is considered for identification of nonlinear model under Gaussian Process framework. The main difference between the Gaussian Process modelling and regular linear modelling approaches is that the former assumes that the output follows a multivariate Gaussian distribution rather than noise following a Gaussian distribution. When output itself is considered to be stochastic and follows a multivariate distribution, it is not applicable to formulate the likelihood function directly based on the flat-topped noise distribution as proposed previously. Due to the nature of Gaussian Process that its output is assumed to be multivariate distributed and the posterior distribution is attainable, Gibbs Sampling is utilized to mitigate the influence of additional uncertainty. A hybrid modelling approach which combines Gaussian Process and Gibbs Sampling is established in this chapter. To verify the efficiency of proposed approach, a numerical example on nonlinear system is presented.

### 3.1 Introduction

In Chapter 2, to deal with additional uniformly distributed measurement uncertainty in linear output noise model, a novel flat-topped distribution is proposed as modified noise distribution and objective function for parameter estimation is established accordingly. As it is mentioned in Section 1.6, one of the potential extensions of problem formulation is to switch model structure from linear parametric structures such as ARX models to more sophisticated non-parametric models like Gaussian Process to model nonlinear processes.

Gaussian Process is a non-parametric modelling approach which is easily implementable,

flexible in handling different types of nonlinearity and capable of providing additional uncertainty and distribution information of predictions. It has been profoundly studied by multidisciplinary researchers [39, 40, 41, 42] and intensively applied to various real-world problems such as soft sensor development for chemical plant [43], surrogate modelling optimizations [44], electrical load forecasting [45] and model predictive control [46]. From regression perspective, infinite mixture of Gaussian Process model is considered to deal with discontinuity, multimodality and other peculiarities [47]; warped Gaussian Process is proposed for non-Gaussian output and non-Gaussian noise [48]; Bayesian tree is incorporated to attain a non-stationary Gaussian Process modelling [49]. Meanwhile, it is also widely applied to solve outstanding issues in classification such as semi-supervised data set [50] and achieving better distribution-free error bounds [51]. Owing to its infrastructural flexibility, kernels and parameters of Gaussian Process can be learned and formulated by Bayesian Inference [52], spectral analysis [53] and non-stationary covariance functions can also be included [54]. However, due to its nature as a multivariate Gaussian distribution, Gaussian Process is computationally heavy which will be revealed later on. To reduce its high demand in computation and memory, sparse approximation of Gaussian Process is proposed by Quiñonero-Candela and Rasmussen [55]. Gaussian Process Latent Variable Model (GP-LVM) is also proposed for high dimensional dataset by Lawrence [56].

The very foundation of Gaussian Process modelling approach is the multivariate Gaussian distribution. However, additional uniform noise is non-Gaussian and will yield discontinuous contaminated output. Therefore, Gibbs Sampling, a Markov Chain Monte Carlo sampling approach, is incorporated to solve Gaussian Process model with additional Flat-topped Gaussian distributed noise. This approach was originated from Geman and Geman [57] and enjoyed a prosperous development afterwards. Rejection sampling in Gibbs Sampling is established by Devroye [58] and Ripley [59]. To move one step further, Gilks and Wild proposed adaptive rejection sampling in [60]. Blocked Gibbs Sampling is developed for graphical modelling by Jensen, et al [61]. Liu proposed a collapsed Gibbs Sampler with the ability of sampling from different variables simultaneously [62]. Specifically, Gibbs Sampling approach has been successfully applied previously to Gaussian Process in the sense of infinite mixtures of Gaussian Processes [63, 64], which enhances its feasibility in solving this Gaussian Process with non-Gaussian noise problem.

In the rest of this chapter, Section 3.2 presents a detailed review of methodologies involved in this problem, namely Gaussian Process and Gibbs Sampling. Section 3.3 formulates the problem to be solved and specifies the dissimilarity between aforementioned

linear processes modelling with flat-topped noise distribution and Gaussian Process modelling with additional uniform noise. A de-noising algorithm based on Gibbs Sampling is proposed in Section 3.4 and mathematical justification is provided. A numerical example is given to validate the efficiency of proposed solution in Section 3.5. Section 3.6 concludes this chapter and remarks on some future extensions of this work.

## 3.2 Review of Methodologies

### 3.2.1 Gaussian Process Modelling

#### Formulation and Structure

Gaussian Process is a non-parametric modelling and prediction approach which is based on the correlation between input measurements. Typically the relationship between input and output of a Gaussian Process is shown in Eqn 3.1. The function  $gp$  consists of two parts. Similar to the deterministic model,  $f(x)$  is a deterministic function of  $x$ ; meanwhile  $\epsilon(x)$  is considered to be a Gaussian distributed random variable with zero mean and variances related to  $x$  which is correlated between different samples.

$$y = gp(x) = f(x) + \epsilon(x) \quad (3.1)$$

Because of the correlation in output variables, there is no explicit model structure of  $gp$  function at each sample point and a multivariate Gaussian distribution is formed instead. Denote the observed data set as  $D = \{X, Y\}$  in which  $X = \{x_1, x_2, \dots, x_n\}$  is  $n \times m$  deterministic input and  $Y = \{y_1, y_2, \dots, y_n\}$  is  $n \times 1$  output measurement. The distribution of  $Y$  is assumed to be multivariate Gaussian with mean function  $f(X)$  and covariance matrix  $\Sigma$  accordingly.

$$Y \sim N(f(X), \Sigma) \quad (3.2)$$

$$\text{where } \Sigma_{ij} = \text{cov}(\epsilon(x_i), \epsilon(x_j))$$

For a better understanding, the covariance matrix  $\Sigma$  can be decomposed as  $\Sigma = \sigma^2 C$ . In this decomposed formulation,  $\sigma^2$  is the variance of marginalized distribution of each  $y_i$ , i.e. Eqn 3.3 holds:

$$\begin{aligned} P(y_i) &= \int \int \dots \int P(y_1, \dots, y_n | X, f, k) dy_1 \dots dy_{i-1} dy_{i+1} \dots dy_n \\ &= \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(y_i - f(x_i))^2}{2\sigma^2}\right) \end{aligned} \quad (3.3)$$

In this formulation, marginal distribution of  $y_i$  is determined by  $f$  and  $\sigma^2$ .  $C$  is the correlation matrix reflecting interactions of output variables with value of each element restrained between  $0 \sim 1$ .

$$\begin{aligned}
 Y &\sim N(f(X), \sigma^2 C) \\
 \text{where } C_{ij} &= \text{corr}(\epsilon(x_i), \epsilon(x_j)) \\
 &= k(x_i, x_j) \\
 \text{s.t. } C_{ij} &\in [0, 1]
 \end{aligned}
 \tag{3.4}$$

Obviously, the structure of  $f$  and  $k$  function are the two key factors in Gaussian Process. Hence the  $gp$  function can be described as:

$$gp \sim GP(f, k) \tag{3.5}$$

which means that function  $gp$  is distributed as a Gaussian Process with mean function  $f$  and covariance function  $k$  [65]. Selections of  $f$  and  $k$  characterizes the property of Gaussian Process. Possible selections of mean and covariance functions are listed but not limited to the content in Table 3.1. Modelling flexibility is significantly enhanced based on combination of different structures which enable Gaussian Process assumption to handle processes with various nonlinear features.

Table 3.1: Examples of Mean & Covariance Functions in Gaussian Process

<b>Mean Function</b>	
Name	Structure
Constant	$f(x) = c, c \in \mathbb{R}$
Linear	$f(x) = \beta^T x, \beta \in \mathbb{R}^m$
Polynomial	$f(x) = \sum_d \beta_d^T x^d, \beta \in \mathbb{R}^{m \times d}$
<b>Covariance Function</b>	
Name	Structure
Squared Exponential	$k(x_i, x_j) = \exp\left(-\frac{1}{\theta}(x_i - x_j)^T(x_i - x_j)\right)$
Periodic	$k(x_i, x_j) = \exp\left(-\frac{2}{l^2} \sin^2\left[\pi\ x_i - x_j\ /p\right]\right)$

### Training a Gaussian Process

After selection of the structures of mean and covariance functions, to train a Gaussian Process model is to obtain parameter estimation of its hyperparameters appearing in mean and covariance functions. Commonly in training regular parametric models, when no prior knowledge of parameters is available, Maximum Likelihood Estimation is obtained by Eqn 3.6:

$$\hat{\theta} = \arg \max_{\theta} P(Y|X, \theta) \quad (3.6)$$

Likely in Gaussian Process, MLE of hyperparameters can be calculated by Eqn 3.7:

$$\hat{\Phi} = \arg \max_{\Phi} P(Y|X, \Phi, f, k) \quad (3.7)$$

where  $\Phi$  denotes all the hyperparameters involved in  $f$  and  $k$ .

Due to different mathematical properties of  $f$  and  $k$  functions, this optimization problem can be solved by either analytical or numerical approaches. Take the constant mean function  $f(x) = \mu$  and squared exponential function  $k(x_i, x_j) = \exp\left(-\frac{1}{\theta}(x_i - x_j)^T(x_i - x_j)\right)$  which are the most widely used selections in practice, for instance. The distribution of  $Y$  given  $X$ ,  $f$  and  $k$  is shown in Eqn 3.8.

$$Y \sim N(\boldsymbol{\mu}, \sigma^2 C) \quad (3.8)$$

where  $C_{ij} = \exp\left(-\frac{1}{\theta}(x_i - x_j)^T(x_i - x_j)\right)$

Hyperparameters to be trained are defined as  $\Phi = \{\mu, \sigma, \theta\}$ . To get the MLE of  $\Phi$ , optimization problem in Eqn 3.9 is established:

$$\begin{aligned} \hat{\Phi} &= \arg \max_{\Phi} P(Y|X, \Phi, f, k) \\ &= \arg \max_{\Phi} L(\Phi) \\ &= \arg \max_{\Phi} \log L(\Phi) \\ &= \arg \max_{\Phi} \log \left[ \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}} \sqrt{|C|}} \exp\left(-\frac{(Y - \boldsymbol{\mu})^T C^{-1} (Y - \boldsymbol{\mu})}{2\sigma^2}\right) \right] \\ &= \arg \min_{\Phi} \left[ \frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2} \log(|C|) + \frac{(Y - \boldsymbol{\mu})^T C^{-1} (Y - \boldsymbol{\mu})}{2\sigma^2} \right] \end{aligned} \quad (3.9)$$

Notice that hyperparameter  $\mu$  and  $\sigma$  appear explicitly in the objective function while  $\theta$  lies implicitly in matrix  $C$ , taking the first order derivative of the objective function over  $\mu$  and  $\sigma$  and set them to be zero, the following MLE solutions can be obtained as:

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{\mathbf{1}^T C^{-1} Y}{\mathbf{1}^T C^{-1} \mathbf{1}} \\ \hat{\sigma}^2 &= \frac{(Y - \hat{\boldsymbol{\mu}})^T C^{-1} (Y - \hat{\boldsymbol{\mu}})}{n} \end{aligned} \quad (3.10)$$

By substituting Eqn 3.10 into the original objective function,  $\hat{\theta}$  can also be solved numerically or by optimization packages.

## Prediction with Gaussian Process

Aforementioned formulation is Gaussian Process Regression. When it comes to the prediction part, conditional distribution function is obtained given a query input  $x^*$  and training data set  $\{X, Y\}$  with parameter  $\Phi$  by Bayes' Rule:

$$P(y^*|x^*, X, Y, \Phi) = \frac{P(Y, y^*|X, x^*, \Phi)}{P(Y|X, x^*, \Phi)} \quad (3.11)$$

Based on Gaussian Process assumption, joint distribution of  $[Y, y^*]$  is  $(n + 1) \times 1$  multivariate Gaussian as the output vector is augmented. Also, Given  $X$  and  $\Phi$ ,  $Y$  is independent of  $x^*$ . Then the conditional distribution function can be expanded:

$$P(y^*|x^*, X, Y, \Phi) = \frac{1}{(2\pi\sigma^2)^{\frac{n+1}{2}} \sqrt{|\Lambda|} P(Y|X, \Phi)} \exp \left[ -\frac{1}{2} \begin{pmatrix} y^* - \mu \\ Y - \mu \end{pmatrix}^T \Lambda^{-1} \begin{pmatrix} y^* - \mu \\ Y - \mu \end{pmatrix} \right] \quad (3.12)$$

$$P(Y|X, \Phi) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}} \sqrt{|C|}} \exp \left[ -\frac{(Y - \mu)^T C^{-1} (Y - \mu)}{2\sigma^2} \right] \quad (3.13)$$

Hence the conditional distribution of  $y^*$  is:

$$P(y^*|x^*, X, Y, \Phi) = (2\pi)^{-\frac{1}{2}} \sigma^n \left( \frac{|C|}{|\Lambda|} \right)^{-\frac{1}{2}} \exp \left[ \frac{(Y - \mu)^T C^{-1} (Y - \mu)}{2\sigma^2} - \frac{1}{2} \begin{pmatrix} y^* - \mu \\ Y - \mu \end{pmatrix}^T \Lambda^{-1} \begin{pmatrix} y^* - \mu \\ Y - \mu \end{pmatrix} \right] \quad (3.14)$$

$\Lambda$  is the covariance matrix of  $[Y, y^*]$  calculated with respect to  $\theta$  and Eqn 3.15 holds:

$$\Lambda = \sigma^2 \begin{bmatrix} C & r^T \\ r & 1 \end{bmatrix} \quad (3.15)$$

$$\text{where } r = \sigma^2 [k(x_1, x^*), k(x_2, x^*), \dots, k(x_n, x^*)]$$

By applying partitioned matrix inversion method,  $\Lambda^{-1}$  can be obtained. Notice that since both joint distribution of  $[Y, y^*]$  and prior distribution of  $Y$  are Gaussian, conditional distribution of  $y^*$  is also Gaussian. Hence it is possible for Eqn 3.14 to be written in a univariate Gaussian distribution function form after proper transformation.

$$P(y^*|x^*, X, Y, \Phi) = c_0 \exp \left[ -\frac{(y^* - \mu - r^T C^{-1} (Y - \mu))^2}{\sigma^2 (1 - r^T C^{-1} r)} \right] \quad (3.16)$$

where  $c_0$  is the normalization constant of a distribution function. Usually the posterior mean  $\hat{p}$  is considered to be the point prediction value of  $y^*$  given training set  $\{X, Y\}$  and

query point  $x^*$ . However, this procedure can also indicate the uncertainty of prediction by predicted variance  $\hat{s}^2$ .

$$\hat{p} = \mu + r^T C^{-1} (Y - \mu) \quad (3.17)$$

$$\hat{s}^2 = \sigma^2 (1 - r^T C^{-1} r) \quad (3.18)$$

### 3.2.2 Gibbs Sampling

Gibbs Sampling is a Markov Chain Monte Carlo approach of sampling from multivariate distributions. Generally speaking, it is not always easy to sample directly from a multivariate distribution especially when the dimension is large and/or distribution function is complicated and alternative sampling methods are necessary. For example, suppose there are  $n$  random variables  $x_1, x_2, \dots, x_n$  with joint distribution function  $P(x_1, x_2, \dots, x_n)$ . When it is expensive or intractable to sample directly from this joint distribution function, Gibbs Sampling is considered as an alternative when the conditional distribution  $P(x_i | x_{\setminus i})$  can be easily obtained and sampled in which  $x_{\setminus i}$  denotes all variables excluding  $x_i$  for arbitrary  $i \in \{1, \dots, n\}$ . It consists of two parts including sampling and updating, and is conducted iteratively with respect to the conditional distributions of all  $x_i$ .

Basic Gibbs Sampling procedure can be described as follows:

1. Initialization: assign initial values to  $X^{(0)} = \{x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}\}$ ;
2. For the first iteration:
  - Sample  $x_1^{(1)}$  from its conditional distribution  $P(x_1 | x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)})$ ;
  - Update current data set to be  $\{x_1^{(1)}, x_2^{(0)}, \dots, x_n^{(0)}\}$ ;
  - Sample  $x_2^{(1)}$  from its conditional distribution with updated value  $P(x_2 | x_1^{(1)}, x_3^{(0)}, \dots, x_n^{(0)})$ ;
  - Update current data set:  $\{x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(0)}\}$ ;
  - ...
  - Sample  $x_i^{(1)}$  from its conditional distribution with updated value  $P(x_i | x_1^{(1)}, x_2^{(1)}, \dots, x_{i-1}^{(1)}, x_{i+1}^{(0)}, \dots, x_n^{(0)})$ ;
  - Update current data set:  $\{x_1^{(1)}, x_2^{(1)}, \dots, x_i^{(1)}, x_{i+1}^{(0)}, \dots, x_n^{(0)}\}$ ;
  - ...

- Sample  $x_n^{(1)}$  from its conditional distribution with updated value  $P(x_n|x_1^{(1)}, x_2^{(1)}, \dots, x_{n-1}^{(1)})$ ;
  - Update current data set:  $\{x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}\}$ ;
3. In  $t$ -th iteration, obtain samples  $\{x_1^{(t)}, x_2^{(t)}, \dots, x_n^{(t)}\}$ . Continue till converges.

The evidence of convergence in Gibbs Sampling is roughly defined as: after  $t$  iterations, if the conditional distribution  $P(x_i|x_{\setminus i})$  remains unchanged, then it can be claimed that all the samples of  $x_i$  obtained afterwards are sampled from the true marginal distribution  $P(x_i)$ . However, it is difficult to verify if this condition is satisfied during sampling procedure. It is pointed out by Casella and George that the distribution of  $x_i^{(t)}$  converges to  $P(x_i)$  as  $t \rightarrow \infty$  [66]. Hence the maximum number of iterations is usually set to be reasonably large in order to achieve better performance.

