

Uncertainty is the only certainty there is.
- John Allen Paulos

University of Alberta

**Particle Filter for Bayesian State Estimation and Its
Application to Soft Sensor Development**

by

Xinguang Shao

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

Doctor of Philosophy

in

Process Control

Department of Chemical and Materials Engineering

©Xinguang Shao
Spring 2012
Edmonton, Alberta

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

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

This thesis is dedicated to ...

Yingdan You

Abstract

For chemical engineering processes, state estimation plays a key role in various applications such as process monitoring, fault detection, process optimization and model based control. Thanks to their distinct advantages of inference mechanism, Bayesian state estimators have been extensively studied and utilized in many areas in the past several decades. However, Bayesian estimation algorithms are often hindered by severe process nonlinearities, complicated state constraints, systematic modeling errors, unmeasurable perturbations, and irregular with possibly abnormal measurements. This dissertation proposes novel methods for nonlinear Bayesian estimation in the presence of such practical problems, with a focus on sequential Monte Carlo sampling based particle filter (PF) approaches. Simulation studies and industrial applications demonstrate the efficacy of the developed methods.

In practical applications, nonlinear and non-Gaussian processes subject to state constraints are commonly encountered; however, most of the existing Bayesian methods do not take constraints into account. To address this inadequacy, a novel particle filter algorithm based on acceptance/rejection and optimization strategies is proposed. The proposed method retains the ability of PF in nonlinear and non-Gaussian state estimation, while taking advantage of optimization techniques in handling complicated constrained problems.

Dynamical systems subject to unknown but bounded perturbations appear in numerous applications. Considering that the performance of the conventional particle filter can be significantly degraded if there is a systematic modeling error or poor prior knowledge on the noise characteristics, this thesis proposes a robust PF approach, in which a deterministic nonlinear set membership filter is used to define a feasible set for particle sampling that guarantees to contain the true state of the system.

Furthermore, due to the imperfection of modeling and the nature of process uncertainty, it is important to calibrate process models in an adaptive way to achieve better state estimation performance. Motivated by a question of how to use the multiple observations of quality variables to update the model for better estimate, this thesis proposes a Bayesian information synthesis approach based on particle filter for utilizing multirate and multiple observations to calibrate data-driven model in a way that makes efficient use of the measured data while allowing robustness in the presence of possibly abnormal measurements.

In addition to the theoretical study, the particle filtering approach is implemented in developing Bayesian soft sensors for the estimation of froth quality in oil sands Extraction processes. The approach synthesizes all of the existing information to produce more reliable and more accurate estimation of unmeasurable quality variables. Application results show that particle filter requires relatively few assumptions with ease of implementation, and it is an appealing alternative for solving practical state estimation problems.

Acknowledgements

During my PhD work in the past five years, I have met and interacted with many interesting people who have, in one way or another, influenced the path of my research. First of all, I would like to express my deepest gratitude towards my thesis advisor Dr. Biao Huang. I feel very fortunate to have joined his research group five years ago. Indeed, I am very lucky to have him as my advisor who has always been supportive, encouraging and patient on my work. I gratefully acknowledge Dr. Huang's invaluable guidance, he has been a never-ending source of inspiration and ideas, and I just wish I could make use of them all.

I am very grateful to my co-advisor Dr. Jong Min Lee for his advising on my research work from 2006 to 2010. I learned so much from him during these years: how to understand and tackle the problem, how to effectively present work in both writing and oral presentation, and how to set high standards for myself in order to achieve constant improvement. Dr. Lee has been a mentor for me and I wish all the best for his faculty career at the Seoul National University and wish him enjoy the life with family in South Korea.

The Computer Process Control (CPC) group at the University of Alberta has been resourceful and filled with some of the nicest persons I ever come across. The courses and seminars given by Dr. Sirish L. Shah, Dr. Amos Ben-Zvi, Dr. Fraser Forbes, Dr. Vinay Prasad and others professors were helpful in acquiring the necessary background for research in process control area. The graduates in CPC are always available for discussing and sharing new ideas. I sincerely thank all my friends and group mates: Fei Qi, Moshood Olanrewaju, Shima Khatibisepehr, Elom Domlan, Natalia Marcos, Xing Jing, Chunyu Liu, Yu Zhao, Fan Yang, Yu Yang, Ruben Gonzales, Yuri Shardt. I wish you all the best in the future.

Great acknowledgements are also due to NSERC and Syncrude Canada Ltd. for providing the funding for my PhD work, and the process control personnel at Syncrude Extraction and Upgrading groups. I sincerely thank Dan Brown, Aris Espejo, Edgar Tamayo, Fangwei Xu and Bo Li for all the support and help during my stay in Fort McMurray.

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

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Contributions	5
1.2.1	Contributions via theoretical developments	5
1.2.2	Contributions via industrial applications	6
1.3	Organization	6
2	Review of Recursive Bayesian State Estimation	9
2.1	Problem Statement	9
2.2	Recursive Bayesian Estimation	10
2.3	Bayesian Interpretation of Existing Methods	13
2.3.1	Kalman filtering based methods	13
2.3.2	Moving horizon estimator	16
2.4	Particle Filter	16
2.4.1	Resampling	20
2.4.2	Choice of importance density	22
2.5	Case Studies	23
2.5.1	Two-state CSTR	23
2.5.2	Tennessee Eastman benchmark process	26
2.6	Conclusions	28
3	Constrained Bayesian State Estimation	33
3.1	Introduction	34
3.2	Constrained Bayesian Estimation	35

3.3	Constrained Particle Filter	36
3.3.1	Acceptance/Rejection	37
3.3.2	Optimization formulation	37
3.3.3	Constrained PF algorithm	43
3.4	Case Studies	45
3.4.1	Two-state batch reaction	45
3.4.2	Three-state batch reaction	51
3.4.3	Three-state continuous stirred-tank reaction	52
3.5	Conclusion	56
4	Robust Particle Filter for Unknown But Bounded Uncertainties	61
4.1	Introduction	61
4.2	Preliminaries of Ellipsoidal Techniques	63
4.3	Ellipsoidal Bound for Nonlinear Systems	66
4.4	Guaranteed Robust Particle Filter	69
4.4.1	Extended Set Membership Filtering	69
4.4.2	ESMF based PF algorithm	71
4.5	Simulation Studies	72
4.5.1	Nonlinear numeric example	72
4.5.2	Continuous fermentation process	74
4.6	Conclusion	75
5	Particle Filter for Multirate Data Synthesis and Model Calibration	79
5.1	Introduction	80
5.2	Data-driven models	82
5.2.1	Dynamic modeling based on fast-rate input/output data	82
5.2.2	Static modeling based on slow-rate input/output data .	83

5.2.3	Dynamic modeling based on fast-rate input and slow-rate output data	83
5.3	Bayesian calibration of data-driven models	85
5.3.1	Model calibration	86
5.3.2	Bayesian information synthesis	86
5.4	Bayesian information synthesis with abnormal observation data	91
5.5	An Illustrative Example	94
5.5.1	Algorithm characteristics	95
5.5.2	Model calibration	97
5.6	Industrial Application	99
5.6.1	Background	99
5.6.2	Model estimation	102
5.6.3	Bayesian calibration	105
5.7	Conclusion	106

6 Industrial Contribution: Estimation of Bitumen Froth Quality

	Using Bayesian Information Synthesis	113
6.1	Introduction	114
6.2	Process Description	115
6.2.1	Aurora Bitumen Froth Transportation	116
6.2.2	Existing Water Content Measurements	118
6.3	Soft Sensor Development	119
6.3.1	Variable Selection	120
6.3.2	Synthesis of Secondary Variables Using PCR	121
6.3.3	Bayesian Model calibration	124
6.4	Soft Sensor Performance Assessment	126
6.4.1	Preliminary Step Test	126
6.4.2	Performance Assessment Using Lab Data	127
6.4.3	Soft Sensor Based Water Content Control	129
6.5	Conclusion	130

7 Conclusion and Future Work	133
7.1 Conclusion	133
7.2 Future Work	135
APPENDIX	136
A Constrained PFs based on Equations (3.9) and (3.10)	136

List of Tables

3.1	Constrained state particles for the generic PF	41
3.2	Comparison of estimation performances for example 3.4.1 . . .	50
3.3	Comparison of estimation performances for example 3.4.3 . . .	56
4.1	Nominal fermenter parameters and operating conditions	75
5.1	Performance comparison for water-content estimate	107
6.1	Selected secondary variables for froth line modeling	121
6.2	Error comparisons between PCR model and hardware sensor .	122
6.3	Error comparisons between soft sensor and hardware sensor . .	125

List of Figures

2.1	Illustration of recursive Bayesian state estimation.	11
2.2	Illustration of point state estimation.	12
2.3	Illustration of multinomial resampling strategy.	21
2.4	Illustration of non-Gaussian property for CSTR case.	24
2.5	State estimation under different approaches for CSTR case. . .	25
2.6	Absolute error and root-mean-square error (RMSE) compar- isons under different approaches for CSTR case.	25
2.7	The Tennessee Eastman process flowsheet (Downs and Vogel (1993)).	26
2.8	Estimation results for TE process (regular measurements). . .	27
2.9	Estimation results for TE process (infrequent measurements and states).	28
3.1	An example for a multimodal pdf.	36
3.2	An example for constraints on states.	38
3.3	Illustration of projection (\circ : <i>valid particle</i> , \bullet : <i>violated particle</i> , \star : <i>true state/measurement</i>).	40
3.4	Illustration example of differences among Equations (3.7), (3.8), (3.9) and (3.10) (\circ : <i>valid particle</i> , \bullet : <i>violated particle</i> , \diamond : <i>estimated mean</i> , \star : <i>true state</i>).	42
3.5	EKF estimates for example 1.	46
3.6	UKF estimates for example 1.	47
3.7	MHE estimates for example 1.	47
3.8	PF estimates for example 1.	49

3.9	Simulation results for example 2.	52
3.10	EKF estimates for example 3.	54
3.11	UKF estimates for example 3.	54
3.12	MHE estimates for example 3, $h = 2$	55
3.13	Optimization based constrained PF estimates for example 3.	55
4.1	Geometry illustration of ellipsoid summation.	64
4.2	Geometry illustration of ellipsoid intersection.	65
4.3	Illustration of ellipsoidal bound of linearization error.	68
4.4	Estimation results of x_1 using PF, EPF and ESMPF.	73
4.5	Error comparisons for estimate of x_1 using PF, EPF and ESMPF.	73
4.6	Estimation results for continuous fermentation process. (a) estimation results of biomass concentration X; (b) estimation results of substrate concentration; (c) estimation results of the product concentration P.	76
5.1	Bayesian filter based data fusion strategies: (a) distributed approach; (b) centralized approach; (c) hybrid (sequential) approach.	88
5.2	Graphical representation of Equation (5.5); grey nodes represent known variables.	90
5.3	Prediction step for Bayesian inference.	90
5.4	Update validation step for Bayesian inference.	91
5.5	Observation validity given a sensor reading.	93
5.6	Evolution of the simulated output for the numeric example.	95
5.7	Comparison of two different measuring approaches.	96
5.8	Illustration of observation fusion.	97

5.9	Observation fusion with one possible abnormal reading ($y_k^1 = -1.8$). (a). poor fusion result with prefixed constant measurement noise variances ($\sigma_1^2 = 0.5^2$, $\sigma_2^2 = 0.2^2$); (b). improved fusion result with time-varying variances calculated based on Equation (5.10) ($\sigma_1^2(k) = 0.83^2$, $\sigma_2^2(k) = 0.2^2$).	98
5.10	Comparisons of existing measurements and model prediction with the true output.	100
5.11	Estimate results with different model calibration approaches; (a). EKF based approach; (b). PF based approach.	100
5.12	Schematic diagram for an Inclined Plates Settler (IPS) unit.	102
5.13	Input and output data for NLARX modeling of the investigated IPS unit.	103
5.14	A general structure of Nonlinear ARX model.	104
5.15	Model simulation and validation results for the investigated IPS unit; (a). training; (b). validation.	104
5.16	NLARX estimates without model calibration for new data set collected in July 2009.	105
5.17	NLARX estimate with model calibration for new data set collected in July 2009.	107
6.1	Simplified schematic of Aurora bitumen froth transportation pipeline.	117
6.2	Hardware readings for water content measurements.	118
6.3	Hot process water addition in Aurora froth pipeline.	120
6.4	PCR model testing results (trends plot).	123
6.5	PCR model testing results (scatter plot).	123
6.6	Soft sensor testing results (trends plot).	125
6.7	Soft sensor testing results (scatter plot).	126
6.8	Online hot water flowrate step test for soft sensor model validation.	127

6.9	Comparisons of soft sensor model, unit lab data with NMR lab results.	128
6.10	Soft sensor online implementation performance.	128
6.11	Inferential control for water content.	129
6.12	Inferential control performance.	130

List of Symbols

$f_k(\cdot)$ system transition function at time k

$h_k(\cdot)$ measurement function at time k

$p(\cdot)$ probability distribution function (pdf)

$q(\cdot)$ pdf for importance sampling function

u_k process input at time k

w_k^i normalized weight for i^{th} particle at time k

\tilde{w}_k^i unnormalized weight for i^{th} particle at time k

x_0 initial state

x_k system state at time k

\hat{x}_k state estimate at time k

\hat{x}_k^- prior state estimate at time k

x_k^i i^{th} particle at time k

$x_k^{i,-}$ i^{th} prior particle at time k

x_k^e state estimate error at time k

y_k observation at time k

y_k^n n^{th} observation at time k

\hat{y}_k observation estimate at time k

$\chi_{k,i}$ i^{th} sigma point at time k

ω_k process noise at time k

ν_k measurement noise at time k

ϕ_k regressor vector at time k

θ_k process model parameter

A_k, B_k, C_k, D_k linear system matrices at time k

F_k, H_k linearized system matrices at time k

$L_c(\cdot)$ constrained likelihood function

N particle sample size

N_o number of observation sources

N_{eff} effective particle sample size

Q_k process noise variance at time k

P_k state covariance at time k

R_k measurement noise variance at time k

$Tr(P)$ trace of matrix P

X_k $X_k = \{x_0, \dots, x_k\}$ states trajectory

X_k^i $X_k^i = \{x_0^i, \dots, x_k^i\}$ i^{th} particle trajectory

Y_k $Y_k = \{y_1, \dots, y_k\}$ observations trajectory

\mathbb{C}_k process constraint region at time k

$\mathbb{E}(c, P)$ ellipsoid with center c and positive-definite matrix P

\mathbb{X}_k state constraint region at time k

\mathcal{D}_k $\mathcal{D}_k = \{\mathcal{Y}_1, \dots, \mathcal{Y}_k\}$ multiple observations trajectory set

$\mathcal{N}(\mu, \sigma^2)$ Gaussian distribution with mean μ and variance σ^2

\mathcal{Y}_k $\mathcal{Y}_k = \{y_k^1, \dots, y_k^{N_o}\}$ multiple observations at time k

Ψ_s summation of ellipsoids

Ψ_i intersection of ellipsoids

Θ_k augmented process model parameter

List of Abbreviations

<i>AIC</i>	Akaike Information Criterion
<i>ARMAX</i>	Autoregressive Moving Average with eXogenous
<i>ARX</i>	Autoregressive eXogenous
<i>BJ</i>	Box-Jenkins
<i>CSTR</i>	Continuous Stirred Tank Reactor
<i>EKF</i>	Extended Kalman Filter
<i>EM</i>	Expectation-Maximization
<i>EPF</i>	Extended Kalman Particle Filter
<i>ESMF</i>	Extended Set-Membership Filter
<i>HMM</i>	Hidden Markov Model
<i>IPS</i>	Inclined Plates Settler
<i>KF</i>	Kalman Filter
<i>MAP</i>	Maximum <i>a posteriori</i>
<i>MCMC</i>	Markov Chain Monte Carlo
<i>MHE</i>	Moving Horizon Estimator
<i>NFL</i>	Natural Froth Lubricity
<i>NLARX</i>	Non-linear ARX

<i>NLP</i>	Nonlinear Programming
<i>NMR</i>	Nuclear Magnetic Resonance
<i>ODE</i>	Ordinary Differential Equations
<i>OE</i>	Output-Error
<i>OLS</i>	Ordinary Least Squares
<i>PCA</i>	Principal Component Analysis
<i>PCR</i>	Principal Component Regression
<i>pdf</i>	Probability Density Function
<i>PF</i>	Particle Filter
<i>PLS</i>	Partial Least Squares
<i>PSV</i>	Primary Separation Vessel
<i>QP</i>	Quadratic Programming
<i>SMF</i>	Set-Membership Filter
<i>SMC</i>	Sequential Monte Carlo
<i>SIS</i>	Sequential Importance Sampling
<i>SIR</i>	Sequential Importance Resampling
<i>TE</i>	Tennessee Eastman
<i>UKF</i>	Unscented Kalman Filter
<i>UPF</i>	Unscented Particle Filter

Chapter 1

Introduction

1.1 Motivation

In today's competitive process industries, the pressure to improve the performance of processing facilities is intensive. Modern industrial enterprises have invested significant resources for process automation to collect and distribute data, with the expectation that it will enhance productivity and profitability via better decision making. However, it is not uncommon that real-time information on critical process variables is unavailable due to various causes such as sensor reading errors, sensor failures, and sensor unavailability (Fortuna, 2007). For a typical example, in a polymerization reactor, measurement of moments of molecular weight distribution cannot be obtained frequently because of high costs and long analysis times involved in measurement using gel permeation chromatography. In such cases, extracting useful hidden variable information using measured variables, process models, and/or expert knowledge is becoming more and more important to sustain plant safety, productivity and profitability. This need is general for many process engineering tasks including, process control, process monitoring, fault detection and diagnosis. Due to the importance of these tasks, many methods have been developed under the name of state estimation (Lehmann and Casella, 1998; Simon, 2006).

An ideal approach for state estimation should have the following features. First, it should be capable of handling all kinds of data, models and prior

information. This includes missing, abnormal and multi-rate process data, first-principle and data-driven models with possibly systematic modeling inaccuracies, linear and nonlinear dynamics, physical constraints and prior knowledge of the investigated process. Second, the approach should be able to extract maximum information with mathematical rigor. Furthermore, it should permit efficient computation for on-line or off-line applications. Finally, this ideal approach should provide information about level of uncertainty (or confidence/probability) in the decision making.

In this thesis, we take completely statistical view and put an emphasis on how to use Bayesian approaches to solve state estimation problems. As opposed to the frequentist approach, all the variables in Bayesian approach are treated as random, and the inference of interested variables is based on their distributions conditional on observed data (Box and Tiao, 1973).

Recall that the well-known Bayes' theorem states that the posterior distribution of some signal B given some signal A is equal to the prior distribution of B times the likelihood of A given B , divided by a normalizing constant:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \quad (1.1)$$

The Bayesian method considers the contribution from both the observed data in the form of the likelihood $p(A|B)$ and the prior knowledge in the form of the prior distribution $p(B)$. As the data become more abundant, then for any unknown variables which are observable from the measured data we expect the posterior distribution $p(B|A)$ to become progressively more concentrated around a single value. In some sense, Bayes' theorem reflects dynamics of learning and accumulation of the knowledge. Prior distribution encapsulates the state of our current knowledge, which can be updated after observing new data, and then posterior distribution reflects the change. When we observe another data then our current posterior distribution becomes prior for the new estimate. Thus every time using our current knowledge we estimate the state, observe data and store gained information in the form of new prior

knowledge. The sequential nature of Bayesian approach elegantly reflects the learning dynamics, and which will be introduced in Chapter 2.

A primary difficulty in the application of Bayes' theorem is the need to perform extensive computation of integration. Despite significant efforts from both scientific and engineering communities, solving Bayesian state estimation problem effectively and efficiently has been quite challenging, especially in the case of nonlinear and non-Gaussian problems, where analytical solution is intractable; most of the developed methods rely on a variety of simplifying assumptions. For example, extended Kalman filtering (EKF) relies on Gaussian approximation and local linearization to find a computationally efficient solution. Unscented Kalman filtering (UKF) avoids the linearization step, but relies on Gaussian assumption approximated by a set of deterministic points. Ensemble Kalman filter (EnKF) uses sampling technique for nonlinear estimation, but it assumes that all probability distributions involved are Gaussian. Moving horizon estimator (MHE) mostly relies on Gaussian approximation so that a least-squares expression can be formulated, and the nature of multi-stage optimization incurs excessive computational burden for on-line applications. These simplifying assumptions may work fine for unconstrained linear dynamic systems without large uncertainties, but can be easily violated in nonlinear dynamic systems, with constraints or large process/observation/modeling uncertainties.

Among many available state estimation approaches, particle filter (PF) based on a rigorous Bayesian formulation that uses sequential Monte Carlo sampling has recently shown promises in providing accurate and efficient estimation for nonlinear and non-Gaussian problems (Ristic et al., 2004). As opposed to conventional Bayesian estimators, PF does not assume Gaussian distributions. It can be implemented for arbitrary (multimodal or highly skewed) posterior pdf, when faced with nonlinear models, non-Gaussian noises, or constrained problems.

However, application of PF to practical chemical engineering processes is still in infancy due to several outstanding issues. For example, process constraints commonly exist in engineering practices, e.g., non-negative concentration or mass balance equation, but conventional PF does not take constraints into account. Another challenging problem, like other Bayesian estimators, is that conventional PF requires availability of an accurate process model and known noise characteristics, which are not realistic in practical applications. Therefore, it is indeed necessary to develop a novel PF algorithm that is robust against model uncertainties or unknown disturbances. Furthermore, in many practical processes, a single sensor is usually unable to provide full knowledge about the hidden state. It is necessary to achieve the fusion of several observations provided by multiple sensors, and these observations are likely to have multiple sampling rates, and each sensor's functioning condition may change, resulting in abnormal observations. This commonly happens because of, e.g., external environmental changes or sensor damage. In such cases, it is necessary to develop an estimation method that can synthesize multirate and multiple observations with abnormal observation detection to avoid dramatic estimation errors, and possibly improve estimation performance.

In summary, the motivation for this research arises from the following facts:

- (i) nonlinear and non-Gaussian Bayesian estimation is a challenging problem (Prakash et al., 2011), while it is of paramount importance in many practical applications;
- (ii) particle filtering approach is an emerging technique to handle nonlinear and non-Gaussian state estimation problems; however it is still in infancy for practical applications;
- (ii) constraints are commonly encountered in practical processes and this kind of prior knowledge can be used by PF for better estimation performance;

- (iii) unaccounted modeling inaccuracies or unknown process/measurement errors can degrade PF performance drastically, and there is a lack of effective, generalized robust PF estimation algorithm;
- (iv) multiple observations with different sampling rate processes are abundant in process industries, and it is important to develop a unified information synthesis approach to achieve better estimation performance.

Motivated by the above factors, this thesis intends to investigate particle filtering approach for nonlinear and non-Gaussian Bayesian state estimation problems of chemical engineering processes. In addition to the development of the particle filtering algorithms, industrial soft sensors are developed specifically for Oil Sands Extraction processes, where lack of suitable hardware instruments has been a critical challenge, and state estimation and soft sensor techniques play important roles.

1.2 Contributions

This thesis presents both theoretical development and industrial application oriented studies, in which the interplay between the theory and application provides interesting and valuable insights and allows for a balanced and systematic view of the investigated topic. The main contributions are listed below in the order of appearance.

1.2.1 Contributions via theoretical developments

The main theoretical contributions include:

- (i) extensive study of recursive Bayesian state estimation problems, with a detailed analysis of particle filtering algorithms;
- (ii) comparative study of constrained Bayesian state estimation problems, with proposal of novel constrained particle filter approaches;

- (iii) development of a robust particle filtering approaches in the presence of unknown but bounded uncertainties;
- (iv) detailed study of multi-rate data-driven process modeling with proposal of a practical Bayesian model calibration strategy based on PF approach.

1.2.2 Contributions via industrial applications

The main practical contributions include:

- (i) study of Oil Sands Extraction process for advanced process monitoring and systematic development of soft sensor design schemes for various extraction processes;
- (ii) on-line application of PF based soft sensor for an Inclined Plate Settler process;
- (iii) on-line application of PF based soft sensor and inferential control for a froth transportation process.

1.3 Organization

The rest of this thesis is organized as follows. Chapter 2 gives a general formulation of the recursive Bayesian estimation, and reviews existing commonly used approaches and the state of the art sequential Monte Carlo sampling based particle filtering approach. Comparative study of constrained Bayesian estimation is presented in Chapter 3, in which novel constrained particle filtering approaches are proposed to handle complicated state constraint. In Chapter 4, a robust particle filter approach is presented for tackling state estimation in the presence of unknown but bounded uncertainties. The proposed approach guarantees that the true state stays within a predefined particle sample set. Chapter 5 describes a Bayesian model calibration strategy using multiple-source observations. A practical robust estimation formulation is derived for handling abnormal observations and implemented within particle

filtering framework. Chapter 6 presents an industrial application, where soft sensor is developed for an oil sands froth transportation process and utilized for inferential control of a key quality variable. Finally, the thesis concludes in Chapter 7, with a discussion of the most important results and suggestions for future research both theoretically and practically.

Bibliography

- Box, G., Tiao, G., 1973. Bayesian Inference in Statistical Analysis. Wiley Classics.
- Fortuna, L., 2007. Soft Sensors for Monitoring and Control of Industrial Processes. Springer.
- Lehmann, E., Casella, G., 1998. Theory of Point Estimation. Springer.
- Prakash, J., Gopaluni, R. B., Patwardhan, S. C., Narasimhan, S., Shah, S. L., 2011. Nonlinear Bayesian state estimation: review and recent trends. In: The 4th International Symposium on Advanced Control of Industrial Processes.
- Ristic, B., Arulampalam, S., Gordon, N., 2004. Beyond the Kalman Filter: Particle Filters for Tracking Applications. Artech House.
- Simon, D., 2006. Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches. Wiley-Interscience.

Chapter 2

Review of Recursive Bayesian State Estimation

¹ State estimation deals with the problem of inferring knowledge about process variables (or state) indirectly measurable from a possibly noisy observation of a real world process, and the state is a physical quantity that affects the observation in a known manner represented by a certain process model. In recursive estimation, the inferred knowledge about the state is updated continuously as new measurements are collected. This recursive processing of observation is suitable in problems where the state dynamics change with time, or when the application demands most updated estimates based on the sequence of measurements observed so far. With the Bayesian view on estimation, both the state and the observation are stochastic entities. This fundamental paradigm yields a unifying framework for estimation problems where the inference result is a conditional density function of the states given the observational outcome.

This chapter is a review of recursive Bayesian estimation theory and it serves as a theoretical platform for the sequel of the thesis.

2.1 Problem Statement

Consider a discrete time system given by

$$x_k = f_k(x_{k-1}, u_{k-1}) + \omega_{k-1}, \quad (2.1)$$

1. Part of this chapter has been published in “X. Shao, B. Huang, J.M. Lee, Practical issues in particle filters for state estimation of complex chemical processes. IFAC SysId, 2009.”

$$y_k = h_k(x_k) + \nu_k, \quad (2.2)$$

where x_k , u_k , y_k , ω_k and ν_k are state, input, output, process noise and measurement noise, respectively; $f_k(\cdot)$, $h_k(\cdot)$ are nonlinear functions; both ν_k and ω_k are white noise of possibly non-Gaussian; initial state x_0 may also follow a non-Gaussian distribution $p(x_0)$; the variables x_k , y_k , ω_k and ν_k are random, while the input term u_k is usually deterministic. For the simplicity, input term is dropped in the remainder of this chapter as it does not affect the derivations. Note that the system can then be alternatively presented in a probabilistic form as

$$x_k \sim p(x_k|x_{k-1}), \quad (2.3)$$

$$y_k \sim p(y_k|x_k), \quad (2.4)$$

2.2 Recursive Bayesian Estimation

The objective of Bayesian estimation is to reconstruct the conditional *a posteriori* probability density function (pdf) $p(X_k|Y_k)$, where $X_k = \{x_0, \dots, x_k\}$ is the vector of states up to time k , and $Y_k = \{y_1, \dots, y_k\}$ is the vector of noisy measurements up to time k .

For many problems, an estimate of the state is required at each time point. Hence a recursive estimation method to construct the posterior pdf, $p(x_k|Y_k)$, is needed. Using a recursive method, received data can be processed sequentially rather than as a batch, eliminating the need to store large amounts of data to be reprocessed at a later time. The solution is obtained by recursively solving the following equations based on the Bayes' rule, also known as recursive prediction and update procedures (Gordon et al., 1993):

$$p(x_k|Y_{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|Y_{k-1})dx_{k-1}, \quad (2.5)$$

$$p(x_k|Y_k) = \frac{p(y_k|x_k)p(x_k|Y_{k-1})}{\int p(y_k|x_k)p(x_k|Y_{k-1})dx_k} = \frac{p(y_k|x_k)p(x_k|Y_{k-1})}{p(y_k|Y_{k-1})}, \quad (2.6)$$

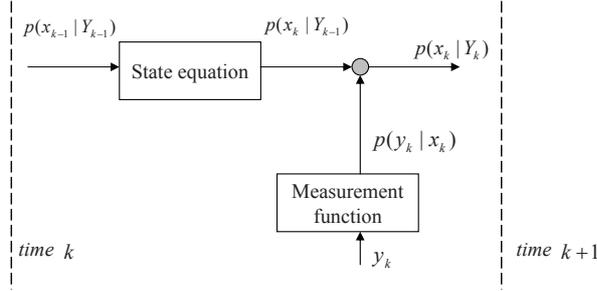


Figure 2.1: Illustration of recursive Bayesian state estimation.

where $p(x_k|Y_{k-1})$ is called prior distribution of x_k before the measurement is taken into account, $p(y_k|x_k)$ likelihood distribution of y_k given a certain x_k , and $p(y_k|Y_{k-1})$ normalizing constant.

The recursive algorithm of Bayesian estimation can be visualized as in Figure 2.1. Suppose that the posterior pdf at time $k - 1$, $p(x_{k-1}|Y_{k-1})$, is available. The prediction stage uses the probabilistic form of the state equation to obtain the prior pdf of the state at time k using the Chapman-Kolmogorov equation (Ristic et al., 2004) as shown in Equation (2.5). The update step is carried out at time k when the measurement y_k becomes available. The prior pdf is updated via Bayes' rule, as show in Equation (2.6).

It should be noted that two assumptions are used during the derivations of Equations (2.5) and (2.6):

(i) the states follow a first-order Markov process:

$$p(x_k|X_{k-1}, Y_{k-1}) = p(x_k|x_{k-1}), \text{ where } X_{k-1} = \{x_0, \dots, x_{k-1}\};$$

(ii) the observations are conditionally independent given the state:

$$p(y_k|X_k, Y_{k-1}) = p(y_k|x_k).$$

Since $p(x_k|Y_k)$ embodies all the statistical information contained in the observations about x_k , the posterior pdf $p(x_k|Y_k)$ is the complete solution of state estimation problem. After the posterior is available, the optimal point estimate, \hat{x}_k , corresponding to the loss function, $L(x_k, \hat{x}_k)$, may be obtained by optimizing the objective function

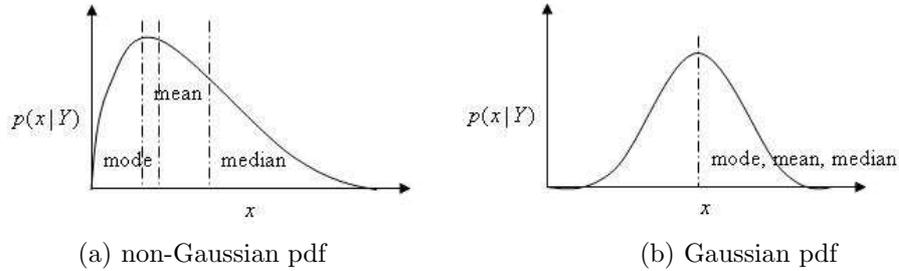


Figure 2.2: Illustration of point state estimation.

$$\min_{\hat{x}_k} E[L(x_k, \hat{x}_k)] = \min_{\hat{x}_k} \int L(x_k, \hat{x}_k) p(x_k|Y_k) dx_k \quad (2.7)$$

Bayesian estimation can use any loss function without changing its basic formulation and can readily provide error bounds. Various kinds of loss function exist, providing popular choice of optimal estimates, such as mode estimation (i.e., maximum of *a posteriori*, MAP), mean estimation (i.e., minimum variance, MV) or median estimation. Figure 2.2 demonstrates that these estimates are generally different except for Gaussian distributions.

It is important to note that, in general, there is no closed-form solution for Equations (2.5) and (2.6), as:

- (i) direct integration is computationally expensive and may not be practical for high-dimensional systems;
- (ii) the implementation of these equations requires storage of the entire pdf, possibly non-Gaussian, which in many cases is equivalent to an infinite dimensional vector.

Hence, for such intractable cases, approximations have to be made in order to proceed. Most estimation approaches address the challenge by making simplifying assumptions about the nature of the model and/or distributions at the cost of accuracy and computational efficiency. However, recent theoretical advance coupled with fast computation provides the foundation of building a feasible Bayesian approach even for large-scale systems. This computationally

efficient algorithm is based on sequential Monte Carlo sampling, also known as particle filtering, and will be discussed in Section 2.4.

2.3 Bayesian Interpretation of Existing Methods

This section provides a Bayesian view of existing methods by focusing on the approach for solving the Equations (2.5) and (2.6) in Section 2.2. Each method is interpreted as a variation of Bayesian estimation depending on approximations for making the solution more convenient. One common assumption underlying the existing methods is the Gaussian distributions for various pdfs, since closed-form solutions may be obtained where only two parameters, mean and variance, are required to describe the entire distribution. Although the assumption is often acceptable in linear unconstrained systems, it can be easily violated in nonlinear and/or constrained dynamic systems.

2.3.1 Kalman filtering based methods

Kalman-filter type of estimators are widely used for state estimation problems. These methods assume that all the system variables follow Gaussian distributions whose statistical information can be fully described by mean and covariance, and the estimate is given by

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - \hat{y}_k), \quad (2.8)$$

where \hat{x}_k^- is the prior estimate based on the information of $p(x_k|Y_{k-1})$, and K_k is the Kalman gain at time k . When $f_k(\cdot)$ and $h_k(\cdot)$ in Equations (2.1) and (2.2) are both linear functions, Gaussianity is kept all the time, and Kalman filter gives the optimal solution.

Extended Kalman filter

For nonlinear systems, Gaussianity is no longer guaranteed, and thus approximate solutions are needed. The most popular approximation method based

on model linearization is Extended Kalman Filter (EKF), in which the mean and covariance of the posterior distribution approximated as Gaussian are calculated as:

$$\begin{aligned}
\hat{x}_k^- &= f_k(\hat{x}_{k-1}), \\
F_k &= \left. \frac{\partial f_k}{\partial x} \right|_{\hat{x}_{k-1}}, \\
P_k^- &= F_k P_{k-1} F_k^T + Q_{k-1}, \\
\hat{y}_k &= h_k(\hat{x}_k^-), \\
H_k &= \left. \frac{\partial h_k}{\partial x} \right|_{\hat{x}_k^-}, \\
S_k &= H_k P_{k-1} H_k^T + R_k, \\
K_k &= P_k^- H_k^T S_k^{-1}, \\
\hat{x}_k &= \hat{x}_k^- + K_k (y_k - \hat{y}_k), \\
P_k &= (I - K_k H_k) P_k^-,
\end{aligned} \tag{2.9}$$

where Q_{k-1} and R_k are the covariance matrices of the system noise, ω_{k-1} , and measurement noise, ν_k , respectively.

The main disadvantages of EKF include: (i) approximated linear model can be inaccurate for highly nonlinear systems, in which estimate may fail to converge to the true state; (ii) update of covariance needs calculation of Jacobian matrices, which can be cumbersome in practice.

Unscented Kalman filter

Instead of approximating nonlinear models, Unscented Kalman Filter (UKF) approximates the posterior distribution by Gaussian distribution directly. It uses a set of deterministically chosen “sigma points” to represent mean and covariance. An estimation procedure for a fully augmented UKF is shown as

follows,

$$\begin{aligned}
\chi_{k-1} &= \left[\hat{x}_{k-1}^a \quad \hat{x}_{k-1}^a + \sqrt{(n^a + \kappa)P_{x_{k-1}}^a} \quad \hat{x}_{k-1}^a - \sqrt{(n^a + \kappa)P_{x_{k-1}}^a} \right], \\
\chi_{k,i}^{x,-} &= f_k(\chi_{k-1,i}^x) + \chi_{k-1,i}^\omega, \\
\hat{x}_k^- &= \sum_{i=0}^{2n^a} W_i^x \chi_{k,i}^{x,-}, \\
P_k^- &= \sum_{i=0}^{2n^a} W_i^c (\chi_{k,i}^{x,-} - \hat{x}_k^-)(\chi_{k,i}^{x,-} - \hat{x}_k^-)^T, \\
\gamma_{k,i} &= h_k(\chi_{k,i}^{x,-}, \chi_{k,i}^\nu), \\
\hat{y}_k &= \sum_{i=0}^{2n^a} W_i^x \gamma_{k,i}, \\
P_{y_k y_k} &= \sum_{i=0}^{2n^a} W_i^c (\gamma_{k,i} - \hat{y}_k)(\gamma_{k,i} - \hat{y}_k)^T, \\
P_{x_k y_k} &= \sum_{i=0}^{2n^a} W_i^c (\chi_{k,i}^{x,-} - \hat{x}_k^-)(\gamma_{k,i} - \hat{y}_k)^T, \\
K_k &= P_{x_k y_k} P_{y_k y_k}^{-1}, \\
\hat{x}_k &= \hat{x}_k^- + K_k (y_k - \hat{y}_k), \\
P_k &= P_k^- + K_k P_{y_k y_k} K_k^T,
\end{aligned} \tag{2.10}$$

where $\chi_{k-1} = [\chi_{k-1}^{xT} \quad \chi_{k-1}^{\omega T} \quad \chi_{k-1}^{\nu T}]^T$ is the vector of ‘‘sigma points’’ of the augmented state, $x_{k-1}^a = [x_{k-1}^T \quad \omega_{k-1}^T \quad \nu_{k-1}^T]^T$, with mean and covariance as

$$\begin{aligned}
\hat{x}_{k-1}^a &= [\hat{x}_{k-1}^T \quad 0 \quad 0]^T, \\
P_{x_{k-1}}^a &= \begin{bmatrix} P_{k-1} & 0 & 0 \\ 0 & Q_{k-1} & 0 \\ 0 & 0 & R_{k-1} \end{bmatrix}.
\end{aligned}$$

$n^a = n^x + n^\omega + n^\nu$ is the dimension of the augmented state; κ is a tuning parameter; W_i^x and W_i^c are weights for state and covariance.

Note that fully augmented UKF is not always necessary for all the situations, and reduction of computational complexity is possible for specific problems. Readers are referred to Kolås et al. (2009) for discussions on selections of UKF algorithms.

2.3.2 Moving horizon estimator

An alternative method for Bayesian approximation is to maximize a conditional *a posteriori* pdf for a sequence of the state trajectory,

$$\{\hat{x}_{k-h}, \dots, \hat{x}_k\} := \arg \max_{x_{k-h}, \dots, x_k} p(x_{k-h}, \dots, x_k | Y_k), \quad (2.11)$$

where $h \in \{0, k\}$ is known as a time horizon parameter.

Using Bayes' rule and Markov assumption, one can have

$$p(x_{k-h}, \dots, x_k | Y_k) \propto \prod_{j=k-h}^k p(y_j | x_j) \prod_{j=k-h}^{k-1} p(x_{j+1} | x_j) p(x_{k-h} | Y_{k-h-1}), \quad (2.12)$$

where $p(x_{k-h} | Y_{k-h-1})$ is the *a priori* information.

By assuming Gaussian distributions, a quadratic optimization problem can be formulated for solving Equation (2.12):

$$\begin{aligned} \min_{x_{k-h}^e, \hat{\omega}_{k-h}, \dots, \hat{\omega}_{k-1}} \quad & x_{k-h}^e{}^T P_{k-h}^{-1} x_{k-h}^e + \sum_{j=k-h}^{k-1} \hat{\omega}_j^T Q^{-1} \hat{\omega}_j + \sum_{j=k-h}^k \hat{v}_j^T R^{-1} \hat{v}_j \\ \text{s.t.} \quad & \hat{x}_{k-h} = \hat{x}_{k-h}^- + x_{k-h}^e, \\ & \hat{x}_{j+1} = f_j(\hat{x}_j) + \hat{\omega}_j, \quad j = k-h, \dots, k-1, \\ & y_j = h_j(\hat{x}_j) + \hat{v}_j, \quad j = k-h, \dots, k, \\ & \hat{x}_j \in \mathbb{X}, \quad \hat{\omega}_j \in \mathbb{W}, \quad \hat{v}_j \in \mathbb{V}. \end{aligned} \quad (2.13)$$

Equation (2.13) is known as Moving Horizon Estimator (MHE), which can be viewed as a form of iterative EKF (Bell and Cathey, 1993) for unconstrained system with a horizon size $h = 1$ (Rao, 2002). The advantage of MHE is that constraints for state or noise can be naturally incorporated into the problem formulation. However, the major problem for MHE is the computational load (see Robertson et al. (1996); Rao (2002); Rao and Rawlings (2002); Rawlings and Bakshi (2006); Zavala et al. (2008), and reference therein).

2.4 Particle Filter

Although the integrals in Equations (2.5) and (2.6) are intractable for nonlinear and non-Gaussian estimation problems, sampling methods can be used to numerically evaluate them.

Particle filter (PF) is a suboptimal Bayesian estimation algorithm that falls into the general class of Sequential Monte Carlo (SMC) sampling techniques. Interesting work in SMC integration methods was carried out by various individuals in the 1960s and 1970s (Ho and Lee, 1964; Yoshimura and Soeda, 1972; Akashi and Kumamoto, 1977). However, due to their severe computational complexity and the limited capability of computers, SMC algorithms have been neglected until recent years, especially after the introduction of the fundamental resampling step by Gordon et al. (1993). SMC algorithms have the great advantage of not being limited by nonlinearity and non-Gaussianity in the state model.

Unlike most other Bayesian estimators, particle filter does not rely on linearization technique or Gaussian assumption. It approximates a probability density by a set of samples or particles, x_k^i , and their associated weights, $w_k^i \geq 0$, in a discrete summation form:

$$\hat{p}(x_k|Y_k) = \sum_{i=1}^N w_k^i \delta(x_k - x_k^i), \quad (2.14)$$

where $\delta(\cdot)$ is the Dirac delta function, and N is the number of particles.

The ideal case for Monte Carlo sampling is to generate particles directly from the true posterior pdf $p(X_k|Y_k)$, which is unknown. Thus an easy-to-implement distribution, the so called *importance density* denoted by $q(X_k|Y_k)$, is defined before sampling, and the unnormalized importance weight for the sample drawn from $q(X_k|Y_k)$ is describes as

$$\tilde{w}_k^i \propto \frac{p(X_k^i|Y_k)}{q(X_k^i|Y_k)}. \quad (2.15)$$

Given the samples and associated normalized weights $\{X_{k-1}^i, w_{k-1}^i\}$ approximating the posterior density $p(X_{k-1}|Y_{k-1})$ at time $k-1$, choose the importance density so that it can be factorized as

$$q(X_k|Y_k) \triangleq q(x_k|X_{k-1}, Y_k)q(X_{k-1}|Y_{k-1}). \quad (2.16)$$

Using Bayes' rule one can express the posterior density at time k as:

$$\begin{aligned}
p(X_k|Y_k) &= \frac{p(y_k|X_k, Y_{k-1})p(X_k|Y_{k-1})}{p(y_k|Y_{k-1})} \\
&= \frac{p(y_k|X_k, Y_{k-1})p(x_k|X_{k-1}, Y_{k-1})p(X_{k-1}|Y_{k-1})}{p(y_k|Y_{k-1})} \\
&= \frac{p(y_k|x_k)p(x_k|x_{k-1})}{p(y_k|Y_{k-1})}p(X_{k-1}|Y_{k-1}) \\
&\propto p(y_k|x_k)p(x_k|x_{k-1})p(X_{k-1}|Y_{k-1}).
\end{aligned} \tag{2.17}$$

By substituting Equations (2.16) and (2.17) into (2.15), it can be shown that the weights associated with the samples at time k can be derived as

$$\tilde{w}_k^i \propto \tilde{w}_{k-1}^i \frac{p(y_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|X_{k-1}^i, Y_k)}. \tag{2.18}$$

The above equation provides a mechanism to sequentially update the importance weights based on a set of particles, therefore, by propagating the associated particles one can perform the recursive Bayesian estimation as each measurement is received sequentially. This method, called sequential importance sampling (SIS), forms the basis of most particle filtering methods. The SIS algorithm is presented in Algorithm 2.1. As the number of samples N becomes very large, the approximation, Equation (2.14), approaches the true posterior density and the SIS algorithm approaches the optimal Bayesian estimator.

Algorithm 2.1: The SIS algorithm

step a. initialization: generate initial particles $\{x_0^i\}_{i=1}^N$ from a *priori* distribution $p(x_0)$, and set $k = 1$;

step b. importance sampling: generate prior particles, $\{x_k^{i,-}\}_{i=1}^N$, from importance sampling distribution $q(x_k|X_{k-1}^i, Y_k)$;

step c. weighting: evaluate weights of each particle according to Eq. (2.18) once new measurement is available and normalize the weights as $w_k^i = \tilde{w}_k^i / \sum_{j=1}^N \tilde{w}_k^j$;

step d. output: estimate the state by calculating $\hat{x}_k = \sum_{i=1}^N w_k^i \cdot x_k^{i,-}$, set $k = k + 1$ and go back to step b.

Note that for particle filter output step, other than the weighted summation, one can also choose mode estimate (i.e., the particle with the largest weight) and robust mean estimate (i.e., weighted summation of particle subset around the mode).

Ideally the importance density function should be the posterior distribution itself. In a such case, the mean and variance of the importance weights will be 1 and 0, respectively. However, for most importance functions, the variance of importance weights will increase over time. The variance increase has a harmful effect on the accuracy and leads to a common problem with the SIS particle filter known as degeneracy problem. In practical terms this means that after certain number of recursive steps, all but one particle will have negligible weights, and a large computational effort will be devoted to updating particles whose contribution to the approximation of $p(x_k|Y_k)$ is almost zero. The degeneracy is difficult to avoid in the SIS framework and hence it was a major stumbling block in the development of sequential MC methods. A suitable measure of degeneracy of an algorithm is the effective sample size and can be estimated as follows:

$$N_{eff} = \frac{1}{\sum_{i=1}^N w_k^2} \quad (2.19)$$

It is straightforward to verify that $1 \leq N_{eff} \leq N$ with the following two extreme cases:

- (i) if the weights are uniform (i.e., $w_k^i = 1/N$ for $i = 1, \dots, N$) then $N_{eff} = N$;
- (ii) if there exists a $j \in \{1, \dots, N\}$ such that $w_k^j = 1$ and $w_k^i = 0$ for all $i \neq j$, then $N_{eff} = 1$.

Hence, small N_{eff} indicates a server degeneracy and vice versa.

To overcome the degeneracy problem, next subsection presents a resampling strategy to propagate the particles in regions with high probability.

2.4.1 Resampling

Whenever N_{eff} falls below a threshold, N_{thr} , resampling is required. The resampling procedure consists of regenerating particles according to the estimated pdf: eliminates samples with low importance weights and multiplies samples with high importance weights.

Resampling involves a mapping of random measure of prior particles into a random measure of posterior particles with uniform weights. The new set of random samples $\{x_k^j, \frac{1}{N}\}$ is generated by resampling (with replacement) N times from an approximate discrete representation of $p(x_k|Y_k)$ given by

$$p(x_k|Y_k) \approx \sum_{i=1}^N w_k^i \delta(x_k - x_k^{i,-}), \quad (2.20)$$

so that $P(x_k^j = x_k^{i,-}) = w_k^i$. The resulting sample is an independent identically distributed (i.i.d.) sample from the discrete density given in Equation (2.20), hence the new weights are uniform, i.e., $w_k^j = \frac{1}{N}$, and the approximation of $p(x_k|Y_k)$ becomes

$$p(x_k|Y_k) \approx \sum_{j=1}^N \frac{1}{N} \delta(x_k - x_k^j), \quad (2.21)$$

Figure 2.3 shows a schematic representation of multinomial resampling strategy (Douc et al., 2005), in which the left side of the figure represents the cumulative density function of the samples and right side shows the random variable $u_l \sim U[0, 1]$, which is mapped into the new sampling index j . Due to the high value of w_k^i , the corresponding particle $x_k^{i,-}$ has a good chance of being selected as the new sample x_k^j if we draw u_l uniformly.

For other resampling strategies, readers are referred to Boloc et al. (2004); Douc et al. (2005) and references therein.

Now that we have defined the main steps of a generic particle filter (GPF), and estimation steps are summarized as follows:

Algorithm 2.2: The GPF algorithm

step a. initialization: generate initial particles $\{x_0^i\}_{i=1}^N$ from *a priori* distribution $p(x_0)$, and set $k = 1$;

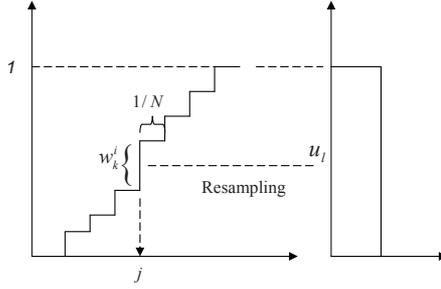


Figure 2.3: Illustration of multinomial resampling strategy.

step b. importance sampling: generate prior particles, $\{x_k^{i,-}\}_{i=1}^N$, from importance sampling distribution $q(x_k|X_{k-1}^i, Y_k)$;

step c. weighting: evaluate weights of each particle once new measurement is available and normalize the weights as $w_k^i = \tilde{w}_k^i / \sum_{j=1}^N \tilde{w}_k^j$;

step d. resampling: if $N_{eff} \leq N_{thr}$, then generate posterior particles, $\{x_k^i\}_{i=1}^N$, based on weighting information and resampling strategy, and set $w_k^i = 1/N$;

step e. output: estimate the state by calculating $\hat{x}_k = \sum_{i=1}^N w_k^i \cdot x_k^i$, set $k = k + 1$ and go back to step b.

Although the resampling step reduces the effects of degeneracy, it introduces other practical problems. If the particles that have high weights w_k^i are statistically selected many times it leads to a loss of diversity among the particles as the resultant samples will contain many repeated points. This problem, known as particle impoverishment, is severe when process noise in the state dynamics is very small. It leads to the situation where all particles will collapse to a single point within a few iterations. Also, since the diversity of the paths of the particles is reduced, any smoothed estimates based on the particles' path degenerate. Intentionally adding disturbance to the prior, or utilize Markov chain Monte Carlo (MCMC) move step, or regularization step, may reduce the impoverishment problem.

2.4.2 Choice of importance density

The choice of importance density $q(x_k|X_{k-1}^i, Y_k)$ is one of the most critical issues in the design of particle filter (Prakash et al., 2011; Shenoy et al., 2011). The samples are drawn from this distribution and it is used to evaluate the importance weights. The support of this proposal distribution should include the support of true posterior distribution. This distribution must also include the most recent measurement. The optimal importance density function that minimizes the variances of importance weights, conditioned upon x_{k-1}^i and y_k is given by Doucet et al. (2000)

$$\begin{aligned} q(x_k|X_{k-1}^i, Y_k)_{opt} &= p(x_k|x_{k-1}^i, y_k) \\ &= \frac{p(y_k|x_k, x_{k-1}^i)p(x_k|x_{k-1}^i)}{p(y_k|x_{k-1}^i)} \end{aligned} \quad (2.22)$$

Substitution of Equation (2.22) into Equation (2.18) yields

$$\tilde{w}_k^i \propto \tilde{w}_{k-1}^i p(y_k|x_{k-1}^i), \quad (2.23)$$

which states that importance density at time k can be computed before the particles are propagated to time k . In order to use the optimal importance function one has to be able to sample from $p(x_k|x_{k-1}^i, y_k)$ and evaluate $p(y_k|x_{k-1}^i) = \int p(y_k|x_k)p(x_k|x_{k-1}^i)dx_k$ up to a normalizing constant. In general, determining either of these two may not be a simple task.

However, there are some special cases where the use of the optimal importance density is possible. The first case is when x_k is a member of a finite set where the integral of $\int p(y_k|x_k)p(x_k|x_{k-1}^i)dx_k$ becomes a sum and sampling from $p(x_k|x_{k-1}^i, y_k)$ is possible. The second case is a class of models of which $p(x_k|x_{k-1}^i, y_k)$ is Gaussian.

The most popular suboptimal choice is using the transitional prior:

$$q(x_k|X_{k-1}^i, Y_k) = p(x_k|x_{k-1}^i). \quad (2.24)$$

If an additive zero-mean Gaussian process noise model is used the transitional prior is simply:

$$p(x_k|x_{k-1}^i) = \mathcal{N}(x_k; f_{k-1}(x_{k-1}^i), Q_{k-1}). \quad (2.25)$$

This choice of importance density does not satisfy the requirement that it must incorporate the latest measurement. However, it is easy to implement. One way of improving this importance density is to use a local estimator to update the particles using the latest measurement and use

$$q(x_k|X_{k-1}^i, Y_k) = \mathcal{N}(x_k; \hat{x}_k^i, \hat{P}_k^i) \quad (2.26)$$

where \hat{x}_k^i and \hat{P}_k^i are the estimates of the mean and covariance computed by a local estimator. This method for approximation of the importance density propagates the particles towards the likelihood function and consequently performs better than general PF. The additional computation cost of using such an importance density is often more than offset by reduction in the number of samples required to achieve a certain level of performance.

2.5 Case Studies

In this section, a two-state adiabatic Continuous Stirred Tank Reactor (CSTR) is studied first to test the effectiveness of the particle filter. Then we apply PF to the Tennessee Eastman (TE) process, which is a well known benchmark example for process monitoring and control. The results show that PF algorithm has potential to be applied in practical chemical engineering processes.

2.5.1 Two-state CSTR

Consider an adiabatic CSTR described by the following equations:

$$\begin{aligned} \frac{dC}{dt} &= \frac{q}{V}(C_0 - C) - kC e^{-E_A/T} \\ \frac{dT}{dt} &= \frac{q}{V}(T_0 - T) - \frac{\Delta H}{\rho C_p} kC e^{-E_A/T} - \frac{UA}{\rho C_p V}(T - T_c) \end{aligned} \quad (2.27)$$

where C is the concentration of product, T the temperature, q the flow rate, V the volume of the reactor, C_0 and T_0 inflow concentration and temperature, $kC_A e^{-E_A/T}$ the reaction rate, ΔH the reaction heat, ρ the density, C_p the specific heat, U and A the effective heat-transfer coefficient and area, respectively,

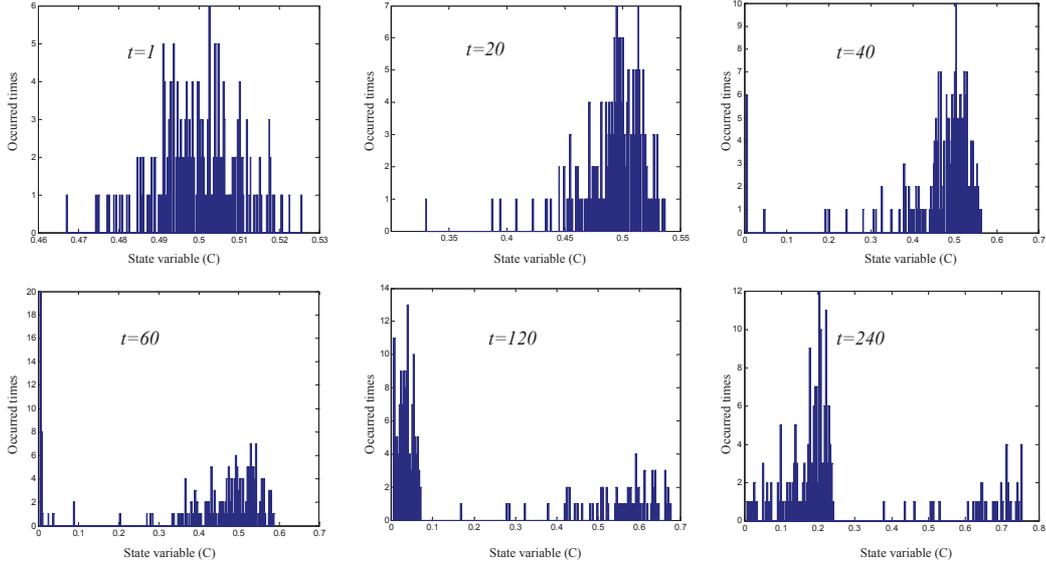


Figure 2.4: Illustration of non-Gaussian property for CSTR case.

and T_c the temperature of the coolant. The detailed parameter specifications can be found in Chen et al. (2004).

Only the temperature is routinely measured at one-second sampling interval, concentration is estimated based on the noisy temperature measurements, an poor initial guess is used, and the case is comparatively studied by EKF, UKF, MHE, and PF.

Figure 2.4 shows the dynamic evolution of the posterior distribution given a Gaussian initial guess. The non-Gaussian shapes of these distributions indicate that state estimation by Gaussian or other fixed-shape distribution can be inaccurate. Figures 2.5 and 2.6 illustrate the estimation results under different methods, i.e. EKF, UKF, MHE ($h = 2$), and generic PF ($N = 100$). From figures we can see that all the methods work fairly well as the simulation progresses. Particle filter is one of the best among these methods as it is more suitable for non-Gaussian estimation problems. An interesting point here is that EKF works very well and in fact even better than UKF. We believe that this is due to the fact that the linearization approximates the process very well and Gaussian noises are used.

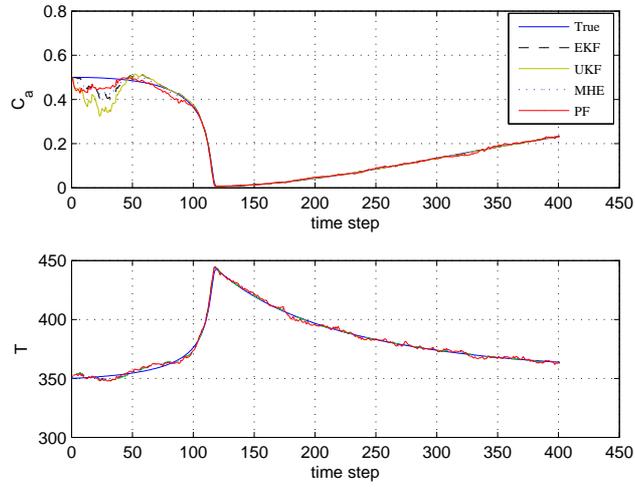


Figure 2.5: State estimation under different approaches for CSTR case.

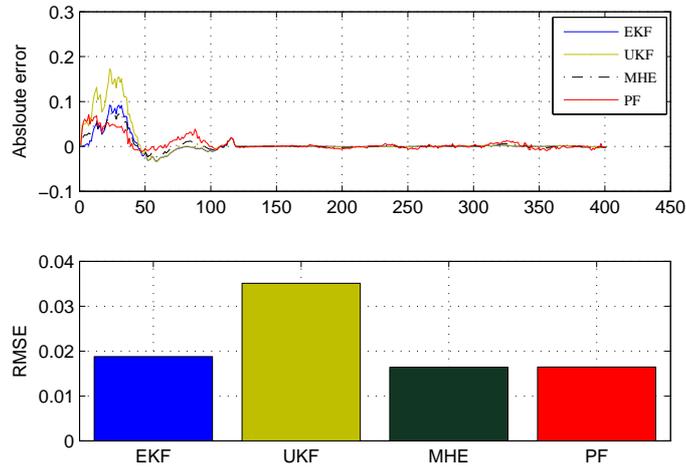


Figure 2.6: Absolute error and root-mean-square error (RMSE) comparisons under different approaches for CSTR case.

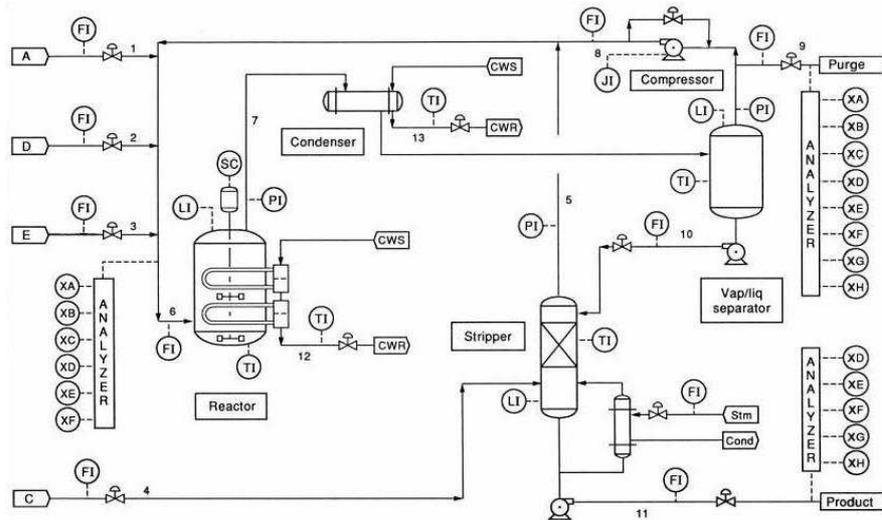
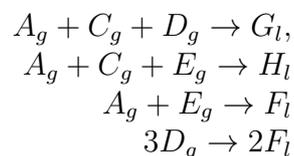


Figure 2.7: The Tennessee Eastman process flowsheet (Downs and Vogel (1993)).

2.5.2 Tennessee Eastman benchmark process

To illustrate the applicability of PF approach to practical chemical processes, a complicated, highly nonlinear and open-loop unstable system, the Tennessee Eastman process (see Figure 2.7), is considered. This process has been widely studied as a challenge problem in process control community.

The process consists of five main units: an exothermic two-phase CSTR, a vapor-liquid separator, a product condenser, a stripper and a recycle compressor. There are totally eight components present in the process, including two products, G and H , four reactants, A , C , D , E , one inert component, B , and one byproduct, F . The reactions are:



where the subscripts g and l denote gas or liquid phase, respectively. More details about the process can be found in Downs and Vogel (1993).

Since the plant is open-loop unstable, many papers have discussed stabilizing controllers for this process. In order to apply advanced control strategies, accurate estimation for interesting but unavailable or infrequently sampled

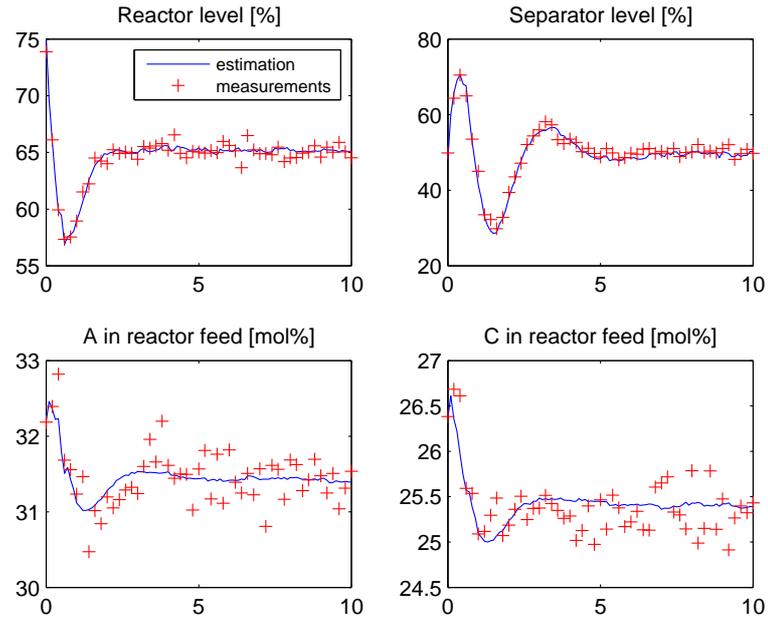


Figure 2.8: Estimation results for TE process (regular measurements).

variables are necessary.

The original TE plant has 41 measurements and 12 manipulated variables. In this chapter, we use a simplified model developed by Ricker and Lee (1995a) as the model for the estimator design. The simplified model contains 26 states, 10 manipulated variables, 16 outputs and 15 adjustable parameters, which are used to compensate the unmeasured disturbance and model error.

In our simulation, the TE process is stabilized by PI controllers described in Ricker and Lee (1995b). Figures 2.8 and 2.9 show sample results based on PF ($N = 500$) estimation after 10 hours of process operation. The first two subplots in Figure 2.8 show the reactor and separator levels, which are the measurements used to update the estimator every 6 *mins*; the second two subplots in Figure 2.8 show the component A and C in reactor feed (stream 6), which are also measured but not used for the estimator updating. The first two subplots in Figure 2.9 show the measurements of G and H for the products (stream 11), which are only available at every 15 *mins* and play important roles for quality control. The second two subplots in Figure 2.9 show the molar

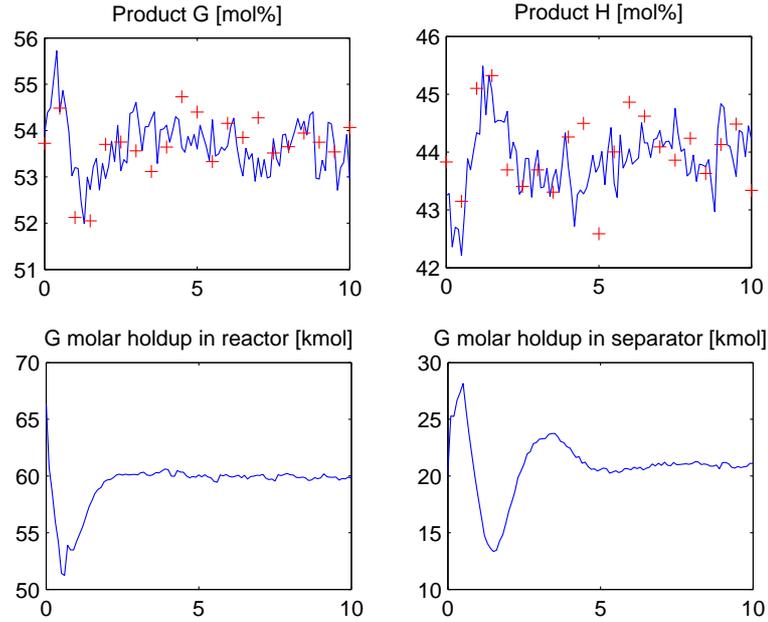


Figure 2.9: Estimation results for TE process (infrequent measurements and states).

holdup of product G in the reactor and separator, which are unmeasured states but important to be monitored. From these figures, we can see that the PF estimator can track the process dynamic responses. The CPU time needed for the simulation is generally in the range of 3 to 5 seconds; therefore, we conclude that PF approach is able to provide real-time estimation for the TE process.

2.6 Conclusions

This chapter introduced the problem of recursive Bayesian state estimation, and reviewed some existing approaches for non-linear Bayesian approximation. A novel particle filtering approach is introduced for non-linear and non-Gaussian cases. The approach is based on a rigorous Bayesian formulation that uses sequential Monte Carlo (SMC) sampling to propagate all information, while minimizing assumptions about the system and probability distribution functions. The resulting PF approach does not rely on common assumptions of

Gaussian or fixed-shape distributions, which are readily violated in nonlinear dynamic systems. Illustrative examples show that PF outperforms many commonly used estimation approaches, including EKF, UKF, MHE, and it has a good potential for real applications in complex chemical engineering processes.

In the next few chapters of this thesis, PF based Bayesian state estimation will be further discussed and industrial applications will be introduced.