Typical steps of Gibbs Sampling are presented in Fig 3.1. It is obvious that the essential knowledge in Gibbs Sampling is the exact formulation of conditional distribution  $P(x_i|x_{\setminus i})$ . With such information available, instant sampling from a multivariate distribution is converted to an iterative procedure with rotative sampling from  $n$  univariate conditional distributions. When the stopping criterion is reached, samples at last few iterations are accepted as the “true” samples generated from the original multivariate distribution.

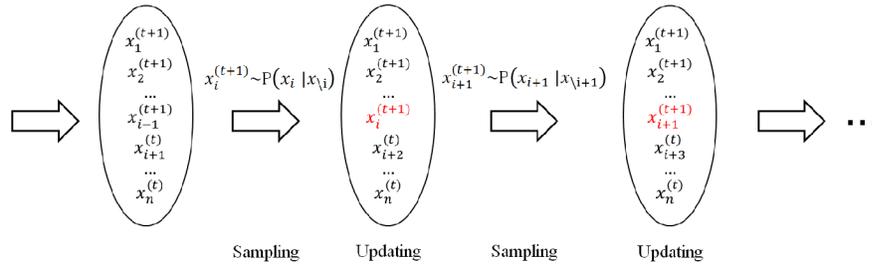


Figure 3.1: Illustration of Gibbs Sampling Procedure

### 3.3 Problem Formulation

Fig 3.2 shows the diagram of a process with additional uniform noise under Gaussian Process assumption.

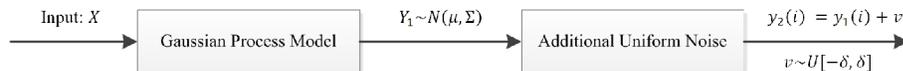


Figure 3.2: Process Diagram under Gaussian Process Framework

In this formulation,  $Y_1 = \{y_1(1), y_1(2), \dots, y_1(n)\}$  in Eqn 3.19 denotes the output of Gaussian Process which is  $n$  dimensional Gaussian distributed. The mean function is assumed to be constant and correlation function is Squared Exponential. Corresponding parameters of this distribution are determined by input  $X$  and hyperparameters  $\Phi_1 = \{\mu, \sigma^2, \theta\}$ . As it is introduced in Section 3.1, the objective of training Gaussian Process is to obtain the Maximum Likelihood Estimation of its hyperparameters  $\Phi_1$ .

$$Y_1 \sim N(\mu, \sigma^2 C), \quad (3.19)$$

where  $C(i, j) = \exp(-\frac{1}{\theta}(x_i - x_j)^2)$ .

However, direct measurement of  $Y_1$  is unavailable due to process mechanism and  $Y_2$  is attained instead with additional independently distributed uniform noise. Eqn 3.20 shows the expression of each  $Y_2$  measurement at  $i$ -th sample point.

$$y_2(i) = y_1(i) + v \quad (3.20)$$

where  $v \sim U[-\delta, \delta]$

Recall the description of quantization procedure in Section 2.1, it will provide discretized output and result in a conflict with the fundamental assumption of Gaussian Process that similar inputs yield similar but different outputs. Fig 3.3 shows the Squared Exponential correlation between two input variables  $x_i$  and  $x_j$ . It can be inferred that outputs with respect to different input variables are supposed to be different regardless of the selection of  $\theta$ . Therefore, the accuracy of hyperparameter learning deteriorates as the resolution of quantization decreases when quantized data is involved indifferently in training a Gaussian Process model. Intrinsic inferiority of contaminated data with additional uniform noise in training and prediction of Gaussian Process will be demonstrated by numerical examples afterwards.

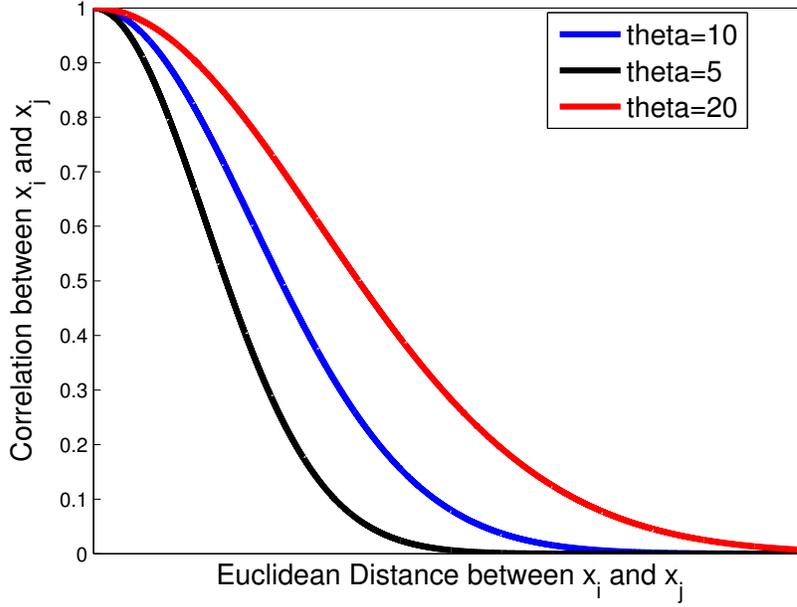


Figure 3.3: Descending Trend of Correlation between  $x_i$  and  $x_j$

More rigorous way to calculate the MLE of  $\Phi_1$  is shown in Eqn 3.21 given all the information available ( $X$  and  $Y_2$ ) instead of training Gaussian Process directly with inaccurate measurement:

$$\begin{aligned}
\hat{\Phi}_1 &= \arg \max_{\Phi_1} P(Y_2|X, \Phi_1) \\
&= \arg \max_{\Phi} \int P(Y_2, Y_1|X, \Phi_1) dY_1 \\
&= \arg \max_{\Phi} \int P(Y_2|Y_1, X, \Phi_1) P(Y_1|X, \Phi_1) dY_1 \\
&= \arg \max_{\Phi} \int \prod_{i=1}^n [P(y_2(i)|y_1(i))] P(Y_1|X, \Phi_1) dY_1
\end{aligned} \tag{3.21}$$

Unlike the independent output model assumption under which  $P(Y_1|X, \Phi_1)$  can also be separated into multiplication of individual samples, here the likelihood of  $Y_1$  is inseparable since all the outputs are considered correlated with each other. So the additional uniform noise cannot be consolidated with original distribution function of  $y_1(i)$  to make it Flat-topped Gaussian distributed. Influenced by uniform distribution, Eqn 3.21 becomes an  $n$  dimensional integration of truncated Gaussian distribution which is intractable. As a consequence,  $Y_1$  should be recovered as intermediate information before training Gaussian Process and prediction with it. From informatics point of view, it is a de-noising procedure that can reconstruct the original output signal  $Y_1$  of Gaussian Process from measurements

with additional uniform noise  $Y_2$  for the purpose of attaining better-trained hyperparameters for Gaussian Process. Fig 3.4 sketches the objective of proposed algorithm so as to solve this problem.

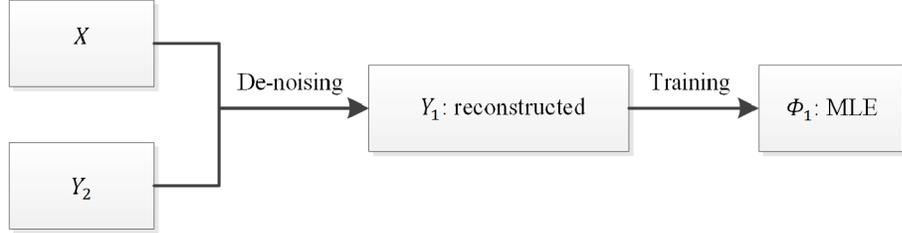


Figure 3.4: Illustration of Solving Gaussian Process Model with Additional Noise

For the de-noising stage, Gibbs Sampling can be employed since it is relatively easy to obtain posterior distribution of individual output in Gaussian Process and its applicability has been shown by literatures. Detailed sampling procedure and summary of algorithm will be discussed in next section.

### 3.4 Data Reconstruction: Gibbs Sampling Approach

#### 3.4.1 Mixture Gaussian Approximation

Generally in Gibbs Sampling approach, the essential step is to sample a value of  $i$ -th sample point  $y_1(i)$  from its posterior distribution  $P(y_1(i)|y_1(1), \dots, y_1(i-1), y_1(i+1), \dots, y_1(n), y_2(i))$ . Therefore, based on Eqn 3.22 and definition of Flat-topped Gaussian distributed random variable, it can be inferred that  $y_2(i)$  is Flat-topped Gaussian distributed with unknown parameters.

$$\begin{aligned}
 y_2(i) &= y_1(i) + v, \\
 \text{where } y_1(i)|y_1(1), \dots, y_1(i-1), y_1(i+1), \dots, y_1(n) &\sim N(p(i), s^2(i)) \quad (3.22) \\
 v &\sim U[-\delta, \delta]
 \end{aligned}$$

Due to the characteristic of sampling procedure, it is easier to sample from Gaussian distribution rather than non-Gaussian ones. In Chapter 2, it is demonstrated that in a composed Flat-topped Gaussian random variable, when the variance of uniform part is dominant, which provides the possibility to use Flat-topped Gaussian distribution to approximate uniform distribution. On the other hand, two ways of representing Flat-topped Gaussian distribution are proposed and justified simultaneously: 1) analytical approximation; 2) Mixture Gaussian approximation. Intuitively, the Mixture Gaussian approximation

is superior to the analytical approximation when to sample from. Eqn 3.23 presents the Mixture Gaussian approximation of noise distribution  $v$  which facilitates the upcoming sampling procedure. Fig 3.5 is an illustrative example of approximating uniform noise by Mixture Gaussian distribution.

$$v \sim \begin{cases} N(-\delta, \sigma^2) & w.p. \frac{1}{m} \\ N\left(-\delta + \frac{2\delta}{m-1}, \sigma^2\right) & w.p. \frac{1}{m} \\ \dots & \\ N(\delta, \sigma^2) & w.p. \frac{1}{m} \end{cases} \quad (3.23)$$

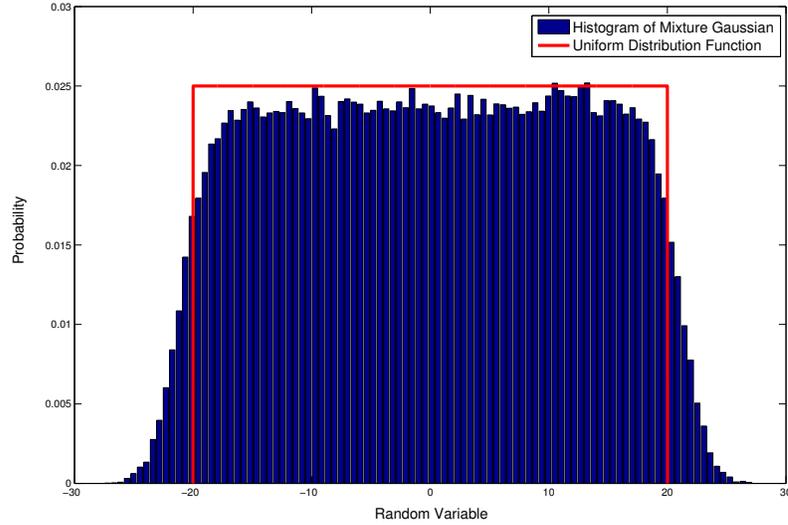


Figure 3.5: Comparison of  $v \sim U[-20, 20]$  and Corresponding Gaussian Mixture with  $\sigma^2 = 4, m = 20$

Therefore, the original uniform uncertainty is approximated by proposed Flat-topped Gaussian uncertainty, which is easier to be handle mathematically. If assuming the quantization resolution or the trust range of measurement  $\delta$  and number of Gaussian Mixture components  $m$  are both known, then the only extra parameter introduced to be estimated is the variance  $\sigma^2$  of each Gaussian component. It is obvious that as  $\sigma^2$  decreases, the Flat-topped Gaussian noise will approach a uniform one. Hence the trick in obtaining a relatively small estimation of  $\sigma^2$  will be covered afterwards since the estimation accuracy of  $\sigma^2$  is insignificant.

### 3.4.2 Posterior Distribution

Here to distinguish the “real” output and reconstructed ones, sampled value of  $y_1(i)$  is denoted as  $Y_1^{(i)}$ . In each and every iteration of Gibbs Sampling, the intermediate objective before sampling is to obtain the posterior distribution of  $Y_1^{(i)}$  in  $Y_1$  given all the rest  $Y_1^{(\setminus i)} = \{Y_1^{(1)}, \dots, Y_1^{(i-1)}, Y_1^{(i+1)}, \dots, Y_1^{(n)}\}$ . Under Gaussian Process assumption, posterior distribution of  $Y_1^{(i)}$  can be derived with respect to Bayes’ Rule. Moreover, the denominator  $P(Y_2|Y_1^{(\setminus i)}, X)$  is a normalization constant which can be omitted since in each sampling step,  $Y_1^{(\setminus i)}$  is sampled in previous steps and assumed to be known.

$$\begin{aligned} P(Y_1^{(i)}|Y_1^{(\setminus i)}, X, Y_2) &= \frac{P(Y_2|Y_1^{(i)}, Y_1^{(\setminus i)}, X)P(Y_1^{(i)}|Y_1^{(\setminus i)}, X)}{P(Y_2|Y_1^{(\setminus i)}, X)} \\ &\propto P(Y_2|Y_1^{(i)}, Y_1^{(\setminus i)}, X)P(Y_1^{(i)}|Y_1^{(\setminus i)}, X) \end{aligned} \quad (3.24)$$

Notice that the likelihood function of  $Y_2$  can be decomposed by chain rule, then Eqn 3.25 holds that:

$$P(Y_1^{(i)}|Y_1^{(\setminus i)}, X, Y_2) \propto P(Y_1^{(i)}|Y_1^{(\setminus i)}, X)P(Y_2^{(i)}|Y_2^{(\setminus i)}, Y_1^{(i)}, Y_1^{(\setminus i)}, X)P(Y_2^{(\setminus i)}|Y_1^{(i)}, Y_1^{(\setminus i)}, X) \quad (3.25)$$

where  $Y_2^{(i)}|Y_2^{(\setminus i)}, Y_1^{(i)}, Y_1^{(\setminus i)}$  is independent of  $X$  and  $Y_2^{(\setminus i)}|Y_1^{(\setminus i)}$  is independent of  $X$  and  $Y_1^{(i)}$ . Therefore Eqn 3.25 can be further simplified:

$$P(Y_1^{(i)}|Y_1^{(\setminus i)}, X, Y_2) \propto P(Y_1^{(i)}|Y_1^{(\setminus i)}, X)P(Y_2^{(i)}|Y_1^{(i)}, \{Y_2^{(\setminus i)}, Y_1^{(\setminus i)}\}) \quad (3.26)$$

In this formulation,  $P(Y_1^{(i)}|Y_1^{(\setminus i)}, X)$  is the predicted distribution obtained from Gaussian Process. Hyperparameter estimation  $\hat{\Phi}_1 = \{\hat{\mu}_0, \hat{\sigma}_0^2, \hat{\theta}\}$  in Gaussian Process is obtained by MLE with  $\{X, Y_1^{(\setminus i)}\}$  as training set:

$$\begin{aligned} \hat{\mu} &= \frac{\mathbf{1}^T \hat{C}^{-1} Y_1^{(\setminus i)}}{\mathbf{1}^T \hat{C}^{-1} \mathbf{1}} \\ \hat{\sigma}_0^2 &= \frac{\left(Y_1^{(\setminus i)} - \hat{\mu}\right)^T \hat{C}^{-1} \left(Y_1^{(\setminus i)} - \hat{\mu}\right)}{n-1} \end{aligned} \quad (3.27)$$

$$\text{where } \hat{C}(k, l) = c(x_k, x_l) = \exp\left(\frac{\|x_k - x_l\|_2}{\hat{\theta}}\right)$$

And corresponding posterior distribution of  $Y_1^{(i)}$  given  $Y_1^{(\setminus i)}$  and  $X$  is also Gaussian distributed with mean and variance:

$$\begin{aligned}\hat{p} &= \hat{\mu} + r_i^T \hat{C}^{-1} \left( Y_1^{(\setminus i)} - \hat{\mu} \right) \\ \hat{s}^2 &= \hat{\sigma}_0^2 \left( 1 - r_i^T \hat{C}^{-1} r_i \right)\end{aligned}\tag{3.28}$$

where  $r_i = [c(x_1, x_i), \dots, c(x_{i-1}, x_i), c(x_{i+1}, x_i), \dots, c(x_n, x_i)]$ .

One of the shortcomings of aforementioned parameters of posterior distribution in Gibbs Sampling is that the predicted variance  $\hat{s}^2$  according to Eqn 3.28 will not approach zero as the number of iterations increases and corresponding sampled value  $Y_{(i)}$  will remain strongly stochastic. As long as the sampling procedure is efficient, the improvement of reconstructed  $Y_1^{(\setminus i)}$  can be anticipated so that the predicted mean value  $Y_1^{(i)}$  becomes more trustworthy. Therefore an exponential decay factor is introduced in order to compel the predicted variance to decrease to a preset small value as the number of iterations increases.

$$\hat{s}_{up}^2 = \exp\left(-\frac{t}{b}\right) \hat{s}^2 + \left(1 - \exp\left(-\frac{t}{b}\right)\right) \epsilon\tag{3.29}$$

where  $t$  is the number of iteration,  $b$  is a tuning parameter which controls the decay rate and  $\epsilon$  is a small positive value comparing with the magnitude of  $\hat{s}^2$ . With incorporating this exponential decayed variance  $\hat{s}_{up}^2$ , the convergence rate of Gibbs Sampling can be improved.