Bibliography

- Akashi, H., Kumamoto, H., 1977. Random sampling approach to state estimation in switching environments. *Automatica* 13, 429–434.
- Bell, F., Cathey, B., 1993. The iterated Kalman filter update as a Gauss-Newton method. *IEEE Transactions on Automatic Control* 38(2), 294–297.
- Boloc, M., Djuric, P., Hong, S., 2004. Resampling algorithms for particle filters: A computational complexity perspective. *EURASIP Journal of Applied Signal Processing* 15, 2267–2277.
- Chen, W., Bakshi, B., Goel, P., Ungarala, S., 2004. Bayesian estimation via sequential Monte Carlo sampling: unconstrained nonlinear dynamic systems. *Industrial & Engineering Chemistry Research* 43, 4012–4025.
- Douc, R., Cappe, O., Moulines, E., 2005. Comparison of resampling schemes for particle filtering. In: *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*.
- Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential simulation-based methods for Bayesian filtering. *Statistics and Computing* 10(3), 197–208.
- Downs, J., Vogel, E., 1993. A plant-wide industrial process control problem. *Computers and Chemical Engineering* 17, 245–255.
- Gordon, N., Salmond, D., Smith, A., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: *IEE Proceedings F Radar and Signal Processing*.

- Ho, Y. C., Lee, R. C. K., 1964. A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control* 9(5), 333–339.
- Kolås, S., Foss, B., Schei, T., 2009. Constrained nonlinear state estimation based on the UKF approach. *Computers & Chemical Engineering* 33(8), 1386–1401.
- Prakash, J., Patwardhan, S. C., Shah, S. L., 2011. On the choice of importance distributions for unconstrained and constrained state estimation using particle filter. *Journal of Process Control* 21(1), 3–16.
- Rao, C., 2002. Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems. Ph.D. thesis, University of Wisconsin-Madison.
- Rao, C., Rawlings, J., 2002. Constrained process monitoring: Moving-horizon approach. *AIChE Journal* 48, 97–109.
- Rawlings, J. B., Bakshi, B. R., 2006. Particle filtering and moving horizon estimation. *Computers and Chemical Engineering* 30, 1529–1541.
- Ricker, N., Lee, J., 1995a. Nonlinear model predictive control of the tennessee eastman challenge process. *Computers and Chemical Engineering* 19, 961–981.
- Ricker, N., Lee, J., 1995b. Nonlinear modeling and state estimation for the tennessee eastman challenge process. *Computers and Chemical Engineering* 19, 983–1005.
- Ristic, B., Arulampalam, S., Gordon, N., 2004. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House.
- Robertson, D. G., Lee, J. H., Rawlings, J. B., 1996. A moving horizon-based

- approach for least-squares state estimation. *AIChE Journal* 42(8), 2209–2224.
- Shenoy, A. V., Prakash, J., McAuley, K. B., Prasad, V., Shah, S. L., 2011. Practical issues in the application of the particle filter for estimation of chemical processes. In: 18th IFAC World Congress.
- Yoshimura, T., Soeda, T., 1972. The application of monte carlo methods to the nonlinear filtering problem. *IEEE Transactions on Automatic Control* 17(5), 681–1684.
- Zavala, V., Larid, C., Biegler, L., 2008. A fast moving horizon estimation algorithm based on nonlinear programming sensitivity. *Journal of Process Control* 18(9), 876–884.

Chapter 3

Constrained Bayesian State Estimation

¹ Chapter 2 gives a review of recursive Bayesian estimation theory and techniques, including an introduction of particle filter for non-linear and non-Gaussian estimation. This chapter investigates constrained Bayesian state estimation problems by using particle filter (PF) approaches. Constrained systems with nonlinear model and non-Gaussian uncertainty are commonly encountered in practice. However, most of the existing Bayesian methods do not take constraints into account and require some simplifications. In this chapter, a novel constrained PF algorithm based on acceptance/rejection and optimization strategies is proposed. The proposed method retains the ability of PF in nonlinear and non-Gaussian state estimation, while take advantage of optimization techniques in constraints handling. The performance of the proposed method is compared with other accepted Bayesian estimators. Extensive simulation results from three examples show the efficacy of the proposed method in constraints handling and its robustness against poor prior information.

-
1. A version of this chapter has been published as “X. Shao, B. Huang, J.M. Lee, Constrained Bayesian State Estimation - a Comparative Study and a New Particle Filter Based Approach, *Journal of Process Control*, 20(2), pp.143-157, 2010.”

3.1 Introduction

Though nonlinear and non-Gaussian processes subject to state constraints are commonly encountered in practical applications, most of the existing Bayesian methods do not take constraints into account and require assumptions of linearity or Gaussianity. Therefore, development of Bayesian estimators that can handle nonlinear and non-Gaussian problems with constraints would be useful and has recently become an active research area (Rao, 2002; Vachhani et al., 2004; Haseltine and Rawlings, 2005; Vachhani et al., 2006; Rawlings and Bakshi, 2006; Kandepu et al., 2008; Kolås, 2008; Teixeira et al., 2009; Prakash et al., 2010).

Preliminary contributions of constrained PF can be found in Lang et al. (2007), in which an acceptance/rejection method is used for dealing with inequality constraint. However, the proposed approach is difficult to handle complicated constraints, such as nonlinear constraints, equality-inequality mixed constraints. Furthermore, the nature of the acceptance/rejection approach is simply removing all the particles outside the constraint region. In some cases (e.g., with a very poor prior estimate), the method could fail due to insufficient number of valid particles. Rajamani and Rawlings (2007) gives some preliminary discussions on the combination of PF and MHE; the method is based on optimization technique, which is more sophisticated to handle different types of constraints, but their optimization scheme is applied to the sample mean only, which may not take the full advantage of particle filter, especially when the posterior distribution is non-Gaussian. For instance, in a multimodal case, it is more appropriate to track and constrain individual particles instead of the sample mean only, since the mean could be located between the modes in a feasible region, but as a very poor estimate.

In this chapter, a constrained PF algorithm based on hybrid use of acceptance/rejection and optimization strategies is proposed. The proposed method combines the ability of PF to handle nonlinear and non-Gaussian problems

and the advantages of optimization techniques in constraints handling. Furthermore, simulation results show that the proposed method enhances the robustness of PF algorithm against poor prior information.

The remainder of this chapter is organized as follows: Section 2 introduces the constrained Bayesian state estimation problem. In Section 3, two constraint handling strategies are discussed within the generic PF framework, and a novel constrained PF algorithm is proposed. Three examples are illustrated in Section 4. Section 5 gives the conclusions.

3.2 Constrained Bayesian Estimation

In practical applications, constraints stem from the physical laws or model restrictions, e.g. non-negative mole fractions, limited liquid levels, mass balance, bounded parameters/disturbances, etc., and they are usually in the form of algebraic equality and inequality relationships, or simply upper and lower bounds. Incorporation of such constraints into estimation will be useful for improving estimation performance.

Take a multimodal posterior pdf (Figure 3.1) as an example; for the maximum *a posteriori* (MAP) state estimation where state x_k is concentration,

$$\hat{x}_k := \arg \max_{x_k} p(x_k | Y_k). \quad (3.1)$$

Mathematically, two optimal solutions can be obtained: one is negative (mode 1) and the other is positive (mode 2). From the knowledge on the constraint (i.e., $x \geq 0$), it is easy to find the correct estimate (mode 2).

Constraints of stochastic variables affect estimation by reshaping their pdfs in Bayesian framework. For instance, constraints on process noise restrict the transition distribution, $p(x_k | x_{k-1})$; constraints on measurement noise have an influence on the likelihood distribution, $p(y_k | x_k)$; and constraints on states alter the posterior distribution, $p(x_k | y_{1:k})$, as well as the transition and likelihood distributions (Ungarala et al., 2008). Therefore, use of these constraints confines the distributions, leading to improvement of estimation accuracy.

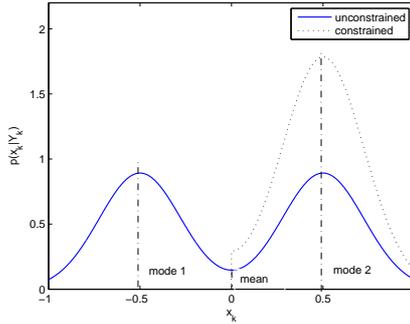


Figure 3.1: An example for a multimodal pdf.

A common way to handle constraints is “clipping” (Haseltine and Rawlings, 2005), where the estimated state is set equal to some predefined bounds if outside the constraint region. A more advanced way to solve constrained estimation problem is to use optimization techniques. Rao and Rawlings (2002) cast the constrained state estimation problem as a series of optimal control problems, and proposed to solve constrained estimation problem by using optimization techniques with a moving horizon fashion. However the MHE generally assumes Gaussian distribution and does not provide full distribution function of the estimated state. Vachhani et al. (2006) and Kolås (2008) adopted the optimization techniques into UKF framework to deal with constraints; however, the nature of the deterministic choice of “sigma points” in UKF restricts its applications for non-Gaussian problems.

3.3 Constrained Particle Filter

As shown in Chapter 2, the generic PF does not consider constraints. In this section, two methods are introduced to handle constraints in the PF framework, and then a new constrained PF algorithm is proposed. The work discussed here can be applied to variants of PF, such as Auxiliary Particle Filter (APF)(Pitt and Shephard, 1999), Unscented Particle Filter (UPF)(van der Merwe et al., 2000), and Kernel Particle Filter (KPF) (Cheng and Ansari, 2003).

3.3.1 Acceptance/Rejection

The nature of sample based representation of PF facilitates incorporating constraints into the estimation procedure. Lang et al. (2007) and Kyriakides et al. (2005) discuss how to accept/reject the particles in the PF algorithm based on constraint knowledge. As a minor modification from their work, a constrained likelihood function is defined as:

$$L_c(x_k^i, y_k^i, \omega_k^i, \nu_k^i) = \begin{cases} 1, & \text{if } \{x_k^i, y_k^i, \omega_k^i, \nu_k^i\} \in \mathbb{C}_k, \\ 0, & \text{if } \{x_k^i, y_k^i, \omega_k^i, \nu_k^i\} \notin \mathbb{C}_k, \end{cases} \quad i = 1, \dots, N, \quad (3.2)$$

where \mathbb{C}_k represents a constraint region at time k . Then the weight calculation step, Equation (2.19), is modified as

$$\tilde{w}_k^i = w_{k-1}^i \frac{p(y_k | x_k^i) \cdot L_c(x_k^i, y_k^i, \omega_k^i) \cdot p(x_k^i | x_{k-1}^i)}{q(x_k^i | X_{k-1}^i, Y_k)}. \quad (3.3)$$

This modification enables the algorithm to discard all the particles violating constraints. Figure 3.2 depicts an example for constraints on state. Take the equality constraint case, i.e. $g(x) = 0$, as an example, only the particles on the constraint surface will be accepted and reproduced, and all the rest particles will be rejected.

The advantage of acceptance/rejection scheme is twofold. First, it guarantees the particles to stay in constraint region and nearly no extra computation cost is needed. Second, the method retains the Monte Carlo sampling feature of PF which makes it suitable for non-Gaussian problems. However, the disadvantage is that it reduces the number of particles and may yield poor estimation. With poor prior information or complicated constraint conditions (e.g., nonlinear constraints), it is possible that all the particles lie outside the constraint region, which fails the PF algorithm.

3.3.2 Optimization formulation

A more systematic way to deal with constraints without discarding any particles is to employ optimization technique. In this section, optimization methods used to handle constraints are discussed in PF framework.

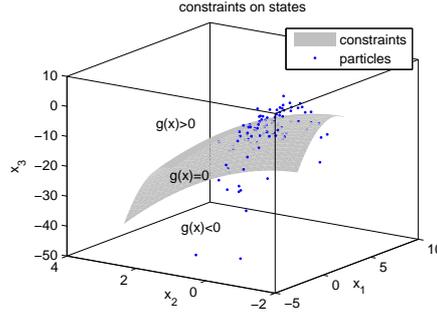


Figure 3.2: An example for constraints on states.

Interpretation of Bayesian estimation as an optimization problem

The estimate in Equation (3.1) can be further written as

$$\begin{aligned}
 \hat{x}_k &:= \arg \max_{x_k} p(x_k | Y_k) \\
 &\propto \arg \max_{x_k} p(y_k | x_k) p(x_k | Y_{k-1}) \\
 &= \arg \max_{x_k} (p_{\nu_k}(y_k - h_k(x_k)) p_{x_k^e}(x_k - x_k^-)).
 \end{aligned} \tag{3.4}$$

Note that the measurement noise ν_k follows distribution p_{ν_k} . Let $x_k = x_k^- + x_k^e$, where x_k^- is the optimal estimate of x_k according to $p(x_k | Y_{k-1})$, and x_k^e is the estimation error which follows distribution $p_{x_k^e}$. Note that exponential and double exponential (Laplacian) distributions are usually used to prescribe pdfs of ω , ν and x (Kotz et al. (2001), p.278; Robertson and Lee (2002); Ungarala et al. (2008)).

The above equation can be rewritten as the following constrained optimization problem by taking negative logarithm:

$$\begin{aligned}
 \min_{x_k} & -\log(p_{x_k^e}(x_k - x_k^-)) - \log(p_{\nu_k}(y_k - h_k(x_k))) \\
 s.t. & \quad x_k = x_k^- + x_k^e, \\
 & \quad y_k = h_k(x_k) + \nu_k, \\
 & \quad x_k \in \mathbb{X}_k,
 \end{aligned} \tag{3.5}$$

where \mathbb{X}_k denotes a general state constraint region.

If both $p_{x_k^e}$ and p_{ν_k} are further assumed as Gaussian distributions, Equation

(3.5) becomes a constrained nonlinear least square problem:

$$\begin{aligned}
& \min_{x_k^e} x_k^{eT} P_k^{-1} x_k^e + \nu_k^T R_k^{-1} \nu_k \\
& s.t. \quad x_k = x_k^- + x_k^e, \\
& \quad y_k = h_k(x_k) + \nu_k, \\
& \quad x_k \in \mathbb{X}_k,
\end{aligned} \tag{3.6}$$

where P_k^{-1} and R_k^{-1} may be treated as the weighting matrices, which are quantitative measures of our belief in the prior estimate and the observation model, respectively. Note that for linear models without constraints, solution of Equation (3.6) is equivalent to the well known Kalman filter estimate (Jazwinski, 1970, p. 205-208).

State constraints imposed on particles

According to the steps of the general PF algorithm (Algorithm 2.2), state constraints in PF can be imposed onto prior particles (particles before resampling procedure), $x_k^{i,-}$, posterior particles (particles after resampling procedure), x_k^i , or estimated mean value, \hat{x}_k . The constrained optimization problem presented in Equation (3.5) can be adapted as one of the followings:

$$\min_{\tilde{x}_k^{i,-}} -\log(p_{x_k^e}(\tilde{x}_k^{i,-} - x_k^{i,-})), \tag{3.7}$$

$$\min_{\tilde{x}_k^{i,-}} -\log(p_{x_k^e}(\tilde{x}_k^{i,-} - x_k^{i,-})) - \log(p_{\nu_k}(y_k - h_k(\tilde{x}_k^{i,-}))), \tag{3.8}$$

$$\min_{\tilde{x}_k^i} -\log(p_{x_k^e}(\tilde{x}_k^i - x_k^i)) - \log(p_{\nu_k}(y_k - h_k(\tilde{x}_k^i))), \tag{3.9}$$

$$\min_{\tilde{x}_k} -\log(p_{x_k^e}(\tilde{x}_k - \hat{x}_k)) - \log(p_{\nu_k}(y_k - h_k(\tilde{x}_k))), \tag{3.10}$$

where the diacritic mark “ \sim ” placed above $x_k^{(\cdot)}$ indicates a projected particle/mean. Figure 3.3 shows an illustration of particle projection, in which the rectangle represents the state(or output) space where particles (or corresponding outputs) located; the ellipse in state space denotes the state constraint region. Each particle corresponds to one possible state trajectory. If a particle violates the constraint, such a particle will be brought within the constraint region to a most likely location based on Equations (3.7) to (3.10).

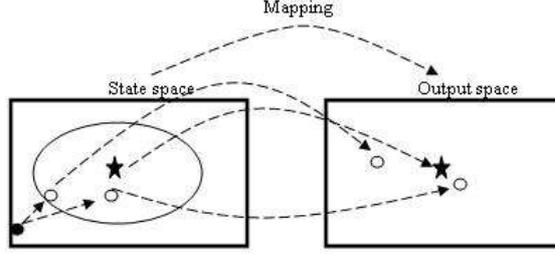


Figure 3.3: Illustration of projection (\circ : *valid particle*, \bullet : *violated particle*, \star : *true state/measurement*).

Generally, $p_{x_k^e}$ and $p_{\hat{v}_k}$ can be any distributions. However, for a tractable solution when dealing with constraints, truncated Gaussian, double half Gaussian or Gaussian mixture pdfs are often used to prescribe pdfs of noise and state particles during the implementation (Robertson and Lee, 2002; Rao, 2002; Kotecha and Djuric, 2003a,b). Hence, a quadratic form of objective function can be formed.

The sampling nature of PF has an advantage that covariance of estimated state error can be computed directly from samplers. For the prior particles, covariance can be estimated as:

$$\begin{aligned}\bar{x}_k^- &= \sum_{i=1}^N w_k^i x_k^{i,-}, \\ P_k^- &= \frac{\sum_{i=1}^N w_k^i (x_k^{i,-} - \bar{x}_k^-)(x_k^{i,-} - \bar{x}_k^-)^T}{1 - \sum_{i=1}^N w_k^i{}^2}.\end{aligned}\tag{3.11}$$

For posterior particles, all the weights are set uniform, then the sample covariance can be computed as

$$\begin{aligned}\bar{x}_k &= \frac{1}{N} \sum_{i=1}^N x_k^i, \\ P_k &= \frac{1}{N-1} \sum_{i=1}^N (x_k^i - \bar{x}_k)(x_k^i - \bar{x}_k)^T.\end{aligned}\tag{3.12}$$

Table 3.1 lists a summary of imposing constraints onto state particles for the generic PF algorithm.

Table 3.1: Constrained state particles for the generic PF

constrained particles	objective function
prior particles: $x_k^{i,-}$	Equations (3.7), (3.8)
posterior particles: x_k^i	Equation (3.9)
estimated mean: \hat{x}_k	Equation (3.10)

Constraints of other variables could also be imposed onto corresponding particles, such as estimated output, \hat{y}_k^i . The choice of which objective function to use, namely which step in Algorithm 1 to implement optimization, depends on the specific system and available computational resources.

Discussions

In the previous section, several variants of constrained PF algorithms have been presented based on optimization formulations. Illustrations of the differences of these formulations are shown in Figure 3.4. In the figures, the rectangle represents the space where particles are generated; the ellipse denotes the state constraint region.

As the figure shows, some of the prior particles are outside the constraint region. By using Equation (3.7), as shown in Figure 3.4(a), violated particles are projected onto the boundary, while particles that are already within the constraints remain unchanged. This equation is equivalent to “clipping”, which requires low computational load but probably yields poor performance.

Figure 3.4(b) shows that measurement information is used when imposing constraints onto prior particles. A trade-off between output error and state deviation is made to project the particles into a feasible region before resampling procedure. As in Figure 3.4(c), constraints are imposed onto particles after resampling procedure. Both Equations (3.8) and (3.9) reshape posterior distribution by projecting a set of particles, which could provide more accurate estimates when the state distribution is non-Gaussian. However, it requires much higher computational resource.

Equation (3.10) imposes constraints onto the estimated mean, as shown in

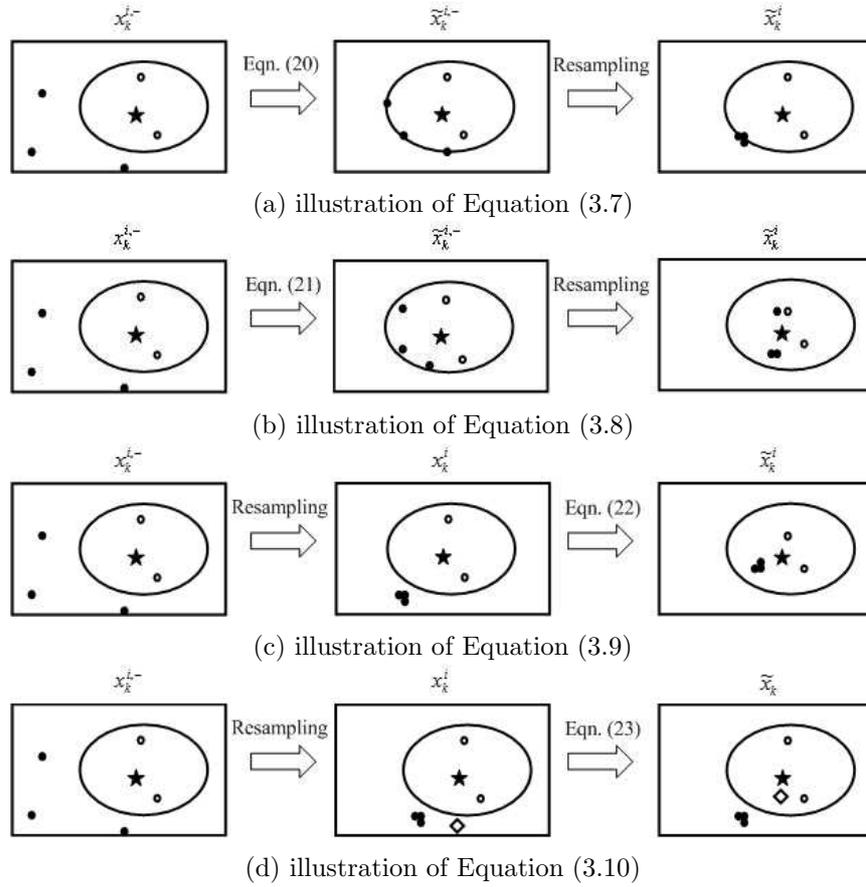


Figure 3.4: Illustration example of differences among Equations (3.7), (3.8), (3.9) and (3.10) (\circ : *valid particle*, \bullet : *violated particle*, \diamond : *estimated mean*, \star : *true state*).

Figure 3.4(d), without considering the constraints on each particle. Compared to Equations (3.8) and (3.9), it is computationally less demanding, but it has limitations due to the consideration of mean value only. For instance, the mean could be located between the modes without violating any constraint, but with a very low posterior probability. Note that a fixed-length moving horizon can be added straightforward, leading to a combination of PF and MHE (Rajamani and Rawlings, 2007).

3.3.3 Constrained PF algorithm

The main concern for the optimization based PFs is the online computation requirement. In order to reduce the computational cost and make the algorithm robust in the presence of poor prior information, a constrained PF algorithm based on hybrid use of acceptance/rejection and optimization strategies is proposed. The proposed scheme executes optimization only when the estimation performance based on the particles inside constraint region fails a performance test; otherwise, acceptance/rejection method (denoted as constrained PF1) is used.

In order to decide if the performance of the acceptance/rejection-based PF is satisfactory, a chi-square test is used. The rationale is that if the particles inside constraint region provide a good state estimate, the innovation term, $e_k = y_k - \hat{y}_k$, will have mean zero and covariance of Σ . In other words, the squared residual is checked if it follows a central chi-square distribution with p degree of freedom when the measurement error follows Gaussian distribution

$$e_k^T \Sigma^{-1} e_k \sim \chi^2(p), \quad (3.13)$$

where $e_k \sim \mathcal{N}(0, \Sigma)$, and $p = \dim(y)$ is the dimension of output.

Given past history data on estimation performance, a sliding time window l can be adopted in Equation (3.13):

$$\sum_{j=k-l+1}^k e_j^T \Sigma^{-1} e_j \sim \chi^2(l \times p). \quad (3.14)$$

To reduce the computational cost, the optimization procedure is executed only when Equation (3.13) (or Equation (3.14)) fails the statistical testing with a given significance level, e.g. $\alpha = 5\%$.

Note that, if the measurement error is assumed as non-Gaussian, the Chi-square test can be simply treated as a quadratic (2nd moment) test of the residual. In that case, a problem-specific threshold would be heuristically chosen instead of using Chi-square table.

The idea of the proposed constrained PF algorithm is summarized in the following, in which Equation (3.8) is chosen as an objective function for the optimization procedure (the algorithm is denoted as constrained PF2 in the following section). Similar algorithms based on Equations (3.9) and (3.10) (denoted as constrained PF3 and constrained PF4 respectively) are provided in the Appendix A.

Algorithm 3.1: A novel constrained PF algorithm based on Equation (3.8)

step a. initialization: generate initial particles $\{x_0^i\}_{i=1}^N$ from *a priori* distribution $p(x_0)$, and set $k = 1$;

step b. importance sampling: generate prior particles, $\{x_k^{i,-}\}_{i=1}^N$, from importance sampling distribution $q(x_k|X_{k-1}^i, Y_k)$;

step c. weighting: calculate constrained likelihood and importance weights according to Equations (3.2) and (3.3), then normalize the weights as $w_k^i = \tilde{w}_k^i / \sum_{j=1}^N \tilde{w}_k^j$;

step d. Chi-square test: calculate the weighted sample mean of the valid particles, $\bar{x}_k^- = \sum_{i=1}^{N_1} w_k^i x_k^{i,-}$, where N_1 is the particle number inside the constraint region; and compute the output residual, $e_k = y_k - h(\bar{x}_k^-)$; test the Chi-square criteria with a preset covariance Σ ;

step e. optimization: project the violated particles into constraint region by solving Equation (3.8) if performance test in step d fails; recalculate importance weights and normalization;

step f. resampling: if $N_{eff} \leq N_{thr}$, then generate posterior particles, $\{\tilde{x}_k^i\}_{i=1}^N$, based on resampling strategy, and set $w_k^i = 1/N$;

step g. output: estimate the state by calculating $\hat{x}_k = 1/N \cdot \sum_{i=1}^N \tilde{x}_k^i$, set $k = k + 1$ and go back to step b.

3.4 Case Studies

In order to investigate the efficacy of the proposed method, several examples with constraints on state are studied in this section. All the simulations were run on a 2.2 GHz CPU with 1 GB RAM PC using MATLAB 2008a. The mean square error (MSE) and CPU time presented below are based on 100 simulations.

3.4.1 Two-state batch reaction

Process description

Consider a gas-phase reaction well studied by Vachhani et al. (2006), Rawlings and Bakshi (2006), Ungarala et al. (2007), Kandepu et al. (2008) and Kolås (2008):



with a stoichiometric matrix

$$v = [-2, \quad 1],$$

and a reaction rate

$$r = kP_a^2.$$

The state and measurement vectors are defined as

$$x = [P_a \quad P_b]^T, \quad y = [1, \quad 1] x,$$

where P_j denotes the *nonnegative* partial pressure of species j . It is assumed that the ideal gas law holds and that the reaction occurs in a well-mixed

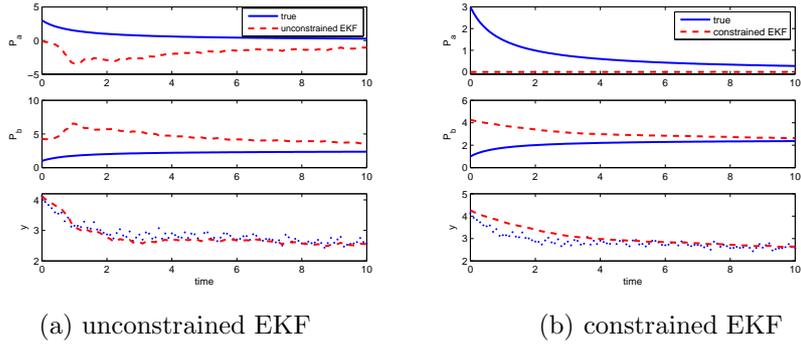


Figure 3.5: EKF estimates for example 1.

isothermal batch reactor. Then, from first principles, the process model can be written as

$$\dot{x} = f(x) = v^T r. \quad (3.15)$$

The system is discretized with a sampling interval of $\Delta t = 0.1s$, and simulated for 100 time steps from the initial condition $x_0 = [3, 1]^T$, and corrupted by Gaussian noise given by $\omega \sim \mathcal{N}\{[0, 0]^T, 10^{-6}I_2\}$, and $\nu \sim \mathcal{N}\{0, 10^{-2}\}$. Estimation starts from a poor initial guess $\bar{x}_0 = [0.1, 4.5]^T$ with a large covariance matrix $P_0 = 6^2I_2$.

This problem has been popularly studied in the literature because without considering constraints the state estimator can experience a multimodal pdf, which may lead to unphysical estimates.

Simulation results

The proposed constrained PF algorithms are tested on the reactor problem. For fair comparisons, $Q_{\hat{\omega}} = Q_{\omega}$, $R_{\hat{\nu}} = R_{\nu}$, and the same constraints and noise sequences are used for all the simulations in this example.

Figures 3.5(a) and 3.6(a) show that due to the poor initial guess and the multimodal nature, neither unconstrained EKF nor UKF converges to true states within the given simulation time despite good estimates of the output.

Figure 3.5(b) shows the estimate of the constrained EKF using clipping method; it does restrict the state to the constraint region, but the estimation

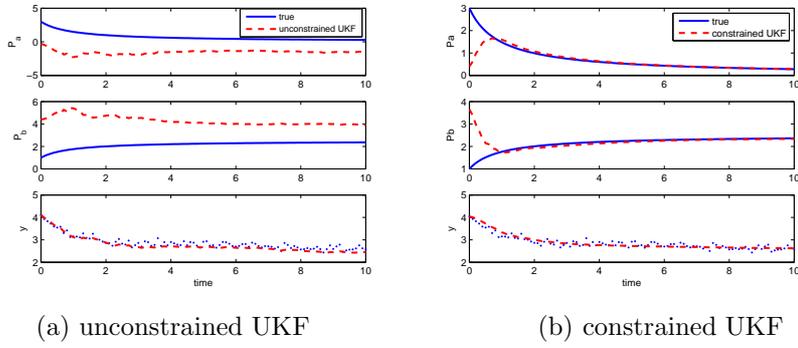


Figure 3.6: UKF estimates for example 1.

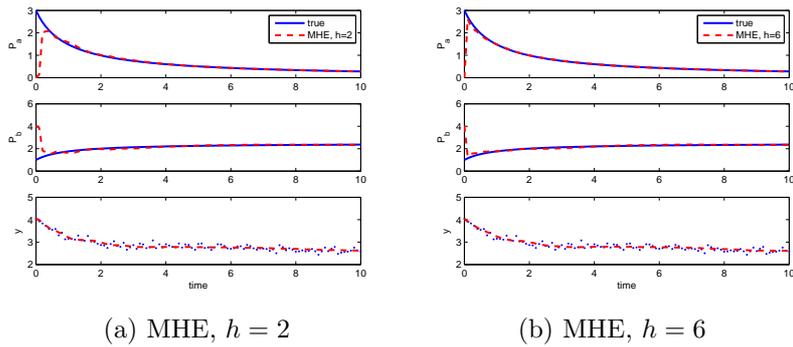


Figure 3.7: MHE estimates for example 1.

result is still poor. This is because the constraint knowledge was not properly used in updating the covariance matrix.

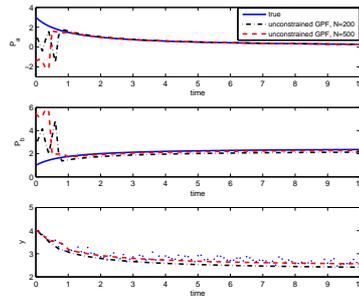
A quadratic programming (QP) based UKF, proposed by Kolås et al. (2009), is used to incorporate state constraints by constraining “sigma points”. Figure 3.6(b) shows the estimation performance is improved compared to the unconstrained case. However, a large increase of computation time is observed in solving the optimization problem

Compared to EKF/UKF based approaches, MHE provides improved estimates in terms of accuracy, see Figures 3.7(a) and 3.7(b), but computation time increases with the horizon size.

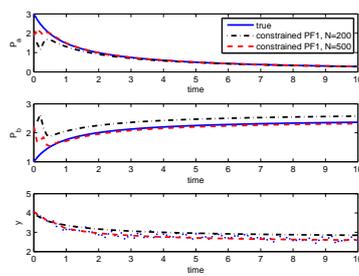
Figure 3.8(a) shows the estimation results of unconstrained generic PF with particle size $N = \{200, 500\}$. Compared to its counterparts of unconstrained EKF and unconstrained UKF, the Monte Carlo sampling based PF yields much more accurate estimate in this example; however, it still gives estimates violating physical constraints during initial time points. Compared to MHE, PF shows the advantage in computation time due to its single-horizon formulation.

Results of constrained PFs are shown in Figures 3.8(b) to 3.8(e), in which constrained PF1 denotes the constrained PF based on acceptance/rejection scheme (Lang et al., 2007); constrained PF2 denotes the constrained PF using hybrid scheme with optimization on prior particles (i.e. Equation (3.8)); constrained PF3 denotes the constrained PF using hybrid scheme with optimization on posterior particles (i.e. Equation (3.9)); and constrained PF4 denotes the constrained PF using hybrid scheme with optimization on estimated mean (i.e. Equation (3.10)). The figures show that all of these constrained methods provide physically valid estimates.

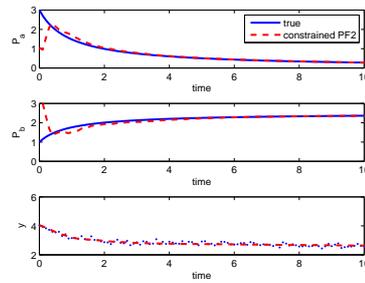
Table 3.2 shows the detailed performance comparisons. It can be seen that optimization-based methods generally yield better estimation, but with much higher computational cost. The table also shows that hybrid use of



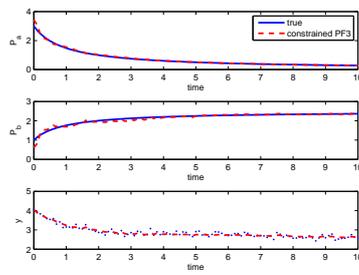
(a) unconstrained GPF



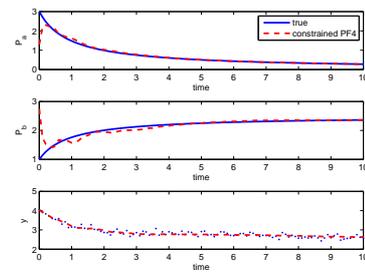
(b) constrained PF1



(c) constrained PF2



(d) constrained PF3



(e) constrained PF4

Figure 3.8: PF estimates for example 1.

Table 3.2: Comparison of estimation performances for example 3.4.1

Estimators		Schemes	MSE P_a	MSE P_b	CPU time (s)
EKF		N/A	6.8743	6.4232	1.547×10^{-4}
Constrained EKF		clipping	0.8431	1.3061	1.547×10^{-4}
UKF ($\alpha = 1, \beta = 2, \kappa = 1$)		N/A	4.6786	4.3234	3.282×10^{-4}
Constrained UKF		optimization	0.1532	0.1843	0.1213
MHE	h=2	optimization	0.1089	0.1186	0.1379
	h=6	optimization	0.0836	0.0949	0.8446
GPF	N=200	N/A	0.3907	0.4614	0.0192
	N=500	N/A	0.3614	0.3853	0.0493
Constrained PF1	N=200	accept/reject	0.0578	0.1496	0.0204
	N=500	accept/reject	0.0183	0.0242	0.0538
Constrained PF2 (N=50)		hybrid	0.0463	0.0565	0.0398
Constrained PF3 (N=50)		hybrid	0.0038	0.0055	0.0448
Constrained PF4	N=50, h=2	hybrid	0.0147	0.0192	0.0297
	N=50, h=6	hybrid	0.0043	0.0011	0.0929

constraint handling strategies provide the best estimates. In this example, the optimization was only necessary in the first few time steps to compensate the poor initial guess; for most time, acceptance/rejection procedure was used.

It should be also noted that: (i) choice of a particular method should depend on the available computational resource and the accuracy requirement; (ii) both unconstrained GPF and constrained PF1 are sensitive to the poor prior information; therefore, a larger particle size should be chosen. Methods based on the hybrid scheme are more robust; thus a smaller particle size can be used to reduce computational cost; (iii) optimization in constrained PF2 and constrained PF3 is not necessary to be applied to the whole particle set; for instance, optimization in constrained PF2 can be only applied to the prior particles violating constraints; optimization in constrained PF3 can be only applied to the parent particles (i.e. the subset particles selected for resampling); (iv) constrained PF4 is actually a combination of constrained PF1 and MHE; thus its estimates will be no poorer than MHE of the same horizon size.

3.4.2 Three-state batch reaction

Process description

Consider a batch reactor system adapted from Ungarala et al. (2008)



where $k = [k_1 \ k_2 \ k_3] = [0.06 \ 0.03 \ 0.001]$. The total number of moles remains constant in the reactor. A set of ODEs is used to describe the process dynamics,

$$\frac{dx}{dt} = \begin{bmatrix} -k_1 & k_2 & 0 \\ k_1 & -k_2 - k_3 & 0 \\ 0 & k_3 & 0 \end{bmatrix} x, \quad (3.16)$$

where $x = [x_A \ x_B \ x_C]^T$ is the vector of model fractions, which must obey the constraints as:

$$\begin{aligned} 0 &\leq x_i \leq 1, \\ \sum x_i &= 1. \end{aligned} \quad (3.17)$$

The system is discretized with a sampling interval of $\Delta t = 1$, and simulated for 50 time steps from the initial condition $x_0 = [1 \ 0 \ 0]^T$. A discretized process function can be obtained as

$$x_k = \begin{bmatrix} 1 - k_1 & k_2 & 0 \\ k_1 & 1 - k_2 - k_3 & 0 \\ 0 & k_3 & 1 \end{bmatrix} x_{k-1} + \omega_{k-1}, \quad (3.18)$$

where ω is zero-mean Gaussian noise with $Q_\omega = \text{diag}([0.01^2 \ 0.01^2 \ 0.0001^2])$. Noisy measurements of mole fractions are only available for species A and B :

$$y_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x_k + \nu_k, \quad (3.19)$$

where $\nu \sim \mathcal{N}(0, 0.02^2 I_2)$. The objective is to filter the measurements and estimate the unmeasured state x_C . Estimation starts from a poor initial guess $\bar{x}_0 = [0.8 \ 0.1 \ 0.1]^T$ with a covariance matrix $P_0 = \text{diag}([1^2 \ 1^2 \ 0.01^2])$.

Simulation results

Although it is a linear problem, this system is interesting to study because without considering the equality constraint, the system is not observable since

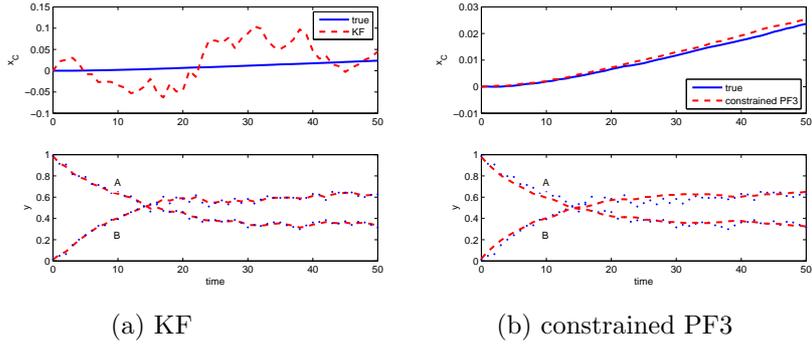


Figure 3.9: Simulation results for example 2.

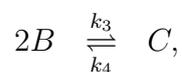
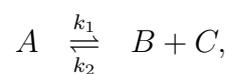
the measurement matrix does not have full column rank. One may estimate x_A and x_B by using Kalman filter with a two-dimension model and then compute x_C from the equality constraint condition. However, the result obtained in Figure 3.9(a) shows that the KF estimate of x_C can easily violate the non-negativity constraint.

Constrained PF based on acceptance/rejection scheme fails due to the poor prior information and the stringent constraint region. The optimization techniques allow for incorporating both equality and inequality constraints in Equation (3.17) into the estimation formulation. Constrained PF3 with particle size $N = 100$ is chosen as an estimator for this example. Figure 3.9(b) shows that the estimate accuracy has been significantly improved. The computation time is also reasonable, average CPU time is 0.04 seconds for each time step in this example.

3.4.3 Three-state continuous stirred-tank reaction

Process description

Consider a three-state CSTR gas-phase reaction taken from Haseltine and Rawlings (2005); Teixeira et al. (2008); Kolås et al. (2009)



$$k = [k_1 \quad k_2 \quad k_3 \quad k_4] = [0.5 \quad 0.05 \quad 0.2 \quad 0.01],$$

with a stoichiometric matrix

$$v = \begin{bmatrix} -1 & 1 & 1 \\ 0 & -2 & 1 \end{bmatrix},$$

and a reactional rate

$$r = \begin{bmatrix} k_1 C_A - k_2 C_B C_C \\ k_3 C_B^2 - k_4 C_C \end{bmatrix}.$$

The states and measurements are defined as to be

$$x = [C_A \quad C_B \quad C_C]^T,$$

$$y = [RT \quad RT \quad RT] x,$$

where C_j denotes the *nonnegative* concentration of species j , R is the ideal gas constant, T is the reactor temperature, and $RT = 32.84$. It is assumed that the ideal gas law holds. From first principles, the process model for a well-mixed, isothermal CSTR reactor is

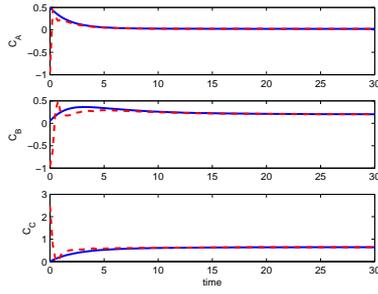
$$\dot{x} = f(x) = \frac{Q_f}{V_R} C_f - \frac{Q_0}{V_R} x + v^T r, \quad (3.20)$$

where $Q_f = Q_0 = 1$, $V_R = 100$ and $C_f = [0.5 \quad 0.05 \quad 0]$.

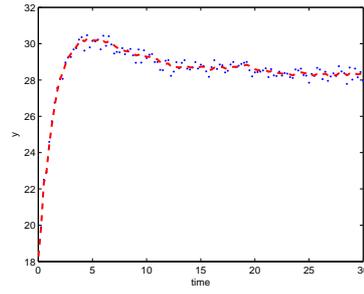
The system is discretized with a sampling interval of $\Delta t = 0.25$, and simulated for 120 time steps from the initial condition $x_0 = [0.5 \quad 0.05 \quad 0]$, and corrupted by Gaussian noise given by $\omega \sim \mathcal{N}([0 \quad 0 \quad 0]^T, 10^{-6} I_3)$ and $\nu \sim \mathcal{N}(0, 0.25^2)$. Estimation starts from a poor initial guess $\bar{x}_0 = [0 \quad 0 \quad 3.5]^T$ with a covariance matrix $P_0 = 4^2 I_3$.

Simulation results

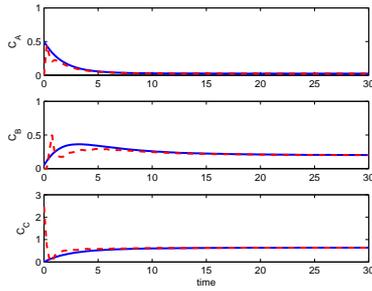
With $Q_{\hat{\omega}} = Q_{\omega}$, $R_{\hat{\nu}} = R_{\nu}$, and the same constraints and noise sequences, different nonlinear estimators are compared. Figure 3.10 shows EKF estimates, where neither unconstrained nor constrained EKF provides satisfactory results. Figure 3.11 shows the UKF estimates, where constrained UKF gives good results but with a huge increase of computation time; see Table 3.3 for



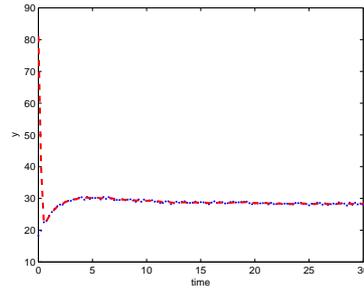
(a) unconstrained EKF (state)



(b) unconstrained EKF (output)

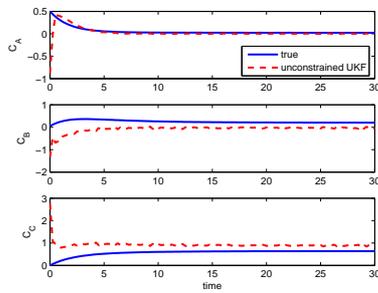


(c) constrained EKF (state)

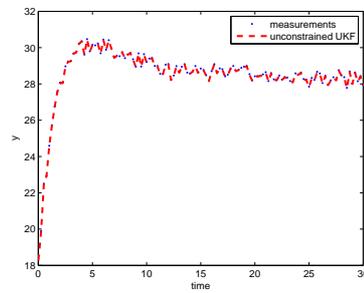


(d) constrained EKF (output)

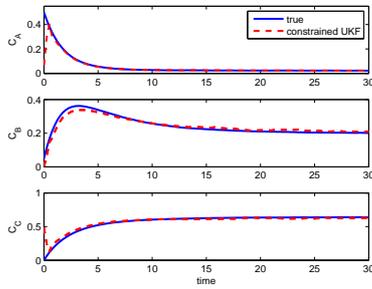
Figure 3.10: EKF estimates for example 3.



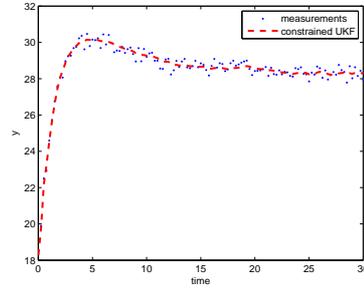
(a) unconstrained UKF (state)



(b) unconstrained UKF (output)



(c) constrained UKF (state)



(d) constrained UKF (output)

Figure 3.11: UKF estimates for example 3.

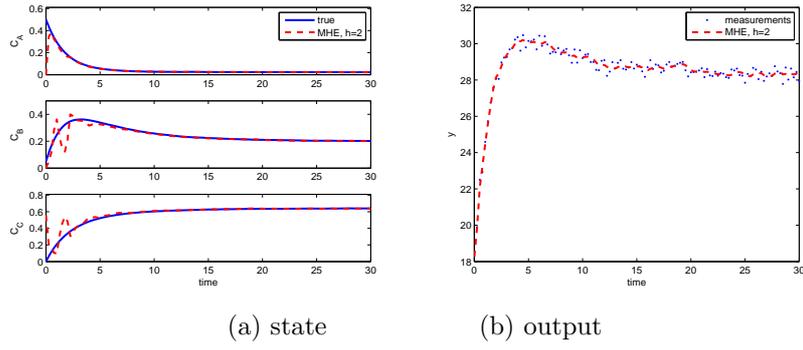


Figure 3.12: MHE estimates for example 3, $h = 2$.

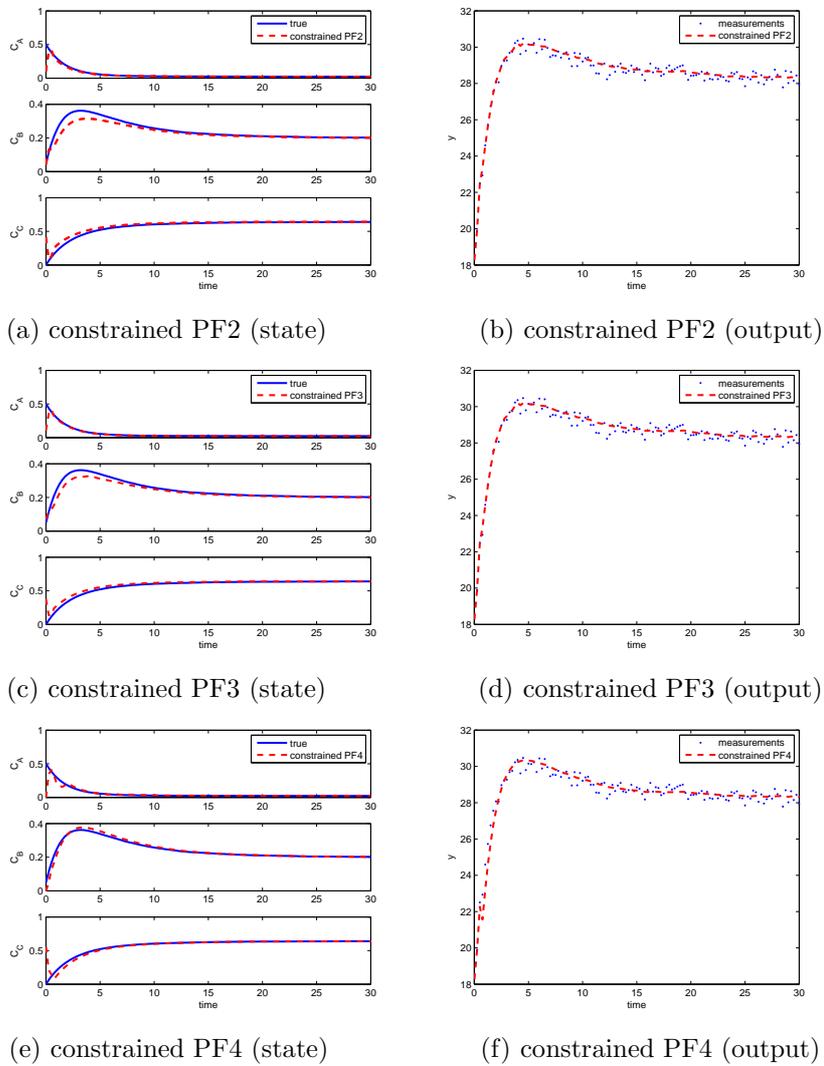


Figure 3.13: Optimization based constrained PF estimates for example 3.

Table 3.3: Comparison of estimation performances for example 3.4.3

Estimators	Schemes	MSE C_A	MSE C_B	MSE C_C	CPU time (s)	
EKF	N/A	0.0176	0.0150	0.0568	2.27×10^{-4}	
Constrained EKF	clipping	0.0024	0.0023	0.0568	2.27×10^{-4}	
UKF ($\alpha = 0.001, \beta = 2, \kappa = 0$)	N/A	0.0184	0.1322	0.1881	0.0012	
Constrained UKF	optimization	0.0017	2.17×10^{-4}	0.0025	0.5330	
MHE	$h = 2$	optimization	9.80×10^{-4}	0.0039	0.3436	
	$h = 6$	optimization	1.68×10^{-4}	0.0029	2.3896	
GPF ($N = 500$)	N/A				fail	
Constrained PF1 ($N = 500$)	accept/reject				fail	
Constrained PF2 ($N = 100$)	hybrid	0.0014	3.32×10^{-4}	0.0017	0.0379	
Constrained PF3 ($N = 100$)	hybrid	0.0014	2.23×10^{-4}	0.0017	0.0330	
Constrained PF4	$N = 100, h = 2$	hybrid	0.0023	2.71×10^{-4}	0.0026	0.0188
	$N = 100, h = 6$	hybrid	1.82×10^{-4}	1.31×10^{-4}	7.6×10^{-4}	0.1241

detailed comparisons. MHE with a horizon size of $h = 2$ also provides good estimates, see Figure 3.12. By increasing the horizon size, MHE estimates can become better, with a further increase of computational cost (see Table 3.3). Constrained PF1 with $N = 500$ fails in this example, due to the poor initial guess. Figure 3.13 shows the estimation results of the proposed constrained PF methods, which provides the best results for this example in terms of computation time and estimation accuracy.

From a large number of simulation runs, it was observed that constrained PF based on acceptance/rejection scheme requires the least amount of computation time, but it easily failed with poor prior information or stringent constraints. Under the same conditions, optimization-based constrained PFs yielded better estimation and showed more robustness; however, they require much higher computation time, which may not be suitable for on-line applications. Hybrid use of the acceptance/rejection and optimization schemes can combine the complementary advantages, and work more efficient in most situations.

3.5 Conclusion

Proper use of constraint knowledge is critical for the successful implementation of Bayesian estimators, since it can confine distribution domains of related variables, and make the estimation more accurate. In this chapter, two

different constraints handling strategies are discussed under the generic PF framework. Several new constrained PF algorithms are implemented based on hybrid use of acceptance/rejection and optimization schemes. Simulation results show that the proposed methods work efficiently for the investigated examples as they combine the advantages of Monte Carlo sampling nature of PF and the benefits of optimization techniques in handling constraints and poor prior information.

It is recommended that different methods should be considered depending on the available computational resource and the accuracy requirements. When one has good initialization knowledge with simple constraints, constraint PF1 (Lang et al., 2007) should be chosen; when one needs to handle complicated constraints with very limited computational resource, constraint PF4 with single horizon window should be considered; if computational cost is not the concern and the state distribution is believed as non-Gaussian, then constraint PF2 and PF3 may be selected.

The main contributions of this chapter are: (i) different and more efficient ways of incorporating state constraints in PF framework have been discussed and implemented; (ii) variant constrained Bayesian estimators are comparatively studied through several simulation examples. The proposed constrained PFs provide some interesting flexibility for constrained nonlinear/non-Gaussian Bayesian state estimation problems.

Bibliography

- Cheng, C., Ansari, R., 2003. Kernel particle filter: iterative sampling for efficient visual tracking. In: Proceedings of International Conference on Image Processing.
- Haseltine, E. L., Rawlings, J. B., 2005. Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Industrial & Engineering Chemistry Research* 44, 2451–2460.
- Jazwinski, A., 1970. *Stochastic Processes and Filtering Theory*. Academic Press, New York.
- Kandepu, R., Foss, B., Imsland, L., 2008. Applying the unscented Kalman filter for nonlinear state estimation. *Journal of Process Control* 18, 753–768.
- Kolås, S., 2008. Estimation in nonlinear constrained systems with severe disturbances. Ph.D. thesis, Norwegian University of Science and Technology.
- Kolås, S., Foss, B., Schei, T., 2009. Constrained nonlinear state estimation based on the UKF approach. *Computers & Chemical Engineering* 33(8), 1386–1401.
- Kotecha, J. H., Djuric, P. M., 2003a. Gaussian particle filtering. *IEEE Transactions on Signal Processing* 51(10), 2592–2601.
- Kotecha, J. H., Djuric, P. M., 2003b. Gaussian sum particle filtering. *IEEE Transactions on Signal Processing* 51(10), 2603–2613.

- Kotz, S., Kozubowski, T., Podgorski, K., 2001. The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance. Birkhäuser Boston.
- Kyriakides, I., Morrell, D., Papandreou-Suppappola, A., 2005. A particle filtering approach to constrained motion estimation in tracking multiple targets. In: The 39th Asilomar Conference on Signals, Systems and Computer.
- Lang, L., Chen, W., Bakshi, B., Goel, P., Ungarala, S., 2007. Bayesian estimation via sequential Monte Carlo sampling-constrained dynamic systems. *Automatica* 43, 1615–1622.
- Pitt, M., Shephard, N., 1999. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association* 94(446), 590–599.
- Prakash, J., Patwardhan, S. C., Shah, S. L., 2010. Constrained nonlinear state estimation using ensemble kalman filters. *Industrial & Engineering Chemistry Research* 49, 2242–2253.
- Rajamani, M., Rawlings, J., 2007. Improved state estimation using a combination of moving horizon estimator and particle filters. In: Proceedings of the 2007 American Control Conference.
- Rao, C., 2002. Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems. Ph.D. thesis, University of Wisconsin-Madison.
- Rao, C., Rawlings, J., 2002. Constrained process monitoring: Moving-horizon approach. *AIChE Journal* 48, 97–109.
- Rawlings, J. B., Bakshi, B. R., 2006. Particle filtering and moving horizon estimation. *Computers and Chemical Engineering* 30, 1529–1541.
- Robertson, D. G., Lee, J. H., 2002. On the use of constraints in least squares estimation and control. *Automatica* 38(7), 1113–1123.

- Teixeira, B. O. S., Chandrasekar, J., Trres, L. A. B., Aguirre, L. A., Bernstein, D. S., 2009. State estimation for linear and nonlinear equality-constrained systems. *International Journal of Control* 82(5), 918–936.
- Teixeira, B. O. S., Trres, L. A. B., Aguirre, L. A., Bernstein., D. S., 2008. Unscented filtering for interval-constrained nonlinear systems. In: *Proceedings of the 47th IEEE Conference on Decision and Control*.
- Ungarala, S., Dolence, E., Li, K., 2007. Constrained extended Kalman filter for nonlinear state estimation. In: *Proceedings of 8th International IFAC Symposium on Dynamics and Control Process Systems*.
- Ungarala, S., Li, K., Chen, Z., 2008. Constrained Bayesian state estimation using a cell filter. *Ind. Eng. Chem. Res.* 47, 7312–7322.
- Vachhani, P., Narasimhan, S., Rengaswamy, R., 2006. Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation. *Journal of Process Control* 16, 1075–1086.
- Vachhani, P., Rengaswamy, R., Gangwal, V., Narasimhan, S., 2004. Recursive estimation in constrained nonlinear dynamical systems. *AIChE Journal* 51, 946–959.
- van der Merwe, R., Doucet, A., de Freitas, N., Wan, E., 2000. The unscented particle filter. Tech. rep., Cambridge University Engineering Department.

Chapter 4

Robust Particle Filter for Unknown But Bounded Uncertainties

Despite the large number of papers published on the particle filter in recent years, one issue that has not been addressed to any significant degree is the robustness. For example, the standard approach to particle filter does not address the issue of robustness against modeling errors, or unknown process and measurement noises. This chapter presents a deterministic approach that has emerged in the area of robust filtering, and incorporates it into particle filtering framework. In particular, the deterministic approach is used to define a feasible set for particle sampling that contains the true state of the system, and makes PF robust against unknown but bounded uncertainties. Simulation results show that the proposed algorithm is superior to the standard particle filter and its variants such as the extended Kalman particle filter.

4.1 Introduction

In conventional particle filtering methods, a set of particles are drawn from the importance density (state transition density is mostly used in the generic particle filtering), which is the distribution of predicted state. If the predicted particles do not include the true state or the observations do not contain valuable information, filter distractions will occur and the estimated states

will gradually deviate from the true states, resulting in estimation divergence.

To improve the robustness of the PF estimation, several techniques have been proposed in the literature. For example, in de Freitas et al. (2000), the EKF Gaussian approximation is used as the proposal distribution for PF; van der Merwe et al. (2000) follows the similar idea, using the Unscented Kalman filter (UKF) as the proposal distribution; Rajamani and Rawlings (2007) propose to combine moving horizon estimation (MHE) with PF to improve the robustness of the algorithm. All of the mentioned strategies require the uncertainties expressed in terms of stochastic models. However, due to incomplete information of the noise statistics and the presence of systematic errors resulted from aggregation and obscurity of the process dynamics, a stochastic error based approach is questionable as many of these uncertainties are inherently non-stochastic. For example, in the case with considerable model-plant mismatch, the residual of the estimated model may have a component caused by deterministic structural errors, and purely random error assumptions can lead to unsatisfactory results.

This chapter studies an appealing alternative based on a deterministic approach by assuming that all of the uncertain quantities (including modeling errors, measurement noise, initial condition, process as well as future input perturbations, etc.) are unknown but bounded to a known set. In this case, all information about the system state is summarized by a set of possible states consistent with both observations and bound constraints on the uncertain quantities, and the true state is guaranteed to be in the resulted set. Under such a deterministic framework, the main interest consists of describing and constructing the feasibility set for particle sampling. The exact shape of such a set is, in general, very complicated and hard to obtain. Therefore, it is usually approximated by some simple geometry shapes, such as box, ball, ellipsoid, orthotope and zonotopes (Alamo et al., 2008). Among them the ellipsoidal estimation seems to be more popular because of its analogy to the

covariance in the stochastic methods (Schweppe, 1968) .

Much work has been done on the development of set membership approaches for linear system, extensions to nonlinear systems have also been made, but are limited in several ways. This chapter proposes a novel robust algorithm where particle filtering techniques and nonlinear set-membership approach are incorporated together in one framework; therefore, the advantages of each method are characterized in the new algorithm. To the best of our knowledge, there are few literatures reporting the synthesis of Monte Carlo sampling approach and nonlinear set membership theory. Simulation results show the proposed method guarantees a minimized outer bound on the particle set despite the model uncertainties as well as linearization errors.

The remainder of this chapter is organized as follows: Section 2 introduces preliminaries of the ellipsoidal techniques for set membership approach. In Section 3, ellipsoidal bound analysis is derived for nonlinear systems. Section 4 presents the novel robust PF based on the nonlinear set membership approach. Two examples are illustrated in Section 5. Section 6 gives the conclusions.

4.2 Preliminaries of Ellipsoidal Techniques

Denote a non-degenerate ellipsoid as:

$$\mathbb{E}(c, P) = \{x \in \mathbb{R}^n : (x - c)^T P^{-1}(x - c) \leq 1\} \quad (4.1)$$

where c is the center of the the ellipsoid, x is any point within the ellipsoid, and P is a positive-definite matrix that characterizes its shape and size.

(i) Summation of Two Ellipsoids

Assume that two ellipsoids are defined as $\mathbb{E}_1(c_1, P_1)$ and $\mathbb{E}_2(c_2, P_2)$, the summation of \mathbb{E}_1 and \mathbb{E}_2 is defined as

$$\begin{aligned} \Psi_s &= \mathbb{E}_1 \oplus \mathbb{E}_2 \\ &= \{x : x = x_1 + x_2, x_1 \in \mathbb{E}_1, x_2 \in \mathbb{E}_2\} \end{aligned} \quad (4.2)$$

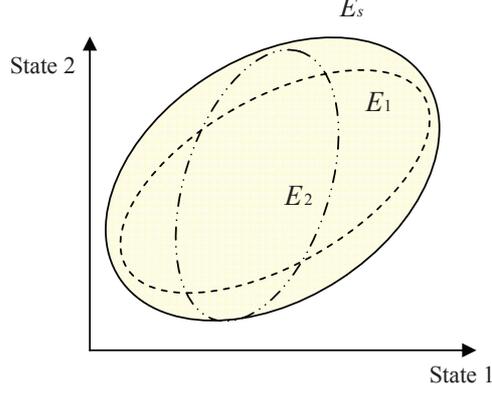


Figure 4.1: Geometry illustration of ellipsoid summation.

In most cases, Ψ_s is not an ellipsoid; its outer bounding ellipsoid is denoted as

$$\begin{aligned}
 \mathbb{E}_s &\ni \Psi_s \\
 \mathbb{E}_s &= \{x : (x - c_s)^T P_s^{-1} (x - c_s) \leq 1\} \\
 c_s &= c_1 + c_2 \\
 P_s &= \frac{P_1}{1 - \alpha} + \frac{P_2}{\alpha}
 \end{aligned} \tag{4.3}$$

where $\alpha \in (0, 1)$ is the scalar parameter depending on the optimality criterion chosen for the resultant ellipsoid \mathbb{E}_s (Chernousko, 1980; Becis-Aubry et al., 2008).

Figure 4.1 shows the geometry description of the vector sum of two ellipsoids. To obtain a compact \mathbb{E}_s , a computationally efficient criterion can be chosen to minimize the size of the ellipsoid. Commonly used optimality criterion includes the minimization of the volume ($Det(P_s)$) and the minimization of the sum of squared semiaxes (i.e., the trace of the positive-definite matrix $Tr(P_s)$). Take the minimal trace criterion for instance, the target is to find the minimum of

$$f(\alpha) = Tr(P_s) = Tr((1 - \alpha)^{-1} P_1 + \alpha^{-1} P_2) \tag{4.4}$$

The above function can be differentiated using the following formula:

$$\frac{d}{d\alpha} f(\alpha) = (1 - \alpha)^{-2} Tr(P_1) - \alpha^{-2} Tr(P_2) \tag{4.5}$$

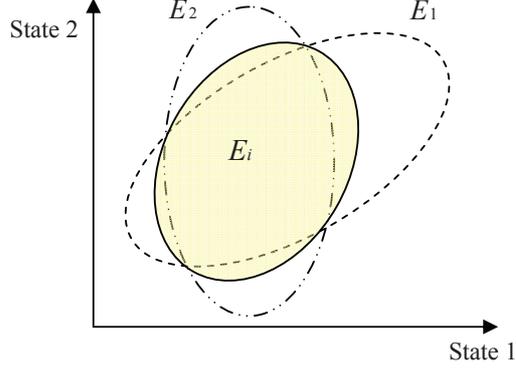


Figure 4.2: Geometry illustration of ellipsoid intersection.

Hence the optimal scalar parameter can be computed from $\frac{d}{d\alpha}f(\alpha) = 0$:

$$\alpha^* = \frac{\sqrt{\text{Tr}(P_2)}}{\sqrt{\text{Tr}(P_1)} + \sqrt{\text{Tr}(P_2)}} \quad (4.6)$$

(ii) Intersection of Two Ellipsoids

Assume that two ellipsoids are defined as $\mathbb{E}_1(c_1, P_1)$ and $\mathbb{E}_2(c_2, P_2)$, the intersection of \mathbb{E}_1 and \mathbb{E}_2 is defined as

$$\begin{aligned} \Psi_i &= \mathbb{E}_1 \cap \mathbb{E}_2 \\ &= \{x : x \in \mathbb{E}_1 \text{ and } x \in \mathbb{E}_2\} \end{aligned} \quad (4.7)$$

In most cases, Ψ_i is not an ellipsoid; its outer bounding ellipsoid is denoted as

$$\begin{aligned} \mathbb{E}_i &\ni \Psi_i \\ \mathbb{E}_i &= \{x : (x - c_i)^T P_i^{-1} (x - c_i) \leq 1\} \\ c_i &= c_1 + P_1 \left(P_1 + \frac{1-\rho}{\rho} P_2 \right)^{-1} (c_2 - c_1) \\ P_i &= \beta(\rho) \left(I - P_1 \left(P_1 + \frac{1-\rho}{\rho} P_2 \right)^{-1} \right) \frac{P_1}{1-\rho} \\ \beta(\rho) &= 1 - (c_2 - c_1)^T \left(\frac{P_1}{1-\rho} + \frac{P_2}{\rho} \right)^{-1} (c_2 - c_1) \end{aligned} \quad (4.8)$$

where $\beta(\rho) > 0$ for all $\rho \in (0, 1)$ when the ellipsoids have a non-empty intersection. Choice of ρ depends on the minimization criterion for the resultant ellipsoid \mathbb{E}_i (Schweppe, 1968; Becis-Aubry et al., 2008).

Figure 4.2 shows the illustration of ellipsoid intersection. Due to the complexity of the optimization, in this thesis the criterion for minimizing the upper bound of $\beta(\rho)$ is selected to provide a sub-optimal solution as:

$$\rho^* = \underset{\rho \in (0,1)}{\operatorname{argminsup}} \beta(\rho) \quad (4.9)$$

The upper bound of $\beta(\rho)$ is

$$\bar{\beta} = 1 - \frac{\|c_2 - c_1\|^2}{\frac{p_{1,max}}{1-\rho} + \frac{p_{2,max}}{\rho}} \quad (4.10)$$

where $p_{1,max} = \lambda_{max}(P_1)$ and $p_{2,max} = \lambda_{max}(P_2)$ are maximum singular values of the matrices P_1 and P_2 , respectively. The minimal upper bound can be computed as

$$\bar{\beta}_{min} = 1 - \frac{\|c_2 - c_1\|^2}{(\sqrt{p_{1,max}} + \sqrt{p_{2,max}})^2} \quad (4.11)$$

when

$$\rho^* = \frac{\sqrt{p_{2,max}}}{\sqrt{p_{1,max}} + \sqrt{p_{2,max}}} \in (0, 1). \quad (4.12)$$

The linear transform of $E(c, P)$ is defined as follows:

$$A(E(c, P)) = E(Ac, A^T P A). \quad (4.13)$$

4.3 Ellipsoidal Bound for Nonlinear Systems

Consider a nonlinear discrete system given by

$$\begin{aligned} x_k &= f(x_{k-1}) + \delta_{k-1} + \omega_{k-1}, \\ y_k &= h(x_k) + \nu_k, \end{aligned} \quad (4.14)$$

where $x_k \in R^{n_x}$ is the system state; $y_k \in R^{n_y}$ is the measurement output; $f(\cdot)$, $h(\cdot)$ are general nonlinear functions; δ_k represents an explicit systematic modeling error; ω_k and ν_k are the process and measurement noise, respectively.