As for  $Y_2^{(i)}$ , when the Mixture Gaussian approximation of  $v$  holds, it is also considered to be Gaussian Mixture distributed given  $Y_1^{(i)}$  and distribution parameters of  $v$  shown in Eqn 3.30:

$$P(Y_2^{(i)} | Y_1^{(i)}, \sigma^2) = \frac{1}{m} \sum_{j=1}^m \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y_2^{(i)} - Y_1^{(i)} - \mu_j)^2}{2\sigma^2}\right)\tag{3.30}$$

The only parameter to be estimated of  $v$  is its variance  $\sigma^2$  of each Gaussian component. It can be obtained by calculating the sample variance of data set  $\{Y_1^{(\setminus i)}, Y_2^{(\setminus i)}\}$ :

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k \neq i} (Y_1^{(k)} - Y_2^{(k)} - \mu^{(k)})^2\tag{3.31}$$

where  $\mu^{(k)} \in \{\mu_1, \mu_2, \dots, \mu_m\}$  is considered to be the ‘‘true’’ mean value of  $v$  at  $k$ -th sample. Since the accuracy of  $\sigma^2$  is insignificant, the following trick in Eqn 3.32 is applied in order to reduce the magnitude of  $\hat{\sigma}^2$ .

$$\begin{aligned}\mu^{(k)} &= \arg \min_{\mu} (Y_1^{(k)} - Y_2^{(k)} - \mu)^2 \\ \text{where } \mu &\in \{\mu_1, \mu_2, \dots, \mu_m\}\end{aligned}\tag{3.32}$$

Hence posterior distribution of  $Y_1^{(i)}$  can be further reformed as:

$$\begin{aligned}
P(Y_1^{(i)}|Y_1^{(\setminus i)}, X, Y_2) &\propto N(\hat{f}, \hat{s}_{up}^2) \times \frac{1}{m} \sum_{j=1}^m \frac{1}{\sqrt{2\pi\hat{\sigma}}} \exp\left(-\frac{(Y_2^{(i)} - Y_1^{(i)} - \mu_j)^2}{2\hat{\sigma}^2}\right) \\
&\propto N(\hat{f}, \hat{s}_{up}^2) \sum_{j=1}^m \frac{1}{\sqrt{2\pi\hat{\sigma}}} \exp\left(-\frac{(Y_2^{(i)} - Y_1^{(i)} - \mu_j)^2}{2\hat{\sigma}^2}\right)
\end{aligned} \tag{3.33}$$

To sample from this distribution, notice this original distribution of  $Y_1^{(i)}$  can be taken as a weighted Gaussian Mixture if the predicted continuous distribution  $N(\hat{f}, \hat{s}_{up}^2)$  is simplified to be a series of Dirac Delta function with a weighted value at  $Y_2^{(i)} - \mu_j$  shown in Eqn 3.34:

$$P(Y_1^{(i)}|Y_1^{(\setminus i)}, X, Y_2) \approx \alpha \delta(Y_1^{(i)} - Y_2^{(i)} + \mu_j) N(\hat{f}, \hat{s}_{up}^2) \sum_{j=1}^m \frac{1}{\sqrt{2\pi\hat{\sigma}}} \exp\left(-\frac{(Y_1^{(i)} - Y_2^{(i)} + \mu_j)^2}{2\hat{\sigma}^2}\right) \tag{3.34}$$

$$Y_1^{(i)}|Y_1^{(\setminus i)}, X, Y_2 \sim \begin{cases} N(Y_2^{(i)} - \delta, \hat{\sigma}^2) & w.p. p_1 \\ N\left(Y_2^{(i)} - \delta + \frac{2\delta}{m-1}, \hat{\sigma}^2\right) & w.p. p_2 \\ \dots & \\ N(Y_2^{(i)} + \delta, \hat{\sigma}^2) & w.p. p_m \end{cases} \tag{3.35}$$

And the weight  $p_j$  of each model  $j$  is assumed to be the probability of its centre  $Y_2^{(i)} - \mu_j$  with respect to the predicted distribution of  $P(Y_1^{(i)}|Y_1^{(\setminus i)}, X)$ :

$$p_j = \frac{P(Y_1^{(i)} = \mu_j|Y_1^{(\setminus i)}, X)}{\sum_{j=1}^m P(Y_1^{(i)} = \mu_j|Y_1^{(\setminus i)}, X)} \propto \frac{1}{\sqrt{2\pi\hat{s}}} \exp\left(-\frac{(\mu_j - \hat{f})^2}{2\hat{s}_{up}^2}\right) \tag{3.36}$$

Fig 3.6 sketches the relationship of posterior distribution of  $Y_1^{(i)}$  given  $Y_1^{(\setminus i)}, X$  and the distribution of  $Y_1^{(i)}$  given  $Y_2^{(i)}$  under Mixture Gaussian assumption (magnitude of probability may be inconsistent). When sampling from the distribution of  $Y_1^{(i)}|Y_1^{(\setminus i)}, X, Y_2$ , both of them should be taken into consideration.

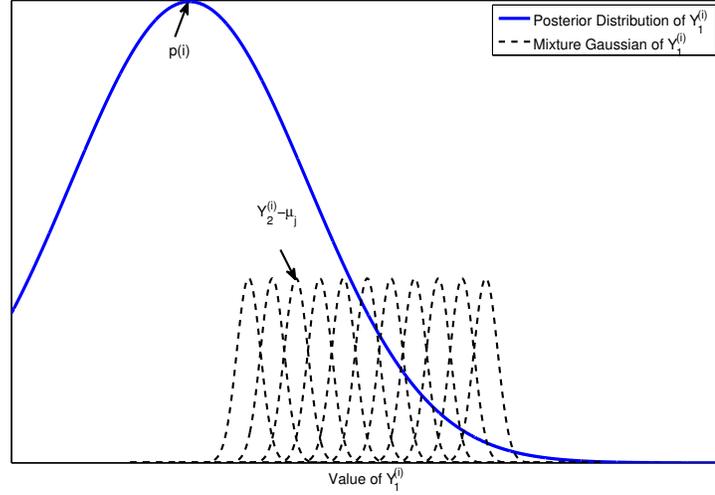


Figure 3.6: Illustration of Sampling from Distributions of  $Y_1^{(i)}$

Detailed sampling procedure is established as follow:

1. Obtain prior MLE of hyperparameters in Gaussian Process with respect to input  $X^{(\setminus i)}$  and  $Y_1^{(\setminus i)}$ .
2. Calculate predicted distribution of  $Y_1^{(i)}$  given  $X$  and  $Y_1^{(\setminus i)}$  based on Gaussian Process prediction step.
3. Get the potential modes of  $Y_1^{(i)}$  with respect to  $Y_2^{(i)}$  and calculate their weight from predicted distribution  $P(Y_1^{(i)} | Y_1^{(\setminus i)}, X)$  respectively.
4. First, sample from a uniform distribution and combine the weights of each identity to determine which mode identity  $j$  that  $Y_1^{(i)}$  belongs to.
5. Second, sample from the distribution  $N(\mu_j, \sigma^2)$  to get  $y_1^*$  as a sample of  $Y_1^{(i)}$  and use it to update  $Y_1$ .

The efficiency of this Gibbs Sampling procedure has been tested the data set generated from a nonlinear model structure.

### 3.4.3 Summary of Algorithm

To conclude, the whole iterative procedure of proposed algorithm is presented in Fig 3.7. To obtain an improved Gaussian Process output  $Y_1$  from available measurements  $X$  and  $Y_2$ , Gibbs Sampling is applied in each and every iteration. In a certain iteration,  $i$ -th sample is

preset to be sampled; values are assigned to the  $Y_1^{(\setminus i)}$  either according to previous sampling results (if sampled) or corresponding sampled value in  $Y_2$ . Posterior distribution of  $Y_1^{(i)}$  given all the rest information is calculated and new value of  $Y_1^{(i)}$  is sampled accordingly. For simplicity, maximum number of iteration is set subjectively to be sufficiently large as the stopping criterion of this algorithm.

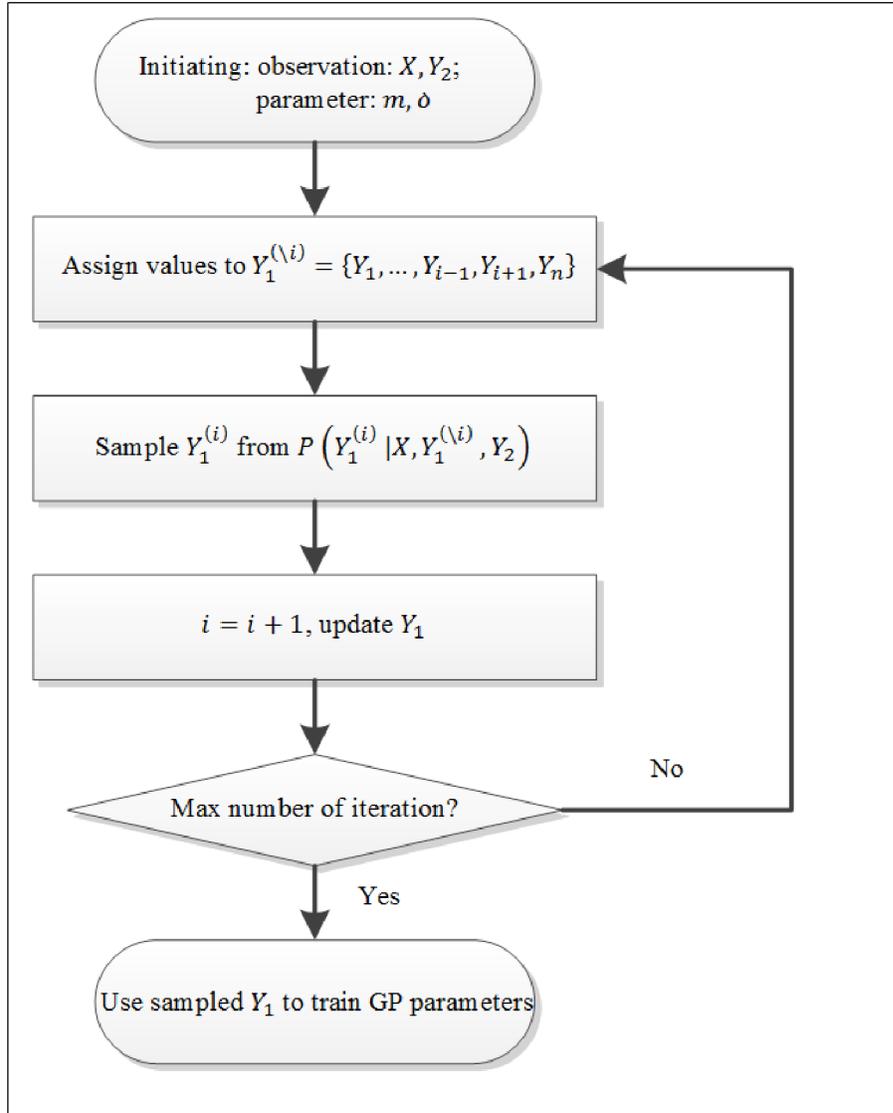


Figure 3.7: Flowchart of Gibbs Sampling Procedure

### 3.5 Simulation & Results

To verify the efficiency of proposed Gibbs Sampling approach as a de-noise method in identification of nonlinear process with quantization under Gaussian Process framework,

the following nonlinear model in Eqn 3.37 is adopted. Fig 3.8 shows the function value of this nonlinear model and it is of significant nonlinearity.

$$y = \sin x + 2 \sin^2(x - 5) + \sin\left(\frac{x}{2} - 10\right) + 2 \sin \frac{x}{2} \quad (3.37)$$

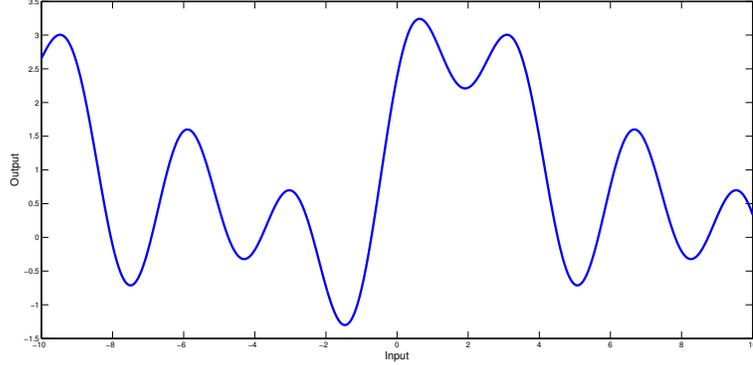


Figure 3.8: Trend Plot of Nonlinear Model Used

Similar to the numerical example as in Chapter 2, different step sizes of quantization module  $\Delta$  are added to the output of this nonlinear process. The input is designed to be white Gaussian noise. For efficiency, the training data set is generated through Latin Hypercube Sampling in order to obtain a comprehensive description of process characteristic. After obtaining the original output from this model, a quantization module is added. The proposed Gibbs Sampling method is conducted on the quantized data in order to reconstruct the original output. The maximum number of iterations is set to be 100 which is sufficiently large with acceptable computational time. When the maximum number is reached, the mean value of last 50 iterations of each sample is taken as the reconstructed output.

To evaluate the proposed algorithm, performances of original quantized data and reconstructed data by proposed algorithm in training the Gaussian Process model are compared. Since Gaussian Process is a non-parametric and surrogate modelling approach, the performance of identification is validated by prediction on cross-validation data instead of directly comparing model parameters with the “real” one. Root Mean Square Error (RMSE) and Correlation Coefficient (CorrCoef) which are widely applied in evaluation of soft sensor performance, are selected as two evaluation criteria for comparing the predicted and real output. These two indicators are defined as Eqn 3.38. Also, to ensure fair and valid comparisons, the same optimization function from Matlab with default settings is used to train

the hyperparameters of Gaussian Process model in both cases.

$$\begin{aligned}
 RMSE &= \sqrt{\frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2} \\
 CorrCoeff &= \frac{\sum_{i=1}^N (y(i) - \bar{y})(\hat{y}(i) - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y(i) - \bar{y})^2} \sqrt{\sum_{i=1}^N (\hat{y}(i) - \bar{\hat{y}})^2}}
 \end{aligned} \tag{3.38}$$

where  $y$  is the reference and  $\hat{y}$  is the predicted value.

Table 3.2 shows the prediction performance on the cross-validation dataset of two Gaussian Process models trained from quantized data and reconstructed data respectively with step size  $\Delta = \{0.5, 1, 1.5, 2, 2.5, 3\}$ . Similar to the performance of Flat-topped Gaussian noise distribution in Chapter 2, as the quantization step size increases (resolution decreases), the prediction performance of Gaussian Process model trained from quantized data decays significantly. Meanwhile the reconstructed training set can still provide a Gaussian Process model with relatively satisfactory prediction performance. In addition, as an illustration of the efficiency in reconstruction of proposed algorithm, Fig 3.9 shows the comparison of quantized output data, reconstructed data and original output data before quantization from training set when the step size is set as  $\Delta = 2$ . It can be seen that in training set, even if the quantization procedure has eliminated some of the local trends in the original output, the proposed de-noising algorithm is capable of reconstructing this information from quantized data and provides a better knowledge of model output. Moreover, Fig 3.10 shows that when quantization effect is not significant ( $\Delta = 0.5$ ), reconstructed data stays close to the original one and corresponding prediction performances of two data sets are similar. Therefore, it can be claimed that when the influence of uniform uncertainty is significant, the true characteristic of this nonlinear model can be better learned by Gaussian Process with the reconstructed data which will yield more accurate prediction performances. In addition, when the influence of uniform uncertainty is not obvious, the proposed de-noising algorithm can also maintain a good prediction performance without introducing extra error to the training data set.

Table 3.2: Prediction Performance of Gaussian Process Trained by Quantized and Reconstructed Data

	Contaminated Data		Reconstructed Data	
	RMSE	CorrCoef	RMSE	CorrCoef
$\Delta = 0.5$	0.6823	0.8759	0.6760	0.8678
$\Delta = 1$	0.7058	0.8647	0.6916	0.8699
$\Delta = 1.5$	1.2398	0.6453	0.9436	0.7970
$\Delta = 2$	1.1843	0.6956	0.7679	0.8499
$\Delta = 2.5$	1.7446	0.4102	0.9015	0.7842
$\Delta = 3$	1.6100	0.5503	1.0944	0.7561

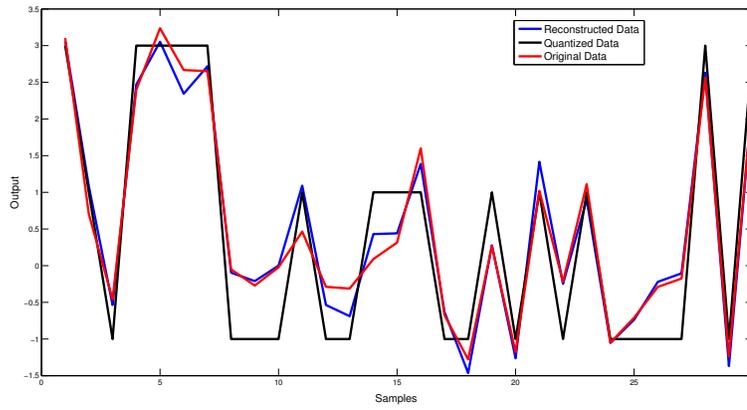


Figure 3.9: Reconstruction Performance in Training Set ( $\Delta = 2$ )

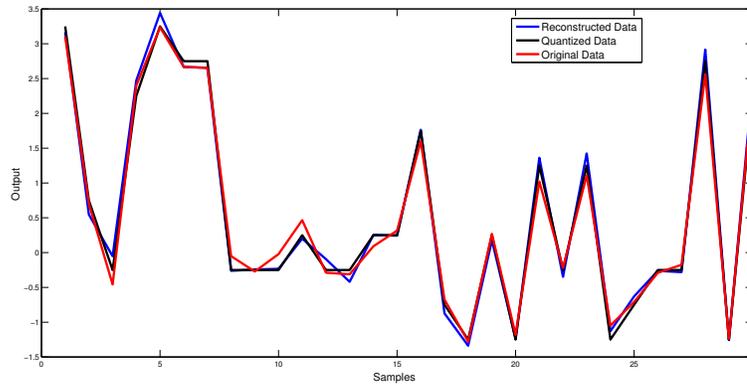


Figure 3.10: Reconstruction Performance in Training Set ( $\Delta = 0.5$ )

### 3.6 Discussion & Future Work

The key finding of this chapter is the establishment of a de-noising algorithm for Gaussian process modelling with Gibbs Sampling in order to achieve better modelling accuracy in data set with additional Flat-topped Gaussian distributed noise. Unlike the revised noise distribution function proposed in Chapter 2, Gibbs Sampling approach is incorporated to sample from the posterior distribution of every output measurement modelled as Gaussian Process. After maximum number of iterations is reached, use the average of last few samples to reconstruct the original output of Gaussian Process and train its hyperparameters correspondingly.