It is assumed that the modeling error, process and measurement noises, as well as the initial state guess are unknown but bounded to ellipsoids:

$$\begin{aligned} \delta_{k-1} \in \mathbb{E}(0, \Delta_{k-1}) &\Leftrightarrow \delta_{k-1}^T \Delta_{k-1}^{-1} \delta_{k-1} \leq 1 \\ \omega_{k-1} \in \mathbb{E}(0, Q_{k-1}) &\Leftrightarrow \omega_{k-1}^T Q_{k-1}^{-1} \omega_{k-1} \leq 1 \\ \nu_k \in \mathbb{E}(0, R_k) &\Leftrightarrow \nu_k^T R_k^{-1} \nu_k \leq 1 \\ x_0 \in \mathbb{E}(\hat{x}_0, P_0) &\Leftrightarrow x_0^T P_0^{-1} x_0 \leq 1 \end{aligned} \quad (4.15)$$

At time step k , the goal is to characterize a set of states represented by a minimized ellipsoid that are consistent with the available measurements and *a priori* bound constraints; the true state is guaranteed to be contained in a resultant compact ellipsoid,

$$x_k \in \mathbb{E}(\hat{x}_k, P_k) \quad (4.16)$$

Note that no assumptions on the structure of the noise or state have been made except the bounds; hence, many types of uncertainties are included within this framework including Gaussian and non-Gaussian uncertainties.

Assuming that $f(\cdot)$ and $h(\cdot)$ are continuously differentiable, and for all estimated values \hat{x}_{k-1} or \hat{x}_k^- , Eq.(4.14) can be linearized using Taylor expansion,

$$\begin{aligned} x_k &= f(x_{k-1})|_{x_{k-1}=\hat{x}_{k-1}} + \frac{f^{(n_r)}(x_{k-1})}{n_r!}|_{x_{k-1}=\hat{x}_{k-1}}(x_{k-1} - \hat{x}_{k-1})^{n_r} \\ &\quad + R_f^{n_r}(x_{k-1} - \hat{x}_{k-1}) + \delta_{k-1} + \omega_{k-1} \\ y_k &= h(x_k)|_{x_k=\hat{x}_k^-} + \frac{h^{(n_r)}(x_k)}{n_r!}|_{x_k=\hat{x}_k^-}(x_k - \hat{x}_k^-)^{n_r} + R_h^{n_r}(x_k - \hat{x}_k^-) + \nu_k \end{aligned} \quad (4.17)$$

where $f^{(n_r)}(\cdot)$ and $h^{(n_r)}(\cdot)$ are n_r -th derivatives, and $R_f^{n_r}(\cdot)$ and $R_h^{n_r}(\cdot)$ are higher order remainder terms, which are equivalent to linearization errors.

Using interval analysis (Moore, 1966; Zemke, 1999), the Lagrange remainder term can be expressed as

$$R_f^{n_r}(x_{k-1} - \hat{x}_{k-1}) = \frac{f^{(n_r+1)}(\bar{X}_{k-1})}{(n_r + 1)!}(x_{k-1} - \hat{x}_{k-1})^{n_r+1} \quad (4.18)$$

where \bar{X}_{k-1} is the state interval bound in which $(x_{k-1} - \hat{x}_{k-1})$ is defined:

$$\bar{X}_{k-1}^i = \left[\hat{x}_{k-1}^i - \sqrt{P_{k-1}^{i,i}}, \hat{x}_{k-1}^i + \sqrt{P_{k-1}^{i,i}} \right], i = 1, \dots, n_x \quad (4.19)$$

For a one-state (i.e., $n_x = 1$) linearization case with first order approximation (i.e., $n_r = 1$), the state function in Equation (4.17) can be rewritten as

$$\begin{aligned} x_k &= f(x_{k-1})|_{x_{k-1}=\hat{x}_{k-1}} + \frac{\partial f(x_{k-1})}{\partial x}|_{x_{k-1}=\hat{x}_{k-1}}(x_{k-1} - \hat{x}_{k-1}) \\ &\quad + \frac{1}{2} \frac{\partial^2 f(\bar{X}_{k-1})}{\partial x^2}(x_{k-1} - \hat{x}_{k-1})^2 + \delta_{k-1} + \omega_{k-1} \end{aligned} \quad (4.20)$$

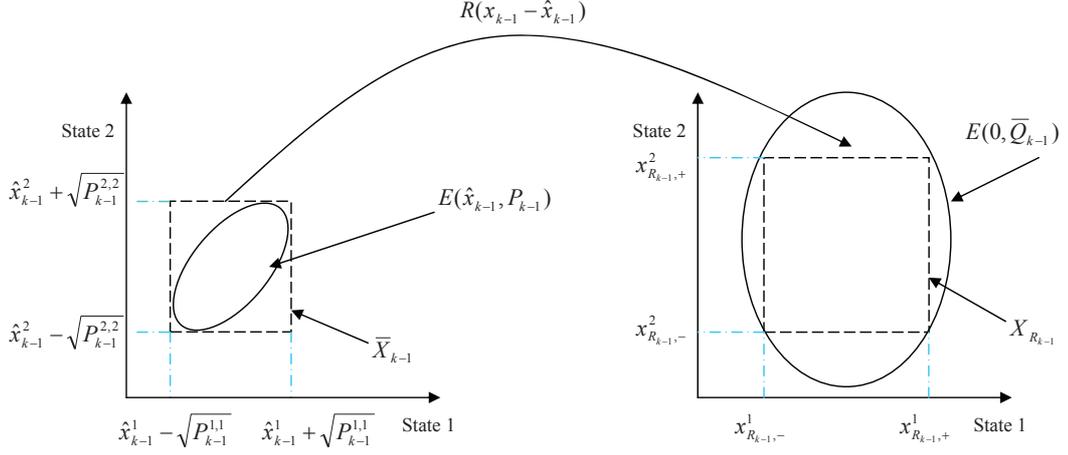


Figure 4.3: Illustration of ellipsoidal bound of linearization error.

For more general multi-state cases, according to Scholte and Campbell (2003) and Zhou et al. (2008), the linearization error is bounded to an ellipsoid $\mathbb{E}(0, \bar{Q}_{k-1})$, with

$$X_{R_{k-1}} = \frac{1}{2} \text{diag}(\bar{X}_{k-1}^T) \begin{pmatrix} Hes_1 \\ \vdots \\ Hes_n \end{pmatrix} \bar{X}_{k-1} \quad (4.21)$$

$$[\bar{Q}_{k-1}]^{i,i} = 2(X_{R_{k-1}}^i)^2, \quad [\bar{Q}_{k-1}]^{i,j} = 0 \quad (i \neq j)$$

where Hes_i represents the Hessian matrix of the nonlinear function $f(\cdot)$.

Figure 4.3 shows the illustration of ellipsoidal bound for the linearization error using interval mathematics for a two-state case.

Using the idea of ellipsoidal summation, the state function in Equation (4.17) can be simplified to

$$x_k = f(x_{k-1})|_{x_{k-1}=\hat{x}_{k-1}} + \frac{f^{(n_r)}(x_{k-1})}{n_r!}|_{x_{k-1}=\hat{x}_{k-1}} (x_{k-1} - \hat{x}_{k-1})^{n_r} + \hat{\omega}_{k-1} \quad (4.22)$$

where $\hat{\omega}_{k-1}$ incorporates modeling inaccuracies, linearization errors and process noise; its outer bound is defined as

$$\hat{\omega}_{k-1} \in \mathbb{E}(0, \hat{Q}_{k-1}) \supset \mathbb{E}(0, Q_{k-1}) \oplus \mathbb{E}(0, \bar{Q}_{k-1}) \oplus \mathbb{E}(0, \Delta_{k-1}) \quad (4.23)$$

$$\hat{Q}_{k-1} = (1 - \alpha_2)^{-1} \left(\frac{Q_{k-1}}{1 - \alpha_1} + \frac{\bar{Q}_{k-1}}{\alpha_1} \right) + \alpha_2^{-1} \Delta_{k-1} \quad (4.24)$$

where α_1 and α_2 are scalar parameters to be chosen according to Equation (4.6).

The linearization of the measurement function is dealt with in the same way, and the incorporated approximation error and measurement noise are bounded to an ellipsoid, i.e., $\hat{\nu}_k \in \mathbb{E}(0, \hat{R}_k)$.

The above analysis forms the basis for the development of the robust filtering algorithm to be introduced in the next section.

4.4 Guaranteed Robust Particle Filter

Ellipsoid based set-membership approach produces an entire set of states as the estimation result, in which the unknown true state is guaranteed to be contained. Normally, the center of the ellipsoid is selected as the point state estimate; however, this selection is not always appropriate. In this section, a novel particle filter (PF) based on the extended set-membership filtering (ESMF) approach is proposed. The combination of ESMF and PF allows the new algorithm to be able to incorporate the latest observations into a prior update routine. Furthermore, the ESMF generates proposal distributions that guarantee the inclusion of the true state, and as result, robust state estimation performance is achieved.

4.4.1 Extended Set Membership Filtering

Like the Bayesian stochastic estimator, set membership approach consists of prediction (time update) and correction (observation update) steps. The algorithm in this section establishes a recursive procedure for computing the sequence of ellipsoid $\mathbb{E}_k(\hat{x}_k, P_k)$.

Prediction:

Assume an ellipsoidal estimate $\mathbb{E}(\hat{x}_{k-1}, P_{k-1})$ is known at time $k - 1$. The prediction step at time k is carried out by linearly transforming the ellipsoid at time $k - 1$ to $\mathbb{E}(f(\hat{x}_{k-1}), F_{k-1}P_{k-1}F_{k-1}^T)$. This is followed by a vector sum of the resulting ellipsoid and the virtual process noise $\hat{\omega}_{k-1}$ to yield an outer

bounding ellipsoid $\mathbb{E}(\hat{x}_k^-, P_k^-)$:

$$\mathbb{E}(\hat{x}_k^-, P_k^-) \supseteq \mathbb{E}(f(\hat{x}_{k-1}), F_{k-1}P_{k-1}F_{k-1}^T) \oplus \mathbb{E}(0, \hat{Q}_{k-1}) \quad (4.25)$$

$$\hat{x}_k^- = f(\hat{x}_{k-1}) \quad (4.26)$$

$$P_k^- = \frac{F_{k-1}P_{k-1}F_{k-1}^T}{(1 - \alpha_k)} + \frac{\hat{Q}_{k-1}}{\alpha_k} \quad (4.27)$$

where the optimal α_k minimizing the bounding ellipsoid can be calculated as Equation (4.6).

Update:

Observation update step is to compute an ellipsoid containing the intersection of predicted ellipsoid $\mathbb{E}(\hat{x}_k^-, P_k^-)$ and the observation set S_k defined by

$$S_k = \{x \in R^n : (y_k - h(x))\hat{R}_k^{-1}(y_k - h(x)) \leq 1\} \quad (4.28)$$

The ellipsoid $\mathbb{E}(\hat{x}_k, P_k) \supset \mathbb{E}(\hat{x}_k^-, P_k^-) \cap S_k$ is the result of the observation based correction. It is essential that $\mathbb{E}(\hat{x}_k^-, P_k^-)$ and S_k have a non-empty intersection, i.e. that the predicted feasible set is consistent with y_k and the observation noise bounds. If not, bound tuning is needed.

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - h(\hat{x}_k^-)) \quad (4.29)$$

$$K_k = \frac{1}{1 - \rho_k} P_k^- H_k^T \left(\frac{H_k P_k^- H_k^T}{1 - \rho_k} + \frac{\hat{R}_k}{\rho_k} \right)^{-1} \quad (4.30)$$

$$P_k = \sigma_k^2 (I - K_k H_k) \frac{P_k^-}{1 - \rho_k} \quad (4.31)$$

$$\sigma_k^2 = 1 - (y_k - h(\hat{x}_k^-))^T \left(\frac{H_k P_k^- H_k^T}{1 - \rho_k} + \frac{\hat{R}_k}{\rho_k} \right)^{-1} (y_k - h(\hat{x}_k^-)) \quad (4.32)$$

where the value of ρ_k is solved by minimizing σ_k^2 as a sub-optimal solution expressed in Equation (4.12).

The linearization model is defined to be the following Jacobians

$$F_{k-1} = \frac{\partial f(x_{k-1})}{\partial x} \Big|_{x_{k-1}=\hat{x}_{k-1}}, \quad H_k = \frac{\partial h(x_k)}{\partial x} \Big|_{x_k=\hat{x}_k^-} \quad (4.33)$$

Note that the linearization and virtual noise bounds are recursively calculated at each time step.

4.4.2 ESMF based PF algorithm

The idea of the combination of PF with ESMF is to use the nonlinear set-membership ellipsoid boundary as the constraint of the feasible particles. Since the set-membership approach ensures that the unknown true state lies in the resulted ellipsoid, a simple strategy is to delete all the particles lying outside the ellipsoid as they are not valid estimate. In this case, the weight updating equation for PF can be expressed as

$$\tilde{w}_k^i = \begin{cases} 0, & \text{if } (x_k^i - c_k)^T P_k^{-1} (x_k^i - c_k) > 1, \\ \tilde{w}_{k-1}^i \cdot \frac{1}{\|y_k - y_k^i\|}, & \text{otherwise,} \end{cases} \quad i = 1, \dots, N, \quad (4.34)$$

However, as mentioned in Chapter 3, in some cases that all particles would lie outside the ellipsoid, and the algorithm would be failed to resample particles. Therefore, it is reasonable to sample new particles from the resulted ellipsoid once the particle violates the boundary conditions predefined by the ESMF estimate.

Estimation steps of the ESMF based PF algorithm are summarized as follows:

Algorithm 4.1: The ESMPF algorithm

step a. initialization: generate initial particles $\{x_0^i\}_{i=1}^N$ from *a priori* distribution $p(x_0)$, and set $k = 1$;

step b. ESMF estimation: calculate state interval based on prior bounded ellipsoid $\mathbb{E}(c_{k-1}, P_{k-1})$; calculate the Jacobian and Hessian matrices, and find the Lagrange remainder using the interval analysis; calculate ellipsoidal summation and intersection, and obtain the optimized bounding ellipsoid $\mathbb{E}(c_k, P_k)$;

step c. importance sampling: generate predicted particles, $\{x_k^{i,-}\}_{i=1}^N$, from importance sampling distribution $p(x_k | x_{k-1}^i)$;

step d. bound checking: check whether the predicted particle falls in the ellipsoid $\mathbb{E}(c_k, P_k)$; discard and regenerate particles, from the resultant

ellipsoid, that do not pass the boundary check;

step e. weighting: evaluate weights of each particle once new measurement

is available and normalize the weights as $w_k^i = \tilde{w}_k^i / \sum_{j=1}^N \tilde{w}_k^j$;

step f. resampling: if $N_{eff} \leq N_{thr}$, then generate posterior particles, $\{x_k^i\}_{i=1}^N$,

based on weighting information and resampling strategy, and set $w_k^i = 1/N$;

step g. output: estimate the state by calculating $\hat{x}_k = \sum_{i=1}^N w_k^i \cdot x_k^i$, set

$k = k + 1$ and go back to step b.

4.5 Simulation Studies

In this section, two simulation examples are used to demonstrate the effectiveness of the proposed algorithm. All the Monte Carlo simulations were run on a 2.4 GHz CPU with 3 GB RAM PC using MATLAB 2009a.

4.5.1 Nonlinear numeric example

We first use a nonlinear numeric example to illustrate the robustness of the algorithm. Consider a system described by the expression below

$$\begin{aligned} x_1(k+1) &= -0.7x_2(k) + 0.1x_2^2(k) + 0.1x_1(k)x_2(k) + 0.2x_2(k)e^{x_1(k)} + \delta(k) + \omega_1(k) \\ x_2(k+1) &= x_1(k) + x_2(k) - 0.1x_1^2(k) + 0.2x_1(k)x_2(k) + \delta(k) + \omega_2(k) \\ y(k) &= x_1(k) + x_2(k) + \nu(k) \end{aligned} \tag{4.35}$$

where $|\omega_1(k)| \leq 0.1$ and $|\omega_2(k)| \leq 0.1$, $|\nu(k)| \leq 0.2$, modeling error $|\delta(k)| \leq 0.2$, and the initial state is bounded by $3I$, where I is the identity matrix. The state to be estimated is $x_1(k)$.

Figure 4.4 shows estimation results by using generic particle filter (PF), extended Kalman particle filter (EPF) and the proposed ESMPF approach, with the same parameter settings: $N = 50$, $\delta = 0.2$, $x_0 = [2, 0]^T$. After 50 Monte Carlo runs, it has been observed that ESMPF provides better estimate

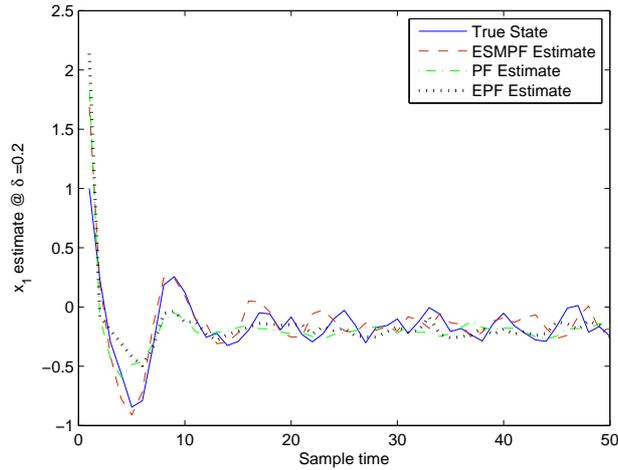


Figure 4.4: Estimation results of x_1 using PF, EPF and ESMPF.

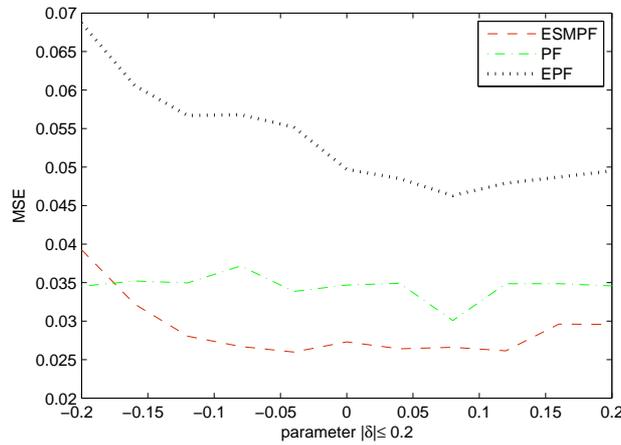


Figure 4.5: Error comparisons for estimate of x_1 using PF, EPF and ESMPF.

results. Figure 4.5 shows the mean square error (MSE) comparison under the choice of different δ values, from the result it can be seen that ESMPF is more robust than the other two approaches against unknown but bounded uncertainties. An interesting observation in this example is that EPF gives worse performance than PF, and we believe one of the main reasons is the unaccounted linearization error.

4.5.2 Continuous fermentation process

In the second case, a nonlinear continuous fermenter is considered for further performance comparison. A simplified unstructured fermentation model is taken from Henson and Seborg (1992),

$$\begin{aligned}\frac{dX}{dt} &= -DX + \mu(P, S)X \\ \frac{dS}{dt} &= D(S_f - S) - \frac{1}{Y_{X/S}}\mu(P, S)X \\ \frac{dP}{dt} &= -DP + (\eta\mu(P, S) + \gamma)X\end{aligned}\quad (4.36)$$

where X , S and P represent biomass concentration, substrate concentration and product concentration, respectively, D the dilution rate, and S_f the feed substrate concentration, $Y_{X/S}$ the cell-mass yield, η and γ yield parameters for the product, and $\mu(P, S)$ the specific growth rate exhibiting both substrate and product inhibition as:

$$\mu = \frac{\mu_m(1 - \frac{P}{P_m})S}{K_m + S + \frac{S^2}{K_i}}\quad (4.37)$$

where μ_m is the maximum specific growth rate, P_m the product saturation constant, K_m the substrate saturation constant, and K_i the substrate inhibition constant.

Table 4.1 shows the nominal model parameters and operating conditions used in this section. In order to make the estimation problem better conditioned, the states, $[X, S, P]^T$, were normalized by dividing them with their nominal values. The measurement is the noisy observation of product concentration. Process and measurement noises are $|\omega(1)| \leq 1e^{-3}$, $|\omega(2)| \leq 1e^{-2}$, $|\omega(3)| \leq 1e^{-3}$ and $|\nu| \leq 1$. In this study, we assume that the dilution rate and the feed substrate concentration are subject to changes with known bounds $|\Delta D| \leq 0.06$ and $|\Delta S_f| \leq 3$. For simulation purpose, an unmodeled disturbance is introduced at the 40th hours with $D = 0.10h^{-1}$ and $S_f = 22g/l$.

The conventional PF, EPF and ESMPF are implemented with $N = 100$. The estimation results are shown in Figure 4.6. As seen in the plots, ESMPF

Table 4.1: Nominal fermenter parameters and operating conditions

Variable	Nominal value
$Y_{X/S}$	0.4 g/g
η	2.2 g/g
γ	0.2 h^{-1}
μ_m	0.48 h^{-1}
P_m	50 g/l
K_m	1.2 g/l
K_i	22 g/l
D	0.15 h^{-1}
S_f	20 g/l
X	7.038 g/l
S	2.404 g/l
P	24.87 g/l

recovers robustly from the unmodeled disturbance due to the consideration of the uncertainty bound for the modeling inaccuracy, while both conventional PF and EPF provides deviated state estimation.

4.6 Conclusion

A well known limitation in the application of Bayesian estimator to real-world problems is the assumption of known *a priori* statistics for the uncertainties. Robustness to unknown noises in estimation is important. This chapter has presented a robust approach for state estimation, applicable where the description of uncertainty due to modeling error, measuring noise, etc., is unknown but bounded. Interesting geometrical insights into the prediction and updating mechanisms are discussed. A robust solution has been obtained for nonlinear uncertain systems based on Monte Carlo sampling and extended set membership approaches.

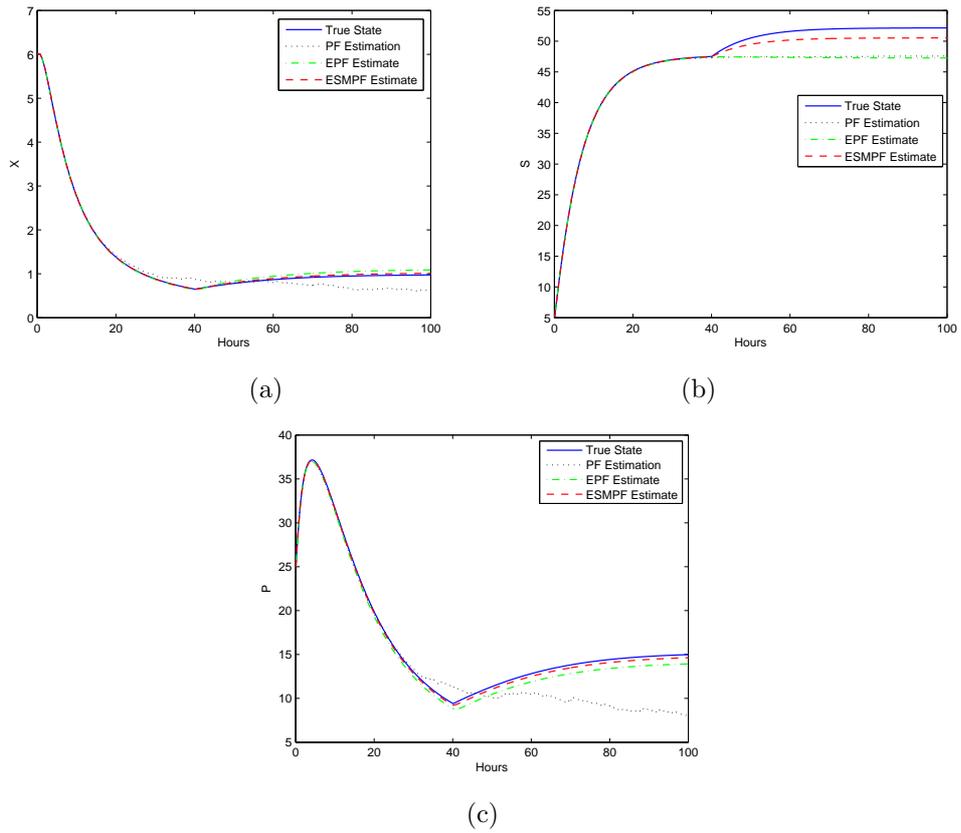


Figure 4.6: Estimation results for continuous fermentation process. (a) estimation results of biomass concentration X ; (b) estimation results of substrate concentration; (c) estimation results of the product concentration P .

Bibliography

- Alamo, T., Bravo, J., Redondo, M., Camacho, E., 2008. A set-membership state estimation algorithm based on dc programming. *Automatica* 44, 216–224.
- Becis-Aubry, Y., Boutayeb, M., Darouach, M., 2008. State estimation in the presence of bounded disturbances. *Automatica* 44(7), 1867–1873.
- Chernousko, F., 1980. Optimal guaranteed estimates of indeterminacies with the aid of ellipsoids, i. *Engineering Cybernetics* 18(3), 1–9.
- de Freitas, J., Niranjan, M., Gee, A., Doucet, A., 2000. Sequential monte carlo methods to train neural network models. *Neural Computation* 12, 955–993.
- Henson, M., Seborg, D., 1992. Nonlinear control strategies for continuous fermenters. *Chemical Engineering Science* 47(4), 821–835.
- Moore, R. E., 1966. *Interval analysis*. NJ: Prentice-Hall.
- Rajamani, M., Rawlings, J., 2007. Improved state estimation using a combination of moving horizon estimator and particle filters. In: *Proceedings of the 2007 American Control Conference*.
- Scholte, E., Campbell, M., 2003. A nonlinear set-membership filter for on-line applications. *International Journal of Robust and Nonlinear Control* 13(15), 1337–1358.
- Schweppe, F., 1968. Recursive state estimation: Unknown but bounded errors and system inputs. *IEEE Transactions on Automatic Control* 13, 22–28.

van der Merwe, R., Doucet, A., de Freitas, N., Wan, E., 2000. The unscented particle filter. Tech. rep., Cambridge University Engineering Department.

Zemke, J., 1999. b4m: A free interval arithmetic toolbox for matlab.

URL <http://www.ti3.tu-harburg.de/zemke/b4m/>

Zhou, B., Han, J., Liu, G., 2008. A ud factorization-based nonlinear adaptive set-membership filter for ellipsoidal estimation. *International Journal of Robust and Nonlinear Control* 18, 1513–1531.

Chapter 5

Particle Filter for Multirate Data Synthesis and Model Calibration

¹ A crucial part in the design of Bayesian estimator is the acquisition of the process model. Due to the complexity of developing accurate first-principle models, data-driven models are becoming more and more common in modern process industries. This chapter presents a brief overview of the most popular techniques and some experiences we have in data-driven modeling relevant to soft sensor development. We show how the flexibility of the Bayesian approach can be exploited to account for multiple-source observations with different degrees of belief, and utilized for data-driven model calibration. A practical Bayesian fusion formulation with time-varying variances is proposed to deal with possibly abnormal observations. Particle filter is used for simultaneously handling systematic and non-systematic errors (i.e., bias and noise), in the presence of process constraints. The proposed method is illustrated through a simulation example and a data-driven soft sensor application in an oil sands froth treatment process.

1. A version of this chapter has been published as “X. Shao, B. Huang, J.M. Lee, F. Xu, A. Espejo, Bayesian Method for Multirate Data Synthesis and Model Calibration, *AIChE Journal*, 57(6), pp. 1515-1525, 2011.”

5.1 Introduction

For process safety and reliability reasons, simultaneous use of multiple measurement methods for critical variables is a common practice in industry. A typical scenario in chemical processes is that both on-line instruments and off-line laboratory analyses are used to monitor the key product quality variable. Generally, on-line instruments have fast sampling rates (such as 1 minute for control purpose) but with low accuracy; furthermore, these hardware sensors could easily fail, leading to information loss. In contrast, the off-line laboratory analysis involves trained personnel manually collecting samples and performing a series of experiment steps for calculations; therefore the result is relatively accurate but the sampling rate is slow (ranging from 30 minutes to 24 hours) with irregular time delays. Overall, each method alone has its own deficiency, and may not be appropriate for real-time monitoring and control purposes.

In order to obtain more accurate and reliable real-time process information, soft sensors (a.k.a. virtual sensor) have been investigated in many process industries to synthesize relevant variables (Rao et al., 1993; Qin et al., 1997; de Assis and Filho, 2000; Chen et al., 2004; Yan et al., 2004; Fortuna et al., 2005; Khatibisepehr and Huang, 2008; Kadleca et al., 2009). The idea of soft sensors is to use a process model that provides online estimates of difficult-to-measure quality variables (e.g., melt index, pH value, concentration) from readily available process variables (e.g., temperature, pressure, flowrate). To achieve a successful soft sensor application, the process model is the key. Generally, there are two well known approaches to building a process model: first-principle approaches (Grantham and Ungar, 1990; Friedman et al., 2002) and data-driven approaches (MacGregor, 2004; Kano and Nakagawa, 2008; Kadleca et al., 2009). A first-principle model is based on good understanding of the underlying fundamental principles such as mass and energy balances. A data-driven model is based on limited process knowledge, and mainly relies on the

historical data describing input (i.e., process variable) and output (i.e., quality variable) characteristics. Increased complexity of the process dynamics often prevents one from building accurate first-principle models. On the other hand, data-driven approaches have been extensively employed for modeling of complex systems since the process related signals are rather easy to obtain from instruments and experiments (Kadleca et al., 2009).

The main challenge of the data-driven modeling arises from the lack of data for good representations of process dynamics. Since the available training data only describes a period of process historical behavior, the investigated process could have changed over time; therefore large validation errors may still exist between the model estimate and actual observation even though the model initially may be sufficient. A natural question then arises: how to use the latest observations of quality variables to update the model for better estimate.

Motivated by the above question, this chapter focuses on the development of a data-driven model update approach for soft sensor applications based on multiple-source quality variable observations. The proposed approach is built on a Bayesian framework (Huang, 2008), which facilitates the inclusion of additional information in the form of prior knowledge and the synthesis of fast sampled but less accurate observations with more accurate but slow sampled observations to derive more accurate posterior distribution for the unknown state and parameters. To enhance the robustness, a practical Bayesian fusion formulation with time-varying variances is proposed and observation validity is taken into account. The Bayesian model calibration strategy is finally implemented by using the particle filtering approach (Doucet and Godsill, 1998), and applied to an industrial soft sensor design.

The remainder of this chapter is organized as follows: Section 2 gives a literature review on data-driven modeling using different sampling rate of input/output data for soft sensor developments. Section 3 introduces model calibration strategies with Bayesian information synthesis. Section 4 introduces a

robust Bayesian fusion formulation for handling abnormal observations. Section 5 implements the Bayesian model calibration strategy as a sequential Monte Carlo sampling based constrained particle filter. Section 6 presents a simulation example to show the characteristics and benefits of the proposed approach. An oil sands froth treatment process is introduced and data-driven soft sensor application results are illustrated in Section 7. Section 8 gives the conclusions.

5.2 Data-driven models

Both scientific and engineering communities have acquired extensive experience in developing and using data-driven modeling techniques. Despite a variety of model structures, two types of data-driven models are widely seen in the literature. One is dynamic model, and the other is static model. This section presents a brief overview of the most popular techniques and some experiences of the author in data-driven modeling using multirate process data.

5.2.1 Dynamic modeling based on fast-rate input/output data

When on-line instruments are available for both input and output variables, a set of valid input and output data can be collected from the historical database. In this case, a fast-rate dynamic model can be generally identified for the investigated process (Wang et al., 2004). The book by Ljung (1999) is considered as a milestone in the field of dynamic identification theory. The identification methods described therein are commonly used for linear dynamic modeling, including autoregressive models (e.g., ARX, ARMAX), Output-Error (OE) models, Box-Jenkins (BJ) models, state space models, etc. Estimation techniques include prediction-error minimization schemes and various subspace methods. When linear models are not sufficient to capture system dynamics, one can resort to nonlinear models, such as non-linear ARX (NLARX) (Chen and Tsay, 1993) and Hammerstein-Wiener models (Bai, 1998).

5.2.2 Static modeling based on slow-rate input/output data

In practice, instrumentation readings for output variables (i.e., difficult-to-measure quality variables) are usually unreliable and inaccurate. If the amount of valid fast-rate output data is insufficient, an alternative way is to use the slow-rate lab data as the output, and resample the fast-rate input data according to the known lab data time stamps; techniques such as moving average could be used to reduce input uncertainties. In this case, process dynamics may be lost during the data collection stage, due to the large sampling intervals. However, more operating conditions are likely to be contained in the originally collected data sets as they come in a more abundant quantity.

For static data-driven modeling, linear regression methods (e.g., ordinary least squares, OLS, Åke Björck (1996)) are commonly used. However, OLS may suffer from numerical problems when a data set is collinear, which is not uncommon in chemical processes. Principal component regression (PCR) and partial least squares (PLS) address the collinearity by projecting the original process variables onto a low dimensional space of orthogonal latent variables. PCR and PLS techniques are well reviewed in Nelson et al. (1996); Dayal and MacGregor (1997); Kresta et al. (1994) and references therein. For the nonlinear case, nonlinear regression methods, such as artificial neural network (Bishop, 1995), support vector machine (Yan et al., 2004), and fuzzy logic (Nagai and Arruda, 2005) could be used.

5.2.3 Dynamic modeling based on fast-rate input and slow-rate output data

Ignoring process dynamics in a static model is one of the causes of model inaccuracy (Zhu et al., 2009). To improve this, dynamic modeling using fast-rate input and slow-rate output has received considerable attention in both academic and industrial communities.