One of the potential limitations in the proposed approach is that it is computationally expensive. In order to attain a relatively accurate reconstruction of the original output, it is necessary to undergo a large number of iterations while Gaussian Process is trained in every iteration which calls for solving a nonlinear optimization problem. Hence the computational load increases significantly and such deficiency will prevent it from on-line implementation.

Moreover, there is still much space of improvement for the proposed algorithm. A few potential improvements are listed as follows. Both mathematical rigorousness and scope of applicability of the proposed algorithm can be improved.

1. Development of multivariate Flat-topped Gaussian distribution: if similar marginal and conditional distribution properties can be proved for Flat-topped Gaussian distribution, then it is possible to formulate multivariate Flat-topped Gaussian distribution imitating multivariate Gaussian distribution and allows so-called “Flat-topped Gaussian Process” to be applied in process modelling.
2. Arbitrary form of additional noise distribution: since Gaussian Mixture model is capable of approximating arbitrary distributions, an enlarged set of additional noise distributions can be handled by revision of the Mixture Gaussian approximation part in the proposed algorithm.
3. Analytical formulation of de-noising step: Gibbs Sampling is usually considered to be a randomized alternative to the deterministic inferential algorithms such as Expectation Maximization (EM) and Variational Bayes (VB). These approaches can be considered in the de-noising step so as to obtain analytical reconstruction of proposed distribution.

## Chapter 4

# Design of Emulsion Flow Soft Sensor for SAGD Process

An accurate and granular measurement of key process variable is vital in monitoring and subsequent advanced control of industrial processes. However, when it is difficult or impossible to measure the key variable directly, an inferential model based on relevant measurements can be established as an alternative. As an example of statistical modelling method applied in industrial problems, development of emulsion flow soft sensor in Steam Assisted Gravity Drainage (SAGD) process is introduced in this chapter. Details of both off-line modelling approaches and practical issues in on-line implementation are covered in order to attain a representative coverage of industrial applications.

### 4.1 Introduction

In the last one and half decade, Steam-assisted Gravity Drainage (SAGD) which is an enhanced in-situ oil recovery technique, has grown to be one of the primary thermal recovery methods in extraction of Athabasca Oils Sands located in northern Alberta [67]. It has been determined based on the geological and physical features of Athabasca Oils Sands that only 10% of it is located at the surface and can be extracted by surface mining method in [68]. Therefore, SAGD process combines the advances in well drilling and steam injection to enhance the extraction ability and improve its efficiency [69]. It has been claimed to be twice as thermally efficient as the Cyclic Steam Stimulation (CSS) method [70]. The primary production performance parameter in SAGD process is the rate of heavy oil produced which is of profound interest. Tracing back to 1980s, Dr. Butler pioneered in exploration of SAGD process and proposed equations for predicting the maximum oil recovery rate [71]. In practice, real-time flow rate measurement of produced fluids from wells is directly related

to the amount of oil extracted and vital to achievement of production objectives, while sometimes it is expensive or even impossible to attain this measurement by hardware sensors due to the multiphase mixture nature of the fluid that consists of various components.

On the other hand, soft sensor incorporates mathematical modelling technology to make it possible to infer the value of interested variables based on information about other relevant variables as well as process dynamics. As an economical alternative to hardware sensors, it is widely applied in industrial practice such as in steel industries [72], machining processes [73], and chemical plants [74]. Soft sensors have enjoyed rapid development during last few decades. Recently, as statistical modelling and data analysis techniques developed more sophisticated approaches, such as Bayesian Inference [12], Just-in-Time methods [75], Expectation Maximization [76] and Variational Bayesian algorithms [23] are incorporated in soft sensor development. Specifically in oil sands industry, soft sensor projects also have been conducted including SAGD process [77, 78, 79, 76]. The soft sensor development procedure can be described by Fig 4.1. In addition to testing better advanced statistical modelling approaches proposed in previous chapters, a number of practical issues that arise in different stages of soft sensor development are introduced and corresponding solutions are proposed by either referring to previous work or utilizing other techniques. Owing to the collaboration with industrial partners in oil sands industry, detailed P&ID diagrams and process descriptions as well as first hand on-site data are available for model establishment and validation throughout this work and its on-line implementation is in progress.

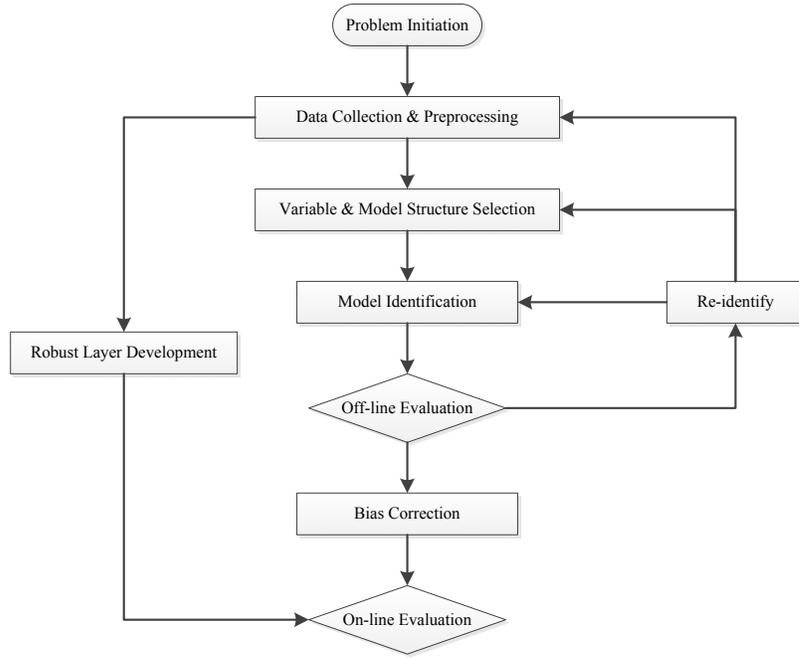


Figure 4.1: Flowchart of Soft Sensor Development Procedure

The remainder of this chapter is organized as follows: Section 4.2 gives a brief description of SAGD process. Section 4.3 is mainly concerned about the model identification procedure and accompanying practical issues. An additional robust layer is developed in Section 4.4. Data reconciliation and bias correction techniques are introduced and incorporated for on-line implementation in Section 4.5. Performances of each approach is considered in corresponding section. Section 4.6 draws a conclusion of present work and suggests potential improvements for the future. Section 4.7 concludes this chapter.

## 4.2 Process Description

Fig 4.2 is an illustrative figure of a typical well pair infrastructure in SAGD process. A well pair is composed of two horizontal wells drilled underground in parallel, of which one is injection well and the other is producer well. Steam generated from upstream steam generator is injected into the underground oil-sand layer through injection well in order to create a steam chamber. Due to the high temperature and pressure of steam injected, viscosity of bitumen is reduced, which makes it possible for bitumen to detach from sands and get carried away by water flow. A fluid mixture composed of bitumen, water and gas, is

pumped out by mechanical lifting pump through producer well to central well pad. A well pad is a collection of well pairs close to each other. Emulsion flow streams from these well pairs converge to the central well pad. The cumulative flow is further transported to a group separator that separates components of emulsion for further treatment. An accurate and granular measurement of the emulsion flow is essential to realization of control objectives as well as process reliability.

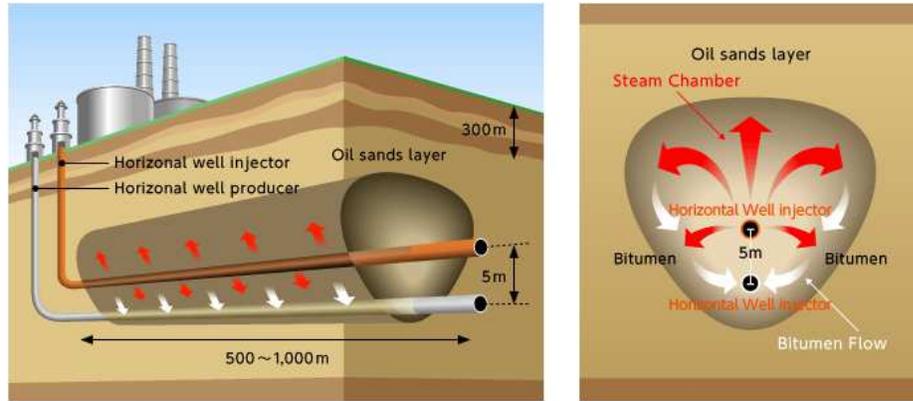


Figure 4.2: Well Pair in SAGD Process. Source: JAPEX

At present, for some well pads, a Vx meter has been installed after producer well of each and every well pair which provides a continuous and fast-rate individual emulsion flow rate measurement. However, on some other pads, well pairs have no access to this hardware sensor and their emulsion flow rate is measured through a test separator with flow meter before converging to well pad once every two weeks. This measurement procedure rotates between well pairs and only one well pair flow measurement can be obtained each time. A fast-rate overall emulsion flow measurement from a well pad is however always available regardless of hard sensor availability in individual well pairs, which provides indirect information about individual flows. Since fast rate measurements of other relevant variables such as pressure, temperature and pump frequency are always available, the objective of this soft sensor project is to establish an inferential model of emulsion flow rate as an alternative to hardware sensors.

## 4.3 Model Identification

### 4.3.1 Data Preprocessing

Generally speaking, the three fundamental steps in data-driven modelling framework are data collection, data preprocessing and model identification. Importance of data prepro-

cessing has been emphasized by Sharmin [80] since it plays a key role in off-line modelling and successive on-line application. Several data preprocessing strategies listed here are applied to raw industrial data set collected from on-site DCS in order to filter disruptive information and achieve better modelling performance.

### Outlier Detection & Replacement

In raw data collected directly from industrial site, there exist possible outliers and unreliable measurements caused by extrinsic disturbances, instrumentation failures and/or process fluctuations. The points circled in Fig 4.3 are a few examples of potential outliers in one variable measurement. The genuine trend of this variable is concealed in comparison to these extreme value. Reliable data should be guaranteed in advance in order to establish a good model and obtain accurate predictions. Here a  $3\sigma$  rule is applied to detect and replace potential outliers shown in Eqn 4.1.  $3\sigma$  is a rule of thumb in data analysis in different areas from epidemiology [81] to chemical reactions [82]. Assuming that the data measurements of  $x_k$  are samples drawn from a Gaussian distribution and denote  $\mu_k$  and  $\sigma_k$  as sample mean and variance estimated, then  $|\frac{x_k(t) - \mu_k}{3\sigma_k}| \leq 3$  will have approximately 99.7% probability. Any measurement outside this empirical boundary is considered as outliers and replaced with the measurements in front of them. Fig 4.4 is the same variable measurement as 4.3 after outlier detection & replacement procedure. Though it is still noisy, influence of extreme points is eliminated and the real trend can be seen.

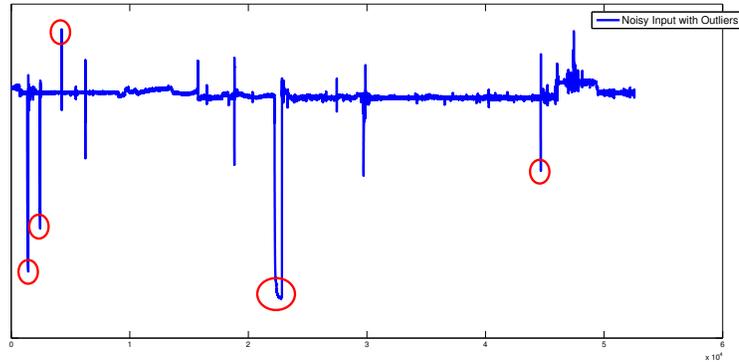


Figure 4.3: Example of Noisy Input Data with Outliers

$$x_k^{(update)}(t) = \begin{cases} x_k(t) & \text{if } |x_k(t) - \mu_k| \leq 3\sigma_k \\ x_k^{(update)}(t-1) & \text{otherwise} \end{cases} \quad (4.1)$$

$$(4.2)$$

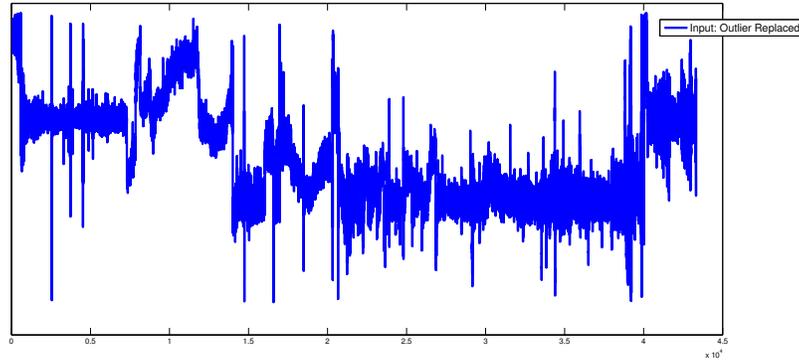


Figure 4.4: Example of Noisy Input Data with Outlier Replaced

### Smoothing

Smoothing is a specific approach that “massages” the original measurement to make it smoother so as to avoid abrupt changes in both input and output of soft sensor. This strategy is employed here because it is observed that in fast-rate 10-min data set, there are severe peaks appearing frequently, especially in pressure measurement, which by nature does not reflect the genuine change in process variables. For simplicity, a smoothing equation is defined in Eqn 4.3. For a fixed small time window  $v$ , starting at time  $t$ , all samples within the window are replaced by the median of  $v + 1$  values. Particularly for this peak phenomenon, median is superior to mean since it is less influenced by extreme observations. Fig 4.5 displays the effect of smoothing procedure on noisy pressure measurement.

However, this procedure is not required, if the measurement is averaged over a longer period of time, say 5 hours, in which case peaks will be eliminated automatically. To conclude, smoothing is only necessary when the sample rate is fast enough.

$$x(t, t + 1, \dots, t + v) = \text{median}(x(t), x(t + 1), \dots, x(t + v)) \quad (4.3)$$

After conducting aforementioned preprocessing approaches, data is ready for identification.

### Model Structure & Variable Selection

Unlike first principle models which are based on complex process mechanisms, model structure in data-driven models is relatively simple and effective for most industrial applications. After interpreting the process and performing correlation analysis between available input

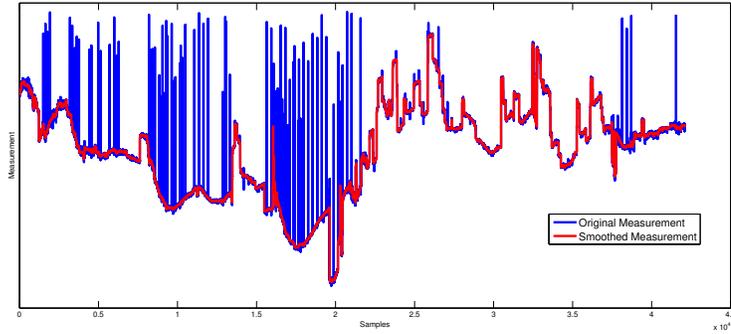


Figure 4.5: Comparison of Pressure Measurement Before/After Smoothing

variables and the target emulsion flow rate variable, the most relevant input variables were selected for model building. These are listed in Table 4.1. A model structure in Eqn 4.4 was chosen. Fig 4.6 is an illustration of the process diagram in this project with labels depicting target and input variables.

Table 4.1: Process Variables in Model Building

Notation	Variable Type	Description
$y$	Output	Emulsion Flow Rate ( $sm^3/hr$ )
$x_1$	Input	Mechanical Lift Pump Frequency PV (Hz)
$x_2$	Input	Produce Well Casing Discharge (kPa)
$x_3$	Input	Produce Well Down Hole $N_2$ HDR (kPa)
$x_4$	Input	Produce Well Down Hole (Maximum Temperature) ( $^{\circ}C$ )

$$y(t) = \theta_1 x_1(t) + \theta_2 x_2(t) + \theta_3 x_3(t) + \theta_4 x_4(t) + \theta_0 \quad (4.4)$$

### 4.3.2 Training Parameters

In addition to Ordinary Least Square method for parameter identification, a variety of other data-driven model identification algorithms were applied and evaluated and are discussed. Efficiency and advantages of these advanced identification algorithms are reflected by comparison with OLS methods. However, it is necessary to point out that in terms of applicability there are some limitations of the advanced identification algorithms, such as, 1) Flat-topped Gaussian Noise distribution is only applied in off-line training procedure to obtain model parameters; 2) Due to the non-parametric nature and matrix inversion requirements of Gaussian Process, it is not recommended in real-time prediction especially when the data set is large.