A special case widely investigated is known as dual-rate system identifi-

cation, where the output sampling time is slower than the input sampling time. Early contributions can be found in Lu and Fisher (1988, 1989), in which a polynomial transformation technique and a least squares algorithm are presented to produce fast-rate output based on the measurements of fast-rate input and slow-rate output. The main disadvantage of their algorithm is that additional parameters are introduced. Li et al. (2001) and Wang et al. (2004) use a so-called lifting technique to extract the original fast single-rate system by identifying a higher dimension lifted model. However, this technique becomes impractical when the output sampling rate is very slow and irregular. Ding and Chen (2004) propose to use an auxiliary finite impulse response (FIR) model to predict the noise-free fast-rate output, and then identify a single-rate dynamic model based on the fast-rate input and the estimated output. Raghavan et al. (2006) use an Expectation-Maximum (EM) based approach to interpolate fast-rate output, and then apply a single-rate dynamic identification method; both regular and irregular sampled slow-rate output data can be treated in this approach. However, implementation of the EM algorithm can be expensive for practical applications and the solution may converge to a local optimum. Zhu et al. (2009) propose to use an OE method to identify a dynamic model directly from the fast-rate input and slow-rate output by minimizing the summation of the squared error between the model output and the measurement at the slow rate; the method has the potential to deal with irregular output, and the authors demonstrated their work through an industrial case study. However, it requires a good initial model to avoid the local optimum. Mo et al. (2009) propose to use a FIR model as an initial model, and then apply a fast single-rate OE model for the dynamic identification. Lu et al. (2004) developed a multirate dynamic inferential model based on multiway PLS approach, and demonstrated its efficacy through the Tennessee Eastman process. Tun et al. (2008) developed a method called Data Selection and Regression (DSAR) for identifying irregularly sampled systems

and applied it to soft sensor development on a two-reactor train system.

5.3 Bayesian calibration of data-driven models

Despite the various modeling approaches, a general form of a data-driven model can be described as

$$y_{k+1} = \hat{f}(\phi_k, \theta) + \epsilon_k, \quad (5.1)$$

where $\phi_k = [y_k, \dots, y_{k-n_y}, u_k, \dots, u_{k-n_u}]^T$ is a regressor vector consisting of output and input. n_y and n_u are the model order parameters, which can be determined by minimizing Akaike information criterion (Akaike, 1974); $\hat{f}(\cdot)$ is a selected model structure describing a linear or nonlinear relationship between the input and output variables; θ is the model parameter estimated from the training data; and ϵ_k is the output residual. Note that for a static model, the regressor only contains one input term.

In many practical application, the mismatch between model prediction and actual observation could be significant in a data-driven model, and the error mainly arises in two stages. One is in the modeling stage, such as misuse of model structure, or insufficiency of training data; the other is in the application stage, such as the drift of operating conditions, or the degradation of equipment efficiency. In order to obtain a better estimate of the true quality information, it is important to synthesize all the available observations, and then use them to update the existing model with the consideration of uncertainty.

Strategies for model updating roughly fall into two categories: model refinement and model calibration (Xiong et al., 2009). Model refinement involves the change of model structure, for example, using a nonlinear model to replace a linear model, which is desirable for fundamentally improving the predictive capability; however, the practical feasibility of refinement is often restricted by available knowledge and computing resource. In contrast, model calibra-

tion utilizes mathematical means to match model predictions with reliable observations, which is a cheaper way for practical applications.

5.3.1 Model calibration

Various model calibration strategies exist, and a conventional way is to consider the model parameters adaptation in the model form of

$$\hat{y}_{k+1} = \hat{f}(\phi_k, \theta_k), \quad (5.2)$$

where θ_k represents time-varying model parameters.

However, in many situations, calibrating model parameters is still unable to compensate model-plant mismatch, for example, due to the use of incorrect model structure. Then the following bias correction form could be used (Singh, 1997; Mu et al., 2006),

$$\hat{y}_{k+1} = \hat{f}(\phi_k, \theta_k) + \gamma_k, \quad (5.3)$$

where γ_k is the discrepancy term to capture the systematic error (i.e., bias).

In addition to using an additive bias, a multiplicative correction could also be considered as

$$\hat{y}_{k+1} = \rho_k \hat{f}(\phi_k, \theta_k) + \gamma_k, \quad (5.4)$$

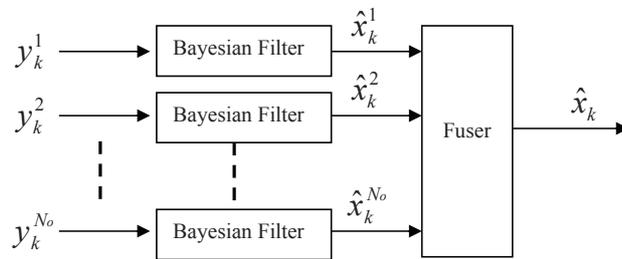
where the scaling parameter ρ_k brings more flexibility to the model-plant mismatch compensation.

The choice of a model calibration form is problem-specific and requires insight into the error sources, while more interesting question remains: how to synthesize the multiple-source quality variable observations in an optimal manner to reduce the uncertainty and achieve more accurate estimation.

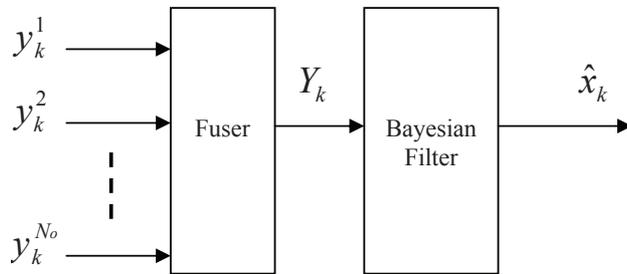
5.3.2 Bayesian information synthesis

There are a few data fusion (or information synthesis) approaches to resolve the above question (Kewley, 1992; Braun, 2000; Koks and Challa, 2003), of

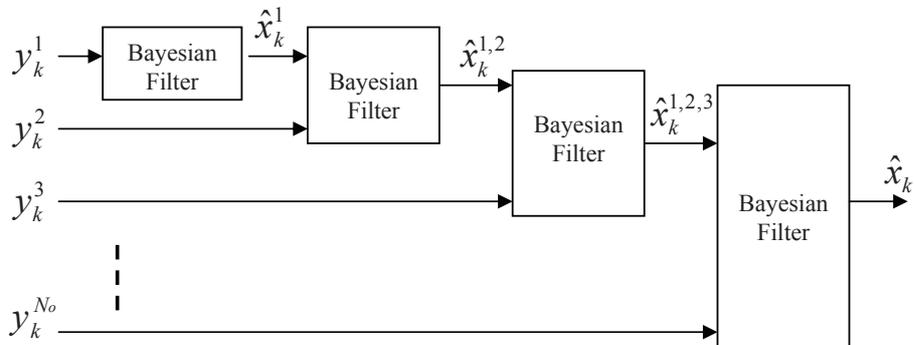
which the Bayesian inference based approach is the most unified one. Figure 5.1 shows three most popular strategies for Bayesian information synthesis. Figure 5.1(a) shows the state vector fusion method, also known as the distributed approach, where a group of Bayesian filters are used to obtain individual observation based estimates, and then fused together (e.g., linear combination) to obtain an improved joint estimate. It is a favorable choice for processes with numerous observation sources, because of computation cost as well as the parallel implementation and fault-tolerance issues (Saha and Chang, 1998). However, this approach requires consistent Bayesian filters, and inappropriate combination of individual estimate can deteriorate the final result (Gan and Harris, 2001). Figure 5.1(b) shows the measurement fusion approach, also known as the centralized approach, in which all the observations are directly fused to obtain synthesized process information, and then a single Bayesian filter is used to obtain the final estimate. Figure 5.1(c) shows a hybrid use of the distributed and centralized approaches, resulting in a more complicated sequential fusion scheme. It yields the same result as centralized one when the number of observation sources equals to two. In this chapter, the centralized approach is selected since it is the best way to synthesize observations in the sense that no information will be lost during the fusion procedure (Koks and Challa, 2003) and the number of observation sources for the problem investigated in this chapter is not large.



(a)



(b)



(c)

Figure 5.1: Bayesian filter based data fusion strategies: (a) distributed approach; (b) centralized approach; (c) hybrid (sequential) approach.

The investigated problem can be put into a state-space form as follows:

$$\begin{aligned}
x_{k+1} &= \begin{bmatrix} 0 & \cdots & 0 & 0 \\ I & & & 0 \\ & \ddots & & \vdots \\ & & I & 0 \\ & & & 0 & \cdots & 0 & 0 \\ & & & I & & & 0 \\ & & & & \ddots & & \vdots \\ & & & & & I & 0 \end{bmatrix} x_k + \begin{bmatrix} \rho_k \hat{f}(x_k, u_k, \theta_k) + \gamma_k \\ 0 \\ \vdots \\ 0 \\ u_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \omega_k^x, \\
\theta_{k+1} &= \theta_k + \omega_k^\theta, \\
\rho_{k+1} &= \rho_k + \omega_k^\rho, \\
\gamma_{k+1} &= \gamma_k + \omega_k^\gamma, \\
y_{T_k^n}^n &= H x_{T_k^n} + \nu_{T_k^n}^n \\
&= [1 \ 0 \ \cdots \ 0] x_{T_k^n} + \nu_{T_k^n}^n, \quad n = 1, \dots, N_o,
\end{aligned} \tag{5.5}$$

where $x_k = [y_k, \dots, y_{k-n_y}, u_{k-1}, \dots, u_{k-n_u}]^T$; ω_k^x , ω_k^θ , ω_k^ρ and ω_k^γ are random variables representing process and model uncertainties; $\nu_{T_k^n}^n$ is a random variable for capturing the non-systematic error (i.e., observation noise) associated with sensor n ; it is assumed that the observation noise is subject to a Gaussian distribution, $\nu_{T_k^n}^n \sim \mathcal{N}(0, \sigma_n^2)$, when the sensor (or observation source) works under normal conditions; T_k^n indicates a time-varying sampling rate for the n^{th} observation source.

With the calibration parameter vector denoted as $\Theta_k = [\theta_k, \rho_k, \gamma_k]^T$, Equation (5.5) can also be represented by a probabilistic graph as shown in Figure 5.2, where all the unknown nodes are considered as random variables. (Note that the arc between x_{k-1} and x_k is left out if $\hat{f}(\cdot)$ is a static model.)

The objective of Bayesian information synthesis is to construct the *a posteriori* distribution, $p(x_k, \Theta_k | \mathcal{D}_k)$, of the state (or unknown true quality variable), x_k , and the calibration parameter, Θ_k , simultaneously, based on available multiple-source noisy observations, $\mathcal{D}_k = \{\mathcal{Y}_1, \dots, \mathcal{Y}_k\}$, where $\mathcal{Y}_k = \{y_k^1, \dots, y_k^{N_o}\}$.

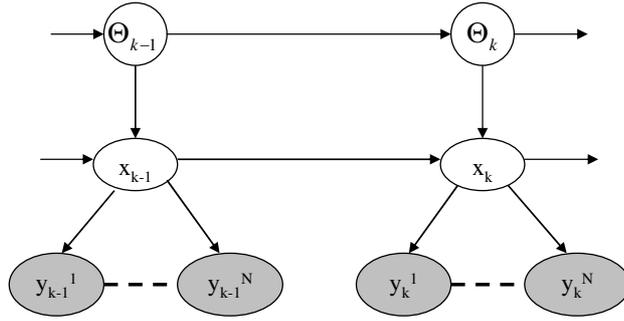


Figure 5.2: Graphical representation of Equation (5.5); grey nodes represent known variables.

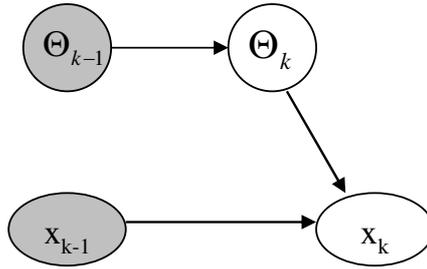


Figure 5.3: Prediction step for Bayesian inference.

As per conventional Bayesian estimation, the required posterior distribution can be obtained by recursively following two steps: prediction and update.

Prediction: At time $k - 1$, all the evidence up to time $k - 1$ has been taken into account, and the posterior distribution $p(x_{k-1}, \Theta_{k-1} | \mathcal{D}_{k-1})$ has been estimated. Then the prior distribution at time k can be obtained as:

$$p(x_k, \Theta_k | \mathcal{D}_{k-1}) = \int p(x_k, \Theta_k | x_{k-1}, \Theta_{k-1}) p(x_{k-1}, \Theta_{k-1} | \mathcal{D}_{k-1}) dx_{k-1} d\Theta_{k-1}. \quad (5.6)$$

Here the probabilistic models $p(x_k | x_{k-1}, \Theta_{k-1})$ and $p(\Theta_k | \Theta_{k-1})$ are defined by the system equations and the associated statistics of ω_k^x , ω_k^θ , ω_k^ρ and ω_k^γ . A graphical interpretation is shown in Figure 5.3.

Update: At time k , the latest observation $\mathcal{Y}_k = \{y_k^1, \dots, y_k^{N_o}\}$ is available (see Figure 5.4), then the posterior distribution can be obtained via Bayes'

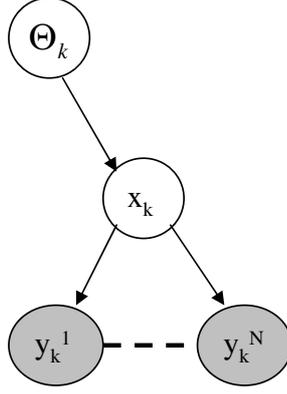


Figure 5.4: Update validation step for Bayesian inference.

rule,

$$\begin{aligned}
 p(x_k, \Theta_k | \mathcal{D}_k) &= \frac{p(\mathcal{Y}_k | x_k, \Theta_k) p(x_k, \Theta_k | \mathcal{D}_{k-1})}{p(\mathcal{Y}_k | \mathcal{D}_{k-1})} \\
 &= \frac{p(y_k^1 | x_k, \Theta_k) p(y_k^2 | x_k, \Theta_k) \cdots p(y_k^{N_o} | x_k, \Theta_k) p(x_k, \Theta_k | \mathcal{D}_{k-1})}{p(y_k^1, y_k^2, \cdots, y_k^{N_o} | \mathcal{D}_{k-1})} \\
 &\propto p(x_k, \Theta_k | \mathcal{D}_{k-1}) \prod_{n=1}^{N_o} p(y_k^n | x_k, \Theta_k),
 \end{aligned} \tag{5.7}$$

where observations from different sources are considered as independent given the state information, and the normalizing denominator is given by

$$p(\mathcal{Y}_k | \mathcal{D}_{k-1}) = \int p(\mathcal{Y}_k | x_k, \Theta_k) p(x_k, \Theta_k | \mathcal{D}_{k-1}) dx_k d\Theta_k. \tag{5.8}$$

Once the posterior distribution is obtained, it can be used for point state inference, such as mode, mean or median estimate. Note that Bayesian approach can handle the varying size of \mathcal{Y}_k (i.e., missing data) naturally caused by the multirate sampling mechanism.

5.4 Bayesian information synthesis with abnormal observation data

In reality, no sensor (or observation source) can provide precise measurements continuously. Due to sensor malfunction, transmission error, or human data entry error, one may obtain “unexpected” values for a measured variable. Such

abnormal data can be propagated through the fusion procedure and cause a divergent estimate. To achieve a robust estimate in the presence of abnormal data, this section describes a variance adaptation scheme for Bayesian information fusion.

It is well known that the observation noise variance is important for information fusion, since it directly determines the relative weight assigned to the observation source (Punska, 1999). However, in the real world, the variance of the true observation noise is rarely known; it is generally pre-estimated and kept unchanged during the application. This will yield the same weight to an observation source regardless of its measurement quality. To circumvent this, we assume the noise is subject to a Gaussian distribution with a time-varying variance, namely,

$$p(y_k^n | x_k, \Theta_k) \sim \mathcal{N}(Hx_k, \sigma_n^2(k)), \quad (5.9)$$

where $\sigma_n^2(k)$ can increase significantly when the observation is becoming abnormal, therefore reducing its influence on the information fusion.

Thus, the problem is how to define the normality or abnormality. Hua and Wu (2006) suggests that the distance between the n^{th} sensor's observation with respect to the rest of the sensors can be used to quantify the abnormality. Their method requires at least three observation sources and assumes that the majority of the sources provide correct and consistent measurements.

In this work, motivated by the approach widely used by practicing engineers in the actual operations, a variance adaptation scheme is developed. For an individual sensor, we partition its measurements into three categories: *valid*, *possibly valid*, and *invalid*. (See Figure 5.5 for an illustration.)

A validity state, λ_k^n , is introduced to indicate the observation validity (i.e. normality) of the n^{th} sensor at time k . $\lambda_k^n = 1$ indicates that the observation data is valid (i.e. normal), and $\lambda_k^n = 0$ indicates that the observation data is invalid (i.e. abnormal). Then the time-varying noise variance $\sigma_n^2(k)$ is defined

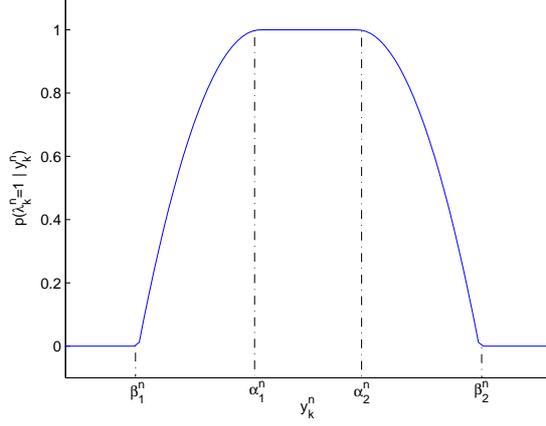


Figure 5.5: Observation validity given a sensor reading.

as

$$\sigma_n^2(k) = \begin{cases} \sigma_n^2, & \text{if } y_k^n \in [\alpha_1^n, \alpha_2^n], \text{ i.e., valid,} \\ \frac{1}{p(\lambda_k^n=1)}\sigma_n^2, & \text{if } y_k^n \in [\beta_1^n, \alpha_1^n] \text{ or } y_k^n \in (\alpha_2^n, \beta_2^n], \text{ i.e., possibly valid,} \\ \infty, & \text{if } y_k^n \in (-\infty, \beta_1^n) \text{ or } y_k^n \in (\beta_2^n, +\infty), \text{ i.e., invalid,} \end{cases} \quad (5.10)$$

where σ_n^2 is the pre-determined variance for sensor n under normal working conditions; α_1^n and α_2^n are the lower and upper bounds of the n^{th} observation to be believed as *valid*; β_1^n and β_2^n are the tolerable bounds of the n^{th} observation to be believed as *possibly valid*; and all the sensor readings smaller than β_1^n or larger than β_2^n are considered as *invalid*.

In Equation (5.10), the probability function $p(\lambda_k^n = 1 | y_k^n) \in [0, 1]$ is a user specified function, which can have different formulations. One option is

$$p(\lambda_k^n = 1 | y_k^n) = \begin{cases} \frac{(\beta_1^n - \alpha_1^n)^2 - (y_k^n - \alpha_1^n)^2}{(\beta_1^n - \alpha_1^n)^2}, & \text{if } y_k^n \in (\beta_1^n, \alpha_1^n), \\ \frac{(\beta_2^n - \alpha_2^n)^2 - (y_k^n - \alpha_2^n)^2}{(\beta_2^n - \alpha_2^n)^2}, & \text{if } y_k^n \in (\alpha_2^n, \beta_2^n). \end{cases} \quad (5.11)$$

The rationale for Figure 5.5 and Equation (5.11) is based on common industrial practice: (i) specification range for a quality variable does not change substantially during a continuous operation, although input variables can have different operating points; (ii) when a measurement is unusually large or small, the measurement is regarded as abnormal and discarded.

Substituting Equations (5.9) and (5.10) into Equation (5.7), one can obtain

the posterior distribution as

$$p(x_k, \Theta_k | \mathcal{D}_k) \propto p(x_k, \Theta_k | \mathcal{D}_{k-1}) e^{-\left\{ \frac{(Hx_k - y_k^1)^2}{2\sigma_1^2} \cdot p(\lambda_k^1=1) + \dots + \frac{(Hx_k - y_k^{N_o})^2}{2\sigma_{N_o}^2} \cdot p(\lambda_k^{N_o}=1) \right\}}. \quad (5.12)$$

In Equation (5.12), the contribution of an individual sensor to the estimate is decreased (i.e., increasing the variance), if its measurement has low probability to be valid. The influence of a particular sensor will be negligible, as the variance goes to infinity, meaning that its measurement is invalid.

To implement the Bayesian model calibration strategy, sequential Monte Carlo sampling based particle filter is utilized as analytical solutions for Equation (5.12) are unavailable in general except for special cases such as unconstrained linear systems with Gaussian noise.

By choosing the system equation as the importance sampling function, one can derive that the unnormalized importance weight, $\tilde{w}_k^{(i)}$, as

$$\begin{aligned} \tilde{w}_k^{(i)} &\propto w_{k-1}^{(i)} p(\mathcal{Y}_k | x_k^{(i)}, \Theta_k^{(i)}) \\ &\propto w_{k-1}^{(i)} e^{-\sum_{n=1}^{N_o} \left\{ \frac{(Hx_k - y_k^n)^2}{2\sigma_n^2} \cdot p(\lambda_k^n=1) \right\}}. \end{aligned} \quad (5.13)$$

5.5 An Illustrative Example

In this section, an illustrative example is presented to show the characteristics and benefits of our proposed method. Consider a nonlinear system given by

$$x_k = 0.9 \cdot x_{k-1} - 0.5 \cdot x_{k-2} \cdot (1 + x_{k-1}^2) + u_{k-1} + 0.5 \cdot u_{k-2} + d_{k-1} \quad (5.14)$$

where $u_{(\cdot)}$ is the input with a sampling time of 1 minute; $d_{(\cdot)}$ is the unknown process disturbance (or modeling mismatch term); $x_{(\cdot)}$ denotes the process quality variable (or model output) which has two approaches to measure its values. The first approach has fast sampling rate (1 minute), but with low accuracy (controlled by the measurement noise, see Equation (5.17)), while the second one has slow sampling rate (4 hours), but with high accuracy.

The process is simulated for 2400 minutes with its input defined as follows

$$u_k = \frac{0.1}{1 - 0.978q^{-1}} \cdot e_k \quad (5.15)$$

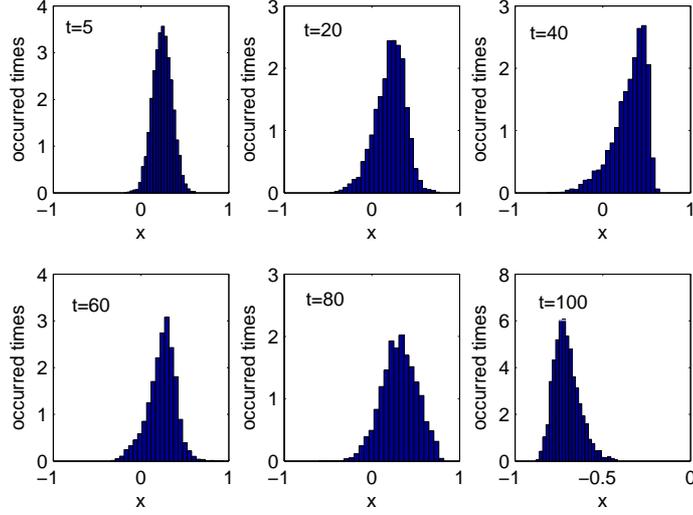


Figure 5.6: Evolution of the simulated output for the numeric example.

where e_k is white noise generated from a normal distribution $\mathcal{N}(0, 0.1^2)$.

The unmodeled disturbance term d is designed as

$$d_k = 0.5 \cdot \cos\left(\frac{k}{10\pi}\right) + 0.2 \cdot n_k \quad (5.16)$$

where n_k is non-Gaussian noise generated from a bimodal distribution such that with 70% of the time it is generated from a Gaussian distribution with a mean value of -0.2 and variance of 0.1^2 , and with 30% of the time it is generated from a Gaussian distribution with a mean value of 0.2 and variance of 0.1^2 .

5.5.1 Algorithm characteristics

Non-Gaussianity: Figure 5.6 shows the evolution of the true unknown output x_k for the above simulation example, in which we can see that the distribution for the output is non-Gaussian due to process nonlinearity as well as the unmodeled disturbance. Traditional Gaussian filters are not suitable to estimate the posterior distribution of x_k , and Monte Carlo sampling based approach is therefore selected.

Multirate observation fusion: Figure 5.7 shows a scatter plot of 30 days

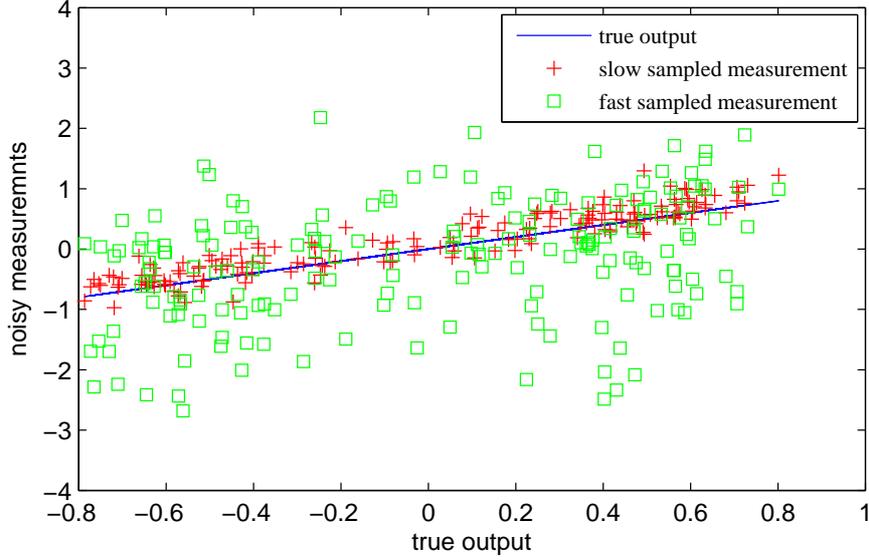


Figure 5.7: Comparison of two different measuring approaches.

measurements from two different observation sources, in which the slow sampled source is designed with higher accuracy (i.e., smaller uncertainty), while the fast sampled source is designed with lower accuracy (i.e., larger uncertainty). Traditional filtering approaches usually use only one of the two observation sources for posterior estimation, namely either $p(x_k|Y_k^1)$ or $p(x_k|Y_k^2)$, due to the implementation difficulties of multirate data. Monte Carlo sampling based approach allows one to use both observation sources for posterior estimation, namely $p(x_k|Y_k^1, Y_k^2)$. Figure 5.8 shows that fused observation (at time step k) can yield less uncertainty information (i.e., smaller variance), and therefore is more likely to produce a better posterior estimate.

Robust to abnormal readings: Due to the measurement uncertainties, abnormal readings are inevitable in practice, especially for those sensors with large uncertainties. Figure 5.9 shows the benefit of using time-varying variance to control the influence of a particular measurement (e.g., a possible abnormal reading $y_k^1 = -1.8$). Figure 5.9(a) shows the fused observation is unable to support the true distribution well when using a prefixed constant variance for each observation source; while Figure 5.9(b) shows an improved

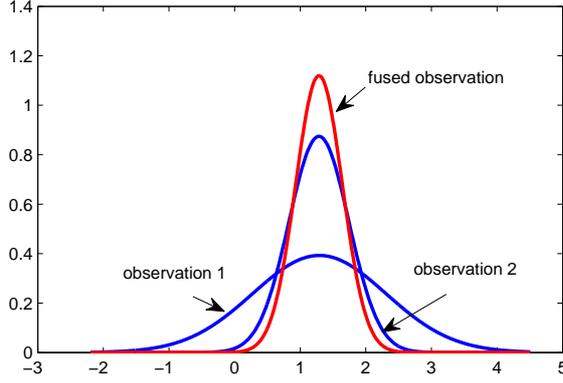


Figure 5.8: Illustration of observation fusion.

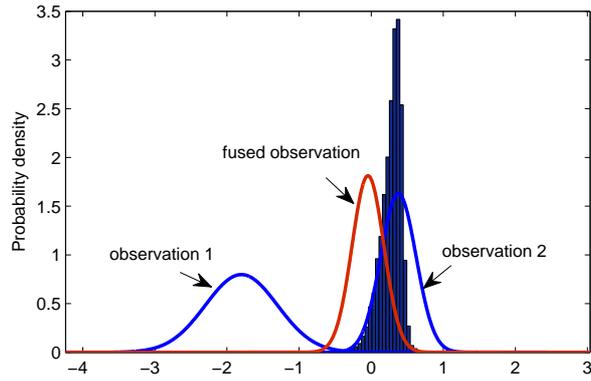
fusion result by adjusting each variance according to Equation (5.10) with parameter set as $\alpha_1^1 = \alpha_1^2 = -1$, $\alpha_2^1 = \alpha_2^2 = 1$, $\beta_1^1 = \beta_1^2 = -2$, $\beta_2^1 = \beta_2^2 = 2$.

Constraint handling: Another benefit of using Monte Carlo sampling approach is that it can easily incorporate lower and upper bound constraints of uncertain variables, which is helpful to confine the distribution shape of the related variables and improve the estimation performance. Further information of Bayesian constrained estimation can be found in Shao et al. (2010) and reference therein.

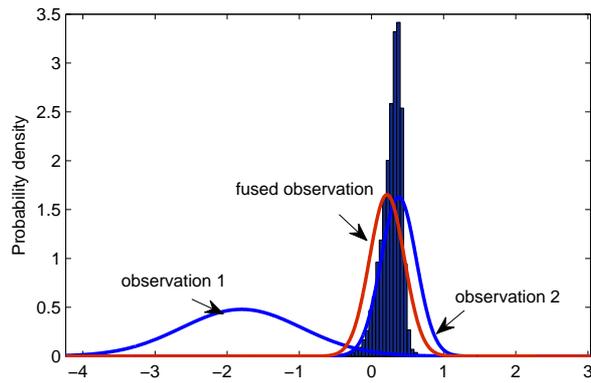
5.5.2 Model calibration

Given the data-driven model as Equation (5.14) excluding the unmodeled term d , Figure 5.10 shows the model prediction (without calibration), the fast rate measurements, the slow rate measurements and the true output. From the figure it is observed that there is a large mismatch between the true output and the model prediction. Note that this kind of comparison is only possible in simulation.

In order to compensate the model-plant mismatch as much as possible, the proposed Bayesian model calibration strategy is applied to the following



(a)



(b)

Figure 5.9: Observation fusion with one possible abnormal reading ($y_k^1 = -1.8$). (a). poor fusion result with prefixed constant measurement noise variances ($\sigma_1^2 = 0.5^2$, $\sigma_2^2 = 0.2^2$); (b). improved fusion result with time-varying variances calculated based on Equation (5.10) ($\sigma_1^2(k) = 0.83^2$, $\sigma_2^2(k) = 0.2^2$).

reconstructed system:

$$\begin{aligned}
x_{k+1}^a &= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x_k^a + \begin{bmatrix} \rho_k \hat{f}(x_k^a, u_k) + \gamma_k \\ 0 \\ u_k \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \omega_k^{x^a}, \\
\rho_{k+1} &= \rho_k + \omega_k^\rho, \\
\gamma_{k+1} &= \gamma_k + \omega_k^\gamma, \\
y_k^1 &= [1 \ 0 \ 0] x_k^a + \nu_k^1, \\
y_{Tk}^2 &= [1 \ 0 \ 0] x_{Tk}^a + \nu_{Tk}^2,
\end{aligned} \tag{5.17}$$

where $x_k^a = [x_k, x_{k-1}, u_{k-1}]^T$ is the augmented state variable; y_k^1 is the fast sampled measurement with large uncertainty $\nu_k^1 \sim \mathcal{N}(-0.1, 1^2)$; y_{Tk}^2 is the slow sampled measurement with small uncertainty $\nu_{Tk}^2 \sim \mathcal{N}(0.1, 0.2^2)$ and $T = 240$ in this example. $\omega_k^x \sim \mathcal{N}(0, 0.5^2)$, $\omega_k^\rho \sim \mathcal{N}(0, 0.1^2)$, $\omega_k^\gamma \sim \mathcal{N}(0, 0.1^2)$; parameters for sensor validation range are chosen as $\alpha_1^1 = \alpha_1^2 = -1$, $\alpha_2^1 = \alpha_2^2 = 1$, $\beta_1^1 = \beta_1^2 = -2$, $\beta_2^1 = \beta_2^2 = 2$; lower and upper bound constraints for output are set as $[-2, 2]$; 100 particles are used for Monte Carlo sampling.

Figure 5.11 shows the estimate results using different model calibration approaches. From the comparison, we can see that particle filter based Bayesian calibration approach gives better estimate than the multirate EKF based approach Gudi et al. (1995). In fact, it is also much easier to implement the proposed algorithm, since it is applicable to nonlinear functions without the need of linearization.

5.6 Industrial Application

In this section, the proposed method is applied to a data-driven soft sensor development in an oil sands bitumen froth treatment process.

5.6.1 Background

Oil sands are mixtures of quartz, clay, water, bitumen and accessory minerals. Athabasca oil sands in Northern Alberta, Canada, is one of the largest oil sands reserve in the world, and currently produces over one million barrels

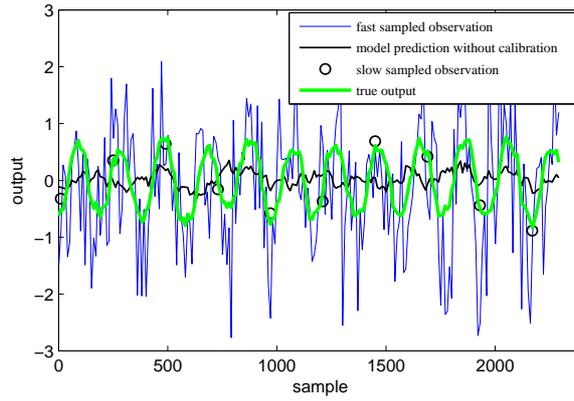
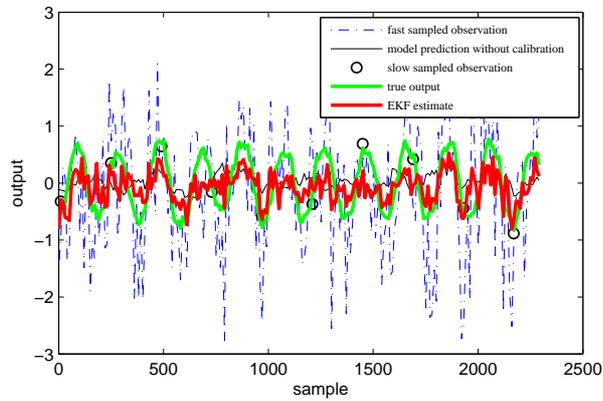
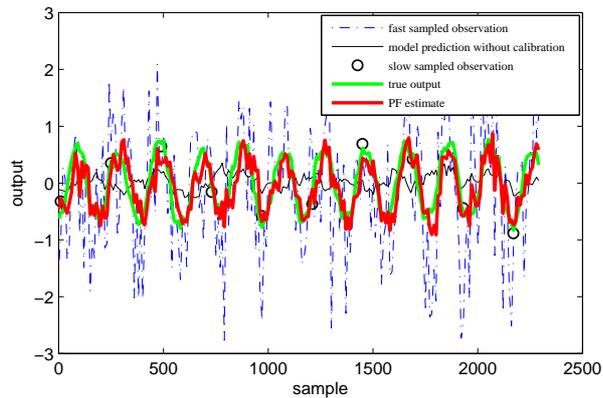


Figure 5.10: Comparisons of existing measurements and model prediction with the true output.



(a)



(b)

Figure 5.11: Estimate results with different model calibration approaches; (a). EKF based approach; (b). PF based approach.

of oil per day. In the process of producing oil from oil sands, the main task is to separate bitumen from other components. The separation is performed through a chain of industrial clarifying units, among which the Inclined Plates Settler (IPS) units are one of the main components of the secondary separation process.

The principle that underlines the functioning of an IPS unit is the space efficient gravity separation, which relies on the density difference between the different components. In order to enhance the density difference, the feed of the IPS unit is diluted with some process aids, e.g., Naphtha and Demulsifier.

The gravity movement in an IPS unit leads to the hydrocarbon-rich phase (light phase) float up and be collected by the outlet boxes to be discharged as overflow product. The denser water and solid-rich phase (heavy phase) settles down the plate and is collected in the hopper to be discharged as underflow tailings. Figure 5.12 gives a schematic representation of an IPS unit.

The water percentage (a.k.a. water content) in the overflow product is a particularly important variable as it reflects the bitumen froth quality and process performance. In practice, both laboratory analysis (e.g., using Karl Fischer titration Scholz (1984)) and hardware instrument (e.g., water-cut meter) are available in the overflow stream. Although the laboratory analysis provides more accurate measurements, the sampling rate, which is 2 hours in this case, is too slow to serve for monitoring and control purposes. Water-cut meter readings are fast sampled, but not accurate enough. Due to the considerable variability in sands, water, clay and bitumen content, the water-cut meter occasionally needs to be removed for maintenance, which leads to the unavailability of online water content information. This poses challenge to control the Naphtha and Demulsifier additions. Therefore, there is an economic necessity to develop a soft sensor to obtain more accurate and reliable real-time water content information.

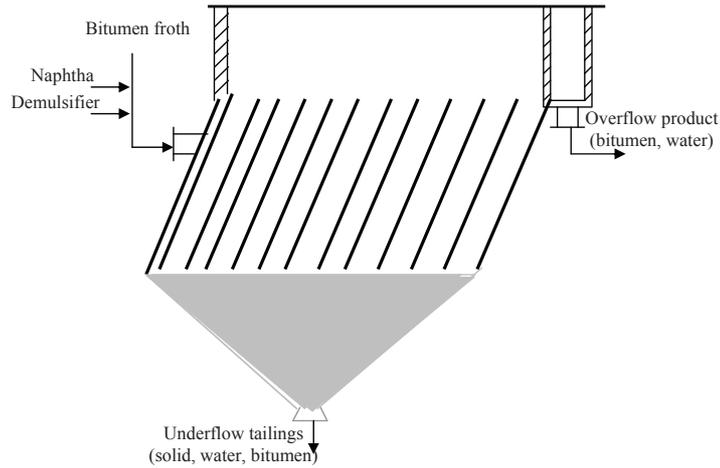


Figure 5.12: Schematic diagram for an Inclined Plates Settler (IPS) unit.

5.6.2 Model estimation

Despite the unreliability and inaccuracy of the water-cut meter readings at most time, we are able to collect a sufficient amount (about one month) of valid fast-rate output data (i.e., water-cut meter readings). Naphtha flowrate, Demulsifier flowrate, inflow flowrate, underflow flowrate, and overflow flowrate are selected as the fast-rate input variables. The idea is that the water-cut meter reading alone as the output variable may not provide a good model but the identified model will be calibrated on-line by the lab data as will be discussed next.

Figure 5.13 shows the collected raw input and output data. For proprietary reason, the actual operating ranges have been modified. After data preprocessing, a second-order NLARX model was identified to represent the process dynamics.

Figure 5.14 shows the structure of the NLARX model, which describes nonlinear dynamics using a parallel combination of nonlinear and linear blocks. A Sigmoid network is used for the nonlinear part, and the estimated model is

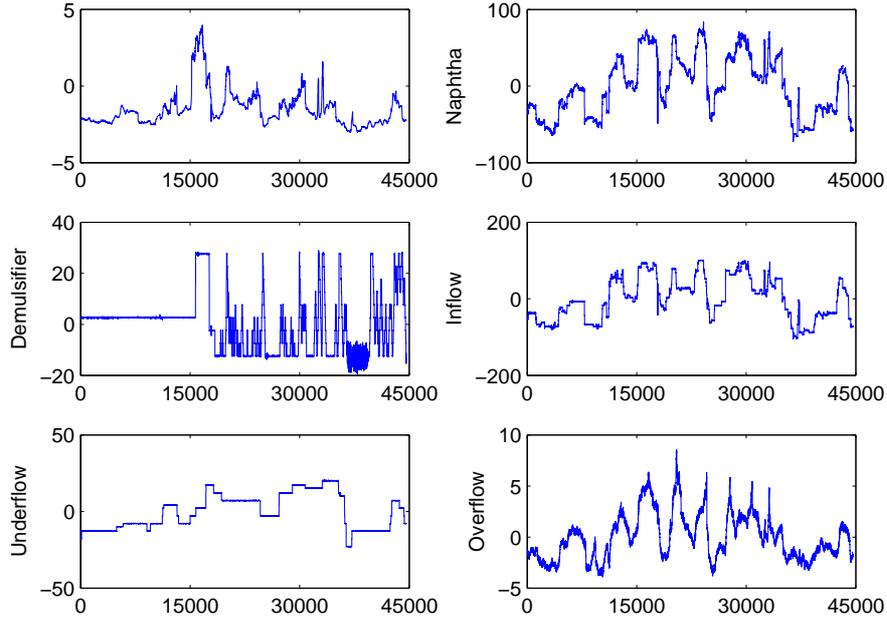


Figure 5.13: Input and output data for NLARX modeling of the investigated IPS unit.

described as

$$\begin{aligned} \hat{y}_{k+1} &= \hat{f}(\phi_k, \theta) \\ &= (\phi_k - r) \cdot P \cdot L + (1 + \exp(-(\phi_k - r) \cdot Q \cdot b - c))^{-1} \cdot a + d, \end{aligned} \quad (5.18)$$

where $\phi_k = [y_k, y_{k-1}, u_k, u_{k-1}]$ is the regressor; $y_{(\cdot)}$ is the true unknown quality information; r the mean of the regressor; Q the nonlinear subspace; P the linear subspace; L the linear coefficient; b the dilation; c the translation; a the output coefficient and d the output offset.

Training and validation results are both shown in Figure 5.15. From the figure we can see that the identified NLARX model is able to capture the process dynamics and provide fairly reasonable estimate. A testing result on a set of fresh data is shown in Figure 5.16. From the figure, we can see that there are a certain amount of mismatches between the lab data and the NLARX estimates as expected.

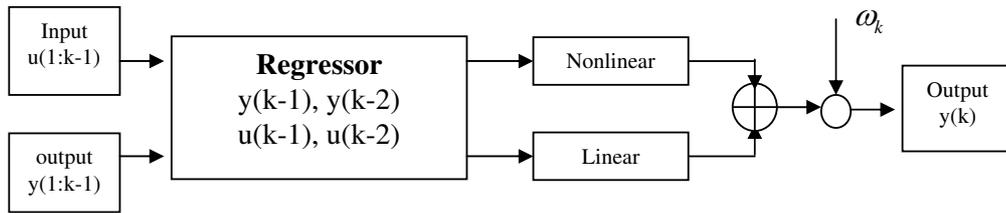
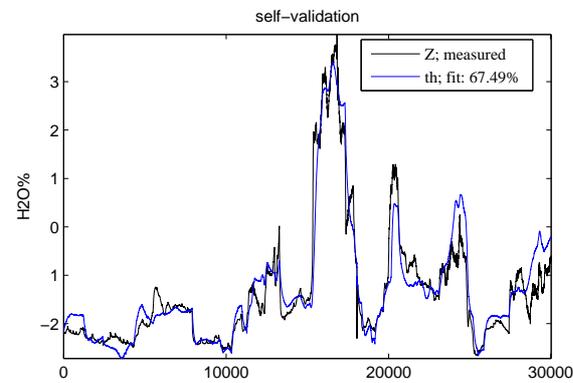
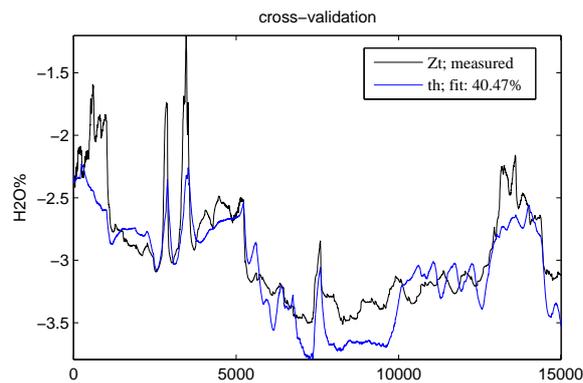


Figure 5.14: A general structure of Nonlinear ARX model.



(a)



(b)

Figure 5.15: Model simulation and validation results for the investigated IPS unit; (a). training; (b). validation.

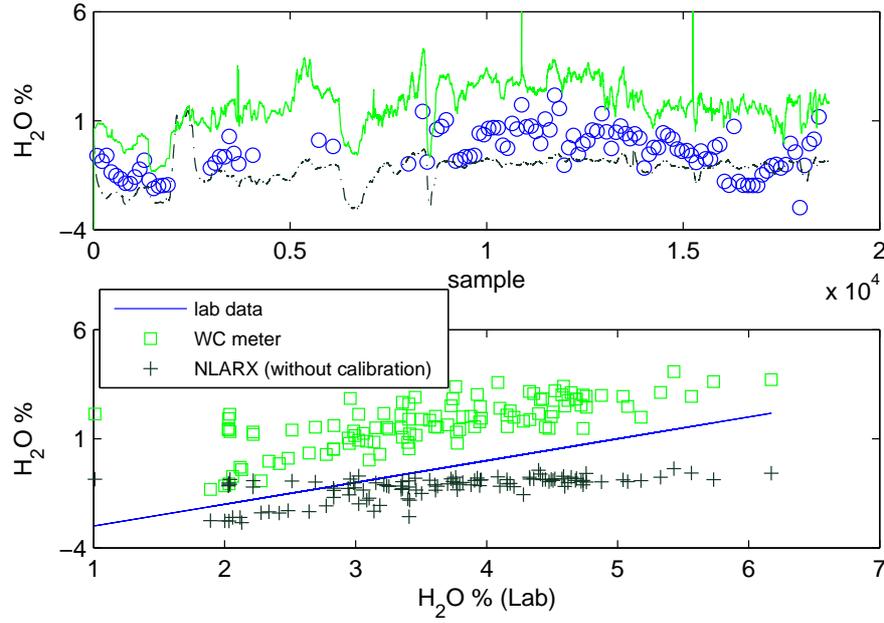


Figure 5.16: NLARX estimates without model calibration for new data set collected in July 2009.

5.6.3 Bayesian calibration

In order to palliate the mismatch as much as possible, we use the proposed Bayesian calibration approach by constructing the problem as

$$\begin{aligned}
 x_{k+1} &= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x_k + \begin{bmatrix} \rho_k \hat{f}(x_k, u_k, \theta_k) + \gamma_k \\ 0 \\ u_k \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \omega_k^x, \\
 \theta_{k+1} &= \theta_k + \omega_k^\theta, \\
 \rho_{k+1} &= \rho_k + \omega_k^\rho, \\
 \gamma_{k+1} &= \gamma_k + \omega_k^\gamma, \\
 y_k^1 &= [1 \ 0 \ 0] x_k + \nu_k^1, \\
 y_{T_k^2 k}^2 &= [1 \ 0 \ 0] x_{T_k^2 k} + \nu_{T_k^2 k}^2,
 \end{aligned} \tag{5.19}$$

where $x_k = [y_k, y_{k-1}, u_{k-1}]^T$; $\theta_k = [r_k, L_k, a_k, d_k]^T$ is a subset vector of NLARX model parameters; y_k^1 is the water-cut meter reading and $y_{T_k^2 k}^2$ is the laboratory analysis.

During the estimation, non-negativity (based on actual operating range)

constraint is imposed on state variable x_k ; 5% perturbations are added to depict the process and model uncertainties; observation noises are chosen as $\nu_k^1 \sim \mathcal{N}(0, 1^2)$ and $\nu_{T_k^2}^2 \sim \mathcal{N}(0, 0.05^2)$; parameters (based on virtual operating range) for sensor validation are set as $\alpha_1^1 = \alpha_1^2 = -3$, $\alpha_2^1 = \alpha_2^2 = 1$, $\beta_1^1 = \beta_1^2 = -4$, $\beta_2^1 = 4$, $\beta_2^2 = 6$; 100 particles are used for sequential Monte Carlo filtering.

Figure 5.17 shows the soft sensor result after model calibration. From the figure, we can see that the overall estimation performance has been improved significantly. Since the modified posterior distribution has taken abnormal observations into account, the estimates are not affected by the sudden abnormal changes in water-cut meter readings.

Table 5.1 presents the comparisons of soft sensor estimates and water-cut meter readings with the lab data as the reference in terms of accuracy (i.e., mean absolute error, MAE), variability (i.e., standard deviation, STD) and overall performance (i.e., rooted mean square error, RMSE). We can clearly see that the soft sensor with model calibration provides the best prediction of the water content. This soft sensor has now been put on-line. Compared to traditional measuring techniques (i.e., hardware sensor and lab analyzer), the developed soft sensor requires much less maintenance effort, thanks to the inclusion of model calibration strategy. Given the obtained benefits, more Bayesian soft sensors are planned for additional processes.

5.7 Conclusion

This chapter presents a practical approach for data-driven model calibration using multiple-source observations. The approach is built within a Bayesian framework to synthesize fast sampled but low accurate observations with high accurate but slow sampled observations to obtain more accurate process information. To enhance the robustness in the presence of abnormal data, a robust Bayesian fusion formulation with time-varying observation noise vari-

Table 5.1: Performance comparison for water-content estimate

	MAE	STD	RMSE
Soft sensor with calibration	0.5704	0.8543	0.8581
Soft sensor without calibration	1.0180	0.7977	1.1598
Water-cut meter	2.1083	0.8594	2.2753

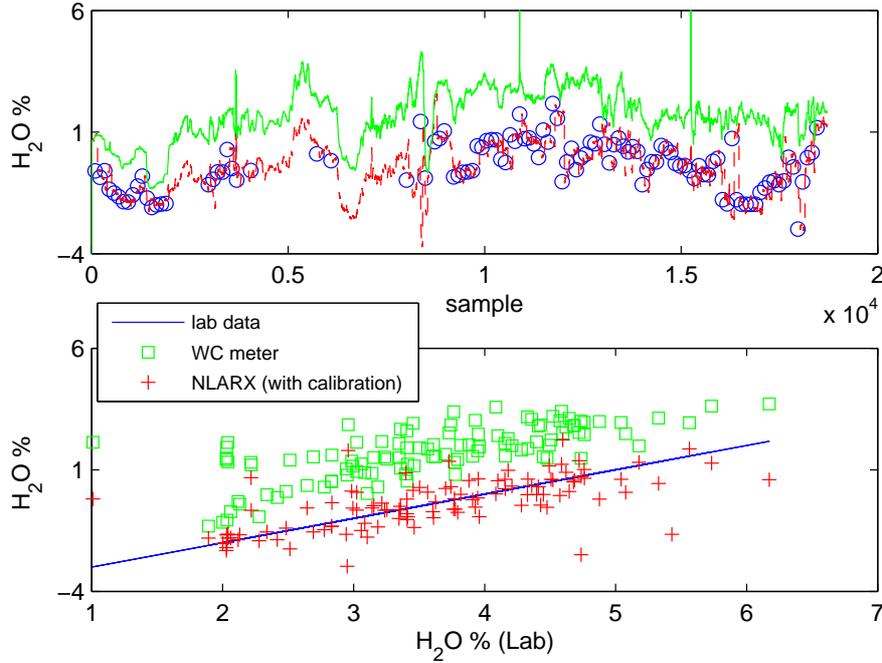


Figure 5.17: NLARX estimate with model calibration for new data set collected in July 2009.

ance is proposed. A sequential Monte Carlo sampling based particle filter is then applied to carry out the Bayesian model calibration strategy. Compared to other approaches, the nature of sample based representation of PF facilitates the handling of constrained non-linear and non-Gaussian estimation problems. The proposed approach is used for a data-driven soft sensor development, which has been successfully demonstrated for water-content monitoring in an oil sands plant.

Bibliography

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.
- Bai, E., 1998. An optimal two-stage identification algorithm for hammerstein-wiener nonlinear systems. *Automatica* 34(3), 333–338.
- Bishop, C. M., 1995. *Neural networks for pattern recognition*. USA: Oxford University Press.
- Braun, J., 2000. Dempster-Shafer theory and Bayesian reasoning in multisensor data fusion. In: *Sensor Fusion: Architectures, Algorithms and Applications IV; Proceedings of SPIE 4051*.
- Chen, L., Nguang, S., Li, X., Chen, X., 2004. Soft sensors for on-line biomass measurements. *Bioprocess and Biosystems Engineering* 26(3), 191–195.
- Chen, R., Tsay, R., 1993. Nonlinear additive arx models. *Journal of American Statistical Association* 88, 955–967.
- Dayal, B., MacGregor, J., 1997. Improved PLS algorithms. *Journal of Chemometrics* 11, 73–85.
- de Assis, A., Filho, R., 2000. Soft sensors development for on-line bioreactor state estimation. *Computers and Chemical Engineering* 24(2), 1099–1103.
- Ding, F., Chen, T., 2004. Identification of dual-rate systems based on finite impulse response model. *International Journal of Adaptive Control and Signal Process* 18, 589–598.

- Doucet, A., Godsill, S., 1998. On sequential simulation-based methods for Bayesian filtering. Tech. rep., Department of Engineering, Cambridge university.
- Fortuna, L., Graziani, S., Xibilia, M., 2005. Virtual instruments in refineries. *IEEE Instrumentation and Measurement Magazine* 8, 26–34.
- Friedman, Y., Neto, E., Porfirio, C., 2002. First principles distillation inference models for product quality prediction. *Hydrocarbon Process* 81(2), 53–58.
- Gan, Q., Harris, C., 2001. Comparison of two measurement fusion methods for kalman-filter-based multisensor data fusion. *IEEE Transactions on Aerospace and Electronic Systems* 37(1), 273–279.
- Grantham, S., Ungar, L., 1990. A first principles approach to automated troubleshooting of chemical plants. *Computers and Chemical Engineering* 14, 783–798.
- Gudi, R., Shah, S., Gray, M., 1995. Adaptive multirate state and parameter estimation strategies with application to a bioreactor. *AIChE Journal* 41, 2451.
- Hua, G., Wu, Y., 2006. Measurement integration under inconsistency for robust tracking. In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Huang, B., 2008. Bayesian methods for control loop monitoring and diagnosis. *Journal of Process Control* 10(9), 829–838.
- Kadleca, P., Gabrys, B., Strandtb, S., 2009. Data-driven soft sensors in the process industry. *Computers and Chemical Engineering* 33, 795–814.
- Kano, M., Nakagawa, Y., 2008. Data-based process monitoring, process control, and quality improvement: recent developments and applications in steel industry. *Computers and Chemical Engineering* 32, 12–24.

- Kewley, D., 1992. Notes on the use of Dempster-Shafer and fuzzy reasoning to fuse identity attribute data. Tech. rep., Defence Science and Technology Organisation, Adelaide.
- Khatibisepehr, S., Huang, B., 2008. Dealing with irregular data in soft sensors: Bayesian method and comparative study. *Industrial & Engineering Chemistry Research* 47, 8713–8723.
- Koks, D., Challa, S., 2003. An introduction to Bayesian and Dempster-Shafer data fusion. Tech. rep., Defence Science and Technology Organisation, Australian Government.
- Kresta, J., Marlin, T., MacGregor, J., 1994. Development of inferential process models using PLS. *Computers and Chemical Engineering* 18(7), 597–611.
- Li, D., Shah, S., Chen, T., 2001. Identification of fast-rate models from multi-rate data. *International Journal of Control* 74, 680–689.
- Ljung, L., 1999. *System identification: theory for the user*, 2nd ed. T. Kailath, Ed. Englewood Cliffs, Prentice Hall.
- Lu, N., Yang, Y., Guao, F., Wang, W., 2004. Multirate dynamic inferential modeling for multivariable processes. *Chemical Engineering and Science* 59, 855–864.
- Lu, W., Fisher, D., 1988. Output estimation with multi-rate sampling. *International Journal of Control* 48(1), 149–160.
- Lu, W., Fisher, D., 1989. Least-squares output estimation with multirate sampling. *IEEE Transactions on Automatic Control* 34(6), 669–672.
- MacGregor, J., 2004. Data-based latent variable methods for process analysis, monitoring and control. *Computer Aided Chemical Engineering* 18, 87–98.

- Mo, S., Chen, X., Zhao, J., Qian, J., Shao, Z., 2009. A two-stage method for identification of dual-rate systems with fast input and very slow output. *Industrial & Engineering Chemistry Research* 48, 1980–1988.
- Mu, S., Zeng, Y., Liu, R., Wu, P., Su, H., Chu, J., 2006. Online dual updating with recursive PLS model and its application in predicting crystal size of purified terephthalic acid (PTA) process. *Journal of Process Control* 16, 557–566.
- Nagai, E., Arruda, L., 2005. Soft sensor based on fuzzy model identification. In: 16th IFAC World Congress.
- Nelson, P., Taylor, P., MacGregor, J., 1996. Missing data methods in PCA and PLS: score calculations with incomplete observations. *Chemometrics and Intelligent Laboratory Systems* 35, 45–65.
- Punska, O., 1999. Bayesian approaches to multi-sensor data fusion. Master's thesis, Department of Engineering, University of Cambridge.
- Qin, S., Yue, H., Dunia, R., 1997. Self-validating inferential sensors with application to air emission monitoring. *Industrial & Engineering Chemistry Research* 36, 1675–1685.
- Raghavan, H., Tangirala, A., Gopaluni, R., Shah, S., 2006. Identification of chemical processes with irregular output sampling. *Control Engineering Practice* 14, 467–480.
- Åke Björck, 1996. *Numerical Methods for Least Squares Problems*. SIAM.
- Rao, M., Corbin, J., Wang, Q., 1993. Soft sensors for quality prediction in batch chemical pulping processes. In: In Proceedings of the IEEE international symposium on intelligent control.
- Saha, R., Chang, K., 1998. An efficient algorithm for multisensor track fusion. *IEEE Transactions on Aerospace and Electronic Systems* 43(1), 200–210.

- Scholz, E., 1984. Karl Fischer titration. Springer, Berlin.
- Shao, X., Huang, B., Lee, J., 2010. Constrained Bayesian state estimation - a comparative study and a new particle filter based approach. *Journal of Process Control* 20, 143–157.
- Singh, A., 1997. Modeling and model updating in the real-time optimization of gasoline blending. Master's thesis, University of Toronto.
- Tun, M., Lakshminarayanan, S., Emoto, G., 2008. Data selection and regression method and its application to softsensing using multirate industrial data. *Journal of Chemical Engineering of Japan* 41(5), 374–383.
- Wang, J., Chen, T., Huang, B., 2004. Multirate sampled-data systems: computing fast-rate models. *Journal of Process Control* 4(1), 79–88.
- Xiong, Y., Chen, W., Tsui, K., Apley, D., 2009. A better understanding of model updating strategies in validating engineering models. *Computer Methods in Applied Mechanics and Engineering* 198, 1327–1337.
- Yan, W., Shao, H., Wang, X., 2004. Soft sensing modeling based on support vector machine and Bayesian model selection. *Computers and Chemical Engineering* 28, 1489–1498.
- Zhu, Y., Telkamp, H., Wang, J., Fu, Q., 2009. System identification using slow and irregular output samples. *Journal of Process Control* 19, 58–67.

Chapter 6

Industrial Contribution: Estimation of Bitumen Froth Quality Using Bayesian Information Synthesis

¹ This chapter presents the design of soft sensors for estimation of bitumen froth quality in an oil sands froth transportation process. One of the most important quality indexes for bitumen froth is the water content. Due to the variation in oil sands composition and the nature of multi-phase process conditions, existing hardware sensors are not reliable enough to provide on-line accurate water content measurement. Laboratory analysis result is obtained off-line with large sampling interval and irregular time delay. Therefore, it is not sufficient for real-time monitoring and control. To overcome these limitations, Bayesian information synthesis approach is proposed to fuse all the existing information to produce more reliable and more accurate real-time froth quality information. This technique has been applied in Syncrude Canada Extraction operations; both monitoring and control performance illustrate the promising perspectives of the proposed approach.

1. A version of this chapter has been accepted for publication as “X. Shao, F. Xu, B. Huang, A. Espejo, Estimation of Bitumen Froth Quality Using Bayesian Information Synthesis: An Application to Froth Transportation Process, The Canadian Journal of Chemical Engineering, in press, 2012.”

6.1 Introduction

Crude oil is used for a diverse range of products, including fuels, plastics, solvents, waxes, lubricants, and dyes, among others, and is, therefore, a vital resource for many industries. As the worldwide demand for petroleum continues to grow, previously nonviable sources of oil are increasingly pursued. One such source is the Athabasca oil sands in northern Alberta in Canada. With 170 billion barrels of bitumen available using current technology, it represents the second largest known oil reserve in the world, and currently produces over 1.4 million barrels of oil per day (Government of Alberta, 2009). Syncrude Canada Ltd., one of the world's largest oil sands companies, has production capacity of 350,000 barrels per day of light, high-quality synthetic crude oil.

Oil sands are mixtures of quartz, clay, water, bitumen and accessory minerals. The bitumen is extracted from the oil sands raw material prior to being upgraded to synthetic oil. Syncrude operation mainly consists of surface mining, extraction, upgrading and utility facilities (Dougan and McDowell, 1997). Much of the technology used in the mining, upgrading and utility operations is common to many similar industries. However, extraction operation is quite unique, not all the physical and chemical mechanisms are fully understood (Kresta, 1997). Among many extraction processing steps, bitumen froth transportation is one of the most important units. Syncrude strives for innovative ways of froth transportation to maximize bitumen recovery rate and reduce unit cost. A recent innovation is the introduction of the so called natural froth lubricity (NFL) technology to ship froth by pipeline from Aurora site to its Mildred Lake processing facilities 35 kilometers away (Joseph et al., 1999).

Considering the large amount of bitumen froth being transported, there is a strong incentive to optimize, or even incrementally improve operational performance for the transportation process. At current production levels, an improvement in bitumen froth quality of 1% can result in million dollars of economic benefit while utilizing the same equipment and ore throughput. More-

over, improving froth quality (e.g., reducing water percentage) can further reduce unit cost and increase equipment life in downstream facilities. However, due to the lack of process monitoring capability, the dynamics of the NFL process are not known, and the operational performance is not optimized. One of the major challenges encountered is the lack of suitable hardware instruments specifically developed for oil sands processes as the market for such sensors is small and the requirements are fairly unique. Syncrude has been addressing this challenge by utilizing available technology intended for other applications where possible; adapting available technology where feasible; and developing novel technology where necessary (Espejo, 2011; Domlan et al., 2011).

In this chapter, a Bayesian method is utilized to synthesize all the related information from existing measurements, including secondary variables (e.g., density, flowrate, etc.) and primary variables (e.g., water content) from multiple observation sources (e.g., hardware sensor, laboratory analysis), to provide more reliable and more accurate real-time froth quality information. After verifying the monitoring performance, an inferential controller is proposed for maintaining the water content value within a desired range as per operation requirements.

The organization of this chapter is as follows: Section 2 provides a brief background description of the investigated process. The design of the soft sensor using Bayesian information synthesis approach is reported in Section 3. Soft sensor based water content monitoring and control results are demonstrated in Section 4. Section 5 gives conclusion.

6.2 Process Description

In the process of producing oil from oil sands, the main task is to separate bitumen present in the oil sands from the other components that are roughly solids and water. The separation is performed through a chain of industrial units, mainly consisting of primary separation vessel (PSV), floatation, froth

treatment unit, and solvent recovery unit. The quality of produced bitumen is determined by its purity and quantified typically by its bitumen content and water content.

6.2.1 Aurora Bitumen Froth Transportation

Syncrude separates bitumen froth from oil sands in both Mildred Lake (Base Plant) and Aurora sites. The bitumen froth from Aurora site is transported via a 35km froth pipeline to Base Plant for further processing. This transportation line is one of the most essential processes to Syncrude as more than 60% of bitumen froth is transported through this pipeline.

As part of the Aurora low energy extraction processes, instead of adding a diluent, such as naphtha, Syncrude developed a new technology, called Natural Froth Lubricity (NFL), for Aurora froth transportation, using the naturally formed sheath of water that forms a sleeve in the pipeline, allowing the relatively less viscous bitumen froth to be transported more easily.

Figure 6.1 shows a simplified flow chart for the Aurora froth pipeline. The system consists of two separate trains (known as Train 1 and Train 2), and is fed from three froth tanks (D-1/2/3) by two primary pumps (G-1/2), one for each train, and discharged by two sets of booster pumps; the pumps are stopped and started to maintain the levels L_1, L_2 and L_3 in the froth tanks, as well as maintaining the minimum critical flows for F_3^1 and F_3^2 in the froth pipeline; two pipelines combine prior to being shipped to Base Plant. Note that the rest nomenclatures in the figure will be introduced in the next section.

To maintain the froth temperature, hot water is added into froth pipeline prior to the primary pump. The efficiency of the NFL process is dependent on many factors, including the temperature of the froth, the quality of feed stream and the amount of hot water added. There is a strict requirement on the froth quality with respect to water content, namely, the in-line water content is not

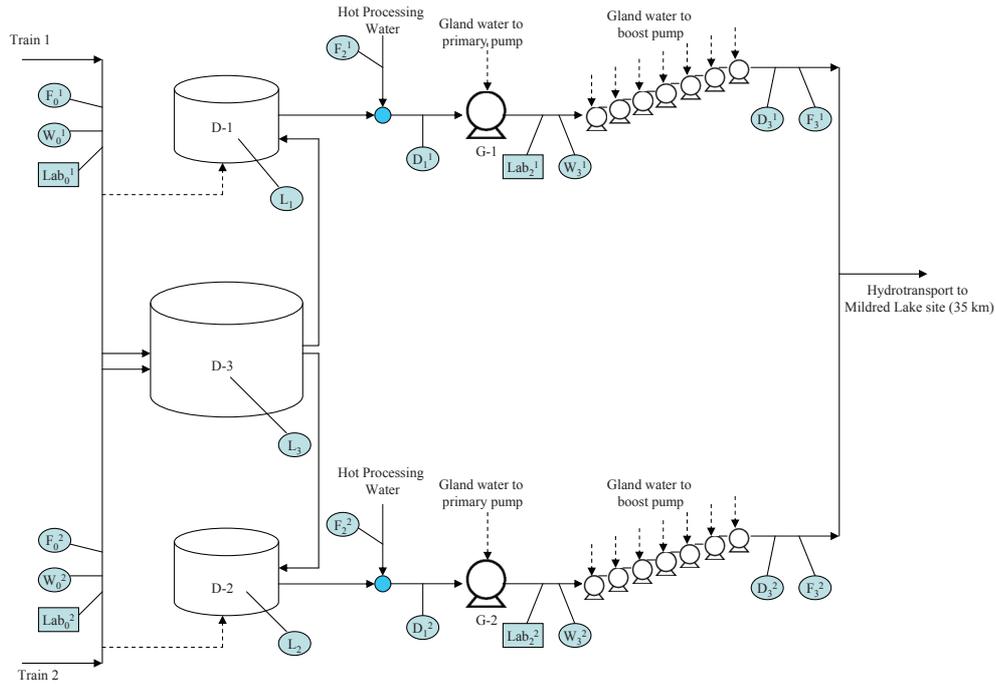


Figure 6.1: Simplified schematic of Aurora bitumen froth transportation pipeline.

allowed to be lower than its low-low specification to avoid restrictions of froth transportation, and should not be any higher than its high specification when it reaches the froth treatment plant. Otherwise, it has to be redirected to the primary separation vessels in Base Plant, which could cause additional bitumen loss and consequently result in reduction of bitumen recovery rate. In contrast, if the water content is too low (e.g., less than its low-low specification), it can cause pipeline plug, which leads to week-long outage and more serious financial loss.

Therefore, optimal control of froth water content through hot process water addition is extremely important for the NFL process operation, as it can increase pipeline uptime, improve froth quality, and reduce operating cost. To achieve this objective, the first and foremost issue is to obtain real-time, reliable, accurate and consistent water content information.