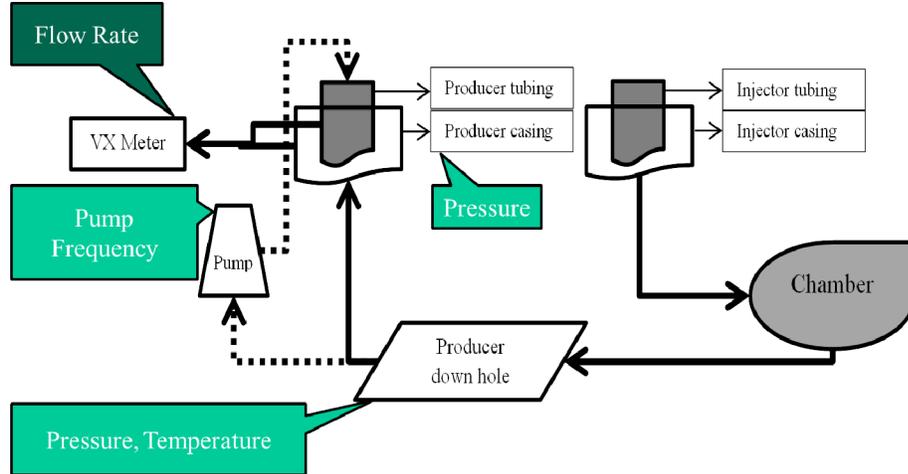


Figure 4.6: Emulsion Flow Soft Sensor Diagram

### Application of Flat-topped Gaussian Noise Distribution

Flat-topped Gaussian noise distribution proposed in Chapter 2 is tested on slow-rate data obtained in this emulsion flow soft sensor project. Although quantization or other additional noise does not exist in this process, test separator measurement is not reliable enough according to process engineers. Consequently, an appropriate tolerance of estimation error in comparison to reference is desirable when learning model parameters, which makes Flat-topped Gaussian noise assumption applicable in this case. Input and output data from two well pairs were used for constructing models using the proposed Flat-topped Gaussian noise distribution as well as ordinary Gaussian noise distribution. Subsequently, the performance of these models was compared and evaluated. Detailed description of the data is shown in Table 4.2. Due to the insufficiency and inaccuracy of slow-rate emulsion flow measurement, a certain level of tolerance is selected and the corresponding Flat-topped Gaussian distribution is used. Note that according to the nature of inaccurate measurement of the test separator, prediction error is no longer defined simply as the squared error between model output and reference. Instead, a bounded error with tolerance  $d$ , defined in Eqn 4.5 is used for calculating prediction error. This tolerance definition does not punish prediction error as long as the error is within a pre-specified error band, whose width is controlled by  $d$ . In addition to the bounded error, a performance index is defined as a count of samples that their predictions lie within the error band. Trend plots in Fig 4.7 and 4.8 qualitatively compare the Flat-topped Gaussian model performance with Gaussian noise model on two well pairs when  $d = 0.5$ . Table 4.3 shows the quantitative performance, i.e. the MSE defined in 4.5, as well as the performance index for the two models.

Table 4.2: Data Description in Model Building

Sample time	5-hr average, once every 2 weeks
Sample size	Around 60
Time range	Apr 2014 ~ Apr 2015
Partition	$\frac{1}{3}$ training and $\frac{2}{3}$ cross-validation

$$e(i) = \begin{cases} 0 & \text{if } |y(i) - \hat{y}(i)| \leq d \\ (|y(i) - \hat{y}(i)| - d)^2 & \text{otherwise} \end{cases}$$

$$MSE = \frac{1}{N} \sum_{i=1}^N e(i) \quad (4.5)$$

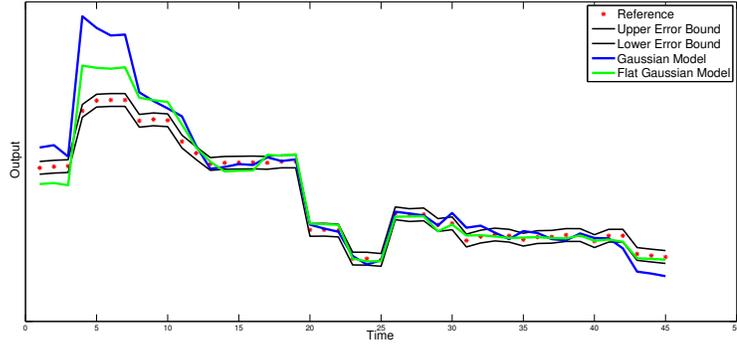


Figure 4.7: Model Performances on Well Pair 1

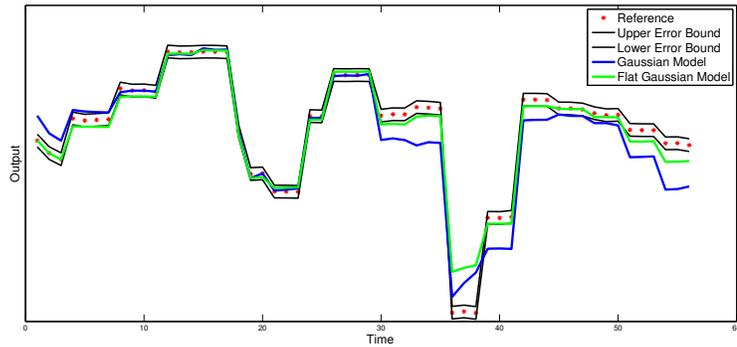


Figure 4.8: Model Performances on Well Pair 2

It can be seen from the comparison results that for both two well pairs, the model trained based on Flat-topped Gaussian noise assumption outperforms the Gaussian one. In other word, if the reference itself has mild inaccuracy issue that reduces its reliability, proposed

Table 4.3: Quantitative Model Performance

	Well Pair 1		Well Pair 2	
	Gaussian	Flat-topped Gaussian	Gaussian	Flat-topped Gaussian
MSE	2.8611	0.6230	1.6302	0.5567
Performance Index	25/45	29/45	24/56	37/56

Flat-topped Gaussian noise assumption can provide more reasonable parameter estimation with a certain degree of tolerance to reference error and therefore improve regression performance.

### Gaussian Process Modelling

As discussed it is introduced in Chapter 3, Gaussian Process modelling is a typical non-parametric modelling approach which basically relies on a distance based similarity measure between values of input variables to make predictions. Owing to its property that the correlation between input variables determines the corresponding prediction, Gaussian Process is applied to model the residual caused by deviated input variable measurements.

The data set tested is similar to that used in previous subsection of Flat-topped Gaussian noise distribution. The most outstanding feature of the well pair tested here is that there is a relatively long abnormal period during which variable measurements except pump frequency behave oddly and should not be considered in model prediction. Here the prediction of Gaussian Process, can be more accurate since it reduces to the prior mean value if input variables deviate significantly from the reliable range defined by the training set. Eqn 4.6 is the prior distribution of model structure and Eqn 4.7 is the posterior/prediction stage. In formulation of Eqn 4.7, output is composed of two parts.  $\hat{y}_1(t)$  is obtained by pump frequency, which is assumed to be normal in the whole time range.  $\hat{f}(t)$  and  $\hat{s}^2(t)$  are posterior distribution parameters of  $\hat{y}_2(t)$  calculated from Gaussian Process model, which is considered to be the residual between  $\hat{y}_1(t)$  and  $y(t)$ .  $r(t)$  is the correlation vector of  $x(t)$  with respect to the Euclidean distance.

$$\mathbf{y} - \mathbf{y}_1 \sim GP(\boldsymbol{\mu}, \sigma^2 C)$$

$$\text{where } C(i, j) = \exp\left(-\frac{1}{\theta} \|x_i - x_j\|_2\right) \quad (4.6)$$

$$\begin{aligned}
\hat{y}(t) &= \hat{y}_1(t) + \hat{y}_2(t) \\
\text{where } \hat{y}_1(t) &= \theta_1 x_1(t) + \theta_0 \\
\hat{y}_2(t) &\sim N(\hat{f}(t), \hat{s}^2(t)) \\
\hat{f}(t) &= \hat{\mu} + r^T(t)C^{-1}r(t)
\end{aligned} \tag{4.7}$$

The efficiency of Gaussian Process in residual prediction of a certain well pair with deviated measurements in comparison to Ordinary Least Square method is presented in Fig 4.9 and corresponding quantified result in Table 4.4. Evaluation criterion is defined the same as Eqn 4.8 in next subsection. It can be inferred that Gaussian Process works well in this situation. However, Gaussian Process modelling method is not suitable for on-line implementation of soft sensor especially when data set is huge since it requires complicated matrix manipulations like matrix inverse and determinant calculation. In order to handle this deviated measurement issue in input variables more efficiently, back-up model approach is proposed in Section 4.2 as part of robust layer of soft sensor.

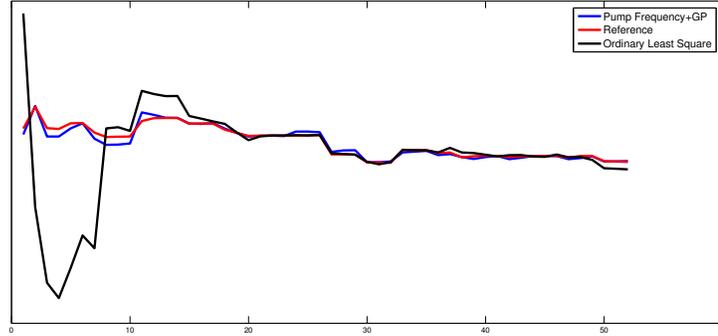


Figure 4.9: Performance of Gaussian Process on Well Pair 1

	Gaussian Process	Ordinary Least Square
RMSE	0.9521	0.0700
CorrCoef	0.9766	13.7378

### 4.3.3 OLS Performance Evaluation

Off-line testing results for soft sensors developed for individual well pairs are presented in this section. The tested sensor model is built under linear black-box model assumption with variable selection explained in Section 4.2 and its parameters are trained by Ordinary

Least Square method. To attain a more comprehensive evaluation of variable selection, model structure and regression method, it is first tested on individual well pairs with fast-rate reference, then applied to a number of well pairs for which only slow-rate reference is available, and finally on all well pairs from a certain well pad with no fast-rate individual reference.

### Well Pairs with Fast-rate Reference

Soft sensor model is first trained and validated on two well pairs from Pad 1, in which all individual well pairs have the access to fast-rate granular flow rate measurement. Hence both training and validation procedure are accomplished on fast-rate 10-min average data set. Table 4.5 shows detailed data description for this off-line test.

Table 4.5: Data Description in Fast-rate Off-line Test

	Well Pair 1	Well Pair 2
Sample time	10-min average	
Sample size	49290	43266
Identification Set	May ~ July 2014	
Cross-validation Set	Jan ~ Apr; Aug ~ Dec 2014	

Quantified cross-validation results are presented in Table 4.6. Root Mean Square Error (RMSE) and Correlation Coefficient (CorrCoef) criteria defined in Eqn 4.8, were used for evaluating the soft sensor performance. Trend comparisons shown in Fig 4.10 and 4.11 can be used to visualize the sensor performance. In the case of fast-rate reference, prediction of developed sensor is of satisfactory accuracy on cross-validation sets and matches the trend well, though there exists a possibility of constant bias that can be easily handled by bias correction step in on-line implementation.

$$\begin{aligned}
 RMSE &= \sqrt{\frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2} \\
 CorrCoef &= \frac{\sum_{i=1}^N (y(i) - \bar{y})(\hat{y}(i) - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y(i) - \bar{y})^2} \sqrt{\sum_{i=1}^N (\hat{y}(i) - \bar{\hat{y}})^2}}
 \end{aligned} \tag{4.8}$$

### Well Pad with Slow-rate Reference

After the developed soft sensors are validated by fast-rate well pair data, their performance on all 14 well pairs from Pad 2 is tested and detailed data description is given in Table 4.7. Table 4.8 gives a quantitative description of sensor performance based on correlation

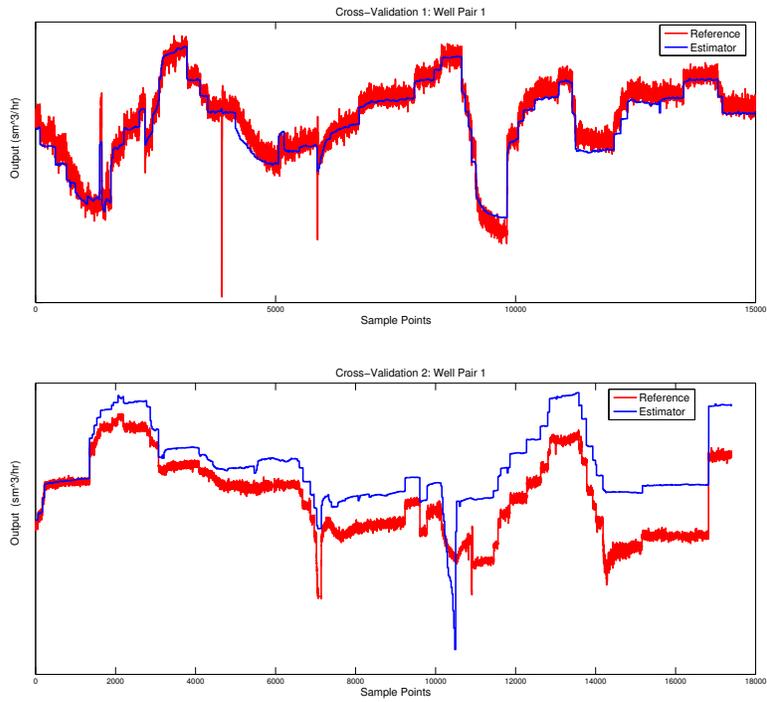


Figure 4.10: Cross-validation Performance of Pad 1 Well Pair 1

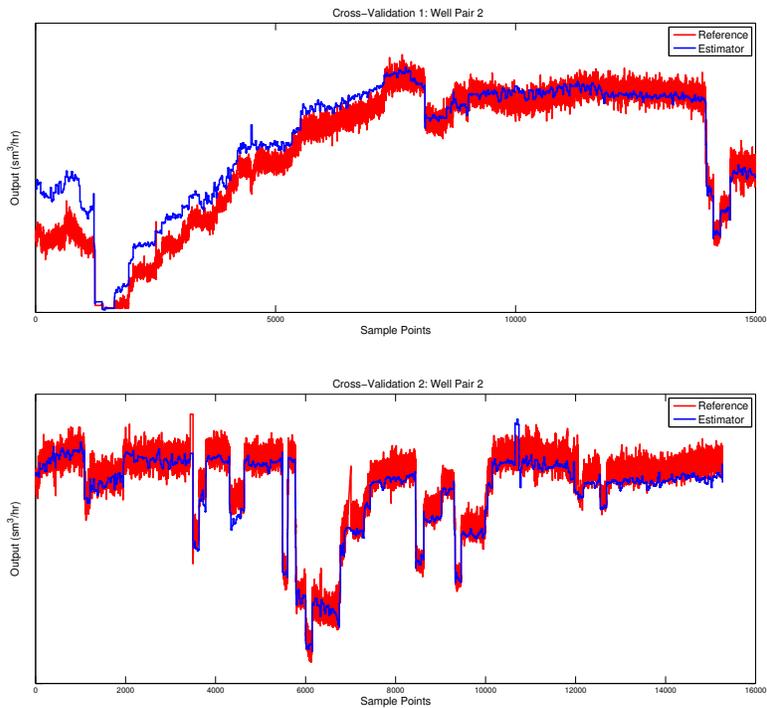


Figure 4.11: Cross-validation Performance of Pad 1 Well Pair 2

Table 4.6: Quantitative Model Performance in Fast-rate Off-line Test

	Well Pair 1		Well Pair 2	
	CorrCoef	RMSE	CorrCoef	RMSE
Auto-validation	0.9878	0.5695	0.9789	0.6014
Cross-validation 1	0.9726	0.6057	0.9770	1.1665
Cross-validation 2	0.8749	5.0901	0.9685	0.7345

and RMSE. Obviously, sensor performance differs from well pair to well pair. In “good” well pairs (coloured in red), the sensor performance is almost as good as the fast rate ones; while in “less satisfactory” well pairs (coloured in blue) sensor performance is relatively poor and thus it is not very reliable. In Section 4.5, these poor well pairs will be corrected and improved through data reconciliation and bias correction strategy in order to simultaneously achieve good individual and global prediction performance.

Table 4.7: Data Description in Pad 2

	Input Variables	Flow Rate Reference
Sample time	5-hr average, continuous	Once every 2 weeks, 1 well pair each time
Sample size	More than 1,000	Around 60
Time range	Jan 2014 ~ June 2015	
Partition	$\frac{1}{3}$ training and $\frac{2}{3}$ cross-validation	

Table 4.8: Quantitative Model Performance of Pad 2

	CorrCoef	RMSE
Overall	0.7042	89.8191
WP1	0.3301	11.9376
WP2	0.1840	19.370
WP3	0.4467	14.3449
WP4	0.4464	9.2376
WP5	0.8334	7.5466
WP6	0.9267	2.2470
WP7	0.2888	5.5988
WP8	0.8357	2.9589
WP9	0.9851	0.8842
WP10	0.8571	4.5152
WP11	0.4225	9.8918
WP12	0.6887	9.0595
WP13	0.3544	15.4104
WP14	0.3814	11.0190

## 4.4 Robust Layer Development

### 4.4.1 Sensor Robustness

In on-line implementation of soft sensor, the counterpart of data preprocessing in off-line modelling is the construction of robust layer, as both of them aim at eliminating or mitigating the influence of outliers in measurement on sensor prediction. Sensor robustness is vital. The most straightforward approach to building a robust layer involves employing a  $3\sigma$  criterion for the input measurements to generate a reliability indicator for soft sensor prediction [78, 83]. Student  $t$  distribution is often used to describe measurement noise owing to its nature of long tails and subsequent robustness to outliers [23, 11]. Advanced statistical approaches based on PCA and PLS are employed to enhance model robustness [84, 85]. Application of Just-in-Time and other locally-weighted or iterative weighted methods [86] has also been proposed.

There is a trade-off between sensor robustness and number of warnings or alarms. As the robust layer of soft sensor becomes more conservative, reliability of prediction it provides will be enhanced; nevertheless, number of warnings and other unreliability indicators increases significantly in the meantime. In this section, a 1-variable model based solely on mechanical lifting pump frequency measurement is proposed as a backup for the original 4-variable model so as to maintain a fair prediction performance even if process variables excluding pump frequency behave abnormally. As an alternative to the original sensor, this backup model can give fair predictions when temperature and/or pressure measurements are abnormal without affecting its reliability.

### 4.4.2 Back-up Model Development

In Section 4.3.2, dominance of pump frequency as the most influential process variable in the sense of flow rate has been demonstrated primarily according to the correlation coefficients between flow rate reference and pump frequency measurement. On the other hand, data analysis in Section 4.3.1 points out that the other three process variable measurements, namely two pressure measurements and one temperature measurement, are more prone to be contaminated by outliers and severely fluctuating abnormal samples, while the pump frequency measurement is supposed to be the most stable and reliable one. Consequently, a simplified model with pump frequency measurement as the only input variable is a practical choice for a back-up in case severe input failure affects any of the other three process variables. Another advantage of back-up model is that its structure is even simpler than

the 4-variable linear model shown in Eqn 4.9:

$$y = \theta_1 x_1 + \theta_0 \quad (4.9)$$

The dominance of pump frequency can be further verified based on slow-rate reference. Fig 4.13 displays the trends predicted by 4-variable model and back-up model in comparison to the slow-rate emulsion flow reference of Well Pair 7 on Pad 2. By examining the standardized trends of all four input variables as well as the output shown in Fig 4.12, it can be concluded that due to potential fluctuations or abnormalities in process variables, 4-variable model gives a deviated prediction. Meanwhile, since there is no obvious jump in pump frequency, a fair prediction is given by back-up model.

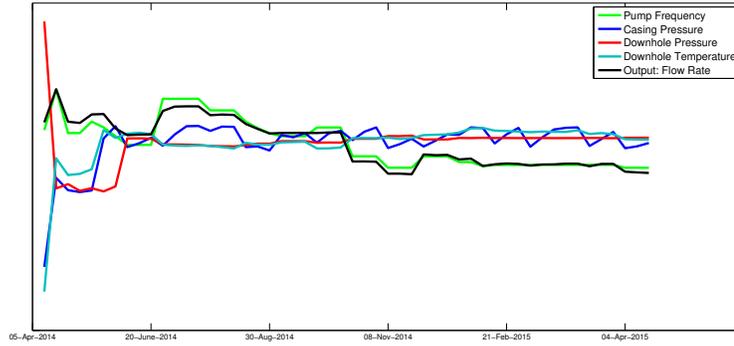


Figure 4.12: Input and Output Trends of Well Pair 7

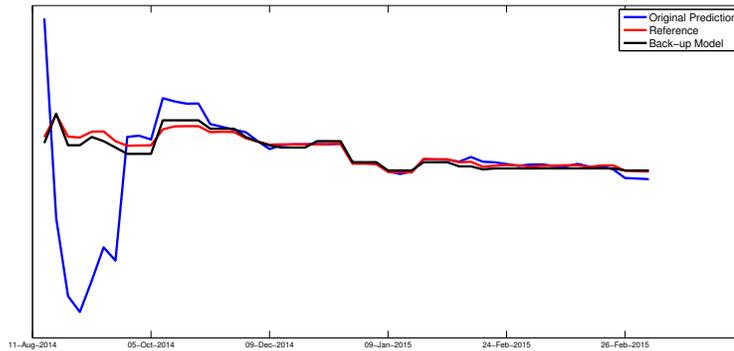


Figure 4.13: Input and Output Trends of Well Pair 7

However, it is not practical to simply remove the other three input variables if we are aiming at a good fast-rate prediction. In other words, pump frequency, which is piece-wise

constant, can give only a rough prediction of slow-rate emulsion flow, but cannot predict rapid changes in flow rate that may be triggered by other process variables. In Fig 4.14 and 4.15, segments of trend plots enclosed within green circles represent the situations where pump frequency measurement remains constant but the 4-variable soft sensor can catch the trend of flow rate measurement. The red curve is reference obtained by hard sensor, blue one is the 4-variable model prediction and black one is the back-up pump frequency model.

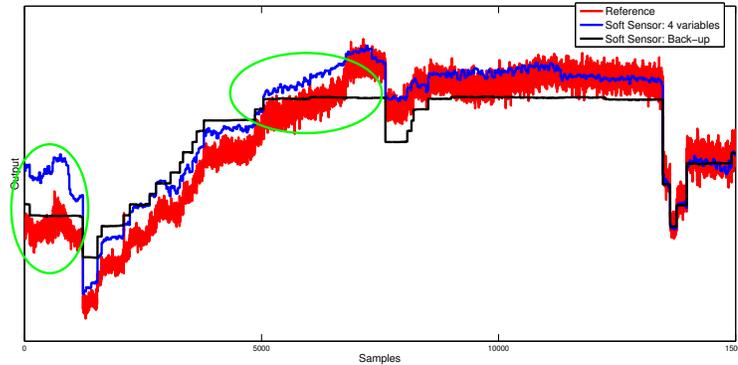


Figure 4.14: Example 1 of Limitation of Back-up Model

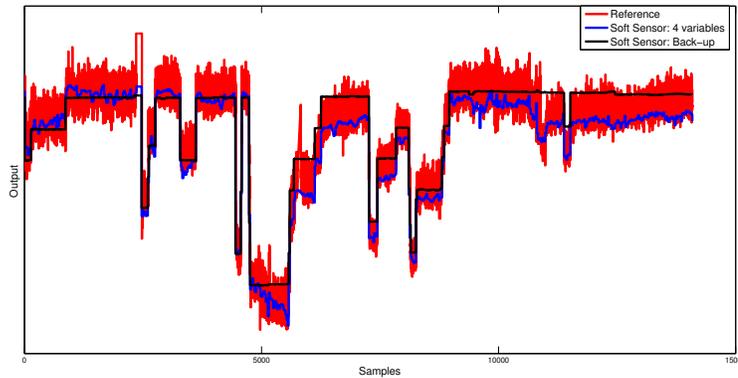


Figure 4.15: Example 2 of Limitation of Back-up Model

To make a short conclusion of this study on feasibility and limitation of the simplified model, different features of flow rate can be roughly categorized with respect to its speed of variance.

1. Slow-rate trend which is mainly introduced by adjustments in pump frequency. The dominance of pump frequency in flow rate is proved by analysing data from a couple of different well pairs. As a consequence, pump frequency model is guaranteed to give

a fair prediction of emulsion flow, regardless of abnormality in other variables, as long as the pump is working properly.

2. Moderate-rate variation caused probably by changes in other process variables such as casing pressure and downhole temperature. These variations are reflected in the trend of emulsion flow rate in addition to the basic slow-rate trend which usually last for several days. However, the test separator is not able to record this type of variation due to the infrequent sampling procedure. As a frequent estimation of emulsion flow, the developed soft sensor is required to capture these moderate-rate variations as well.
3. Fast-rate fluctuation which is considered as instrumentation noise and neglected. This fluctuation can be filtered by averaging over time and soft sensor is not designed to catch it.

#### 4.4.3 Performance Evaluation

The soft sensor output performance is tested on two well pairs from Pad 2 shown in Table 4.9. To integrate the back-up model in the soft sensor, it is trained and executed in parallel with the original 4-variable model. For now, a criterion mentioned in Eqn 4.10 is used to determine model switch in on-line sensor testing. Performance comparison shown in Fig 4.16 demonstrates that after applying back-up model, the number of unreliable sensor predictions reduces considerably and prediction performance in comparison to the slow-rate reference is improved significantly.

Table 4.9: Data Description in Back-up Model Testing

Sample time	5-hr average
Sample size	Around 1500
Number of reference	Around 60
Time range	Apr 2014 ~ Apr 2015

$$\hat{y}(t) = \hat{y}_{back}(t) \tag{4.10}$$

*if*  $\hat{y}_{original}(t) < 0.$

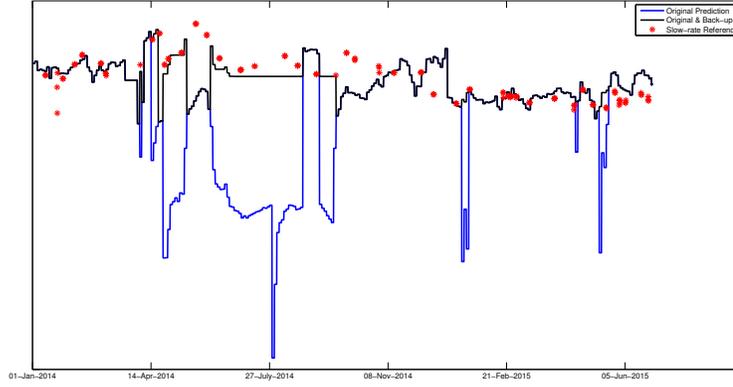


Figure 4.16: Performance Comparison of Original Model and Original Model w/ Back-up

## 4.5 Data Reconciliation in Bias Correction

### 4.5.1 Data Reconciliation & Bias Correction Approaches in Soft Sensor Development

In on-line implementation of a soft sensor, bias correction is a simple and widely used approach which can incorporate additional information regarding objective variable obtained from existing hard sensor and/or lab analysis and improve the soft sensor performance.

#### Bias Correction

Bias correction, also known as bias update, is the most widely used strategy in on-line correction stage of soft sensor due to its nature of simplicity and ease of implementation [80, 87, 17]. A typical bias correction equation is formulated as Eqn 4.11, which is based on the reference  $y(t)$  at time  $t$  and utilizes bias term to correct the original sensor output  $\hat{y}_o(t + 1)$  at next time point.

$$\begin{aligned}
 b(t) &= b(t - 1) + \alpha (y(t) - \hat{y}_u(t)) \\
 \hat{y}_u(t + 1) &= \hat{y}_o(t + 1) + b(t)
 \end{aligned}
 \tag{4.11}$$

where  $\alpha$  is a coefficient that balances the weight of historical bias  $b(t - 1)$  and current bias  $y(t) - \hat{y}_u(t)$ . It is employed in order to avoid jumps and fluctuations in corrected output and provide a smoother result.

In this project, the most challenging issue with bias correction is that individual sensor performance differs from well pair to well pair. As it is shown in Table 4.8, “good” well pairs require little correction such as Well Pair 8 and 9, while prediction of “less satisfactory” ones like Well Pair 4 deviate significantly from reference. Intuitively, it is not reasonable

to equally correct different well pairs regardless of their original performance. Moreover, although the fast-rate flow measurement of individual well pairs is unavailable, overall emulsion flow from all well pairs in a well pad is granularly measured by a hard sensor and qualifies as a supplementary of slow-rate references. Therefore, data reconciliation is introduced to reinforce the bias correction approach in presence of measurement redundancy.

### Data Reconciliation

Widely applied in industrial practice, data reconciliation is a technology that incorporates process mechanisms and mathematical techniques in order to automatically correct raw measurements associated with measurement noise. Measurement noise is handled from both optimization and probability perspectives as mentioned in Chapter 2, whereas data reconciliation approach is preferable when additional process knowledge and redundant measurements are available. The objective of data reconciliation is to harmonize original measurements with physical constraints, say mass or energy balance, in a process. Primitive concept and solution of data reconciliation is introduced in a variety of literatures [88, 89]. Yu [90], Chen [91] and Llanos [92] also indicate that more powerful ways of data reconciliation can be developed under advanced statistical frameworks such as Bayesian Inference, Principle Component Analysis and objective functions with enhanced robustness.

Fundamental data reconciliation strategy can be mathematically formulated as an optimization problem, presented in Eqn 4.12, which serves as the foundation for more sophisticated strategies.

$$\min_{y^*} \sum_{i=1}^n \left( \frac{y_i - y_i^*}{\sigma_i} \right)^2 \quad (4.12)$$

$$s.t. \quad F(y^*) = 0 \quad (4.13)$$

where  $y = \{y_1, y_2, \dots, y_n\}$  are raw measurements while  $y^* = \{y_1^*, y_2^*, \dots, y_n^*\}$  are reconciled ones.  $\sigma_i$  is standard deviation of individual measurement noise which reflects measurement accuracy and reliability.  $F(y^*) = 0$  represents additional constraints implanted to ensure that process mechanisms are satisfied.

#### 4.5.2 Problem Formulation

As introduced in Section 4.2, due to the following features of available on-line measurements, additional data reconciliation approach should be applied when performing a bias correction step in emulsion flow soft sensor development.

- Measurement redundancy: both fast-rate overall and slow-rate individual emulsion flow measurements can be obtained, which results in a situation of information redundancy;
- Desynchronized individual sampling: measurements of individual emulsion flow are not obtained simultaneously and
- Sparse objective variable measurement: ultimate goal of this project is to develop a soft sensor for estimation of individual emulsion flows; however, individual reference is obtained only once every two weeks.

Two constraints can be formulated according to available measurements in bias correction step:

$$\text{Fast-rate: } \sum_{i=1}^p (\hat{y}_i(t) + b_i(t)) = y(t) \quad (4.14)$$

$$\text{Slow-rate: } \hat{y}_i(T_i) + b_i(T_i) = y_i(T_i) \quad (4.15)$$

Notations are declared as follow:

$\hat{y}_i(t)$ : individual flow estimation of  $i$ -th well pair at time  $t$ ;

$b_i(t)$ : corresponding bias;

$y(t)$ : fast-rate overall flow measurement;

$T_i$ : time point at which output flow from  $i$ -th well pair is measured through test separator;

$y_i(T_i)$ : flow rate measurement of  $i$ -th well pair at time  $T_i$ ;

$p$ : number of well pairs in a certain well pad.

Objective variables to be reconciled are fast-rate individual bias terms  $b_i(t)$  which are going to be utilized to correct original sensor outputs. Strictly following the typical formulation of data reconciliation, an optimization problem should be established to minimize the weighted summation of difference between original measurement and reconciled output with hard constraints such as mass balance. According to process configuration in Section 4.1, all individual measurements are obtained by the same test separator. Therefore measurement noise levels of individual well pairs are supposed to be identical and reconciliation

deviation should hence be equivalently weighted. However, process mass balance equation cannot be attained due to desynchronized sampling and therefore as an alternative, other constraints should be formulated with respect to process knowledge. Since it is suggested by process engineers that overall flow measurement is considered to be more accurate than the individual one, overall fast-rate bias was adopted as the hard constraint whereas slow-rate individual bias was included in objective function. Modified data reconciliation formulation is shown in Eqn 4.16:

$$\begin{aligned} \min_{b_i(T_i)} \quad & \sum_{i=1}^p (b_i(T_i) - (y_i(T_i) - \hat{y}_i(T_i)))^2 \\ \text{s.t.} \quad & \sum_{i=1}^p b_i(t) = b(t) \end{aligned} \quad (4.16)$$

It is obvious that this formulation is still affected by mis-synchronization issue, therefore additional approximation should be considered in order to simplify this problem. Since the prediction duration is over one year while individual samples are obtained once every two weeks, individual bias term  $b_i(t)$  of well pair  $i$  can be kept the same during the sample interval between  $T_i^{(k)}$  and  $T_i^{(k+1)}$ , in other words Eqn 4.17 holds:

$$b_i(t) = b_i(T_i), \quad \text{when } t \in [T_i^{(k)}, T_i^{(k+1)}] \quad (4.17)$$

Simplified problem statement is shown in Eqn 4.18, in which the only manipulated variables at time  $t$  are the individual bias  $b_i(t)$ .

$$\begin{aligned} \min_{b_i(t)} \quad & \sum_{i=1}^p (b_i(t) - (y_i(T_i) - \hat{y}_i(T_i)))^2 \\ \text{s.t.} \quad & \sum_{i=1}^p b_i(t) = b(t) \end{aligned} \quad (4.18)$$

Lagrange Multiplier [93] is included to solve this constrained optimization problem. After introducing a Lagrange Multiplier, the dual problem is generated as:

$$\min_{b_i(t)} \sum_{i=1}^p (b_i(t) - (y_i(T_i) - \hat{y}_i(T_i)))^2 + \lambda \left( \sum_{i=1}^p b_i(t) - b(t) \right) \quad (4.19)$$

After taking first order derivative over every  $b_i(t)$  and  $\lambda$  then setting them to be zero, the symbolic solution of this problem is as follows:

$$\begin{aligned}
b_i(t) &= b_i(T_i) - \frac{1}{p} \left( \sum_{i=1}^p b_i(T_i) - b(t) \right) \\
\text{where } b_i(T_i) &= y_i(T_i) - \hat{y}_i(T_i) \\
b(t) &= b(t-1) + \alpha \left( y(t) - \sum_{i=1}^p \hat{y}_i(t) \right)
\end{aligned} \tag{4.20}$$

As the dual problem shown in Eqn 4.19 is solved whenever there is fast-rate overall measurement  $y(t)$  available, Eqn 4.20 will provide real-time fast-rate individual bias terms that can be utilized to correct original sensor output. To summarize, bias correction works simultaneously with data reconciliation approach in order to allow both individual and overall predictions to be corrected as frequently as overall measurement is collected.