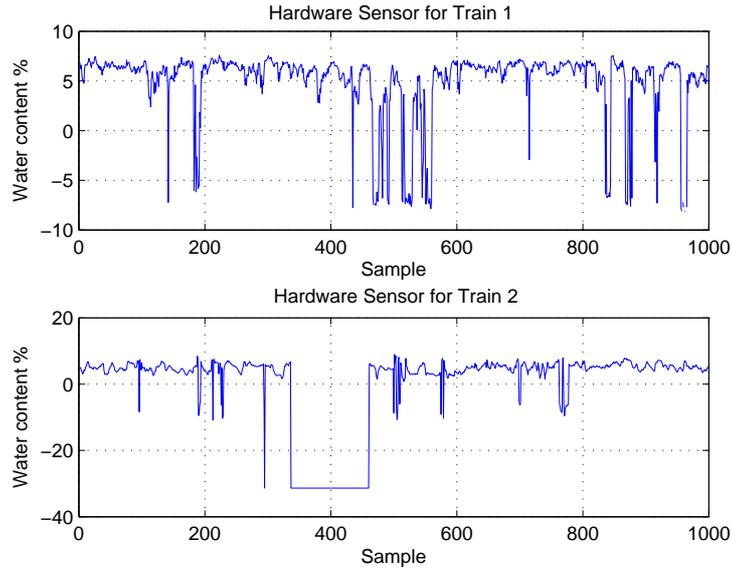


Figure 6.2: Hardware readings for water content measurements.

6.2.2 Existing Water Content Measurements

There are two water content hardware sensors installed on froth pipeline, one for each train. They were commissioned to monitor water content of the froth discharged by primary pumps, and configured to implement feedback control for the hot process water additions. However, the variation of the oil sands composition and the nature of multi-phase processing conditions create harsh environment for in-line instruments. Historical observation shows that these two meters have reliability issues, which could cause serious upsets if they are used for automatic hot process water control. Figure 6.2 shows historic readings from two meters, it can be seen that both meters are not reliable; in fact, sometimes the readings even drop to negative. Note that, for proprietary reason, the actual operating ranges have been modified/removed in the chapter.

Due to the importance of the water content information, lab data is available hourly from the Aurora unit lab. Froth samples are collected manually and put in a centrifuge machine to separate the bitumen, solid and water, and then the water content is calculated by visually reading the amount of

different components in a test tube. This procedure is carried out off-line, and usually takes one hour to complete. Furthermore, human error can be introduced during the lab analysis procedures. A preliminary test shows that different technicians can easily produce 5% error in reading the same sample. Nevertheless, unit lab result is still considered as the most trustful information to operators, and the hourly averaged lab data is being used in operations as the indication to manually adjust the setpoint for hot process water addition.

6.3 Soft Sensor Development

To achieve better monitoring performance, soft sensor technique (Chen et al., 2004; Khatibisepehr and Huang, 2008; Kadleca et al., 2009; Shao et al., 2011) is investigated. The froth transportation process appears to be an ideal candidate for the application of soft sensor technique for some of the following reasons:

- (i) The process is very dynamic; multi-phase mixtures of bitumen, coarse solids, fine solids, water and air that can exhibit time-dependent behaviors, wherein pipeline friction losses increase drastically with time;
- (ii) Froth compositions are complex as the oil sand deposits are naturally highly variable in bitumen and clay content. Due to the large volumes processed and the primary extraction techniques used, most of the oil sand variability is passed through to the bitumen froth. While the bitumen liberation mechanism can be very complex, the froth pipeline operation itself is quite simple with few controllable parameters;
- (iii) The outcomes of the NFL process are highly sensitive to the characteristics of the feedstock stream (e.g., density, water content, etc.) and the addition of hot dilution water;
- (iv) The major difficulty faced when attempting to better understand the dynamics of the process has been the lack of sensors capable of monitoring performance.

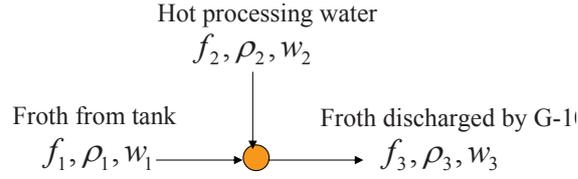


Figure 6.3: Hot process water addition in Aurora froth pipeline.

6.3.1 Variable Selection

Since the lab data is considered as the most trustful information source, it is selected as the output variable for soft sensor modeling. To choose closely related secondary variables as input variables, mass balance principle is used for process analysis. Take train 1 as an example, the investigated process can be simplified as shown in Figure 6.3.

From Figure 6.3, it can be seen that froth from storage tank has flowrate F_1 , density ρ_1 and water content W_1 ; it is diluted by hot process water with flowrate F_2 , density ρ_2 and water content W_2 , then discharged by a primary primer pump; the discharged froth has flowrate F_3 , density ρ_3 and water content W_3 . A mass balance equation can be obtained as,

$$F_1 \cdot \rho_1 \cdot W_1 + F_2 \cdot \rho_2 \cdot W_2 = F_3 \cdot \rho_3 \cdot W_3 \quad (6.1)$$

Therefore, the water content for the discharged froth is calculated as,

$$W_3 = \frac{F_1 \cdot \rho_1 \cdot W_1 + F_2 \cdot \rho_2 \cdot W_2}{F_3 \cdot \rho_3} \quad (6.2)$$

Unfortunately, in Equation (6.2), only F_2, ρ_2, W_2 and F_3 are known, ρ_3 is not directly measured, but can be approximately inferred from two existing density readings. Critical missing information includes F_1, ρ_1 and W_1 , which implies infeasibility of using the first principle model to estimate W_3 . Hence, the only solution is to use statistical approach to retrieve the missing information.

The following assumptions are made in the modeling analysis:

- (i) water content of the froth in different storage tanks are the same, and

Table 6.1: Selected secondary variables for froth line modeling

Input	Description
$F_0^1 + F_0^2$	Total froth flow to storage tanks
L_1, L_2, L_3	Tank volume based weighted level
F_2	Flowrate of hot process water
ρ_2	Density prior to primary pump
F_3	Pipeline discharge flowrate
ρ_3	Density of discharged froth
$\frac{1}{\tau} \sum_{\tau} W_0^1 + W_0^2$	Average water content in storage tanks

the value does not change significantly within one hour (considering the normal residence time τ is around 4 hours);

- (ii) the flowrate of gland water added to the primary pump and boost pump set is small enough to be neglected.

Based on the above assumptions, the following variables, as shown in Table 6.1, are selected as the input variables for soft sensor modeling.

6.3.2 Synthesis of Secondary Variables Using PCR

Synthesis of secondary variables is also known as process modeling, which is one of the key steps to achieve a successful soft sensor application. Depending on the studied process, the model of the soft sensor could be first-principle or data-driven, dynamic or static, linear or nonlinear, and the parameters are estimated using historical data.

Considering the collinearity among the selected variables, a latent variable technique, Principle Component Regression, (PCR) (Jolliffe, 1982), is chosen for data modeling. The calculation process is described as follows: First, the normalized input variable matrix $U(n \times m)$ is analyzed by using principle component analysis (PCA) approach (Jolliffe, 2002). The factor score matrix $T(n \times m)$ and loading matrix $L(m \times m)$ can be obtained as

$$U = T \cdot L^T \quad (6.3)$$

$$T = U \cdot W \quad (6.4)$$

Table 6.2: Error comparisons between PCR model and hardware sensor

	MAE	STD	RMSE
Hardware sensor	3.23	4.96	5.66
Model prediction	1.77	2.05	2.17

$W(m \times m)$ is the factor score coefficient matrix, where

$$W = (L^T)^+ = (L^T L)^{-1} L \quad (6.5)$$

Second, replace the input matrix U by factor score matrix T , and perform least squares regression with the normalized output vector over the factors,

$$Y = T \cdot \beta + e \quad (6.6)$$

To learn the PCR parameters, output and input data are collected from historical database; robust regression (Rousseeuw and Leroy, 2003) method is used to obtain the model parameters β .

As the factors are the combinations of the input variables, Equation (6.6) can be written as a direct regression model between input and output variables

$$\begin{aligned} Y &= U \cdot W \cdot \beta + e \\ &= U \cdot \Theta + e \end{aligned} \quad (6.7)$$

To validate the PCR model, a set of new data is collected and compared with the simulated result from the PCR model. The results are shown in Figures 6.4 and 6.5, from which we can see that the estimated model is able to capture water content dynamics in general.

Table 6.2 shows the performance comparisons in terms of mean absolute error (MAE), standard deviation (STD), and root mean square error (RMSE). It clearly shows that PCR model prediction overall outperforms existing hardware sensors.

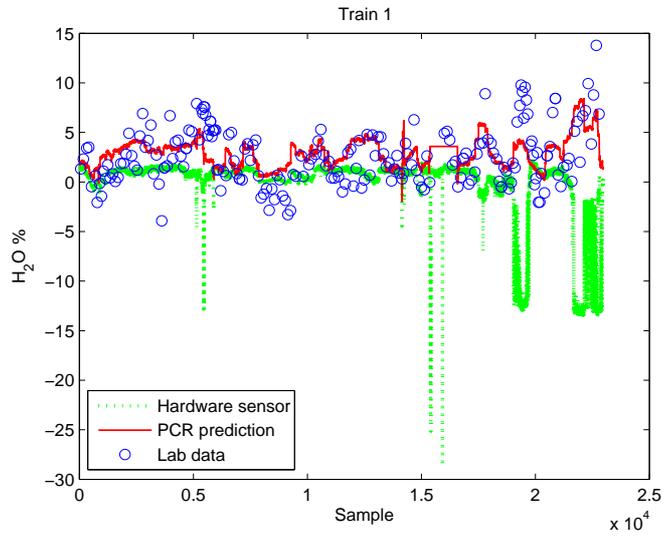


Figure 6.4: PCR model testing results (trends plot).

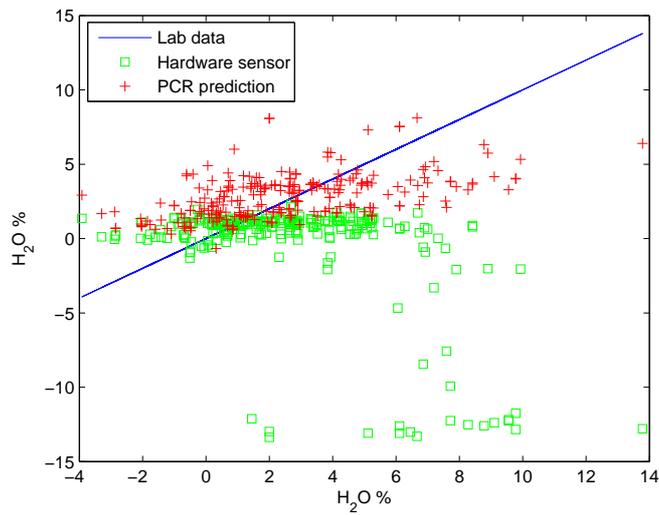


Figure 6.5: PCR model testing results (scatter plot).

6.3.3 Bayesian Model calibration

Equation (6.7) can also be represented in a state-space form as follows

$$\begin{aligned}x_k &= u_k \cdot \theta_{k-1} + \omega_k^x \\ \theta_k &= \theta_{k-1} + \omega_k^\theta \\ y_k &= x_k + \nu_k\end{aligned}\tag{6.8}$$

where x_k is the unknown true process output (i.e., noise-free water content) at time step k ; θ_{k-1} is the PCR model parameter; ω_k^x and ν_k are process noise and measurement noise, respectively; ω_k^θ is a random variable representing model parameter uncertainty.

Due to the modeling error or presence of process uncertainties (e.g., process drifting), model prediction can be deviated from the true output as time increases. To ensure the soft sensor performance, multiple observation sources of various sampling rates for primary variable are synthesized to update model parameters within the Bayesian framework (Shao et al., 2011). The objective is to construct *a posteriori* distribution of the unknown variable by recursively solving the following steps (Huang, 2008).

Prediction:

$$p(x_k, \theta_k | D_{k-1}) = \int p(x_k, \theta_k | x_{k-1}, \theta_{k-1}) p(x_{k-1}, \theta_{k-1} | D_{k-1}) dx_{k-1} d\theta_{k-1}.\tag{6.9}$$

Update:

$$\begin{aligned}p(x_k, \theta_k | D_k) &= \frac{p(Y_k | x_k, \theta_k) p(x_k, \theta_k | D_{k-1})}{p(Y_k | D_{k-1})} \\ &= \frac{p(y_k^1 | x_k, \theta_k) p(y_k^2 | x_k, \theta_k) \cdots p(y_k^{N_o} | x_k, \theta_k) p(x_k, \theta_k | D_{k-1})}{p(y_k^1, y_k^2, \cdots, y_k^{N_o} | D_{k-1})} \\ &\propto p(x_k, \theta_k | D_{k-1}) \prod_{n=1}^{N_o} p(y_k^n | x_k, \theta_k),\end{aligned}\tag{6.10}$$

where $p(x_k, \theta_k | x_{k-1}, \theta_{k-1})$ and $p(y_k^n | x_k, \theta_k)$ are the probabilistic forms of Equation (6.8); $D_k = \{\mathcal{Y}_1, \cdots, \mathcal{Y}_k\}$ represents all the observations up to time k ; $\mathcal{Y}_k = \{y_k^1, \cdots, y_k^{N_o}\}$ denotes the measurement set from N_o observation sources. In this chapter, N_o equals to 2, which indicates both water content hardware

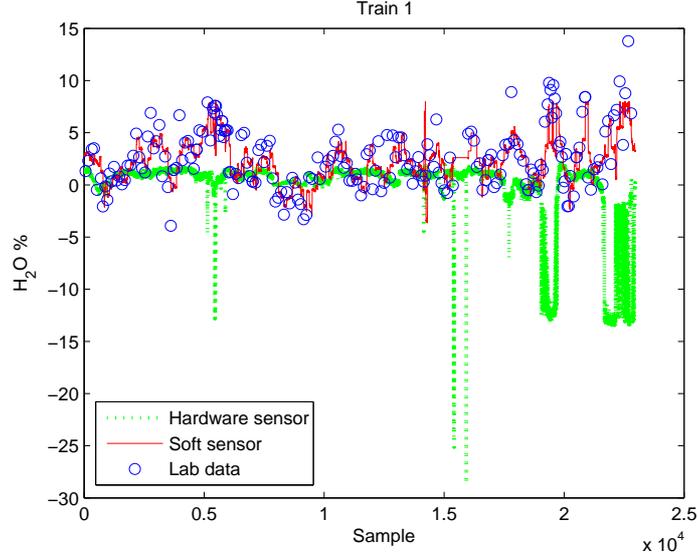


Figure 6.6: Soft sensor testing results (trends plot).

Table 6.3: Error comparisons between soft sensor and hardware sensor

	MAE	STD	RMSE
Hardware sensor	3.23	4.96	5.66
Soft sensor	0.89	1.04	1.09

sensor reading and lab analysis data are synthesized. In the implementation, initial guess of x_0 was obtained based on mean value of historical data, and θ_0 is the preidentified PCR model parameter.

Considering the nonlinear and non-Gaussian nature for the investigated process, a sequential Monte Carlo sampling based particle filter (PF) (Gordon et al., 1993) is used for Bayesian model calibration. Readers can refer to Shao et al. (2011) for more details about the calibration strategy, while in this section, only main results are presented. From Figures 6.6 and 6.7, as well as Table 6.3, it can be clearly seen that the performance of soft sensor has been further improved due to the combination of additional measurement information.

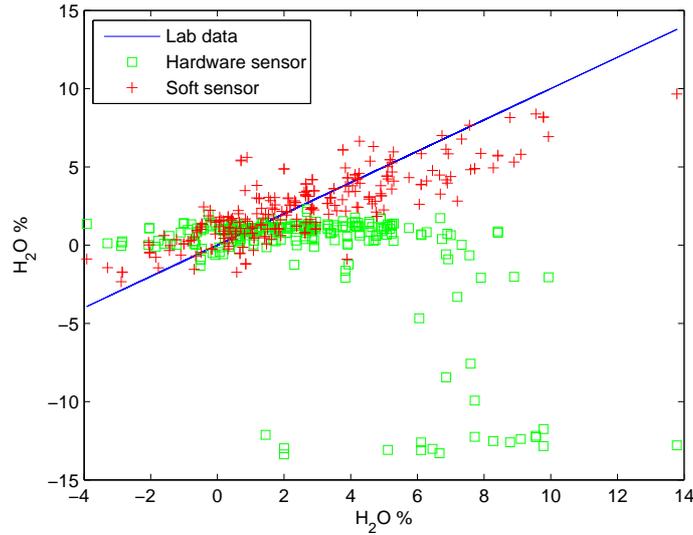


Figure 6.7: Soft sensor testing results (scatter plot).

6.4 Soft Sensor Performance Assessment

To assess the soft sensor estimation performance, some tests are carried out and described below.

6.4.1 Preliminary Step Test

Preliminary step test was first conducted on the variables (e.g., hot process water flow) identified having direct influences on the soft sensor model outputs. Figure 6.8 shows OSI PI readings of the online step test results, from which it can be concluded that: (i) soft sensor and hardware water content sensor give the same trend when both work reliably; (ii) hardware sensor gives abnormal reading when increasing hot process water flow to a certain amount, while soft sensor is able to work reliably and captures the operating condition changes; (iii) hardware sensor could give abnormal reading (e.g., negative value) without any obvious reasons.

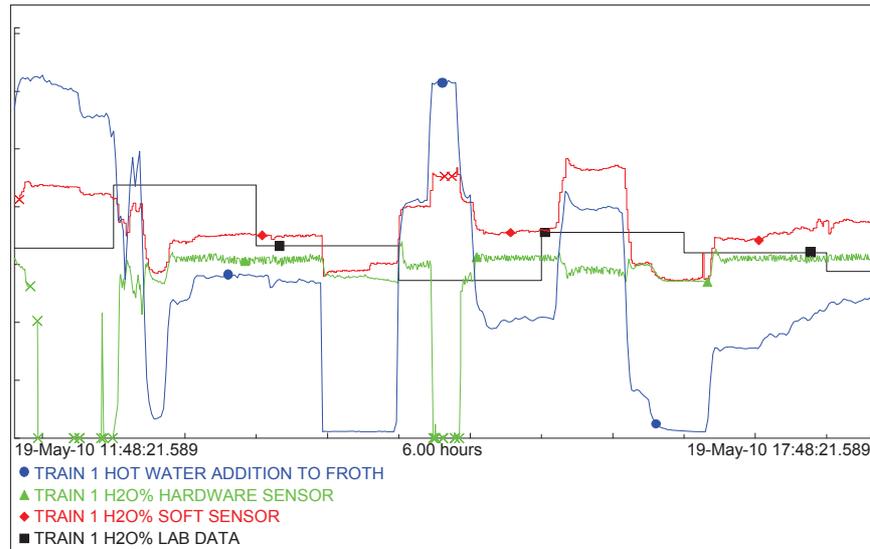


Figure 6.8: Online hot water flowrate step test for soft sensor model validation.

6.4.2 Performance Assessment Using Lab Data

To further assess soft sensor performance, both Aurora unit lab result (using centrifuge machine separation approach) and Base Plant main lab results (using Nuclear Magnetic Resonance analyzer, NMR) are used to compare with soft sensor model prediction. Figure 6.9 shows the test result, from which we can see that the trend of the soft sensor model output is consistent with Aurora unit lab result as well as Base Plant NMR result, and soft sensor model output has less variation than Aurora unit lab data. The values of soft sensor model output are generally located within the NMR upper and lower bounds, except for the points with extremely high water content (this mismatch is expected to be compensated by the model calibration strategy).

Figure 6.10 shows the soft sensor on-line implementation results, from which we can see that soft sensor estimate is reliable and accurate in comparison with Aurora unit lab data and much better than hardware water content sensors.

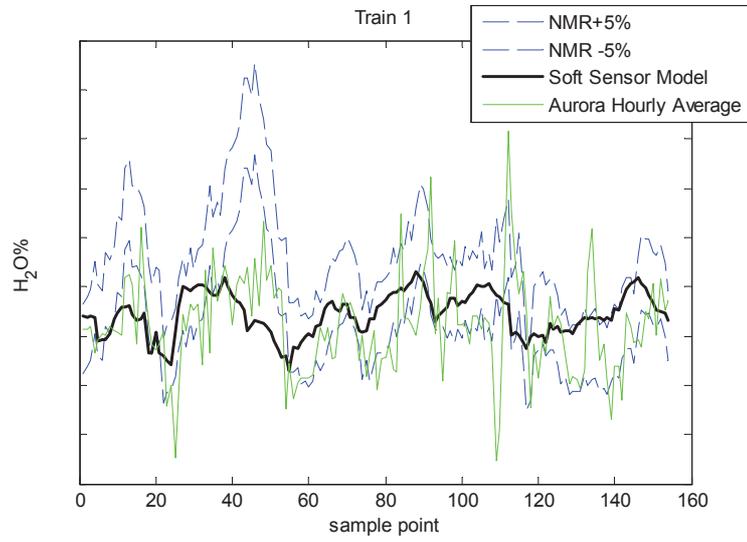


Figure 6.9: Comparisons of soft sensor model, unit lab data with NMR lab results.

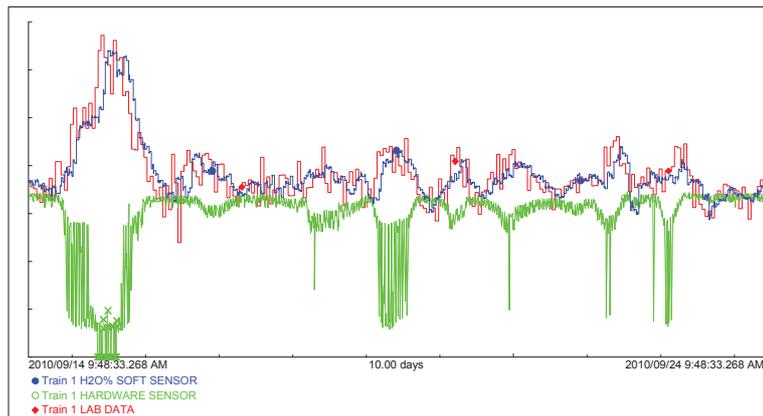


Figure 6.10: Soft sensor online implementation performance.

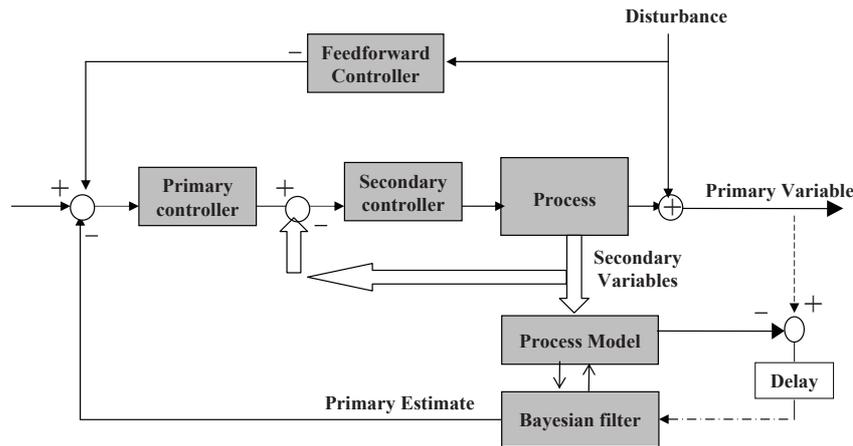


Figure 6.11: Inferential control for water content.

6.4.3 Soft Sensor Based Water Content Control

To take the full advantages of the developed soft sensor, an inferential control strategy is proposed in this section to control the water content within a desired range. Soft sensor estimate is chosen as the primary control variable (CV), hot process water is chosen as the manipulated variable (MV), froth flow and density are chosen as the disturbance variables (DVs). A feedforward plus feedback cascade control loop is designed as shown in Figure 6.11. In the implementation, proportional-integral (PI) type controllers were used for both inner and outer loops with sampling rate of 1 second and 30 seconds, respectively. Furthermore, in order to improve stability, a gap option was practically configured for the primary controller (i.e., water content controller) to achieve range control philosophy. By doing this, the setpoint values for secondary controller (i.e., hot water addition) will remain unchanged if the primary CV (i.e., water content) stays within the desired range. The result of on-line implementation in the actual process is shown in Figure 6.12. Based on the data analysis, we noticed that water content off-spec time has been reduced by 17.3% after the implementation of soft sensor and inferential control, and the quality variable (QV) variation has been reduced from 1.492 to 0.802.

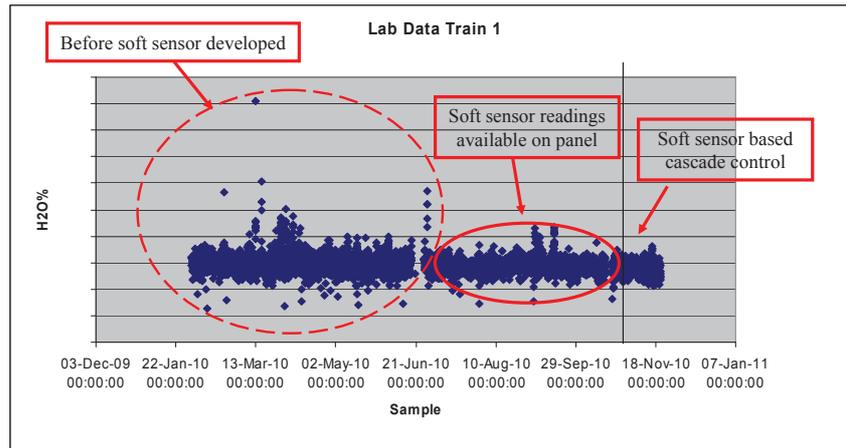


Figure 6.12: Inferential control performance.

6.5 Conclusion

A Bayesian information synthesis approach is proposed to develop soft sensors for the estimation of froth quality in oil sands bitumen froth transportation process. The approach synthesizes all of the existing information to produce more reliable and more accurate estimation. With the implementation of Bayesian model calibration, the developed soft sensor is sufficient for closed-loop control. An inferential control strategy is designed and tested for online froth quality control and the results obtained from the industrial application show effectiveness of the developed soft sensor.

Bibliography

- Chen, L., Nguang, S., Li, X., Chen, X., 2004. Soft sensors for on-line biomass measurements. *Bioprocess and Biosystems Engineering* 26(3), 191–195.
- Domlan, E., Huang, B., Xu, F., Espejo, A., 2011. A decoupled multiple model approach for soft sensor design. *Control Engineering Practice* 19:2, 126–134.
- Dougan, P., McDowell, K., 1997. Sensor development in oil sand processing. In: *Proceeding of 1997 Dynamic Modeling Control Applications for Industry*.
- Espejo, A., 2011. Managing & leveraging a large control system. In: *International Symposium on Advanced Control of Industrial Processes*.
- Gordon, N., Salmond, D., Smith, A., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: *IEE Proceedings F Radar and Signal Processing*.
- Government of Alberta, 2009. Alberta energy: Oil sands. In: www.energy.gov.ab.ca/OurBusiness/oilsands.asp. Government of Alberta.
- Huang, B., 2008. Bayesian methods for control loop monitoring and diagnosis. *Journal of Process Control* 10:9, 829–838.
- Jolliffe, I., 2002. *Principal Component Analysis*. Springer-Verlag.
- Jolliffe, I. T., 1982. A note on the use of principal components in regression. *Journal of the Royal Statistical Society* 31:3, 300–303.
- Joseph, D., Bai, R., Mata, C., Sury, K., Grant, C., 1999. Self-lubricated transport of bitumen froth. *Journal of Fluid Mechanics* 386, 127–148.

- Kadleca, P., Gabrys, B., Strandtb, S., 2009. Data-driven soft sensors in the process industry. *Computers and Chemical Engineering* 33, 795–814.
- Khatibisepehr, S., Huang, B., 2008. Dealing with irregular data in soft sensors: Bayesian method and comparative study. *Ind. Eng. Chem. Res.* 47, 8713–8723.
- Kresta, J., 1997. Advanced process control of extraction : Sensors and models. In: International heavy oil symposium.
- Rousseeuw, P., Leroy, A., 2003. Robust regression and outlier detection. Wiley-IEEE.
- Shao, X., Huang, B., Lee, J., Xu, F., Espejo, A., 2011. Bayesian method for multirate data synthesis and model calibration. *AIChE Journal* 57:6, 1514–1525.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this dissertation, particle filter (PF) is investigated for solving nonlinear state estimation problems. The PF approach is based on a rigorous Bayesian formulation and uses sequential Monte Carlo (SMC) sampling technique to propagate all information recursively. As opposed to other Bayesian estimators, PF do not rely on common assumptions of Gaussian or fixed-shape distributions; therefore it is more suitable to handle nonlinear and non-Gaussian estimation problems.

Applications of PF to practical chemical engineering processes however are restrained by (i) complicated process constraint, (ii) unknown but bounded uncertainty, (iii) imperfect model, (iv) multirate and possibly abnormal observations, etc. This research addresses practical issues and applies the PF to soft sensor developments in oil sands Extraction processes.

The following items summarize the main results of this thesis:

- (i) Chapter 2 reviews both optimal and sub-optimal Bayesian algorithms for nonlinear state estimation problems, with a focus on the state-of-the-art particle filtering approach. Illustrative examples show that PF outperforms many commonly used estimation approaches, including EKF, UKF, MHE, and it has a good potential for real applications in complex chemical engineering processes.

- (ii) Proper use of constraint knowledge is critical for the successful implementation of Bayesian estimators. In Chapter 3, two different constraint handling strategies are discussed under the generic PF framework. Several new constrained PF algorithms are implemented based on hybrid use of acceptance/rejection and optimization schemes. Three case studies demonstrate the efficacy of the proposed approaches in complicated constraint handling.
- (iii) Chapter 4 presents a robust PF algorithm that is applicable to where the description of uncertainty, due to modeling error or measuring noise, is unknown but bounded. A robust solution has been obtained for nonlinear uncertain systems based on Monte Carlo sampling and nonlinear set membership approach.
- (iv) A novel application of particle filter is presented in Chapter 5 for data-driven model calibration using multiple-source observations. The approach is built within a PF framework to synthesize fast sampled but low accurate observations with high accurate but slow sampled observations to obtain more accurate process information. To enhance the robustness in the presence of abnormal data, a robust Bayesian fusion formulation with time-varying observation noise variance is proposed. Simulation study and industrial application demonstrate that PF can provide improved estimation by fusing multirate observations.
- (v) Chapter 6 introduces a PF based approach to develop industrial soft sensors, with a focus on froth quality estimation in oil sands froth transportation process. The approach synthesizes all of the existing information to produce more reliable and more accurate quality variable estimation. Furthermore, an inferential control strategy has been designed based on the soft sensor estimate, and online application results illustrate the promising potential of the PF approach.

7.2 Future Work

New and open research problems have been identified throughout this dissertation writing. These problems have potential theoretical and practical values to process control community, and hence are summarized as follows:

- (i) Improvement of robust particle filtering algorithms, including stability studies, estimation of the minimum number of samples required, and a practical formulation of particle filters with uniform convergence property.
- (ii) Further studies on the data fusion technique when the measurement noises are not independent. As discussed in Chapter 5, particle filter based data fusion is used with the assumption that noises of different measurements are independent of each other. In practice the measurement modes may be correlated, and the correlation information can potentially be used to improve the results.
- (iii) Online application of the developed estimation algorithms to more complex processes, including oil sands Upgrading processes. The developed PF estimation algorithm can be further tested on more complex processes to demonstrate the efficacy of the methods.
- (iv) Extending the research to multirate inferential control. Soft sensor based inferential control has been introduced in Chapter 6, but without extensive further development. Practical or theoretical study of PF approaches with closed-loop feedback control is challenging and needs extra attentions.

Appendix A

Constrained PFs based on Equations (3.9) and (3.10)

Since Equation (3.7) is the same with clipping, hereby only the constrained PFs based on Equation (3.9) and (3.10) are summarized as follows:

Algorithm 3: A novel constrained PF algorithm based on Equation (3.9)

step a. initialization: generate initial particles $\{x_0^i\}_{i=1}^N$ from *a priori* distribution $p(x_0)$, and set $k = 1$;

step b. importance sampling: generate prior particles, $\{x_k^{i,-}\}_{i=1}^N$, from importance sampling distribution $q(x_k|X_{k-1}^i, Y_k)$;

step c. weighting: calculate constrained likelihood and importance weights according to Equations (3.2) and (3.3), then normalize the weights as $w_k^i = \tilde{w}_k^i / \sum_{j=1}^N \tilde{w}_k^j$;

step d. resampling: if $N_{eff} \leq N_{thr}$, then generate posterior particles, $\{\tilde{x}_k^i\}_{i=1}^N$, based on resampling strategy, and set $w_k^i = 1/N$;

step e. Chi-square test: calculate the sample mean of the posterior particles, $\bar{x}_k = \frac{1}{N} \sum_{i=1}^N x_k^i$, and compute the output residual, $e_k = y_k - h(\bar{x}_k)$; test the Chi-square criteria with a preset Σ ;

step f. optimization: project the parent particles (i.e. the subset particles

selected for resampling) to new locations by solving Equation (3.9) if performance test in step e fails; recalculate the weights and resampling;

step g. output estimate the state by calculating $\hat{x}_k = 1/N \cdot \sum_{i=1}^N \tilde{x}_k^i$, set $k = k + 1$ and go back to step b.

Algorithm 4: A novel constrained PF algorithm based on Equation (3.10)

step a. initialization: generate initial particles $\{x_0^i\}_{i=1}^N$ from *a priori* distribution $p(x_0)$, and set $k = 1$;

step b. importance sampling: generate prior particles, $\{x_k^{i,-}\}_{i=1}^N$, from importance sampling distribution $q(x_k|X_{k-1}^i, Y_k)$;

step c. weighting: calculate constrained likelihood and importance weights according to Equations (3.2) and (3.3), then normalize the weights as $w_k^i = \tilde{w}_k^i / \sum_{j=1}^N \tilde{w}_k^j$;

step d. resampling: if $N_{eff} \leq N_{thr}$, then generate posterior particles, $\{\tilde{x}_k^i\}_{i=1}^N$, based on resampling strategy, and set $w_k^i = 1/N$;

step e. Chi-square test: calculate the sample mean of the posterior particles, $\hat{x}_k = \frac{1}{N} \sum_{i=1}^N x_k^i$, and compute the output residual, $e_k = y_k - h(\hat{x}_k)$; test the Chi-square criteria with a preset Σ ;

step f. optimization: calculate the projected mean, \tilde{x}_k , by solving Equation (3.10) if performance test in step e fails;

step g. output yield the projected mean as PF output; calculate state covariance, \tilde{P}_k , by using EKF method, and regenerate particles from a normal distribution $N(\tilde{x}_k, \tilde{P}_k)$; set $k = k + 1$ and go back to step b.