### 4.5.3 Performance Evaluation

In Section 4.3, individual well pair models were established and corresponding sensor estimations can be obtained. Fast-rate real-time reconciled bias terms obtained by solving the optimization problem defined in Eqn 4.18 are combined with original sensor output. Performance of sensor output corrected with data-reconciled bias is presented and compared with the one with blind correction.

The underlying assumption in data reconciliation is that all measurement noise are zero-mean [94]. Thus as a preceding step, various approaches of gross error detection and management are developed in order to eliminate potential impact of gross error in data reconciliation. Fig 4.17 demonstrates potential existence of gross error between overall and individual measurement instruments. A constant gross error  $e_{gross}$  is roughly estimated by the difference between the mean of overall prediction and summation of original individual predictions in Eqn 4.21.

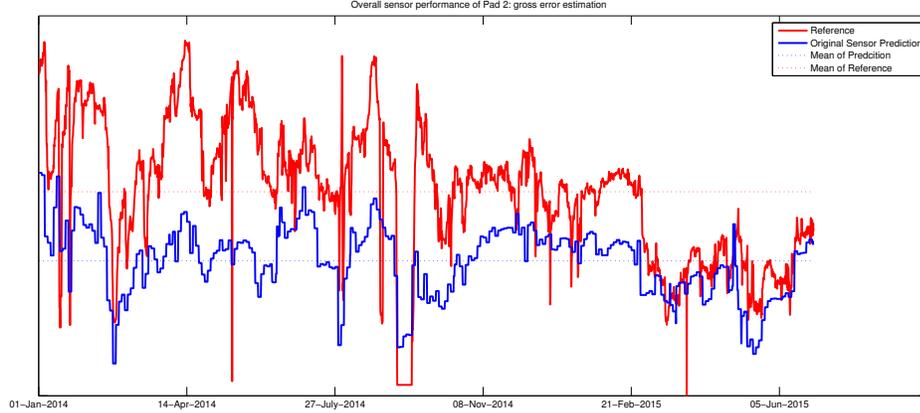


Figure 4.17: Existence of Gross Error between Overall Reference and Prediction

$$e_{gross} = \frac{1}{N} \sum_{t=1}^N y(t) - \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^p \hat{y}_i(t) \quad (4.21)$$

where  $N$  is the number of samples and  $p = 14$  is the number of well pairs on Pad 2.

Table 4.10 shows the performance comparison of original soft sensor output, blind correction result and data reconciled correction with gross error removed when the weight of current bias term is set to be  $\alpha = 0.2$ . It is obvious that the performance of data reconciliation bias correction strategy is superior both in the sense of correlation as well as mean square error in comparison to slow-rate reference.

Trend plots of all 14 well pairs along with overall flow rate are also presented to prove the efficiency of bias correction. It can be inferred from the results that for individual well pairs, the proposed bias correction method with data reconciliation will maintain good prediction performance for “good” well pairs such as Well Pair 9 and improve significantly for “less satisfactory” ones like Well Pair 4.

## 4.6 Discussion & Future Work

A few points are proposed here regarding future work that can further improve the soft sensor performance.

- The linear black-box model structure is validated in off-line testing procedure so far. However, further literature review indicates the existence of a nonlinear relationship between viscosity of bitumen, which is the most important objective component in

Table 4.10: Sensor Performance Evaluation Before/After Bias Correction

	<b>Blind Correction</b>		<b>Correction w/ DataRec</b>	
	CorrCoef	RMSE	CorrCoef	RMSE
Overall	0.8889	32.9932	0.8817	33.9909
WP1	0.3099	11.7908	0.7109	9.3359
WP2	0.2369	20.4643	0.7543	9.7723
WP3	0.5110	14.8324	0.8418	6.0244
WP4	0.4227	9.6382	0.6283	8.6695
WP5	0.8351	7.3922	0.9176	5.3936
WP6	0.8732	4.1366	0.8386	3.5873
WP7	0.4985	5.3870	0.8010	4.2897
WP8	0.7970	4.2693	0.8300	3.5259
WP9	0.8983	3.4288	0.8389	2.9935
WP10	0.7731	6.5466	0.8509	4.6129
WP11	0.5254	10.3475	0.9145	4.5869
WP12	0.7065	7.9238	0.8312	4.5290
WP13	0.4249	15.2720	0.6987	9.7499
WP14	0.4202	10.0941	0.8344	6.1980

emulsion flow, and the fluid temperature [95]. A nonlinear model structure may be utilized to improve soft sensor performance.

- More sophisticated and rigorous algorithms such as dynamic modelling and Bayesian inference [96], can be considered so as to solidify the reasoning of data reconciliation and bias correction. Also, a smoothing coefficient should be included in order to prevent the sensor from making noisy predictions.
- To avoid unnecessary or blind correction, reliability indicator of measurements will be incorporated in future and influence of outliers and/or failures in reference can be prevented.
- At present, on-line testing of soft sensor developed is ongoing. Aforementioned strategies should be integrated and transformed into easily implementable modules for real-time prediction and correction.
- To further make this soft sensor an eligible alternative to hardware sensors, additional practical situations require to be taken into consideration. For instance, abnormality in process variable measurements lasting for a while is usually observed when mechanical pump fails. Subsequent pump change will result in well pair shut down and influence emulsion flow severely. This pattern of behaviour is observed in different well pairs. Hence it is necessary for soft sensor to either recognize pump failure/change

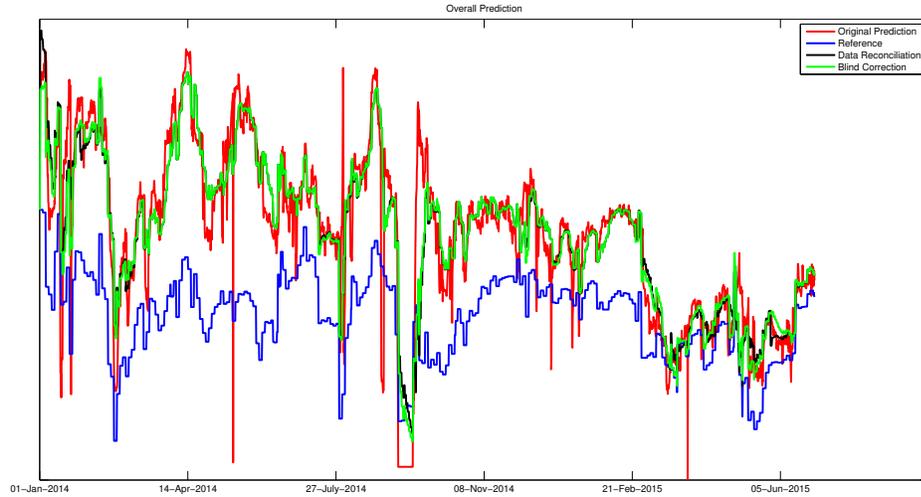


Figure 4.18: Overall Sensor Performance

behaviour automatically or read manual signals indicating the pump failure situation from operators. Therefore, fault detection and diagnosis approaches are promising to be an extension of current soft sensor.

## 4.7 Conclusion

In this work, an industrial application case study is presented as an example of methodologies in statistical modelling, process integration and industrial implementation. By synthesis of theory and hands-on experiences of engineers, an emulsion flow soft sensor is developed with embedded robust layer and bias correction stage. Off-line test of this prototype soft sensor has been successfully completed and on-line testing is in progress.

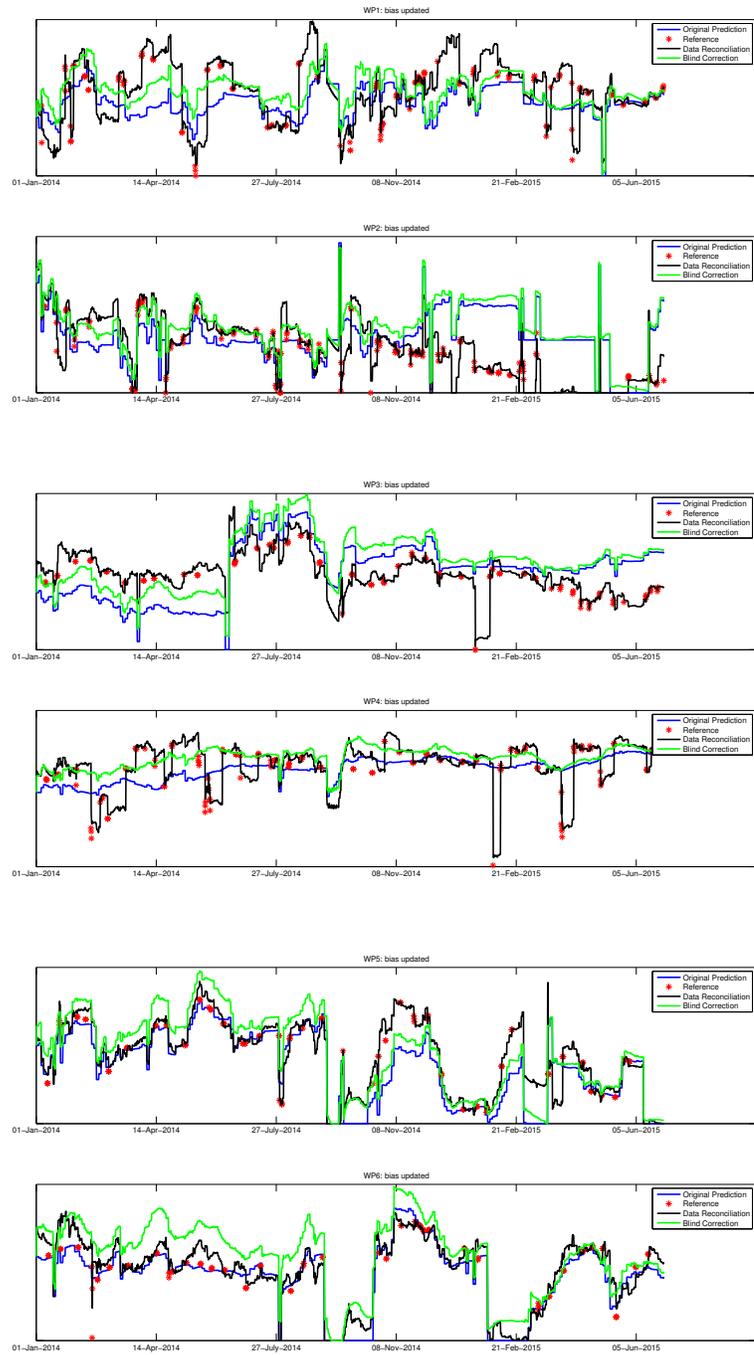


Figure 4.19: Individual Sensor Performance Before/After Bias Update: WP1-6

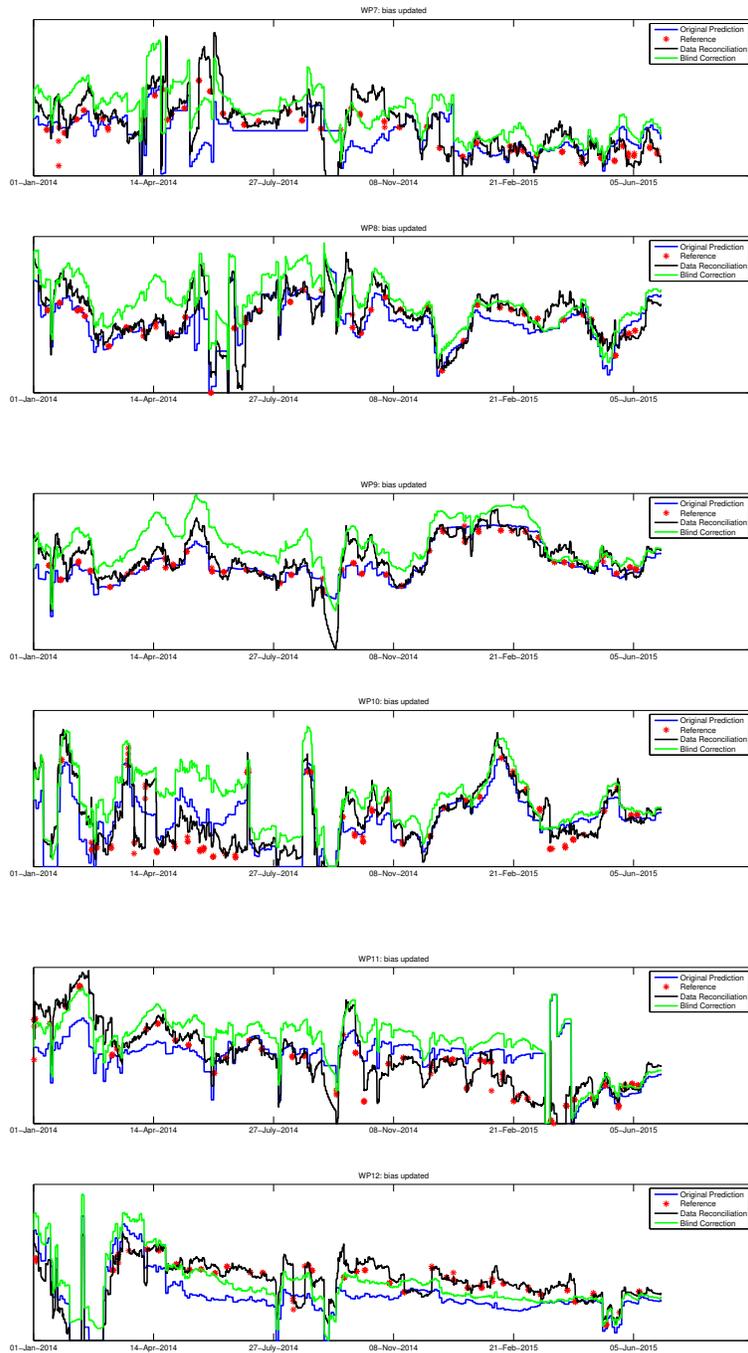


Figure 4.20: Individual Sensor Performance Before/After Bias Update: WP7-12

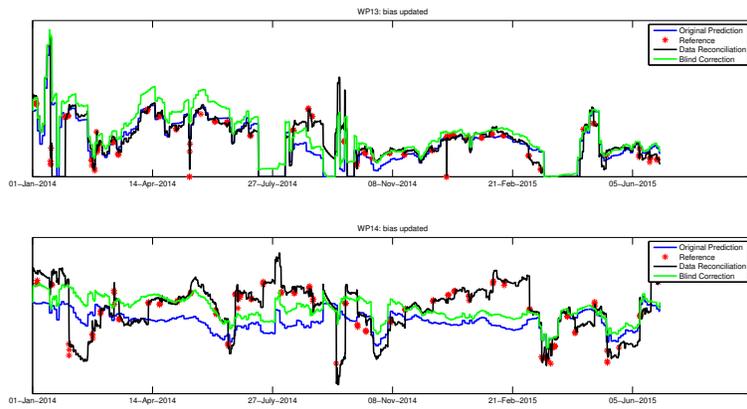


Figure 4.21: Individual Sensor Performance Before/After Bias Update: WP13-14

## Chapter 5

# Conclusions

### 5.1 Summary of This Thesis

Data-driven identification of output noise model with Flat-topped Gaussian uncertainty was investigated in this thesis. This Flat-topped Gaussian uncertainty in both linear models and nonlinear models with output which follows multivariate Gaussian distribution was studied and corresponding solutions were proposed respectively.

In the sense of parametric linear output noise model with independent Gaussian and additive uniform noise, Flat-topped Gaussian distribution was proposed in Chapter 2 as the noise distribution as an alternative to the regular Gaussian assumption. It is considered to be the distribution of the summed random variable of a Gaussian and a uniformly distributed random variable. Both functional structure and parameter estimation were discussed in this chapter. Moment fitting strategy can be applied to approximate the distribution of the summation of random variables with other distributions. Robustness of Flat-topped Gaussian distribution was studied and Flat-topped  $t$  distribution was investigated.

Chapter 3 introduced a de-noising algorithm with Gibbs Sampling method in order to solve Gaussian Process model with additive uniform noise. Mixture Gaussian approximation inherited from Chapter 2 was applied to facilitate the reconstruction of original output from Gaussian Process. Since the posterior distribution of each individual sample given the rest is available based on Gaussian Process assumption, an iterative procedure of Gibbs Sampling was formulated to sequentially reconstruct the original output from the measurement with extra uniform noise. This sampling method was further used to obtain an accurate estimation of hyperparameters in Gaussian Process.

A case study of emulsion flow soft sensor was discussed in Chapter 4 as a synthesised example of application of advanced statistical modelling approaches in real industrial objects. In the off-line modelling step, datasets collected on-site were utilized to test the effectiveness

of Flat-topped Gaussian distribution and Gaussian Process modelling approach introduced in previous chapters. In addition, various practical issues which exist both in off-line model identification and on-line implementation were considered in this work.

## 5.2 Directions for Future Work

Throughout this thesis, statistical modelling and parameter estimation strategies are proposed and discussed in presence of additional uniformly distributed uncertainty in output noise model. For deterministic model structure with white Gaussian and additional uniform noise, Flat-topped Gaussian noise distribution is established by moment fitting strategy. To move one step further, Gibbs Sampling method is incorporated to solve nonlinear and correlated output model with additional uniform uncertainty under Gaussian Process framework. There are a few points which are still open for investigation in the sense of both theoretical and practical perspectives.

From statistical aspect, in order to enhance the robustness of the identification algorithm, functional structure and parameter estimation of Flat-topped  $t$  distribution can be obtained by rigorously following the moment fitting strategy. Another possible statistical extension is the formulation of multivariate Flat-topped Gaussian distribution in case of correlated output with additional uniform noise. In this case, the “Flat-topped Gaussian” Process may emerge as an alternative to the Gaussian Process as the Flat-topped Gaussian distribution replaces the Gaussian distribution.

Meanwhile, due to the flexibility of Mixture Gaussian assumption, non-Gaussian noise distributions other than uniform one can be handled similarly in Gaussian Process modelling by going through Gibbs Sampling procedure. Furthermore, better statistical inferential approaches regarding posterior distributions, such as Expectation Maximization and Variational Bayes, can be utilized to substitute Gibbs Sampling method and provide analytical solutions instead of sampling so as to improve the computational efficiency.

Efficiency of proposed statistical modelling approaches should be improved for practical applications. For Flat-topped Gaussian noise distribution, if the analytical solution of Maximum Likelihood Estimation can be obtained, then a recursive parameter estimation algorithm may be achieved similarly as Recursive Least Square method. For Gaussian Process modelling with Gibbs Sampling, if potential sparseness of the correlation matrix is taken into consideration, the optimization step within each iteration will be simplified; therefore the computational complexity of this reconstruction algorithm can be reduced, which makes it on-line implementable.

Additional practical issues in Emulsion Flow soft Sensor project should also be considered. Before on-line implementation, it is necessary to convert the off-line data preprocessing techniques into real-time implementable procedure. More sophisticated approaches in data reconciliation, (e.g., estimating the noise covariance matrix in the objective function) may contribute to the performance improvement of bias corrected soft sensor especially when the original one is unsatisfactory.

# Bibliography

- [1] David Peel and Geoffrey J McLachlan. Robust mixture modelling using the t distribution. *Statistics and computing*, 10(4):339–348, 2000.
- [2] Hagai Attias. Inferring parameters and structure of latent variable models by variational bayes. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 21–30. Morgan Kaufmann Publishers Inc., 1999.
- [3] Markus Svensén and Christopher M Bishop. Robust bayesian mixture modelling. *Neurocomputing*, 64:235–252, 2005.
- [4] C Radhakrishna Rao. *Linear statistical inference and its applications*, volume 22. John Wiley & Sons, 2009.
- [5] Ake Björck. *Numerical methods for least squares problems*. Siam, 1996.
- [6] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [7] Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.
- [8] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
- [9] Svante Wold, Michael Sjöström, and Lennart Eriksson. Pls-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2):109–130, 2001.
- [10] Zhiqiang Ge, Biao Huang, and Zhihuan Song. Nonlinear semisupervised principal component regression for soft sensor modeling and its mixture form. *Journal of Chemometrics*, 28(11):793–804, 2014.
- [11] Jinlin Zhu, Zhiqiang Ge, and Zhihuan Song. Robust semi-supervised mixture probabilistic principal component regression model development and application to soft sensors. *Journal of Process Control*, 32:25–37, 2015.
- [12] Ming Ma, Shima Khatibisepehr, and Biao Huang. A bayesian framework for real-time identification of locally weighted partial least squares. *AIChE Journal*, 61(2):518–529, 2015.
- [13] Koji Hazama and Manabu Kano. Covariance-based locally weighted partial least squares for high-performance adaptive modeling. *Chemometrics and Intelligent Laboratory Systems*, 146:55–62, 2015.

- [14] Allen Gersho and Robert M Gray. *Vector quantization and signal compression*, volume 159. Springer Science & Business Media, 2012.
- [15] Allen Gersho. Quantization. *Communications Society Magazine, IEEE*, 15(5):16–16, 1977.
- [16] Yi Liu and Junghui Chen. Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes. *Journal of Process Control*, 23(6):793–804, 2013.
- [17] Huaiping Jin, Xiangguang Chen, Jianwen Yang, Hua Zhang, Li Wang, and Lei Wu. Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes. *Chemical Engineering Science*, 131:282–303, 2015.
- [18] K De Brabanter, P Karsmakers, F Ojeda, C Alzate, J De Brabanter, K Pelckmans, B De Moor, J Vandewalle, and JAK Suykens. Ls-svmlab toolbox users guide. *ESAT-SISTA Technical Report*, pages 10–146, 2011.
- [19] Thomas Hofmann, Bernhard Schölkopf, and Alexander J Smola. Kernel methods in machine learning. *The annals of statistics*, pages 1171–1220, 2008.
- [20] Gerd Wübbeler, Michael Krystek, and Clemens Elster. Evaluation of measurement uncertainty and its numerical calculation by a monte carlo method. *Measurement science and technology*, 19(8):084009, 2008.
- [21] James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.
- [22] Richard A Redner and Homer F Walker. Mixture densities, maximum likelihood and the em algorithm. *SIAM review*, 26(2):195–239, 1984.
- [23] Yaojie Lu and Biao Huang. Robust multiple-model lpv approach to nonlinear process identification using mixture t distributions. *Journal of Process Control*, 24(9):1472–1488, 2014.
- [24] David L Shealy and John A Hoffnagle. Beam shaping profiles and propagation. In *Optics & Photonics 2005*, pages 58760D–58760D. International Society for Optics and Photonics, 2005.
- [25] Yajun Li. Light beams with flat-topped profiles. *Optics letters*, 27(12):1007–1009, 2002.
- [26] Juan Blázquez, A García-Berrocal, C Montalvo, and M Balbás. The coverage factor in a flatten-gaussian distribution. *Metrologia*, 45(5):503, 2008.
- [27] IEC BIPM, ILAC IFCC, IUPAP IUPAC, and OIML ISO. Evaluation of measurement dataguide for the expression of uncertainty in measurement. jcgmm 100: 2008, 2008.
- [28] Peter W Buchen and Michael Kelly. The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, 31(01):143–159, 1996.
- [29] Miodrag Lovric. *International encyclopedia of statistical science*. Springer London, 2011.

- [30] Monica Borda. *Fundamentals in information theory and coding*. Springer Science & Business Media, 2011.
- [31] Sung Y Park and Anil K Bera. Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics*, 150(2):219–230, 2009.
- [32] Pran Nath. Confluent hypergeometric function. *Sankhyā: The Indian Journal of Statistics*, pages 153–166, 1951.
- [33] FR Helmert. Über die berechnung des wahrscheinlichen fehlers aus einer endlichen anzahl wahrer beobachtungsfehler. *Z. Math. U. Physik*, 20:300–303, 1875.
- [34] J Lüroth. Vergleichung von zwei werthen des wahrscheinlichen fehlers. *Astronomische Nachrichten*, 87(14):209–220, 1876.
- [35] Student. The probable error of a mean. *Biometrika*, pages 1–25, 1908.
- [36] Ronald Aylmer Fisher et al. Applications of students distribution. *Metron*, 5(3):90–104, 1925.
- [37] Geoffrey McLachlan and David Peel. *Finite mixture models*. John Wiley & Sons, 2004.
- [38] YC Zhu and Zuhua Xu. A method of lpv model identification for control. In *Preprint of IFAC World Congress*, volume 11, pages 5018–5023, 2008.
- [39] Marc Peter Deisenroth, Carl Edward Rasmussen, and Jan Peters. Gaussian process dynamic programming. *Neurocomputing*, 72(7):1508–1524, 2009.
- [40] Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392, 2009.
- [41] Edmund Jackson, Matthieu Davy, Arnaud Doucet, and William J Fitzgerald. Bayesian unsupervised signal classification by dirichlet process mixtures of gaussian processes. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 3, pages III–1077. IEEE, 2007.
- [42] Jonathan Ko and Dieter Fox. Gp-bayesfilters: Bayesian filtering using gaussian process prediction and observation models. *Autonomous Robots*, 27(1):75–90, 2009.
- [43] Ratko Grbić, Dražen Slišković, and Petr Kadlec. Adaptive soft sensor for online prediction and process monitoring based on a mixture of gaussian process models. *Computers & chemical engineering*, 58:84–97, 2013.
- [44] Eric Brochu, Vlad M Cora, and Nando De Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010.
- [45] Douglas J Leith, Martin Heidl, and John V Ringwood. Gaussian process prior models for electrical load forecasting. *Probabilistic Methods Applied to Power Systems*, pages 112–117, 2004.

- [46] Juš Kocijan, Roderick Murray-Smith, Carl Edward Rasmussen, and Agathe Girard. Gaussian process model based predictive control. In *American Control Conference, 2004. Proceedings of the 2004*, volume 3, pages 2214–2219. IEEE, 2004.
- [47] Edward Meeds and Simon Osindero. An alternative infinite mixture of gaussian process experts. *Advances in Neural Information Processing Systems*, 18:883, 2006.
- [48] Edward Snelson, Carl Edward Rasmussen, and Zoubin Ghahramani. Warped gaussian processes. *Advances in neural information processing systems*, 16:337–344, 2004.
- [49] Robert B Gramacy, Herbert KH Lee, and William G Macready. Parameter space exploration with gaussian process trees. In *Proceedings of the twenty-first international conference on Machine learning*, page 45. ACM, 2004.
- [50] Neil D Lawrence and Michael I Jordan. Semi-supervised learning via gaussian processes. In *Advances in neural information processing systems*, pages 753–760, 2004.
- [51] Matthias Seeger. Pac-bayesian generalisation error bounds for gaussian process classification. *The Journal of Machine Learning Research*, 3:233–269, 2003.
- [52] Mark N Gibbs. *Bayesian Gaussian processes for regression and classification*. PhD thesis, Citeseer, 1998.
- [53] Andrew Gordon Wilson and Ryan Prescott Adams. Gaussian process kernels for pattern discovery and extrapolation. *arXiv preprint arXiv:1302.4245*, 2013.
- [54] C Paciorek and M Schervish. Nonstationary covariance functions for gaussian process regression. *Advances in neural information processing systems*, 16:273–280, 2004.
- [55] CE Rasmussen. Gaussian processes for machine learning. 2006.
- [56] Neil D Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. *Advances in neural information processing systems*, 16(3):329–336, 2004.
- [57] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741, 1984.
- [58] Luc Devroye. Nonuniform random variate generation. *Handbooks in operations research and management science*, 13:83–121, 2006.
- [59] Brian D Ripley. *Stochastic simulation*, volume 316. John Wiley & Sons, 2009.
- [60] Walter R Gilks and Pascal Wild. Adaptive rejection sampling for gibbs sampling. *Applied Statistics*, pages 337–348, 1992.
- [61] Claus S Jensen, Uffe Kjærulff, and Augustine Kong. Blocking gibbs sampling in very large probabilistic expert systems. *International Journal of Human-Computer Studies*, 42(6):647–666, 1995.
- [62] Jun S Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966, 1994.

- [63] Carl Edward Rasmussen and Zoubin Ghahramani. Infinite mixtures of gaussian process experts. *Advances in neural information processing systems*, 2:881–888, 2002.
- [64] Carl Edward Rasmussen. The infinite gaussian mixture model. In *NIPS*, volume 12, pages 554–560, 1999.
- [65] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer, 2004.
- [66] George Casella and Edward I George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.
- [67] Q Jiang, B Thornton, JR Houston, S Spence, et al. Review of thermal recovery technologies for the clearwater and lower grand rapids formations in the cold lake area in alberta. In *Canadian International Petroleum Conference*. Petroleum Society of Canada, 2009.
- [68] Jason A McLennan and Clayton V Deutsch. Sagd reservoir characterization using geostatistics: Application to the athabasca oil sands, alberta, canada. *Press, Canadian Heavy Oil Association Handbook*, 2004.
- [69] Tawfik Noaman Nasr, Oluropo Rufus Ayodele, et al. Thermal techniques for the recovery of heavy oil and bitumen. In *SPE International Improved Oil Recovery Conference in Asia Pacific*. Society of Petroleum Engineers, 2005.
- [70] James G Speight. *The chemistry and technology of petroleum*. CRC press, 2014.
- [71] RM Butler, DJ Stephens, et al. The gravity drainage of steam-heated heavy oil to parallel horizontal wells. *Journal of Canadian Petroleum Technology*, 20(02), 1981.
- [72] Manabu Kano and Yoshiaki Nakagawa. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry. *Computers & Chemical Engineering*, 32(1):12–24, 2008.
- [73] M Chandrasekaran, M Muralidhar, C Murali Krishna, and US Dixit. Application of soft computing techniques in machining performance prediction and optimization: a literature review. *The International Journal of Advanced Manufacturing Technology*, 46(5-8):445–464, 2010.
- [74] JCB Gonzaga, LAC Meleiro, C Kiang, and Rubens Maciel Filho. Ann-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. *Computers & Chemical Engineering*, 33(1):43–49, 2009.
- [75] Koichi Fujiwara, Manabu Kano, Shinji Hasebe, and Akitoshi Takinami. Soft-sensor development using correlation-based just-in-time modeling. *AIChE Journal*, 55(7):1754–1765, 2009.
- [76] Xing Jin, Siyun Wang, Biao Huang, and Fraser Forbes. Multiple model based lpv soft sensor development with irregular/missing process output measurement. *Control Engineering Practice*, 20(2):165–172, 2012.

- [77] Shima Khatibisepehr, Biao Huang, Elom Domlan, Elham Naghoosi, Yu Zhao, Yu Miao, Xinguang Shao, Swanand Khare, Marziyeh Keshavarz, Enbo Feng, et al. Soft sensor solutions for control of oil sands processes. *The Canadian Journal of Chemical Engineering*, 91(8):1416–1426, 2013.
- [78] Li Xie, Yu Zhao, Daniel Aziz, Xing Jin, Litao Geng, Errol Goberdhansingh, Fei Qi, and Biao Huang. Soft sensors for online steam quality measurements of otsgs. *Journal of Process Control*, 23(7):990–1000, 2013.
- [79] Jing Deng, Li Xie, Lei Chen, Shima Khatibisepehr, Biao Huang, Fangwei Xu, and Aris Espejo. Development and industrial application of soft sensors with on-line bayesian model updating strategy. *Journal of Process Control*, 23(3):317–325, 2013.
- [80] Rumana Sharmin, Uttandaraman Sundararaj, Sirish Shah, Larry Vande Griend, and Yi-Jun Sun. Inferential sensors for estimation of polymer quality parameters: Industrial application of a pls-based soft sensor for a ldpe plant. *Chemical Engineering Science*, 61(19):6372–6384, 2006.
- [81] Ruth Tennant, Mohammed A Mohammed, Jamie J Coleman, and Una Martin. Monitoring patients using control charts: a systematic review. *International Journal for Quality in Health Care*, 19(4):187–194, 2007.
- [82] Jean-Paul Mardon, Daniel Charquet, and Jean Senevat. Influence of composition and fabrication process on out-of-pile and in-pile properties of m5 alloy. In *Zirconium in the Nuclear Industry: Twelfth International Symposium*. ASTM International, 2000.
- [83] Hancong Liu, Sirish Shah, and Wei Jiang. On-line outlier detection and data cleaning. *Computers & chemical engineering*, 28(9):1635–1647, 2004.
- [84] Shengwei Wang and Fu Xiao. Ahu sensor fault diagnosis using principal component analysis method. *Energy and Buildings*, 36(2):147–160, 2004.
- [85] BM Wise and NL Ricker. Recent advances in multivariate statistical process control: improving robustness and sensitivity. In *IFAC Symposium on Advanced Control of Chemical Processes. Toulouse, France*, pages 125–130. Citeseer, 1991.
- [86] Masoud Behnasr and Hooshang Jazayeri-Rad. Robust data-driven soft sensor based on iteratively weighted least squares support vector regression optimized by the cuckoo optimization algorithm. *Journal of Natural Gas Science and Engineering*, 22:35–41, 2015.
- [87] Faisal Ahmed, Salman Nazir, and Yeong Koo Yeo. A recursive pls-based soft sensor for prediction of the melt index during grade change operations in hdpe plant. *Korean Journal of Chemical Engineering*, 26(1):14–20, 2009.
- [88] Jose A Romagnoli and Mabel Cristina Sanchez. *Data processing and reconciliation for chemical process operations*, volume 2. Academic Press, 1999.
- [89] Shankar Narasimhan and Cornelius Jordache. *Data reconciliation and gross error detection: An intelligent use of process data*. Gulf Professional Publishing, 1999.

- [90] Jie Yu. A bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. *Computers & Chemical Engineering*, 41:134–144, 2012.
- [91] Hongwei Tong and Cameron M Crowe. Detection of gross errors in data reconciliation by principal component analysis. *AIChE Journal*, 41(7):1712–1722, 1995.
- [92] Claudia E Llanos, Mabel C Sánchez, and Ricardo A Maronna. Robust estimators for data reconciliation. *Industrial & Engineering Chemistry Research*, 54(18):5096–5105, 2015.
- [93] SAMUEL D Silvey. The lagrangian multiplier test. *The Annals of Mathematical Statistics*, pages 389–407, 1959.
- [94] Lloyd PM Johnston and Mark A Kramer. Maximum likelihood data rectification: Steady-state systems. *AIChE Journal*, 41(11):2415–2426, 1995.
- [95] Anil K Mehrotra and William Y Svrcek. Viscosity of compressed athabasca bitumen. *The Canadian Journal of Chemical Engineering*, 64(5):844–847, 1986.
- [96] Ruben Gonzalez, Biao Huang, Fangwei Xu, and Aris Espejo. Dynamic bayesian approach to gross error detection and compensation with application toward an oil sands process. *Chemical Engineering Science*, 67(1):44–56, 2012.